

ФОРМИРОВАНИЕ ВЕКТОРА ЗНАЧИМЫХ ПОКАЗАТЕЛЕЙ КЛИНИКО-БИОХИМИЧЕСКИХ АНАЛИЗОВ ПАЦИЕНТОВ С ПСОРИАЗОМ С ПОМОЩЬЮ ИСКУССТВЕННОЙ НЕЙРОННОЙ СЕТИ ДЛЯ ВЫЯВЛЕНИЯ ЗАБОЛЕВАНИЯ НА РАННИХ СТАДИЯХ

Е.В. ВЫСОЦКАЯ, А.Н. БЕЛОВОЛ, Ю.В. КИРИЧЕНКО

В статье был рассмотрен подход к формированию вектора значимых показателей с помощью нейросетевых технологий. Была создана автоассоциативная нейронная сеть, позволяющая выделять наиболее важные биохимические показатели пациентов с псориазом.

Ключевые слова: вектор значимых показателей, искусственная нейронная сеть, клинико-биохимические показатели, псориаз.

ВВЕДЕНИЕ

Псориаз — распространенный хронический дерматоз, популяционная частота которого по данным различных авторов составляет от 1 до 5% [1]. В последние годы отмечены рост заболеваемости этим дерматозом, увеличение числа тяжелых, атипичных, инвалидизирующих, резистентных к проводимой терапии форм заболевания. Псориазический процесс носит хронический рецидивирующий характер течения. Ремиссия может быть длительной — от нескольких месяцев до десятиков лет, а у ряда пациентов ремиссия вообще не наступает. Прогноз для жизни при псориазе чаще всего благоприятный. Исключение составляют случаи тяжелого артропатического и пустулезного псориаза. Болезнь значительно ухудшает качество жизни, снижает работоспособность и социальную активность пациентов, что определяет не только медицинскую, но и социальную значимость проблемы [1, 2].

Поэтому задача выявления псориаза на ранних стадиях, т.е. до появления характерных признаков на коже, является в настоящее время актуальной.

Как было показано в [3], диагностика дерматологических заболеваний, и в частности, псориаза, по клинико-биохимическим показателям возможна с помощью искусственной нейронной сети, но для этого необходимо наличие несколько десятков значений показателей пациента, которые отражают функциональное состояние различных органов и систем.

Формирование вектора значимых показателей позволит выбрать показатели, которые наиболее полно характеризуют исследуемый объект при сокращении параметрической избыточности и сохраняют особенности всей выборки показателей.

1. ОБЗОР СУЩЕСТВУЮЩИХ МЕТОДОВ ФОРМИРОВАНИЯ ВЕКТОРА НАИБОЛЕЕ ИНФОРМАТИВНЫХ ПОКАЗАТЕЛЕЙ

В настоящее время известны различные методы формирования вектора информативных

признаков. В статистике эти методы известны также как методы сокращения размерности. Все эти методы основаны на трех основных идеях принципиальной возможности уменьшения пространства признаков [4]:

- дублирование информации в исходной выборке за счет взаимосвязанности признаков;
- неинформативность признаков, мало меняющихся при переходе от одного образа к другому;
- возможность агрегирования (простое или «взвешенное» суммирование) по некоторым признакам.

К основным методам можно отнести следующие методы [4, 5, 6, 8, 9]:

- метод главных компонент;
- факторный анализ;
- метод экстремальной группировки признаков;
- регрессионный анализ;
- дискриминантный анализ;
- методы последовательного увеличения и уменьшения группы признаков;
- метод контрастных групп;
- методы коллективной оценки, средне-взвешенной оценки и т.п.;
- методы, основанные на использовании искусственных нейронных сетей.

Метод главных компонент — это один из способов выделения значимых показателей, состоящий в переходе к новому ортогональному базису, оси которого ориентированы по направлениям максимальной дисперсии набора входных данных. Вдоль первой оси y_1 нового базиса $y_1(x)$ дисперсия максимальна, вторая ось максимизирует дисперсию при условии ортогональности первой оси, и т.д. Последняя ось имеет минимальную дисперсию из всех возможных.

Геометрически это выглядит как ориентация новой координатной оси вдоль направления наибольшей вытянутости эллипсоида рассеивания объектов исследуемой выборки в пространстве признаков X .

Такое преобразование позволяет убирать показатели путем отбрасывания координат, соответствующих направлениям с минимальной дисперсией. Предполагается, что если нужно отказаться от одного из базисных векторов, то лучше, если это будет тот вектор, вдоль которого набор входных данных меняется менее значительно.

