

Изучение влияния клинико-генетических факторов на течение дисциркуляторной энцефалопатии с использованием методов распознавания

Кузнецова А.В.¹, Костомарова И.В.²
Водолагина Н.Н.², Малыгина Н.А.², Сенько О.В.³

¹Институт биохимической физики им. Н.М. Эмануэля РАН, Москва, Россия

²Филиал ГОУ ВПО РГМУ Росздрава "НКЦ геронтологии", Москва, Россия

³Вычислительный центр им. А.А. Дородницына РАН, Москва, Россия

Аннотация. В статье представлены результаты исследований влияния клинических, биохимических показателей, а также полиморфизма генов, кодирующих ангиотензинпревращающий фермент (АПФ), липопротеинлипазу (ЛПЛ), белок-переносчик эфиров холестерина (СЕТР), на тяжесть течения дисциркуляторной энцефалопатии. При исследованиях использован метод интеллектуального анализа данных, основанный на построении оптимальных разбиений признакового пространства, а также различные методы распознавания. В ходе исследований была установлена статистически достоверная связь между компьютерной оценкой тяжести ДЭП, рассчитанной с помощью методов распознавания по набору клинических и биохимических параметров, и полиморфизмом генов АПФ и СЕТР.

Ключевые слова: дисциркуляторная энцефалопатия, полиморфизм генов, интеллектуальный анализ данных, распознавание.

ВВЕДЕНИЕ

Цереброваскулярная патология занимает второе место в ряду основных причин смертности и инвалидизации населения в экономически развитых странах, что определяет ее как одну из важнейших медицинских и социальных проблем [1, 2]. Смертность от цереброваскулярных заболеваний в России – одна из самых высоких в мире: в 2000 г. стандартизированный показатель составил 319,8 на 100 000 населения [3].

В нашей стране неуклонно растет число пациентов с явлениями хронической ишемии головного мозга, составляя не менее 700 на 100 000 населения [1, 2].

По данным Научного центра неврологии РАМН дисциркуляторная энцефалопатия (ДЭП) занимает одно из ведущих мест, в структуре всех сосудистых поражений головного мозга у лиц старше 60 лет [4]. ДЭП – одна из форм хронической цереброваскулярной недостаточности с прогредиентным течением – прогрессирующее диффузное поражение головного мозга, обусловленное нарастающим ухудшением кровоснабжения мозговой ткани. По данным различных авторов, когнитивные нарушения сосудистого генеза выявляются у 5–22% пожилых лиц.

К числу основных факторов риска ДЭП относятся: артериальная гипертония с резкими колебаниями артериального давления и нарушениями нормального

суточного ритма АД (отсутствие или недостаточное снижение АД в ночное время; на поздних стадиях иногда, наоборот, резкое снижение АД ночью); атеросклероз магистральных артерий головы, стеноз сонных артерий, патология сердца, сахарный диабет, нарушения липидного метаболизма, курение, гиперагрегация тромбоцитов, повышение вязкости крови и гематокрита, и др. В большинстве случаев сосудистое поражение головного мозга является мультифакториальным, обусловленным сочетанием нескольких факторов риска [5].

Однако не у всех больных с ДЭП выявляются так называемые классические факторы риска данного заболевания. Многие из пациентов, имея по стандартным критериям общеизвестных факторов риска низкую вероятность развития ДЭП, часто оказываются подвержены крайне неблагоприятному течению болезни, переносят инсульт неоднократно или погибают от его осложнений. В связи с этим становится очевидной необходимость поиска новых маркёров, которые позволяли бы выделять группы пациентов с высоким риском неблагоприятного прогрессирующего течения ДЭП.

В настоящее время все больше внимания уделяется изучению генетических факторов развития дисциркуляторной энцефалопатии. Изучается значение ряда генов ренин-ангиотензиновой системы (АПФ, АТГ), генов, регулирующих метаболизм липидов (ЛПЛ, СЕТР, ароА-I, ароА-II, ароЕ), систему гемостаза и др.

Особое внимание в генезе инсульта уделяется роли полиморфизма гена ангиотензинпревращающего фермента (АПФ). Известно не менее 12-ти полиморфизмов, три из которых находятся в кодирующей последовательности [6]. Ген АПФ картирован на хромосоме 17q23. Наиболее изучен I/D (insertion/deletion) полиморфизм гена АПФ. Под полиморфизмом понимают наличие в генофонде популяции нескольких аллелей какого-либо гена. Аллели – это сохранившиеся в популяции варианты одного гена, возникшие в результате генных мутаций и отличающиеся друг от друга последовательностью нуклеотидов. Данный полиморфизм дуаллельный и основан на наличии (аллель I) или отсутствии (аллель D) в 16 интроне гена АПФ вставки из 287 пар нуклеотидов (н.п.) [7].

Доказано, что полиморфизм типа вставка/отсутствие (инсерция/делеция) в гене АПФ коррелирует с уровнем АПФ в крови и повышенным риском ишемического инсульта (ИИ). Установлено, что у пациентов с генотипом DD гена АПФ достоверно чаще встречались стенозы магистральных артерий головы [8]. При этом носительство аллеля D положительно коррелирует с толщиной сосудистой стенки сонных артерий. Однако отсутствие различий в распределении полиморфных вариантов гена АПФ у больных с ИИ и длительной хронической ишемической болезнью мозга (без формирования инфарктов), вероятно, свидетельствует об относительно независимом характере процесса атеротромбогенеза [8].

Следует подчеркнуть, что мутация в данном гене предрасполагает к развитию гипертонии, и в отдельных популяциях – к ИИ, сахарному диабету II типа. Имеющиеся данные о связи артериальной гипертонии с I/D полиморфизмом противоречивы. Продемонстрировано существенное повышение частоты генотипа DD гена АПФ и снижение частоты генотипа II у больных артериальной гипертонией в ряде европейских и азиатских популяций. Однако в значительном числе работ предположение о возможной связи I/D полиморфизма гена АПФ с артериальной гипертонией не подтверждается [9].

Липопротеинлипаза (ЛПЛ) играет ключевую роль в метаболизме триглицериднасыщенных липопротеинов, она гидролизует хиломикроны, ХС ЛПОНП и ХС ЛПНП, тем самым влияя на уровень липопротеидов высокой плотности (ЛПВП) в плазме крови. Дефекты ЛПЛ приводят к

гипертриглицеридемии и гиперлипидемии, которые ассоциируют с ранним проявлением атеросклероза и сердечно-сосудистых заболеваний [10].

Ген ЛПЛ локализован на хромосоме 8p22, длиной 35 тысяч нуклеотидных пар, содержит 10 различных экзонов [11]. Описано более 60 мутаций гена ЛПЛ, приводящих к образованию слабоактивных вариантов фермента, и несколько его полиморфных вариантов [12], одним из которых является HindIII полиморфизм. HindIII полиморфизм гена ЛПЛ связан с наличием (H⁺) или отсутствием (H⁻) сайта рестрикции для рестриктазы HindIII в интроне 8. Поскольку полиморфизм локализован в интроне, то маловероятно, что он оказывает прямое влияние на структуру или количество белка, но, тем не менее, он связан с уровнем триглицеридов и ЛПВП в плазме крови, что позволяет считать его функциональным.

Многочисленными исследованиями показана связь аллеля H⁺ и генотипа H⁺H⁺ со снижением уровня ЛПВП и гипертриглицеридемией [10]. Последняя ассоциирована с гиперкоагуляцией и снижением фибринолиза, что увеличивает риск развития тромбозов. Выявлена связь HindIII полиморфизма с атеросклерозом. G.C. Chamberlain [13], R.K. Mattu [14] и другие исследователи обнаружили связь аллеля H⁺ с наибольшей степенью тяжести атеросклероза

Итальянскими исследователями показано, что носительство гомозиготного генотипа H⁺H⁺ увеличивает риск множественного атеросклеротического поражения сосудов [15]. В российской популяции также были выявлены связи генотипа H⁺H⁺ гена ЛПЛ с сердечно-сосудистыми заболеваниями [16].

Белок-переносчик эфиров холестерина (СЕТР – cholesteryl ester transfer protein) участвует в метаболизме ЛПВП, он переносит эфиры холестерина к триглицерид-богатым липопротеинам. Ингибирование активности СЕТР ведет к повышению концентрации ХС-ЛПВП, нормализации их транспортных свойств и, как следствие, к благоприятным сдвигам в липидном спектре плазмы.

Ген СЕТР локализован в хромосоме 16q21, длиной 25 кб, содержит 16 экзонов. Полиморфизм гена СЕТР по сайту рестрикции Taq IB рестриктазы влияет на концентрацию СЕТР и ЛПВП. Показано снижение прогрессирования атеросклероза у носителей генотипа В1В1, но не у носителей генотипа В2В2. Генотип В2В2 значительно чаще определяется у долгожителей (> 90 лет), что позволяет оценить его вклад в позитивный прогноз жизни человека [16].

Обращает на себя внимание, что в большинстве случаев связь генетической компоненты с характером течения цереброваскулярной патологии изучается только на примере инсульта.

Цель работы – изучение влияния клинических факторов и полиморфных вариантов генов ангиотензинпревращающего фермента, липопротеинлипазы и белка-переносчика эфиров холестерина на течение дисциркуляторной энцефалопатии у больных старших возрастных групп.

В настоящем исследовании используются методы распознавания и интеллектуального анализа данных [17–19]. Преимуществом методов распознавания является то, что они предоставляют объективную оценку точности диагноза или прогноза путем проверки в режиме скользящего контроля, что дает возможность повысить надежность диагностики.

МАТЕРИАЛЫ И МЕТОДЫ

1. База данных

База данных содержала информацию о 358 пациентах отделения неврологии "НКЦ геронтологии". 273 женщины и 85 мужчин в возрасте от 35 до 102 лет

(средний возраст 71,9). Больные были разделены на четыре возрастные группы. I группа – 107 человек (35-59 лет), II группа – 95 человек (60–75 лет), III группа – 33 человека (75–89 лет), IV группа – 123 человека (старше 90 лет).

Все пациенты были также разделены по тяжести течения ДЭП: I стадия – 47 человек, II стадия – 195 человек, III стадия – 116 человек.

Рассматривались анамнестические и клинико-лабораторные показатели, результаты инструментальных методов исследования (МРТ, ЭКГ, УЗИ и др.), а также молекулярно-генетического типирования генов ангиотензинпревращающего фермента (I/D полиморфизм), липопротеинлипазы (Hind III полиморфизм) и белка-переносчика эфиров холестерина (Taq I полиморфизм), полученные в лаборатории возрастной популяционной иммуногенетики "НКЦ геронтологии". Исследование липидного метаболизма проводилось в лаборатории липидного обмена "НКЦ геронтологии" Росздрава. В общей сумме число показателей в базе данных достигало 354.

2. Методы анализа данных

Для выявления зависимости тяжести течения ДЭП от клинико-генетических показателей использовали методы интеллектуального анализа данных, включающие набор методов распознавания, разработанный в Вычислительном центре им. А.А. Дородницына РАН и в Институте биохимической физики им. Н.М. Эмануэля РАН [18, 19].

В ходе анализа данных использовали на некоторых этапах стандартные статистические тесты Хи-квадрат (χ^2) и Уилкоксона–Манна–Уитни (WMW), реализованные в программе «Статистика» (StatSoft).

Основным инструментом в рамках данного исследования служили методы распознавания образов (система «РАСПОЗНАВАНИЕ») [20] и метод оптимально достоверных разбиений (ОДР) (см. Приложение 1) [17]. Методы распознавания позволяют решать задачи диагностики и/или производить прогноз с помощью алгоритма, заранее настроенного (обученного) по обучающей выборке случаев заболевания с известными значениями диагностических параметров и экспертной оценкой состояния пациента.

Система методов интеллектуального анализа данных «РАСПОЗНАВАНИЕ», включает набор методов распознавания и анализа данных, основанных на различных хорошо зарекомендовавших себя подходах:

- стандартные статистические методы, включая линейный дискриминант Фишера (ЛДФ),
- метод k ближайших соседей (k -NN);
- комбинаторно-логические методы;
- нейросетевые методы, включая стандартный многослойный перцептрон, а также мультипликативную нейронную сеть (МНС);
- методы, основанные на голосовании по логическим закономерностям (ЛЗ);
- метод, основанный на статистически взвешенном голосовании по системе синдромов (СВС);
- классифицирующие деревья;
- метод опорных векторов (МОВ);
- линейная разделяющая машина.

Система «РАСПОЗНАВАНИЕ» позволяет генерировать коллективные решения, при которых учитываются результаты распознавания сразу нескольких алгоритмов.

Метод интеллектуального анализа данных ОДР основан на построении разбиений областей допустимых значений объясняющих переменных,

индуцирующих разбиение исходной выборки на группы объектов с максимальными различиями уровней прогнозируемой величины. Верификация выявленных закономерностей производится с помощью перестановочного теста (см. Приложение 1). Метод позволяет достоверно оценивать совместное влияние на прогнозируемую величину сочетаний переменных. Метод реализован в рамках программной системы «РАЗБИЕНИЯ».

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

1. Исследование влияния различных показателей на тяжесть течения ДЭП с помощью метода ОДР

Исследование проводилось с использованием одномерных и двумерных моделей разбиений метода ОДР на выборке из 47 пациентов с I стадией и 43 пациентов с III стадией ДЭП старших возрастных групп, как наиболее полярных групп по клиническим проявлениям.

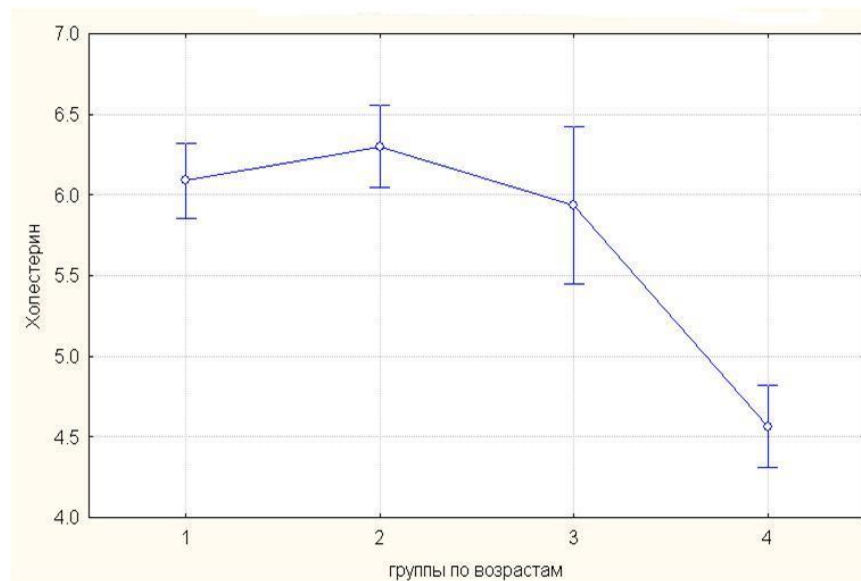


Рис. 1. Зависимость уровня холестерина (ммоль/л) в крови от возраста больных ДЭП.

