

В.В. Дроботкин, И.С. Машоткин, С.В. Краско, Л.И. Зинило,  
**ПРИМЕНЕНИЕ РАЗВЕДОЧНЫХ МЕТОДОВ АНАЛИЗА ДАННЫХ  
В ДИФФЕРЕНЦИАЛЬНОЙ ДИАГНОСТИКЕ ПОДТИПОВ  
ТРАНЗИТОРНЫХ ИШЕМИЧЕСКИХ АТАК**

<sup>1</sup>Белорусский государственный медицинский университет,

<sup>2</sup>Республиканский научно-практический центр неврологии и нейрохирургии,

<sup>3</sup>Объединенный Институт Проблем Информатики НАН Беларуси

Понятие «разведочные методы анализа данных» еще достаточно ново в медицинских исследованиях и не получило широкого распространения, а между тем, в технических и биологических исследованиях оно уже прочно обосновалось и используется как термин, обозначающий методы анализа для «добычи данных» и решения различных классификационных задач. В рамках этого понятия в наших работах мы применяем факторный и дискриминантный анализы (ФА и ДА) [1, 2, 4-9]. Дискриминантный анализ, применительно к нашим исследованиям, мы рассматриваем как «цифровую дифференциальную диагностику» подтипов транзиторных ишемических атак (ТИА) с учетом гетерогенности этой нозологической единицы [5-7]. Мы стараемся избегать понятия «вариант» так, как оно, по нашему мнению, излишне жестко, и не позволяет учитывать своеобразие дифференциальной диагностики цереброваскулярной патологии. Например, если констатируется первый вариант течения заболевания, то уже второй и последующие исключаются из рассмотрения, первый – значит не второй и не третий. В наших же исследованиях акцент делается на наибольшую вероятность одного вида (подтипа) возникновения и течения ТИА не исключая вероятности присутствия и других подтипов, более того, допуская их одновременное присутствие. Речь, таким образом, идет о преимущественном лидирующем положении одного из них. В этом отношении мы полностью разделяем мнение, приведенное в публикациях сотрудников НИИ неврологии РАМН по подтипам ишемического инсульта [15-19].

Абстрактно-математически варианты какой-либо патологии – это непересекающиеся множества с четкими границами и четким отграничением друг от друга, что реальных условиях далеко не всегда имеет место. Подтипы ТИА, по нашему мнению, в этом аспекте скорее «пушистые множества», «множества с нечеткими границами», “fuzzy sets” [12-14, 25]. Поэтому в наших исследованиях мы пользуемся понятием «подтип ТИА», а не «вариант ТИА». Так, этиопатогенетически гемодинамически значимые атероматозные изменения стенок кровеносных сосудов, кардиозмболия и артериальная гипертензия «беспроblemно уживаются» друг с другом и могут сосуществовать в клинической картине одного пациента, но одно из этих проявлений, как правило, лидирует.

С подобной проблемой нередко приходится сталкиваться и при тестировании, когда из трех-четырех вариантов ответа на вопрос правильными могут быть два или даже все варианты.

Такую ситуацию можно смоделировать на конкретном примере, если необходимо ответить на вопрос, какой фактор риска создает условия и определяет возникновение эпизода ТИА:

- 1) Атеросклеротические изменения стенки сосудов?
- 2) Нарушение сердечной деятельности (мерцательная аритмия)?
- 3) Артериальная гипертензия?
- 4) Экстравазальное стенозирование сосуда?

На клиничко-интуитивном уровне понятно, что все эти факторы риска в той или иной степени являют-

