

СИСТЕМНЫЙ АНАЛИЗ, УПРАВЛЕНИЕ И ОБРАБОТКА ИНФОРМАЦИИ

УДК 004.051

АЛГОРИТМ ОЦЕНКИ КАЧЕСТВА РАБОТЫ НЕЙРОННЫХ СЕТЕЙ ПРИ РЕШЕНИИ ЗАДАЧ КЛАССИФИКАЦИИ В МЕДИЦИНСКОЙ ДИАГНОСТИКЕ

Статья поступила в редакцию 21.11.2017, в окончательном варианте – 25.01.2018.

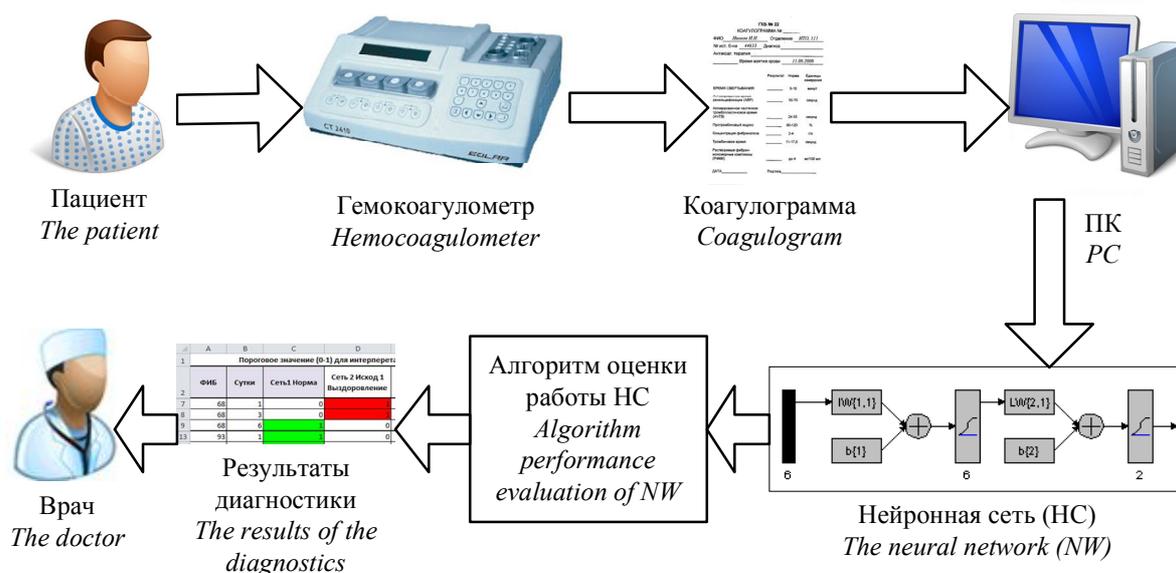
Сидорова Маргарита Александровна, Пензенский государственный технологический университет, 440039, Российская Федерация, г. Пенза, пр. Байдукова / ул. Гагарина, 1а/11, кандидат технических наук, доцент, идентификатор автора в SCOPUS: 26425316100, AuthorID elibrary: 435014, e-mail: sidorova_mailbox@mail.ru

Сержантова Наталья Александровна, Пензенский государственный технологический университет, 440039, Российская Федерация, г. Пенза, пр. Байдукова / ул. Гагарина, 1а/11, кандидат технических наук, доцент, идентификатор автора в SCOPUS: 40462363600, AuthorID elibrary: 564853, e-mail: itmmbspgta@yandex.ru

Рассмотрены некоторые существующие критерии оценки качества работы нейронных сетей (НС) при решении задач классификации. К таким критериям относятся следующие: функции ошибок обучения (сумма квадратов ошибок, средняя квадратичная ошибка, регулируемая ошибка и средняя абсолютная ошибка обучения); погрешность оценки толщины сосуда; зависимость вероятности возникновения ошибок первого и второго рода; доля неправильно классифицированных объектов в тестовой выборке. Указаны основные недостатки этих критериев. Для оценки результатов работы НС (на примере задач классификации параметров гемостаза таких, как свертываемость и вязкость крови) был использован расширенный аппарат статистического анализа, опирающийся на байесовский подход. Предложен собственный критерий оценки качества работы НС при решении задач классификации в медицинской диагностике, включающий учет порогового значения принятия НС прогностически верного решения. Разработан алгоритм оценки качества работы НС, основанный на предложенном критерии качества. Этот алгоритм позволяет варьировать значение уровня значимости для различных задач, тем самым, определяя необходимые и достаточные значения, при которых работа исследуемой НС считается эффективной.

Ключевые слова: нейронная сеть, классификация, медицинская диагностика, критерий, качество, уровень значимости, порог принятия решения, эффективность, информативность, ошибка, пропуск, ложное срабатывание

Графическая аннотация (Graphical annotation)



THE ALGORITHM FOR ASSESSING THE PERFORMANCE OF NEURAL NETWORKS IN SOLVING CLASSIFICATION PROBLEMS IN MEDICAL DIAGNOSIS

The article was received by editorial board on 21.11.2017, in the final version – 25.01.2018.

Sidorova Margarita A., Penza State Technological University, Baydukov Av. / 1a/11 Gagarin St., Penza, 440039, Russian Federation

Cand. Sci (Engineering), Associate Professor, author ID 26425316100, e-mail: sidorova_mailbox@mail.ru

Serzhantova Natalia A., Penza State Technological University, Baydukov Av. / 1a/11 Gagarin St., Penza, 440039, Russian Federation

Cand. Sci. (Engineering), Associate Professor, author ID 40462363600, e-mail: itmmbspgta@yandex.ru

The article discusses some of the existing criteria for assessing the performance of neural networks (ANNs) to solve classification problems. These criteria include the following: training error functions (sum of squared errors, root-mean-square error, variable error and mean absolute training error); the error of vessel thickness estimation; the probability of first and second type errors occurrence; the share of incorrectly classified objects in the testing sample. The authors reveal the main disadvantages of these criteria. To assess the results of the work of the ANN (exemplified by the problems of classification parameters of hemostasis, such as coagulation and blood viscosity), the paper uses advanced statistical analysis based on Bayesian approach. It goes on to propose an original criterion for assessing the quality of ANN performance to solve classification problems in medical diagnosis; the criterion takes into consideration the threshold value of ANN making prognostically right decision. The authors developed an algorithm that estimates the quality of ANN performance based on the proposed quality criteria. This algorithm allows varying the value of significance level for different tasks, thereby establishing the necessary and sufficient values that define the performance of ANN under study as efficient.

Keywords: neural network, classification, medical diagnosis, criteria, quality, significance level, the decision threshold, efficiency, information capacity, error, omission, false alarm

Введение. В настоящее время в медицине активно используются различные информационные технологии, в т.ч. информационные системы, автоматизированные диагностические системы и пр. [6]. Одним из наиболее востребованных и перспективных направлений внедрения разработок в данной области является решение задач диагностики, разработка и внедрение систем поддержки принятия решений (СППР) врача – в т.ч. при клинической диагностике. Прежде всего, это обусловлено большим объемом и сложным характером анализируемых данных (особенно при массовых диагностических обследованиях [13]), результаты которых клиницист не может в полной мере осмыслить и учесть в своих выводах [14].