В описанном выше методе главных компонент критерий информативности пространства признаков подразумевает, что ценную для диагностики информацию можно отразить в линейной модели, которая соответствует новой координатной оси в данном пространстве с максимальной дисперсией распределения проекций исследуемых объектов.

Факторный анализ, в отличие от метода главных компонент, основан не на дисперсионном критерии автоинформативности системы признаков, а ориентирован на объяснение имеющихся между признаками корреляций. Основное предположение факторного анализа заключается в том, что корреляционные связи между большим числом наблюдаемых переменных определяются существованием меньшего числа гипотетических ненаблюдаемых переменных или факторов. Общей моделью служит следующая линейная модель:

$$X_i = \sum_{j=1}^k a_{ij} f_j + b_i U_i + \varepsilon_i,$$

где случайные величины f_j — общие факторы, случайные величины U_i — факторы, специфические для величин X_i и не коррелированные с f_j , а ε_i — случайные ошибки. Предполагается, что $k < n$ задано, случайные величины независимы между собой и с величинами f_j и U_i и имеют $\sum \varepsilon_i = 0$ и $D_{\varepsilon_i} = \sigma_i^2$. Постоянные коэффициенты a_{ij} называются факторными нагрузками (нагрузка i -й переменной на j -й фактор). Значения a_{ij} , b_i и σ_i^2 считаются неизвестными параметрами, подлежащими оценке.

В отличие от метода главных компонент факторный анализ подает корреляционную структуру матрицы входных переменных в терминах латентной (гипотетической) модели, но он не дает однозначного решения задачи, потому что представление корреляционной матрицы факторами можно осуществить бесконечным количеством способов — ортогональное преобразование матрицы входных данных приводит к новой факторизации.

Метод экстремальной группировки признаков опирается на представление о том, что изменение какого-либо общего фактора неодинаково сказывается на измеряемых признаках. В этом случае параметры разбивают на группы таким образом, чтобы параметры каждой из групп в основном отражали изменение какой-либо одной характеристики объекта и мало бы зависели от его других существенных характеристик. Сле-

довательно, параметры, принадлежащие одной группе — сильно коррелированы, в то время как принадлежащие к различным группировкам — слабо. После этого строится случайная величина для каждой группы признаков, которая наиболее сильно коррелирует с параметрами данной группы и принимается за фактор, от которого существенно зависят параметры этой группы.

Регрессионный анализ рассматривает критериальный показатель z как «зависимую» переменную, которая выражается функцией от «независимых» признаков x_1, \dots, x_p . Для оценки эффективности регрессионной диагностической модели вводится вектор остатков $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n)'$, который отражает влияние на z совокупности неучтенных случайных факторов, либо меру достижимой аппроксимации значений критериального показателя z_j функциями типа $y(x_i)$. Линейная функция регрессии записывается следующим образом

$$z_i = w_0 + w'x_i + \varepsilon_i,$$

где w_0 называется свободным членом, а элементы вектора $w' = (w_1, \dots, w_p)$ называются коэффициентами регрессии. Недостатком регрессионного анализа является то, что модели, имеющие слишком малую сложность, могут оказаться неточными, а в моделях, имеющих избыточную сложность, когда количество параметров более ста, возникает явление мультиколлинеарности, приводящее к плохой обусловленности и в предельном случае вырожденности матрицы ковариации.

Дискриминантный анализ применяется, если связь показателя z с исходными признаками является нелинейной и носит неизвестный характер. В этом случае испытываемые параметры разбиваются на классы, а эффективность модели рассматривается на основании ее способности разделять (дискриминировать) эти классы. Как и при построении регрессионных моделей, большое количество и наличие групп связанных показателей обуславливают применение «грубых» алгоритмов нахождения дискриминантных функций. Данные алгоритмы также в основном сводятся к отбору информативных признаков с помощью эвристических процедур. Отличие указанных процедур заключается в иных критериях оптимальности признаков, чем при построении регрессионных моделей.

Методы последовательного увеличения и уменьшения группы признаков основаны на полном переборе всех параметров, что делает их малоприменимыми в данном случае из-за длительности комбинаторного перебора всех параметров.

В методе контрастных групп исходной информацией, помимо таблицы экспериментальных данных с результатами обследования, является «черновая» версия линейного правила вычисления тестируемого показателя. Эта «черновая» версия может быть составлена экспериментатором, исходя из его теоретических представлений о том,

какие признаки и с какими весами должны быть включены в модель. В данном случае применение этого метода невозможно из-за трудности построения «черновых» версий для более сотни показателей даже эксперту.

