

Н.А. Новоселова, Е.Н. Апанель, Б.В. Дривотинов, А.С. Мاستыкин
**«Консилиум» адаптивных нейросетевых классификаторов для
дифференциальной диагностики подтипов транзиторных
ишемических атак**

*Объединенный Институт Проблем Информатики НАН Беларуси,
Республиканский научно-практический центр неврологии и нейрохирургии,
Белорусский государственный медицинский университет*

Ангионеврологические заболевания обладают одной особенностью: по жалобам пациента и симптомам они маскируются один под другого. Примечательны в этом отношении симптомокомплексы акустической душевной глухоты, доброкачественного пароксизмального головокружения, острого вестибулярного головокружения. Пациентам с этими заболеваниями ошибочно ставится диагноз острого нарушения мозгового кровообращения и они не получают лечения, адекватного заболеванию. Действительно, дифференциальная диагностика этих патологических состояний с лидирующим симптомом головокружения крайне затруднительна, они нередко прячутся под маской острого нарушения мозгового кровообращения [11].

Кроме того, в этом дифференциально-диагностическом контексте не следует упускать из внимания антифосфолипидный синдром (АФС, антифосфолипидные АТ, синдром Хьюджа), который также может прятаться под маской нарушений мозгового кровообращения и предстать как одно из неврологических проявлений транзиторной ишемической атаки (ТИА) [12].

К этому же списку трудно диагностируемых состояний следует добавить еще патологические изменения в кровеносных сосудах, связанные с дисфункцией эндотелия и активностью закиси азота (NO). В этом случае проблема дифференциальной диагностики ТИА с другими сердечно-сосудистыми заболеваниями становится практически неразрешимой традиционными клинико-лабораторными методами. В этой связи, в частности, следует упомянуть о сравнительно недавно открытом маркере сердечно-сосудистой патологии – асимметричном диметиларгinine (АДМА). АДМА является эндогенным ингибитором NO-синтазы, участвует в патогенезе атеросклероза и является слабым независимым предиктором острого инсульта и сильным предиктором ТИА [23].

В статьях [11, 12] только формулируется проблема дифференциальной диагностики, возможно, может быть, с акцентом легкого замешательства о возможности ее решения традиционным клиническим путем, и в этом контексте ответа на вопрос «что под маской?» напрашивается мысль о необходимости поиска других путей разрешения проблемы.

Из общей мешанины преходящих жалоб и нечетких скоротечных досимптомных проявлений дискомфорта необходимо вычленить тот наиболее угрожающий «концентрат» наиболее информативных предикторов [7, 10], дальнейший анализ которых позволит с достаточной приемлемой вероятностью определить вектор (или наиболее вероятные векторы) траектории дальнейшего развития патологического процесса.

Актуальность поиска методов распознавания образа начинающегося патологического процесса, в частности, оговорена Медицинским информационно-аналитическим центром РАМН (МИАЦ РАМН) в анализе прогноза развития медицинской науки на период до 2025 года: «получит качественно новый уровень диагностика заболеваний центральной нервной системы на пресимптоматической стадии патологического процесса». В этом ключе применение адаптивной нечеткой нейросетевой модели в качестве диагностического инструмента представляется наиболее перспективным именно в тех случаях, когда разбалансировка нормально устойчиво протекающего гомеостаза здорового организма только начинает заявлять о себе *in statu nascendi* в до-и/или субклинической форме своего проявления, когда больше сомнений, чем уверенности в том, что это уже начало развития предпатологической стадии заболевания [1, 3, 4, 5, 8, 14].

Одним из подходов такого решения может быть составление «консилиума нейросетевых классификаторов» по субматрицам Mp_i , $i=1, \dots, k$. В предыдущих работах адаптивный нейросетевой классификатор строился с использованием «наилучших», наиболее информативных признаков* (предикторов), остальные отбрасывались, считаясь «дефективными» и о них в дальнейшем ничего не упоминалось [4, 5, 9].

Любой статистический анализ медицинских данных начинается с составления матрицы исходных данных. Матрица данных должна быть корректной и выглядеть с обязательным соблюдением условия ее размерности по M – числу предикторов (жалобы, симптомы, другие признаки) заболевания, и N – числу наблюдений (случаев) рис. 1а.

