2006 июль-сентябрь № 3

УДК 681.327.11:615.8

В.В. Анищенко¹, И.В. Бегун², В.В. Кулага³, С.В. Спиридонов³, К.В. Чижиков¹, Д.Л. Можейко¹

ТЕХНОЛОГИЯ АВТОМАТИЗИРОВАННОЙ ДИАГНОСТИКИ ЗАБОЛЕВАНИЙ ЛИМФАТИЧЕСКИХ УЗЛОВ ПО УЛЬТРАЗВУКОВЫМ ИЗОБРАЖЕНИЯМ

Рассматривается информационная технология автоматизации процесса диагностики заболеваний лимфатических узлов (ЛУ) шейной группы в процессе эхографии. Описывается комплексная методика ультразвукового (УЗ) исследования опухолевого процесса ЛУ. Предлагаются алгоритмы предварительной обработки ультразвуковых изображений (УЗИ), описываются структура специальной нейронной сети (НС) и алгоритм ее обучения по информации, полученной при обследовании.

Введение

Одной из основных задач в медицине является диагностика. Постановка диагноза часто является сложной задачей и зависит от большого количества факторов, в частности от имеющихся данных о симптомах и признаках заболевания, полноты и результатов лабораторных исследований, общего объема медицинской информации о предполагаемом заболевании, от опыта и квалификации врача-диагноста. Наличие субъективного фактора повышает вероятность ошибки диагностики заболевания.

В последние десятилетия благодаря развитию и применению современных методов диагностики, появлению новейшей диагностической аппаратуры, развитию информационных технологий существенно повысились возможности постановки правильного диагноза. Среди основных проблем, которые возникают при попытках автоматизации процесса диагностики, можно отметить выбор параметров, необходимых для постановки достоверного диагноза, и построение надежного алгоритма, устойчиво работающего в условиях значительной зашумленности и изменчивости входных данных.

Хорошие перспективы при постановке диагноза имеют нейросетевые технологии. Однако при традиционных способах использования НС возникают значительные трудности при необходимости учета априорных вероятностей появления того или иного заболевания и учета риска при постановке неправильного диагноза.

В настоящей работе предлагается новый метод использования НС в системах медицинской диагностики, заключающийся в применении ее в качестве обучаемого формирователя признаков с заданными статистическими характеристиками. Сформированные признаки использует байесовский классификатор, позволяющий принимать решение с учетом априорных вероятностей и степени риска постановки неправильного диагноза.

Решение задачи автоматизации процесса диагностики окажет существенную помощь врачам при диагностике заболеваний ЛУ и принятии решения о наиболее эффективных методах лечения и объемах дополнительного гистологического обследования лимфоидной ткани.

1. Комплексная методика диагностики заболеваний ЛУ по УЗИ

Основными методами, дающими право на формулировку диагноза «лимфома», являются морфологическое и иммунологическое исследования субстрата опухоли. Роль и место УЗ-метода в диагностике лимфом определяются первичной локализацией опухоли и закономерностями распространения опухолевого процесса. Не менее важны данные анализа линейных скоростных показателей кровотока, резистентности циркуляторного русла и индекса васкуляризации. В некоторых публикациях подтверждается перспектива в разграничении реактивной гиперплазии и лимфом при помощи УЗ-технологий к моменту постановки морфологического диагноза. Эти данные могут быть использованы как дополнительные критерии дифференциальной диагностики доброкачественных

лимфаденопатий (ЛАП) и злокачественных лимфопролиферативных форм детского рака ЛУ (лимфогранулематоза (ЛГМ) и неходжкинских лимфом (НХЛ)) на этапе первичной диагностики.

По данным исследователей из политехнического университета Гонг-Конга [1], с увеличением размеров ЛУ увеличивается внутриузловая скорость кровотока без существенного изменения резистентности циркуляторного русла ЛУ.

