УДК 608

ИНФОРМАЦИОННО-АНАЛИТИЧЕСКАЯ МОДЕЛЬ ПОДДЕРЖКИ ПРИНЯТИЯ РЕШЕНИЙ В ПРОЦЕССЕ ДИАГНОСТИКИ РАКА ЖЕЛУДКА

Артеменко М.В., Серебровский В.В., Бабков А.С.

ФГБОУ ВПО «Юго-Западный государственный университет», Курск, e-mail: maloi@swsu.org

Представлена информационно-аналитическая модель синтеза и применения диагностических решающих правил онкологических заболеваний желудка в процессе скрининга на основе анализа частных и интегральных показателей крови (включая онкомаркеры). Рассматривается новая методика анализа уверенности в принятии адекватного диагностического решения на основе анализа функциональных изменений в основном и альтернативном классах заболеваний (не болеющих рассматриваемой онкологией и болеющих, соответственно), отражаемых структурами и параметрами идентифицированных с помощью самоорганизационных алгоритмов методом группового учета аргументов математических моделей взаимосвязей между регистрируемыми значениями показателей, характеризующих системное состояние организма в процессе возникновения анализируемого заболевания. Отличительной особенностью информационно-аналитической системы является возможность осуществления ранней диагностики онкологических заболеваний желудка на основе совместного использования объективной информации по результатам системных изменений в организме, регистрируемых по изменениям состава крови, и субъективной информации, полученной в процессе обследования пациента.

Ключевые слова: онкологические заболевания желудка, скрининг, самоорганизационные алгоритмы моделирования, уверенность принятия решений

INFORMATION AND ANALYTICAL MODEL OF DECISION SUPPORT IN THE STOMACH CANCER DIAGNOSIS

Artemenko M.V., Serebrovskiy V.V., Babkov A.S.

Southwest State University, Kursk, e-mail: maloi@swsu.org

The structure of and considered the information-analytical model of synthesis and application of diagnostic decision rules gastric cancer in the screening process based on the analysis of private and integral parameters of blood (including tumor markers). Author's method of analysis is presented of confidence in accepting adequate diagnostic decisions based on analysis of the differences and functional changes in the classroom, «not with cancer of the stomach» and «ill people». Training samples are confirmed clinically. It is proposed to carry out the calculation of the coefficient of confidence in making decisions based on the following indicators of confidence: confidence in the availability of necessary reprezantivnosti training sample, confidence in the adequacy of identified through self-organization simulation algorithms – the group method of data handling; confidence in the correctness of classification, the estimated values of the errors on the first and second kind. The proposed information-analytical system for early diagnosis of cancer of the stomach during the screening process differs sharing of objective information about the patient's condition at the stage of pre-clinical examination of the patient in terms of screening of blood analysis and subjective information obtained during the examination of the patient.

Keywords: gastric cancer, screening, self-organization algorithms for modeling, decision-making confidence

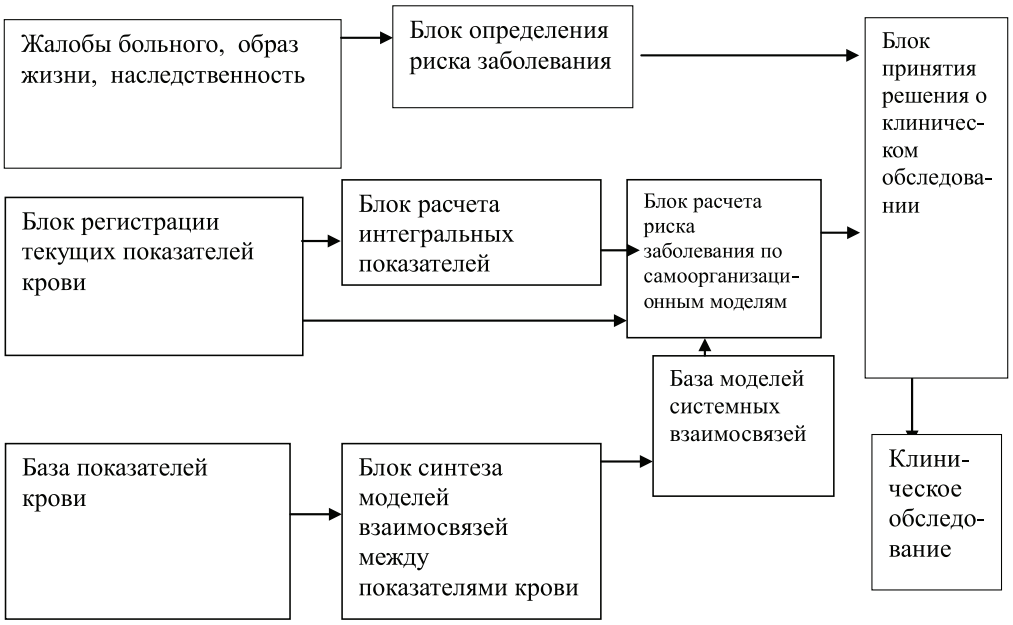
В настоящее время актуальна проблема диагностики онкологических заболеваний желудка на ранней стадии. В странах с большим количеством населения территориально отдаленного последних от специализированных центров своевременной диагностики и лечения, диагностика осуществляется, как правило, в процессе скрининга или средств телемедицины [1]. В основе предлагаемого диагностического подхода ранней диагностики рака желудка предлагается применять концепцию рассмотрения крови как соединительной ткани, системно присутствующей во всех участках внутренних и внешних физиологических систем и органов и несущей в значениях регистрируемых характеристик информацию об изменениях в организме [2].

