



Ю.Б.Котов, В.М.Гурьева, Т.А.Семенова

**Когнитивное исследование поведения
сложных систем на примере
медицинских объектов**

Рекомендуемая форма библиографической ссылки

Котов Ю.Б., Гурьева В.М., Семенова Т.А. Когнитивное исследование поведения сложных систем на примере медицинских объектов // Проектирование будущего. Проблемы цифровой реальности: труды 5-й Международной конференции (3-4 февраля 2022 г., Москва). — М.: ИПМ им. М.В.Келдыша, 2022. — С. 128-138. — <https://keldysh.ru/future/2022/11.pdf>
<https://doi.org/10.20948/future-2022-11>

Размещено также [видео выступления](#)

Когнитивное исследование поведения сложных систем на примере медицинских объектов

Ю.Б. Котов¹, В.М. Гурьева¹, Т.А. Семенова²

¹ *Московский областной научно-исследовательский институт акушерства и гинекологии*

² *Национальный исследовательский ядерный университет «МИФИ»*

Аннотация. В статье проведен краткий обзор математических методов, разработанных для исследования поведения открытых и неравновесных физических систем. Приведены некоторые результаты применения этих методов в области медицины. При обработке массивов медицинских данных используются непараметрическая и нечисловая статистика и трехзначная логика. Для сравнения нечисловых множеств авторами введена мера их близости, названная *неотличимостью*. Предлагаемый метод обработки данных в открытых системах – «метод логических симптомов» – обеспечен авторским комплексом вычислительных программ и может быть использован в других областях науки.

Ключевые слова: непараметрическая и нечисловая статистики, трехзначная логика, прототип (маска), логические симптомы, «неотличимость»

Cognitive investigation of the complex systems behavior using medical objects as an example

Yu.B. Kotov¹, V.M. Guryeva¹, S.T. Alekseevna²

¹ *Moscow Regional Research Institute of Obstetrics and Gynecology*

² *National Research Nuclear University MEPhI*

Abstract. The article gives a brief review of mathematical methods developed to study the behavior of open and nonequilibrium physical systems. Some results of these methods are used in medical field. Three-valued logic, nonparametric and nonnumerical statistics applied to medical data processing. To compare nonnumerical sets, the authors have introduced a measure of their proximity, called indistinguishability. The proposed data processing method in open systems, the “logical symptoms method”, is provided by the author's computer programs complex and can be used in other fields of science.

Keywords: nonparametric and nonnumerical statistics, three-digit logic, prototype (mask), logical symptoms, “indistinguishability”

Введение

Сложный объект состоит из многих автономных подсистем. Некорректно исследовать автономную подсистему отдельно, так как она испытывает влияние со стороны всей сложной системы и сама влияет на поведение системы в целом. Живые организмы и человек — это сложные системы, непрерывно общающиеся с внешней средой.

Врач умеет ставить диагноз пациенту и чем опытнее врач, тем точнее его диагноз. Однако важно не только поставить диагноз конкретному пациенту и вылечить его, но и передать этот опыт врачам следующего поколения. Не всегда врач может четко сформулировать, почему он поставил именно такой диагноз. Он перечислит разные признаки заболевания, многие из которых сочтет важными, а многие не назовет, хотя чувствует их интуитивно, но не может явно сформулировать.

Врач при изучении пациента использует результаты инструментальных измерений (клинических и биохимических анализов крови, мочи, данные об артериальном давлении и его динамике, биометрические сведения и т.д.). Это числовые данные. Но есть и другие данные: результаты внешних наблюдений, текстовое описание состояния организма типа «слишком твердая печень, мягкий живот, бледные покровы, сухая кожа, замедленная реакция, жалобы пациента» и т.д. Эти текстовые высказывания – не числовые данные, которые тоже важны в постановке окончательного диагноза. Полученный медицинский материал никогда не бывает полным, могут отсутствовать какие-то измерения.

В совместной работе врачей и математиков главная цель математика помочь врачу вербализовать его интуитивное знание, понять структуру собственного знания.