Значения клинико-генетических показателей пациентов со II стадией (145 чел.) служили в качестве контрольной группы.

Группа долгожителей (IV) достоверно отличалась от трех первых по всем показателям. Учитывая существующие большие различия значений показателей у пациентов-долгожителей по сравнению с больными других возрастных групп, было решено проводить исследования связи тяжести течения ДЭП с клинико-биохимическими и генетическими показателями, исключив всех пациентов-долгожителей. В качестве иллюстрации можно показать зависимость содержания в крови холестерина от возраста (рис. 1). На рисунке хорошо видно, что для первых трех возрастных групп среднее значение уровня холестерина находится в районе ~ 6,0 ммоль/л, в то время как у долгожителей оно резко снижено (~ 4,5 ммоль/л).

Проанализировано влияние 113 клинических и биохимических параметров на тяжесть течения ДЭП, а также полиморфных вариантов генов АПФ, ЛПЛ, СЕТР. Метод ОДР позволил выявить большое число статистически достоверных закономерностей в рамках одномерных и двумерных моделей разбиений. Рассмотрим сначала одномерные модели.

1.1. Влияние клинических и биохимических факторов, выявленное в рамках одномерных моделей

На рисунке 2 приведен пример одномерной закономерности, отражающей связь уровня холестерина липопротеинов высокой плотности (ХС-ЛПВП) с тяжестью течения ДЭП. Граница разбиения по ЛПВП найдена компьютерным способом по методу ОДР таким образом, что ниже и выше нее преобладают значения преимущественно одной из исследуемых групп.

Рисунок 2 наглядно иллюстрирует различия в распределении значений уровня ХС-ЛПВП в группах пациентов с I и III стадиями ДЭП. Видно, что в область слева от границы попали 13 значений, соответствующих пациентам с III стадией заболевания, и 2 значения, соответствующие пациентам с I стадией ДЭП. Средний уровень ХС-ЛПВП слева от границы (0,35 ммоль/л) равен 0,23 ммоль/л (в этой области преобладают пациенты с III стадией ДЭП), в то время как в интервале от 0,35 до 2,4 ммоль/л средний уровень ХС-ЛПВП равен 1,13 ммоль/л (в данной области преобладают пациенты с III стадией ДЭП). Таким образом, общее количество пациентов с III стадией ДЭП слева от границы составляет 87%. Справа от границы – составляет 37%. Статистическая значимость выявленных различий между группами больных ($p = 0,006$) рассчитана с помощью перестановочного теста (см. Приложение 1).

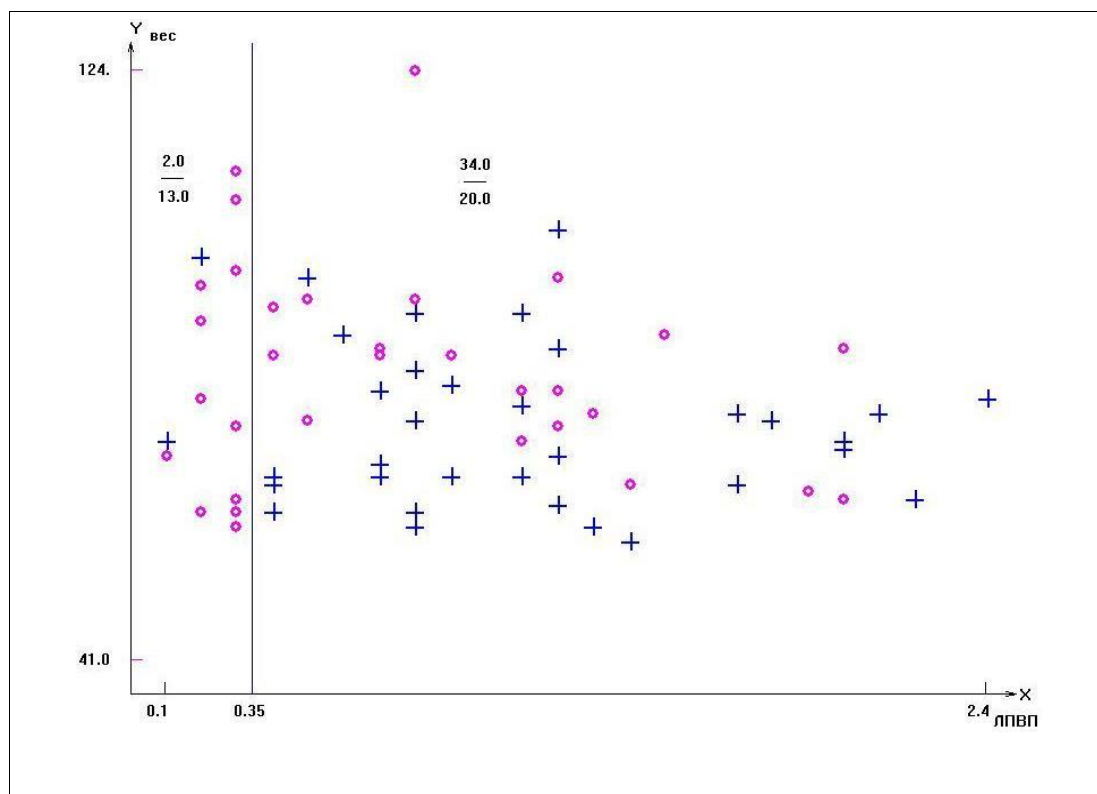


Рис. 2. Диаграмма рассеяния для уровня ХС-ЛПВП и веса пациентов (вспомогательная ось) (по осям X и Y, соответственно) у групп пациентов с I стадией ДЭП («+») и пациентов с III стадией ДЭП («o»). Значимость $p = 0,006$.

Следовательно, среди пациентов с более низким уровнем ХС-ЛПВП ($< 0,35$ ммоль/л) преобладают больные с III стадией ДЭП, что свидетельствует о наличии более выраженного атеросклеротического процесса у данных пациентов. Наши данные соответствуют результатам большого количества эпидемиологических исследований, где было показано, что низкий уровень ХС-ЛПВП в плазме считается

независимым фактором риска атеросклероза, а увеличение концентрации ХС-ЛПВП на 1 мг/дл ассоциируется со снижением относительного риска развития сердечно-сосудистых заболеваний на 2–3% [21, 22].

В таблице 1 приведены одномерные закономерности, связывающие клинические и биохимические показатели с тяжестью течения ДЭП и имеющие статистическую значимость не хуже, чем 0,01.

При использовании метода одномерных разбиений (таб. 1) выявлены наиболее информативные показатели, достоверно связанные с тяжестью течения заболевания. ОНМК и ПНМК – абсолютные признаки III стадии заболевания.

Таблица 1. Информативные клинические и биохимические показатели с достоверным различием при одномерных разделениях для групп с I и III стадиями ДЭП, выявленные методом ОДР. Показатели расположены по убыванию функционала качества (информативности) (описание – см. Приложение 1)

Клинико-лабораторные показатели (в скобках через наклонную черту приведены значения по данному признаку – ниже границы/выше границы)	Ниже границы (чел., %)		Граница	Выше границы (чел., %)		p ≤
	I ст. ДЭП	III ст. ДЭП		I ст. ДЭП	III ст. ДЭП	
1. Острое нарушение мозгового кровообращения (ОНМК) (нет/есть)	47 87%	7 13%	1,5	0 0%	36 100%	0,0001
2. Преходящие нарушения мозгового кровообращения (ПНМК) (нет/есть)	47 63%	28 37%	1,5	0 0%	15 100%	0,0001
3. Давление систолическое, мм рт. ст. (ниже/выше)	45 73%	17 27%	195	2 7%	26 93%	0,0002
4. Возраст (меньше/больше)	43 75%	18 25%	68,5	4 14%	25 86%	0,0002
5. Ишемическая болезнь сердца атеросклеротического генеза в анамнезе (нет/есть)	21 95%	1 5%	1,5	26 40%	42 60%	0,0002
6. Группы по возрасту (моложе 60 лет /старше)	32 74%	11 26%	1,5	15 32%	32 68%	0,0002
7. Гипертрофия левого желудочка по данным ЭКГ (нет/есть)	22 81%	5 19%	1,5	25 40%	38 60%	0,0002
8. Атеросклероз аорты (по данным рентгенологического исследования) (нет/есть)	14 87%	2 13%	1,5	14 32%	30 68%	0,0002
9. Сахарный диабет 2 типа (нет/есть)	46 59%	32 41%	1,5	1 8%	11 92%	0,001
10. Наличие гепатоза по данным УЗИ (нет/есть)	37 64%	21 36%	0,5	9 29%	22 71%	0,0032
11. Нарушение проводимости по данным ЭКГ (нет/есть)	34 65%	18 35%	1,5	13 34%	25 66%	0,004
12. Уровень липопротеидов высокой плотности (ЛПВП) (ммоль/л) (ниже/ выше)	2 14%	13 86%	0,35	34 63%	20 37%	0,005
13. Давление диастолическое (ниже/ выше)	39 73%	23 27%	105	8 37%	20 63%	0,0098

14. Инфаркт миокарда в анамнезе (нет/есть)	45 58%	33 42%	1,5	2 17%	10 23%	0,0134
15. Лейкоэроз по данным МРТ (нет/есть)	9 35%	17 65%	1,5	0 0%	14 100%	0,0188
16. Масса тела (кг) (меньше/больше)	35 65%	19 35%	77,5	12 33%	24 67%	0,0312
17. Гипертоническая болезнь в анамнезе (нет/есть)	11 79%	3 21%	1,5	36 47%	40 53%	0,0352
18. Давление диастолическое (выше 90 мм рт.ст./ниже)	21 42%	29 58%	1,5	26 65%	14 35%	0,0378
19. Стабильная стенокардия в анамнезе (нет/есть)	38 59%	26 41%	1,5	9 35%	17 65%	0,038
20. Атеросклероз сосудов по данным УЗДГ (нет/есть)	5 100%	0 0%	1,5	16 47%	18 53%	0,0446
Тенденция (0,05 < p < 0,1)						
21. Индекс массы тела (ИМТ) (ниже/выше)	47 56%	36 44%	34,45	0 0%	7 100%	0,053
22. АПФ генотип DD (отсутствует/присутствует)	34 59%	24 41%	1,5	7 26%	14 74%	0,0706 (0,047)
23. Трийодтиронин (ниже/выше)	23 85%	4 15%	24,4	2 33%	4 67%	0,0756
24. Глюкоза (ниже/выше)	47 56%	37 44%	10,25	0 0%	6 100%	0,0818

* последний столбец – достоверность p по перестановочному тесту – методом Монте-Карло (см. Приложение 1), в скобках – достоверность различия p по критерию χ^2 .

У больных с ДЭП III стадии достоверно выше частота встречаемости ИБС, гипертонической болезни, инфаркта миокарда в анамнезе, сахарного диабета 2 типа, гепатоза. У таких больных отмечены признаки атеросклеротического поражения сосудов (по данным УЗДГ и рентгенологического исследования). Среди пациентов с III стадией заболевания преобладали более пожилые пациенты (старше 68,5 лет), имеющие уровень систолического давления выше 195 мм рт. ст., уровень диастолического давления выше 77,5 мм рт. ст. и уровень ХС-ЛПВЛ ниже 0,35 ммоль/л. Большинство проспективных исследований демонстрируют большую роль повышения систолического АД. В исследовании MRFIT [1] относительный риск инсульта у пациентов с повышенным систолическим и диастолическим АД составлял 8,2 и 4,4 соответственно. По данным Фрамингемского исследования показано значение изолированной гипертонии, как независимого фактора риска сердечно-сосудистых осложнений [23,24].

Полученные нами результаты о влиянии артериальной гипертензии на тяжесть течения ДЭП соответствуют также данным, полученным Российским центром профилактической медицины. Артериальная гипертензия рассматривается как фактор риска основных сердечно-сосудистых осложнений и инсульта от любых причин [25].

1.2. Влияние клинических и биохимических факторов, выявленное в рамках двумерных моделей (применение метода ОДР)

В рамках двумерных моделей на уровне значимости $p < 0,05$ было выявлено 289 закономерностей.

В качестве примера на рис. 3 приведём одну из наиболее наглядных и ярко выраженных закономерностей.

На диаграмме представлена двумерная закономерность, показывающая влияние возраста пациента и уровня его систолического давления на тяжесть течения ДЭП.

Рисунок 3 иллюстрирует различия в распределении значений уровня систолического давления и возраста пациентов с I и III стадиями ДЭП. Так, среди пациентов моложе 68,5 лет с уровнем систолического давления ниже 195 мм рт. ст. преобладают больные с I стадией ДЭП (IV квадрант диаграммы).

