

УДК 519.23  
ББК 2.2.22.172

## **ИНТЕЛЛЕКТУАЛЬНЫЙ АНАЛИЗ ТЕРМОМЕТРИЧЕСКИХ ДАННЫХ В ДИАГНОСТИКЕ МОЛОЧНЫХ ЖЕЛЕЗ**

**Лосев А. Г.<sup>1</sup>, Левшинский В. В.<sup>2</sup>**

*(Волгоградский государственный университет, Волгоград)*

*Работа посвящена разработке нового метода интеллектуального анализа медицинских термометрических данных, предназначенного для создания на его основе консультативных интеллектуальных диагностических систем. Рассмотрены подходы к анализу данных микроволновой радиотермометрии, полученных для диагностики заболеваний молочной железы. Представлен метод формирования информационных признаков на базе количественного описания медицинских знаний о поведении температурных полей молочных желез.*

Ключевые слова: интеллектуальный анализ данных, микроволновая радиотермометрия, консультативные интеллектуальные системы.

### **1. Введение**

Разработка диагностических технологий, основанных на интеграции современных достижений в инженерии, информационных технологиях, медицине и математических методах искусственного интеллекта, является крайне актуальной задачей. В первую очередь это связано с необходимостью создания новых методик функциональной диагностики, основанных на динамическом картировании физических полей и излучений

---

<sup>1</sup> Александр Георгиевич Лосев, доктор физико-математических наук, профессор (alexander.losev@volsu.ru).

<sup>2</sup> Владислав Викторович Левшинский, студент (vladi.lev.email@gmail.com).

организма человека. Однако применение современного медицинского оборудования в процессе решения одних проблем зачастую 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 обеспечивается созданием систем интерпретации и анализа медицинских данных. При этом наибольший интерес вызывает разработка консультативных интеллектуальных систем, т.е. экспертных систем, содержащих механизм объяснения и обоснования предлагаемых решений на языке, понятном пользователю [5].

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

## **2. Обзор литературы**

Первые попытки обоснования возможности применения радиотермометрии для диагностики рака молочной железы были предприняты в середине 70-х годов 20 века [11, 12, 14]. В указанных работах были предложены первые математические модели, описывающие температурные поля молочных желез в терминах решений уравнений с частными производными. Дальнейшие исследования показали, что в подобных моделях следует учитывать физико-химические свойства биотканей [21], выбор реологической модели крови [15], сложное внутреннее строение молочных желез и ряд других параметров [9]. Однако, учитывая достаточно широкий разброс даже основных параметров указанных моделей, их применение в диагностических системах пока не представляется возможным.

Одним из наиболее популярных средств повышения эффективности диагностики является автоматизация обработки данных с использованием методов искусственного интеллекта. Наиболее популярным здесь является применение искусственных нейронных сетей [10, 13], байесовского классификатора [19], генетических алгоритмов [16], алгоритмов символьного обучения [18], нечеткой логики [20].

В настоящее время большинство экспертных систем предлагает врачу свои решения либо в детерминированной форме однозначного заключения, либо в виде вероятностных оценок каждого из возможных диагнозов. Одновременно идет процесс формирования новых подходов к созданию диагностических систем, «объясняющих» предлагаемые ими решения [4, 5].

### **3. Описание задачи**

Микроволновая радиотермометрия – биофизический метод неинвазивного обследования, заключающийся в измерении внутренних и поверхностных температур тканей по интенсивности их теплового излучения в микроволновом (РТМ) и инфракрасном (ИК) диапазонах соответственно. В течение последнего десятилетия данный метод получил распространение в различных областях медицины. Одновременно возник и ряд проблем. 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 использования термометрической аппаратуры 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 [6], на будущее качество алгоритмов классификации влияют качественный и количественный составы пространства информационных признаков.

Основной задачей данного исследования является разработка метода формирования пространства информационных признаков.

#### 4. 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 начинается с измерения температур в опорных точках  $T1$  и  $T2$ : первая расположена 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 (схема представлена на рис. 1).