Методика исследования медицинского знания разработана математиками группы И.М. Гельфанда [1,2] на основе многолетнего опыта работы авторов с врачами разных специальностей. В этой статье кратко описаны некоторые методы обработки разнородных данных – метода логических симптомов – и приведены примеры их использования для целей медицинского прогнозирования.

Подготовка медицинского материала

При постановке медицинской задачи и решении ее большую роль играют диагностические игры [1,2]. Диагностическая игра — это диалог двух специалистов – врача и математика. Каждый из них решает свою профессиональную задачу. Врач видит картину в целом (гештальт-образ), математик изучает детали, структурные связи. В диалоге обсуждаются детали из истории болезни уже леченных пациентов. Диалог протоколируется и используется затем для извлечения необходимых сведений. В результате игры можно сформулировать термины и понятия, имеющие точный смысл.

Для поиска прогностических признаков конкретного заболевания врач в соответствии со своим опытом делит пациентов на классы: выделяет группы с тяжелой формой заболевания, с умеренной и легкой формами. Обязательно формирует контрольную группу – класс пациентов, не имевших этого заболевания («условно здоровых»). Все полученные разнородные (числовые, нечисловые и пробелы) сведения переносятся в базы данных.

Математик проводит формальную обработку собранного материала для выделения наиболее важных прогностических признаков заболевания и формулировки диагностических правил. Если при решении задачи обрабатываются только числовые параметры, используется только непараметрическая статистика. Как указано в белорусском руководстве по статистике [3]: «...в реальности нормальное распределение данных в биомедицинских исследованиях встречается чуть чаще, чем никогда, ввиду чего непараметрические методы составляют основу математического аппарата медицинского статистика» (выделено нами). В случаях (их большинство), когда в диагностике использованы разнородные параметры – и числовые, и не числовые – мы применяем нечисловую статистику [4]. В этом случае невозможны привычные математические операции сложения, умножения и т.д. Все переменные кодируются как текстовые. Поскольку в медицине обычно система данных не полна, есть пробелы (пропуски), при обработке данных применяется слегка модифицированная трехзначная логика Лукасевича [2,5].

Нормативы процессов

Приведем сначала примеры методики работы с числовым материалом. Любые сложные системы, особенно живые организмы, непрерывно изменяют свое состояние, подстраиваясь под внешнюю среду. Динамика процесса особенно важна при лечении организма. Для того чтобы выделить особенности болезненного поведения организма, нужно знать, по каким правилам развивается здоровый организм. Для этого по контингенту пациентов, не имеющих исследуемой болезни, строятся зависимости медицинских параметров от времени – динамические модели процессов.

Например, в задаче о раннем прогнозе задержки внутриутробного развития плода (ЗВУР) [2] проводились исследования беременных женщины в возрасте от 16 до 40 лет. Были использованы результаты ультразвуковой плацентографии и биометрии плода у беременных женщин, концентрации гормонов в крови беременных (плацентарных лактогена, прогестерона и эстрадиола), записи частоты сокращений сердца плода и т.д. Тенденция нормального роста или отставания плода проявляется достаточно рано и, если нет лечения, сохраняется до рождения ребенка. На рис. 1 приведен пример норматива для концентрации прогестерона в крови беременных в зависимости от срока беременности (гестационного срока в неде-

лях). Числа рядом с каждой кривой указывают перцентильные значения квантилей: числом 50 отмечена медиана, числами 25 и 75 – кривые, соответствующие нижнему и верхнему квартилям. На каждой неделе беременности у пациенток с нормальным развитием плода были построены распределения концентрации прогестерона. Приведенные на рис. 1 гладкие нормативы построены по квантилям этих распределений.