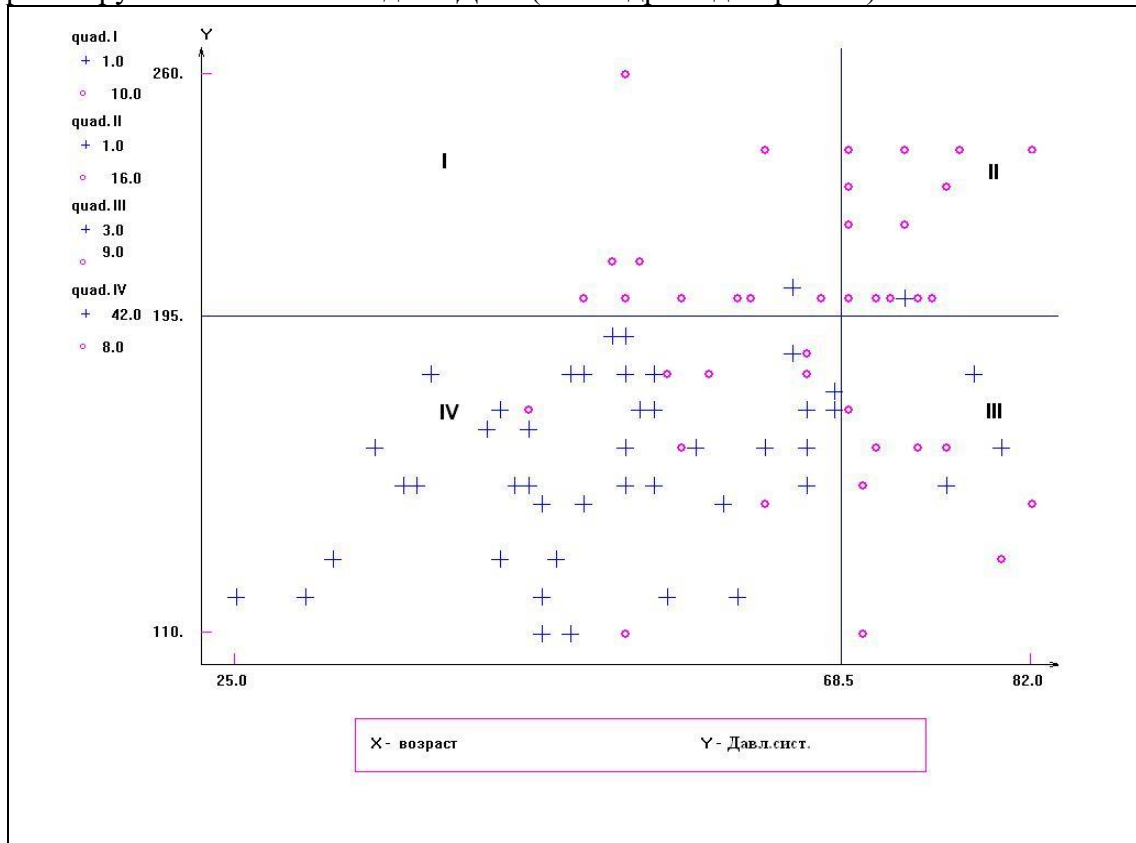


Рис. 3. Диаграмма рассеяния для показателей возраста и уровня систолического давления (по осям X и Y соответственно) у групп пациентов с I стадией ДЭП («+») и пациентов с III стадией ДЭП («o»). $p_1 < 0,003$, $p_2 = 0,006$ (объяснение в тексте).

В квадранты I и II попали значения (с уровнем систолического давления выше 195 мм рт. ст.), соответствующие пациентам с III стадией заболевания. Следовательно, имеется достоверная связь уровня систолического давления с тяжестью течения ДЭП, но только для возрастной группы моложе 68,5 лет ($p < 0,0001$). Относительный риск прогрессирования течения заболевания у пациентов, в возрастной группе моложе 68,5 лет, с уровнем систолического давления выше 195 мм рт. ст. почти в 6 раз выше (OR = 5,68 ДИ 95% [2,93-11,02]), чем у больных с более низкими цифрами давления соответствующего возраста. В группе пациентов старше 68,5 лет не показано влияния крайне высоких цифр систолического давления на тяжесть процесса. Возможно, на тяжесть течения ДЭП оказывают влияние другие факторы. Также на рис.3 видно, что пациенты с I стадией ДЭП (IV квадрант) моложе пациентов с III стадией заболевания (III квадрант), $p < 0,0003$, что дает основание говорить о возрастзависимом нарастании тяжести течения ДЭП.

Для того, чтобы показать, что тяжесть ДЭП зависит именно от пары факторов используется два значения достоверности – p_1 и p_2 . p_1 – p -значение, рассчитанное не путём проверки нулевой гипотезы о полной независимости прогнозируемой

величины (стадии ДЭП в нашем случае) от показателей, а исходя из проверки нулевой гипотезы о том, что зависимость может быть объяснена только переменной X (возрастом). $p_1 < 0,003$. Соответственно, p_2 вычисляется исходя из проверки нулевой гипотезы о том, что зависимость может быть объяснена только переменной Y (систолическим давлением), $p_2 = 0,006$. Подробное описание дано в Приложении 1.

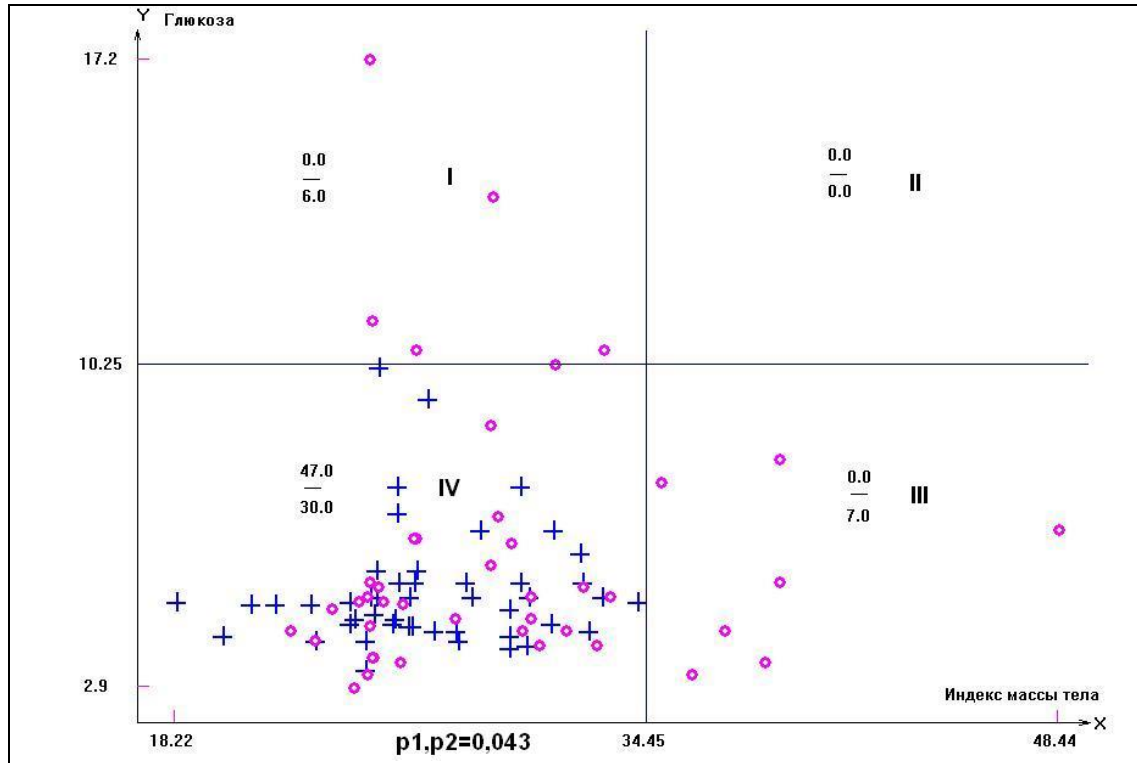


Рис. 4. Диаграмма рассеяния для индекса массы тела (ИМТ, $\text{кг}/\text{м}^2$) (ось X) и уровня глюкозы (ммоль/л) (ось Y) у групп пациентов с I стадией ДЭП (обозначены «+») и пациентов с III стадией ДЭП («o»). $p_1, p_2 = 0,043$.

На рисунке 4 отображена зависимость стадии ДЭП от индекса массы тела (ось X) и содержания уровня глюкозы в крови. Так, среди пациентов с ожирением (III квадрант) только больные с III стадией вне зависимости от уровня глюкозы, а среди пациентов с индексом массы тела до $34,45 \text{ кг}/\text{м}^2$ (квадрант I) – примерно равное количество больных с I и III стадиями (47/30). Среди пациентов с высокими значениями уровня глюкозы (квадрант I) и ИМТ более $34,45 \text{ кг}/\text{м}^2$ (квадрант III) выявлены больные только с III стадией ДЭП. Таким образом, можно предположить, что уровень глюкозы более 10,25 ммоль/л и ожирение (ИМТ $> 34,45 \text{ кг}/\text{м}^2$) являются независимыми факторами, влияющими на тяжесть течения заболевания.

Значимость двумерной закономерности оценивалась аналогично выше описанному способу, однако в данном случае одномерные закономерности отсутствовали, как для показателей ИМТ, так и для показателей глюкозы. Поэтому величины p_1 и p_2 совпали; $p_1, p_2 = 0,043$.

На рисунке 5 отображена зависимость стадии ДЭП от возраста и наличия сахарного диабета II типа. Среди больных с сахарным диабетом II типа встречаются больные с III стадией ДЭП вне зависимости от возраста (I, II квадрант). У пациентов без СД II типа (III, IV квадрант) выявляется возрастзависимое нарастание тяжести заболевания.

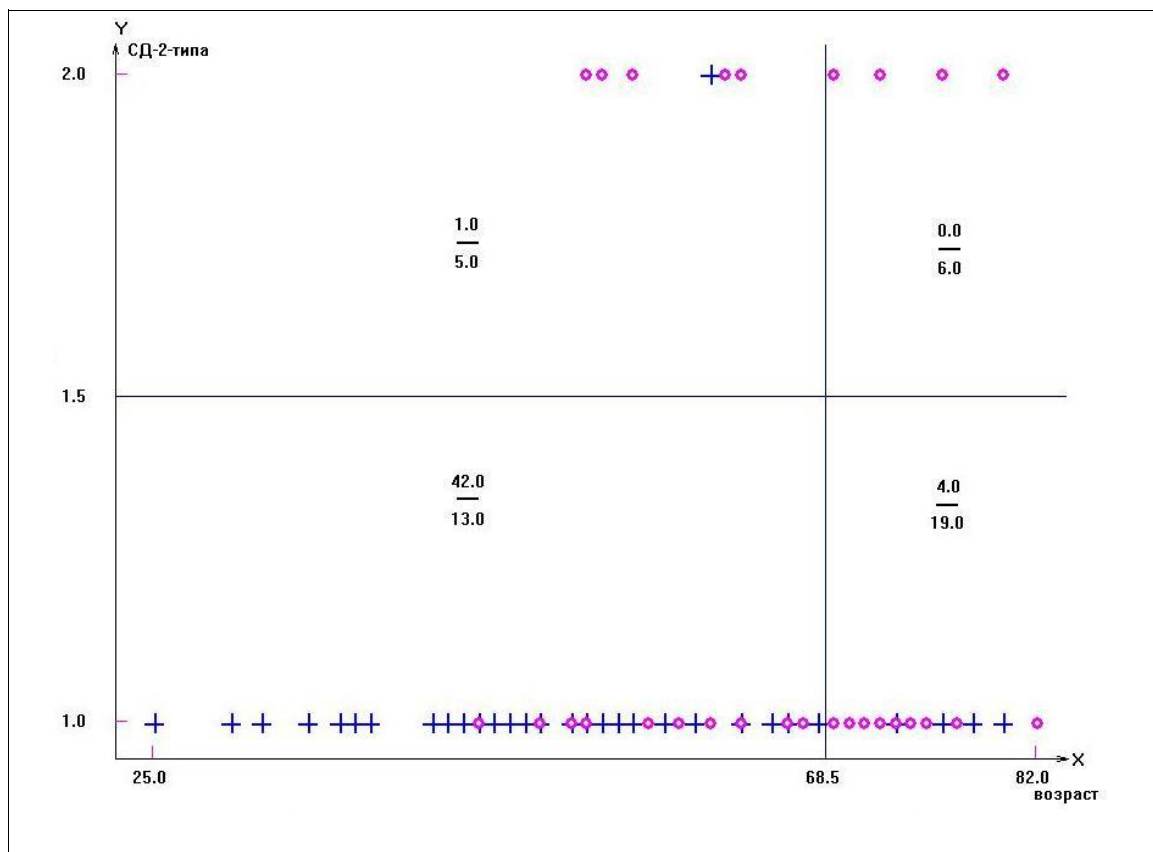


Рис. 5. Диаграмма рассеяния для показателей возраста и наличия у больных сахарного диабета II типа (1 – нет СД, 2 – есть СД) (по осям X и Y соответственно) для групп пациентов с I стадией ДЭП (обозначены «+») и пациентов с III стадией ДЭП («o»). $p_1 = 0,021$, $p_2 < 0,0002$.

1.3. Применение метода ОДР для изучения влияния на течение ДЭП генетических факторов

Для изучения связи полиморфных вариантов генов АПФ, ЛПЛ и СЕТР с тяжестью течения ДЭП производилось разбиение пациентов на две группы в зависимости от генотипов каждого гена. При этом в первую группу входили пациенты, у которых был установлен определенный генотип. Во вторую группу входили пациенты, у которых

выявлялись два других генотипа. Таким образом, было получено всего 9 дихотомических разбиений. Статистические тесты использовались для выявления статистических различий между распределениями оценок тяжести для каждой из дихотомий. Кроме этого дополнительно для описания полиморфизма трёх рассматриваемых генов использовались трёхзначные параметры. Для гена АПФ генотипу DD соответствовала 1 (единица), ID – 2, II – 3. Для гена ЛПЛ генотипу H^+H^+ соответствовала 1, H^+H^- – 2, H^-H^- – 3. Для гена СЕТР генотипу B1B1 соответствовала 1, B1B2 – 2, B2B2 – 3. Наряду с отдельной оценкой влияния полиморфных вариантов генов АПФ, ЛПЛ, СЕТР на тяжесть ДЭП, изучалось также совместное сочетание с генотипом 125 клинических и биохимических параметров. Исследование проводилось с использованием одномерных и двумерных моделей разбиений на выборке из 47 пациентов с I стадией и 43 пациентов с III стадией ДЭП.

Одномерных закономерностей, имеющих уровень статистической достоверности не хуже 0,05, с помощью методики ОДР выявлено не было. Однако была обнаружена тенденция на уровне 0,07 для бинарного индикаторного показателя для генотипа DD.

На рисунке 6 представлено одномерное разбиение по оси X, отображающее варианты генотипов АПФ (1,0 – генотипы II и ID, 2,0 – генотип DD). Ось Y, отображающая возраст пациентов, не участвует в разбиении. Соотношение пациентов с генотипом АПФ – II и ID по стадиям было: 34 – I стадия ДЭП, 24 – III стадия ДЭП. В то время как это соотношение для пациентов с генотипом DD гена АПФ было соответственно 7 против 14. Критерий χ^2 – выявил достоверное различие с $p = 0,047$.

