

# АВТОМАТИЗИРОВАННАЯ СИСТЕМА ДИАГНОСТИКИ ЗАБОЛЕВАНИЙ ЛЕГКИХ<sup>1</sup>

Н.С. Безруков<sup>(1)</sup>, Е.Л. Еремин<sup>(2)</sup>, Ю.М. Перельман<sup>(1)</sup>

<sup>(1)</sup>Дальневосточный научный центр физиологии и патологии дыхания СО РАМН, г. Благовещенск;

<sup>(2)</sup>Амурский государственный университет, г. Благовещенск

Предложена автоматизированная система диагностики бронхиальной астмы и хронической обструктивной болезни легких по показателям изокапнической гипервентиляции холодным воздухом и реоэнцефалографии. Система реализована на основе нейро-нечеткой сети.

## ВВЕДЕНИЕ

Компьютерные системы диагностики развиваются в двух основных направлениях [1]: по пути создания экспресс-систем, основанных на обработке опросников, и по пути создания систем на базе алгоритмов обработки медицинской информации, которая представляется, как правило, в виде результатов инструментальных обследований пациента.

Перед экспертом-медиком встает проблема поиска нужной информации и выбора алгоритмов извлечения знаний из нее для решения задачи диагностики. При анализе «вручную» большого количества информации в поисках нужной врачи теряют время, необходимое для оперативной медицинской помощи [2].

Современные способы автоматизированного извлечения знаний непосредственно из исходных данных отчасти способны снять остроту указанной проблемы. Тем более, что они уже успешно применяются для анализа рынка, в маркетинге и дру-

гих бизнес-приложениях [3]. Однако в области медицины методы интеллектуальной поддержки, основанные на компьютерной обработке и анализе данных, применяются пока недостаточно. Между тем автоматизированный анализ информации позволяет решать ряд медицинских задач, связанных, в частности, со стратегией лечения больного (относительно цены и качества), равномерной загрузкой объектов медицинского обслуживания (поликлиник, лабораторий, диагностических центров), а в особенности — с диагностикой заболеваний.

Актуальность задачи диагностики заболеваний, имеющих сходную симптоматику, подтверждается наличием различных программ. Например, одно из направлений федеральной целевой программы «Развитие пульмонологической службы России на 2002—2007 годы» состоит в разработке и внедрении в практическое здравоохранение России методов ранней диагностики таких заболеваний, как бронхиальная астма (БА) и хроническая обструктивная болезнь легких (ХОБЛ), а также выявления ошибок их диагностики. Решать данную проблему предлагается мультидисциплинарным объединением знаний из различных направлений медицины.

## 1. МЕТОДЫ ИССЛЕДОВАНИЯ

Бронхиальная астма, наряду с ХОБЛ, относится к числу наиболее распространенных хронических заболеваний органов дыхания, она встречается бо-

<sup>1</sup> Работа выполнена в рамках программы «Системный анализ и моделирование влияния экологических и социальных факторов на здоровье населения Дальневосточного региона на основе применения современных информационных технологий с целью прогнозирования и разработки систем принятия управленческих решений в социальной сфере для обеспечения национальной безопасности», утвержденной постановлением Президиума ДВО РАН №147 от 15 декабря 2005 г.

лее чем у 5 % населения, варьируя в различных регионах мира от 1 до 30 %. Данные официальной статистики относительно низкой распространенности заболеваний легких в России (БА выявлена у 1–2 % населения страны) не отражают истинной картины. Проведенные в последние годы эпидемиологические исследования в Москве, Новосибирске и других городах страны свидетельствуют о наличии симптомов БА у 5–7 % населения [4].

Основная причина гиподиагностики обусловлена длительное время существовавшими взглядами на БА как заболевание, обязательным проявлением которого является приступ экспираторного удушья. Другие симптомы болезни (преходящий дыхательный дискомфорт, затруднение дыхания на выдохе, приступообразный кашель, сухие, преимущественно экспираторные дистантные хрипы) расценивались как проявления ХОБЛ или предастмы. Такое представление неизбежно сужало круг больных БА, усложняло диагностику и препятствовало адекватному лечению. Как следствие — преобладание тяжелых форм заболевания, высокий процент выхода на инвалидность, низкое качество жизни больных и связанное с этим бремя, ложащееся на семью больного и общество в целом.