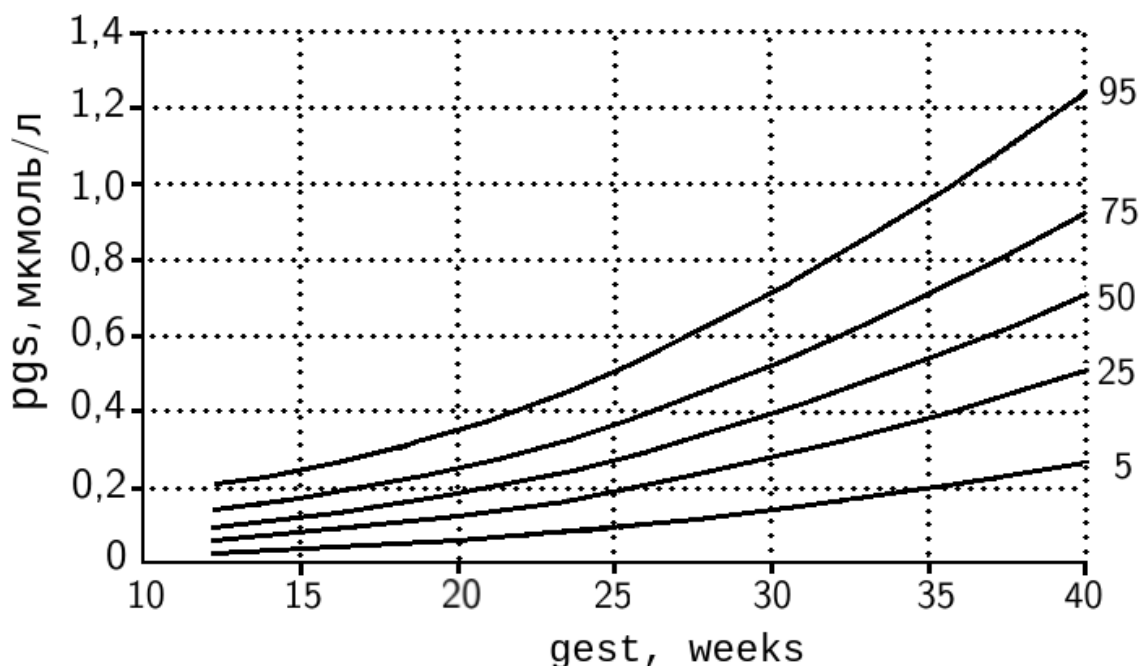


Рис. 1. Зависимость концентрации прогестерона от срока беременности

Недавние работы авторов [6,7] посвящены поиску ранних диагностических признаков тяжелого осложнения беременности – преэклампсии (ПЭ). Силами врачей МОНИИАГ проводился в 2019-21 гг. скриннинг беременных женщин Московской области. Исследовалась связь между динамикой плацентарных белков SFLT-1 и PLGF и развитием преэклампсии у беременных. На рис. 2 представлены графики зависимости этих белков от гестационного срока у женщин без преэклампсии (ПЭ). Белок PLGF (placental growth factor) стимулирует рост в плаценте матери кровеносных сосудов, питающих плод. Концентрация белка начинается с первых дней беременности, достигает максимума к 28-30 неделям и затем убывает вплоть до родоразрешения. Белок sFLT-1 (soluble fms-like tyrosine kinase 1) очень медленно нарастает примерно до 28 недели, а затем рост заметно ускоряется. Этот белок является ингибитором роста кровеносных сосудов.

На рис. 2 прозрачные прямоугольники указывают квартильные границы двухнедельных распределений, черные пояски – медианы распределений. Сплошная линия – аппроксимация медиан (*med*), пунктирные – аппроксимация соответственно нижних (q_1) и верхних (q_2) квартилей.

