

УДК 57+615.47]: 616-074/.-006/.33 + 004.891.3

ПРЕВЕНТИВНАЯ ДИАГНОСТИКА ОНКОЛОГИЧЕСКИХ ЗАБОЛЕВАНИЙ ЖЕЛУДКА ПРИ СКРИНИНГЕ С ПОМОЩЬЮ СИСТЕМЫ ПОДДЕРЖКИ ПРИНЯТИЯ КЛАССИФИКАЦИОННЫХ РЕШЕНИЙ

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

ФГБОУ ВО Юго-западный государственный университет, Курск, email: artem1962@mail.ru

В статье рассматриваются вопросы формирования базы знаний системы поддержки принятия диагностических решений по ранней диагностике рака желудка на основе моделей, отражающих связи между показателями крови, регистрируемых при скрининге. Предлагаются: информационно-аналитическая модель скрининг диагностики, идентифицированные в математические модели, результаты апробирования на клиническом материале, схема использования диагностического модуля в учебном процессе повышения квалификации. В качестве интегрального информативного диагностического показателя формируется «ЭрГеМо», вычисляемый по концентрации моноцитов и значению цветового показателя в крови. Показано, что: характер линейных связей в процессе развития заболевания классах практически не меняется; у больных людей - уменьшается в среднем на 16% концентрация эритроцитов и уровень гемоглобина в крови, примерно на 20% возрастает степень влияния на него концентрации эритроцитов; при синтезе математических моделей рекомендуется использовать меры доверия к используемым показателям и диагностическим процедурам, используемому клиническому материалу. Доказано, что: при возникновении рака желудка снижаются диагностические возможности «СОЭ» в условиях изменения связи с ЭрГеМо с отрицательно степенной на положительную; в функциональных зависимостях ЭрГеМо проявляется обратная зависимость от цветового показателя; при заболевании раком желудка снижается роль насыщенности кислородом крови в регуляции процесса кроветворения.

Ключевые слова: онкологические заболевания желудка, превентивная скрининг диагностика, моделирование, автоматизированные системы поддержки принятия решений

PREVENTIVE DIAGNOSIS OF ONCOLOGICAL DISEASES OF A STOMACH WHEN SCREENING BY MEANS OF SYSTEM OF SUPPORT OF ADOPTION OF CLASSIFICATION DECISIONS

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

Southwest state university, Kursk, email: artem1962@mail.ru

In article questions of formation of the knowledge base of system of support of adoption of diagnostic decisions on early diagnosis of cancer of stomach on the basis of the models reflecting communications between indicators of blood, registered when screening are considered. Are offered: information and analytical model diagnostics screening, identified in mathematical models, results of approbation on clinical material, the scheme of use of the diagnostic module in educational process of professional development. As an integrated informative diagnostic indicator ErGemo calculated on concentration of monocytes and value color indicators in blood is formed. It is shown that: the nature of linear communications in development of a disease classes practically does not change; at sick people - concentration of erythrocytes and level of hemoglobin in blood decreases on average by 16%, approximately for 20% extent of influence of concentration of erythrocytes on it increases; at synthesis of mathematical models it is recommended to use trust measures to the used indicators and diagnostic procedures, the used clinical material. It is proved that: when developing cancer of stomach diagnostic opportunities of «BSR» (blood sedimentation rate) in the conditions of change of communication with ErGemo with negatively sedate on positive decrease; in functional dependences of ErGemo inverse relation from a color indicator is shown; at a disease of cancer of a stomach the saturation role blood oxygen in blood formation process regulation decreases.

Keywords: oncological diseases of a stomach, preventive screening diagnostics, modeling, the automated systems of support of decision-making

Профилактика, превентивная диагностика и лечение онкологических заболеваний желудка представляют актуальную проблему во всем мире. По сводным эпидемиологическим данным, это заболевание в мире является четвертым наиболее частым после рака легкого, молочной железы, толстой кишки. Ежегодно диагностируется более 930 тысяч новых случаев выявления заболевания, одногодичная летальность составляет более 700 тыс. (более 75%) [27]. В Российской Федерации ежегодно онкология желудка фиксируется у 46 тысяч человек, а одногодичная летальность составляет 56%

[28]. Среднее время от появления специфических симптомов до постановки диагноза «рак желудка» (РЖ) составляет 3 месяца, что свидетельствует как о поздней обращаемости пациентов к врачу, и об объективных сложностях диагностики и о необходимости оптимизации тактики обследования на превентивном или поликлиническом уровнях.

В настоящее время мероприятия в области превентивной медицины [26] и профилактики рака желудка [28] сводится к опросу потенциальных больных, принадлежащих к группе риска во время плановой диспансеризации, скрининга с использованием и

без средств телемедицины [13] путем сбора жалоб на существующее состояние, анкетирования и проведения простейших анализов и проб. Симптоматика онкологических заболеваний желудка на ранней стадии развития заболеваний проявляется слабо и, следовательно, пациент своевременно не обращается за медицинской помощью. Регистрация и анализ онкомаркеров крови в условиях массового скрининга не всегда возможно (особенно в сельской местности).

Между тем, применение существующих для решения диагностических задач в технических и биотехнических системах, теоретические исследования и практика в области систем поддержки принятия решений (СППР) и различных способов прогнозирования поведения систем [4, 14] позволяет существенным образом снизить летальность и инвалидность, уменьшить экономические затраты на лечение и последующую реабилитацию больных.

Тем самым, актуальность научных и практических изыскания в рассматриваемой предметной области биологических наук обуславливается необходимостью разработки и применения автоматизированных систем поддержки принятия решений диагностики рака желудка при скрининге и-или профилактических осмотрах на основе достижений искусственного интеллекта и информационно-компьютерных технологий.