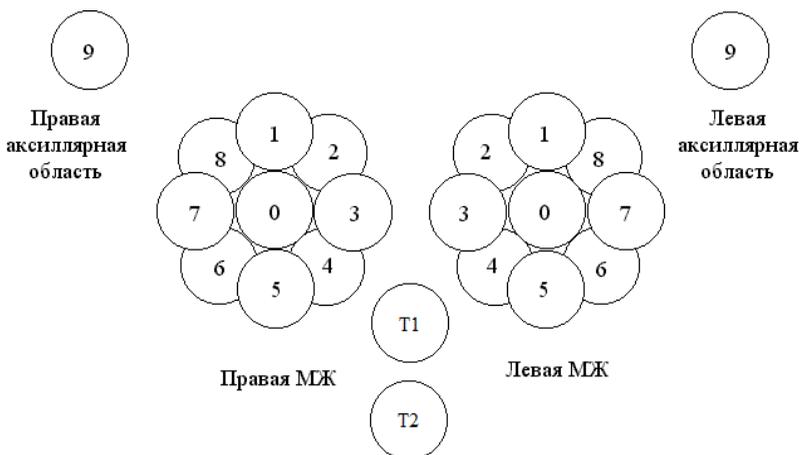


Рис. 1. Схема обследования молочной железы

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. Статистический анализ используемых термометрических данных был проведен в [8].

В том числе доказано, что существующий объем выборки является достаточным для проведения исследований и анализа данных.

В ходе исследований и анализа данных специалистами были выявлены следующие признаки рака молочной железы [1, 2, 7] (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, а также выявление их количественных характеристик.

На первом этапе, опираясь на существующие медицинские знания, а также модели поведения температурных полей, попытаемся классифицировать существующие термометрические диагностические признаки, а также дополнить их.

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

Заметим, что 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  $|t_{i,пр} - t_{i,л}|$  или  $(t_{i,пр} - t_{i,л})$ , где

$t_{i,пр}$  и  $t_{i,л}$  – температуры в  $i$ -х точках правой и левой молочных желез соответственно.

Отметим, что в данном признаке медики традиционно используют  $i = 1, \dots, 8$ , выделяя разность температур сосков в отдельный признак.

1.2. Повышенная разница температур сосков правой и левой молочных желез:  $|t_{0,пр} - t_{0,л}|$ , или  $(t_{0,пр} - t_{0,л})$ .

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

$$(1) \sqrt{\sum_{i=0}^8 \frac{(t_{i,пр} - t_{i,л})^2}{9}}.$$

На самом деле асимметрия полей температур молочных желез может быть описана различными функциями вида

$$(2) g(f(t_{0,пр}, \dots, t_{n,пр}) - f(t_{0,л}, \dots, t_{n,л})),$$

где  $f(t_{0,пр}, \dots, t_{n,пр})$  – значение функции температур для точек  $t_0, \dots, t_n$  правой молочной железы,  $f(t_{0,л}, \dots, t_{n,л})$  – 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:

$$(3) \left| \frac{t_{0,пр} + t_{i,пр} + t_{i(\text{mod}8)+1,пр}}{3} - \frac{t_{0,л} + t_{i,л} + t_{i(\text{mod}8)+1,л}}{3} \right|,$$

где  $i = 1, \dots, 8$ .

1.5. Разница среднеквадратичных отклонений температур молочных желез:

$$(4) \left| \left( \sqrt{\sum_{i=1}^8 \frac{(t_i - t_{cp})^2}{8}} \right)_{пр} - \left( \sqrt{\sum_{i=1}^8 \frac{(t_i - t_{cp})^2}{8}} \right)_{л} \right|,$$

где

$$(5) \quad t_{\text{cp}} = \sum_{i=1}^8 \frac{t_i}{8}.$$

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

2.1. Повышенное среднеквадратичное отклонение температур в одной из молочных желез:

$$(6) \quad \sqrt{\sum_{i=0}^8 \frac{(t_i - \overline{t_{\text{cp}}})^2}{9}},$$

где

$$(7) \quad \overline{t_{\text{cp}}} = \sum_{i=0}^8 \frac{t_i}{9}.$$

2.2. Повышенный разброс температур между отдельными точками в пораженной молочной железе:  $t_{\text{cp}} - t_i$ , где  $t_i$  – температура в  $i$ -й точке молочной железы.