Методы коллективной оценки, средневзвешенной оценки, ранжирование, парное сравнение и другие методы экспертных критериев в данном случае также не применимы, так как требуют привлечения специалистов в области медицины: врачей, биохимиков и т.д. Объективизация достигается увеличением числа экспертов.

Методы, основанные на использовании искусственных нейронных сетей, способны выявлять скрытые закономерности в многомерных данных. Исходные данные для сети составляются таким образом, чтобы на выходах был всегда тот же набор сигналов, что и на входе. В процессе работы алгоритм обратного распространения ошибки минимизирует ошибку. Это означает, что веса связей от входного слоя нейронов и, примерно, до серединного слоя будут работать на компрессию сигнала, а остальные — на его декомпрессию.

Исходя из рассмотренных выше математических методов формирования вектора информативных показателей, можно сделать вывод, что аппарат нейронных сетей позволяет выявлять информативные признаки для «бедной» выборки. Также применение аппарата искусственных нейронных наиболее оправданно при наличии большого количества входных параметров и их нелинейности.

2. ПОСТАНОВКА ЗАДАЧИ

Целью работы является формирование вектора значимых показателей клиничко-биохимических анализов пациентов с псориазом с помощью искусственной нейронной сети для выявления заболевания на ранних стадиях.

Было рассмотрено 274 пациента с псориазом, у которых были взяты различные показатели белкового (ast, alt, urea и т.д.), углеводного (gluc, Mg, Г-6ФФГ и т.д.), микроэлементного (Fe, Ca, Mg и т.д.) и жирового (ИБХЛ, МДА, Диены и т.д.) обменов (всего 141 показатель), которые объединили в один вектор $X = \{x_1, x_2, \dots, x_i\}$ — массив исходных данных (табл. 1), где $x_i = (x_{i1}, x_{i2}, \dots, x_{ik})$, $k = \overline{1, 274}$ — вектор значений одного из показателей для каждого пациента.

Таблица 1

Таблица исходных данных биохимических показателей больных псориазом

Пациент	Параметры биохимических показателей				
	1	2	3	...	i
1	X_{11}	X_{12}	X_{13}	...	X_{1i}
2	X_{21}	X_{22}	X_{23}	...	X_{2i}
3	X_{31}	X_{32}	X_{33}	...	X_{3i}
...
k	X_{k1}	X_{k2}	X_{k3}	...	X_{ki}

Тогда задача сводится к тому, чтобы найти такой вектор $X' = \{x_{s1}, \dots, x_{sm}\}$, $s_{s1}, \dots, s_{sm} \in [1, 141]$, что его размерность меньше размерности исходного вектора, и который обладает некоторым желаемым свойством, а именно:

$$\varepsilon(X') \leq \varepsilon_{\text{порог}}$$

где $\varepsilon(X')$ — значение критерия оптимальности для набора параметров X' , $\varepsilon_{\text{порог}}$ — пороговое значение критерия оптимальности.

3. ПРИМЕНЕНИЕ АВТОАССОЦИАТИВНОЙ НЕЙРОННОЙ СЕТИ К ФОРМИРОВАНИЮ ВЕКТОРА ЗНАЧИМЫХ ПОКАЗАТЕЛЕЙ

Автоассоциативная сеть является универсальным средством для решения задачи редукции размерности как для линейно структурированных, так и для варианта нелинейно структурированных данных.

Будем использовать автоассоциативную нейронную сеть (рис. 1), в которой число нейронов во входном и выходном слое одинаково и равно N , число нейронов в сжимающем скрытом слое равно M ($M \leq N$). В принципе такая нейронная сеть представляет собой многослойный персептрон, причем при его обучении в качестве требуемых целевых значений используются те же значения, которые подаются на вход сети; в качестве процедуры обучения такой сети можно использовать любой градиентный метод [7].