В настоящее время, превентивная диагностика онкологических заболеваний на ранней стадии базируется в основном на применении определенных онкомаркеров

(обладающих диагностической чувствительностью не более 0,7) и узкоспециализированного осмотра с учетом анализа косвенных признаков, выявленных в процессе беседы с обследуемым.

Учитывая достижения в представлении организма как целостной интегральной системы [29, 32], превентивной медицины и медицинских информационных технологий [16, 18, 23], предлагается использовать в ходе массового профилактического медицинского осмотра населения общие показатели крови. Рассматриваемый подход основывается на гипотезе об информативности крови как соединительной ткани, выполняющей транспортные функции, о функционировании различных функциональных и физиологических систем и организма в целом (особенно в случае системных нарушений, которые, безусловно, обуславливают онкологические заболевания) [20]. Применение подобных систем позволит существенным образом повысить эффективность массовых медицинских осмотров населения для выявления тяжелых заболеваний на ранней стадии и своевременно принимать решение о необходимости госпитализации или клинического обследования.

Материалы и методы

В процессе скрининг диагностики и в ходе массовой профилактической диспансеризации населения в обязательном порядке осуществляется общий анализ крови – регистрируются такие показатели как: концентрации эритроцитов, лейкоцитов, лимфо-

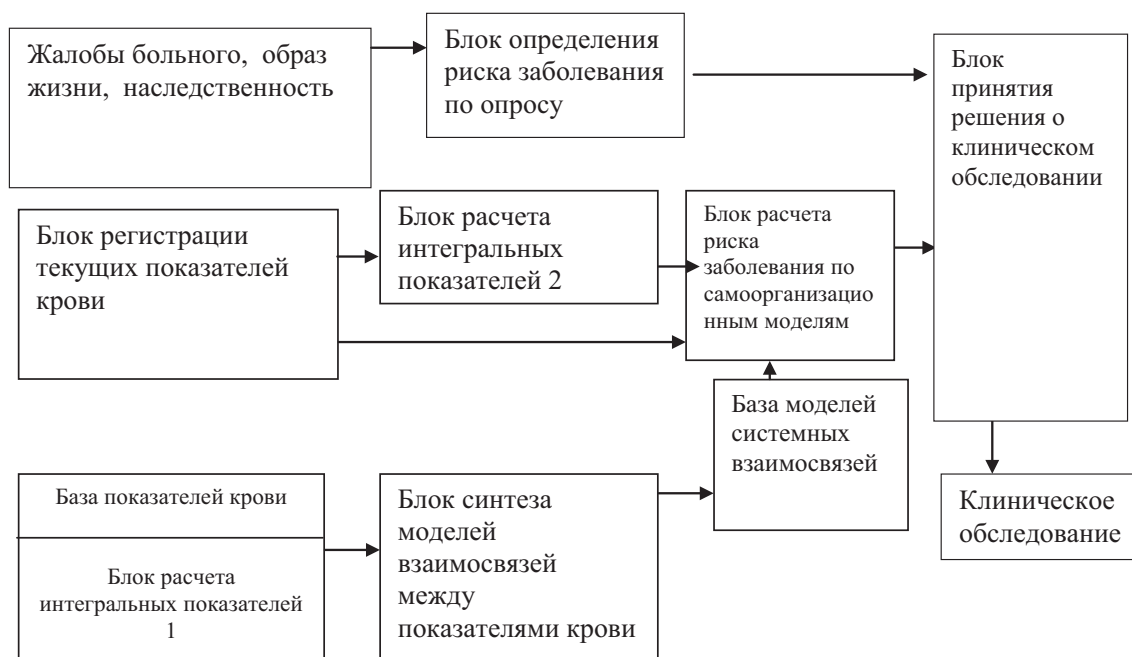


Рис. 1. Информационно-аналитическая модель скрининг диагностики рака желудка

цитов, гемоглобина, цветовой показатель, СОЭ, содержание глюкозы, натрия, калия, билирубин и т.д. Эти показатели определим как частные (непосредственно регистрируемые) - множество $\{X\}$.

Кроме того, значения показателей крови на ранних этапах развития злокачественных заболеваний организма в целом (желудка, в частности) отличаются от значений на этапах выявленного заболевания при обращении больного в клинику [27,28,21].

На основе изученного материала предлагается информационно-аналитическая модель поддержки принятия решений на этапе скрининг диагностики рака желудка, представленная на рисунке 1 [10] (аналогичный подход использовался при обработке результатов скрининг – диагностики в гастроэнтерологии [33]).

На принятие управленческого решения о необходимости проведения клинического обследования пациента на предмет выявления рака желудка оказывают влияние различные факторы, выявленные в ходе беседы (анкетирования) с обследуемым. В «Блоке определения риска заболевания по опросу» реализуется методика вычисления значения риска соотношения обследуемого к группе с возможным присутствием заболевания изложенная авторами в работах [1, 2].

Основной группой анализируемых показателей превентивной скрининг диагностики возможности наличия онкологических заболеваний желудка являются, согласно принятой гипотезы, значения показателей, полученных в процессе лабораторного анализа крови и онкомаркеров (РЭА, СА242, СА72.4, СА19.9). Необходимый информационный архив (клинически подтвержденный) формируется в «Базе показателей крови».

Для синтеза диагностических решающих классов формируются обучающие выборки. Процедура обучения (формирование базы знаний) реализуется в «Блоке синтеза моделей взаимосвязей между показателями крови» предлагаемой информационно-аналитической модели. В качестве исходных данных используются: множество векторов $\bar{X}=x_1, \dots, x_n$ и множество интегральных (латентных) показателей (вектор $\bar{Y}=y_1, \dots, y_n$). Значения интегральных показателей определяется по формуле (1):

$$Y_{k,j} = \frac{\sum_{i=1}^n (\alpha_{i,k} \cdot \frac{x_{i,j} - \bar{X}_{\omega_{0i}}}{\sqrt{D_{\omega_{0i}}}})^2}{\sum_{i=1}^n \alpha_{i,k}}, \quad (1)$$

где: $Y_{k,j}$ - j-ое – значение k-го интегрального показателя, $x_{i,j}$ - j-ое значение i-го частного показателя крови у пациента, $\bar{X}_{\omega_{0i}}$ - среднее значение i-го частного показателя в классе ω_0 , $D_{\omega_{0i}}$ - дисперсия i-го частного показателя крови, $\alpha_{i,k} \in [0,1]$ - весовые коэффициенты, определяющие информационный вклад показателя x_i в формирование интегрального показателя Y_k . Диагностические возможности показателей, рассчитанных по формуле (1), исследовались в работах [5, 6, 11, 34].