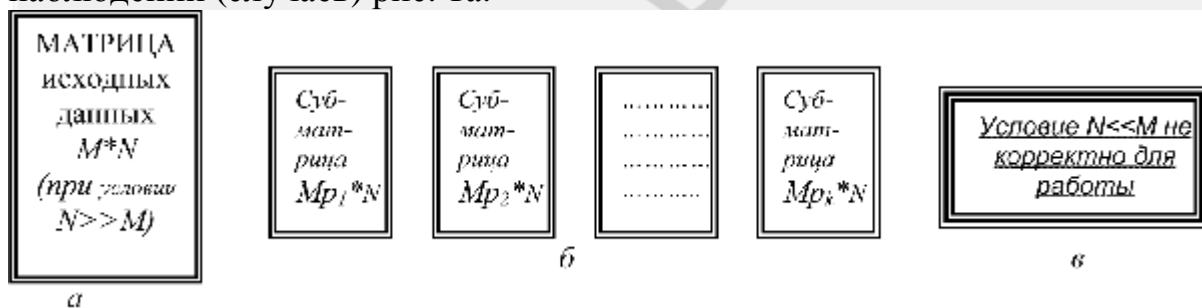


Рис. 1. а – обычный корректный вид матрицы исходных данных; б – представление исходных данных в виде субматриц с различными комбинациями предикторов из их общего числа принятых в рассмотрение; в – вид матрицы с нарушением корректного соотношения числа наблюдений (N) и количества предикторов (M); «проклятие размерности».

Но далеко не всегда представляется возможным выполнить условие $N \gg M$ и получить достаточно большую выборку данных. Число N , как минимум, должно превышать M в три и более раз. В действительности, особенно в медицинской практике, в лучшем случае $N=M$, что, чаще всего, обусловлено трудностью набора необходимого числа наблюдений. При размерности N При построении моделей решения задач классификации или прогноза методами многомерного статистического анализа в рамках проведения разведочного анализа данных (РАД) есть большая вероятность переобучения модели [3, 6, 9]. Результатом этого является низкое качество работы модели применительно к новым поступающим данным.

Одним из общих подходов и первым этапом решения этой проблемы является выделение наиболее информативных признаков-предикторов относительно решаемой задачи, в нашем случае задачи распознавания подтипов ТИА [7, 10]. В проводимом исследовании рассматривается возможность учета всей имеющейся информации, а именно всех M предикторов при построении модели решаемой задачи в случае, когда количество наблюдений N не намного превышает количества предикторов M , характеризующих каждое из них, рис. 1. Чтобы гарантировать высокую обобщающую способность построенной модели (высокую точность ее работы на новых данных) все имеющиеся предикторы распределяются по k субматрицам исходных данных, удовлетворяющим условию $N > M$ (рис. 1б). Каждая такая субматрица рассматривается как обучающая матрица, на основе которой строится модель классификации или распознавания подтипов ТИА. В данном исследовании в качестве модели используется адаптивный нейросетевой классификатор (АНК). В результате осуществления такого подхода получается несколько классификаторов, каждый из которых построен с использованием отличающегося от других подмножества предикторов. Такую комбинацию классификаторов называют ансамблем, где каждая субматрица представляет собой «виртуального эксперта» и участника «консилиума». Результаты работы каждого из классификаторов ансамбля комбинируются с использованием операторов агрегирования, позволяя получить окончательное решение.

После «посубматричного» анализа каждой матрицы $M_i \times N$, $i=1, \dots, k$ выносятся обобщенное решение т.е. классификационное решение их «консилиума». Обобщенное решение позволяет повысить точность распознавания подтипов ТИА.

1. Поиск решения проблемы распознавания подтипов ТИА по этиопатогенетическим особенностям

Если вспомнить средневековую схоластическую сентенцию о «Поиске дьявола, спрятавшегося в деталях», то «дьяволом» можно считать начинающийся патологический процесс ТИА по своему этиопатогенетическому сценарию, «трудный поиск» которого с целью подавления его активности, осуществляется в «дебрях» нечетко выраженных жалоб и признаков (M предикторов). Эта аллегория вполне укладывается в процесс постановки дифференциального диагноза заболевания на самой ранней стадии его развития, который по типу «восхождения кресчендо» может продолжать свое дальнейшее развитие. В этом контексте диагноз, тем более, дифференциальный, был и остается быть кантианской «вещью в себе», а стопроцентная достоверная точность недостижима, как в квантовомеханических исследованиях с учетом принципа неопределенности Гейзенберга, так и аналогично положению в макромире медицинской работы [5].

Цель настоящей работы – дальнейшая отладка и накопление опыта работы по технологии АНК и подтверждение (или опровержение): действительно ли оставшиеся за пределом внимания малоинформативные с низким классификационным потенциалом «предикторы-аутсайдеры» бесполезны в выявлении пациентов группы риска развития эпизодов ТИА по их подтипам?