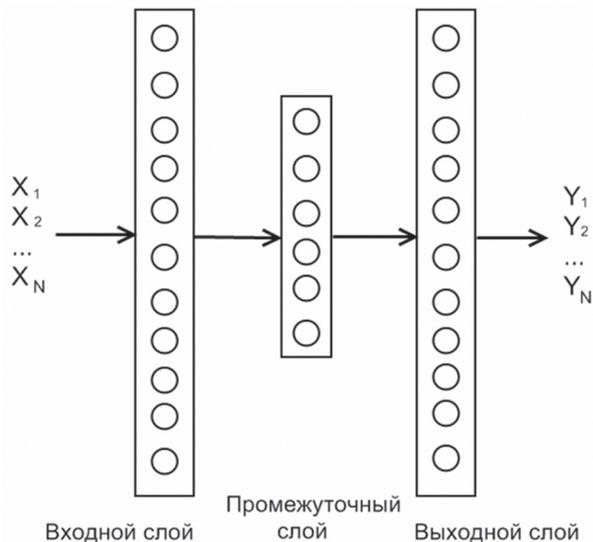


Рис. 1. Условное изображение нейронной сети

Основная идея работы заключается в том, что при наличии m главных переменных, присутствие «узкого горла» с количеством нейронов M ($M \leq N$) не должно сказываться на способности сети обучаться и воспроизводить на своем выходе входную информацию. Однако в работе [7] показано, что такая нейронная сеть не способна эффективно выявить большее, чем $M = 1$, количество переменных. Поэтому для выявления наиболее информативных признаков поступим сле-

Таблица 2

Список выделенных наиболее значимых клинко-биохимических показателей больных псориазом

2,4-ДНФ-К	LDH	ИБХЛ
Ca(волосы)	Mg(волосы)	ИЛ-4
CD-19	Na(волосы)	ИЛ-6
CD-3	Na(моча)	ИЛ-8
Cd-4	Na(сыв крови)	Кортизол
CD-8	T4	КФК
Fe(сыв крови)	Zn(волосы)	Мелатонин
IgA	АКТГ	Мочевина
IgG	БХЛ	Пролактин
IgM	вит А	ТДО
K(волосы)	вит С	ФНО-А
K(эритроц крови)	Глутамин	

ВЫВОДЫ

В данной статье описан способ выделения из всего множества значений клинко-биохимических показателей наиболее информативных с помощью искусственной нейронной сети. Созданная нейронная сеть позволяет сократить количество параметров, необходимых для раннего обнаружения псориаза, с 141 до 35, т.е. 24,8% от исходного размера, таким образом, уменьшив материальные и временные затраты на их получение. Выделенные информативные признаки также могут использоваться для синтеза моделей классификации.

Возможно применение созданной нейронной сети для дальнейшего «сжатия» пространства характеристик с целью их визуализации, выявления причинно-следственных связей, построения обобщенных характеристик и т.д.

Литература.

- [1] Бакулев А.Л., Шагова Ю.В., Козлова И.В. Псориаз как системная патология // Саратовский научно-медицинский Журнал. – 2008. – №1(19). – С.13-20.
- [2] Кутасевич Я.Ф. Современный взгляд на проблему псориаза // Дерматология и венерология. – 2002. – №2 (16). – С. 3-10.
- [3] Высоцкая Е.В., Жукова Н.В., Кириченко Ю.В. Обучение искусственной нейронной сети для диагностирования дерматологических заболеваний пациентов. Информационные технологии и автоматизация – 2009, II всеукраинская научно-практическая конференция. – С. 20-21.
- [4] Дюк В.А. Компьютерная психодиагностика. – СПб: Братство. – 2004. – 364 с.
- [5] Коваленко Н. В., Воронников В. В. Формирование вектора первичных значимых параметров для оценки технического состояния сложных информационных систем // Збірник наукових праць ЖВІ НАУ. Випуск 1. – С. 63-69.

дующим образом: будем выявлять параметры по одному и исключать из обучения.

Критерием останова процедуры обучения возьмем величину коэффициента множественной корреляции R, характеризующего тесноту связи между входными значениями и значениями на выходе нейронов сжимающего слоя.

Алгоритм выделения наиболее информативных признаков состоит из следующих этапов.

Шаг 1. Синтез искусственной ассоциативной нейронной сети с архитектурой [141-1-141].

Шаг 2. Нормализация данных исходной выборки.

$$x_{\text{норм}} = \frac{x - x_{\text{min}}}{x_{\text{max}} - x_{\text{min}}}$$

Шаг 3. Инициализация весовых коэффициентов начальными значениями [0,1].

Шаг 4. Обучение сети методом обратного распространения ошибки.

Шаг 5. Выделение наиболее значимого нейрона из входного слоя, т.е. такого нейрона, у которого сумма значений весов к нейронам скрытого слоя будет максимальной.

Шаг 6. Исключение показателя, соответствующему найденному нейрону, из обучающей выборки и включение его в массив наиболее информативных признаков.

Шаг 7. Определение коэффициента множественной корреляции R для текущего массива наиболее информативных признаков.