В процессе скрининг-диагностики или общей диспансеризации осуществляется лабораторный анализ крови. Определим их как частные, первично регистрируемые, принадлежащие некоторому множеству {X}. Для решения задач доклинической диагностики на основе анализа указанных показателей в процессе скрининга предлагается применять следующую информационно-аналитическую модель (рисунок).

На принятие управленческого решения о необходимости проведения клинического обследования на предмет выявления рака желудка оказывают влияние различные факторы: образ жизни (прежде всего характер питания и стрессовость), жалобы пациента и другие показатели функционирования определенных физиологических систем. Эти показатели группируются в блоке «Жалобы больного, образ жизни, наследственность».

Основной группой анализируемых показателей доклинической диагностики возможности наличия онкологических заболеваний желудка являются значения показателей, полученных в процессе анализа крови (включая значения показателей

онкомаркеров) – в информационно-аналитической модели представлены блоком «База показателей крови». В этом блоке весь информационный архив показателей лабораторного анализа крови состоит из двух групп. В первую – включены показатели крови людей с клинически подтвержденным диагнозом рака желудка. Вторую группу составляют значения тех же показателей у людей, не больных раком желудка. Таким образом, в базе показателей крови представлены показатели для двух альтернативных классов. Под альтернативным классом будем понимать не больных онкологическими заболеваниями людей – класс w0, больных раком желудка – класс w1.



Информационно-аналитическая модель доклинической диагностики рака желудка

Результаты лабораторного анализа крови людей, по которым решается задача ранней диагностики, представлены «Блоком регистрации текущих показателей крови».

Для построения классификационных правил формируются обучающие выборки, по которым синтезируются правила принятия решений с использованием самоорганизационных алгоритмов, основанных на математических моделях, идентифицированных с помощью программных средств метода группового учета аргументов (МГУА) [3, 4].

Процедура обучения реализуется в «Блоке синтеза моделей взаимосвязей между показателями крови». В качестве исходных данных для этого блока используются как традиционные показатели анализа состава крови (вектор $X = x_1, \ldots, x_n$), так и комплекс интегральных показателей (вектор $Y = y_1, \ldots, y_m$). Значения интегральных показателей определяются по формуле вида:

$$Y_k = \frac{\sum_{i=1}^{n}\left(\alpha_{i,k} \cdot \left(\frac{x_i - \overline{Xw0_i}}{\sigma w0_i}\right)^2\right)}{\sum_{i=1}^{n}\alpha_{i,k}}, \quad (1),$$

где k – индекс интегрального показателя; Y_k – интегральный показатель; x_i – значение i-го частного показателя крови у пациента, $\overline{Xw0_i}$, $\sigma w0_i$ – соответственно, модальное значение (в случае нормального распределения совпадает с математическим ожиданием и средними значениями) и средне-квадратичное отклонение (СКО) i-го частного показателя крови, полученное на обучающей выборке для класса «здоровые», $\alpha_{i,k} \in [0,1]$ – весовые коэффициенты, определяющие информационный вклад показателя x_i в формирование интегрального показателя Y_k.