В предлагаемой системе предусмотрено несколько вариантов определения указанных коэффициентов: экспертное заключение, на основе дисперсионного анализа, использование информационного критерия Кульбака или критерия Стьюдента.

По сути, значения интегральных показателей характеризуют нормированное по дисперсии взвешенное расстояние «образа обследуемого» в гиперпространстве от центра альтернативного класса. Под альтернативным классом понимается класс не больных онкологическими заболеваниями людей (в дальнейшем именуемый как «здоровые») - класс ω_0 . Значения интегральных показателей вычисляются в «Блоке расчета интегральных показателей 1» при формировании обучающих выборок и в «Блоке расчета интегральных показателей 2» при решении диагностической задачи.

Результатом работы «Блока синтеза моделей взаимосвязей между показателями крови» является набор моделей, характеризующих множественные связи между векторами признаков $\bar{X}$ и интегральных показателей $\bar{Y}$ для основного (ω_0) и альтернативного (ω_1) классов. Работа «Блока синтеза моделей взаимосвязей между показателями крови» основана на использовании самоорганизационных алгоритмов структурно-параметрической идентификации метода группового учета аргументов (МГУА) [19, 24, 25,37]. Применяется ортогональный многорядный алгоритм [30], позволяющий синтезировать математические модели вида (2):

$$z_{il} = A_{il,0}^{w_l, \{X\} \vee \{Y\}} + \sum_{i=1}^{nr} (A_{il,i}^{w_l, \{X\} \vee \{Y\}} \cdot \prod_{j=1, j \neq i}^{mr} z_{j,il}^{p_{il,j}^{w_l, \{X\} \vee \{Y\}}}), \quad (2)$$

где: z_{il} - переменная (из множеств $\{X\}$, $\{Y\}$); $A_{il,i}$ - весовой коэффициент термина i для модели отклика функции z_{il} , $A_{il,0}^{w_l, \{X\} \vee \{Y\}}$ – свободный член для модели отклика z_{il} в классе w_l ($l=0,1$) на множествах $\{X\}$ или $\{Y\}$, $A_{il,i}^{w_l, \{X\} \vee \{Y\}}$ - весовой коэффициент термина i для отклика функции z_{il} в классе w_l ($l=0,1$) на множестве $\{X\}$ или $\{Y\}$, $p_{il,i,j}^{w_l, \{X\} \vee \{Y\}}$ - степень аргумента j в терме i для модели отклика функции отклика z_{il} в классе w_l ($l=0,1$) на множестве $\{X\}$ или $\{Y\}$; nr

– количество рядов селекции (термов полинома); m_T – количество переменных z .

Для оптимизации вычислительных процедур формируется «База моделей системных взаимосвязей», в которой множество идентифицированных моделей располагается по критерию информативности.

В «Блоке расчета риска заболевания по самоорганизационным моделям» основного класса заболеваний ω_l , используя вектора $\bar{X}$ и $\bar{Y}$, по математическим моделям вида (2) вычисляются значения модифицированных векторов класса ω_l вида:

$$\begin{aligned}\bar{X}_{\omega_l}^* &= F_{X1}(A_1, \bar{X}_{\omega_l}), \\ \bar{Y}_{\omega_l}^* &= F_{Y1}(B_1, \bar{Y}_{\omega_l}), \\ \bar{Y}_{\omega_l}^* &= F_{Y1 \leftarrow \{X1, Y1\}}(C_1, \bar{X}_{\omega_l}),\end{aligned}\quad (3)$$

где A_1 , B_1 и C_1 вектора настраиваемых параметров, F_{X1} , F_{Y1} и $F_{Y1 \leftarrow \{X1, Y1\}}$ – функционалы, соответствующие структурам моделей взаимосвязей исследуемых показателей (модели 2), идентифицированные соответственно на множествах $\{X\}$, $\{Y\}$ класса ω_l .

Аналогично определяются модифицированные вектора для класса ω_0 :

$$\begin{aligned}\bar{X}_{\omega_0}^* &= F_{X0}(A_0, \bar{X}_{\omega_0}), \\ \bar{Y}_{\omega_0}^* &= F_{Y0}(B_0, \bar{Y}_{\omega_0}), \\ \bar{Y}_{\omega_0}^* &= F_{Y0 \leftarrow X0, Y0}(C_0, \bar{X}_{\omega_0}).\end{aligned}\quad (4)$$

По полученным векторам определяются пары мер близостей D_1 и D_0 между векторами $\{\bar{X}, \bar{Y}\}$, $\{\bar{X}_1^*, \bar{Y}_1^*\}$ и $\{\bar{X}, \bar{Y}\}$, $\{\bar{X}_0^*, \bar{Y}_0^*\}$.

Полученные значения мер близости являются аргументами для расчета риска появления исследуемого заболевания соответствующим «Блоком расчета риска заболевания по самоорганизационным моделям».

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

Выбранный математический аппарат исследования относится к классу систем распознавания образов с учителем, предполагающий использование на этапе обучения репрезентативные выборки (обучающая – для идентификации моделей и экзаменационная – для отбора наилучших моделей, наиболее удовлетворяющих субъектов диагностического процесса [12] и расчета критериев качества работы СППР).

