

Введение

Одной из актуальных задач медицины и информационных технологий является разработка и внедрение эффективных систем поддержки принятия решений, которые, используя методы интеллектуального анализа данных, помогают специалистам в задачах постановки диагнозов, прогнозирования развития заболеваний и т. д. В подавляющем большинстве случаев применение современного медицинского оборудования, решая одни проблемы — порождает другие. В настоящее время трудности диагностики возникают не из-за дефицита информации, а из-за недостаточной эффективности методов ее обработки. В какой-то мере решение указанных проблем обеспечивает создание систем интерпретации и анализа медицинских данных. При этом наибольший интерес вызывает разработка консультативных интеллектуальных систем, то есть экспертных систем, содержащих механизм объяснения и обоснования предлагаемых решений на языке, понятном пользователю [4].

Целью данного исследования является разработка некоторых методов анализа и интерпретации медицинских данных, получаемых с помощью микроволновой радиотермометрии.

1. Описание задачи

Одним из действенных способов повышения эффективности диагностики является автоматизация обработки медицинских данных с помощью методов искусственного интеллекта [11–15]. Однако большинство экспертных систем свои решения предлагает врачу или в детерминированной форме однозначного заключения, или в виде вероятностных оценок каждого из возможных диагнозов. В настоящее время идет процесс формирования новых подходов к созданию диагностических систем, «объясняющих» предлагаемые ими решения [4; 5].

Одним из наиболее перспективных методов функциональной диагностики, на основе которого вполне возможно создание эффективной консультативной интеллектуальной системы, является микроволновая радиотермометрия. Это биофизический метод неинвазивного обследования, заключающийся в измерении внутренних и поверхностных температур тканей по интенсивности их теплового излучения, соответственно в микроволновом (РТМ) и инфракрасном (ИК) диапазонах. В течение последнего десятилетия данный метод получил распространение в различных областях медицины [9]. Однако существующий на данный момент диагностический комплекс РТМ-01-РЭС является системой поддержки принятия решений специалиста высокой квалификации. Сложность восприятия информации, возникающая у медицинского персонала без специальной длительной подготовки, значительно снижает потенциальную возможность его широкого использования в скрининге. Таким образом, актуальной задачей является создание экспертной системы, обладающей возможностями обоснования предполагаемого диагностического решения.

Особую сложность в данной проблеме вызывает нахождение высокоинформативных признаков заболеваний. Как отмечается большинством специалистов, на будущее качество алгоритмов классификации влияют качественный и количественный составы пространства информационных признаков. Основной задачей данного исследования является разработка метода формирования пространства информационных признаков.

2. Качественные составляющие информационных признаков

Вначале опишем подробнее существующую методику диагностики рака молочной железы по данным микроволновой радиотермометрии. Комплекс РТМ-01-РЭС позволяет оценивать функциональное состояние тканей путем измерения внутренней температуры (РТМ) на глубине до 5 см и температуры кожи (ИК). Обследование пациентки (см., например, [2]) проводится в горизонтальном положении, обнаженной по пояс, руки под головой. Обследование начиналось с измерения температур в опорных точках Т1 и Т2, расположенных: первая — в центре грудной клетки сразу под и между молочными железами, вторая — непосредственно под мечевидным отростком. Далее измерения проводятся в 10 точках на каждой железе, и в аксиллярной области. При этом получаемые данные сразу выводятся на экран монитора (рис. 1).

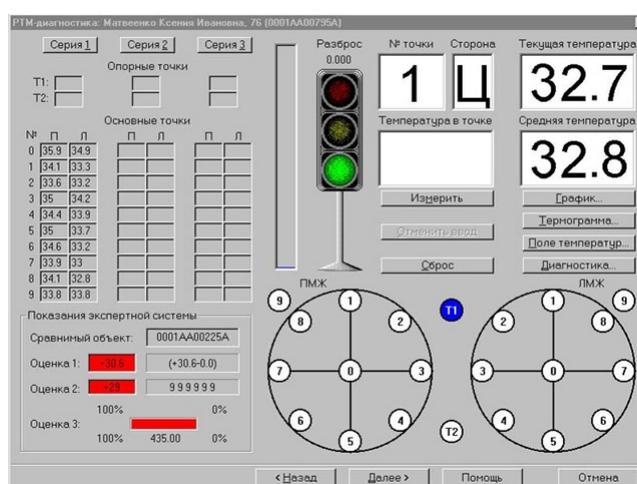


Рис. 1. Методика обследования молочных желез

После анализа информации о температурных полях молочных желез, в частности с помощью термокарт (см., например, рисунок 2), врач ставит диагноз пациентке. Здесь каждое значение температуры передается на экране монитора своим цветом. Участки с пониженной температурой передаются «холодными» цветами (синим), а с повышенной температурой — «теплыми» цветами (розовым, красным).