В предлагаемой информационно-аналитической модели предусмотрено несколько вариантов определения указанных коэффициентов: экспертное; заключение, на основе дисперсионного анализа; использование информационного критерия Кульбака.

Интегральные показатели используются для решения задач ранней диагностики рака желудка и вычисляются для анализируемого пациента в «Блоке расчета интегральных показателей».

Результатом работы «Блока синтеза моделей взаимосвязей между показателями

крови» является набор математических моделей, характеризующих множественные связи между векторами признаков $\overline{X}$ и интегральных показателей $\overline{Y}$ для основного (рак желудка) и альтернативного классов.

В целях оптимизации применения моделей формируется «База моделей системных взаимосвязей».

В «Блоке расчета риска заболевания по самоорганизационным моделям» основного класса заболеваний w_1, используя вектора $\overline{X}$ и $\overline{Y}$, определяются модифицированные вектора $\overline{X}_1^*$ и $\overline{Y}_1^*$ как значения аппроксимантов идентифицированных моделей. Аналогично по моделям альтернативного класса w_0, используя те же вектора $\overline{X}$ и $\overline{Y}$, определяются вектора $\overline{X}_0^*$ и $\overline{Y}_0^*$. По полученным векторам определяются меры близостей – вектора: $\overline{D_1}$ и $\overline{D_0}$ соответственно между векторами $\left\{\overline{X}, \overline{Y}\right\}$, и $\left\{\overline{X}_1^*, \overline{Y}_1^*\right\}$ и $\left\{\overline{X}_0^*, \overline{Y}_0^*\right\}$.

Полученные значения мер близости являются базовыми переменными для решения задачи расчета рисков соотнесения состояния пациента к различным альтернативным классам Rk_{w0} и Rk_{w1} в блоке расчета риска заболевания по самоорганизационным моделям.

С целью улучшения качества классификации кроме показателей крови используются значения признаков, формируемых в блоке «Жалобы больного, образ жизни, наследственность». По этой группе признаков в блоке оценки риска заболевания рассчитываются показатели Rt_{w0} и Rt_{w1}, характеризующие степени риска возникновения рака желудка (или не возникновения соответственно), значения которого выбираются из «Блока оценки риска заболевания».

Окончательное решение о проведении клинического обследования осуществляется путем агрегации указанных показателей риска в «Блоке принятия решения о клиническом обследовании».

Учитывая факт наличия малых выборок при решении рассматриваемой диагностической задачи, предлагается использовать самоорганизационный алгоритм структурно-параметрической идентификации МГУА, синтезирующий модели вида (2):

$$z_i = A_0 + \sum_{i=1}^{n}\left(A_i \cdot \prod_{j=1, j\neq i}^{m} z_j^{q_j}\right), \qquad (2)$$

где A_i – весовой коэффициент; z_j – переменная (из множеств $\{X\},\{Y\}$); q_j – степень аргумента; n – количество рядов селекции (термов полинома); m – количество переменных.

Затем в каждом из альтернативных классов формируются наборы адекватных моделей, не пересекающихся друг с другом по параметрам или структурам.

Таким образом, согласно предлагаемой информационно-аналитической модели доклинической диагностики рака желудка предлагается следующая методика классификации:

1. Формируется репрезентативная выборка исходных данных (частные и интегральные показатели крови – множества $\{X\}$ и $\{Y\}$ соответственно), на основе которой после селекции артефактной информации осуществляется формирование из нее обучающей и экзаменационной подвыборок.

2. Для каждого диагностируемого альтернативного класса формируются пакеты математических моделей, отражающих связи между показателями соответственно:

$$F1_{w0}(\{X\}):\ \hat{X}_i = F1w0_i\left(\{X_l\},\ Kcm_i^{w0} = R_l^2 / l \neq i,\ l = 1, n1\right); \qquad (3)$$

$$F2_{w0}(Y\}):\ \hat{Y}_k = F2w0_k\left(\{Y_l\},\quad Kcm_k^{wo} = R_l^2 / l \neq k,\ k = 1, n2\right); \qquad (4)$$

$$F3_{w0}(\{Y \leftarrow X\}):\ \widehat{Y}_k = F3w0_k\left(\{X_l\},\quad Kcm_k^{w0} = R_l^2 /\ l = 1, n1\right):\ k - 1, n2); \qquad (5)$$