## ☆ Лечебно-профилактические вопросы

ся и предпосылкой и причиной ТИА. Положительная корреляция между ТИА и каждым из этих факторов не вызывает сомнения. Можно удовлетвориться этой констатацией. Но это мало что дает для проведения конкретных целенаправленных превентивных профилактических мероприятий. И совершенно ничего не дает для их унификации. Каждый врач волен по-своему характеризовать эти факторы и истолковывать их приоритетность, и исходя из своих личных представлений и предпочтений, назначать лечебно-профилактические мероприятия. Бесспорно, и многолетний опыт и профессиональные знания определяют правильность и эффективность решения. К тому же применение традиционной проверки статистических гипотез (например, по  $t$ -критерию Стьюдента и значению  $p < 0,05$ ) вселяет определенную долю уверенности в правильности выбранного решения и чувство удовлетворенности от осознания его непогрешимости, которое далеко не всегда оправдано. Многолетний опыт клинической работы во всех странах и различных научных школах неоспоримо показывает недостаточность такого подхода к распознаванию истинного этиопатогенеза заболевания у каждого конкретного пациента. Особенно демонстративны сомнения и противоречия между истинной клинико-терапевтической ситуацией и сложившимся представлением на базе только клинического опыта и мышления проявляется в результатах межгосударственных многоцентровых рандомизированных контролируемых исследований (РКИ). Но проведение РКИ осуществляется «где-то там», данные еще долго будут обрабатываться и затем наконец-то будут представлены в систематическом обзоре и других публикациях. Оптимальное же решение безотлагательно необходимо принимать здесь и сейчас. В этом отношении дополнительным, а главное, значительно более доступным, чем проведение дорогостоящих и трудоемких РКИ, обладающим инструментом распознавания может быть разведочный анализ данных (РАД)\* [1-9, 11-18, 20-26]. Доказательность становится доступнее. Располагая собственными наблюдениями, базой данных и необходимой оргтехникой врач на своем рабочем месте может принимать решения, удовлетворяющие понятию «доказательная медицина». Начало такого направления математического обоснования клинико-терапевтических решений по имеющимся клиническим и лабораторным данным пациента было положено еще в 70-х годах прошлого столетия на кафедре нервных болезней Минского медицинского института и Институте технической кибернетики (ИТК) АН БССР [1, 4, 9]. Ряд публикаций тех лет приведен в сборнике «Кибернетика в неврологии и психиатрии. Мн., 1976».

В 2002 году ИТК был переименован в Объединенный институт проблем информатики (ОИПИ) НАН Беларуси, где продолжают проводиться исследования по применению информационных технологий, РАД, теории распознавания образов в медицине. В контексте статьи наиболее интересны некоторые из них.

Построение регрессионных моделей в медицине очень часто сопровождается нарушением корректности их составления, что приводит к искажению интерпретации результатов. Чтобы избежать возможных некорректностей предлагается алгоритм, использующий специфику «температурного кодирования», когда один параметр (качественный или порядковый) заменяется на несколько бинарных параметров. Этот метод предложен для исследования зависимости раннего ответа на лечение и выхода на ремиссию больных детей с различными типами лейкозов. По построенным интерпретируемым регрессионным моделям прогнозируется результат (выход на ремиссию) проводимого лечения [3].

Буквально на каждом шагу принятие решений осуществляется в условиях неполноты и нечеткости исходных данных. Для таких ситуаций создана система разработки компактного комплекса правил для решения классификационных задач на базе нечеткой нейросетевой модели. Принятие решений в ситуациях с нечеткими и неполными исходными данными особенно актуально в медицинской диагностике, бизнесе, военном деле [11].

Изображения сосудистого рисунка на традиционных ангиограммах нередко расплывчаты и «смазаны». Предложен алгоритм, позволяющий значительно улучшить качество изображения сосудистого рисунка на нечетких полутонных ангиограммах сосудов мозга. Алгоритм может быть реализован на компьютере с минимальной конфигурацией на базе процессора Celeron 1300 [10].

Для случаев, когда необходимо интерактивное общение с пациентом, создана информационно-аналитическая система с выходом в Интернет, которая предназначена для решения классификационных и прогностических задач при лейкомиях. Она состоит из пяти частей: базы данных и четырех Веб-приложений. Общение с пациентом в Интернете осуществляется по схеме «Пациент – Веб-сервер – сервер базы данных». Проводится анализ прогностически важных факторов риска и особо выделяется группа пациентов с повышенным риском исхода заболевания. Правильность прогноза исхода лечения у таких больных варьирует в пределах 75-97% [27].

Эти и многие другие широкопрофильные алгоритмы и методы могут быть использованы для решения конкретных классификационных и прогностических задач в ангионеврологии, в частности, для определения характера и прогноза возникновения эпизода ТИА.

В настоящее время дальнейшее развитие такого подхода предполагает и продолжает вторжение в область прикладного использования разработок по искусственному интеллекту [3, 11-18, 21-26].

Прикладной медицинский аспект анализа данных в настоящее время успешно развивается и применяется в различных отраслях знаний и практи-

\* В англоязычной литературе используется термин *Exploratory Data Analysis (EDA)*.