При формировании выборок рекомендуется придерживаться соотношения «золотого сечения»: $N_2/N_1 = N_1/(N_1+N_2)$, где N_1 – объем обучающей выборки, N_2 – объ-

ем контрольной выборки ($N_1 \approx 0.62(N_1+N_2)$, $N_2 \approx 0.38(N_1+N_2)$).

Для формирования статистически репрезентативных выборок применяется следующий подход. Значения всех используемых показателей нормируются в интервал $[0+e, 1-e]$ ($e=1/(N_1+N_2)$) путем линейного преобразования. Затем вычисляется для каждого вектора значений различных показателей средняя величина значений показателей. По полученным значениям осуществляется упорядочивание элементов исходного множества по возрастанию. В соответствии с выбранным соотношением N_1/N_2 формируются обучающие и экзаменационные выборки по следующей технологии. Все «четные» элементы входят в обучающую выборку, «нечетные» – в экзаменационную. Из полученной экзаменационной выборки в обучающую выборку переводится каждый пятый элемент.

С учетом идеологии GMDH общий процесс синтеза правил принятия решений осуществляется в соответствии со следующим методом:

1. Формируется пространство информативных признаков: $X = x_1, x_2, \dots, x_{n1}$ – частные показатели общего анализа крови; $Y = y_1, y_2, \dots, y_{n2}$ – интегральные показатели; $Q = q_1, q_2, \dots, q_{n3}$ характеристики образа жизни, предыдущих и существующих заболеваний, жалоб больного, наследственность и т.п.;

2. Формируются обучающая и контрольная выборки, осуществляется расчет меры доверия к исходным данным (МДД).

3. Для классов ω_0 и ω_l путем применения специализированного программного обеспечения [12] формируется пакет математических моделей типа (3), (4), по которым рассчитываются коэффициенты детерминации моделей аппроксимантов R_t^2 и R_s^2 (где $t=1, \dots, T$ номер моделей (3) в общем их списке (класс ω_l), $s=1, \dots, S$ – номер моделей (4)) в общем их списке моделей класса ω_0). Коэффициенты детерминации выступают в качестве мер доверия к адекватности математических моделей МДМ0 и МДМ1 соответствующих альтернативных классов. В пакеты моделей базы знаний СППР включаются только те, у которых значения коэффициентов детерминации превышают определенные пороговые уровни R_t^{Π} и R_s^{Π} .

4. Определяется общая мера доверия к адекватности по отношению к возможностям моделей описывать структуры связей между используемыми признаками для каждого по выражениям (5):

$$МДМ_0 = 1 - \prod_{t=1}^T (1 - МДМ_{0t})$$

$$МДМ_1 = 1 - \prod_{s=1}^S (1 - МДМ_{0s}). \quad (5)$$

5. Определяется мера доверия к адекватности моделей взаимосвязи МДМ для двух альтернативных классов по формуле (6):

$$МДМ = \frac{\gamma_0 \cdot МДМ_0 + \gamma_1 \cdot МДМ_1}{\gamma_0 + \gamma_1}. \quad (6)$$

где: γ_0, γ_1 - весовые коэффициенты, определяющие предпочтение риска соотношения пациента к классу здоровых или возможно болеющих людей.

6. Синтезируются формулы определения мер соответствия результатов текущих измерений x_i и значений интегральных показателей y_k с результатом расчетов по математическим моделям (3)-(4). Для расчета мер соответствия по аналогии с классическим понятием функций принадлежности введем функцию соответствия результатов модельных вычислений и измерений: $f_{wl}(d_{i,l}), f_{wl}(d_{k,l})$. В качестве базовой переменной выступают меры близости $d_{i,l}$ и $d_{k,l}$ между измеренными x_i и y_k и вычисленными по формулам (3) - (4) x_i^* и y_k^* значениями признаков и интегральных показателей – определяются по формулам (7):

$$d_{i,\ell} = \left| \frac{x_i - x_i^*}{x_i} \right|, d_{k,\ell} = \left| \frac{y_k - y_k^*}{y_k} \right| \quad (7),$$

где ℓ - номер исследуемых классов ($\ell=0,1$).

Функции соответствия с базовыми переменными $d_{i,\ell}$ и $d_{k,\ell}$ определяются в виде кусочно-линейной зависимости (8):

$$f_{\omega_\ell}(d_\ell) = \begin{cases} 1, & \text{если } d_\ell \leq a \\ \frac{b-d_\ell}{b-a}, & \text{если } b > d_\ell > a, \\ 0, & \text{если } d_\ell > b \end{cases} \quad (8)$$

где d_ℓ - базовая переменная, определяемая по формулам (7).

7. Определяются уверенности UM_{ω_ℓ} в отнесении неизвестного объекта, описываемого наборами x_i и показаний y_k к исследуемым классам состояний ω_0 и ω_1 (ω_ℓ), по моделям МГУА согласно формулы (9):

$$UM_{\omega_\ell} = \frac{\sum_{i=1}^{m_1} f_{\omega_\ell}(d_{i,\ell}) + \sum_{k=1}^{m_2} f_{\omega_\ell}(d_{k,\ell})}{m_1 + m_2}, \quad (9)$$

где $\ell=0,1$, m_1 - количество моделей, построенных по множеству X; m_2 - количество моделей, построенных по множеству Y.

С целью учета «желаемых» взаимоотношений между значениями ошибок первого и второго рода принятие классификационных решений рекомендуется осуществлять с использованием порогов P_{ω_0} и P_{ω_1} , относительно которых строится таблица 1 принятия решений. Если значение UM_{ω_ℓ} превышает значения порога P_{ω_ℓ} , то элемент таблицы равен нулю. В противном случае - единице. Величины порогов P_{ω_ℓ} определяются экспертами на этапе обучения.