Фармако-эпидемиологические исследования [4], проведенные в России, свидетельствуют о значительном несоответствии между научно-обоснованными рекомендациями и практическим положением дел в лечении этой категории больных. Поэтому актуальна задача дальнейших эпидемиологических исследований заболеваний легких в различных регионах страны и последующая разработка на их основе методов диагностики БА и ХОБЛ.

Проведенное в Дальневосточном научном центре физиологии и патологии дыхания СО РАМН исследование биоэлектрической активности головного мозга и церебральной гемодинамики у больных БА и ХОБЛ выявило отклонения электроэнцефалографической и реоэнцефалографической картины по сравнению с контрольной группой практически здоровых лиц. Была выдвинута гипотеза о возможности диагностики БА, используя информацию о параметрах реоэнцефалограммы (РЭГ). В исследовании принимали участие группа здоровых людей и группы больных БА легкой (БА1) и средней степени тяжести (БА2).

Реоэнцефалография — это метод исследования пульсового кровенаполнения тканей головного мозга, основанный на способности организма изменять сопротивление переменному току высокой частоты в зависимости от количества протекающей через него крови [5]. Реограмма отражает

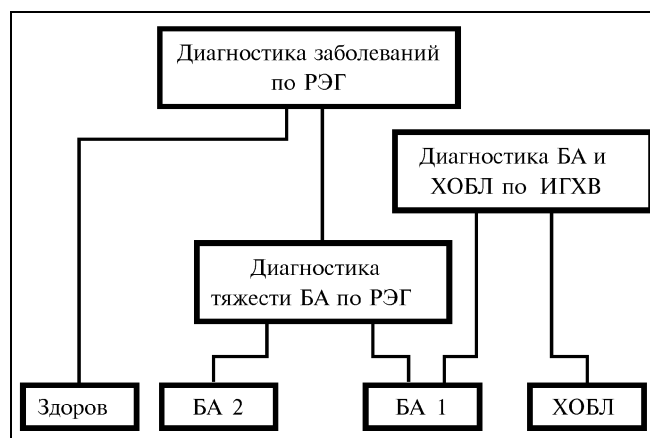


Рис. 1. Схема диагностики заболеваний легких

комплексное сопротивление живых тканей. Реографическое исследование дает 28 диагностических показателей, анализ которых позволит принимать решение о состоянии пациента.

Была исследована также реакция легких на изокпапническую гипервентиляцию холодным воздухом (ИГХВ) у больных БА и больных ХОБЛ легкой степени тяжести. Больным со средней степенью тяжести указанных заболеваний ИГХВ не проводилась в связи с принятыми ограничениями на применение бронхопровокационных проб.

В основе бронхоспастической реакции на ИГХВ лежит рефлекторное сокращение гладкой мускулатуры дыхательных путей под влиянием раздражения холодовых рецепторов гортани [6]. Для усиления охлаждающего эффекта воздуха дыхание во время пробы осуществляется в форсированном режиме. Реакция дыхательных путей на охлаждение оценивается через 1 мин после проведения пробы по изменению четырех диагностических показателей.

Предлагается следующая схема диагностики заболеваний легких (рис. 1), состоящая из трех подсистем. Диагностика по РЭГ обнаруживает заболевание у обследуемого, причем данная подсистема позволяет искать болезнь в режиме скрининга при массовом обследовании, так как отсутствуют ограничения на ее применение. В случае обнаружения заболевания диагностика по ИГХВ позволяет определить вид болезни — БА или ХОБЛ. Наконец, если известно, что у пациента БА, то подсистема диагностики по РЭГ позволит более точно дифференцировать тяжесть болезни.

Данная схема диагностики была реализована в автоматизированной системе диагностики заболеваний легких (АСДЗЛ).