Шаг 8. Определение коэффициента множественной корреляции. Остановить процедуру формирования, если $R = 1 - \epsilon (\epsilon = 0.03)$, в противном случае перейти к шагу 3.

На рис. 2 показан график корреляции R в зависимости от количества выделенных признаков.

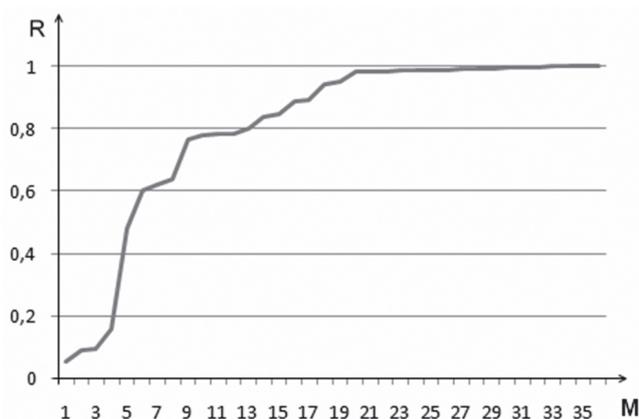


Рис. 2. Зависимость коэффициента R от количества выделенных показателей

Список сформированных параметров показан в табл. 2.

Таким образом, в результате исследования был сформирован вектор значимых клинко-биохимических показателей. Размерность вектора равна 35.

- [6] *Иваськив Ю.Л., Левченко В.В.* Вопросы использования репликативных нейронных сетей в задачах сжатия изображений // Математичні машини і системи, 2006, № 4. С. 42-50.
- [7] *Филаретов Г.Ф., Джордан Б.* Применение автоассоциативных нейронных сетей для сжатия информации. XXX международная конференция. Информационные технологии в науке, образовании, телекоммуникации, бизнесе. Украина, Крым, Ялта-Гурзуф, 2003. – С.12-15.
- [8] *Хайкин С.* Нейронные сети: полный курс, 2-е изд., испр.: Пер. с англ. – М.: ООО «И.Д. Вильямс», 2006. – 1104 с.
- [9] *Осовский С.* Нейронные сети для обработки информации: Пер. с польск. Н.Д. Рудинского. – М.: Финансы и статистика, 2004. – 344 с.

Поступила в редколлегию 13.04.2010



Высоцкая Елена Владимировна, кандидат технических наук, доцент, доцент кафедры Биомедицинских электронных устройств и систем Харьковского национального университета радиоэлектроники. Область научных интересов: аппараты и системы замещения утраченных органов и функций организма человека.



Беловол Алла Николаевна, кандидат медицинских наук, доцент, заведующая кафедрой дерматологии, венерологии и медицинской косметологии. Область научных интересов: системные дерматозы в сочетании с гипертонической болезнью.



Кириченко Юрий Владимирович, MS Tech, GlobalLogic Ukraine. Область научных интересов: разработка автоматизированных систем управления, нейронные сети.

УДК 519:616-079.4:616.5

Формування вектора значущих показників клініко-біохімічних аналізів пацієнтів з псоріазом за допомогою штучної нейронної мережі для виявлення захворювання на ранніх стадіях / О.В. Висоцька, А.М. Біловол, Ю.В. Кириченко // Прикладна радіоелектроніка: наук.-техн. журнал. – 2010. Том 9. № 2. – С. 280-284.

У статті було розглянуто підхід до формування вектора значущих показників за допомогою нейромережових технологій. Була створена автоасоціативна нейронна мережа, що дозволяє виділяти найбільш важливі біохімічні показники пацієнтів з псоріазом.

Ключові слова: вектор значущих показників, штучна нейронна мережа, клініко-біохімічні показники, псоріаз.

Табл.03. Лл.02. Бібліогр.: 09 назв.

UDC 519:616-079.4:616.5

Forming the vector of meaningful indicators of clinical and biochemical analyses of patients with psoriasis using an artificial neural network to detect the disease at early stages / E.V. Vysotskaya, A.A. Belovol, Y.V. Kirichenko // Applied Radio Electronics: Sci. Mag. – 2010. Vol. 9. № 2. – P. 280-284.

The paper considers an approach to the formation of the vector of significant indicators using neural network technologies. An autoassociative neural network highlighting important biochemical parameters of psoriasis patients is created.

Key words: vector of meaningful parameters, artificial neural network, clinical and biochemical parameters, psoriasis.

Tab. 03. Fig. 02. Ref.: 09 items.