На основе данных, предоставленных онкологическими центрами России, была сформирована экспертная база термометрических данных. В настоящее время она включает в себя информацию о 734 молочных железах пациенток, которые делятся на два контрольных класса: «Здоровые» — 148 молочных желез и «Больные» — 586 молочных желез. Класс «Больные» в свою очередь делится на несколько групп: «Узловой рак» (185 молочных желез), «Диффузный рак» (13 молочных желез), «Узловые изменения, но не рак» (90 молочных желез), «Диффузные изменения, но не рак» (125 молочных желез), «Неотдифференцированные гистологически» (8 молочных желез), «Норма-2» (165 молочных желез). В класс «Норма-2» входят термометрические данные здоровых молочных желез больных пациенток. В рамках данного исследования молочные железы пациенток в базе были разделены на два новых класса: «Рак» — 326 молочных желез и «Не рак» — 408 желез. Статистический анализ используемых термометрических данных был проведен в работе [7].

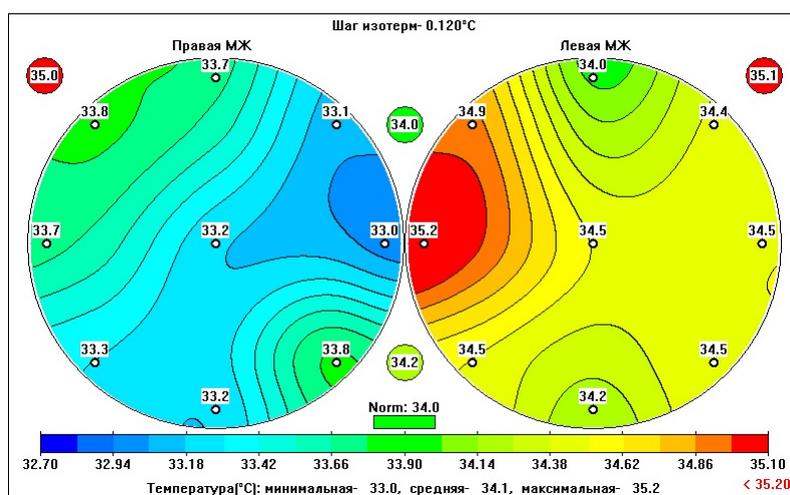


Рис. 2. Термокарты

В ходе исследований и анализа данных специалистами были выявлены следующие признаки рака молочной железы [2; 3; 8] (далее будем называть их качественными):

- повышенная величина термоасимметрии между одноименными точками молочных желез;
- повышенный разброс температур между отдельными точками в пораженной молочной железе;
- разница температур сосков;
- повышенная температура соска в пораженной молочной железе по сравнению со средней температурой молочной железы с учетом возрастных изменений температуры;
- соотношение кожной и глубинной температур и некоторые другие.

Важным этапом создания эффективной консультационной интеллектуальной системы является математическое описание данных признаков, выявление их количественных характеристик, а также выявление новых признаков.

- 1) Группа признаков, характеризующая асимметрию температурных полей молочных желез.

Данный класс признаков исходит из гипотезы о «зеркальной» симметрии температурных полей правой и левой молочных желез здоровых пациенток. Данная гипотеза используется при анализе термометрических данных не только молочных желез, но практически всех парных органов человека [9]. Опишем вначале известные признаки заболеваний, относящиеся к данному классу.

- 1.1) Повышенное значение разности температур между одноименными точками правой и левой молочных желез.

В качестве характеристик, описывающих этот эффект, могут быть использованы функции вида

$$|t_{i,r} - t_{i,l}|,$$

или

$$t_{i,r} - t_{i,l},$$

где $t_{i,r}$ и $t_{i,l}$ — температуры в i -х точках правой и левой молочных желез соответственно.

- 1.2) Повышенная разница температур сосков правой и левой молочных желез:

$$|t_{0,r} - t_{0,l}|,$$

или

$$t_{0,r} - t_{0,l}.$$

- 1.3) Повышенное среднеквадратичное значение разностей температур между одноименными точками правой и левой молочных желез:

$$\sqrt{\sum_{i=0}^8 \frac{(t_{i,r} - t_{i,l})^2}{9}}.$$

В общем виде асимметрия полей температур молочных желез может быть описана различными функциями вида

$$g(f(t_{i,r}, \dots, t_{n,r}) - f(t_{i,l}, \dots, t_{n,l})),$$

где $f(t_{i,r}, \dots, t_{n,r})$ — значение функции температур для точек t_0, \dots, t_n правой молочной железы, $f(t_{i,l}, \dots, t_{n,l})$ — значение аналогичной функции температур для точек t_0, \dots, t_n левой молочной железы, а $g(x)$ — некоторая функция одного переменного. Таким образом, значительно расширяется множество исследуемых параметров. В частности, таким способом можно описать следующие характеристики.

- 1.4) Разница средних значений температур «зеркально-симметрично» расположенных подобластей молочных желез, например:

$$\left| \frac{t_{0,MW,r} + t_{i,MW,r} + t_{(i \bmod 8)+1,MW,r}}{3} - \frac{t_{0,MW,l} + t_{i,MW,l} + t_{(i \bmod 8)+1,MW,l}}{3} \right|,$$

$$i = 1, \dots, 8.$$

- 1.5) Разница среднеквадратичных отклонений температур молочных желез:

$$\left| \left(\sum_{i=1}^8 \frac{(t_{i,MW} - t_{MW,m})^2}{8} \right)_r - \left(\sum_{i=1}^8 \frac{(t_{i,MW} - t_{MW,m})^2}{8} \right)_l \right|,$$

где $t_{MW,m} = \sum_{i=1}^8 \frac{t_i}{8}$ — среднее значение точек $1, \dots, 8$ правой или левой молочных желез соответственно и т. д.