Во многих работах подтверждается возможность разделения доброкачественных гиперплазий и метастатически пораженных ЛУ при оценке внутриузловой резистентности: в метастатически пораженных ЛУ индекс резистентности (RI) имеет значения более 0,8 [2–5]. Другие авторы считают признаком злокачественного поражения ЛУ значение RI более 0,72 [6]. Известны попытки определения качественных дифференциально-диагностических признаков разграничения лимфом и реактивных гиперплазий (РГ) на основании полуколичественных и качественных показателей, таких как форма ЛУ, выделения области ворот, сосудистая картина [7–9]. Есть работы по применению метода контрастирования при изучении перфузии пораженных ЛУ [10] и метода энергетического доплеровского сканирования при оценке васкуляризации непораженных ЛУ [11]. Имеются единичные работы по изучению сосудистой картины пораженных ЛУ при лимфомах у детей [9].

Известно также, что определенные патологии характеризуются коэффициентом формы ЛУ и его средней плотностью. Данные параметры могут быть успешно найдены при обработке визуальных графических изображений. Принятие решения о необходимости дополнительного исследования лимфоидной ткани определяется данными как клинических, так и инструментальных методов исследования.

Проведенный анализ позволил выделить круг параметров, которые могут быть использованы при автоматизированной диагностике заболеваний ЛУ по УЗИ.

2. Общее описание технологии автоматизированной диагностики заболеваний ЛУ

Исходя из основных признаковых характеристик заболеваний и проведенного выше анализа разработан перечень диагностических параметров, которые можно получить с УЗИ, и методика расчета их количественных показателей. На основе комплексной методики УЗ-исследования опухолевого процесса ЛУ была разработана информационная технология автоматизированной диагностики нозологической формы ЛГМ, НХЛ и ЛАП и создан программный комплекс, реализующий данную технологию.

Программный комплекс работает с УЗИ, полученными в режимах полутоновой шкалы (GRAY), цветного доплеровского картирования (CFM) и энергетического доплеровского сканирования (PDI), в режиме импульсноволнового доплера (PWD) при измерении скоростных характеристик кровотока и индекса интранодальной резистентности (рис. 1).

При оценке характера изменений ЛУ учитываются следующие параметры: линейные размеры, форма, структура (изменения текстуры изображения) ткани, эхогенность, характер и индексы васкуляризации (при цветовом доплеровском и энергетическом картировании), амплитудно-временные характеристики кровотока и резистентность циркуляторного русла ЛУ по данным ИД.

Для каждого режима разработан перечень измеряемых параметров (табл. 1).

Таблица 1 Перечень измеряемых параметров на УЗИ

Название	Обозначение	Описание				
1	2	3				
Для режима GRAY						
Индекс неоднородности	Disp	Усредненная эхогенность, равная среднему значению плотности отраженного потока в пределах выделенной области				
Длина	L	Размер большой оси эллипса				
Ширина	S	Размер малой оси описанного эллипса				
Коэффициент формы ЛУ	Round Shape	Отношение размера большой оси к малой. Рассчитывается по формуле $RS = L/S$				
Площадь ЛУ	Sq	Площадь выделенной области				
Объем ЛУ	V	Объем тела вращения, полученного путем вращения эллипса вокруг большой оси. Рассчитывается по формуле $V = 4/3\pi LS^2$				

продолжение таол.							
1	2	3					
Для режима СFМ							
Относительная площадь кровотока	Color Pixel Density	Отношение общей площади кровотока (цветной области) к общей площади ЛУ					
Коэффициент разнона- правленности кровотока	KoefDD	Отношение прямого и обратного кровотоков					
Для режима PDI							
Относительная площадь кровотока	Color Pixel Density	Отношение общей площади кровотока (цветной области) к общей площади ЛУ					
Индекс васкуляризации	Power Wiegh Pixel	Энергетически взвешенная плотность кровотока, равная отношению общей энергии кровотока (цветной области) к общей площади сечения ЛУ					
Для режима PWD							
Максимальная скорость кровотока	Vmax	Максимальная скорость кровотока за один период					
Минимальная скорость кровотока	Vmin	Минимальная скорость кровотока за один период					
Средняя скорость кровотока	Vmean	Средняя скорость кровотока за один период					
Индекс резистентности	IR	Отношение разности максимальной и минимальной скоростей кровотока за один период к его максимальной скорости. Рассчитывается по формуле IR = (Vmax–Vmin)/Vmax					
		Отношение разности максимальной и минимальной скоростей					