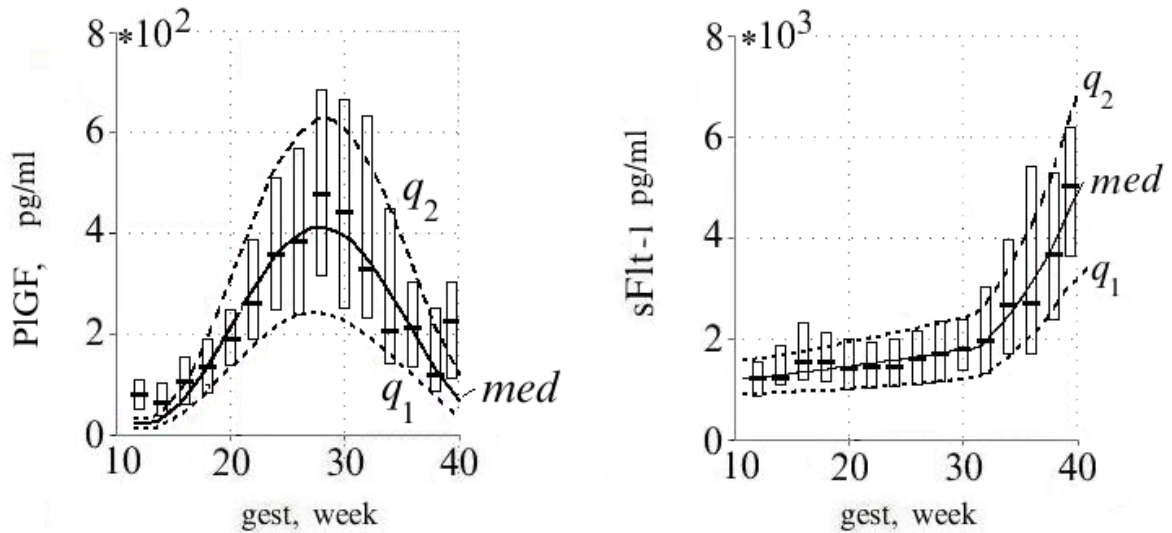


Рис. 2. Зависимость концентрации белков PIGF и sFLT-1 от срока беременности

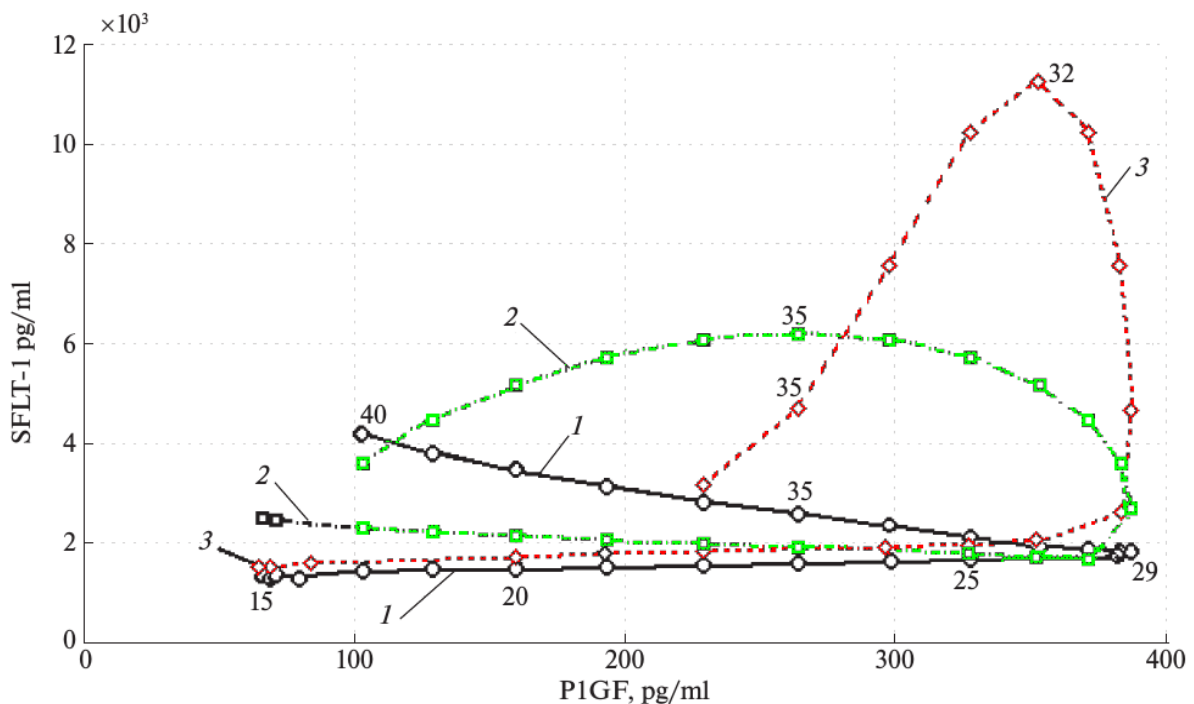


Рис. 3. Модель взаимосвязи белков PIGF и sFLT-1 от срока беременности: кривая 1 – пациентки без ПЭ, 2 – с умеренной формой ПЭ, 3 – тяжелая форма ПЭ