- 2) Группа признаков, характеризующих повышенный разброс температур в пораженной молочной железе.

- 2.1) Повышенное среднеквадратичное отклонение температур в одной из молочных желез:

$$\sqrt{\sum_{i=0}^8 \frac{(t_i - \bar{t}_m)^2}{9}},$$

где $\bar{t}_m = \sum_{i=0}^8 \frac{t_i}{9}$.

- 2.2) Повышенный разброс температур между отдельными точками в пораженной молочной железе:

$$t_m - t_i,$$

где $t_m = \sum_{i=1}^8 \frac{t_i}{8}$, t_i — температура в i -й точке молочной железы.

В общем виде разброс температур молочных желез может быть описан различными функциями вида

$$g(f_1(t_0, \dots, t_n) - f_2(t_0, \dots, t_n)),$$

где $f_1(t_0, \dots, t_n)$ и $f_2(t_0, \dots, t_n)$ — функции температур точек t_0, \dots, t_n соответствующей молочной железы, а $g(x)$ — некоторая функция одного переменного. В частности, таким способом можно описать следующие характеристики.

- 2.3) Повышенные средние значения температур соседних точек вплоть до средней температуры молочной железы, то есть функции вида:

$$t_m - \frac{t_i + t_{(i \bmod 8)+1} + t_{(i \bmod 8)+2}}{3}$$

и т. п.

- 3) Группа признаков, характеризующих повышенное значение температуры соска в пораженной молочной железе.

- 3.1) Аномальная разность температуры соска и средней температуры молочной железы:

$$t_0 - t_m,$$

где $t_m = \sum_{i=1}^8 \frac{t_i}{8}$.

- 3.2) Аномальная разность температуры соска и температур отдельных точек молочной железы:

$$t_0 - t_i, i = 1, \dots, 8.$$

В целом аномальные значения температуры соска по отношению к другим параметрам можно описать функциями вида

$$g(t_0 - f_1(t_0, \dots, t_n)),$$

где $f_1(t_0, \dots, t_n)$ — функция температур точек t_0, \dots, t_n молочной железы, а $g(x)$ — некоторая функция одного переменного. В частности, таким способом можно описать следующие характеристики.

- 3.3) Аномальная разность температуры соска и средней температуры различных подобластей молочной железы, например:

$$t_0 - \frac{t_i + t_{(i \bmod 8)+1}}{2}$$

и т. п.

- 4) Признаки, характеризующие соотношение кожной и глубинной температур:

Аномальное значение разности между кожной и глубинной температурами точки пораженной молочной железы (внутренний градиент):

$$t_{i,MW} - t_{i,IR},$$

где $t_{i,MW}$ — глубинные и $t_{i,IR}$ — кожные температуры в i -й точке молочной железы. Разность температур молочных желез, измеренных в РТМ и ИК-диапазонах (так называемый внутренний градиент), может быть описана функциями вида

$$g(f(t_{0,MW}, \dots, t_{n,MW}) - f(t_{0,IR}, \dots, t_{n,IR})),$$

где $f(t_0, \dots, t_n)$ — функция температур точек t_0, \dots, t_n молочной железы, а $g(x)$ — некоторая функция одного переменного.

Отметим, что сформированный выше набор функций представляет собой расширенное описание известных качественных признаков, и получен на базе уже известных медицинских фактов. Но особенно важной, хотя и достаточно сложной задачей, является выявление новых знаний.

- 5) Признаки, базирующиеся на параметрах физико-математических моделей поведения температурных полей.

В течение последних лет было построено несколько математических моделей, описывающих поведение температурных полей молочных желез с помощью уравнений в частных производных второго порядка [9; 10]. Кроме того, отметим, что в предьявленном выше наборе функций присутствуют как функции температур, так и разностные аналоги их производных по различным направлениям. Например, величина $t_0 - t_i$ является разностным аналогом производной в радиальном направлении (радиальный градиент). Аналогично величина $t_{i,MW} - t_{i,IR}$ является разностным аналогом производной во внутреннем направлении (внутренний градиент).

Учитывая вышесказанное, в работе было решено рассматривать и разностные аналоги вторых производных функций температур, то есть функции вида:

$$\nabla(t_0 - f(t_0, \dots, t_n)) = t_{0,MW} - f(t_{1,MW}, \dots, t_{n,MW}) - t_{0,IR} + f(t_{1,IR}, \dots, t_{n,IR}).$$

Перейдем к формальному описанию задачи поиска информативных признаков. Введем следующие обозначения.