Как показывает практика, традиционные медицинские экспертные системы не находят широкого применения в лечебных учреждениях, несмотря на достаточно высокий уровень использования компьютерной техники. Это в значительной степени связано с особенностями медико-биологической информации. Традиционные системы нуждаются в детальной проработке математических алгоритмов, с помощью которых каждая такая система должна решать задачу классификации и делать необходимые выводы (например, ставить диагноз). Большое количество факторов, оказывающих прямое и косвенное влияние на конечный результат диагностики, затрудняет создание новых алгоритмов, а уже разработанные алгоритмы не обладают необходимой точностью и приемлемым временем работы [4, 15].

Выход видится в применении нейросетевых технологий (НСТ) для решения задач клинической диагностики. Внедрение таких систем позволяет значительно ускорить процесс постановки предварительного диагноза [6, 15, 19, 22–24]. Однако, чтобы оценить достоверность полученных результатов, необходимо проводить процедуру тестирования нейросетевой системы для оценки качества ее работы.

В связи с этим, **целью исследования** является выявление критериев оценки качества работы нейронных сетей (НС), и разработка алгоритма оценки качества работы НС при постановке предварительного диагноза.

Общая характеристика проблематики работы. При решении задач с применением НСТ, разработчик создает информационную модель G некоего реального процесса или системы F . В общем случае модель G является неполной, так как пространства входных и выходных переменных не могут, содержать все параметры, описывающие процесс или систему F . Это происходит из-за отсутствия необходимой информации о моделируемой системе, намеренного сокращения количества параметров (например, выбора только информативных признаков) или ограниченного объема данных, имеющихся в распоряжении у разработчика для обучения и тестирования нейросетевой модели (НСМ). Все это может негативно влиять на результат использования НСМ.

В массиве экспериментальных данных, используемых для обучения и тестирования НС, всегда присутствуют ошибки разной природы, шум, а также противоречия между отдельными измерениями. Особенно это характерно для клинических данных, в которых границы «нормы-патологии» часто перекрываются, применяются понятия нозологической и индивидуальной нормы и т.д. Кроме того, обучающие множества могут содержать пропущенные значения (например, вследствие потери информации, невозможности проведения полного набора анализов, и т.п.). Таким образом, получаются разреженные матрицы, опи-

сывающие обучающие, тестовые и исследуемые множества данных. При этом пропущенные значения могут трактоваться как нулевые, минимальные, средние и т.д. Произвольность в интерпретации этих значений ухудшает свойства модели и результатов ее использования. Такие особенности в данных и в постановке задач требуют особого контроля ошибок информационных моделей [11, 18, 21, 23]. Однако в настоящее время не существует единого подхода к определению того, насколько качественной является нейросетевая информационная модель, насколько эффективно разработанная НС справляется со своими функциями.

Оценка существующих критериев качества работы нейронных сетей. В процессе настройки НС подбираются ее внутренние параметры (веса и смещения) таким образом, чтобы на выходе получался ряд значений, максимально соответствующий матрице желаемых результатов для обучающего множества [2, 22]. В связи с этим наиболее распространенным критерием качества работы НС является оценка ошибки обучения, то есть оценка разности между желаемым результатом и выходными значениями на выходе НС после завершения процесса обучения (функция ошибок).

Следует отметить, что оценки ошибки обучения используются для характеристики НС, реализующих алгоритмы обучения с учителем (сети прямой передачи сигнала *Feed-forward backprop*; сети Элмана *Elman backprop*; радиально-базисные сети *Radial Basis Function* и т.д.). Дальнейшая оценка критериев качества работы НС проводилась на примере базовой архитектуры сети прямой передачи сигнала *Feed-forward backprop*.

Если ошибка сети, выходной слой которой имеет n нейронов, $e_i = y_i - t_i$ представляет собой разность между реальным и желаемым сигналами на выходе i -го нейрона, то в качестве характеристики ошибок могут быть использованы следующие функции [7]:

- сумма квадратов ошибок $sse = \sum_{i=1}^n e_i^2$;

- средняя квадратичная ошибка $mse = \frac{1}{n} \sum_{i=1}^n e_i^2$;

- регулируемая или комбинированная ошибка

$$mse_{reg} = \frac{\gamma}{n} \sum_{i=1}^n e_i^2 + \frac{1-\gamma}{n} \sum_{i=1}^n e_i^2, \text{ где } \gamma - \text{параметр регуляции};$$

- средняя абсолютная ошибка $mae = \frac{1}{n} \sum_{i=1}^n |e_i|$.

Однако методы оценки качества работы НС, основанные на функциях ошибок, имеют ряд недостатков.

1. Отсутствует какая-либо шкала для значений функции ошибок, по которой можно оценить, насколько хорошо работает та или иная НС.

2. При таком подходе минимизируют не ошибку, которую можно ожидать от сети, когда ей будут подаваться совершенно новые наблюдения, а ошибку генерации сетью результатов, соответствующих желаемым для обучающего множества. Так как обучающее множество конечно и ограничено, то данные параметры не соответствуют ошибке в заранее неизвестной модели явления. При достаточно большом количестве эпох (итераций) процесса настройки весов и смещений (НС), ошибку обучения можно практически свести к нулю. Однако при поступлении новых, не использовавшихся при обучении данных на вход НС гарантировать стопроцентную точность результатов? исходя из пренебрежительно малой ошибки обучения невозможно. Добиваясь большей точности решения на обучающем множестве, исследуемая сеть теряет часть обобщающей способности, что негативно сказывается на результате работы НС [1, 16].

Неизвестная ошибка, допускаемая НС на данных, не использовавшихся при обучении, называется ошибкой обобщения модели E_G . Основной целью при построении НСМ является уменьшение именно ошибки обобщения.

Для получения оценки ошибки обобщения анализируется часть данных, из имеющейся базы, не использовавшихся при обучении, для которых известны желаемые результаты. Эта выборка примеров называется тестовой выборкой [1]. Ошибка обобщения оценивается, как характеристика неверных ответов НСМ на множестве примеров из тестовой выборки.

Существуют различные методы оценки ошибки обобщения. Например, Н.Ю. Ильясова, Д.Е. Липка, А.В. Куприянов [5] критерием качества работы сети для определения параметров сосудов сетчатки глаза выбрали погрешность оценки толщины сосуда, вычисляемую по формуле:

$$\bar{\delta} = \left(\sum_{x=1}^m \frac{|\sigma_{\varepsilon}(x) - \sigma(x)|^2}{\sigma(x)^2} \right) 100\%,$$

где $\sigma(x)$ – радиус сосуда в точке трассы, использованный при экспертной оценке тестового изображения, $\sigma_{\varepsilon}(x)$ – оценка радиуса, полученная с помощью разработанного авторами [5] алгоритма.