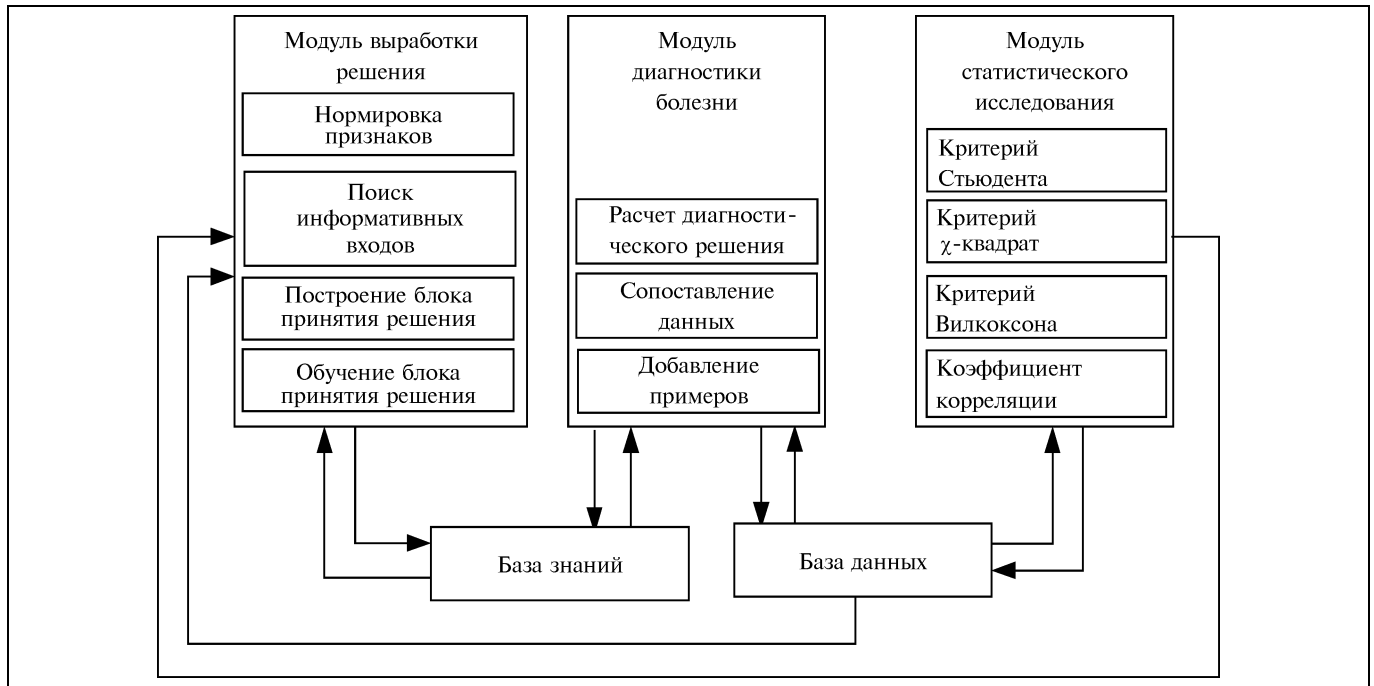


Рис. 2. Структура автоматизированной системы диагностики заболеваний легких

## 2. СТРУКТУРА АВТОМАТИЗИРОВАННОЙ СИСТЕМЫ ДИАГНОСТИКИ ЗАБОЛЕВАНИЙ ЛЕГКИХ

Предлагаемая структура АСДЗЛ (рис. 2) состоит из модулей, позволяющих:

- строить подсистемы диагностики;
- диагностировать болезни;
- проводить статистические исследования.

Модули работают с базой данных, где хранятся результаты проведенных исследований, и базой знаний, в которой находятся подсистемы диагностики.

Основная функция *модуля диагностики* болезни состоит в расчете диагностического решения о возможной болезни пациента, для чего модуль использует подсистемы диагностики и априорно определенные данные. Он позволяет дополнять подсистему по ходу работы новыми примерами и экспортировать их в базу данных.

Функция *модуля выработки решения* заключается в конкретизации структуры вывода диагностического решения с последующим обучением. Структура вывода, в зависимости от начальных условий (числа диагностических признаков и числа обучающих примеров), определяется алгоритмами, описанными в работах [7–9].

При построении структуры вывода диагностического решения необходимо решить три подзадачи: нормировать признаки к единому диапазону и к равномерному распределению; произвести по-

иск информативных входов; построить блок принятия решения.