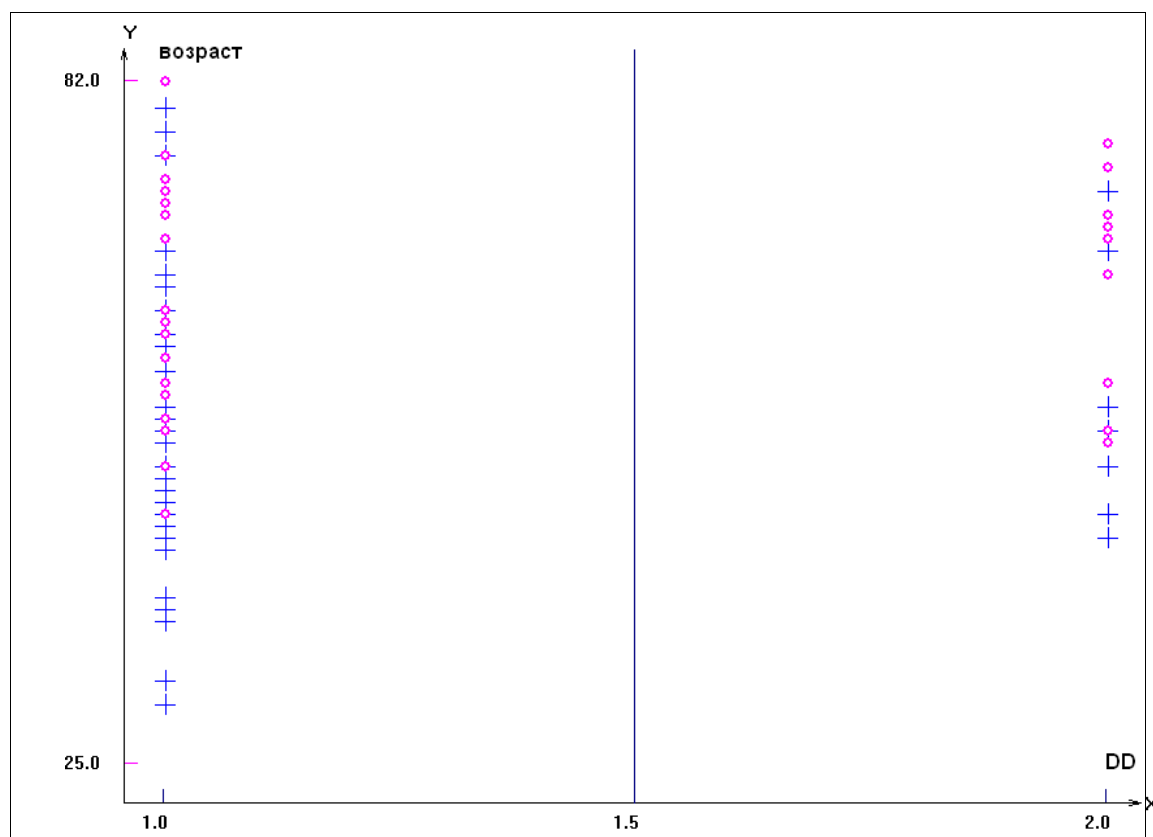


Рис. 6. Диаграмма рассеяния для генотипов АПФ (1.0 – генотипы II и ID, 2.0 – генотип DD) и показателя возраста (вспомогательная ось) (по осям X и Y соответственно) для групп пациентов с I стадией ДЭП («+») и пациентов с III стадией ДЭП («o»), $p < 0,07$.

Также достоверные закономерности были выявлены с использованием двумерных моделей.

На диаграмме (см. рисунок 7) представлена двумерная закономерность, связывающая тяжесть течения ДЭП с генотипами гена АПФ и уровнем преβ-липопротеидов. Видно, что при крайне высоком – 32,4–45,9 мг/дл (1,8–2,6 ммоль/л) уровне преβ-липопротеидов преобладают больные с I стадией ДЭП, независимо от генотипов гена АПФ (I и II квадранты).

Среди пациентов с генотипом DD и повышенным уровнем преβ-липопротеидов (диапазон 12,9–32,4 мг/дл) наблюдается тенденция к снижению количества больных с III стадией ДЭП (85% против 25%). Возможно, происходит элиминация пациентов с генотипом DD гена АПФ и с крайне высоким уровнем преβ-липопротеидов. Среди носителей генотипов ID или II гена АПФ и крайне высокими уровнями преβ-липопротеидов отмечено отсутствие пациентов с III стадией ДЭП (48% против 0). Таким образом, можно говорить о позитивном влиянии генотипов II и ID гена АПФ на течение ДЭП вне зависимости от уровня преβ-липопротеидов.

При изучении связи течения ДЭП с полиморфными вариантами генов выявлено 26 двухфакторных закономерностей с геном ангиотензинпревращающего фермента,

9 закономерностей с геном белка-переносчика эфиров холестерина и 8 закономерностей с геном липопротеинлипазы.

Прицельное изучение связи стадии течения ДЭП с полиморфными вариантами гена АПФ выявило следующее число двухфакторных закономерностей: с генотипом DD – 11, с генотипом ID – 6, с генотипом II – 2. Когда ген АПФ выступал в качестве 3-значного параметра, выявлено 7 двухфакторных закономерностей.

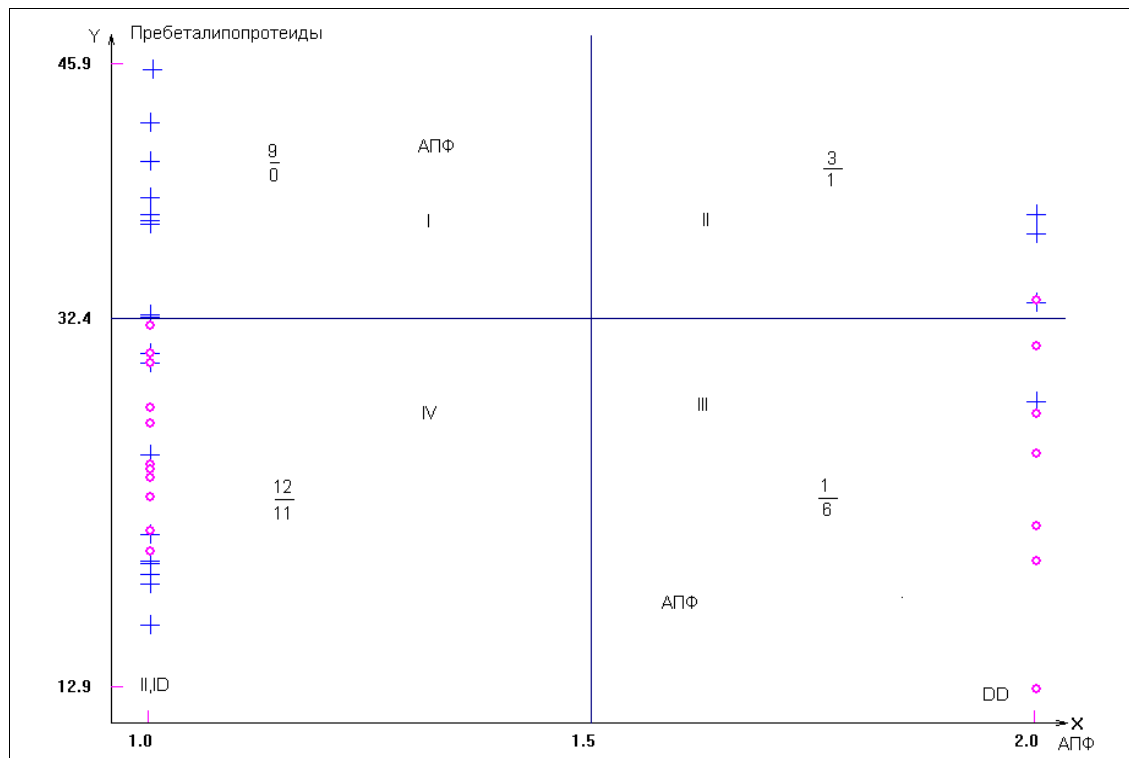


Рис. 7. Диаграмма рассеяния для генотипов АПФ (1.0 – генотипы II и ID, 2.0 – генотип DD) и показателя уровня преβ-липопротеидов (мг/дл) (по осям X и Y, соответственно) для групп пациентов с I стадией ДЭП (обозначения «+») и пациентов с III стадией ДЭП («○»), $p1 < 0,0003$, $p2 = 0,045$.

Полный перечень закономерностей представлен в таблице 2. В таблице представлены автоматически выбранные границы, функционал качества, р-значения и число значений показателей признаков у пациентов двух групп в 4 квадрантах. Квадранты обозначались по часовой стрелке от верхнего левого, аналогично диаграмме на рисунках 3, 4 и 7. Жирным шрифтом выделены достоверности $p \leq 0,025$. Расшифровка обозначений клинических и биохимических параметров из таблицы 2 представлена в Приложении 2.

Можно видеть, что в таблице 2 присутствуют оба типа двумерных закономерностей:

- закономерности, в которых участвуют показатели с ранее выявленными одномерными закономерностями (5, 6, 13, 22, 23);
- закономерности, в которых участвуют показатели, для которых простых закономерностей ранее выявлено не было.

Некоторые иллюстрации – диаграммы рассеяния на двумерных разбиениях с их описанием – представлены в Приложении 3.

Таблица 2. Достоверные двумерные зависимости генетических показателей с клинико-биохимическими показателями при распознавании групп с I и III стадиями ДЭП

N	Генетич. признак	Клинико-биохимический (к.б.х) признак	Граница гено-типа	Граница к.б.х признаков	Функционал качества	p1, p2- значение *	I**	II	III	IV
1	ЛПЛ***	Эф.Хол.	1,500	299	13,576	0,02	0/5	8/8	7/0	7/2
2	СЕТР	Фосфолип.	2,500	65,8	12,240	0,014	15/5	1/1	4/0	2/9
3	АПФ	стадии-ожирен.	1,500	2,5	10,877	0,03	3/4	6/12	28/12	4/10
4	АПФ	Дс-ч	1,500	1,5	11,155	0,008	7/12	26/24	8/0	0/2
5	АПФ	ГБ	1,500	1,5	11,155	p1=0,009, p2=0,027	7/12	26/24	8/0	0/2
6	АПФ	СД-2-типа	1,500	1,5	13,409	p1=0,0024, p2=0,025	0/0	1/8	33/16	7/14
7	АПФ	Преб-ЛП	1,500	32,4	12,517	0,0298	3/1	9/0	12/11	1/6
8	ЛПЛ Н ⁺ Н ⁺	ЛПОНП-ч	1,500	1,5	7,705	0,047	9/8	1/8	9/4	11/8
9	ЛПЛ Н ⁺ Н ⁺	Преб- ЛП-ч	1,500	1,5	8,075	0,042	11/1	3/4	53	6/9
10	ЛПЛ Н ⁺ Н ⁺	Эф.Хол.	1,500	299	13,576	0,012	8/8	0/5	7/2	7/0
11	НН	ЛПОНП-	1,500	1,5	7,672	0,0354	8/16	2/0	2/3	18/9
12	СЕТР В2В2	Фосфолип.	1,500	76,8	12,240	0,0185	15/5	1/1	4/0	2/9
13	СЕТР В2В2	T4	1,500	26,1	11,315	0,043	0/3	0/0	5/0	18/4
14	АПФ DD	стадии-ожирен.	1,500	2,5	10,877	0,018	6/12	3/4	4/10	28/12
15	АПФ DD	Дс-ч	1,500	1,5	11,155	0,006	26/24	7/12	0/2	8/0
16	АПФ DD	ГБ	1,500	1,5	11,155	p1=0,005, p2=0,018	26/24	7/12	0/2	8/0
17	АПФ DD	СД-2-типа	1,500	1,5	13,409	p1=0,0008, p2=0,02	1/8	0/0	7/14	33/16
18	АПФ DD	Глюкоза	1,500	7,85	11,934	0,034	1/6	1/0	6/14	33/18
19	АПФ DD	ЭКГ-глж	1,500	1,5	14,673	p1=0,014, p2=0,04	17/22	6/11	1/3	17/2
20	АПФ DD	Преб-ЛП	1,500	32,4	12,517	0,019	9/0	3/1	1/6	12/11
21	АПФ DD	УЗИ-цит.уз.	1,500	1,5	7,979	0,04	17/4	3/7	2/1	8/3
22	АПФ DD	T4	1,500	24,4	12,058	0,04	1/4	1/0	6/2	16/1
23	АПФ II	Преб-ЛП	1,500	31,65	10,958	0,0333	12/1	0/1	4/2	9/14
24	АПФ II	Эф.Хол.	1,500	299	11,230	0,03	8/10	0/3	3/0	11/2

* Оба р-значения приводятся только в тех случаях, когда они различаются, т.е. для хотя бы одного из двух показателей существует достоверная одномерная закономерность.

** в столбцах I, II, III, IV, которые обозначают квадранты на диаграммах рассеяния по часовой, приведены значения у пациентов с I стадией ДЭП – первое число, после наклонной черты – с III стадией ДЭП.

*** жирным шрифтом выделены строки с достоверностью $p2 \leq 0,02$, а также с достаточно высоким значением функционала качества.

2. Применение методов распознавания

Методы распознавания были использованы для исследования влияния генотипа на тяжесть течения ДЭП. На первом этапе изучалась эффективность различных методов при анализе тяжести течения ДЭП в группах больных с I и III стадиями. При этом диагностика проводилась по клиническим или биохимическим показателям, исключая генетические факторы.

Для достоверной оценки диагностической способности использовался метод скользящего контроля. В результате выделялся набор методов, обладающих достаточной эффективностью. Структура алгоритма распознавания позволяет не только осуществлять диагностику в смысле отнесения пациентов к одной из групп, но производить оценку тяжести в непрерывной шкале [16, 18]. В качестве компьютерной оценки тяжести течения заболевания у пациента была использована соответствующая оценка за класс (группу пациентов), вычисляемая распознающим оператором.

Под компьютерной оценкой тяжести понимается число, вычисляемое алгоритмом для каждого пациента по набору входных показателей. Для рассмотренных методов такая оценка принадлежит отрезку $[0,1]$. Чем ближе к нулю величина оценки, тем более высокая тяжесть заболевания (III стадия ДЭП). Соответственно, близость оценки к 1 свидетельствует о принадлежности объекта к группе с I стадией ДЭП.

Например, для пациента №X алгоритмом вычислена оценка 0,25. Это означает, что для него высока вероятность принадлежности к группе с III стадией ДЭП. При некоторой области, близкой к 0,5, пациент относится к области неопределенности.

Точность алгоритма определяется по общему числу соответствия автоматического диагноза и реального состояния пациента (стадии ДЭП).