В нашем исследовании в качестве обучающей матрицы для построения ансамбля классификаторов мы рассматривали полный набор из 41 предиктора и 101 наблюдения пациентов с тремя подтипами ТИА и нормой (всего 4 класса распознавания).**

2. Постановка математического эксперимента:

подход к комбинации нескольких классификаторов.

Комбинация нескольких классификаторов является одной из составных частей теории распознавания образов и позволяет повысить точность классификации относительно результатов, полученных с использованием каждого из классификаторов в отдельности [19].

Пусть имеется множество

$$\Omega = \{\omega_1, \dots, \omega_c\}$$

меток классов и пусть вектор (матрица)

$$x = [x_1, \dots, x_M]^T \in R^M$$

– M-мерный вектор признаков-предикторов, описывающий объект данных.

Классификатором является отображение следующего вида:

$$D: R^M \rightarrow [0,1]^c,$$

где $D(x)$ – c-мерный вектор, представленный в виде матрицы, у которого i-ый компонент определяет степень принадлежности x классу w_i , $i=1, \dots, c$. В

системах, основанных на комбинации k классификаторов, выходы отдельных классификаторов агрегируются для получения окончательного классификационного решения:

$$D(x) = F(D_1(x), \dots, D_k(x))$$

где F – оператор агрегирования. Выходом каждого отдельного классификатора является c-мерный вектор (матрица)

$$D_i(x) = [d_{i,1}(x), \dots, d_{i,c}(x)]^T$$

$i=1, \dots, k$. Выходом комбинации классификаторов является также c-мерный вектор (матрица)

$$D(x) = [\mu_1(x), \dots, \mu_c(x)]^T$$

Если необходимо определить для объекта x единственную метку класса, то класс w_s соответствует максимальному значению степеней принадлежности:

$$d_{i,s}(x) \geq d_{i,j}(x) \quad \forall j = 1, \dots, c$$

-для отдельных классификаторов

$$\mu_s(x) \geq \mu_t(x), \quad \forall t = 1, \dots, c$$

-для всего ансамбля.

Среди всех имеющихся методов построения ансамбля классификаторов наиболее популярными являются «bagging» и «boosting» [21], которые основаны на манипуляциях с исходным обучающим множеством с целью построения нескольких классификаторов.

«Bagging» (от bag – мультимножество, аналогичное обычному множеству set, в отличие от которого его одинаковые элементы могут входить в это мультимножество bag несколько раз) является специальным мета-алгоритмом машинного обучения классификационных и регрессионных моделей,

предназначенным для улучшения их стабильности и точности. В методе «bagging» k раз осуществляется случайный отбор обучающего множества из исходного набора данных, размерность которого равна размерности самого исходного набора. Так как осуществляется отбор с замещением, то некоторые предикторы могут несколько раз повторяться в составе нового набора данных. Полученные k наборов используются для обучения k классификаторов (например, деревья решений, нейронная сеть, модель дискриминантного анализа и т.д.). Классификация нового объекта данных происходит с использованием методов синтеза результатов классификации по k классификаторам. В методе «boosting» (метод усиления простых классификаторов) ансамбль классификаторов строится путем адаптивного изменения распределения обучающих данных на основе достигнутого уровня точности, полученного с использованием ранее построенных классификаторов ансамбля. Получение обобщенного решения осуществляется с использованием взвешенной процедуры голосования, где в качестве весов используются значения точности классификации с использованием каждого из классификаторов на обучающих данных.

Теоретические и эмпирические результаты показывают, что результат комбинации классификаторов наиболее эффективен, когда классификаторы являются независимыми, т.е. не коррелируют друг с другом [22]. Наиболее эффективным методом в этом случае является обучение отдельных членов ансамбля на качественно различающихся подмножествах предикторов. Таким образом, метод построения ансамбля классификаторов на основе декомпозиции набора признаков в большинстве случаев имеет преимущества. Имеется большое количество публикаций, исследующих эффективность ансамбля классификаторов, которые построены с использованием различных подмножеств признаков. Например, для декомпозиции признаков используются результаты их кластеризации, при которой каждый из кластеров оценивается с использованием обобщенной ошибки классификации ансамбля классификаторов, построенных для каждого из кластеров. В работе [17] была продемонстрирована эффективность рандомизированных подмножеств признаков для построения ансамбля классификаторов.