Продолжение табл. 1

Программный комплекс содержит набор относительно самостоятельных модулей:

по формуле PI = (Vmax–Vmin)/Vmean

кровотока за один период к его средней скорости. Рассчитывается

- определения режима работы и настроек УЗ-аппарата;
- предварительной обработки;

Индекс

- анализа и отбора параметров;
- нейросетевой автоматизированной диагностики.

PΙ

3. Модуль определения режима работы и настроек УЗ-аппарата и модуль предварительной обработки УЗИ

Перед началом обработки УЗИ программный комплекс при помощи предварительно обученной НС автоматически распознает режим работы УЗ-аппарата, его настройки, масштабный коэффициент (рис. 2, a–a) и производит привязку к эталонной шкале соответствующего режима (рис. a, a-a) и контролируются врачом. В случае необходимости, например при ошибочном распознавании, параметры могут быть отредактированы в ручном режиме. Необходимость фиксации настроек УЗ-аппарата обусловлена значительным их влиянием на измеряемые параметры.

Модуль предварительной обработки производит первичную обработку УЗИ с целью фильтрации шумов и подчеркивания полезной составляющей, а в некоторых случаях нормализацию. Для предварительной обработки используются несколько алгоритмов, предназначенных для расчета различных параметров, при этом в качестве основного из них предлагается применять комплексный метод, сочетающий медианную фильтрацию и оператор скользящего среднего.

4. Модуль анализа и отбора параметров

Любой объект, подлежащий распознаванию и классификации, обладает рядом различительных признаков. Начальным этапом любого процесса распознавания является выбор различительных признаков. Количество признаков, необходимое для успешного решения задачи распознавания, зависит от разделяющих качеств выбранных признаков. При решении задачи формирования признакового пространства необходимо выделить из векторов измерений более существенные признаки.