Наряду с отдельными методами рассматривалось также коллективное решение, основанное на голосовании по большинству алгоритмов: объект относится в тот класс, в который его отнесло большее число методов. Для оценки тяжести течения заболевания при использовании коллективного решения, вычисляли оценку тяжести для произвольного пациента по большинству отнесений его алгоритмами из эффективного набора к одной из двух исследуемых стадий ДЭП. Коллективный метод компенсирует ошибки отдельных алгоритмов, входящих в коллектив. Теоретически доказано, что ошибка коллективного решения в задачах прогнозирования не превышает среднюю ошибку, вычисленную для методов по отдельности. Все это позволяет говорить о высокой надежности коллективного решения.

3.2.1. Применение методов распознавания для диагностики I и III стадий ДЭП

Исследования на выборке из 47 пациентов с I стадией и 43 пациентов с III стадией ДЭП показали достаточно высокую (свыше 80% для некоторых методов) диагностическую способность 125 клинических и биохимическим показателей при использовании методов распознавания (см. табл. 3).

Использовали все 125 показателей, точнее, все они подаются на вход алгоритмов распознавания, которые автоматически создают решающее правило. Каждый алгоритм исходит при этом из заложенной в него модели при построении оптимального решающего правила. Не все алгоритмы позволяют определить набор информативных признаков. Из использованных методов такой набор известен только для СВС, ЛЗ и частично ЛДФ. Информативные признаки, вошедшие в решающее правило, полученное методом СВС, приведены в Приложении 4.

Таблица 3. Диагностическая способность методов распознавания в задаче диагностики I и III стадий ДЭП с участием и без участия ОНМК

Методы интеллектуального анализа данных	Полный набор из 125 показателей	Набор без ОНМК
Линейного дискриминанта Фишера (ЛДФ)	77,00%	74,50%
Метод <i>k</i> ближайших соседей (k-NN)	67,80%	74,30%
Статистически взвешенные синдромы (СВС)	85,50%	76,50%
Метод опорных векторов (МОВ)	83,3 %,	80,00%
Мультипликативной нейронной сети (МНС)	81,00%	67,70%
Метод логических закономерностей (ЛЗ)	89,90%	77,80%

Поскольку показатель ОНМК (острое нарушение мозгового кровообращения в анамнезе) был очень информативным, важнейшим показателем, определяющим тяжесть ДЭП, то для того, чтобы выявить роль остальных клинико-биохимических показателей, он был исключён из базы, используемой при изучении связи компьютерной оценки тяжести ДЭП с генотипом. Точность распознавания для большинства (за исключением метода *k* ближайших соседей) методов снизилась, но незначительно.

2.2. Применение методов распознавания для диагностики пациентов со II стадией ДЭП

Проделанная работа позволила получить «синдром», вычисляемый по объективным показателям и характеризующий степень тяжести ДЭП. Можно предположить, что пациенты с промежуточным состоянием (II стадия ДЭП) будут отличаться по степени «близости» к I стадии или к III стадии ДЭП. Это позволило выделить группу риска с отрицательной динамикой по ДЭП.

Полученный коллективный алгоритм применили для диагностики пациентов со II стадией ДЭП (145 человек). В результате в область неопределенности попали 13 человек. К I первой стадии ДЭП было отнесено 83 человека, к III стадии 49 человек. Таким образом, можно считать, что эти 49 человек имеют угрозу ухудшения тяжести состояния, и, следовательно, они должны быть выделены в группу риска с последующим повышенным вниманием и особой схемой ведения больного.

2.3. Связь оценок тяжести течения ДЭП с генетическими факторами

Исследовалась связь оценок тяжести течения ДЭП, полученная с помощью коллективного решения, с генетическими факторами. На первом этапе зависимость изучалась для пациентов со II стадией ДЭП, наблюдения которых не использовались для построения решающего правила. Исследование проводилось с использованием тестов Wilcoxon-Mann-Whitney (WMW) и χ^2 . База данных делилась на группы по наличию вариантов генотипов всех изучаемых генов – АПФ, ЛПЛ и СЕТР. Высокую статистическую значимость связи компьютерной оценки тяжести ДЭП дало только деление пациентов на группы с учетом вариантов гена АПФ. Исследования выявили значимые различия между распределениями оценок тяжести в группах в зависимости от генотипа DD гена АПФ или генотипов ID и II. Для генов ЛПЛ и СЕТР значимой взаимосвязи с компьютерной оценкой тяжести выявлено не было.

В таблице 4 представлена статистическая значимость взаимосвязи между компьютерной оценкой тяжести ДЭП и генотипом DD гена АПФ, рассчитанная с помощью критериев WMW и χ^2 для методов ЛДФ, МОВ, СВС и коллективного

решения по большинству, включающий ЛДФ, k-NN, СВС, МОВ. Под компьютерной оценкой тяжести коллективного решения понимается доля алгоритмов распознавания, отнесших данного пациента к I стадии ДЭП.

Каждый из распознающих алгоритмов, обученных по группам пациентов с I и III стадиями заболевания, а также коллективное решение, далее были использованы при диагностике тяжести течения ДЭП у каждого из пациентов со II стадией. На этом этапе использовали полный набор клинико-биохимических показателей.

Из таблицы 4 видно, что оба критерия (WMW и χ^2) выявили наличие взаимосвязи компьютерной оценки тяжести ДЭП с наличием генотипа DD гена АПФ для методов ЛДФ. При этом критерий WMW выявил зависимость со значимостью $p = 0,011$. Для методов МОВ выявлена тенденция с помощью метода χ^2 . Для метода СВС достоверной значимости не выявлено. Коллективное решение выявляет значимость $p = 0,028$ (меньше 0,05) по критерию χ^2 . Поскольку коллективное решение более надежно, то хороший уровень значимости доказывает наличие зависимости тяжести течения ДЭП с генетическими факторами.

Таблица 4. Статистическая значимость взаимосвязи компьютерной оценки тяжести ДЭП с наличием генотипа DD гена АПФ для пациентов со II стадией ДЭП по всем клинико-биохимическим показателям

Методы распознавания	Критерий WMW	Критерий χ^2
ЛДФ	0,011	0,022
МОВ	нет	0,056
СВС	нет	нет
Коллективное решение	0,054	0,028

В таблице 5 представлены оценки статистической значимости связей компьютерной оценкой тяжести ДЭП с вариантом DD гена АПФ, рассчитанной по объединенной выборке пациентов, включающей пациентов с I, II и III стадиями ДЭП. Доказательство взаимодействия компьютерной оценки тяжести ДЭП с вариантом DD гена АПФ по полной таблице является наиболее надежным и достоверным по сравнению с результатами отдельных групп пациентов.

Таблица 5. Статистическая значимость взаимосвязи компьютерной оценки тяжести ДЭП с наличием варианта DD гена АПФ для полной выборки пациентов по всем клинико-биохимическим показателям (I, II и III стадии ДЭП)

Методы распознавания	Критерий WMW	Критерий χ^2	ОДР-одномерная модель
ЛДФ	0,007	0,07	0,08
МОВ	0,024	0,017	0,05
СВС	0,06	0,07	0,07
Коллективное решение	0,014	0,014	0,007

Здесь наряду со стандартными статистическими критериями WMW и χ^2 использовали также метод оптимальных достоверных разбиений. Из таблицы 5 видно, что оценка по критерию WMW оказалась значимой для всех методов, кроме СВС. Для критерия χ^2 значимая взаимосвязь выявлена для МОВ и коллективного решения. Для критерия ОДР высокая значимость выявлена для коллективного решения, хотя для самих методов выявлены только тенденции. Из математической теории распознавания известно, что коллективные решения обладают более высокой устойчивостью и точностью прогноза по сравнению с устойчивостью и точностью

прогноза отдельных методов коллектива, поэтому высокая корреляция прогноза с индикатором DD генотипа гена АПФ является убедительным доказательством действительного существования такой связи.

Далее изучалась связь оценок тяжести распознающих алгоритмов, обученных по группам пациентов с I и III стадиями заболевания по массиву клинико-биохимических показателей за исключением ОНМК, а также соответствующего коллективного решения, с наличием генотипа DD гена АПФ в группе пациентов со II стадией ДЭП (таб. 6).

Из таблицы 6 видно, что значимые различия на уровне $p < 0,05$ выявлены для оценок тяжести методом ЛДФ, а также для оценок тяжести, полученных коллективным решением ($p = 0,011$).

Таблица 6. Статистическая значимость взаимосвязи компьютерной оценки тяжести ДЭП по клинико-биохимическим показателям, за исключением ОНМК, с наличием генотипа DD гена АПФ у пациентов со II стадией ДЭП

Методы распознавания	Критерий WMW	Критерий χ^2
ЛДФ	0,01	0,04
МОВ	0,08	0,07
СВС	0,098	нет
Коллективное решение	нет	0,011

В таблице 7 представлены результаты исследования взаимосвязи компьютерной оценки тяжести ДЭП с наличием генотипа DD гена АПФ по полной выборке пациентов – с I, II и III стадиями ДЭП. Наряду с результатами применения стандартных статистических критериев представлены также статистические оценки, полученные в рамках метода ОДР, о которых подробнее будет сказано в следующем разделе.

Таблица 7. Статистическая значимость взаимосвязи компьютерной оценки тяжести ДЭП по клинико-биологическим показателям, за исключением ОНМК, с наличием у пациента генотипа DD гена АПФ для полной выборки больных

Метод распознавания	Критерий WMW	Критерий χ^2	ОДР - одномерная модель
ЛДФ	0,01	0,05	0,04
МОВ	0,02	0,01	0,01
СВС	0,02	нет	0,05
Коллективное решение	0,05	0,0038	0,01

Из таблицы 7 хорошо видно, что связь компьютерных оценок тяжести генотипом DD гена АПФ выявляется на уровне $p < 0,05$ всеми тремя критериями для методов ЛДФ, МОВ и коллективного решения. Для метода СВС значимость различий подтверждена на уровне $p < 0,05$ только критерием WMW. Тенденция с $p = 0,0502$ выявлена также перестановочным тестом в рамках метода ОДР.

Наиболее высокая оценка значимости (0,0038) получена с помощью критерия χ^2 для коллективного метода.

Различия между распределениями отнесений к I и III стадиям в группах с генотипом DD гена АПФ и генотипами ID и DD представлены в таблице 8.

Таблица 8. Отнесение пациентов, носителей генотипа DD гена АПФ, коллективом алгоритмов распознавания к I и III стадиям ДЭП из полной выборки случаев. Полная выборка включает все случаи, отнесённые экспертами к I, II и III-й стадиям

Большинство алгоритмов отнесло к стадии ДЭП:	Носители генотипов ID и II гена АПФ абс. число (% случаев)	Носители генотипа DD гена АПФ абс. (% случаев)
I стадия	82 (79%)	22 (21%)
III стадия	39 (58%)	28 (42%)

Из таблицы видно, что количество пациентов – носителей генотипа DD, которые были отнесены к III стадии заболевания, достоверно ($p = 0,0056$) превосходит количество пациентов с генотипом DD, отнесенных к I стадии ДЭП. Следовательно, наличие у больного генотипа DD можно считать маркером более тяжелого течения ДЭП.

3. Изучение взаимосвязи компьютерных оценок тяжести ДЭП и генетических факторов с помощью метода Оптимальных Достоверных Разбиений (ОДР)

Стандартные статистические тесты выявили наличие достоверной однофакторной связи между индикаторной функцией генотипа DD гена АПФ и компьютерной оценкой тяжести. Здесь и далее под индикаторной функцией (индикатором) генотипа XX некоторого гена понимается величина, принимающая два значения. При этом одно значение принимается в случаях с генотипом XX и другое значение принимается для альтернативных генотипов. Однако представляют интерес закономерности, отражающие совместное влияние генетических факторов на компьютерную тяжесть. Для изучения таких зависимостей был использован метод ОДР.

Использование двумерных моделей привело к обнаружению двумерной закономерности, описывающей совместное влияние на тяжесть заболевания пола и индикаторной функции генотипа DD гена АПФ.

Данная закономерность приведена в таблице 9. Статистическая значимость выявлена на уровнях $p_1 = 0,013$ относительно достоверной одномерной закономерности, выявленной для пола, и $p_2 = 0,009$ относительно достоверной одномерной закономерности, выявленной для индикаторной функции генотипа DD гена АПФ.

Таблица 9. Связь совместного влияния пола и генотипа DD гена АПФ на тяжесть течения ДЭП

Генотипы гена АПФ у пациентов различного пола	Количество отнесений к I/III стадиям ДЭП (абс.)
генотип DD, мужчины	3/13
генотип ID и II, мужчины	17/14
генотип DD, женщины	19/15
генотип ID и II, женщины	65/25

Из таблицы 9 видно достоверное ($p = 0,0286$) влияние на тяжесть течения ДЭП полиморфных вариантов гена АПФ у мужчин. Подавляющее большинство пациентов-мужчин, носителей генотипа DD, было отнесено к III стадии ДЭП. Для женщин такой закономерности не выявлено ($p = 0,009$). Генотип DD гена АПФ у женщин не оказывает влияния на тяжесть течения заболевания. Таким образом, генотип DD гена АПФ можно считать маркером тяжести течения ДЭП, но только

для пациентов мужского пола. Это говорит о гендерных особенностях влияния генотипа на тяжесть ДЭП.

На следующем этапе изучалась зависимость индикаторной функции варианта DD гена АПФ от оценок тяжести и генотипов генов СЕТР и ЛПЛ. Сначала были использованы простейшие одномерные модели. Результаты анализа методом ОДР связи показателя оценки тяжести ДЭП, равного доле алгоритмов комитета, отнесших пациентов к I стадии ДЭП, с индикатором генотипа DD гена АПФ фактически воспроизвели результат, приведённый в таблице 8.

Использование же двумерных моделей разбиений позволило выявить дополнительную закономерность, описывающую совместное влияние на компьютерную тяжесть полиморфных вариантов генов АПФ и СЕТР (рис. 8).