Для формирования подмножеств признаков-предикторов в нашем исследовании используется случайный отбор, который соответствует методу «bagging». Отличием является то, что метод в этом случае применяется не к обучающему набору, а к набору всех принятых в рассмотрение признаков-предикторов (жалобы, симптомы, другие признаки). В ходе работы был проведен эксперимент по поиску оптимальной величины их подмножеств, которые смогли бы обеспечить максимальную точность классификации распознавания подтипов ТИА с использованием ансамбля классификаторов. В ходе построения такого ансамбля проверка на независимость случайным образом сформированных подмножеств предикторов не проводится. Получение достаточно высокой точности и стабильности классификации с помощью комбинации классификаторов гарантируется построением достаточно большого, подбираемого экспериментально, количества классификаторов с различными наборами предикторов.

Таким образом, преимуществами использования подмножеств признаков-предикторов при обучении ансамбля классификаторов является:

- 1) сокращение размерности данных, что снижает влияние «проклятия размерности» (т.е. когда количество клинических случаев не намного больше количества предикторов);
- 2) снижение корреляции между классификаторами путем обучения их на различных признаках-предикторах (аналог независимости суждений реальных экспертов);
- 3) улучшение точности классификации с использованием ансамбля классификаторов.

3. Операторы агрегирования классификаторов

Существует большое количество операторов, позволяющих комбинировать выходы отдельных классификаторов ансамбля. К ним относятся: оператор максимума, минимума, произведения, усреднения, решение большинством голосов и т.д.

В нашем исследовании используется комбинация большинством голосов, которая является достаточно популярной и проста в реализации.

Пусть вектор-матрица

$$D_i(x) = [d_{i,1}(x), \dots, d_{i,c}(x)]^T \in [0,1]^c$$

-выход классификатора D_i , $i=1, \dots, k$ для входного объекта x . Значение

$$d_{i,j}(x) \in [0,1]$$

означает поддержку, оказываемую классификатором D_i гипотезы, что x относится к классу w_j . Для того, чтобы определить «голос» классификатора в поддержку единственного класса мы огрубляем классификационное решение, а именно выбираем класс

$$\omega_s \Leftrightarrow d_{i,s}(x) = \max_j \{d_{i,j}(x)\}.$$

Таким образом, классификационное решение для каждого D_i формулируется как бинарный вектор D_i^h , имеющее единицу в позиции s и ноль в остальных позициях:

$$d_{i,j}^h(x) = \begin{cases} 1, & j = s \\ 0, & j \neq s \end{cases}$$

Решение большинством голосов F_{maj} , представленное в виде матрицы размерности $s \times 1$, рассчитывается следующим образом:

$$F_{maj} \equiv D(x) = [d_1(x), \dots, d_c(x)]^T, \quad d_j(x) \in \{0,1\}, j=1, \dots, c$$

и

$$d_j(x) = \begin{cases} 1, & \sum_{i=1}^k d_{i,j}^h(x) = \max_{s=1, \dots, c} \sum_{i=1}^k d_{i,s}^h(x) \\ 0, & \end{cases}$$

где k – количество классификаторов в ансамбле.

4. Результаты математического эксперимента

Исходным исследуемым обучающим набором данных для построения ансамбля классификаторов является набор данных для дифференциальной диагностики подтипов ТИА. Набор состоит из 101 наблюдения по 41 признакам-предикторам, приведенных в анкете опроса пациента. В состав обучающего набора входят: 22 наблюдения с атеротромботическим этиопатогенезом эпизодов ТИА (СубТИА1), 23 наблюдения с кардиоэмболическим этиопатогенезом эпизодов ТИА (СубТИА2), 22 наблюдения с гипертензивным этиопатогенезом эпизодов ТИА (СубТИА3) и контрольная группа НОРМА-34 наблюдения**.

В ходе экспериментальных исследований по построению ансамблей классификаторов для этого набора было определено, что оптимальная размерность генерируемых случайным образом наборов признаков-предикторов для получения стабильной классификации должна равняться от 1/2 до 1/3 величины всего признакового пространства, т.е. от 13 до 20 признаков-предикторов. В нашем эксперименте было сгенерировано 10 подмножеств размерности 14, т.е. состоящих из 14 элементов. Каждый из 10-ти этих наборов признаков-предикторов – «виртуальный эксперт» и участник «консилиума» с достаточно независимым и отличным от других набором знаний в процессе проведения дифференциальной диагностики. В таблице 1 приведены три из 10 случайным образом сгенерированных наборов признаков-предикторов.