Содержательное клиническое описание математически выделенных факторов

Порядковый номер фактора	Предметное содержательное клиническое объяснение фактора	Вклад фактора в %
1 фактор (генеральный g-фактор)	Фактор неотвратимых возрастных изменений (возраст, шейный остеохондроз, возраст с которого начали беспокоить головные боли, головокружения). Генеральный фактор в объяснении этиопатогенеза ТИА.	19.4
2 фактор	Фактор табакокурения: возраста начала и количество сигарет	9.5
3 фактор	Конфликты и перемены места жительства. Фактор служебно-бытовых неудобств, информационного дискомфорта	7.9
4 фактор	Фактор артериальной гипертензии	7.6
5 фактор	Фактор, обусловленный переменой профессии	5.8
6 фактор	Фактор нарушенного сна, бессонница	5.6
7 фактор	Фактор зрительных нарушений, при наличии эпизодов ТИА	5.0
8 фактор	Метеозависимость	4.9
9 фактор	Снижение остроты зрения, при наличии эпизодов ТИА	4.1
Всего объяснено ортогональными факторами всей клинической информации, приведенной 25 показателями.		70.0%

ческой работы для нахождения систематических связей между переменными (в медицинском аспекте между симптомами, показателями, признаками) в ситуациях, когда отсутствуют (или имеются недостаточные) априорные представления о природе этих связей [11-14, 21].

Наш скромный опыт по применению разведочных методов ДА и ФА нам представляется вполне обнадеживающим и позволяющим и в дальнейшем получать необходимые данные нетрадиционным для ангионеврологической клиники путем [2, 4-9].

Применение ДА позволило нам констатировать и математически обосновать следующие положения особенностей ТИА.

- ТИА – это гетерогенная нозологическая единица, включающая в себя несколько подтипов, что предопределяет неоднозначный подход в проведении лечебно-профилактических мероприятий по их предотвращению.

- Этиопатогенетическая сложность происхождения подтипов ТИА с учетом вовлеченности различных систем организма в патологический процесс, формально-математически может быть представлена как пересекающиеся, и в то же время, достаточно обособленные множества, существующие в соответствующем признаковом пространстве.

- Применение дискриминантного анализа дает возможность математически прогнозировать континуум динамического развития вектора начальных проявлений ишемического поражения мозга по одному из изначально исходного подтипа ТИА.

- Зная наиболее вероятный вектор развития патологического процесса у каждого конкретного обследуемого пациента можно давать более конкрет-

ные, адресные лечебно-профилактические рекомендации, обусловленные своеобразием и особенностями соответствующего подтипа ТИА, акцентируя, тем самым, внимание на профилактическое и терапевтическое воздействие на наиболее выраженный этиопатогенетический компонент [4-9].

Применение разведочного метода ФА позволило конкретизировать и проранжировать в порядке убывания факторы риска и основные признаки ТИА [8]. Результаты приведены в таблице.

Многочисленные литературные данные и результаты наших исследований окончательно убеждают в целесообразности использования в ангионеврологии разведочных методов «добычи данных» в решении дифференциально-диагностических задач, как составного компонента прикладных разработок в области создания искусственного интеллекта.

#### Литература

1. Дривотинов Б. В., Лупьян Я.А. Прогнозирование и диагностика дискогенного пояснично-крестцового радикулита. -Мн.-1982.
2. Дривотинов Б.В., Мاستыкин А.С. Математика и вычислительная техника в неврологии. Бел. мед. журн. 2004. № 4.-С. 50 — 52.
3. Красько О.В., Новоселова Н.А., Том И.Э. "Температурное" кодирование переменных в нейросетевых регрессионных моделях. // Искусственный интеллект.-2006.-№ 2.-С. 380-383.
4. Мастыкин А.С., Семан А.Е., Рябцева Т.Д., Гришков Е.Г. Дискриминантный анализ при прогнозировании возникновения ишемического инсульта. В сб.: Кибернетика в неврологии. /Под ред. Н.С. Мисюка, Мн.-1976.-С. 79-83.
5. Мастыкин А.С., Дривотинов Б.В., Апанель Е.Н. Гетерогенность нозологического понятия транзиторной ишемической атаки. // Бел. мед. журн. 2004. № 1. С. 18 — 21.
6. Мастыкин А.С., Апанель Е.Н., Антонов И.П. Применение дискриминантного анализа для выделения подтипов (моделей) транзиторных ишемических атак. // Мед. новости.-2004.-№ 7. - С. 95-98.
7. Мастыкин А.С., Дривотинов Б.В., Апанель Е.Н. Вероятностно-статистический метод в изучении гетерогенности транзиторных ишемических атак.// Вестн НАН Беларуси. Сер. мед. наук 2005, № 1.-С. 76-82