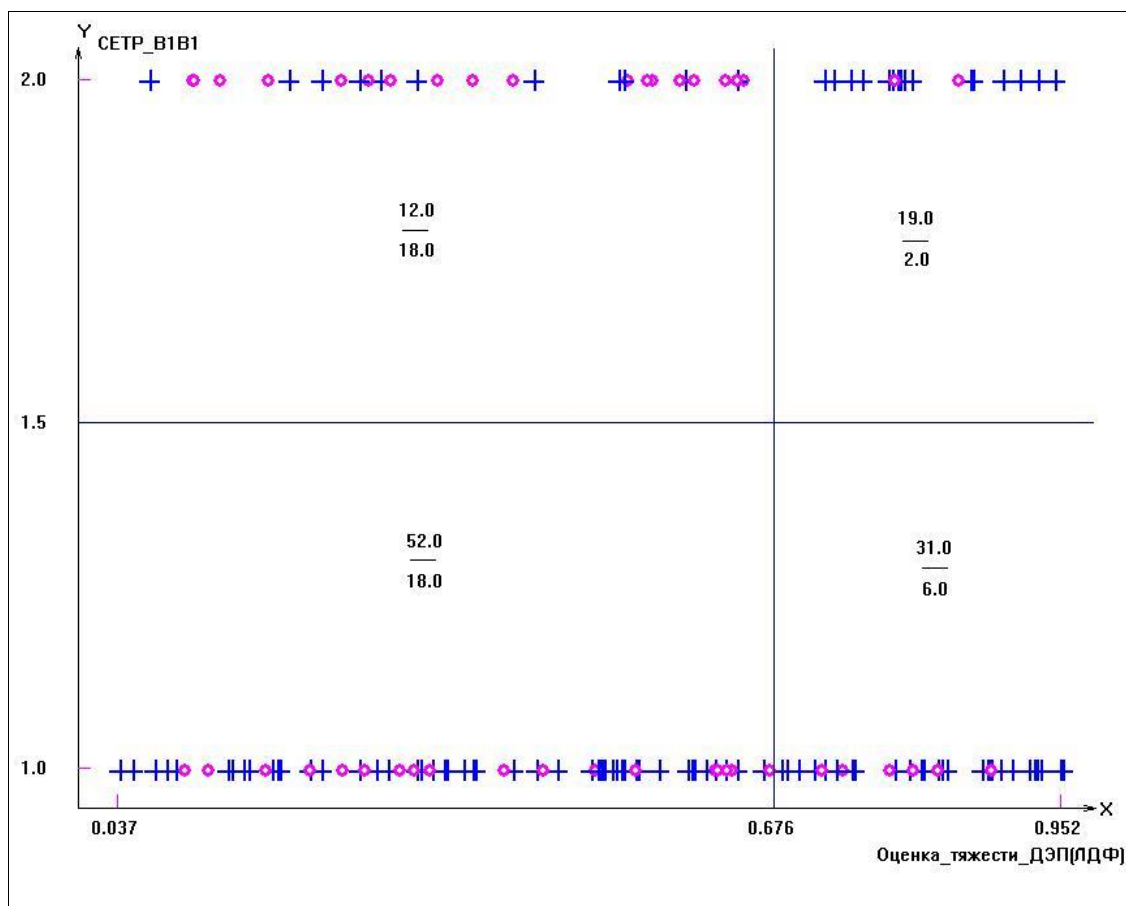


Рис. 8. Диаграмма рассеяния для компьютерных показателей оценки тяжести ДЭП по генотипу DD гена АПФ (методом ЛДФ) и генотипов В1В1 (2.0 по оси Y), В1В2 и В2В2 (1.0 по оси Y) гена СЕТР. «○» соответствует количеству пациентов с генотипом DD гена АПФ, «+» - количество пациентов с генотипами ID и II, $p_1 = 0,02$, $p_2 = 0,006$.

На рисунке 8 изображена двумерная закономерность, связывающая оценку тяжести ДЭП методом ЛДФ, индикаторную функцию генотипа В1В1 гена СЕТР и индикаторную функцию генотипа DD гена АПФ. В данном случае под оценкой тяжести понималась оценка за I стадию ДЭП, рассчитанную распознающим оператором ЛДФ. Напомним, что величина оценки, близкая к 1, в данном случае соответствует более лёгкому течению заболевания.

На рисунке видно, что для больных с генотипом В1В1 гена СЕТР (значение 2,0 по оси Y) наблюдается достаточно отчётливая связь между компьютерной оценкой тяжести течения ДЭП методом ЛДФ и DD вариантом гена АПФ. Значения оценки тяжести, близкие к 1,0 соответствуют лёгкому течению, то есть первой стадии ДЭП.

Оценке тяжести ниже 0,67 и генотипу В1В1 соответствуют 12 случаев с вариантами ID или II и 18 случаев вариантом DD. Оценке тяжести выше 0,67 и генотипу В1В1, напротив, соответствуют 19 случаев с вариантами ID или II и только 2 случая вариантом DD. Для случаев с альтернативными вариантами гена СЕТР (значение 1,0 по оси Y) выраженной связи между вариантами АПФ и оценкой тяжести не наблюдается. Таким образом, можно говорить о сочетанном влиянии генотипов DD гена АПФ и В1В1 гена СЕТР на тяжесть течения ДЭП. Статистическая значимость: $p_1 < 0,05$, $p_2 < 0,05$.

ВЫВОДЫ

1. Исследования показали, что для больных с ДЭП III стадии характерна достоверно более высокая частота встречаемости ИБС, гипертонической болезни, инфаркта миокарда в анамнезе, перенесенных ОНМК и ПНМК, наличие сахарного диабета 2 типа, гепатоза. Преимущественно, это пациенты старше 68 лет, имеющие уровень систолического давления выше 195 мм рт. ст., уровень диастолического давления выше 105 мм рт. ст. и уровень ХС-ЛПВЛ ниже 0,35 ммоль/л.

2. Роль перенесенного ОНМК, возраст старше 68 лет и уровень систолического давления выше 195 мм рт. ст. в прогнозе прогрессирования ДЭП представляется наиболее весомой.

3. Показана достоверная связь систолического давления выше 195 мм рт. ст. с тяжестью течения ДЭП для возрастной группы пациентов моложе 68 лет ($p < 0,0001$). Относительный риск прогрессирования заболевания у таких пациентов, почти в 6 раз выше, чем у больных с более низкими цифрами систолического давления соответствующего возраста. В группе пациентов старше 68 лет не показано влияния крайне высоких цифр систолического давления на тяжесть процесса.

4. Уровень глюкозы более 10,25 ммоль/л и ожирение II–III стадии ($ИМТ > 34,45 \text{ кг/м}^2$) являются независимыми факторами, влияющими на тяжесть течения заболевания.

5. Сахарный диабет II типа является фактором, увеличивающим риск прогрессирования ДЭП вне зависимости от возраста пациентов. У пациентов без сахарного диабета II типа выявляется возрастзависимое нарастание тяжести заболевания.

6. Генотип DD гена АПФ выступает, как предиктор тяжести течения ДЭП, главным образом, для пациентов мужского пола. Это позволяет говорить о гендерных особенностях влияния генотипа на тяжесть ДЭП. Влияние генотипа DD, в большей степени, проявляется у больных ДЭП с уровнем систолического давления ниже 195 мм рт.ст.

7. Выявлено сочетанное влияние генотипов DD гена АПФ и В1В1 гена СЕТР на тяжесть течения ДЭП.

ЗАКЛЮЧЕНИЕ

Дисциркуляторная энцефалопатия – одна из основных причин развития когнитивной дисфункции у пожилых. Раннее распознавание клинико-генетических маркеров, способствующих неблагоприятному течению заболевания и риску развития осложнений, поможет разработать индивидуальные программы профилактики и лечения. Это приведет к улучшению качества жизни больных, снижению инвалидизации и позволит добиться большего контроля над смертностью от цереброваскулярной патологии.

Проведенные исследования с использованием новых математических методов (методом оптимального разделения, методом статистически взвешенных синдромов, методами системы «Распознавание»): позволяют выявить наиболее информативные

показатели для прогнозирования тяжести ДЭП. Наиболее важными клиническими показателями, в порядке убывания информативности, можно считать: 1) перенесенные ОНМК; 2) возраст старше 68 лет; 3) уровень систолического давления выше 195 мм рт.ст.; 4) гипертрофия левого желудочка (ЭКГ-признаки); 5) сахарный диабет II типа; 6) ожирение II-III стадия (ИМТ > 34,45 кг/м²).

Показано влияние генотип DD гена АПФ на тяжесть течения ДЭП.

В процессе исследования:

1. показано, что из совокупности двухфакторных закономерностей, описывающих связь полиморфных вариантов генов АПФ, ЛПЛ, СЕТР с ДЭП, большая часть достоверных закономерностей связана с геном АПФ;
2. получена компьютерная оценка тяжести ДЭП;
3. установлена прямая связь компьютерной оценки тяжести на уровне $p < 0,005$ с генотипом DD гена АПФ;
4. в рамках двухфакторной закономерности установлено совместное влияние на компьютерную оценку тяжести сочетаний генотипов В1В1 гена СЕТР и DD гена АПФ;
5. в рамках двухфакторной закономерности установлено совместное влияние мужского пола пациента и наличия генотипа DD гена АПФ на компьютерную оценку тяжести;
6. показано, что использованный программный комплекс может быть рекомендован для анализа данных в случае большого количества признаков в обучающих выборках, плохо структурированных признаков и пропусков в данных, что характерно для медико-биологических исследований.

ПРИЛОЖЕНИЕ 1.

МЕТОД АНАЛИЗА ДАННЫХ, ОСНОВАННЫЙ НА ОПТИМАЛЬНЫХ ДОСТОВЕРНЫХ РАЗБИЕНИЯХ ПРОСТРАНСТВ ПРОГНОСТИЧЕСКИХ ПЕРЕМЕННЫХ

1. Введение

Целью метода **оптимальных достоверных разбиений (ОДР)** является построение по выборке эмпирических данных по возможности наиболее точного и достоверного описания зависимости некоторой величины ζ от набора потенциальных прогностических переменных X_1, \dots, X_n .

Для решения данной задачи может быть использован широкий круг методов, в число которых входит использование разнообразных статистических тестов для выявления значимых различий в распределениях прогностических переменных для задаваемых ζ групп объектов, корреляционный, ковариационный и дисперсионный методы анализа. Перечисленные подходы требуют предположений о характере существующих распределений, а также априорного выделения сравниваемых групп объектов, что вызывает необходимость разработки дополнительных средств анализа, лишённых указанных недостатков.

Метод ОДР основан на поиске оптимальных разбиений пространства \mathbf{R}^n , соответствующего прогностическим переменным. На самом деле в рамки технологии ОДР укладывается целая совокупность методов анализа зависимостей, различающихся по характеру прогнозируемой величины. Так, в случае задачи исследования закономерностей распределения в пространстве прогностических

переменных классов K_1, \dots, K_L (для краткости назовём её задачей распознавания) в качестве прогнозируемой величины может выступать

- а) бинарный вектор $(\alpha_1, \dots, \alpha_L)$, где компонента $\alpha_l \in \{0, 1\}$ указывает на принадлежность описываемого объекта классу K_l ;
- б) непрерывный вектор $(\zeta_1, \dots, \zeta_L)$, где компонента $\zeta_l \in [0, 1]$ представляет собой вероятность принадлежности описываемого объекта классу K_l , $l \in \{1, \dots, L\}$.

Прогнозируемыми величинами могут быть также вектора непрерывных переменных произвольной размерности, кривые выживаемости и т.д.

Пусть M_x – область многомерного пространства \mathbf{R}^n , из которой принимают значения переменные X_1, \dots, X_n . Нашей целью далее будет являться поиск оптимальных разбиений, позволяющих выделить подобласти M_x :

- а) для которых в максимальной степени выражена тенденция существенных отклонений значений величины ζ от ее средних значений в соседних областях;
- б) которые описываются в виде конъюнкции простейших неравенств над значениями прогностических переменных.

Предполагается наличие обучающей выборки, представляющей собой набор $\tilde{S}_o = \{(\zeta_1, \mathbf{x}_1), \dots, (\zeta_m, \mathbf{x}_m)\}$. Для задач, в которых в качестве зависимой переменной ζ выступают вектора непрерывных переменных, величины $(\zeta_1, \dots, \zeta_m)$ представляют собой просто значения ζ , измеренные при значениях векторов прогностических переменных $\mathbf{x}_1, \dots, \mathbf{x}_m$. Для задач, распознавания величины $(\zeta_1, \dots, \zeta_m)$ представляют собой бинарные вектора индикаторных функций классов. Для задач исследования зависимости выживаемости от переменных X_1, \dots, X_n величины $(\zeta_1, \dots, \zeta_m)$ представляют собой пары $[(\alpha_1, t_1), \dots, (\alpha_m, t_m)]$, где величины (t_1, \dots, t_m) являются временами последнего наблюдения за объектами с соответствующими векторами прогностических переменных $\mathbf{x}_1, \dots, \mathbf{x}_m$, а величины $(\alpha_1, \dots, \alpha_m)$ являются значениями бинарных показателей, указывающих, существовал ли объект ($\alpha_j = 0$) в момент последнего наблюдения, или в данный момент была зафиксирована его гибель ($\alpha_j = 1$). Предполагается также существование процедуры усреднения для произвольного набора значений переменной ζ . Усреднённое значение ζ по множеству значений $\tilde{\zeta}$ будем обозначать как $\hat{\zeta}(\tilde{\zeta})$.

В случае, когда в качестве зависимой переменной ζ выступают вектора непрерывных переменных, а также в случае задачи распознавания, в качестве усреднённого значения будем использовать просто арифметическое среднее или $\hat{\zeta}(\tilde{\zeta}) = (1/|\tilde{\zeta}|) \sum_{\zeta_j \in \tilde{\zeta}} \zeta_j$. При этом для задачи распознавания компоненты вектора $\hat{\zeta}(\tilde{\zeta})$ интерпретируются, как оценки вероятности принадлежности классам K_1, \dots, K_L на генеральной совокупности из которой выбрана \tilde{S}_o . Для задач исследования выживаемости в качестве усредненного значения $\hat{\zeta}(\tilde{\zeta})$ может быть использована оценка кривой выживаемости Каплан-Майера. Упорядочим объекты обучающей выборки по величинам времени последнего наблюдения. Напомним, что оценка вероятности дожития к моменту времени T в методе Каплан-Майера вычисляется

$$\text{Pr}_{K-M}(t < T, \tilde{\zeta}) = \prod_{t_i < T} \left[\frac{m-i}{m-i+1} \right]^\alpha.$$

Обозначим через M_ζ^e множество всевозможных эмпирических оценок. Предполагается, что на множестве $M_\zeta^e \times M_\zeta^e$ задана функция расстояний $\rho(\hat{\zeta}', \hat{\zeta}'')$, соответствующая существующим представлениям о взаимной удаленности объектов и обладающая следующими свойствами:

- а) $\rho(\hat{\zeta}', \hat{\zeta}'') \geq 0, \forall \hat{\zeta}', \hat{\zeta}'' \in M_\zeta^e,$
- б) $\rho(\hat{\zeta}', \hat{\zeta}') = 0, \forall \hat{\zeta}' \in M_\zeta^e,$
- в) $\rho(\hat{\zeta}', \hat{\zeta}'') = \rho(\hat{\zeta}'', \hat{\zeta}'), \forall \hat{\zeta}', \hat{\zeta}'' \in M_\zeta^e.$

Для задач, в которых в качестве зависимых величин выступают скалярные величины или конечномерные вектора, в качестве функции расстояния может быть использован квадрат декартова расстояния. Для задач исследования выживаемости в качестве функции расстояния между кривой вероятности дожития $P_1(t)$, определённой на отрезке $[0, T_1]$, и кривой вероятности дожития $P_2(t)$, определённой на отрезке $[0, T_2]$, может быть задана в виде интеграла, $\rho(P_1, P_2) = \int_0^{T_m} (P_1 - P_2) dt$, где $T_m = \min(T_1, T_2)$.