С использованием 10 сгенерированных наборов признаков-предикторов на обучающих данных было построено 10 классификаторов. В качестве модели классификации использовался адаптивный нейросетевой классификатор. В таблице 2 приведены значения точности классификации с использованием ансамбля и с использованием наилучшего (по точности) классификатора из 10 построенных.

Таблица 1. Три из сгенерированных 10 наборов признаков-предикторов***

Номер предиктора	Набор 1	Набор 2	Набор 10
1	HEPACOLC	METDEPEN	HRLIMITS
2	HEARTACH	HEPACOLC	HYPERTEN
3	HYPERTEN	AHEADECH	SMOKAGE
4	CORCARSC	HRLIMITS	VERTIGO
5	SMOKAGE	HEREDITA	OSTEOCON
6	METDEPEN	INSOMNIA	LIFERESI
7	ALCOHOLI	CHANGRES	CHANGRES
8	HEARTAUS	OPDISODS	AGE
9	GASTRULC	KSDESEAS	LABEFFEC
10	HEREDITA	CLIMAX	ALCOHOLI
11	DIASPRES	IRRITAB	GENDER
12	OSTEOCON	LIFERESI	CORCARSC
13	MEMORYLO	CONFLSAD	METDEPEN
14	INSOMNIA	HRITDDIS	HEADACHE

***Для краткости представления материала в таблице и в дальнейших вычислениях используются не полные названия жалоб и признаков, но их коды. Например, “HEPACOLC” – болезненность печени при пальпации, “HRLIMITS” – расширение границ сердца, “HEARTAUS” – данные аускультации сердца, “GENDER”-пол, “LABEFFEC” – трудоспособность, “MEMORYLO” – снижение памяти, “METDEPEN” – метеозависимость, “HEREDITA” – данные о заболеваниях близких родственников, “GASTRULC” – наличие указаний на язву

желудка “OPDISODS” – зрительные расстройства, “SLEEPATO” – повышенная сонливость, INSOMNIA – бессонница, IRRITAB – повышенная возбудимость и т.д.

Таблица 2. Сопоставление точности по наилучшему отдельному классификатору и по ансамблю классификаторов

Тип классификатора	Точность классификации (%)
Единственный наилучший	77.2
Ансамбль из 10 классификаторов	82.2

Далее приведены 11 правил (сгенерированных гипотез), соответствующих наилучшему единственному классификатору из ансамбля, точность классификации с использованием которого 77.2 %.

1. Если OPDISODS=1 и SLEEPATO>2, то класс=3.
2. Если OPDISODS=1 и SLEEPATO<=2 и AGE>1 и IRRITAB>2, то класс=1.
3. Если OPDISODS=1 и SLEEPATO<=2 и AGE>1 и IRRITAB<=2 и MEMORYLO<=3, то класс=4.
4. Если OPDISODS=1 и SLEEPATO<=2 и AGE>1 и IRRITAB<=2 и MEMORYLO>3, то класс=2.
5. Если OPDISODS=1 и SLEEPATO<=2 и AGE<=1, то класс=2.
6. Если OPDISODS>1 INSOMNIA<=1, то класс=1.
7. Если OPDISODS>1 INSOMNIA>1 и IRRITAB<=1, то класс=2.
8. Если OPDISODS>1 INSOMNIA>1 и IRRITAB>1 и OPDISODS>2, то класс=2.
9. Если OPDISODS>1 INSOMNIA>1 и IRRITAB>1 и OPDISODS<=2 и EDUCATIO<=3, то класс=3.
10. Если OPDISODS>1 INSOMNIA>1 и IRRITAB>1 и OPDISODS<=2 и EDUCATIO>3 и BRONCHRO<=1, то класс=4.
11. Если OPDISODS>1 INSOMNIA>1 и IRRITAB>1 и OPDISODS<=2 и EDUCATIO>3 и BRONCHRO>1, то класс=1.

Таким образом, рассмотренный нами способ распознавания подтипов ТИА с использованием ансамбля («консилиума») классификаторов и всего набора из 41 предиктора позволяет улучшить классификацию подтипов ТИА по сравнению с результатами, полученными в предыдущих работах, где использовался один классификатор и всего 10 наиболее релевантных к ТИА признаков-предикторов.