## ☆ Лечебно-профилактические вопросы

8. Мاستыкин А.С., Апанель Е.Н., Дривотинов Б.В., Антонов И.П. Факторный анализ признаков, симптомов и факторов риска транзиторной ишемической атаки. // Весті НАН Беларусі. Сер. мед. навук. 2005. № 3. С. 31—36.
9. Мисюк Н.С., Баранова А.М., Мастыкин, А.С., Фришков Е.Г., Крылова Е.А. Факторный анализ в изучении системы свертывания крови при тромботических инсультах. // Клиническая медицина. 1977. № 9. С. 65—72.
10. Недзведь А.М., Абламейко С.В., Белоцерковский А.М., Лемман Т. Быстрый метод выделения сети сосудов на полутоновых ангиографических изображениях. // Искусственный интеллект.-2006.-№ 2.-С. 206-210.
11. Новоселова Н.А., Том И.Э., Красько О.В. Нечеткое нейросетевое моделирование для получения интерпретируемого набора классифицирующих правил // Искусственный интеллект.-2006.-№ 2.-С. 211-214.
12. Орловский С.А. Проблемы принятия решений при нечеткой исходной информации. М.: Наука, 1981.
13. Основы теории нечетких множеств. Метод. указания / Сост. И.Л. Коробова, И.А. Дьяков Тамбов: Изд-во Тамбовского гос. техн. ун-та. 2003.
14. Прикладные нечеткие системы / Под ред. Тэтано Т., Асаи К., Сугано М: Мир, 1993.
15. Реброва О.Ю. Статистический анализ медицинских данных. М, 2003.
16. Реброва О.Ю., Максимова М.Ю., Пирадов М.А. Нейросетевой алгоритм диагностики патогенетических подтипов ишемического инсульта // Журнал неврологии и психиатрии им. С.С. Корсакова. Инсульт (Приложение), 2004, вып. 12, с. 23-28.
17. Реброва О.Ю. Информатика в ангионеврологии // В кн. «Очерки ангионеврологии». Под ред. члена-корр. РАМН З.А. Суслиной М.: Атмосфера, 2005, с. 346-359
18. Реброва О.Ю. Применение методов интеллектуального анализа данных для решения задачи медицинской диагностики // Новости искусственного интеллекта, 2004, №3, с. 76-80
19. Суслина З.А., Верещагин Н.В., Пирадов М.А. Подтипы ишемических нарушений мозгового кровообращения: диагностика и лечение // Consilium Medicum.-2001.-Том 3.- № 5.-С.-218-221.
20. Татаринцев П.Б. Разработка систем диагностики, дифференциальной диагностики и прогнозирования заболеваний методами многомерного статистического анализа.- Автореф. канд. дисс. Барнаул, 2006.
21. Aslanyan S, Weir CJ, McInnes GT, Reid JL, Walters MR, Lees KR. Statin administration prior to ischaemic stroke onset and survival: exploratory evidence from matched treatment-control study. Eur J Neurol. 2005.-Vol. 12.-№ 7.-P 493-498.
22. Desrosiers J, Noreau L, Rochette A, Bravo G, Boutin C. Predictors of handicap situations following post-stroke rehabilitation. Disabil Rehabil. 2002.-Vol. 24.-№ 15.-P 774-785.
23. Floor C, Bakker MA, Catharina J.M. et al. Cognitive Impairment Is Related to Cerebral Lactate in Patients With Carotid Artery Occlusion and Ipsilateral Transient Ischemic Attacks Stroke. 2003.-Vol.34.-P 1419.
24. Lyden P, Claesson L, Havstad S, Ashwood T, Lu M. Factor analysis of the National Institutes of Health Stroke Scale in patients with large strokes. Arch Neurol. 2004 Vol. 61.-№11.-P 1677-1680.
25. Sinacore JM, Chang RW, Falconer J. Seeing the forest despite the trees. The benefit of exploratory data analysis to program evaluation research. Eval. Health. Prof. 1992.-Vol. 15.-№2.-P 131-146.
26. Peterson M.G.E. -Intelligent Medical Systems and the Interface with Statistics-, p. 300, 11<sup>th</sup> IEEE Symposium on Computer-Based Medical Systems (CBMS'98), 1998.
27. Tom I.E., Krasko O.V., Новоселова Н.А., Выданов О.И. Experience of the medical-oriented information-analytical system's development on a basis of the three-tier architecture // Искусственный интеллект.-2006.-№ 2.-С. 462-362.

РЕПОЗИТОРИЙ БГУИМ