Пусть $t_i^j, i = 0, \dots, 9$ — температуры i -й точки правой молочной железы j -й пациентки в диапазоне РТМ;

$t_i^j, i = 10, \dots, 19$ — температура $[i - 10]$ точки правой молочной железы j -й пациентки в диапазоне ИК;

$t_i^j, i = 20, \dots, 29$ — температура $[i - 20]$ точки левой молочной железы j -й пациентки в диапазоне РТМ;

$t_i^j, i = 30, \dots, 39$ — температура $[i - 30]$ точки левой молочной железы j -й пациентки в диапазоне ИК;

$t_{40}^j = T1, t_{41}^j = T2$ — глубинные температуры опорных точек j -й пациентки;

$t_{42}^j = T1, t_{43}^j = T2$ — кожные температуры опорных точек j -й пациентки.

Тогда обучающую выборку можно представить в виде матрицы

$$S = \begin{pmatrix} t_0^1 & \dots & t_{43}^1 \\ \vdots & \ddots & \vdots \\ t_0^k & \dots & t_{43}^k \\ t_0^{k+1} & \dots & t_{43}^{k+1} \\ \vdots & \ddots & \vdots \\ t_0^n & \dots & t_{43}^n \end{pmatrix},$$

где t_i^j при $j = 1, \dots, k$ — температуры молочных желез пациенток класса «Не рак», t_i^j при $j = k + 1, \dots, n$ — температуры молочных желез пациенток класса «Рак».

Пусть f^q — q -я исследуемая функция, рассматриваемая на множестве векторов $\Omega_q = (t_0^j, \dots, t_{43}^j)$, где $j = 1, \dots, n$. Обозначим $f_j^q = f^q(t_0^j, \dots, t_{43}^j)$.

Высокоинформативным признаком будем называть тройку (f^q, V, X) , где f^q — функция, описывающая поведение температурных полей; $V = I(f^q, X)$ — информативность признака; X — «информативная» область множества значений функции f^q .

Определим указанные выше характеристики. Под информативностью традиционно понимается количественный параметр, определяющий, насколько хорошо закономерность описывает различия между искомой и отделяемой группами. В качестве $I(f^q, X)$ в данной работе использовался ряд характеристик, в том числе следующие функционалы.

Статистическая информативность, вычисляемая по формуле

$$ST(f^q, x) = -\ln\left(\frac{C_k^h C_{n-k}^s}{C_{k+(n-k)}^{h+s}}\right),$$

где n — количество векторов с температурными данными пациенток в обучающей выборке; k — количество векторов с температурными данными пациенток класса «Не рак» в обучающей выборке; h — количество молочных желез класса «Не рак», для которых $f^q \in X$, а s — количество молочных желез класса «Рак», для которых $f^q \in X$.

Эвристическая информативность, вычисляемая по формуле

$$HR(f^q, x) = \frac{\max\left(\frac{k}{h}, \frac{n-k}{s}\right)}{\min\left(\frac{k}{h}, \frac{n-k}{s}\right)}.$$

Энтропийная информативность, вычисляемая по формуле

$$EN(f^q, x) = H\left(\frac{k}{k+(n-k)}, \frac{n-k}{k+(n-k)}\right) - \frac{k+(n-k)-h-s}{k+(n-k)} H\left(\frac{k-h}{k+(n-k)-h-s}, \frac{(n-k)-s}{k+(n-k)-h-s}\right),$$

где $H(q_0, q_1) = -q_0 \log_2(q_0) - q_1 \log_2(q_1)$ — математическое ожидание количества информации.

Комбинированная информативность, которая вычисляется по формуле

$$CI(f^q, X) = \sqrt[n]{(I_1(f^q, X)I_2(f^q, X) \dots I_n(f^q, X))},$$

где I_1, I_2, \dots, I_n — другие информативности.

Среднее гармоническое информативностей, которое вычисляется по формуле

$$HM(f^q, X) = \frac{n}{\frac{1}{I_1(f^q, X)} + \dots + \frac{1}{I_n(f^q, X)}}.$$

Далее пусть

$$I_s = \sup_{X_\alpha} I(f^q, X),$$

где X_α — все возможные подобласти числовой прямой, на которых указанная информативность имеет смысл. «Информативной» областью множества значений функции f^q

будем называть подмножество числовой прямой, обеспечивающее достижение величины I_g . Для j -х молочных желез признак (f_j^g, V, X) считается выполненным, если $f_j^g \in X$.

Оценка эффективности алгоритмов классификации, в частности и диагностических методов в медицине, традиционно выполняется при помощи следующих показателей.

Чувствительность (Sensitivity), также известная как полнота (Recall), вычисляется по формуле

$$Sens = \frac{TP}{TP + FN},$$

где TP — количество молочных желез класса «Рак», которые были классифицированы как «Рак»; FN — количество молочных желез класса «Рак», которые не были классифицированы как «Рак».

Специфичность (Specificity) вычисляется по формуле

$$Spec = \frac{TN}{TN + FP},$$

где TN — количество молочных желез класса «Не рак», которые не были классифицированы как «Рак»; FP — количество молочных желез класса «Не рак», которые были классифицированы как «Рак».

Точность (Precision) вычисляется по формуле

$$Prec = \frac{TP}{TP + FP}.$$

Достоверность (Accuracy) вычисляется по формуле

$$Acc = \frac{TP + TN}{TP + FP + FN + TN}.$$

Некоторые примеры высокоинформативных признаков приведены в таблице 1. В данном случае при поиске использовался показатель комбинированной информативности статистической и эвристической информативностей.