5. Обсуждение результатов математического эксперимента

Основной клинический акцент в этом исследовании сфокусирован на поиск информационно-технологического метода, который позволил бы получить максимально возможную дифференциально-диагностическую информацию из нечетких до-и субклинических анамнестических данных и данных осмотра. Постановка такой, казалось бы, неразрешимой задачи не нова, она достаточно четко ставилась в кардиологии еще в 60-е годы прошлого столетия. С позиций профилактической направленности в этом контексте ставится предпосылка разграничения медицинского диагноза по двум разновидностям – диагноз болезни и диагноз состояния. Если с диагнозом болезни все ясно, -он необходим для выбора общей стратегии лечения, то диагноз состояния нужен для решения тактических вопросов, например, для распознавания признаков патологии, для индивидуализации терапии, для разработки превентивных методов предотвращения дальнейшего уже клинически состоявшегося развития болезни.

«При этом имеется ввиду оценка динамики функционального состояния <патокинеза>, выявление и оценка подчас незначительных сдвигов в организме» [13]. Применение адаптивных нейросетевых классификаторов, по нашему мнению, как раз и является тем подходящим инструментом для постановки диагноза состояния. В этом контексте еще преследуется цель объединить (состыковать) математический (формально-логический) подход к решению клинической дифференциально-диагностической задачи с самой этой задачей в единый клинико-формально-логический комплекс решения прикладной клинико-неврологической проблемы превентивного предотвращения возникновения эпизодов ТИА с учетом дифференциации по подтипам. Причем, делается это по «незначительным», в основном, нечетко обозначенным анамнестическим предикторам.

Однако, имея перед собой весь набор предикторов и сопоставляя предикторы-«аутсайдеры» в субматрицах с «избранными» предикторами в основной матрице исходных данных, у каждого врача, более или менее искушенного в данной нозологии, могут возникнуть многочисленные «почему?». Например, почему в исходный основной набор предикторов не были включены, допустим, предикторы конфликтности в семье и с администрацией по работе, но включен как важный информативный предиктор «профессия». Вопросов в таком формате может быть много. И вот в этом случае, помимо основной цели, в этом исследовании еще преследуется цель согласования между традиционными клиническими представлениями о дифференциально-диагностической информативности признака-предиктора и о его информативности по результатам формально-логического (математического) анализа.

Траектория континуума вектора развития ишемии имеет свое начало процесса патокинеза и свой конец – инфаркт мозга. В настоящее время все усилия клинико-научных исследований концентрируются у ее начала – здесь и диагностика и прогноз, и разработка профилактических мероприятий. Предфинальная инфарктная часть этой траектории вызывает гораздо меньший интерес. А если такой интерес и есть, то заключается он, прежде всего, в желании не доводить ход событий до этих необратимых предфинальных инфарктных структурно-функциональных изменений. Тем не менее, имеется ряд научных работ, особенно за прошлые годы прошлого столетия, в которых преследуется цель прогнозирования исходов инсульта. В этом отношении есть все основания согласиться с тем, что «основными мишенями информатики в неврологической практике являются разработка методов профилактики и ранней диагностики инсульта, выбор оптимальных терапевтических программ, а не прогнозирование якобы неизбежного летального исхода уже при первичном осмотре больного» [2]. Удовлетворение такого «нездорового интереса», по мнению автора, не этично, ничего не дает практическому здравоохранению и только отвлекает силы и средства от решения насущных актуальных задач по профилактике ишемического поражения мозга на начальном этапе развития патологического процесса. В этом аспекте представляется целесообразным прилагать усилия на разработку инновационных диагностических методик раннего распознавания в ангионеврологии на базе современных логико-статистических методов распознавания образов, нейросетевых методов и

экспертных систем, инновационным по отношению к традиционным клиническим методам.

Практикующему врачу в повседневной работе совершенно нет необходимости вникать в структуру математического изложения этих информационных методов (как, например, нет необходимости вникать в химическую структуру формулы пенициллина или реополиглукина), но знать, что такая структура существует и иметь о ней какое-то представление – желательно.

Учитывая особенность ангионеврологических нозологий маскироваться один под другого, адаптивные нейросетевые классификаторы представляются тем самым наиболее удобным дифференциально-диагностическим инструментом в их распознавании для получения ответа на вопрос «что под маской?», если не со стопроцентной, то с приемлемой точностью вероятности.

По мере накопления знаний о своеобразии и многоликости ТИА, становится все очевиднее актуальность дифференциально-диагностической значимости этой острой цереброваскулярной патологии с целью предотвращения ее эпизодов [1, 3-6, 16, 20]. Именно в этом русле и проводится настоящее исследование поиска наилучшего алгоритма решения дифференциально-диагностической задачи по выделению подтипов ТИА.