$$F1_{w1}(\{X\}):\ \hat{X}_i = F1w1_i\left(\{X_l\},\ Kcm_l^{w1} = R_l^2 / l \neq i,\ l = 1, n1\right); \qquad (6)$$

$$F2_{w1}(Y\}):\ \hat{Y}_k = F2w1_k\left(\{Y_l\},\quad Kcm_k^{w1} = R_l^2 / l \neq k,\ k = 1, n2\right); \qquad (7)$$

$$F3_{w1}(\{Y \leftarrow X\}):\ \widehat{Y}_k = F3w1_k\left(\{X_l\},\ Kcm_k^{w1} = R_l^2 /\ l = 1, n1:\ k - 1, n2\right), \qquad (8)$$

где $n1$ – количество частных показателей; $n2$ – количество интегральных показателей; $F1$ – аппроксиманты, построенные на множестве $\{X\}$; $F2$ – аппроксиманты, построенные на множестве $\{Y\}$; $F3$ – аппроксиманты, построенные на множествах $\{X\}$ и $\{Y\}$, Kcm – коэффициенты уверенности в адекватности применения аппроксимантов в качестве моделей (в соответствующих альтернативных классах $w0$, $w1$); R^2 – коэффициенты детерминации для соответствующих моделей-аппроксимантов.

В пакеты моделей включаются только те, у которых значения коэффициентов детерминации превышает определенный исследователем пороговый уровень.

Уверенность в адекватности применения сформированных пакетов математических моделей в альтернативных классах определяются по формулам (9) и (10) соответственно.

$$\text{КУм}_{w0} = 1 - \prod_{k=1}^{Kw0} \left(1 - Kcm_k^{w0}\right); \qquad (9)$$

$$\text{КУм}_{w1} = 1 - \prod_{k=1}^{Kw1} \left(1 - Kcm_k^{w1}\right), \qquad (10)$$

где $Kw0 = n1$, $Kw1 = n2$.

Значение коэффициента уверенности одновременного применения обоих пакетов моделей вычисляется по формуле (11).

$$\text{КУм}_{w0\&w1} = \min\left(\text{КУм}_{w0}, \text{КУм}_{w1}\right). \qquad (11)$$

3. На экзаменационной подвыборке определяется уверенность квалификационных возможностей для каждого альтернативного класса КУкв$w0$ и КУкв$w1$ как $(1 - \beta_{w0})$ и $(1 - \beta_{w1})$, где β_{w0}, β_{w1} – ошибки второго рода классификации.

В случае недостаточности объемов экзаменационных подвыборок в каждом из альтернативных классов осуществляется экспертный анализ для построения функции принадлежности на базовой переменной Zb, вычисляемой по формуле (12). Затем значение функции принадлежности используется для вычисления соответствующих значений коэффициентов уверенности:

$$Zb_j = \begin{cases} g1\left(\left(\dfrac{(\widehat{X}_i - X_i^*)}{X_i^*}\right)^2\right), & \text{для} \quad \forall \quad F1_i(X); \\ g2\left(\left(\dfrac{(\widehat{Y}_i - Y_i^*)}{Y_i^*}\right)^2\right), & \text{для} \quad \forall \quad (F2_i(Y) \vee F3_k\{(Y \leftarrow X\})), \end{cases} \qquad (12)$$

где $g1$, $g2$ – некоторые функции, определяющие мерность носителя; X_i^*, Y_i^* – значения частных и интегральных показателей для диагностируемого пациента.

В этом случае функция принадлежности рассматривается в виде формулы:

$$\mu(Zb_j) = 1 - Zb_j. \qquad (13)$$

4. Процесс диагностики возможного заболевания раком желудка осуществляется следующим образом.

В ходе осуществления анализа крови формируется экспериментальный вектор (включая значения онкомаркеров) $\overline{X_э}$ (размер вектора $m1$) и характеристики образа жизни – пациента – вектор $\overline{\text{Life}}$. По последнему в блоке определения риска заболевания (рисунок) рассчитывается значение показателя риска Rt (по табличным данным, заранее определенным экспертами с учетом личного опыта и/или рекомендаций научно-практических исследований по данным публикаций). По вектору $\overline{X_э}$ рассчитывается вектор экспериментальных интегральных показателей $\overline{Y_э}$ (размерность вектора m2).