2. Построение оптимальных разбиений

Метод основан на построении оптимальных разбиений по эмпирической обучающей выборке $\tilde{S}_o = \{S_1^o = (\zeta_1^o, \mathbf{x}_1^o), \dots, S_m^o = (\zeta_m^o, \mathbf{x}_m^o)\}$ интервалов допустимых значений одиночных переменных или совместных областей допустимых значений групп переменных в рамках априори заданных моделей. При этом разбиение считается оптимальным, если оно индуцирует разбиение \tilde{S}_o на несколько групп с возможно минимальными расстояниями между объектами внутри одной и той же группы и возможно максимальными расстояниями между объектами из разных групп. Данные различия между группами и внутри групп описываются с помощью специального функционала, который мы далее будем называть функционалом качества разбиений. Задача при этом сводится к поиску разбиений из рассматриваемых моделей, на которых достигается максимум функционала качества.

Функционал качества. Предположим, что R - разбиение обучающей выборки \tilde{S}_o на подвыборки $\tilde{S}_1, \dots, \tilde{S}_q$. Тогда интегральный функционал качества определяется как сумма

$$F_I(\tilde{S}_o, R) = \sum_{i=1}^q \rho[\hat{\zeta}(\tilde{S}_i), \hat{\zeta}(\tilde{S}_o)] m_i,$$

где m_i – число объектов в подвыборке \tilde{S}_i .

Наряду с интегральным функционалом качества может быть использован также локальный функционал качества, в котором оценка проводится по подвыборке максимально отличающейся от исходной обучающей выборки \tilde{S}_o

$$F_L(\tilde{S}_o, R) = \max_{i \in \{1, \dots, q\}} \{\rho[\hat{\zeta}(\tilde{S}_i), \hat{\zeta}(\tilde{S}_o)] m_i\}$$

В то время как интегральный функционал качества не позволяет выделять отдельные элементы разбиения и с его помощью могут быть построены оптимальные разбиения, только с числом элементов равным числу элементов в

используемой модели, локальный функционал основан на выделении наилучшего элемента разбиения и позволяет также производить разбиение признакового пространства на две подобласти. Причём на одной из двух подобластей является элемент разбиения, на котором как раз достигает максимума $F_L(\tilde{S}_o, R)$, а вторая подобласть представляет собой объединение всех остальных $q - 1$ элемента.

В случае рассматриваемой далее в этой главе задачи прогнозирования бинарной величины соответствующей задачи распознавания с двумя классами, интегральный функционал качества будет иметь вид $F_I(\tilde{S}_o, R) = \sum_{i=1}^q (v_i - v_o)^2 m_i$.

Модели разбиений. Под моделью разбиения мы понимаем множество разбиений с числом элементов не превышающем некоторое заранее фиксированное число, которые строятся с помощью априори заданного алгоритма.

Примеры моделей разбиений приведены на рисунках 9–11.

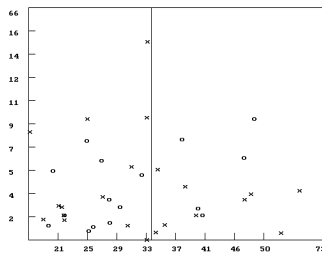


Рис. 9. Модель I.

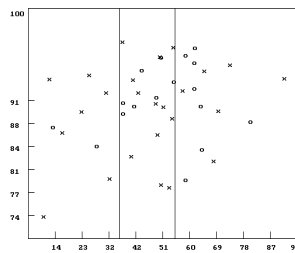


Рис. 10. Модель II.

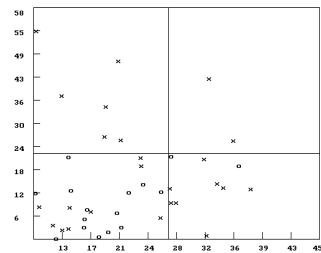


Рис. 11. Модель III.

Модель I включает все разбиения интервалов допустимых значений одиночных переменных с числом элементов (подобластей) не более двух, которые разделены с помощью одной граничной точки. Модель II включает все разбиения интервалов допустимых значений одиночных переменных с числом элементов не более трех, которые разделены с помощью не более двух граничных точек. Модель III включает все разбиения области допустимых значений пары переменных с числом элементов не более четырех. При этом при построении разбиения используется не более одной граничной точки для каждой из двух переменных.

Построение оптимальных разбиений. Поиск оптимальных разбиений производится в рамках выбранной модели путём оптимизации одного из двух описанных выше функционалов. Использование интегрального функционала приводит к разбиению пространства прогностических переменных на число областей, равное числу элементов в используемой модели. Использование локального функционала приводит к дихотомическому разбиению. Оптимизация производится путём перебора всевозможных потенциальных граничных точек. Предположим, что переменная X_i на обучающей выборке принимала значения из монотонно возрастающей последовательности точек $\{a_1^i, \dots, a_{m_i}^i\}$. Тогда в качестве потенциальных граничных точек при разбиении интервала допустимых значений переменной X_i используются точки $\{b_1^i = (a_1^i + a_2^i) / 2, \dots, b_{m_i-1}^i = (a_{m_i}^i + a_{m_i-1}^i) / 2\}$. Поиск оптимальных разбиений в рамках одномерной модели I производится путём последовательного перебора точек $\{b_1^i, \dots, b_{m_i-1}^i\}$. Поиск оптимальных разбиений в рамках одномерной модели II производится путём последовательного перебора пар точек $\{(b_1^i, b_2^i), \dots, (b_1^i, b_{m_i-1}^i), \dots, (b_{m_i-1}^i, b_{m_i-2}^i)\}$. Поиск оптимальных разбиений в рамках

двумерной модели III для совместной области распределений переменных X_i', X_i'' производится путём последовательного перебора пар точек $\{(b_1^{i'}, b_1^{i''}), \dots, (b_{m_i-1}^{i'}, b_{m_i-1}^{i''})\}$. Для повышения скорости поиска на каждом шаге производится не полный пересчёт оптимизируемого функционала, а только коррекция значения найденного на предыдущем шаге.

Верификация полученных результатов. Статистическая верификация построенных оптимальных разбиений может производиться с помощью стандартных статистических тестов (Хи-квадрат, Уилкоксона-Манн-Уитни, Колмогоров-Смирнов и др.). Однако такой подход возможен только в том случае, если объём исходной информации достаточен для формирования двух выборок: обучающей и контрольной. При этом первая из выборок используется для поиска закономерностей, а вторая для вычисления статистики используемого для верификации критерия. В противном случае для верификации используется перестановочный тест, основанный на методе Монте-Карло. Результаты разбиений на реальной выборке сравниваются с результатами разбиений на искусственных выборках, полученных путем случайных перестановок позиций прогнозируемой величины относительно фиксированных позиций прогнозирующих переменных. Автоматически генерируется от 1000 до 5000 таблиц. Доля таких сгенерированных таблиц, для которых функционал качества разбиения оказался выше, чем функционал для реальной выборки, рассматривается как p -значение. Если это значение меньше некоторого выбранного порога (0,1, 0,01, 0,001), то существование закономерностей на реальной выборке считается доказанным на соответствующем уровне.

3. Методы формирования оптимального набора закономерностей

Наиболее простым и естественным методом формирования оптимального набора логических закономерностей является внесение в окончательный список тех закономерности, для которых оценка статистической значимости была не хуже, чем $p \leq \beta$, где β – заданное пороговое значение.

Однако эксперименты с искусственно сгенерированными выборками показали, что использование данного простого метода приводит к включению в окончательный оптимальный набор чрезвычайно большого числа частично ложных закономерностей, выявленных в моделях разбиений II и III. В связи с этим был разработан метод **условных значимостей**.

Рассмотрим ещё один метод селекции логических закономерностей, который мы далее будем называть методом условных значимостей (УЗ). В рамках метода УЗ оценка значимости закономерности Q_c , найденной в рамках моделей II или III для множества переменных $\tilde{X}(Q_c)$, производится исходя из множества простых закономерностей Q_c^l , найденных в рамках модели I для тех же самых переменных. Пусть Q – закономерность из множества Q_c^l . Поставим в соответствие Q гипотезу $H_1(Q)$ о том, что зависимость прогнозируемой величины ζ от переменных из $\tilde{X}(Q_c)$ полностью описывается Q . Иными словами ζ не зависит от переменных из $\tilde{X}(Q_c)$ внутри Q и внутри подобластей разбиения, соответствующего Q . Может быть поставлен вопрос, могут ли различия в группах объектов обучающей выборки, индуцируемые сложной закономерностью Q_c , быть объяснены исходя из гипотезы $H_1(Q)$. Ответ на этот вопрос предполагается также получить с помощью

перестановочного теста. Также как и проверке нулевой гипотезы о полной независимости ζ от $\tilde{X}(Q_c)$ в качестве статистики критерия снова берется оптимальное значение используемого функционала качества разбиений. Множество случайных выборок для оценки p -значений также как и при проверке гипотезы о полной независимости ζ от $\tilde{X}(Q_c)$ формируется из обучающей выборки путём случайных перестановок значений прогнозируемой величины ζ относительно фиксированных векторов прогностических переменных. Однако при проверке гипотезы $H_1(Q)$ перестановки допускаются только внутри подмножеств объектов обучающей выборки, для которых вектора прогностических признаков принадлежат одному и тому же элементу разбиения, соответствующего закономерности Q .

В методе УЗ закономерность Q_c предполагается значимой и включается в выходное множество закономерностей, если она признаётся значимой с точки зрения хотя бы одной из гипотез $H_1(Q)$, где $Q \in Q_c^I$. Так, закономерность, найденная с помощью модели разбиений II для некоторой переменной, вводится в выходной набор, если её значимость, оцениваемая с помощью описанной выше процедуры, исходя из закономерности, найденной для той же самой переменной с помощью модели разбиений I, не хуже заранее заданного порогового значения β . Закономерность, найденная с помощью модели разбиений III для некоторой пары переменных, вводится в выходной набор, если её значимость, оцениваемая с помощью описанной выше процедуры, исходя из закономерностей, найденных для тех же переменных с помощью модели разбиений I, не хуже заранее заданного порогового значения β .

Использование метода УЗ при анализе двумерных закономерностей. При использовании метода УЗ для оценки значимости двумерных закономерностей, выявленных для объясняющих переменных X и Y , вычисляются два p -значения: p_1 и p_2 . При этом p_1 рассчитывается путём проверки нулевой гипотезы о полной независимости прогнозируемой величины от объясняющих переменных, а исходя из проверки нулевой гипотезы о том, что зависимость может быть объяснена в рамках простейшей одномерной закономерности, ранее выявленной для переменной X . Проверка такой нулевой гипотезы производится с помощью перестановочного теста с запрещением перестановок между двумя областями простейшей закономерности для X . Соответственно, p_2 вычисляется, аналогичным образом, исходя из проверки нулевой гипотезы о том, что зависимость может быть объяснена в рамках простейшей одномерной закономерности, ранее выявленной для переменной Y . Если для какой-то из двух переменных простейшей закономерности на фиксированном пользователем уровне не существует, то проверяется нулевая гипотеза о полной независимости прогнозируемой величины от X и Y .

ПРИЛОЖЕНИЕ 2.

НЕКОТОРЫЕ ОБОЗНАЧЕНИЯ, ИСПОЛЬЗУЕМЫЕ В ТАБЛИЦАХ

№	Краткое назв.	Полное название со стандартными границами
1.	Д.сис.	давление систолическое (в цифровых значениях)
	Д.сис.-ч	1 – 110–140 мм рт. с., 2 – больше 140 мм рт. с.
2.	Эф. Хол.	эфиры холестерина (в цифровых значениях)

3.	УЗИ щит.уз.	узлы щитовидной железы: 1 – нет, 2 – да
4.	стадия ожирен.	1-норма (ИМТ 18,5–25 кг/м ²), 2 – 1 ст. (25–29,9 кг/м ²), 3 – 2 ст. (30–39,9 кг/м ²), 4 – 3 ст. (>40 кг/м ²), ИМТ – индекс массы тела
5.	СД II типа	сахарный диабет II типа: 1 – нет, 2 – да
6.	ГБ	гипертоническая болезнь: 1 – нет, 2 – да
7.	ЭКГ- глж	гипертрофия левого желудочка (по данным электрокардиографии): 1 – нет, 2 – да
8.	ЛПОНП	липопротеины очень низкой плотности (в цифровых значениях)
9.	ЛПОНП-ч	1 – <0,63, 2 – >0,63
10.	Фосфолип.	фосфолипиды (в цифровых значениях)
11.	преβ-ЛП	преβ-липопротеиды (в цифровых значениях)
12.	преβ-ЛП-ч	1 – < 30,0, 2 – > 30,0 (мг/дл)
13.	βЛП	β-липопротеиды (в цифровых значениях)
14.	β-ЛП-ч	1 – <50,0 (мг/дл), 2 – >50,0 (мг/дл),
15.	МРТ-ГЦ	магнитно-резонансная томография (данные за наличие гидроцефалии): 1 – нет, 2 – да
16.	T4	свободный тироксин (в цифровых значениях)

ПРИЛОЖЕНИЕ 3.