Появляется реальная возможность проследить динамику патокинеза ТИА. Чем больше предикторов в обучающей выборке (матрице исходных данных) принято в рассмотрение, причем, в условиях хронического дефицита выверенных верифицированных клинических наблюдений, тем четче, точнее, более гибко можно следить за нюансами их изменений и осуществлять тонкую, адресную для каждого пациента, лечебную (управляющую) тактику по алгоритму «диагноз-лечение» [14, 15]. Появляется реальная возможность вернуться опять к устоявшейся истине «чем больше (предикторов), тем лучше», но уже в новых условиях. Если в случае чисто статистического подхода возможности корректного решения жестко регламентируются соотношением $N \gg M$ матрицы исходных данных (влияние «проклятия размерности»: число наблюдений должно значительно превышать число предикторов), то в условиях работы агрегированного «консилиума» адаптивных нейросетевых классификаторов это жесткое условие ослаблено, и практически не влияет на корректность правильного решения.

Таким образом, уже на раннем этапе развития патологического процесса появляется реальная возможность проследить, преимущественно по нечетким анамнестическим данным, континуум «кресчендо» драматического восхождения на эшафот церебрального инфаркта, начиная от едва заметных скоротечных эпизодов ТИА, и прервать его на раннем (субклиническом) этапе. Это дает возможность четче ориентировать превентивный лечебно-профилактический вектор препятствования развитию патокинеза на упреждающий позитивный саногенетический ход динамики коррекции гомеостатических нарушений.

Выводы

1. Распознавание подтипов ТИА по матрице исходных данных, в которой число предикторов превышает число клинических случаев, нарушает жесткое требование по размерности и соотношению этих ее элементов. Решение дифференциально-диагностической задачи традиционным путем, с

использованием методов многомерного статистического анализа, в этом случае некорректно; сказывается влияние феномена «проклятия размерности» (the curse of dimensionality).

2. Учитывая то, что в качестве предикторов распознавания (дифференциальной диагностики) используются неточные и нечетко обозначенные данные (жалобы, симптомы, другие признаки), решение проблемы дифференциальной диагностики ангионеврологических заболеваний (в том числе и транзиторных ишемических атак) на раннем субклиническом этапе пресимптоматической стадии патологического процесса их развития, традиционным клиническим путем практически невозможно, необходим поиск новых, нетрадиционных (инновационных) путей.

3. Применение в этом случае интеллектуального классификатора АНК с предварительным отбором информативных признаков-предикторов позволяет значительно ослабить жесткие требования к матрице исходных данных по соотношению количества клинических наблюдений и релевантных к ним предикторов, и делает решение задачи в этих условиях приемлемым и корректным.

4. Многие, не включенные в исследование предикторы (жалобы, симптомы, другие признаки), могут быть не менее информативными при рассмотрении их в различных комбинациях классификаторов. Поэтому предложенный подход создания ансамбля классификаторов («консилиума», «совета виртуальных экспертов» адаптивных нейросетевых классификаторов) на основе нескольких подмножеств признаков позволяет обнаружить новые взаимосвязи, и, в конечном итоге, повысить точность распознавания подтипов ТИА.

5. Использование предложенного подхода при проведении экспериментальных исследований по распознаванию подтипов ТИА в рамках единой экспертной системы путем агрегирования нескольких нейросетевых классификаторов в единый ансамбль позволило повысить точность распознавания с 77,2 % (результат решения по одному наилучшему АНК) до 82,2 % с учетом «мнения» нескольких таких классификаторов в рамках их единого ансамбля («консилиума»). Полученное повышение точности распознавания дает основание считать целесообразным применение агрегированной экспертной системы на базе генерирования нескольких ансамблей классификаторов для дифференциальной диагностики подтипов ТИА.

Литература

1. Апанель, Е. Н. Комплексный подход к этиотропным лечебно-профилактическим мероприятиям по предотвращению эпизодов транзиторных ишемических атак // Медицинский журнал. 2008. № 2. С. 117 – 120.
2. Виленский, Б. С. Прогнозирование исходов инсульта не правомерно // Неврологический журнал. 2008. № 1. С. 52 – 53.
3. Дривотинов, Б. В., Мастыкин, А. С., Красько, О. В., Апанель, Е. Н. Применение разведочных методов анализа данных в дифференциальной диагностике подтипов транзиторных ишемических атак // Военная медицина. 2006. № 1. С. 51 – 54.
4. Дривотинов, Б. В., Апанель, Е. Н., Мастыкин, А. С. Транзиторные ишемические атаки в свете современных нейропатологических