Другим примером оценки качества работы НС является критерий, разработанный [11] для определения надежности средств биометрической аутентификации с использованием образов «Свой» и «Чужой». Качество работы НС зависит от вероятностей ошибок первого (P_1) и второго (P_2) рода. Комплексный показатель качества работы НС в данном случае имеет вид:

$$Q_{НС} = \frac{1}{\sqrt{P_1 \cdot P_2}}.$$

Также в роли критерия качества работы НС может выступать доля неправильно классифицированных объектов тестовой выборки (P):

$$P = \frac{n}{N} 100\%,$$

где n – число неправильно классифицированных объектов в тестовой выборке, N – общее число объектов в тестовой выборке [8].

Перечисленные выше критерии позволяют количественно оценить качество работы НС при решении задач классификации. Однако они обладают одним общим недостатком: в них не учитывается то, как интерпретируется результат работы сети с конкретным примером, и какую погрешность вносит аналитик, устанавливая границы отнесения результата к конкретному классу.

Например, нормой в двухслойной НС считаются нулевые значения на выходе НС, а патологией – единичные. Допустим, в результате работы НС мы получили некий результат, находящийся в диапазоне от 0 до 1, скажем, 0,81. С одной стороны, данный результат можно интерпретировать как патологический, так как он ближе к «1». С другой – пороговое значение принятия/отвержения гипотезы об отнесении данного результата к патологии должно быть не менее 0,8. Это автоматически снижает точность и надежность разработки и, следовательно, степень доверия к конечному результату.

Таким образом, актуальной является задача создания нового алгоритма оценки качества работы НС для решения задач классификации. По мнению авторов статьи, новый алгоритм должен базироваться на оценке ошибки обобщения НС и учитывать погрешности, связанные с установлением порогов классификации.

Алгоритм оценки качества работы нейронных сетей при решении задач классификации для медицинской диагностики. В вычислительной диагностике приняты следующие термины и определения [17].

Признак – любой заранее сформулированный вопрос, касающийся определенных характеристик, полученных клиническими, лабораторными и инструментальными методами исследования, применяемыми при конкретной болезни. Так признаком могут быть данные анамнеза, измеренные параметры, результаты проведения инструментальных исследований и т.д.

Симптом – заранее сформулированный ответ на любой из вопросов. Симптомом может быть наличие или отсутствие какого-либо признака, увеличение или уменьшение его значения по сравнению с нормой и т.д.

Симптокомплекс – любое сочетание симптомов, относящихся к разным признакам.

Частота симптома – количество случаев в группе наблюдений, в которых встречается рассматриваемый симптом при данной болезни.

Вероятность симптома – отношение частоты симптома при данной болезни к объему группы наблюдений.

Условная вероятность некоторого события относительно другого события – отношение вероятности их совместного появления к вероятности второго события.

Нейронные сети, предназначенные для решения задач классификации (диагностики), обычно работают с ограниченной группой больных d , страдающих различными заболеваниями ($D1, D2, \dots, Dd$). Имеется также список S параметров – результатов анализов, обозначаемых $S1, S2, \dots, Ss$. Для любого анализируемого заболевания Dj , можно на основе накопленного материала определить вероятность $P(S/Dj)$ наблюдения симптома S . Для репрезентативных выборок можно также определить априорную вероятность наличия данного заболевания $P(Dj)$ и апостериорную вероятность $P(Dj/S)$ каждого заболевания при данном симптома комплексе.

Пусть требуется решить задачу диагностики, при которой результаты работы НС группируются в два класса: «норма» и «патология». При этом НС имеет два выхода.

Для оценки результатов работы НС был применен расширенный аппарат статистического анализа, опирающийся на байесовский подход [10, 17]. При оценке качества эксперимента введены (использованы) следующие обозначения: α – уровень значимости или вероятность ошибки первого рода; a – частота регистрации патологии первым выходом НС; b – частота регистрации нормы вторым выходом НС; c – частота отсутствия патологии на первом выходе НС; d – частота отсутствия нормы на втором выходе НС; $Y_{\phi 1i}$ – i -е значение на первом выходе («патология») НС; $Y_{\phi 2i}$ – i -е значение на втором выходе («норма») НС; Y_{p1i} – i -ое значение первого столбца матрицы желаемых результатов (МЖР); Y_{p2i} – i -ое значение второго столбца матрицы желаемых результатов; $D1$ – патология; $D2$ – норма; S – совокупность параметров на основе которых производится нейросетевой анализ (то есть симптокомплекс); n – общее число наблюдений; $P(D1)$ – априорная вероятность развития патологии, определяется исходя из частоты патологических значений, верно интерпретированных НС; $P(D2)$ – априорная вероятность нормы - исходя из частоты нормальных значений, верно интерпретированных НС; $P(S/D1)$ – условная вероятность развития патологии; $P(S/D2)$ – условная вероятность нормы [14].

Для оценки качества эксперимента была предложена математическая структура записи показателей информативности результатов, генерируемых НС:

$$\Delta_{1i} = |Y_{p1i} - Y_{\phi 1i}|;$$

$$\Delta_{2i} = |Y_{p2i} - Y_{\phi 2i}|;$$

$$a_i = \begin{cases} z_1 = Y_{\phi 1i} & \text{при } \Delta_{1i} < \alpha, \\ z_1 = 0 & \text{при } \Delta_{1i} > \alpha; \end{cases} \quad b_i = \begin{cases} z_2 = Y_{\phi 2i} & \text{при } \Delta_{2i} < \alpha, \\ z_2 = 0 & \text{при } \Delta_{2i} > \alpha; \end{cases}$$

$$c_i = \begin{cases} z_3 = 1 & \text{при } Y_{p1i} = 0; \\ z_3 = 0 & \text{при } Y_{p1i} = 1; \end{cases} \quad d_i = \begin{cases} z_4 = 1 & \text{при } Y_{p2i} = 0; \\ z_4 = 0 & \text{при } Y_{p2i} = 1; \end{cases}$$

$$a = \sum_{i=1}^n a_i; \quad b = \sum_{i=1}^n b_i; \quad c = \sum_{i=1}^n c_i; \quad d = \sum_{i=1}^n d_i;$$

$$P(S/D1) = a / (a + c); \quad P(S/D2) = b / (b + d);$$

$$P(D1) = (a + b) / n; \quad P(D2) = (c + d) / n.$$

При этом уровень значимости α определяет пороговое значение, позволяющее установить некоторые доверительные пределы (принятия или отвержения гипотезы) при определении того, к какому из классов принадлежит значение на выходе НС.