Полученные вектора $\overline{X_э}$ и $\overline{Y_э}$ поступают в «Блок расчета риска заболеваний по самоорганизационным моделям», в котором определяются коэффициенты уверенности КУм_{w0}, КУм_{w1}, КУкв_{w0}, КУкв_{w1}, КУв. По данным коэффициентам уверенности определяются обобщенные классификационные коэффициенты уверенности соотнесения к альтернативным классам КУ_{w0} и КУ_{w1} – формулы (14) и (15), соответственно.

$$\text{КУ}_{w0} = \text{КУв} \cdot \min(\text{КУм}_{w0}, \text{КУкв}_{w0}); \qquad (14)$$

$$\text{КУ}_{w1} = \text{КУв} \cdot \min(\text{КУм}_{w1}, \text{КУкв}_{w1}). \qquad (15)$$

Далее, используя формулы (3)–(8), вычисляются для каждого из альтернативных классов множества: $\left\{\overline{X_{эw0}}, \overline{Y_{эw0}}\right\}$ и $\left\{\overline{X_{эw1}}, \overline{Y_{эw1}}\right\}$ соответственно. Затем определяются меры близости по формулам (16) и (17) соответственно.

$$D_{0j} = \begin{cases} \left|\dfrac{Xэw0_i - Xэ_i}{Xэ_i}\right| & \text{для} \quad \{\overline{X_{эw0}}\}, \quad i = 1, m1 \\ \left|\dfrac{Yэw0_i - Yэ_i}{Yэ_i}\right| & \text{для} \quad \{\overline{Y_{эw0}}\}, \quad i = 1, m2 \end{cases} \quad j = 1, m1 + m2; \qquad (16)$$

$$D_{1j} = \begin{cases} \left| \dfrac{Xэw1_i - Xэ_i}{Xэ_i} \right| \text{ для } \{\overline{X_{эw1}}\}, & i = 1, m1 \\ \left| \dfrac{Yэw1_i - Yэ_i}{Yэ_i} \right| \text{ для } \{\overline{Y_{эw1}}\}, & i = 1, m2 \end{cases}, j = 1, m1 + m2. \quad (17)$$

По полученным значениям мер и предварительно заданным экспертным путем (клиницистами) функциям принадлежностей к альтернативным классам w_0 и w_1 ($\left\{\mu_{w0_j}(D_{0_j})\right\}_{m1+m2}$ и $\left\{\mu_{w1_j}(D_{1_j})\right\}_{m1+m2}$ соответственно), определяются показатели уверенности принадлежности к альтернативным классам по мерам близости КУмбw_0 и КУмбw_1 по формулам (18) и (19) соответственно.

$$КУмбw_0 = \frac{\sum_{j=1}^{m1+m2} \mu_{w0_j}\left(D_{0_j}\right)}{m1+m2}; \quad (18)$$

$$КУмбw_0 = \frac{\sum_{j=1}^{m1+m2} \mu_{w0_j}\left(D_{0_j}\right)}{m1+m2}. \quad (19)$$

Примечание. Функции принадлежности рекомендуется строить по формуле (20). (Эксперт определяет численные значения a и b – рекомендуется: $a = 0{,}1 - 0{,}15$, $b > 0{,}5$)

$$\mu(d) = \begin{cases} 1, \text{ если } d \le a \\ \dfrac{b-d}{b-a} 1, \text{ если } b > d > a. \\ 0, \text{ если } d > b \end{cases} \quad (20)$$

По полученным значениям показателей уверенности определяются показатели риска Rk_{w0} и Rk_{w1} по формулам (21) и (22) соответственно:

$$Rk_{w0} = КУмб\, w_0 \times КУ_{w0}; \quad (21)$$

$$Rk_{w1} = КУмб\, w_1 \times КУ_{w1}. \quad (22)$$

Для принятия решения о принадлежности пациента к одному из альтернативных классов в блоке принятия решения о клиническом обследовании осуществляется расчет коэффициентов уверенности в диагностике $КУд_{w0}$ и $КУд_{w1}$ по формулам (23) и (24) соответственно.

$$КУд_{w0} = Rk_{w0} \cdot Rt_{w1}; \quad (23)$$

$$КУд_{w1} = Rk_{w1} \cdot Rt_{w1}. \quad (24)$$

Для каждого из показателей уверенности задаются пороговые отношения P_{w0} и P_{w1} и составляется таблица. Если значение соответствующего показателя не превышает порогового уровня, то значение элемента таблицы равно «0», в противном случае – «1». Каждому соотношению соответствует определенная ситуация.