Отметим, что применяя более сложные конструкции из данных признаков, например, на основе генетических алгоритмов [1], можно получить признаки, обладающие большей информативностью, специфичностью и чувствительностью. Однако резко осложняется возможность их применения в блоке обоснования предлагаемого диагностического решения.

3. Алгоритм классификации

Полученные наборы высокоинформативных признаков в первую очередь предназначены для обнаружения аномалий поведения температурных полей и обоснования предлагаемого диагноза в консультативной интеллектуальной системе. При этом возможно их использование в различных алгоритмах классификации. Для иллюстрации проверим данную возможность на алгоритме логистической регрессии.

В качестве метрики эффективности классификации чаще всего используются различные варианты F -мер, в частности — среднее гармоническое между точностью (Prec) и полнотой, иначе называемой чувствительностью (Sens):

$$F_1 = 2 \frac{Prec \cdot Sens}{Prec + Sens}.$$

Таблица 1

Высокоинформативные признаки

Признак	Диапазон	Sens	Spec	Acc
$\nabla(t_0 - \frac{t_1+t_2+t_3}{3}), 16, 66, (-\infty, -0, 89)$	—	0,13	0,99	0,3
$\nabla(\max_{i=1,8}(t_0 - \frac{\sum_{j=1}^2 t_{((i+j) \bmod 8)+1}}{2})), 15, 43, (-\infty, -0, 6)$	—	0,1	0,99	0,28
$\nabla(t_0 - \frac{t_2+t_3}{2}), 15, 18, (-\infty, -0, 94)$	—	0,12	0,99	0,29
$t_{1,r} - t_{1,l}, 15, 05, (-\infty, -0, 9)$	ИК	0,14	0,99	0,31
$\nabla(t_0 - t_9), 14, 7, (-\infty, -1, 49)$	—	0,11	0,99	0,29
$t_0 - t_6, 13, 72, (1, 4, \infty)$	РТМ	0,11	0,99	0,28
$\sum_{i=1}^8 \frac{t_i}{8} - \frac{t_4+t_5+t_6+t_7}{4}, 13, 48, (0, 3, \infty)$	РТМ	0,1	0,99	0,28
$\sum_{i=1}^8 \frac{t_i}{8} - \frac{t_8+t_1+t_2+t_3}{4}, 13, 48, (-\infty, -0, 31)$	РТМ	0,1	0,99	0,28
$\frac{t_4+t_5+t_6+t_7}{4} - \frac{t_8+t_1+t_2+t_3}{4}, 13, 48, (-\infty, -0, 62)$	РТМ	0,1	0,99	0,28
$t_0 - \frac{t_4+t_5+t_6}{3}, 13, 24, (1, 13, \infty)$	РТМ	0,1	0,99	0,28
$t_0 - \frac{t_4+t_5+t_6+t_7}{4}, 13, 24, (1, 25, \infty)$	РТМ	0,1	0,99	0,28
$\frac{t_0+t_8+t_1+t_2+t_3+t_4}{6} - \frac{t_5+t_6+t_7}{3}, 13, 24, (0, 72, \infty)$	РТМ	0,1	0,99	0,28
$\sum_{i=1}^8 \frac{t_i}{8} - \frac{t_5+t_6+t_7}{3}, 13, 0, (0, 45, \infty)$	РТМ	0,1	0,99	0,28
$\frac{t_5+t_6+t_7}{3} - \frac{t_8+t_1+t_2+t_3+t_4}{5}, 13, 0, (-\infty, -0, 73)$	РТМ	0,1	0,99	0,28
$\sum_{i=1}^8 \frac{t_i}{8} - \frac{t_8+t_1+t_2+t_3+t_4}{5}, 12, 76, (-\infty, -0, 28)$	РТМ	0,09	0,99	0,27
$\nabla(t_0 - \frac{t_2+t_3+t_4}{3}), 12, 76, (-\infty, -0, 93)$	—	0,09	0,99	0,27
$\sqrt{\sum_{i=0}^8 \frac{((t_{i,r}-t_{i,l})-\sum_{j=0}^8 \frac{t_{j,r}-t_{j,l}}{9})^2}{9}}, 12, 27, (0, 59, \infty)$	РТМ	0,14	0,99	0,31

В процессе исследования были проведены вычислительные эксперименты с различным выбором информативности, различными наборами высокоинформативных признаков, различными методами формирования обучающей и тестовой выборок. Ниже приведены результаты расчетов для случая, когда обучающая выборка составляла 62 %, а тестовая — 38 % от всей имеющейся базы данных.

Таким образом, в лучших на тестовой выборке чувствительность и специфичность получена в районе 70 %. Заметим, что в данном случае при классификации не использовались результаты анамнеза и другие медицинские данные, традиционно применяемые при диагностике рака молочных желез.