В качестве решения предлагается использовать заключения: 1 – пациент нуждается в дополнительном обследовании; 2 – вероятность присутствия рака желудка низка (пациент практически здоров); 3 – высока вероятность присутствия заболевания рака желудка.

8. Для признаков, сформированных по анализу обследуемого Q, на основании данных специальной литературы и опыта экспертов согласно рекомендациям [26, 28] определяются соответствующие риски по отношению к заболеванию раком желудка Rt_{ω_1} на основании которых формируются частные уверенности UQ в классе ω_1 :

$$UQ = F_Q(Q, Rt_{\omega_1}) \quad (10).$$

9. На экзаменационной выборке с известным, клинически подтвержденным диагнозом, проверяется качество срабатывания полученных решающих правил и,

Таблица 1

Диагностические заключения относительно порогов $P_{\omega_\ell}/\ell=0,1$

$UM_{\omega_0} > P_{\omega_0}$	$UM_{\omega_0} < P_{\omega_0}$	$UM_{\omega_0} < P_{\omega_0}$	$UM_{\omega_0} > P_{\omega_0}$	№ ситуации	Описание ситуации
$UM_{\omega_1} > P_{\omega_1}$	$UM_{\omega_1} > P_{\omega_1}$	$UM_{\omega_1} < P_{\omega_1}$	$UM_{\omega_1} < P_{\omega_1}$		
0	0	0	0	1	Дополнительное обследование
0	0	0	1	2	Здоров – класс ω_0
0	0	1	0	1	Дополнительное обследование
0	1	0	0	3	Болен – класс ω_1
1	0	0	0	1	Дополнительное обследование

если оно удовлетворительно, то процесс обучения заканчивается. В противном случае, осуществляется корректировка параметров решающих правил в сторону минимизации ошибок классификации.

Результаты исследований

Исходное множество показателей общего анализа крови определялось экспертами – медиками в соответствии с рекомендациями [11, 25]. Была сформирована группа из 9 высококвалифицированных экспертов (четыре доктора наук, два кандидата наук и три врача высшей квалификации). Коэффициент конкордации экспертов выбранной группы составил 0,86. В результате в исходное множество анализируемых показателей были включены: X1 – «эритроциты», X2 – «гемоглобин», X3 – «цветовой показатель», X4 – «лейкоциты», X5 – «эозинофилы», X6 – «палочкоядерные», X7 – «сегментоядерные», X8 – «лимфоциты», X9 – «моноциты», X10 – «СОЭ». На уровне значимости ошибки первого рода $p < 0.01$ выделяем показатели X3 (цветовой показатель) и X9 (моноциты), которые формируют множества информативных показателей $\{X\}$. Затем, согласно рекомендациям [8, 17, 31, 36] было сформировано множество информативных показателей.

Поскольку у двух оставшихся показателей значения коэффициента Стиюдента близки, то принимается решение об объединении их в единый кластер для формирования интегрального показателя Y по формуле (1) с весовыми коэффициентами: $\alpha_{x3}=0,44$, $\alpha_{x9}=0,56$. В этом случае, согласно (1) получаем следующую формулу для вычисления интегрального показателя:

$$Y = \frac{0.44 \cdot (x_3 - 0.96)^2}{0.002} + \frac{0.56 \cdot (x_9 - 4.9)^2}{5.7} \quad (11),$$

где x_3 , x_9 – значения показателей X3 и X9.

Назовем этот показатель Y – «ЭрГеМо» – поскольку его аргументами являются «Цве-

товой показатель» и «Моноциты». Первый, в свою очередь рассчитывается, исходя из значений показателей: «Эритроциты» и «Гемоглобин».

Для его включения в множество показателей, предположительно обладающими диагностическими возможностями, убеждаемся, что ошибка первого рода различий между альтернативными классами менее 0,01.

Таким образом, множество информативных показателей включает следующие показатели крови – «Эритроциты», «Гемоглобин», «Цветовой показатель», «Моноциты», «СОЭ» и интегральный показатель «ЭрГеМо».

На основании анализа историй болезни и «добровольцев» (группа не больных раком желудка) были сформированы две выборки для каждого из альтернативных классов w_0 и w_1 объемами по 115 обследуемых.

Половозрастное распределение «больных раком желудка» и «не больных раком желудка» подчинялось равномерному закону распределения. Объем выборки определялся исходя из рекомендаций [15] для обеспечения чувствительности статистических критериев различий между классами на контрольных выборках на уровне 0,95. Соответственно, мера доверия к выборке МДО = 0,95.

Согласно принципу «золотого сечения» обучающие выборки включали 78 объектов в каждом из классов. В таблице 2 приведены основные статистические характеристики показателей обучающих выборок в классах.

Наилучшие статистические отличия между собой показывают показатели согласно критерию Стиюдента (на уровне $p < 0.01$): «Эритроциты», «Моноциты», «СОЭ», «ЭрГеМо». Это означает, что значения МДП $\approx 0,67$. Поскольку, согласно таблицы 3, наблюдается пересечение классов доверительных интервалов по указанным в ней показателям, то значение меры доверия к выборкам МДВ $\approx 0,83$.

Таблица 3

Статистические характеристики показателей для классов w_0 и w_1

Классы	Класс w0				Класс w1			
Показатели	Статистические характеристики							
	M	D	As	Ex	M	D	As	Ex
Эритроциты	4,24	0,09	0,47	0,29	4,1	0,66	-0,09	-0,176
Гемоглобин	106,8	693	-0,55	-0,009	109	833	-0,02	-0,68
Цветовой показатель	1,17	0,46	1,06	1,14	1,06	0,21	0,42	0,566
Моноциты	6,39	10,3	0,77	0,62	4,1	4,8	2,08	9,24
СОЭ	5,43	5,66	0,96	0,57	10,8	33,3	0,93	0,63
ЭрГеМо	4,09	3,36	0,72	0,45	2,7	1,63	1,82	7,34

В таблице: M, D, As, Ex – соответственно: математическое ожидание, дисперсия, асимметрия и эксцесс соответствующих показателей.