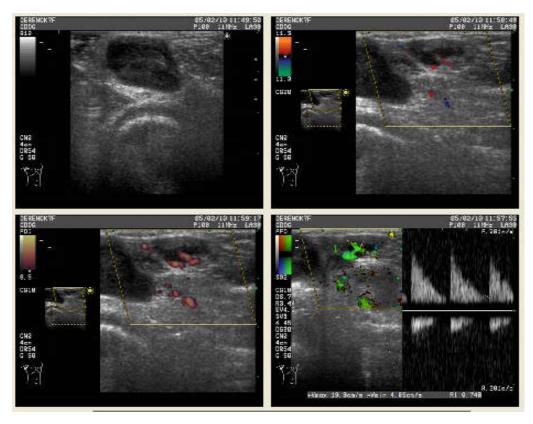


Рис. 1. Основные режимы работы УЗ-аппарата

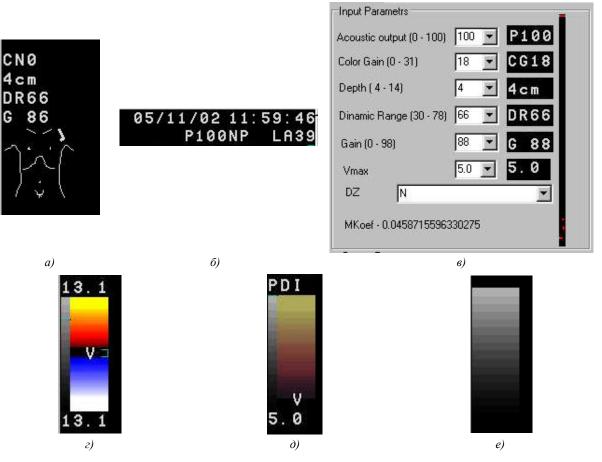


Рис. 2. Примеры работы модуля определения режима работы и настроек УЗ-аппарата

В результате работы программного комплекса оценки УЗИ опухолевого процесса ЛУ врач получает 14 различных параметров. Информативность параметров для оценки того или иного диагноза различна, поэтому при помощи аппарата математической статистики необходимо обработать данные и выявить наиболее характерные для каждого конкретного заболевания параметры, изменения которых с высокой вероятностью позволят утверждать о его наличии или отсутствии.

В качестве метода исследования использованы методы дисперсионного анализа [11–13]. Для каждого из исследуемых заболеваний набрана классифицированная обучающая выборка. Для всех 14 параметров заполняются таблицы, аналогичные табл. 2. В ячейках содержатся относительные значения выбранного параметра (одного из 14) для каждого из подтвержденных диагнозов рассматриваемых заболеваний.

Таблица 2 Значения измеренного параметра пациента с подтвержденным диагнозом заболевания

Количество пациентов	Значения рассматриваемого параметра						
	Заболевание	Заболевание		Заболевание М–1	Заболевание М		
1	X_{II}	X_{12}		X_{lm-1}	X_{lm}		
2	X_{21}	X_{22}		X_{2m-1}	X_{2m}		
3	X_{3I}	X_{32}		X_{3m-1}	X_{3m}		
N	X_{nl}	X_{n2}		X_{nm-1}	X_{nm}		
Среднее	X_{I}	X_2		X_{m-1}	X_m		

Строка «Среднее» содержит средние арифметические значения полученных наблюдений для каждого заболевания:

$$X_{j} = \frac{1}{n} \sum_{i=1}^{n} x_{ij} (j = \overline{1,m}),$$

где X_j – выборочное среднее j-го уровня; x_{ij} – i-е значение j-го уровня. Введем в рассмотрение следующие величины:

$$\bar{x} = \frac{1}{a} \sum_{i=1}^{n} \sum_{j=1}^{m} x_{ij};$$
 $S_A = a \sum_{j=1}^{m} (x_j - \bar{x})^2;$

$$S = \sum_{i=1}^{n} \sum_{j=1}^{m} (x_{ij} - \bar{x})^{2}; \qquad S_{0} = \sum_{i=1}^{n} \sum_{j=1}^{m} (x_{ij} - x_{j})^{2},$$

где $\overset{-}{x}$ — общее выборочное среднее анализируемого параметра по всем заболеваниям; a — общее число всех наблюдений; S — общая сумма квадратов отклонений от общего выборочного среднего; S_A — сумма квадратов отклонений групповых средних x_j от общего выборочного среднего $\overset{-}{x}$ (характеризует рассеяние между группами); S_0 — остаточная сумма квадратов отклонений от своего выборочного среднего (характеризует рассеяние внутри групп).

Очевидно, что

$$S = S_A + S_0$$
.

Сумма S характеризует связь выбранного для анализа параметра с конкретным заболеванием. Если заболевания имеют сильную зависимость с параметром, то групповые средние будут тем больше рассеяны вокруг общего выборочного среднего, чем больше эта зависимость. Отсюда следует, что для оценки связи заболевания с анализируемым параметром необходимо рассматривать взвешенную сумму квадратов отклонений групповых средних от общего среднего, т. е. S_A . Влияние случайных причин определяется значением S_0 .

Для выяснения степени связи заболевания с анализируемым параметром рассматривалась нулевая гипотеза, согласно которой заболевание не связано с изменениями анализируемого параметра. В этом случае статистики рассчитываются по формулам

$$C^2 = \frac{S}{nm-1};$$
 $C_A^2 = \frac{S_A}{m-1};$ $C_0^2 = \frac{S_0}{m(n-1)}.$

Общая, факторная и остаточная дисперсии являются несмещенными оценками генеральной дисперсии анализируемого параметра.