На рис. 3 приведена модель динамического взаимодействия белков PIGF и sFLT-1 у пациенток без ПЭ, с умеренной формой ПЭ и тяжелой формой заболевания. По оси абсцисс отложены медианы концентрации белка PIGF, а по оси ординат – медианы концентрации белка sFLT-1. Сро-

ки беременности в неделях проставлены прямо на кривых вблизи экспериментальных точек. Кривая 1 получена для пациенток без ПЭ: от начала обследования (с 14 недели) и примерно до 29 недели беременности рост концентрации белка PIGF сопровождается медленным ростом концентрации sFLT-1. После 29-й недели концентрация PIGF постепенно уменьшается, а концентрация sFLT-1 продолжает расти ускоренными темпами, что соответствует рис. 2. Кривая 2 описывает развитие процессов у пациенток с умеренной формой ПЭ. Кривая 3 – динамику процессов у пациенток с тяжелой формой ПЭ.

Из рис. 3 хорошо видно, как резко после 29 недели меняется роль белка sFLT-1, подавляющего рост кровеносных сосудов и ухудшающего качество питания плода. Однако в поведении этих белков проявляется некоторое различие и на ранних стадиях беременности. Исследование динамики параметров PIGF и sFLT-1 в начале второго триместра беременности позволило нам выделить признаки риска развития ПЭ [8].

Метод логических симптомов

Метод представляет собой единую процедуру обработки массивов с данными разной природы (числовыми, текстовыми и пробелами).

1. Логические симптомы. Вектор логических симптомов.

Логическим симптомом называется утверждение о некотором свойстве организма пациента или о некоторой особенности процесса лечения. Для обработки в компьютере это утверждение может быть закодировано символом.

Подбор симптомов по базе данных в вектор осуществляет математик, обсуждая с врачами значимость того или иного параметра.

Примеры записи симптомов:

```
@symptom 005(Гипертоническая болезнь 2 стд., гест. срок)  
$ HTON=="Y" && gest >=24 && gest <= 28
```

```
@symptom 010(Сахарный диабет, гиперт. болезнь тяж ст.)  
$ DIA == "Y" && HYPTR == "Y"
```

```
@symptom 021(Концентр. белка SFLT_1 в крови матери)  
$ SFLT_1 >=94
```

```
@symptom 101 ( ) $ #005 && #010 && #021
```

После заголовка стоит номер симптома, далее в круглых скобках комментарий, затем (после знака доллара \$) кодировка симптома и его значение. Например в симптоме 005 диагнозу «гипертоническая болезнь 2 стадии» соответствует код HTON. Его значение равно «Y» – да (yes). Вторая запись после логического «и» представляет собой неравенство, указывающее интервал беременности в неделях, в которых поставлен диагноз «гипертоническая болезнь второй стадии». Симптом может содержать од-

но или несколько утверждений, соединенных логическими операциями. В симптоме 101, например, объединены три ранее вычисленных симптома.

Вся совокупность симптомов, принадлежащих одному j -му пациенту, образует вектор (строку). Длина n вектора логических симптомов (его размерность) и их порядок следования одинаковы в данной задаче у всех пациентов. Длина векторов равна полному числу симптомов, представленных в медицинском материале: $F_j = \{x_{j1}, x_{j2}, x_{j3}, x_{j4}, \dots, x_{jn}\}$.

Вектор логических симптомов может содержать свыше сотни симптомов в зависимости от целей поставленной медицинской задачи.