В ходе оценки качества эксперимента необходимо рассчитать следующее:

- коэффициент корреляции по методу Юла для четырехпольной таблицы

$$r = \frac{a * d - b * c}{\sqrt{(a + c)(b + d)(a + b)(c + d)}} \quad [9];$$

- количество информации по Шеннону $\alpha = \log_2 \frac{P(Di/S)}{P(Di)}$;
- диагностические коэффициенты $DK_i = 10 \lg \frac{P(S/D1)}{P(S/D2)}$;
- информативность по Кульбаку $Ik = 10 \lg \frac{P(S/D1)}{P(S/D2)} (P(S/D1) - P(S/D2))$.

Полученные результаты считаются информативными, если одновременно выполняются следующие условия: $|r| \geq 0,25$; $|DK| \geq 3$; $|Ik| \geq 0,25$ [1, 20].

Для оценки работы НС также подсчитывается эффективность нейросетевой системы, включающая такие показатели, как чувствительность и специфичность [3, 22].

Чувствительность $Se = TP / (TP + FN)$ – доля больных пациентов, для которых был правильно поставлен диагноз наличия болезни (вероятность правильного обнаружения заболевания). TP – число правильных диагнозов «патология»; FN – число пропусков «патологии».

Специфичность $Sp = TN / (TN + FP)$ – доля здоровых пациентов, для которых было правильно диагностировано отсутствие болезни (вероятность отсутствия ложного диагноза у здоровых или больных другой болезнью). TN – число правильных заключений «норма»; FP – число ложных диагнозов «патология».

Диагностическая эффективность (безошибочность) $ДЭ = (TP + TN) / (TP + FN + TN + FP)$ – относительная частота принятия безошибочных решений.

Таким образом, алгоритм оценки качества работы НС при решении задач классификации (диагностики) состоит из следующих этапов:

- формирование тестового множества и МЖР;
- определение уровня значимости α ;
- тестирование НС;
- получение значений на выходах НС, путем определения разности между фактическими и желаемыми результатами работы НС;
- расчет вспомогательных параметров;
- расчет показателей информативности;
- расчет показателей эффективности работы НС;
- интерпретация и визуализация полученных результатов.

Схема алгоритма оценки качества работы НС представлена на рисунке 1.

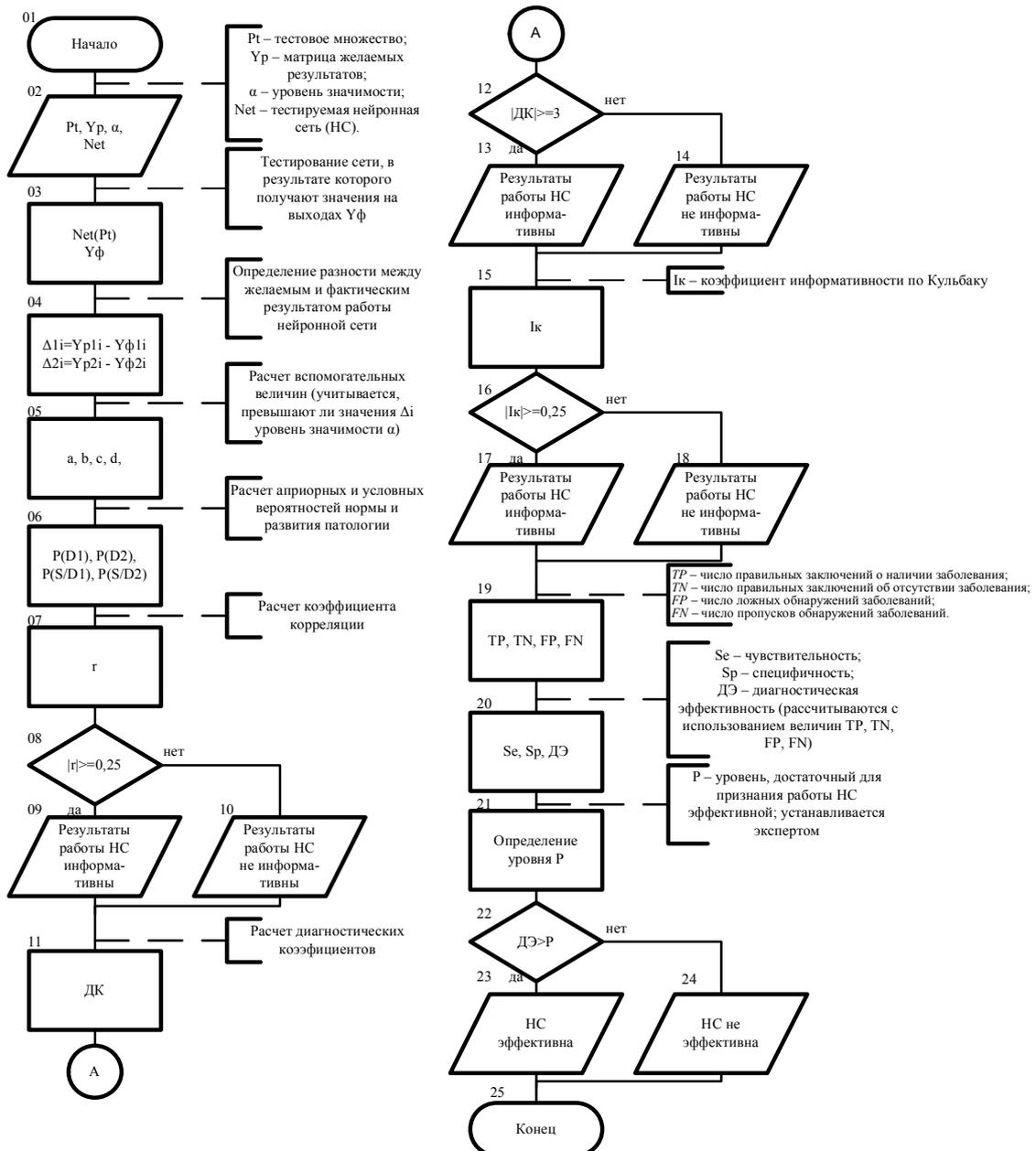


Рисунок 1 – Схема алгоритма оценки качества работы НС

Разработанный алгоритм позволяет оценивать результат нейросетевой диагностики с точки зрения информативности, безошибочности и эффективности с заранее заданной степенью достоверности.

Материал и методика тестирования разработанного алгоритма. Тестирование разработанного алгоритма оценки качества работы НС при решении задачи диагностики проводилось на примере выявления патологий сердечно-сосудистой системы в ходе оценки параметров гемостаза таких, как свертываемость и вязкость крови. Эти параметры в медико-биологической практике анализируются при диагностике наличия у пациентов тромбозов и эмболий [12, 23, 24].

НС должна решать задачу классификации следующим образом. На вход НС подаются значения шести различных, наиболее информативных и статистически обоснованных параметров, характеризующих свертывающую способность крови (время свертывания, активированное время рекальцификации плазмы, протромбиновое время, тромбиновое время, активированное тромбопластиновое время, концентрация фибриногена) [13]. Обучающее и тестовое множества сформированы на основе реальных клинических данных. Обучающее множество содержит 100 наборов значений параметров свертывания. Из них 50 значений норма и 50 – патология различной степени тяжести. Тестовое множество включает 20 наборов значений параметров свертывания, не использовавшихся для обучения НС.