Диагностические ситуации (р – пороговые значения – определяются экспертно)

$КУд_{w0} > P_{w0}$ $КУд_{w1} > P_{w1}$	$КУд_{w0} < P_{w0}$ $КУд_{w1} > P_{w1}$	$КУд_{w0} < P_{w0}$ $КУд_{w1} < P_{w1}$	$КУд_{w0} > P_{w0}$ $КУд_{w1} < P_{w1}$	Идентификатор (№) ситуации	Описание ситуации
0	0	0	0	1	Дополнительное обследование
0	0	0	1	2	Здоров – класс w_0
0	0	1	0	1	Дополнительное обследование
0	1	0	0	3	Болен – класс w_1
1	0	0	0	1	Дополнительное обследование

Таким образом, выделяются 3 ситуации, возникающие в процессе применения предлагаемой информационно-аналитической модели: 1 – пациент нуждается в дополнительном обследовании; 2 – вероятность присутствия рака желудка низка (пациент здоров); 3 – высока вероятность присутствия заболевания рака желудка.

Заметим, что для решения классификационных задач возможного развития заболевания в процессе мониторинга изменений определенных физиологических сигналов,

представленных временными рядами и отражающих системные изменения в организме, возможно применение искусственных иммунных сетей (по методике, например, рассмотренной в работе [5]).

Таким образом, предлагаемый метод доклинической диагностики рака желудка, основываясь на самоорганизационных алгоритмах идентификации математических моделей, отличается анализом межсистемных связей между различными показателями крови, несущей информацию о происходящих в организме системных изменениях в процессе возникновения онкологических заболеваний, в первую очередь внутренних органов – в частности рака желудка.

Список литературы

1. Рак желудка: симптомы, диагностика, лечение, профилактика [Электронный ресурс]. – Режим доступа: http://natural-medicine.ru/4127-rak-zheludka.html (дата обращения 18.02.2014).

2. Шебшаевич Л.Г., Алексеев А.А. Жизнь – кибернетическая медико-биологическая системность («Геном человека», клонирование – критический анализ). – М.: Триада Плюс, 2001. – 608 с.

3. Артеменко М.В., Бабков А.С. Классификация методов прогнозирования поведения систем // Современные проблемы науки и образования. – 2013. – № 6. URL: www.science-education.ru/113-11527 (дата обращения: 18.02.2014).

4. Ивахненко А.Г. Долгосрочное прогнозирование и управление сложными системами. – М.: Техника, 1975. – 313 с.

5. Оболенский А.Н., Кореневский Н.А., Артеменко М.В. Интегральное применение иммунных сетей и корреляционного анализа для выявления патологий в электрокардиосигнале // Известия Юго-Западного государственного университета. Серия: Управление, вычислительная техника, информатика. – 2012. – № 2, Ч.3. – С. 351–356.

References

1. Stomach cancer: symptoms, diagnosis, treatment, prevention [electronic resource]. Available at: http://natural-medicine.ru/4127-rak-zheludka.html (accessed 18 February 2014).

2. Shebshaevich L.G., Alekseev A.A. Life – cybernetic biomedical systematic («Human Genome» cloning – a critical analysis). M.: Triad Plus, 2001. pp. 608.

3. Artyomenko M.V., Money A.S. Classification of methods for predicting the behavior of systems // Modern problems of science and education. 2013. no. 6. URL: www.science-education.ru/113-11527 (accessed 18 February 2014).

4. Ivakhnenko A.G. Long-term forecasting and management of complex systems. M.: Technology, 1975. pp. 313.

5. Obolensky A.N., Korenevskii N.A., Artyomenko M.V. Integral use of immune networks and correlation analysis to identify pathologies in electrocardiosignals // Proceedings of the Southwestern State University. Series: Management, Computer Engineering, Computer Science. 2012. Vol. 3, no. 2. pp. 351–356.

Рецензенты:

Ключиков И.А., д.т.н., профессор, ГНС НИЦ, ФГУП «18 ЦНИИ» МО РФ, г. Курск;

Иванов В.А., д.м.н., профессор, зав. кафедрой медицины и логопедии, КГУ, г. Курск.

Работа поступила в редакцию 04.04.2014.