2. Построение маски (образца, прототипа) совокупности объектов.

Рассмотрим некоторое множество пациентов, разделенных по степени тяжести заболевания на 4 класса: Z – пациенты без заболевания, A – с легкой формой болезни, B – с умеренной формой, C – с тяжелой формой заболевания. Пусть мы хотим выделить пациентов самого тяжелого класса C на фоне остальных объединенных классов $C : ZAB$ (читается: C против объединения классов Z, A, B). Создадим маску (образец, прототип) $C : ZAB$ – новый логический вектор, с которым будем сравнивать векторы логических симптомов остальных векторов. Технология подбора маски подробно описана в работе [9]. Маска используется для проверки принадлежности пациентов предварительно сформированным классам.

3. Неотличимость как мера близости двух множеств.

Для сравнения векторов логических симптомов, построенных в пространстве смешанных данных (числовых, текстовых и пробелов), вводим меру близости двух множеств. Эту меру мы назвали неотличимостью двух векторов в пространстве трехзначной логики.

Неотличимостью “ u ” двух симптомных векторов называется отношение числа координат, совпадающих и имеющих значения «Х» (верно) или “=” (неверно), к полному числу координат «верных» и «неверных».

Координаты со значениями «.» (неизвестно) в расчете « u » не участвуют.

Неотличимость изменяется в пределах $0 \leq u \leq 1$.

При $u = 1$ вектор логических симптомов полностью совпадает с маской. На рис. 4 приведен пример вычисления неотличимости двух фрагментов векторов. Сравниваются фрагменты векторов двух объектов A и B . В третьей сверху строке галочкой отмечены координаты двух верхних векторов, в которых совпадают значения обоих симптомов. Они могут иметь значения «верно» или «неверно», но известны и одинаковы у обоих пациентов. Таких элементов (координат) всего 4. В четвертой строке сверху указаны галочками все координаты, известные у обоих пациентов A и B , но не обязательно совпадающие. Неизвестные значения (изображенные точками) не учитываются. Таких элементов 8. Их отношение – неотличимость равна 0.50.

Пример расчета неотличимости для объектов А и В

объект А	= X = . . . = . X X = X X X	Σ
объект В	X X = = X = . . . X X = =	
совпадения	V V V V	4
Известные значения в обоих векторах	V V V V V V V V	8

Неотличимость равна

$$u = 4/8 = 0.50$$

“X” – положительный результат,
“=” – отрицательный, “.” неизвестный.

Рис. 4. Пример расчета неотличимости двух векторов

4. Работа с матрицей, построенной из векторов логических симптомов.

На рис. 5 приведен фрагмент матрицы, элементами которой являются координаты (логические симптомы) векторов каждого пациента. В этом примере контингент пациентов был разбит на 4 класса. Построена маска класса С (верхняя строка рисунка) – пациентов с самой тяжелой формой заболевания – и сравнивается множество симптомов пациентов этого класса с объединенным множеством симптомов остальных классов (т.е. класса Z пациентов, не имеющих этого заболевания и классов А и В с более легкими формами заболевания). Это указано в левом нижнем углу рис. 5 формулой С: ZAB (С против объединенных классов Z, А,В). Маска записана над матрицей на рис. 5. Два левых столбца указывают класс и идентификационный номер пациента. Далее для каждого пациента приведены значения логических симптомов: симптом верен (знак «X»), неверен (знак «=»), неизвестен (знак «.»).

Три нижних цифровых строки указывают номера симптомов в общем списке исходных логических симптомов (читаются номера сверху вниз). Самый правый столбец в вертикальной рамке содержит значения неотличимости *u* вектора каждого пациента от маски. Значения неотличимости, близкие к единице (например, у пациентов С 181, 182, 183) говорят о том, что эти пациенты правильно отнесены к классу С тяжелых больных. А пациент С 389 далек от своего класса. Это может быть связано и с тем, что он плохо обследован, в строке его симптомов слишком много неизвестных («точек»). После такой формальной обработки начинается содержательное исследование – выяснение медицинского смысла тех симптомов, которые обеспечили самую высокую неотличимость между пациентом и маской, и формулировка диагностического правила.