Нормировка признаков осуществляется с помощью экспоненциальной функции и функционала энтропии по алгоритму, представленному в работе [7]. Поиск информативных входов производится на основе метода независимых компонент, который позволяет обнаруживать скрытые факторы, лежащие в основе набора случайных величин, измерений или сигналов [8]. Подробно задача поиска рассмотрена в Приложении. Блок принятия решения строится на основе каскадной нейро-нечеткой сети по алгоритму, представленному в работе [9]. Достоинство предлагаемой каскадной нейро-нечеткой сети — меньшее по сравнению с общепринятой адаптивной нейро-нечеткой сетью число настраиваемых коэффициентов. Параметры сети определяются в ходе обучения при помощи алгоритма обратного распространения ошибки на априорно определенных данных [10].

Функция *модуля статистических исследований* состоит в подтверждении или опровержении гипотез, выдвигаемых врачом-экспертом при помощи статистических критериев Стьюдента, Вилкоксона и  $\chi$ -квадрат, а также коэффициента корреляции [11]. По данным формализованным статистическим методам выбираются диагностические признаки при построении подсистем диагностики, а при увеличении базы данных отслеживаются изменения значимости признаков для диагностики болезни.

### 3. РЕЖИМЫ РАБОТЫ АВТОМАТИЗИРОВАННОЙ СИСТЕМЫ

Работа АСДЗЛ реализована в двух режимах — основном и обеспечивающем.

В основном режиме АСДЗЛ решает задачу о подтверждении или опровержении выдвигаемого диагноза. Задействуются модуль диагностики болезни, база правил и база данных. Из базы знаний в модуль диагностики подгружается подсистема диагностики, а результат диагностики представляется на априорно определенных примерах из базы данных; в соответствие им ставится новый результат. Подсистема может дополняться и новыми примерами.

В обеспечивающем режиме разрабатываются подсистемы диагностики. Задействуются модули статистического исследования и выработки решения, а также база правил и база данных. Структура функционирования данного режима соответствует рис. 2, но только без модуля диагностики болезни. При построении подсистемы из базы данных в модуль статистического исследования отправляются априорно определенные данные. В модуле с помощью статистических критериев выбираются диагностические признаки, по которым затем строится подсистема диагностики. Выбранные признаки отправляются в модуль выработки решения, где строится подсистема диагностики, сохраняющаяся затем в базе знаний.

В случае увеличения числа примеров обеспечивающий режим способен переобучить подсистему из базы правил в модуле выработки решения. Причем структура вывода решения не меняется, переобучается только каскадная нейро-нечеткая сеть с учетом новых априорно определенных данных. Это снимает проблему разработки новой подсистемы, от пользователя не требуется понимания работы подсистемы.

В рамках предложенной АСДЗЛ были разработаны три подсистемы диагностики: две по параметрам РЭГ [7] и одна по параметрам ИГХВ [12].

### 4. ПРОГРАММНАЯ РЕАЛИЗАЦИЯ АВТОМАТИЗИРОВАННОЙ СИСТЕМЫ

В качестве основного средства реализации АСДЗЛ был выбран математический пакет Matlab. Он представляет собой интерактивную систему, основным объектом которой является массив, не требующий явного указания размерности. Обладая множеством специализированных функций (статистический пакет, нейросетевой пакет и др.), данный программный продукт может быть использован для автоматизации требуемых вычислений

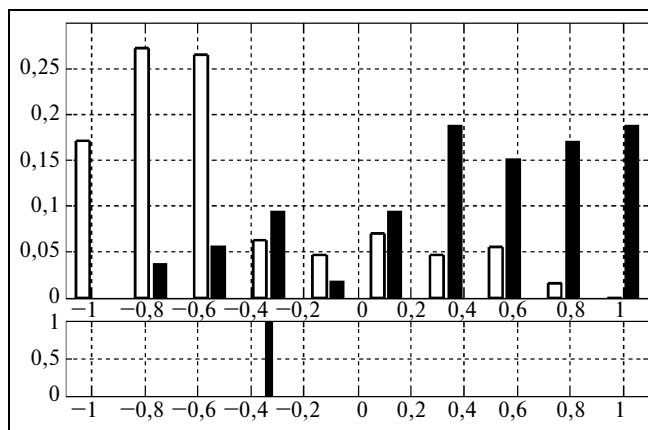


Рис. 3. Графическая интерпретация выхода подсистемы диагностики

практически в любой области науки и техники. Кроме того, по скорости выполнения вычислений его возможности превосходят потенциалы многих подобных систем.