Для выяснения связи каждого заболевания с анализируемым параметром сравнивались дисперсии C_A^2 и C_0^2 . Связь заболевания считается значимой на выбранном уровне значимости α , если выполняется неравенство

$$C_A^2/C_0^2 > F_{1-\alpha;\nu_1;\nu_2}$$
.

Граница критической области $F_{1-\alpha;\nu_1;\nu_2}$ находится по известным таблицам F-распределений для заданного уровня значимости α с $\nu_1=m-1, \nu_2=m(n-1)$ степенями свободы. Если неравенство $C_A^2/C_0^2>F_{1-\alpha;\nu_1;\nu_2}$ выполняется, то существует связь заболевания с анализируемым параметром. После проведения аналогичных вычислений для всех параметров может быть определен набор основных параметров для обучения нейронной сети, предназначенной для диагностики выбранного заболевания.

5. Нейросетевой модуль постановки диагноза

При традиционных способах использования HC в задачах распознавания медицинских объектов возникают значительные трудности, связанные с необходимостью учета априорных вероятностей появления того или иного заболевания и учета риска при постановке неправильного диагноза.

Предлагается новый метод использования НС в системах медицинской диагностики в качестве обучаемого формирователя признаков с заданными статистическими характеристиками. Сформированные признаки использует байесовский классификатор, позволяющий принимать решение с учетом априорных вероятностей и степени риска постановки неправильного диагноза. При таком подходе задача постановки диагноза делится на две относительно самостоятельные подзадачи: первая — формирование вектора признаков с желаемым видом и параметрами законов распределения вероятности отдельных признаков, вторая — принятие решения о классе предъявленного вектора на основе хорошо разработанных методов теории статистических решений с учетом априорных вероятностей появления объектов каждого класса и матрицы риска [14]. Процесс распознавания с учетом такого подхода показан на рис. 3.

Для решения первой задачи используется нейросетевой формирователь признаков (НСФП), который содержит три слоя однотипных нейронов. Число нейронов в выходном слое определяется размером необходимого признакового пространства. Признаковое пространство выбирается из соображений минимизации размерности, с одной стороны, и обеспечения требуемых обобщающих свойств НС, с другой. Все нейроны имеют сигмоидную активационную функцию.

В качестве желаемого закона распределения вероятности признаков выбран нормальный, так как при этом вместо громоздкого и трудоемкого процесса формирования оценок условных

многомерных плотностей вероятности достаточно вычислить вектор выборочных средних и выборочную ковариационную матрицу. На основе полученного вектора признаков проводится оценка вероятности принадлежности предъявляемого образа к некоторому известному классу по известной формуле

$$p(A_k / Y) = \frac{1}{(2\pi)^{\frac{m}{2}} \sqrt{|\Sigma_k|}} \exp(-\frac{1}{2} (M_k - Y)^T \Sigma_k^{-1} (M_k - Y)), \qquad (1)$$

где Y — входной вектор признаков; m — размерность вектора признаков; M_k — вектор математических ожиданий k-го класса; \sum_k — ковариационная матрица k-го класса; k — индекс класса.

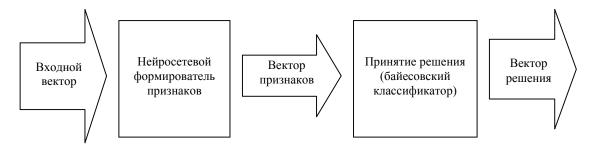


Рис. 3. Структурная схема системы распознавания образов с нейросетевым формирователем признаков