Заключение

В работе предложен новый метод формирования информационных признаков диагностики рака молочной железы на базе количественного описания качественных признаков, получаемых с помощью анализа медицинских знаний и физико-математических моделей поведения температурных полей молочных желез. Отметим, что данный подход позволил получить новую медицинскую информацию об особенностях поведения температурных полей пациенток. А именно, с помощью исследования разностных аналогов вторых производных функции температур по различным направлениям была обнаружена целая группа достаточно эффективных, качественно новых диагностических признаков. Аналогичный способ обнаружения информативных признаков, предназначенных для недифференциальной диагностики заболеваний молочных желез, был применен в [6]. Однако в данном случае метод был изменен, т. к. некоторые используемые в [6] приемы и алгоритмы для диагностики рака молочных желез оказались неэффективными.

Результаты работы алгоритма логистической регрессии

Информативность	Выборка	<i>Sens</i>	<i>Spec</i>	<i>Acc</i>	F_1
Комбинированная эвристической и энтропийной	Обучающая	0,67	0,72	0,7	0,67
	Тестовая	0,67	0,69	0,68	0,64
Комбинированная статистической, эвристической и энтропийной	Обучающая	0,66	0,71	0,69	0,66
	Тестовая	0,66	0,69	0,68	0,64
Энтропийная	Обучающая	0,67	0,67	0,67	0,65
	Тестовая	0,67	0,64	0,65	0,62
Статистическая	Обучающая	0,66	0,68	0,67	0,65
	Тестовая	0,66	0,66	0,66	0,62
Среднее гармоническое статистической и энтропийной	Обучающая	0,68	0,62	0,65	0,64
	Тестовая	0,68	0,64	0,66	0,63
Комбинированная статистической и энтропийной	Обучающая	0,68	0,62	0,65	0,64
	Тестовая	0,68	0,64	0,66	0,63
Комбинированная статистической и эвристической	Обучающая	0,67	0,75	0,72	0,68
	Тестовая	0,57	0,67	0,62	0,56
Среднее гармоническое эвристической и энтропийной	Обучающая	0,39	0,9	0,67	0,52
	Тестовая	0,45	0,88	0,69	0,56
Среднее гармоническое статистической, эвристической и энтропийной	Обучающая	0,4	0,9	0,67	0,53
	Тестовая	0,45	0,86	0,69	0,55
Среднее гармоническое статистической и эвристической	Обучающая	0,51	0,92	0,73	0,63
	Тестовая	0,39	0,82	0,64	0,48
Эвристическая	Обучающая	0,4	0,95	0,7	0,54
	Тестовая	0,34	0,88	0,65	0,45

Применение на базе полученных признаков более мощных алгоритмов классификации улучшает результаты диагностики. Например, применение нейронных сетей каскадной корреляции с методом имитации отжига в качестве обучения [1] позволяет получить чувствительность и специфичность в районе 80%. Более того, в разрабатываемой консультационной системе и предполагается применять алгоритмы классификации на базе нейронных сетей, генетических алгоритмов, нечеткой логики и т. д. Однако обоснование получаемых результатов будет производиться на основе полученных признаков. Отдельно отметим, что реальная чувствительность и специфичность консультационной системы должна выявляться путем опытной эксплуатации, то есть более интересным является вопрос, насколько полученные признаки улучшат диагностику, осуществляемую специалистами.

ПРИМЕЧАНИЕ

¹ Работа выполнена при финансовой поддержке РФФИ и Администрации Волгоградской области (проект № 15-47-02475-р_поволжье_а).

СПИСОК ЛИТЕРАТУРЫ

1. Алгоритмы принятия решений в консультативной интеллектуальной системе диагностики молочных желез / А. В. Зенович, В. А. Глазунов, А. С. Опарин, Ф. Г. Примаченко // Вестник Волгоградского государственного университета. Серия 1, Математика. Физи-