Результаты проведенных исследований и созданных на их основе подсистем диагностики сохраняются в базе данных и базе знаний и отображаются в окне выбора модуля.

Загрузка диагностической подсистемы осуществляется через модуль диагностики заболеваний. Результат работы подсистемы дифференциальной диагностики БА и ХОБЛ по параметрам ИГХВ представляется в виде гистограммы (рис. 3) с априорно определенными данными, на которых подсистема обучалась, и ставящимися им в соответствие новыми данными. Если результат системы менее  $-0,4$ , то пациент скорее всего болен БА, а если более  $0,2$  — ХОБЛ.

Для запуска модуля статистических исследований необходимо выбрать базы данных в зависимости от диагноза в окне выбора модуля. К первому классу отнесли результаты исследований РЭГ для здоровых пациентов, а ко второму — результаты исследований РЭГ для пациентов с БА1 и БА2. Модуль исследований рассчитывает статистические критерии, что позволяет разработчику оценить значимость тех или иных признаков для диагностики болезни. В модуле имеется возможность графического отображения данных в виде топографических карт и гистограмм. Это весьма удобно, поскольку значения критериев свидетельствуют о существовании закономерностей в данных, в то время как рисунки наглядно демонстрируют характер этих закономерностей.

Диагностическая подсистема отражается в окне «Построение диагностической системы» (рис. 4) через меню «Модель/Создать». В окне допускается вести номера диагностических признаков, на ос-

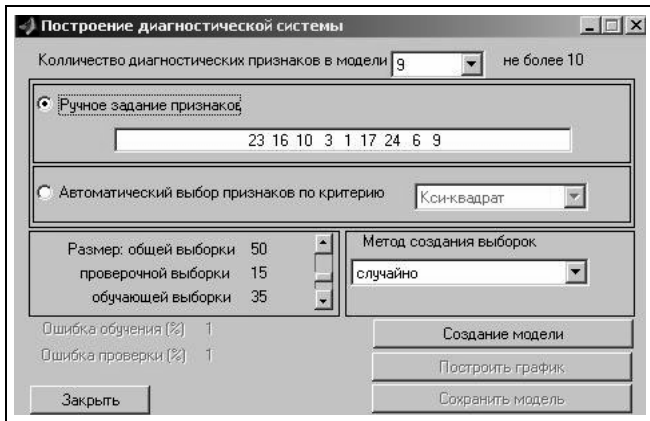


Рис. 4. Окно построения диагностической подсистемы

нове которых строится подсистема. Для проверки полученных результатов все имеющиеся данные разбиваются на проверочную и обучающую выборки. Размер и характер деления определяется пользователем. По разнице в ошибках при обчете системой обучающей и проверочной выборок пользователь судит, насколько адекватна, статистически значима построенная система.

Построенные подсистемы хранятся в базе знаний, откуда их можно в любой момент актуализировать для решения задач диагностики заболеваний легких.

## ЗАКЛЮЧЕНИЕ

Предложенные модули и реализованные на их основе режимы работы автоматизированной системы позволяют решать различные задачи, стоящие перед медицинскими работниками: от статистического анализа до построения подсистем диагностики и их применения для диагностики заболеваний легких на практике.

Способ диагностики бронхиальных заболеваний на основе РЭГ не имеет ограничений на применение, в отличие от способа ИГХВ, который способен вызвать чрезмерную бронхоспастическую реакцию, требующую неотложной помощи. Поэтому описанную автоматизированную систему диагностики можно применять при массовом обследовании (скрининге).

Предложенный алгоритм поиска информативных входов позволяет значительно сократить число входов, по которым строится блок принятия решений, значительно упрощает расчеты как на этапе проектирования, так и при работе системы на практике. Реализуемая в блоке принятия решений каскадная нейро-нечеткая сеть дает возможность

аппроксимировать имеющиеся в данных закономерности. «Мягкий» выход блока в диапазоне от  $-1$  до  $1$  помогает врачу выдвигать гипотезы о состоянии больного.

Дальнейшее развитие рассмотренной системы направлено на охват других заболеваний, связанных с поражением органов дыхания, и разработку для них систем диагностики посредством применения новых методов обследования, что поможет врачу в дифференциальной диагностике бронхолегочных заболеваний, а также в выборе эффективного лечения.