- представлений // В кн.: Нейрогуморальные механизмы регуляции функций в норме и патологии. Минск, 2007. С. 295 – 301.
5. Дривотинов, Б. В., Тарасевич, М. И., Мастыкин, А. С., Апанель, Е. Н., Новоселова, Н. А. К выявлению этиотропного фактора подавления развития кардиоэмболического (кардиогенного) подтипа транзиторной ишемической атаки // Медицинский журнал. 2008. № 1. С. 115 – 119.
 6. Мастыкин, А. С., Дривотинов, Б. В., Апанель, Е. Н. Гетерогенность нозологического понятия транзиторной ишемической атаки // Белорусский медицинский журнал. 2004. № 4. С. 18 – 21.
 7. Новоселова, Н. А. Предварительный отбор информативных признаков для улучшения точности предсказания с помощью нейронной сети // Искусственный интеллект. 2004. № 2. С. 150 – 154.
 8. 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.
 9. Новоселова, Н. А., Апанель, Е. Н., Дривотинов, Б. В., Мастыкин, А. С. Применение адаптивной нейро-нечеткой модели для распознавания подтипов транзиторных ишемических атак // Сборник Международной конференции «Сигнальные механизмы регуляции физиологических функций», 30 ноября 1 декабря 2007. С. 186 – 188. Минск, 2007.
 10. Новоселова, Н. А. Алгоритмы построения гибридного нечеткого классификатора для анализа медицинских данных. Диссертация на соискание ученой степени кандидата технических наук. Минск, 2008.
 11. Парфенов, В. А., Абдулина, О. В., Замерград, М. В. Периферическая вестибулопатия под маской инсульта // Неврологический журнал. 2007. № 6. С. 21 – 25.
 12. Пономарева, Е. Н., Пономарев, В. В. Неврологические маски антифосфолипидного синдрома // Неврологический журнал. 2007. № 4. С. 14 – 17.
 13. Сидоренко, Г. И. Пути применения кибернетики для решения вопросов диагностики и терапии // Кибернетика в медицине. Научная конференция 14 марта 1968 г., Минск, 1968. С. 35 – 38.
 14. Сидоренко, Г. И., Прокопенко, Ю. И. Методологические аспекты предпатологии // Вестник АМН СССР. 1976. № 4. С. 13 – 22.
 15. Сидоренко, Г. И. Творчество и медицина: поиск неочевидных решений. Минск, 2002.
 16. Скворцова, В. И. Механизмы повреждающего действия церебральной ишемии и новые терапевтические стратегии // Журнал. невропатол. и психиатр. 2003. № 9. С. 20 – 22.
 17. Bay, S.D. Combining nearest neighbor classifiers through multiple feature subsets // Proceedings of the 17th International conference on machine learning. Madison, WI. 1998. P. 37 – 45.
 18. Brown, M., Bossley, K.M., Mills, D.J., Harris, C.J. High dimensional neurofuzzy systems: overcoming the curse of dimensionality/ Fuzzy Systems, 1995. International Joint Conference of the Fourth IEEE International Conference on Fuzzy Systems and

The Second International Fuzzy Engineering Symposium. Proceedings of 1995 IEEE International Conference. Yokohama, Japan. 1995. Vol. 4. P. 2139 – 2146.

19. Duda, R.O., Hart, P.H., Stork, D.G. Pattern classification. Willey-Interscience, New York, 2002.

20. Ovbiagele, B., Cruz-Flores, S., Lynn, M., Chimowitz, M. Early Stroke Risk After Transient Ischemic Attack Among Individuals With Symptomatic Intracranial Artery Stenosis // Arch. Neurol. 2008. Vol 65. № 6. P. 733 – 737.

21. Quinlan, J.R. Bagging, boosting and C4.5 // Proceedings of AAA/IAAI. 1996. Vol. 1. P. 725 – 730.

22. Tumer, K., Oza, N.C. Decimated input ensembles for improved generalization // Proceedings of the International Joint Conference on Neural Networks. Washington, DC. 1999.

23. Wanby, P., Teerlink, T., Brudin, L. et al. Asymmetric dimethylarginine (ADMA) as a risk marker for stroke and TIA in a Swedish population // Atherosclerosis. 2006. Vol. 185. № 2. P. 271 – 277.

*Понятие «признак» трактуется двояко: в широком абстрактно-математическом смысле (признаковое пространство) и в прикладном, применительно к решаемой задаче классификации, – «признак-предиктор» как независимая переменная.

** По клиническим данным использован диссертационный материал В.Б. Шалькевича.