Авторами была создана двухслойная НС способная анализировать значения параметров свертывания и классифицировать состояние каждого отдельного пациента в выборке как нормальное или патологическое (рис. 2).

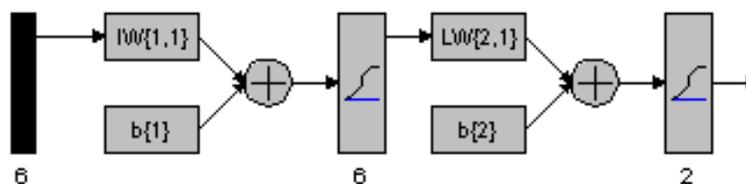


Рисунок 2 – Используемая структура двухслойной НС

После обучения был проведен анализ результатов классификации биомедицинских данных НС, который заключался в сравнении выходных множеств с контрольной матрицей желаемых результатов.

Интерпретация результатов классификации (диагностики) сильно зависит от порогов принятия решения о принадлежности результата к тому или иному классу. Авторы предлагают устанавливать порог в соответствии с выбранным уровнем значимости. А именно – для данного уровня определяется вероятность верной интерпретации патологически значений НС $P(D1)$, вероятность верной интерпретации нормы НС $P(D2)$, вероятности наличия $P(S/D1)$ и отсутствия $P(S/D2)$ заболевания для рассматриваемой совокупности симптомов.

В таблице 1 представлены результаты расчета вспомогательных и основных показателей информативности результатов нейросетевой классификации 20 наборов параметров свертывания крови (тестового множества).

Таблица 1 – Показатели информативности результатов нейросетевого анализа параметров свертывания крови при диагностике тромбоза

α	a	b	c	d	P(S/D1)	P(S/D2)	P(D1)	P(D2)	r	ДК	Ik
0,01	0,01	1,99	13	7	0,00	0,22	0,10	1	-0,38	-7,80	-30,05
0,05	0,11	4,90	13	7	0,01	0,41	0,25	1	-0,50	-5,65	-20,96
0,32	6,94	7	13	7	0,35	0,50	0,70	1	-0,15	-0,52	-9,75

Рисунок 3 иллюстрирует влияние изменения порога принятия гипотезы о принадлежности результата к тому или иному классу.

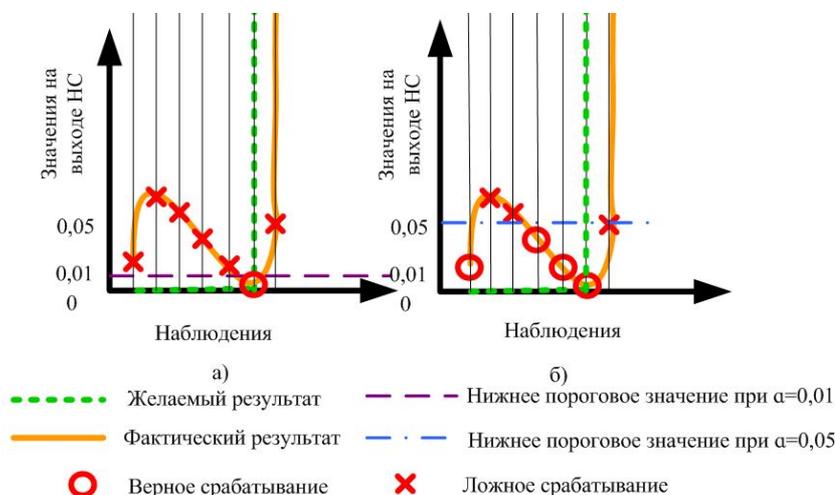


Рисунок 3 – Влияние изменения порога принятия гипотезы о принадлежности результата

Первые 6 наблюдений (т.е. результатов нейросетевой диагностики для 6-ти пациентов) должны быть интерпретированы как «норма», 7 наблюдение для достижения желаемого результата должен быть отнесен к классу «патология». Сеть во всех 7 случаях генерирует довольно близкие результаты в диапазоне от 0 до 0,1. Слева (рисунок 3а) показан результат интерпретации при установке нижнего порогового значения, соответствующего $\alpha = 0,01$. В этом случае, с желаемым результатом совпадает 1 фактическое значение из 7. Справа (рис. 3б) для $\alpha = 0,05$ желаемому результату соответствует 4 из 7 фактических.

Выбор более низких значений α приводит к увеличению вероятности обнаружения в исследуемом множестве наборов значений интерпретируемых как «норма» или «патология», однако информативность результата классификации снижается в 0,5–3 раза, что подтверждается результатами расчетов, приведенными в таблице 1.

В таблице 2 приведены вспомогательные и основные показатели эффективности работы НС для различных уровней значимости α . Также в таблице 2 для сравнения приведен результат расчета доли неправильно классифицированных объектов тестовой выборки [8]. N в данном случае составляет 20 наблюдений, а n представляет собой сумму ложных пропусков и ложных срабатываний, то есть FP и FN.

Таблица 2 – Показатели эффективности работы НС

α	TP	TN	FP	FN	Se	Sp	ДЭ	$P = n / N * 100 \% = (FP + FN) / N * 100 \%$
0,01	2	5	0	18	0,1	1	0,28	90 %
0,05	5	5	0	15	0,25	1	0,4	75 %
0,32	18	7	0	2	0,9	1	0,93	10 %

Из таблицы 2 видно, что относительная частота принятия безошибочных решений (ДЭ) для $\alpha = 0,32$ существенно выше, чем для $\alpha = 0,05$ или $\alpha = 0,01$. Доля неправильно классифицированных объектов тестовой выборки при этом уменьшается, то есть полученные результаты не противоречат друг другу. Однако в [8] не указано, каким образом следует выбирать порог принятия/отвержения гипотезы о принадлежности результата нейросетевого анализа к тому или иному классу, а, следовательно, результат оценки может варьироваться от 90 до 10 % ошибок. Обеспечить воспроизводимость результата оценки одних и тех же данных разными исследователями при этом также не представляется возможным.

Увеличение диагностической эффективности ДЭ при выборе более низкого уровня значимости ($\alpha = 0,05$, $\alpha = 0,1$ или даже $\alpha = 0,32$) объясняется тем, что задание высокого уровня значимости значительно сужает границы отнесения значения на выходе НС к тому или иному классу. Установленные высокие пороги принятия/отвержения обуславливают большое количество результатов, которые считаются «неопределенным», то есть не относятся ни к одному из классов. А это в свою очередь приводит к низким показателям диагностической эффективности НС.

С другой стороны, расширение доверительных интервалов и установление более низких порогов принятия/отвержения позволяет более гибко интерпретировать результаты работы НС, и получать приемлемые значения показателя диагностической эффективности, однако снижает общие технические характеристики нейросетевой системы, в частности достоверность результата диагностики. Уровень статистической значимости таких систем недостаточен для применения в клинической практике.