## ПРИЛОЖЕНИЕ

Поиск информативных входов реализуется в два этапа: на первом формируется уравнение с матрицей смешивания, которое необходимо решить, на втором вводится функционал, ограничивающий решение, и находится матрица смешивания.

### Этап 1. Формирование уравнения с матрицей смешивания

Пусть объект имеет  $n$  признаков  $\{x_1, x_2, \dots, x_n\}$ , по которым можно определить класс объекта из двух возможных  $\{y\}$ .

Необходимо сократить имеющиеся признаки до  $m$   $\{s_1, s_2, \dots, s_m\}$ , для чего объединяем имеющиеся признаки в  $m$  групп. Группы формируются из признаков, которые имеют линейную зависимость или по представлению эксперта:

$$\begin{aligned}\hat{x}_1 &= [x_1 \ x_2 \ \dots \ x_{k_1}]^T \\ \hat{x}_2 &= [x_1 \ x_2 \ \dots \ x_{k_2}]^T \\ &\dots \\ \hat{x}_m &= [x_1 \ x_2 \ \dots \ x_{k_m}]^T,\end{aligned}$$

где  $n = k_1 + k_2 + \dots + k_m$ .

Для полученных групп необходимо решить уравнение вида:

$$\begin{bmatrix} s_1 \\ s_2 \\ \vdots \\ s_m \end{bmatrix} = \begin{bmatrix} w_{11} & \dots & w_{1k_1} & 0 & \dots & 0 & \dots & 0 & \dots & 0 \\ 0 & \dots & 0 & w_{21} & \dots & w_{2k_2} & \dots & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \dots & 0 & 0 & \dots & 0 & \dots & w_{n1} & \dots & w_{nk_m} \end{bmatrix} \cdot \begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \\ \vdots \\ \hat{x}_m \end{bmatrix}$$

или

$$S = WX, \quad (1)$$

где  $W$  — матрица смешивания, подлежащая определению;  $S$  — независимые ненаблюдаемые выходы, подлежащие определению;  $X$  — зависимые наблюдаемые признаки.

### Этап 2. Формирование функционала

В рамках метода box-counting (подсчет ячеек) [3] было предложено понятие кросс-энтропии, позволяющее

качественно оценить энтропию по заполнению ячеек, на которые разбивается все пространство переменных:

$$I_{\text{box}}(x_1, x_2, \dots, x_d, y) = \log \frac{N_{x_1} \dots N_{x_d} N_y}{N_{xy}},$$

где  $N_{x_1}, \dots, N_{x_d}, N_y$  — число промежутков с присутствием координат  $x_1, x_2, \dots, x_d, y$  соответственно;  $N_{xy}$  — число клеток, в которых одновременно содержатся точки с координатами  $(x_1, x_2, \dots, x_d, y)$ .

Так как кросс-энтропия имеет очевидный недостаток (явно не представлена зависимость входов от выхода) и поэтому она способна реагировать на какую-либо зависимость между входами, то для ограничения решения уравнения (1) предлагается функционал вида:

$$F(s_1, s_2, \dots, s_m, y) = \log \frac{N_s}{N_{sy}} \rightarrow \max,$$

где  $N_s$  — число клеток, в которых одновременно содержатся точки с координатами  $(s_1, s_2, \dots, s_m)$ ;  $N_{sy}$  — число клеток, в которых одновременно содержатся точки обоих классов с координатами  $(s_1, s_2, \dots, s_m)$ .

Максимум функционала ищется путем перебора (сканирования) матрицы  $W$ .