Поскольку обучающая и контрольная выборки формировались по предложенной выше методике, то мера доверия к репрезентативности МДР=1. Набор показателей определялся одновременно с анализом литературных источников справочного характера (которые можно воспринимать как экспертов) и по критерию Стьюдента - следовательно, меры доверия экспертов к выборке и к составу признаков (МДЭВ и МДЭП) принимаются за 1. В результате - мера доверия к данным МДД=0,94. Так как, контрольная и обучающая выборки взаимно репрезентативны, то МДДо=МДДк=МДД=0,94.

Значимые парные корреляционные связи в альтернативных классах показаны на рисунках 1 и 2.

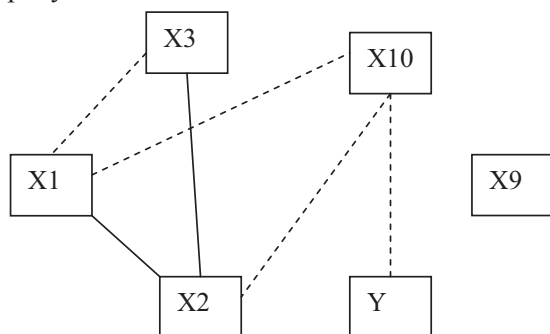


Рис. 1. Корреляционные связи в классе w_0 (сплошные линии – положительная корреляция, пунктирные – отрицательная)

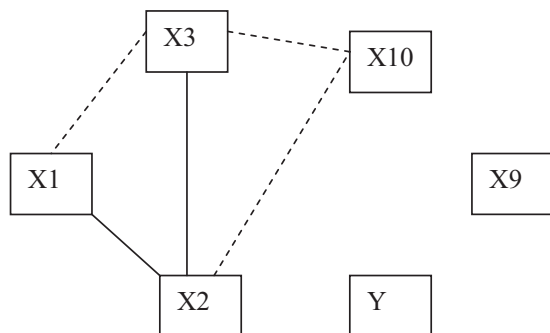


Рис. 2. Корреляционные связи в классе w_1

Таблица 3

Таблица распределения результатов диагностики для класса ω_1

Обследуемые	Результаты срабатывания правил		Всего
	положительные	Отрицательные	
$\eta_{\omega_1}=100\cdot78$	51	27	78
$\eta_{\omega_0}=100\cdot78$	17	61	78
Всего	68	88	156

ДЧ=0,75; ДС=0,69; ДЭ=0,72.

В качестве показателей качества характеризующих статистическую достоверность диагностических правил выбраны: диагностическая чувствительность (ДЧ), диагностическая специфичность (ДС), диагностическая эффективность (ДЭ) [21, 23]. Необходимые для расчета результаты исследования, полученные на репрезентативной экзаменационной выборке обследуемых, приведены в таблице 3.

Обсуждение и заключение

Анализ корреляционных связей (рисунки 1 и 2) подтверждает структурные изменения в крови при заболевании раком желудка: исчезает корреляция между такими показателями как «СОЭ»- «Эритроциты» и «СОЭ»-«ЭрГеМо». Уменьшение количества значимых корреляционных связей позволяет предположить, что анализируемое заболевание приводит к деструктуризации процессов управления соотношений концентраций (количеств) определенных элементов в крови. Уравнения парной регрессии между этими показателями включаются в базу знаний СППР решения диагностических задач. Модели представляются следующими формулами:

$$\overline{X10^*}_{1,w_0} = 20.53 - 3.75 \cdot \overline{X1} \quad R^2_{1,10,w_0} = 0.21,$$

$$\overline{X10^*}_{2,w_0} = 7.14 - 0.667 \cdot \overline{Y} \quad R^2_{2,11,w_0} = 0.21 \quad (12)$$

Заметим, что слабая корреляционная связь или ее отсутствие между показателями сформированного множеств $\{X\}$ $\{Y\}$ гиперпространства, подчеркивает их ортогональность, подтверждая полученные высокие меры доверия.

Анализ полученных моделей (приведены в [3]), позволяет сделать следующие заключения:

1. Характер линейных связей в альтернативных классах не меняется;
2. У больных раком желудка людей уменьшается в среднем на 16% концентрация эритроцитов, что приводит к соответствующему изменению значения цветового показателя.
3. У больных – уровень гемоглобина в крови снижается приблизительно на 16% , примерно на 20% возрастает степень влияния на него концентрации эритроцитов.

4. В анализируемых альтернативных классах в моделях присутствуют как общие для них аргументы, так и оригинальные – это говорит о том, что состав крови определяется как общими для всего организма «правилами», так и специфическими для определенной системной патологии, к которой относится рак желудка;

5. Количество эритроцитов в крови у не болеющих раком желудка людей характеризуется дополнительным влиянием отношением значений показателей ЭрГеМо и СОЭ, что подчеркивает возможность учета данного показателя крови для превентивной скрининг-диагностики рака желудка;

6. При возникновении рака желудка снижаются диагностические возможности показателя «СОЭ» в условиях изменения связи с ЭрГеМо с отрицательно степенной на положительно степенную, что позволяет использовать этот факт в качестве диагностической характеристики;

7. Количество моноцитов в крови при заболевании изменяет свою «связь» с цветовым показателем с прямой на обратную, что подчеркивает происходящие в этом случае изменения содержания «гемоглобина в эритроцитах»;

8. Показатель «СОЭ» при заболевании не имеет выраженной функциональной зависимости с остальными анализируемыми показателями, что подчеркивает дестабилизации системы кроветворения;

9. В функциональных зависимостях интегрального показателя ЭрГеМо на высоком уровне корреляции появляется обратная зависимость от показателя «цветовой показатель», что подчеркивает необходимость учета показателей «Эритроциты» и «Гемоглобин» при скрининг диагностики заболевания;

10. В классе w_1 минимизируется связь показателя ЭрГеМо с «Цветным показателем», что говорит об уменьшении роли насыщенности эритроцитов кислорода регуляции функций кроветворения при заболевании раком желудка.