Варьирование уровня значимости при установлении порогов принятия/отвержения гипотезы о принадлежности результата нейросетевого анализа к тому или иному классу может быть полезно на этапе формирования обучающего множества и выбора архитектуры НС. В случае, если НС способна разграничивать классы с приемлемой ДЭ на достаточно низком уровне значимости (ниже 0,05), то следует сосредоточиться на подборе оптимальных алгоритмов обучения и адаптации. Если же НС генерирует неопределенный результат независимо от порога принятия решения, то это свидетельствует о том, что обучающее множество не является репрезентативным, сбалансированным, и не удовлетворяет требованиям необходимости и достаточности в отношении количества исследуемых параметров.

Выводы.

1. В настоящее время не существует единого подхода к определению того, насколько качественной является нейросетевая информационная модель, насколько эффективно разработанная НС справляется со своими функциями.

2. Ошибка обучения НС не может являться эффективным критерием оценки качества работы НС, т.к. она не отражает обобщающую способность сети.

3. Большинство существующих критериев оценки качества работы НС, основанных на оценке ошибки обобщения не учитывают то, как интерпретируется результат работы сети с конкретным примером, и какую погрешность вносит аналитик, устанавливая границы отнесения результата к конкретному классу.

4. Предлагается использовать алгоритм оценки качества работы НС с возможностью варьирования порога принятия решения о принадлежности результата к тому или иному классу в соответствии с выбранным уровнем значимости α .

5. Выбор более низких значений α приводит к увеличению вероятности обнаружения в исследуемом множестве наборов значений интерпретируемых как «норма» или «патология», однако информативность результата классификации снижается. Поэтому необходимо выбирать уровень значимости таким образом, чтобы обеспечивать баланс между высокой диагностической эффективностью и требуемой достоверностью результата.

6. Разработанный алгоритм позволяет варьировать значение уровня значимости для различных задач, тем самым, определяя необходимые и достаточные значения, при которых работа исследуемой НС удовлетворяет требованиям диагностической эффективности, а результат остается достоверным на принятом уровне значимости.

Список литературы

1. Горбань А. Н. Нейроинформатика / А. Н. Горбань, В. Л. Дунин-Барковский. – Москва : ИНТУИТ, 2016. – С. 331.
2. Грибачев В. Настоящее и будущее нейронных сетей / В. Грибачев // Компоненты и технологии. – 2006. – № 5. – С. 146–150.
3. Диагностическая эффективность методов врачебно-педагогического контроля в оценке функционального состояния спины у детей среднего школьного возраста / А. Н. Богачев, И. Б. Грещкая. – Режим доступа: http://www.sp10vlg.ru/UserFiles/Diagnoz_spin.doc (дата обращения 03.12.2017), свободный. – Заглавие с экрана. – Яз. рус.
4. Зайцев В. М. Прикладная медицинская статистика : учебное пособие / В. М. Зайцев, В. Г. Лифляндский, В. И. Маринкин. – Санкт-Петербург : Фолиант, 2003. – С. 432.
5. Ильясова Н. Ю. Применение искусственных нейронных сетей для оценивания диагностических параметров на биомедицинских изображениях / Н. Ю. Ильясова, Д. Е. Липка, А. В. Куприянов. – Режим доступа: <http://www.computeroptics.smr.ru/KO/PDF/KO25/KO25328.pdf> (дата обращения 03.12.2017), свободный. – Заглавие с экрана. – Яз. рус.
6. Интегральная медицина XXI века: теория и практика. – Режим доступа: http://www.it-med.ru/library/lib_article (дата обращения 03.12.2017), свободный. – Заглавие с экрана. – Яз. рус.
7. Калацкая Л. В. Организация и обучение искусственных нейронных сетей / Л. В. Калацкая, В. А. Новиков, В. С. Садов. – Минск : БГУ, 2002. – С. 76.
8. Краснянчин Я. Н. Некоторые аспекты параметризации искусственных нейронных сетей в задачах качественного химического анализа / Я. Н. Краснянчин, А. В. Пантелеймонов, Ю. В. Холин // Вісник Харківського національного університету. – 2010. – № 932. Хімія, вип. 19(42). – С. 170–181.
9. Лакин Г. А. Биометрия / Г. А. Лакин. – Москва : Высшая школа, 1991. – С. 352.
10. Лапач С. Н. Статистические методы в медико-биологических исследованиях с использованием Excel / П. Н. Бабич, С. Н. Лапач, А. В. Чубенко. – Киев : Морион, 2000. – С. 408.
11. Секретов М. В. Комплексный показатель качества средств нейросетевой биометрической аутентификации / М. В. Секретов // Труды научно-технической конференции кластера пензенских предприятий, обеспечивающих безопасность информационных технологий. – 2012. – Т. 8. – С. 36–37.
12. Сидорова М. А. Система скринингового исследования параметров гемостаза / М. А. Сидорова, Н. А. Ерушова // Известия ЮФУ. Технические науки. – 2008. – № 5 (82). – С. 72–75.
13. Сидорова М. А. Обоснование выбора метода скринингового исследования патологий сердечно-сосудистой системы / М. А. Сидорова, Н. А. Сержантова // Известия ЮФУ. Технические науки. Тематический выпуск. «Медицинские информационные системы». – 2012. – № 9. – С. 139–144.
14. Сидорова М. А. Диагностика и прогнозирование исходов перитонита с помощью нейросетевой системы исследования параметров гемостаза / М. А. Сидорова, Н. А. Сержантова, Л. А. Филиппова // Медицинская техника. – 2011. – № 2. – С. 42–74.
15. Сидорова М. А. Некоторые аспекты применения компьютерных технологий нейросетевого прогнозирования в медицине / М. А. Сидорова, Н. А. Сержантова, В. А. Чулков // XXI век: итоги прошлого и проблемы настоящего плюс. – 2015. – № 4 (26). – С. 94–100.
16. Сорокоумова Д. А. Построение и обучение нейронной сети для решения задачи распознавания речи / Д. А. Сорокоумова, О. Н. Корелин, А. В. Сорокоумов // Труды Нижегородского государственного технического университета им. Р.Е. Алексеева. – 2015. – № 3 (110). – С. 77–84.
17. Спиридонов А. В. Методы и алгоритмы вычислительной диагностики / И. Н. Спиридонов, А. В. Самородов. – Москва : МГТУ им. Н.Э. Баумана, 2006. – С. 50.
18. Старченко И. Б. Распознавание мимических жестов на основе анализа электромиографического сигнала / И. Б. Старченко, Р. Ю. Будко // Прикаспийский журнал: управление и высокие технологии. – 2016. – № 1. – С. 39–50 ([http://hi-tech.asu.edu.ru/files/1\(33\)/39-50.pdf](http://hi-tech.asu.edu.ru/files/1(33)/39-50.pdf)).
19. Хлопкова О. А. Нейроэволюционный метод интеллектуализации принятия решений в условиях неопределенности / О. А. Хлопкова // Прикаспийский журнал: управление и высокие технологии. – 2015. – № 3. – С. 114–129 ([http://hi-tech.asu.edu.ru/files/3\(31\)/114-129.pdf](http://hi-tech.asu.edu.ru/files/3(31)/114-129.pdf)).
20. Юнкеров В. И. Математико-статистическая обработка данных медицинских исследований / В. И. Юнкеров, С. Г. Григорьев. – Санкт-Петербург : ВМедА, 2002. – С. 320.
21. Artificial Intelligence and Knowledge Engineering. – Режим доступа: <http://zeus.sai.msu.ru/7000/hardware> (дата обращения 03.12.2017), свободный. – Заглавие с экрана. – Яз. рус.
22. Krizhevsky A. Imagenet classification with deep convolutional neural networks / A. Krizhevsky, I. Sutskever, G. Hinton // Advances in Neural Information Processing Systems. – 2012. – № 25. – P. 1106–1114.