## ЛИТЕРАТУРА

1. Гусев Л.А., Хуторская О.Е. Об одной оценке эффективности машины диагностики двигательных нарушений // Автоматика и телемеханика. — 2003. — № 12. — С. 112–121.
2. Арсеньев С. Извлечение знаний из медицинских баз данных / 1999 <<http://www.megaputer.ru>>.
3. Ежов А.А., Шумский С.А. Нейрокомпьютинг и его применение в экономике и бизнесе / 1998 <<http://www.neuroproject.ru>>.
4. Луни А.А. и др. Бронхиальная астма. Пособие для врачей. — Смоленск: Смоленская гос. мед. акад., 2003. — 66 с.
5. Зенков Л.Р., Ронкин М.А. Функциональная диагностика нервных болезней (Руководство для врача). — М.: МЕДпресс-информ, 2004. — 488 с.
6. Перельман Ю.М., Приходько А.Г. Диагностика холодовой гиперреактивности дыхательных путей. — Благовещенск: ГУ ДНЦ ФПД СО РАМН, 2004. — 10 с.
7. Безруков Н.С., Ермакова Е.В. Прогнозирование бронхиальной астмы по параметрам реоэнцефалографии на основе гибридной сети. Математические методы в технике и технологиях — ММТТ-19. — Воронеж, Воронеж. гос. технол. акад., 2006. — Т. 7. — С. 211–214.
8. Huvarinen A., Oja E. Independent Component Analysis: A Tutorial. — Laboratory of Computers and Information Science, Helsinki University of Technology. — Finland, 1999 <<http://www.cis.hut.fi/projects/ica/>>.
9. Безруков Н.С., Еремин Е.Л. Построение и моделирование адаптивной нейро-нечеткой системы в задачах медицинской диагностики // Информатика и системы управления. — 2005. — № 2 (10). — С. 36–46.
10. Рутковская Д., Пилинский В., Рутковский Л. Нейронные сети, генетические алгоритмы и нечеткие системы. — М.: Горячая линия; Телеком, 2004. — 452 с.
11. Сырцова Л.Е. и др. Основы эпидемиологии и статистического анализа в общественном здоровье и управлении здравоохранением. Ч. 1. Московская мед. акад. им. И.М. Сеченова. Учебное пособие для ординаторов и аспирантов. — Москва, 2003. — С. 92.
12. Безруков Н.С. и др. Система поддержки принятия решения для диагностики бронхиальной астмы и хронической обструктивной болезни легких // Бюллетень физиологии и патологии дыхания. — 2006. — Вып. 23 (приложение). — С. 32–34.

☎ (4162) 39-46-50, e-mail: bezrukow@mail.ru

Статья представлена к публикации членом редколлегии В.Н. Новосельцевым. □

## Новые книги

- Колесников А.А. Гравитация и самоорганизация. — М.: URSS, 2006. — 108 с.
- Неусыпин К.А. Системы управления летательными аппаратами и алгоритмы обработки информации. — М.: МГОУ, 2006. — 219 с.
- Пановко Я.Г. Устойчивость и колебания упругих систем. — М.: URSS, 2006. — 351 с.
- Системный подход к управлению информацией. — М.: URSS, 2006. — 324 с.
- Системы управления: инжиниринг качества. — М.: Вузовская книга, 2006. — 315 с.
- Экология и экономика природопользования. — М.: ЮНИТИ-ДАНА, 2007. — 591 с.
- Аалст Вил ван дер. Управление потоками работ. — М.: Физматлит, 2007. — 315 с.
- Александров И.О. Формирование структуры индивидуального знания. — М.: Ин-т психологии РАН, 2006. — 559 с.
- Ащепков Л.Т. Универсальные решения интервальных задач оптимизации и управления. — М.: Наука, 2006. — 151 с.
- Бодунов Н.А. Введение в теорию локальной параметрической идентифицируемости. — СПб., 2006. — 143 с.
- Лукьянов В.С. Модели анализа вероятностно-временных характеристик и структур сетей передачи данных. — Волгоград: Политехник, 2006. — 183 с.
- Максимов Н.В. Архитектура ЭВМ и вычислительных систем. — М.: ФОРУМ; ИНФРА-М, 2007. — 511 с.
- Основные философские направления и концепции науки и технoзнания. — СПб., 2006. — 420 с.
- Самарский А.А. Устойчивость разностных схем. — М.: УРСС, 2005. — 384 с.
- Тимошенко С.П. Колебания в инженерном деле. — М.: URSS, 2006. — 439 с.
- Шарапов В.М. Универсальные технологии управления. — М.: Техносфера, 2006. — 494 с.
- Петров К.Н. Как разработать бизнес-план. — М.: Вильямс, 2007. — 384 с. (+CD-Rom).
- Тидвелл Д. Разработка пользовательских интерфейсов. — СПб.: Питер, 2007. — 416 с.