Наиболее часто в качестве методов синтеза нейронных сетей используются алгоритмы обучения на примерах, построенные на основе итерационных процедур градиентного поиска в пространстве весовых коэффициентов и параметров активационной функции нейрона. Однако эти методы обладают рядом существенным недостатков, основным из которых является возможность попадания в локальный экстремум целевой функции. Для преодоления этого недостатка предлагается применять комплексный подход, основанный на совместном использовании генетического алгоритма [14-17] и алгоритма обратного распространения ошибки [18]. На каждом шаге генетического алгоритма текущая популяция подвергается воздействию генетических операторов кроссинговера, мутации и инверсии. Оператор кроссинговера обеспечивает грубый поиск областей решений с высоким уровнем целевой функции, производя комбинаторные обмены частей решений (хромосом). Оператор инверсии производит сбор полезных комбинаций значений генов в компактные области хромосомы, т. е. формирует так называемые строительные блоки, снижая тем самым вероятность их разрыва при кроссинговере. Оператор мутации осуществляет точный поиск путем мелких шагов решений в случайном направлении. После формирования новой популяции для каждого решения включается алгоритм обратного распространения ошибки для поиска локальных оптимумов в каждой подобласти. Сформированные таким образом решения оцениваются в соответствии с целевой функцией, и из них при помощи оператора репродукции производится выбор наиболее пригодных решений для продолжения поиска. Это значительно увеличивает скорость сходимости алгоритма к глобальному экстремуму.

Целевой функцией настройки параметров НСФП является комплексный функционал

$$J = W_1 J_1 + W_2 J_2 + W_3 J_3, (2)$$

где W_1, W_2, W_3 – весовые коэффициенты предпочтения критериев; J_1 – функционал качества вектора признаков; J_2 – функционал степени отклонения от нормального закона распределения; J_3 – степень вырожденности ковариационной матрицы.

Полученный после НСФП вектор признаков используется байесовским классификатором для принятия окончательного решения о классе предъявленного образа.

Заключение

В статье рассмотрены возможности автоматизированной обработки УЗИ ЛУ при реактивных гиперплазиях и лимфомах у детей. Представленная технология разработана в рамках

выполнения проекта МНТЦ В-736 и основана на компьютерном анализе УЗИ в режимах полутоновой шкалы, цветного и энергетического доплеровского картирования, а также оценке скоростных характеристик кровотока и индекса интранодальной резистентности в режиме импульсноволнового доплеровского измерения.

На базе Республиканского научно-практического центра детской онкологии и гематологии была проведена клиническая апробация представленной информационной технологии при обследовании 75 детей обоего пола в возрасте от 2 до 17 лет с диагнозами «неходжкинская лимфома», «лимфогранулематоз» и «реактивная гиперплазия ЛУ шейной группы». При постановке диагноза учитывались линейные размеры ЛУ, скоростные показатели внутриузлового кровотока, индекс резистентности магистральной артерии ЛУ, сосудистая картина ЛУ.

Полученные при клинической апробации данные анализа линейных скоростных показателей кровотока, резистентности циркуляторного русла и индекса васкуляризации подтвердили перспективу в разграничении реактивной гиперплазии и лимфом при помощи доплеровских технологий к моменту постановки морфологического диагноза. Эти данные могут быть использованы как дополнительные критерии дифференциальной диагностики доброкачественных лимфаденопатий и злокачественных лимфопролиферативных заболеваний у детей на этапе первичной диагностики.