23. Escodro P. B. Differential diagnosis between aorto-iliac thrombosis and equine protozoal myeloencephalitis: case report / P. B. Escodro, P. F. R. Lopes, C. G. Gianini // *Arq. bras. Med. veter. Zootec.* – 2010. – Vol. 62, № 5. – P. 1048–1053.
24. Patil S. Neural network in the clinical diagnosis of acute pulmonary embolism / S. Patil, J. W. Henry, M. Rubinfire, et al. // *Chest.* – 1993. – Vol. 104. – P. 1685–1689.

References

1. Gorban A. N., Dunin-Barkovskiy V. L. *Neyroinformatika* [Neuroinformatics], Moscow, INTUIT Publ., 2016, pp. 331.
2. Gribachev V. Nastoyashchee i budushchee neyronnykh setey [Present and future of neural networks]. *Komponenty i tekhnologii* [Components and Technologies], 2006, no. 5, pp. 146–150.
3. *Diagnosticheskaya effektivnost metodov vrachebno-pedagogicheskogo kontrolya v otsenke funktsionalnogo sostoyaniya spiny u detey srednego shkolnogo vozrasta* [Diagnostic efficiency of the methods of medical-pedagogical control in the assessment of the functional condition of the back in children of secondary school age]. Available at: http://www.sp10vlg.ru/UserFiles/Diagnoz_spin.doc (accessed 03.12.2017).
4. Zaytsev V. M., Lifyandskiy V. G., Marinkin V. I. *Prikladnaya meditsinskaya statistika* [Applied medical statistics], Saint Petersburg, Foliant Publ., 2003, pp. 432.
5. Ilyasova N. Yu., Lipka D. E., Kupriyanov A. V. *Primenenie iskusstvennykh neyronnykh setey dlya otsenivaniya diagnosticheskikh parametrov na biomeditsinskikh izobrazheniyakh* [Application of artificial neural networks for estimation of diagnostic parameters on biomedical images]. Available at: <http://www.computeroptics.smr.ru/KO/PDF/KO25/KO25328.pdf> (accessed 03.12.2017).
6. *Integralnaya meditsina XXI veka: teoriya i praktika* [Integral medicine of XXI century: theory and practice]. Available at: http://www.it-med.ru/library/lib_article. (accessed 03.12.2017).
7. Kalatskaya L. V., Novikov V. A., Sadv V. S. *Organizatsiya i obuchenie iskusstvennykh neyronnykh setey* [Organization and training of artificial neural networks], Minsk, BSU Publ. House, 2002, pp. 76.
8. Krasnyanchin Ya. N., Panteleymonov A. V., Kholin Yu. V. Nekotorye aspekty parametrizatsii iskusstvennykh neyronnykh setey v zadachakh kachestvennogo khimicheskogo analiza [Some aspects of the parameterization of artificial neural networks in problems of qualitative chemical analysis]. *Vestnik Kharkovskogo natsionalnogo universiteta* [Bulletin of the Kharkov National University], 2010, no. 932. Chemistry, vol. 19 (42), pp. 170–181.
9. Lakin G. A. *Biometriya* [Biometrics], Moscow, Vysshaya shkola Publ., 1991, pp. 352.
10. Lapach S. N., Lapach S. N., Chubenko A. V. *Statisticheskie metody v mediko-biologicheskikh issledovaniyakh s ispolzovaniem Excel* [Statistical methods in medical-biological researches using Excel], Kiev, Morion Publ., 2000, pp. 408.
11. Sekretov M. V. Kompleksnyy pokazatel kachestva sredstv neyrosetevoy biometricheskoy autentifikatsii [Complex index of quality of neural network for biometric authentication]. *Trudy nauchno-tekhnicheskoy konferentsii klastera penzenskikh predpriyat, obespechivayushchikh bezopasnost informatsionnykh tekhnologiy* [Proceedings of the Scientific Conference of the Cluster of the Penza Enterprises, providing Security of Information Technologies], 2012, vol. 8, pp. 36–37.
12. Sidorova M. A., Erushova N. A. Sistema skrinigovogo issledovaniya parametrov gemostaza [System screening tests of the hemostatic parameters]. *Izvestiya YuFU. Tekhnicheskie nauki* [Proceedings of the YUFU. Technical Sciences], 2008, no. 5 (82), pp. 72–75.
13. Sidorova M. A., Serzhantova N. A. Obosnovanie vybora metoda skrinigovogo issledovaniya patologiy serdechno-sosudistoy sistemy [The rationale for selecting the method of screening studies of pathologies of the cardiovascular system]. *Izvestiya YuFU. Tekhnicheskie nauki. Tematicheskyy vypusk. «Meditsinskie informatsionnye sistemy»* [Izvestiya YuFU. Technical Sciences. Thematic Issue. «Medical Information System»], 2012, no. 9, pp. 139–144.
14. Sidorova M. A., Serzhantova N. A., Filippova L. A. Diagnostika i prognozirovaniye iskhodov peritonita s pomoshchyu neyrosetevoy sistemy issledovaniya parametrov gemostaza [Diagnosis and prediction of outcome of peritonitis by using neural network system study of hemostatic parameters]. *Meditsinskaya tekhnika* [Biomedical Engineering], 2011, no. 2, pp. 42–74.
15. Sidorova M. A., Serzhantova N. A., Chulkov V. A. Nekotorye aspekty primeneniya kompyuternykh tekhnologiy neyrosetevogo prognozirovaniya v meditsine [Some aspects of the use of computer technology neural network forecasting in medicine]. *XXI vek: itogi proshlogo i problemy nastoyashchego plyus* [XXI century: the Results of the Past and Challenges of the Present Plus], 2015, no. 4 (26), pp. 94–100.
16. Sorokoumova D. A., Korelin O. N., Sorokoumov A. V. Postroeniye i obucheniye neyronnoy seti dlya resheniya zadachi raspoznavaniya rechi [Building and training a neural network to solve the problem of speech recognition]. *Trudy Nizhegorodskogo gosudarstvennogo tekhnicheskogo universiteta im. R.E. Alekseeva* [Proceedings of the Nizhny Novgorod State Technical University named R. E. Alekseev], 2015, no. 3 (110), pp. 77–84.
17. Spiridonov A. V., Samorodov A. V. *Metody i algoritmy vychislitelnoy diagnostiki* [Methods and algorithms for computational diagnostics], Moscow, MSTU named N.E. Bauman Publ. House, 2006, pp. 50.
18. Starchenko I. B., Budko R. Yu. Raspoznavaniye mimicheskikh zhestov na osnove analiza elektromiograficheskogo signala [Recognition of facial gestures based on the analysis of electromyographic signal]. *Prikaspiyskiy zhurnal: upravleniye i vysokie tekhnologii* [Caspian Journal: Control and High Technologies], 2016, no. 1, pp. 39–50 ([http://hi-tech.asu.edu.ru/files/1\(33\)/39-50.pdf](http://hi-tech.asu.edu.ru/files/1(33)/39-50.pdf)).
19. Khlopokova O. A. Neyroevolyutsionnyy metod intellektualizatsii prinyatiya resheniy v usloviyakh neopredelenosti [Neuroevolutionary method of intellectualization of decision-making in conditions of uncertainty]. *Prikaspiyskiy zhurnal: upravleniye i vysokie tekhnologii* [Caspian Journal: Control and High Technologies], 2015, no. 3, pp. 114–129 ([http://hi-tech.asu.edu.ru/files/3\(31\)/114-129.pdf](http://hi-tech.asu.edu.ru/files/3(31)/114-129.pdf)).
20. Yunkerov V. I., Grigorev S. G. *Matematiko-statisticheskaya obrabotka dannykh meditsinskikh issledovaniy* [Mathematical-statistical processing of data for medical research], Saint-Petersburg, MMA Publ. House, 2002, pp. 320.