- ка. — 2016. — № 6 (37). — С. 129–142. — DOI: <http://dx.doi.org/10.15688/jvolsu1.2016.6.13>.
2. Вайсблат, А. В. Использование микроволновой радиотермометрии в диагностике рака молочной железы / А. В. Вайсблат, С. Г. Веснин, М. А. Конкин. — Электрон. текстовые дан. — Режим доступа: <http://www.resltd.ru/rus/literature/cancer.htm>. — Загл. с экрана.
 3. Веснин, С. Г. Современная микроволновая радиотермометрия молочных желез / С. Г. Веснин, М. А. Каплан, Р. С. Авакян // Опухоли женской репродуктивной системы. — 2008. — № 3. — С. 28–33.
 4. Кобринский, Б. А. Консультативные интеллектуальные медицинские системы: классификация, принципы построения, эффективность / Б. А. Кобринский // Врач и информационные технологии. — 2008. — № 2. — С. 38–47.
 5. Кобринский, Б. А. Системы искусственного интеллекта в медицине: состояние, проблемы и перспективы / Б. А. Кобринский // Новости искусственного интеллекта. — 1995. — № 2. — С. 65–79.
 6. Лосев, А. Г. Интеллектуальный анализ термометрических данных в диагностике молочных желез / А. Г. Лосев, В. В. Левшинский // Управление большими системами. — 2017. — Вып. 70. — С. 113–135.
 7. Лосев, А. Г. О взаимосвязи некоторых признаков РТМ-диагностики заболеваний молочных желез / А. Г. Лосев, Е. А. Мазепа, Х. М. Сулейманова // Вестник Волгоградского государственного университета. Серия 1, Математика. Физика. — 2015. — № 4 (29). — С. 35–44. — DOI: <http://dx.doi.org/10.15688/jvolsu1.2015.4.4>.
 8. Лосев, А. Г. О некоторых характерных признаках в диагностике патологии молочных желез по данным микроволновой радиотермометрии / А. Г. Лосев, Е. А. Мазепа, Т. В. Замечник // Современные проблемы науки и образования. — 2014. — № 6. — Article ID: 16587.
 9. Проблемы измерения и моделирования тепловых и радиационных полей в биотканях: анализ данных микроволновой радиотермометрии / А. Г. Лосев, А. В. Хоперсков, А. С. Астахов, Х. М. Сулейманова // Вестник Волгоградского государственного университета. Серия 1, Математика. Физика. — 2015. — № 6. — С. 31–71. — DOI: <http://dx.doi.org/10.15688/jvolsu1.2015.6.3>.
 10. Поляков, М. В. Математическое моделирование пространственного распределения радиационного поля в биоткани: определение яркостной температуры для диагностики / М. В. Поляков, А. В. Хоперсков // Вестник Волгоградского государственного университета. Серия 1, Математика. Физика. — 2016. — № 5 (36). — С. 73–84. — DOI: <http://dx.doi.org/10.15688/jvolsu1.2016.5.7>.
 11. Ясницкий, Л. Н. Введение в искусственный интеллект : учебн. пособие для студ. вузов / Л. Н. Ясницкий. — М. : Academia, 2005. — 176 с.
 12. Burke, H. B. Artificial neural networks for cancer research: outcome prediction / H. B. Burke // Seminars in Surgical Oncology. — 1994. — Vol. 10, № 1. — P. 73–79.
 13. Kononenko, I. Application of machine learning to medical diagnosis. / I. Kononenko, I. Bratko, M. Kukar // Machine Learning, Data Mining and Knowledge Discovery: Methods and Applications. — New York : John Wiley & Sons, 1998. — P. 389–408.
 14. Kononenko, I. Machine Learning for Medical Diagnosis: History, State of the Art and Perspective / I. Kononenko // Artificial Intelligence in Medicine. — 2001. — Vol. 23, № 1. — P. 89–109. — DOI: [http://doi.org/10.1016/S0933-3657\(01\)00077-X](http://doi.org/10.1016/S0933-3657(01)00077-X).
 15. Lesmo, L. Learning of Fuzzy Production Rules for Medical Diagnoses / L. Lesmo, L. Saitta, P. Torasso // Approximate Reasoning in Decision Analysis. — 1982. — P. 249–260.

REFERENCES

1. Zenovich A.V., Glazunov V.A., Oparin A.S., Primachenko F.G. Algoritmy prinyatiya resheniy v konsultativnoy intellektualnoy sisteme diagnostiki molochnykh zhelez [Algorithms of Decision-Making in Intelligent Advisory System for Diagnostics of the Mammary Glands]. *Vestnik Volgogradskogo gosudarstvennogo universiteta. Seriya 1, Matematika. Fizika* [Science Journal of Volgograd State University. Mathematics. Physics], 2016, no. 6 (37), pp. 129–142. DOI: <http://dx.doi.org/10.15688/jvolsu1.2016.6.13>.