Список литературы

- 1. Vascularity and grey-scale sonographic features of normal cervical lymph nodes: variations with nodal size / M. Ying [et al.] // Clinical Radiology. 2001. Vol. 56, № 5. P. 416–419.
- 2. Differential diagnosis of cervical lymphadenopathy: usefulness of color Doppler sonography / D.G. Na [et al.] // American journal of roentgenology. -1997. Vol. 168, $N_{2} 5. P. 1311-1316$.
- 3. Usefulness of Doppler spectral analysis and power Doppler sonography in the differentiation of cervical lymphadenopathies / C.H. Wu [et al.] // American journal of roentgenology. 1998. Vol. 171, N₂ 2. P. 503–509.
- 4. Brnic, Z. Usefulness of Doppler waveform analysis in differential diagnosis of cervical lymphadenopathy / Z. Brnic, A. Hebrang // European Radiology. -2003. Vol. 13, N0 1. P. 175–180.
- 5. Power Doppler sonography to differentiate tuberculous cervical lymphadenopathy from nasopharyngeal carcinoma / A. Ahuja [et al.] // American journal of neuroradiology. -2001. Vol. 22, N $\underline{0}$ 4. P. 735–740.
- 6. Doppler sonographic differential diagnosis of superficial lymphadenopathy: metastasis, malignant lymphoma, and benign process / T. Shirakawa [et al.] // J. Ultrasound Med. -2001. Vol. 20, N_{2} 5. P. 525–532.
- 7. Color Doppler sonography in the evaluation of superficial lymphomatous lymph nodes / F. Giovagnorio [et al.] // J. Ultrasound Med. − 2002. − Vol. 21, № 4. − P. 403–408.
- 8. Петрович, С.В. Эпидемиология злокачественных новообразований у детей / С.В. Петрович, О.В. Алейникова. Минск: Беларуская навука, 2004.
- 9. High-resolution and color Doppler ultrasonography of cervical lymphadenopathy in children / O. Papakonstantinou [et al.] // Acta Radio. 2001. Vol. 42(5). P. 470–476.
- 10. Evaluation of lymph node perfusion using continuous mode harmonic ultrasonography with a second-generation contrast agent / L. Rubaltelli [et al.] // J. Ultrasound Med. -2004. Vol. 23, \mathbb{N}_{2} 6. P. 829–836.
- 11. Мацкевич, И.П. Высшая математика. Теория вероятностей и математическая статистика / И.П. Мацкевич, Г.П. Свирид. Минск: Вышэйшая школа, 1993. 269 с.
- 12. Четыркин, Е.А. Вероятность и статистика / Е.А.Четыркин, И.Л. Калихман. М.: Финансы и статистика, 1982. 319 с.
- 13. Митропольский, А.К. Техника статистических вычислений / А.К. Митропольский. М.: Наука, 1971.-576 с.
- 14. Фукунага, К. Введение в статистическую теорию распознавания образов / К. Фукунага. М.: Наука, 1979. 368 с.

- 15. De Jong, K. Evolving Complex Structures via Cooperative Coevolution / K. De Jong, M. Potter // Proc. Fourth Annual Conference on Evolutionary Programming. San Diego, CA, 1995. P. 307–317.
- 16. Holland, J.H. Escaping brittleness: the possibilities of general purpose learning algorithms applied to parallel rule-based systems / J.H. Holland. Los Altos, CA: Machine Learning, 1986. Vol. 2. P. 593–623.
- 17. Чернухин, Ю.В. Особенности использования генетических алгоритмов при обучении перцептронов / Ю.В. Чернухин, М.А. Беляев // Известия ТРТУ. 2001. № 4. С. 49–53.
- 18. Горбань, А.Н. Нейронные сети на персональном компьютере / А.Н. Горбань, Д.А. Россиев. Новосибирск: Наука, 1996. 276 с.

Поступила 27.04.06

¹Объединенный институт проблем информатики НАН Беларуси, Минск, Сурганова, 6 e-mail: anishch@newman.bas-net.by

²Республиканский научно-практический центр детской онкологии и гематологии, пос. Лесное-2, Минская обл. e-mai: begun igor@mail.ru

³Военная академия Республики Беларусь, Минск-57 e-mail: kulagaV@mail.ru

U.V. Anishchanka, I.V. Begun, V.V. Kulaga, S.V. Spiridonov, K.V. Chizhikov, D.L. Mozheyko

COMPUTER-AIDED DIAGNOSTICS TECHNOLOGY OF LYMPH NODES DISEASES BY ULTRASONIC IMAGES

A computer-aided diagnostics technology of diseases of cervical group lymph nodes during echography is considered. The ultrasonic investigation technique of neoplastic tumor process in the lymph nodes is described. Algorithms of ultrasonic image preprocessing, special neural network structure and the algorithm of its training are presented.