Полученные компоненты базы знаний могут быть использованы как в скрининг диагностики рака желудка, так и в обучающем процессе повышения квалификационного уровня медицинских работников, например, по схеме, аналогично рассмотренной в [35] – сравниваются результаты работы диагностического модуля автоматизированной СППР (AMESDD -autonomous module expert support diagnostic decisions) и обучаемого медработника. Схема учебного тестирования приведена на рисунке 3.

УЧИТЕЛЬ на основании информации о рассогласовании действий субъекта обучения с правильной оценкой ситуации (δS и δX) с помощью воздействует на субъект обучения через модуль коррекции приобретаемых знаний и навыков, оценивает уровень профессионализма, формирует статистическую отчетность. В процессе обучения осуществляется активное тестирование субъекта обучения путем генерации тестов,

включающих себя: либо значения характеристик состояния виртуального пациента X , либо его гипотетическое состояние Sit .

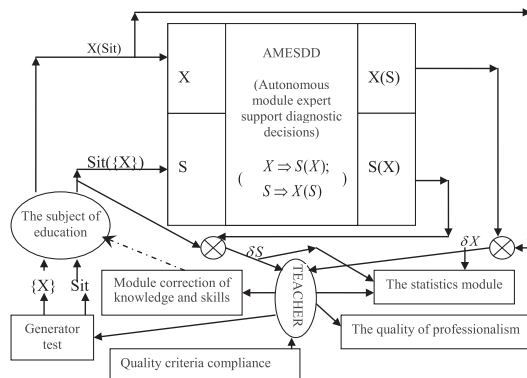


Рис. 3. Схема применения AMESDD в интерактивном тестировании профессиональных знаний и навыков, где: X – множество информативных показателей, $S(X)$ – состояние обследуемого, соотнесенное к определенному классу и полученное по диагностическим правилам импликации $X \Rightarrow S(X)$; $X(S)$ – множество значений X , необходимое для идентификации состояния S , согласно импликациям $X \Rightarrow S(X)$; $\{X\}$ – множество значений показателей X виртуального обследуемого; $Sit(\{X\})$ – состояние виртуального обследуемого согласно мнению субъекта обучения; Sit – «истинное» состояние виртуального обследуемого; δS – рассогласование между $X(Sit)$ и $X(S)$; δX – рассогласование между $Sit(\{X\})$ и $S(X)$; $\otimes$ – блоки вычисления сигналов рассогласования

На основе полученной о состоянии виртуального обследуемого информации, собственных и приобретенных в процессе обучения знаний и опыта, субъект обучения формирует $Sit(\{X\})$ или $X(Sit)$, которые поступают на соответствующие входы AMESDD. AMESDD на основании прямых и обратных решающих правил [7] либо диагностируют состояние виртуального обследуемого (идентифицирует $S(X)$) либо определяют присутствие у него диагностические информативных признаков ($X(S)$). Блоки оценки рассогласования между правильными («эталонными») решениями (сформированными AMESDD) и предложенными обучаемыми формируют соответствующие значения сигналов рассогласования (δS и δX), по которым УЧИТЕЛЬ реализуют процедуры обучения, определенные целями программ повышения квалификации медработника.

Список литературы

1. Артеменко М.В. Синтез диагностических правил заболевания в условиях массового обследования населения //Медико-экологические информационные технологии – 2015: сборник материалов XVIII Международной научно-технической конференции /редкол.: Н.А. Корневский и др.; Юго-Зап. гос. ун-т. Курск, 2015. – стр.126-130