21. *Artificial Intelligence and Knowledge Engineering*. Available at: <http://zeus.sai.msu.ru/7000/hardware> (accessed 03.12.2017).

22. Krizhevsky A., Sutskever I., Hinton G. Imagenet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*, 2012, no. 25, pp. 1106–1114.

23. Escodro P. B., Lopes P. F. R., Gianini C. G. Differential diagnosis between aorto-iliac thrombosis and equine protozoal myeloencephalitis: case report. *Arq. brasil. Med. veter. Zootecn.*, 2010, vol. 62, no. 5, pp. 1048–1053.

24. Patil S., Henry J. W., Rubinfire M., et al. Neural network in the clinical diagnosis of acute pulmonary embolism. *Chest*, 1993, vol. 104, pp. 1685–1689.

РЕДАКЦИОННЫЙ КОММЕНТАРИЙ К СТАТЬЕ

Статья посвящена одному из направлений использования информационных технологий для поддержки принятия решений в сфере медицинской диагностики – применению нейронных сетей (НС) для обработки многопараметрических медицинских данных. В работе описан оригинальный подход к оценкам результатов работы НС, позволяющий варьировать значение уровня значимости для различных задач – с целью адаптации предложенных алгоритмов к особенностям решаемых задач.

По статье можно сделать следующие замечания. 1. При обработке многопараметрических медицинских данных для уменьшения размерности пространства показателей наряду с НС могут использоваться также методы «главных компонент» и «главных факторов». При этом диагностические заключения могут делаться на основе значений некоторых «синтетических показателей», построенных в виде линейных комбинаций значений натуральных показателей. 2. Поскольку речь идет о классификации данных с целью получения диагностических заключений, следовало бы хоть что-то сказать и о методах многомерного дискриминантного анализа. 3. Наряду с «бинарной классификацией», рассматриваемой в данной статье, имеет право на существование и подход с тремя видами «заключений»: есть патология; нет патологии; информации для принятия решения недостаточно. В последнем случае в клинической практике делается следующее: назначаются дополнительные исследования; принимаются во внимание заболевания, которыми страдали (страдают) родители, братья, сестры пациентов; учитывается статистика по заболеваниям лиц аналогичного пола и возраста со сходными результатами диагностики и пр. 4. В статье предлагаемые методы (подходы) рассмотрены применительно только к одному из направлений клинической диагностики, основанном на использовании инструментальных средств. Однако название статьи сформулировано достаточно общее (для медицинской диагностики вообще). Поэтому, наверное, не лишним было бы указать, для каких еще видов медицинской диагностики предлагаемые методы будут эффективны с точки зрения авторов (хотя бы в виде их перечисления).

УДК 378, 004.942

РАЗРАБОТКА СИСТЕМЫ УПРАВЛЕНИЯ ЖИЗНЕННЫМ ЦИКЛОМ ЭЛЕКТРОННЫХ ОБРАЗОВАТЕЛЬНЫХ РЕСУРСОВ¹

Статья поступила в редакцию 02.12.2017, в окончательном варианте – 06.02.2018.

Деев Михаил Викторович, Пензенский государственный университет, 440026, Российская Федерация, Пенза, ул. Красная, 40,
аспирант, e-mail: miqz@yandex.ru

Кравец Алла Григорьевна, Волгоградский государственный технический университет, 400005, Российская Федерация, г. Волгоград, пр. им. Ленина, 28,
доктор технических наук, профессор, e-mail: agk@gde.ru; ORCID <https://orcid.org/0000-0003-1675-8652>

Цель исследования – повышение эффективности принятия решений при управлении жизненным циклом электронных образовательных ресурсов (ЭОР), используемых в процессе непрерывной подготовки специалистов. Для достижения поставленной цели решены следующие задачи: анализ актуальных проблем подготовки специалистов с использованием ЭОР, описание и формализация жизненного цикла ЭОР в виде итеративной модели развития, анализ вариантов проектирования системы управления и поддержки жизненного цикла ЭОР для обеспечения возможностей массовой разработки и актуализации учебных материалов и ресурсов. Авторами проведено изучение актуальных проблем массовой разработки ЭОР, модернизации их в процессе использования. Предложены функциональная модель в виде IDEF0-диаграммы для системы управления жизненным циклом ЭОР; модель репозитория ЭОР в виде IDEF1X-диаграммы. Реализован прототип системы управления жизненным циклом ЭОР. Он базируется на применении системы управления контентом Alfresco, расширенной дополнительными модулями. Разработанный прототип системы обеспечивает поддержку всех этапов жизненного цикла ЭОР, позволяет минимизировать расходы при их создании и модернизации, уменьшить время отклика «системы разработки и поддержки ЭОР» на динамически изменяющиеся требования работодателей и законодательства.

Ключевые слова: электронные образовательные ресурсы, жизненный цикл, управление, модель жизненного цикла, образовательная среда, система управления контентом, электронное обучение

¹ Результаты работы получены при финансовой поддержке РФФИ в рамках грантов № 16-07-00031, 17-307-50010.