2. Vaysblat A.V., Vesnin S.G., Konkin M.A. Ispolzovanie mikrovolnovoy radiotermometrii v diagnostike raka molochnoy zhelezy [The Use of Microwave Radiothermometry in the Diagnosis of Breast Cancer]. URL: <http://www.resltd.ru/rus/literature/cancer.htm>.
3. Vesnin S.G., Kaplan M.A., Avakyan R.S. Sovremennaya mikrovolnovaya radiotermometriya molochnykh zhelez [Modern Microwave Radiothermometry of Mammary Glands]. *Opukholi zhenskoy reproduktivnoy sistemy*, 2008, no. 3, pp. 28-33.
4. Kobrinskiy B.A. Konsultativnye intellektualnye meditsinskie sistemy: klassifikatsiya, printsipy postroeniya, effektivnost [Advisory Intelligent Medical Systems: Classification, Principles of Construction, Efficiency]. *Vrach i informatsionnye tekhnologii*, 2008, no. 2, pp. 38-47.
5. Kobrinskiy B.A. Sistemy iskusstvennogo intellekta v meditsine: sostoyanie, problemy i perspektivy [Artificial Intelligence Systems in Medicine: State, Problems and Prospects]. *Novosti iskusstvennogo intellekta*, 1995, no. 2, pp. 65-79.
6. Losev A.G., Levshinskiy V.V. Intellektualnyy analiz termometricheskikh dannykh v diagnostike molochnykh zhelez [The Thermometry Data Mining in the Diagnostics of Mammary Glands]. *Upravlenie bolshimi sistemami*, 2017, iss. 70, pp. 113-135.
7. Losev A.G., Mazepa E.A., Suleymanova Kh.M. O vzaimosvyazi nekotorykh priznakov RTM-diagnostiki zabolevaniy molochnykh zhelez [On Interrelation of Some Signs of Rtm Diagnostics of Mammary Glands Diseases]. *Vestnik Volgogradskogo gosudarstvennogo universiteta. Seriya 1, Matematika. Fizika* [Science Journal of Volgograd State University. Mathematics. Physics], 2015, no. 4 (29), pp. 35-44. DOI: <http://dx.doi.org/10.15688/jvolsu1.2015.4.4>.
8. Losev A.G., Mazepa E.A., Zamechnik T.V. O nekotorykh kharakternykh priznakakh v diagnostike patologii molochnykh zhelez po dannym mikrovolnovoy radiotermometrii [About Several Typical Traits in the Diagnosis of Mammary Glands Pathology According to the Data of Microwave Radiothermometry]. *Sovremennye problemy nauki i obrazovaniya*, 2014, no. 6, article ID: 16587.
9. Losev A.G., Khoperskov A.V., Astakhov A.S., Suleymanova Kh.M. Problemy izmereniya i modelirovaniya teplovykh i radiatsionnykh poley v biotkanyakh: analiz dannykh mikrovolnovoy radiotermometrii [Problems of Measurement and Modeling of Thermal and Radiation Fields in Biological Tissues: Analysis of Microwave Thermometry Data]. *Vestnik Volgogradskogo gosudarstvennogo universiteta. Seriya 1, Matematika. Fizika* [Science Journal of Volgograd State University. Mathematics. Physics], 2015, no. 6, pp. 31-71. DOI: <http://dx.doi.org/10.15688/jvolsu1.2015.6.3>.
10. Polyakov M.V., Khoperskov A.V. Matematicheskoe modelirovanie prostranstvennogo raspredeleniya radiatsionnogo polya v biotkani: opredelenie yarkostnoy temperatury dlya diagnostiki [Mathematical Modeling of Radiation Fields in Biological Tissues: the Definition of the Brightness Temperature for the Diagnosis]. *Vestnik Volgogradskogo gosudarstvennogo universiteta. Seriya 1, Matematika. Fizika* [Science Journal of Volgograd State University. Mathematics. Physics], 2016, no. 5 (36), pp. 73-84. DOI: <http://dx.doi.org/10.15688/jvolsu1.2016.5.7>.
11. Yasnitskiy L.N. *Vvedenie v iskusstvennyy intellekt: uchebn. posobie dlya stud. vuzov* [Introduction to Artificial Intelligence: a Tutorial for University Students]. Moscow, Academia Publ., 2005. 176 p.
12. Burke H.B. Artificial Neural Networks for Cancer Research: Outcome Prediction. *Seminars in Surgical Oncology*, 1994, vol. 10, no. 1, pp. 73-79.
13. Kononenko I., Bratko I., Kukar M. Application of machine learning to medical diagnosis. *Machine Learning, Data Mining and Knowledge Discovery: Methods and Applications*. New York, John Wiley & Sons, 1998, pp. 389-408.
14. Kononenko I. Machine Learning for Medical Diagnosis: History, State of the Art and Perspective. *Artificial Intelligence in Medicine*, 2001, vol. 23, no. 1, pp. 89-109. DOI: [http://doi.org/10.1016/S0933-3657\(01\)00077-X](http://doi.org/10.1016/S0933-3657(01)00077-X).
15. Lesmo L., Saitta L., Torasso P. Learning of Fuzzy Production Rules for Medical Diagnoses. *Approximate Reasoning in Decision Analysis*, 1982, pp. 249-260.

DATA MINING OF MICROWAVE RADIOMETRY DATA IN THE DIAGNOSIS OF BREAST CANCER

Alexander Georgievich Losev

Doctor of Physical and Mathematical Sciences, Professor, Department of Mathematical Analysis and Function Theory,
Volgograd State University
alexander.losev@volsu.ru
Prosp. Universitetsky, 100, 400062 Volgograd, Russian Federation

Vladislav Viktorovich Levshinskiy

Student,
Volgograd State University
vladi.lev.email@gmail.com
Prosp. Universitetsky, 100, 400062 Volgograd, Russian Federation

Abstract. The paper deals with development of a new method for the thermometry data mining, which is designed as a basis for creation of the intelligent diagnostic-advisory systems. The method is based on quantitative descriptions of medical knowledge about the behavior of the temperature fields of the mammary glands.

Previously gained knowledge has been represented as a base set of grouped functions of temperatures, e.g. qualitative feature “a large temperature difference between nipple and a certain point of mammary gland” can be described by a set of functions of the form $|t_0 - t_i|, i = 1, \dots, 9$, where t_i is the temperature of the point number i . According to the qualitative feature, a large value of such functions may be an indication of pathology. Hence, we will bind a range of unacceptable values with certain functions and denote produced pairs as features. Thus, on the base of features we can determine possible anomalies of the temperature fields of mammary glands. It's clear that some features probably are false features, so a set of features requires preprocessing. One way to do this is to assign a weight to a feature and analyze it only if the weight is larger than some minimum acceptable value. Weighted features were applied in a logistic regression algorithm.

With the features gained, the average breast cancer diagnosis accuracy on the test sample is about 67 percent. Moreover, each received feature has a qualitative description and can be applied in more complex algorithms.

Key words: data mining, microwave radiothermometry, intelligent advisory systems, highly informative signs, breast cancer.