2. Артеменко М.В. Формирование и применение опросника для автоматизированной системы поддержки принятия диагностических решений превентивной медицины скрининга рака желудка // *Международный журнал прикладных и фундаментальных исследований*. – 2015. – № 11-2. – стр. 184-190; URL: <http://www.applied-research.ru/article/view?id=7703> (дата обращения: 15.12.2016).
3. Артеменко М.В. Бабков А.С. Информационная поддержка принятия решений в скрининг диагностики рака желудка // *Научный вестник*, -2016, №1(7), - стр. 115-131; DOI: 10.17117/nv.2016.01.115
4. Артеменко М.В., Бабков А.С. Классификация методов прогнозирования поведения систем // *Современные проблемы науки и образования*. – 2013. – № 6.; URL: <http://www.science-education.ru/ru/article/view?id=11527> (дата обращения: 17.09.2016).
5. Артеменко М.В., Дронова Т.А. Количественная оценка различий соорганизации физиологических функций в диагностическом процессе // *Вестник новых медицинских технологий*. 2006., Т. 13. № 2. - стр. 127-129.
6. Артеменко М.В., Дронова Т.А., Корневский Н.А. Применение показателей системной организации в диагностическом процессе // *Системный анализ и управление в биомедицинских системах*. -2003, -Т.2, -№1. – стр. 16-19
7. Артеменко М.В., Добровольский И.И., Мишустин В.Н. Информационно-аналитическая поддержка автоматизированной классификации на основе прямых и обратных решающих правил на примере прогноза тромбозов. // *Современные наукоемкие технологии*. – 2015. №12 (часть 2). – стр.199-205.
8. Артеменко М. В., Калугина Н. М., Шуткин А. Н. Формирование множества информативных показателей на основании аппроксимирующего полинома Колмогорова-Габора и максимального градиента функциональных различий // *Известия Юго-Западного государственного университета. Серия: Управление, вычислительная техника, информатика. Медицинское приборостроение*, 2016, №1, стр. 116-223
9. Артеменко М.В., Подвальный Е.С., Старцев Е.А. Метод комплексной оценки и выборка состава информативных признаков в задачах оценки состояния биотехнических систем // *Биомедицинская радиоэлектроника*, 2016, №9, стр.38-44
10. Артеменко М.В., Серебровский В.В., Бабков А.С. Информационно-аналитическая модель поддержки принятия решений в процессе диагностики рака желудка // *Фундаментальные исследования*. – 2014. - №6, стр.18-23.
11. Артеменко М.В., Корневский Н.А., Жилинкова Л.А. Диагностика здоровья новорожденного путем системного анализа показателей беременной // *Вестник новых медицинских технологий*. 2003, Т.10, №3. –стр. 50-52
12. Артеменко М.В., Оболенский А.Н. Мультипликативная аппроксимация методом группового учета аргументов // *Свидетельство об официальной регистрации программы для ЭВМ № 2007611654 от 25.04.2007*
13. Атьков О.Ю., Кудряшов Ю. Ю. Персональная телемедицина. Телемедицинские и информационные технологии реабилитации и управления здоровьем. –М.: Практика, 2015. - 248 с.
14. Воронцов, И.М., Шаповалов В.В., Шерстюк Ю.М. Здоровье. Опыт разработки и обоснование применения автоматизированных систем для мониторинга и скринирующей диагностики нарушений здоровья. – СПб.: ООО «ИПК «Коста»Б, 2006. – 432 с.
15. Гланц С. Медико-биологическая статистика. Пер. с англ. – М.: Практика, 1999. - 259 с.
16. Дюк В., Эммануэль В. Информационные технологии в медико-биологических исследованиях. – СПб: Питер, 2003. – 528 с.
17. Дронова Т.А., Калугина Н.М. Щекина Е.Н. Системные изменения структуры информативных показателей как индикаторы кластеров состояний организма // *Научный вестник*. 2016. № 3 (9). С. 37-51; DOI: 10.17117/nv.2016.03.037
18. Илларионов В.Е. Научно-практические основы информационной медицины. – М.: Книжный дом «ЛИБРОКОМ», 2010. – 184 с.
19. Ивахненко А.Г., Юрачковский Ю.П. Моделирование сложных систем по экспериментальным данным. – М. Радио и связь, 1987. – 118с.
20. Кровь – индикатор стояния организма и его систем / Под ред. Р.В. Ставицкой. МНПИ, 1999. – 160 с.
21. Ключин Д., Петунин Ю. Доказательная медицина. Применение статистических методов. Из-тво: Диалектика, Вильянс, 2008.-320 с.
22. Кохонен Т. Самоорганизующие карты. –М.: БИНОМ. Лаборатория знаний, 2016. – 655 с.
23. Омельченко В., Демидова А. Математика. Компьютерные технологии в медицине. Из-тво Феникс, 2010.-592 с.
24. Орлов А.А. Принципы построения архитектуры программной платформы для реализации алгоритмов метода группового учета аргументов / А.А. Орлов // *Управляющие системы и машины*.-2013. - №2. -стр.65-71
25. Орлов А.И. Теория принятия решений. Учебное пособие. - М.: Издательство «Март», 2004. - 656 с.
26. Превентивная медицина: Опыт работы информационного полипараметрического комплекса / Под ред. Н.В. Дмитриевой. –М.: Книжный дом «ЛИБРОКОМ», 2010. – 248 с.
27. Рак желудка: симптомы, диагностика, лечение, профилактика [Электронный ресурс]. – Режим доступа: <http://natural-medicine.ru/4127-rak-zheludka.html> (дата обращения 18.11.2015).
28. Соленова Л.Г., Михайловский Н.Я., Смулевич В.Б. Первичная профилактика рака: Информ. бюл. 2009. URL: http://www.ppr-info.ru/attachments/158_PPR-1_9.pdf (дата обращения 29.12.2015).
29. Судаков К.В., Кузичев И.А., Николаев А.Б. Эволюция терминологии и схем функциональных систем в научной школе П. К. Анохина. –М.: Европейские полиграфические системы, 2010. – 240 с.
30. Справочник по типовым программам моделирования / А.Г. Ивахненко, Ю.В. Коппа, В.С. Степашко и др.; под ред. Ивахненко А.Г. – К.: Техника, 1980.- 184 с.
31. Сотникова А. А., Писарев М. В., Калугина Н.М. Выделение информативных признаков // *Медико-экологические информационные технологии – 2016: сборник научных статей по материалам XIX Международной научно-технической конференции / ред.: Н. А. Корневский [и др.]*; Юго-Зап. гос. ун-т. – Курск, 2016. – 308 с.
32. Шебшаевич Л.Г., Алексеев А.А. Жизнь – кибернетическая медико-биологическая системность («Геном человека», клонирование – критический анализ). - М.: Трида Плюс, 2001. – 608 с.
33. Agarkov N.M., Artemenko M.V., Babkov A.S., Dmitrieva V.V., Gorbatenko S.A. Screening diagnosis of gastric cancer based on fuzzy models of blood composition and on patients anamnesis // *Biomedical Engineering*. 2015. v. 49. № 2. p. 94-97.
34. Artemenko M.V. Quantitative measures for assessing functional state of the human body during the diagnostic process // *Biomedical Engineering*. – 2008, v.42, №2, p.92-96
35. Artemenko M. V., Dobrovolsky I.I. Automated training test-system of doctors trauma on the basis of expert module support diagnostic decisions // 3rd International Conference «Research, Innovation and Education» 25-30 January 2016 / The collection includes the 3rd International Conference «Research, Innovation and Education» by SCIEURO in London, 25-30 January, -2016. – p. 226 – 238
36. Artemenko M.V., Kalugina N.M., Dobrovolsky I.I. The formation of a set of informative features based on the functional relationships between the data structure field observations // *European Journal of Natural History*. 2016. № 6. С. 43-48
37. GMDH //URL: <http://gmdh.net/> (дата обращения 8.12.2016).