фрагменты маски и матрицы

для 4-х классов пациентов: Z – здоровые, А – с легкой формой заболевания,
В – с более тяжелой, С – с самой тяжелой формой заболевания

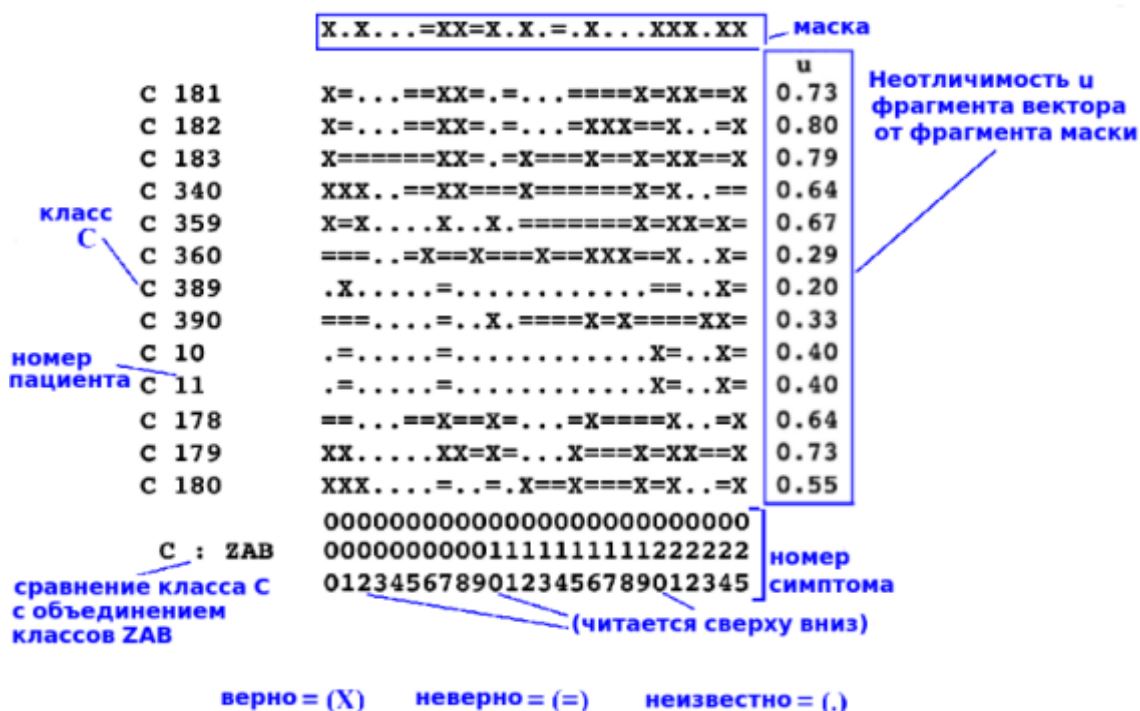


Рис. 5. Пример фрагментов маски и матрицы симптомов

На рис. 6 приведена таблица из работы [10] с результатами анализа состояния плаценты, подтверждающая врачебный диагноз и ход лечения пациенток с хронической болезнью почек (ХБП). Контингент пациенток был разбит на 5 классов: класс Z – без ХБП, А, В, С – классы с ХБП от легкой формы до более тяжелых В и С, и класс D с самой тяжелой формой ХБП. Численность пациенток в классах указана во второй строке сверху. Обратим внимание, что почти все симптомы, указанные в таблице, – текстовые утверждения. В самом тяжелом классе D все пациентки имели артериальную гипертензию и преждевременное созревание ворсин в плаценте. Почти у половины пациенток этого класса наблюдалась тяжелая преэклампсия. В последней строке таблицы приведены результаты измерения суммарной площади поперечного сечения кровеносных сосудов в ворсинах плаценты (просветов в срезах ворсин).