НЕКОТОРЫЕ ДИАГРАММЫ РАССЕЯНИЯ ДЛЯ ДВУМЕРНЫХ РАЗБИЕНИЙ ПО КЛИНИКО-ЛАБОРАТОРНЫМ И ГЕНЕТИЧЕСКИМ ПОКАЗАТЕЛЯМ

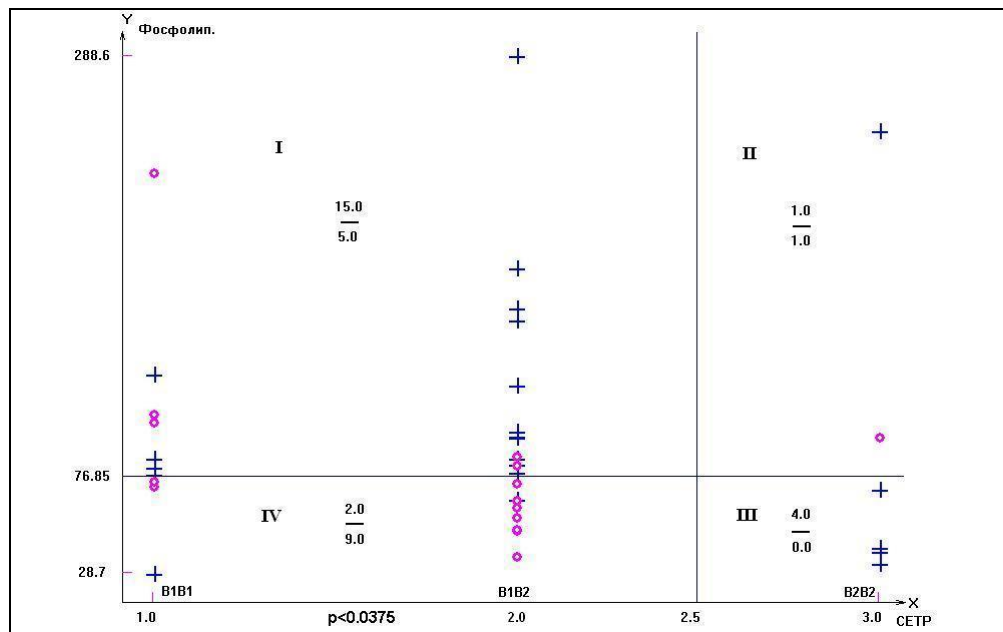


Рис. 12. На рисунке отражено позитивное влияние генотипа B2B2 гена CETP (значение 3.0 по оси X) на тяжесть течения ДЭП даже при низком уровне фосфолипидов (ниже 76,85 мг/дл): отсутствие пациентов с III стадией (III квадрант) – 0% против 82% (IV квадрант) при наличии генотипов B1B1 и B1B2 (значения 1.0 и 2.0 по оси X). Это позволяет говорить о позитивном влиянии генотипа B2B2 на течение заболевания.

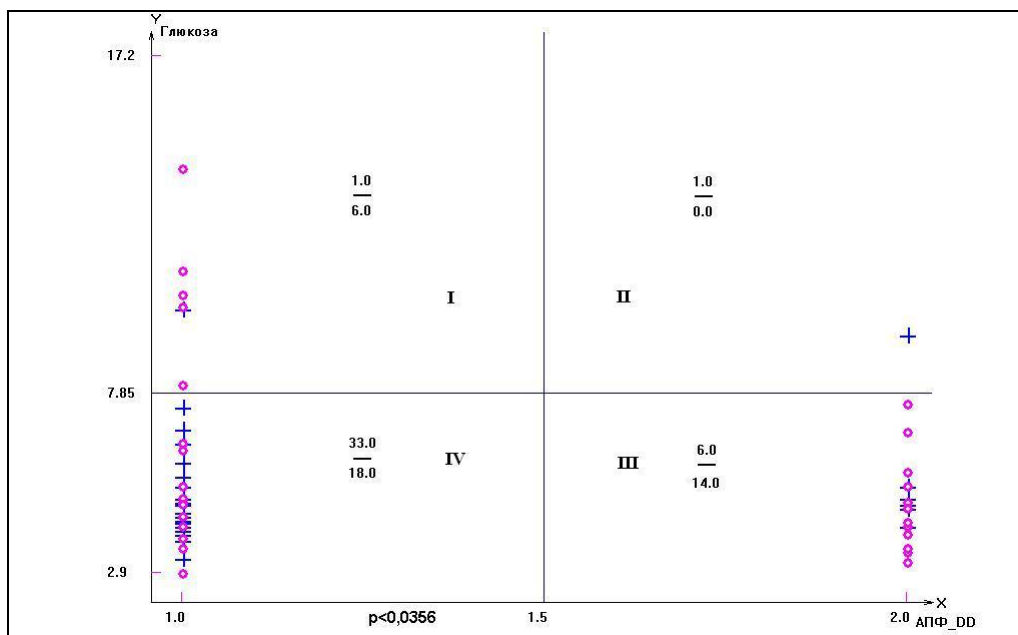


Рис. 13. Граница по глюкозе выставлена машинным способом выше референтного значения нормы уровня глюкозы. В диапазоне 2,9–7,85 ммоль/л (нормальные уровни и чуть выше нормы) наблюдается достоверное различие в соотношении пациентов с I и III стадиями ДЭП в зависимости от генотипа. В III квадранте наблюдается значительное преобладание пациентов с III стадией ДЭП (45% против 15% больных с I стадией), носителей генотипа DD гена АПФ. Таким образом, можно говорить о влиянии генотипа DD на тяжесть течения ДЭП у больных. В тоже время у пациентов с уровнем глюкозы выше (7,85 ммоль/л) влияние генотипа нивелируется.

ПРИЛОЖЕНИЕ 4.

НАИБОЛЕЕ ИНФОРМАТИВНЫЕ ПОКАЗАТЕЛИ (ПО УБЫВАНИЮ ИНФОРМАТИВНОСТИ), ВОШЕДШИЕ В РЕШАЮЩЕЕ ПРАВИЛО ПРОГНОЗА УВЕЛИЧЕНИЯ ТЯЖЕСТИ (КОМПЬЮТЕРНАЯ ОЦЕНКА ТЯЖЕСТИ ДЭП)

№	Краткое название показателя	№ показ.	Функционал
1	ОНМК	25	255.482590
2	МРТ-ЛЮ	98	164.795334
3	МРТ-ОНМК	97	147.063675
4	ИБС-атер	20	120.967804
5	Дсис	11	87.088341
6	Возраст	4	58.240372
7	ТГ	62	54.145458
8	ЭКГ-ГЛЖ	56	47.642784
9	ЛПВП	66	42.918297
10	СД II типа	19	42.828617
11	βЛП	70	40.400249

12	Группы по возрасту	5	38.132450
13	Гепатоз - бинар	110	36.654819
14	T4	126	35.548882
15	ИМТ	8	34.374844
16	Вес	7	31.926586
17	Рост	6	30.690744
18	ЭКГ-нп	58	27.797136
19	Общий белок	35	26.773758
20	ПНМК	24	26.370604
21	МРТ-ГЦ	99	25.210951
22	Rg-аорты	106	23.654018
23	ИБС-ИМ	23	23.574968
24	Поджел. железа	115	22.169552
25	Hb	29	22.134832
26	Глюк-ч	54	21.216473
27	Узи-гепатоз	108	20.664322
28	Гепатоз - ч	109	20.664322
29	ЛПОНП	68	20.285618
30	Стадия ожирения	9	20.081099
31	РЭГтнГт	92	19.842817
32	УЗИ-хол-т	112	19.133516
33	ЭКГ-ИМ	55	19.103443
34	ТГ-ч	63	18.902332
35	Триглицер	40	18.750011
36	ЛПОНП-ч	69	18.545727
37	Фибр-ч	43	18.506529
38	Дс-ч	12	18.381010
39	Дс-из.	13	18.381010
40	Ддиас	14	18.119616
41	МС	10	17.485670
42	ЛПВП-ч	67	16.973644
43	Пол	3	16.822971
44	УЗИ-м.т.	117	16.810574
45	УЗИ-п/ж	113	15.536945
46	подж/железа	114	15.536945
47	ИБС-ст	22	15.408543
48	Бр (билирубин)	47	15.355932
49	Тг	78	15.342134
50	Окулист	107	15.231236
51	Преβ-ЛП	72	15.217369
52	АЛТ	48	15.058541
53	Д д-ч	15	14.252114
54	Миома	122	13.222770
55	Св.х.	83	13.161526
56	Лейкоцит	32	12.554853
57	ТЗ	124	12.465432

58	АЛТ-ч	49	12.388723
59	альфа-ЛП	74	11.251026
60	Фосфолип.	76	11.030509
61	РЭГпкПС	94	10.903591
62	Преβ-ЛП-ч	73	10.895425
63	Рак Предстат Жел	120	10.325603
64	РЭГтнСП	90	10.150853
65	ХС	59	9.554152
66	βЛП-ч	71	8.953374
67	Св.холестерин - ч	84	7.767142
68	Mg	52	5.077855
69	Св.жирные к-ты	85	5.056785
70	ОХС > 8	38	0.921410

СПИСОК ЛИТЕРАТУРЫ

1. Скворцова В.И., Лимборская С.А., Кольцова Е.А., Сломинский П.А. Приложение "Инсульт" к журналу *Неврология и психиатрия*. 2001. Т. 1. С. 23–29.
2. Гусев Е.И., Мартынов М.Ю., Ясаманова А.Н. и др. Этиологические факторы и факторы риска сосудистой мозговой недостаточности и мозгового инсульта. Приложение "Инсульт" к журналу *Неврология и психиатрия*. 2001. Т. 1. С. 41–46.
3. Федеральная служба государственной статистики. 2009. URL: www.gks.ru.
4. Максудов Г.А. Дисциркуляторная энцефалопатия. В: *Сосудистые заболевания нервной системы*. Под ред. акад. АМН СССР Шмидта Е.В. М.: Медицина, 1975. С. 501–512.
5. Mancia G. Prevention and treatment of stroke in patients with hypertension. *Clin. Ther.* 2004. V. 26. № 5. P. 631–48.
6. Clifford C.P., Nunez D.J.R. Racial differences in the frequency of the restriction site polymorphism of the angiotensin converting enzyme (ACE) gene. In: *Abstracts 16th Sci. Meeting Internat. Soc. Hypertension (Glasgow, UK, 1996)*. Abstract 4B.2.0736.
7. Rigat B., Hubert C., Corvol P. et al. PCR detection of the insertion/deletion polymorphism of the human angiotensin converting enzyme gene (DCP1). *Nucl. Acids Res.* 1991. V. 20. P. 1433.
8. Скворцова В.И., Лимборская С.А., Кольцова Е.А., Сломинский П.А., Шетова И.М., Боцина А.Ю., Шамалов Н.А. *Неврология и психиатрия*. 2002. Т. 3. С. 47–51.
9. Rosei E.A. Hypertension, atherosclerosis and LVH in ELSA. *Blood Pressure*. 1996. V. 5. Suppl. 4. P. 44–47.
10. Ahn Y.I., Kamboh M.I., Hamman R.F. et al. Two DNA polymorphism in lipoprotein lipase gene and there associations with factors related to cardiovascular disease. *J. Lipid Res.* 1993. V. 34. P. 421–428.

11. Sparkes R.S., Zollman S., Klisak I. et al. Human genes involved in lipolysis of plasma lipoproteins: mapping of loci for lipoprotein lipase to 8p22 and hepatic lipase to 15q21. *Genomics*. 1987. V. 1. P. 138–144.
12. OMIM – On-line Mendelian Inheritance in Man. URL: <http://www.ncbi.nlm.nih.gov/omim>.
13. Chamberlain J.C., Thorn J.A., Morgan R. et al. Genetic variation at the lipoprotein lipase gene associated with coronary atherosclerosis. *Adv. Exp. Med. Biol.* 1990; 285:275-281
14. Mattu R.K, Needham E.W., Morgan R., Rees A., Hackshaw A.K., Stocks J., Elwood P.C., Galton D.J. DNA variants at the LPL gene locus associate with angiographically defined severity of atherosclerosis and serum lipoprotein levels in a Welsh population. *Arteriosclerosis and Thrombosis*. 1994. V. 14. № 7. P. 1090–1097.
15. Gambino R., Scaglione L., Alemanno N., Pagano G., Cassader M. Human lipoprotein lipase HindIII polymorphism in young patients with myocardial infarction. *Metabolism*. 1999. V. 48. № 9. P. 1157–1161.
16. Костомарова И.В., Водолагина Н.Н., Малыгина Н.А. и др. Связь полиморфизма генов липопротеинлипазы и белка-переносчика эфиров холестерина с продолжительностью жизни больных с хронической ишемией мозга. *Клиническая медицина*. 2008. №4. С. 22-28
17. Журавлёв Ю.И. Рязанов В.В. Сенько О.В. *Распознавание. Математические методы. Программная система. Применения*. Москва: Фазис, 2006.
18. Журавлёв Ю.И. Об алгебраическом подходе к решению задач распознавания или классификации. В: *Проблемы кибернетики*. М.: Наука, 1978. Вып. 33. С. 5–68.
19. Кузнецова А.В., Сенько О.В. Возможности использования методов Data Mining в медико-лабораторных исследованиях для выявления закономерностей в массивах данных. *Врач и информационные технологии*. 2005. № 1. С.49–56.
20. Senko O.V., Kuznetsova A.V. The Optimal Valid Partitioning Procedures, «InterStat», *Statistics in Internet*. URL: <http://interstat.statjournals.net/YEAR/2006/abstracts/0604002.php?Name=604002> (Abstract); <http://interstat.statjournals.net/YEAR/2006/articles/0604002.pdf>
21. Gordon D.J. High-density lipoprotein cholesterol and cardiovascular disease. Four prospective American studies. *Circulation*. 1989. V. 79. P. 8–15.
22. Оганов Р.Г., Шальнова С.А., Калинина А.М. *Профилактика сердечно-сосудистых заболеваний*. М., 2009, 216 с.
23. Franklin S.S., Larson M.G., Khan S.A et al. The Framingham Heart Study. *Circulation*. 2001. V. 103. № 9. P. 1245–1249.
24. Kannel W.P., Wolf P.A. et al. Risk stratification in hypertension: new insights from the Framingham Study. *Am. J. Hypertens*. 2000. V. 13. № 1. Suppl. 1. P. S3-S10.
25. Чазова Е.И., Сметник В.П. и др. *Системные гипертензии*. 2008. № 3. С. 27–29.

Материал поступил в редакцию 30.05.2011, опубликован 22.06.2011.