Чем меньше суммарная площадь просветов, тем меньше крови может попасть от ворсин в пуповину плода, тем хуже питание плода. Это одна из возможных причин задержки внутриутробного развития плода (ЗВУР). В классах с более тяжелой формой заболевания ХБП (В, С и D) площадь просветов в 4 и более раз меньше, чем у пациенток без ХБП.

Пример: выделение плацентарных симптомов у беременных с ХБП

Номер симп.	Название симптома	Класс	Z	A	B	C	D
		Число пациенток	13	70	14	22	5
028	Облитерационная ангиопатия		8%	-	-	-	-
034	Межворсиночная гематома		31%	-	-	-	-
030	Многочисленные синцитиальные узлы		-	37%	-	-	-
041	Гепаринотерапия		-	41%	-	-	-
033	Патология базальной пластинки		-	-	14%	9%	-
035	Инфаркт плаценты		-	-	22%	5%	-
032	Фибриноид распростр. Псевдоинфаркты		-	-	28%	27%	-
020	Плацента с ободком, экстрахориальная		-	-	-	32%	-
038	Артериальная гипертензия		-	17%	64%	70%	100%
025	Преждевременное созревание ворсин		-	3%	-	-	100%
040	Тяжелая проэклампсия		-	-	-	-	40%
S	Суммарная площадь поперечного сечения просветов ворсин в плаценте, 10^3 мкм ²		9	3	1.5	2	1.7

Рис. 6. Таблица плацентарных диагнозов у пациенток с хронической болезнью почек

Заключение

Использование метода логических симптомов в совокупности с непараметрической и нечисловой статистикой позволяет решать важные когнитивные задачи, выделить и вербализовать наиболее важную информацию, которой интуитивно пользуются опытные специалисты. Описанная стадия выделения наиболее важных признаков заболеваний, создание маски, определение меры близости сравниваемых множеств (неотличимости объектов) являются начальной стадией исследования – стадией обучения. На этой стадии удастся сформулировать диагностические правила для конкретной задачи. В сложных случаях всю процедуру, начиная от игр, пересмотр деления на классы, построение масок и формулировку правил приходится повторять неоднократно.

После окончательной формулировки правила оно тестируется на новых контингентах пациентов.

Литература

1. Гельфанд И.М., Розенфельд Б.И., Шифрин М.А. Очерки о совместной работе математиков и врачей. – М.: Наука, 1989. – 272 с.
2. Котов Ю.Б. Новые математические подходы к задачам медицинской диагностики. – М.: Едиториал УРСС, 2004. – 328 с.

3. Жильцов И.В., Семенов В.М., Зенкова С.К. Основы медицинской статистики, Дизайн биомедицинских исследований. Практическое руководство. – Витебск, ВГМУ, 2014. – 154 с.
4. Орлов А.И. Нечисловая статистика. – М.: МЗ-Пресс, 2004. – 513 с.
5. Карпенко А.С. Логика Лукасевича и простые числа. – М.: Наука, 2000. – 319 с.
6. Гурьева В.М., Котов Ю.Б. и др. Математический анализ результатов скрининга пациенток в целях прогнозирования преэклампсии // [Вестник НИЯУ МИФИ 9\(6\), 567–576 \(2020\)](#).
7. Гурьева В.М., Котов Ю.Б. и др. Математическая модель взаимосвязи между sFLT-1 и PlGF при преэклампсии // [Вестник НИЯУ МИФИ 10\(2\), 129–134 \(2021\)](#).
8. Гурьева В.М., Котов Ю.Б. и др. Математическое выделение ранних прогностических признаков преэклампсии // [Вестник НИЯУ МИФИ 10\(4\), 338–348 \(2021\)](#).
9. Гурьева В.М., Котов Ю.Б. и др. Технология подбора маски в методе логических симптомов // [Вестник НИЯУ МИФИ 6\(4\), 368–376 \(2017\)](#).
10. Котов Ю.Б., Баринова И.В. и др. Статистический анализ связи между состоянием плаценты и хронической болезнью почек // [Вестник НИЯУ МИФИ 9\(3\), 262–269 \(2020\)](#).