На самом деле разброс температур молочных желез может быть описан различными функциями вида

$$(8) \quad g(f_1(t_0, \dots, t_n) - f_2(t_0, \dots, t_n)),$$

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, т.е. функции вида:

$$(9) \quad t_{\text{cp}} - \frac{t_i + t_{i(\text{mod}8)+1} + t_{i(\text{mod}8)+2}}{3}.$$

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

3.1. Аномальная разность температуры соска и средней температуры молочной железы:  $t_0 - t_{\text{cp}}$ .

3.2. Аномальная разность температуры соска и температур отдельных точек молочной железы:  $t_0 - t_i$ ,  $i = 1, \dots, 8$ .

На самом деле аномальные значения температуры соска по отношению к другим параметрам можно описать функциями вида

$$(10) \ g(t_0 - f_1(t_1, \dots, t_n)),$$

где  $f_1(t_1, \dots, t_n)$  – функция температур точек  $t_1, \dots, t_n$  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:

$$(11) \ t_0 - \frac{t_i + t_{i(\text{mod}8)+1}}{2}.$$

4. Группа признаков, характеризующих соотношение кожной и глубинной температур:

4.1. Аномальное значение разности между кожной и глубинной температурами точки пораженной молочной железы (внутренний градиент):  $(t_{i,\text{РТМ}} - t_{i,\text{ИК}})$ , где  $t_{i,\text{РТМ}}$  – глубинные и  $t_{i,\text{ИК}}$  – 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

$$(12) \ g(f(t_{0,\text{РТМ}}, \dots, t_{n,\text{РТМ}}) - f(t_{0,\text{ИК}}, \dots, t_{n,\text{ИК}})),$$

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.

Как было отмечено выше, за последние годы было построено несколько математических моделей, описывающих поведение температурных полей молочных желез с помощью уравнений в частных производных второго порядка [9]. Таким образом, для обнаружения аномалий температурных полей пациенток может оказаться интересным изучение поведения разностных аналогов вторых производных функции температур.

Заметим, что в предъявленном 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  $(t_{i,р\text{тм}} - t_{i,и\text{к}})$  является разностным аналогом производной во внутреннем направлении (внутренний градиент). Учитывая вышесказанное, в работе было решено рассматривать и разностные аналоги вторых производных функций температур, т.е. функции вида:

$$(13) \quad \nabla(t_0 - f(t_1, \dots, t_n)) = \\ = t_{0,р\text{тм}} - f(t_{1,р\text{тм}}, \dots, t_{n,р\text{тм}}) - (t_{0,и\text{к}} - f(t_{1,и\text{к}}, \dots, t_{n,и\text{к}}))$$

Были предложены и иные математические описания возможных аномалий поведения температурных полей. Всего при исследовании рассмотрено порядка 900 числовых функций.

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

$$(14) S = \begin{pmatrix} t_0^1 & \cdots & t_{43}^1 \\ \vdots & \ddots & \vdots \\ t_0^k & \cdots & t_{43}^k \\ t_0^{k+1} & \cdots & t_{43}^{k+1} \\ \vdots & \ddots & \vdots \\ t_0^n & \cdots & t_{43}^n \end{pmatrix},$$

где  $t_i^j$  при  $j = 1, \dots, k$  – температуры молочных желез пациенток класса «Здоровые»,  $t_i^j$  при  $j = k + 1, \dots, n$  – температуры молочных желез пациенток класса «Больные».

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  $f_j^q = f^q(t_0^j, \dots, t_{43}^j)$ .

Пусть  $X_\alpha$  – некоторое множество числовой прямой  $R$ , такое что  $f_j^q \in X_\alpha$  для некоторых  $j = k + 1, \dots, n$ , и  $f_j^q \in R \setminus X_\alpha$  для всех  $j = 1, \dots, k$ . Тогда «характерной» областью множества значений  $f^q$  будем называть такое подмножество числовой прямой  $X$ , что  $X_\alpha \subset X$  для всех возможных значений индекса  $\alpha$ , но при этом  $f_j^q \in R \setminus X$  для всех  $j = 1, \dots, k$ .

Характеристическим признаком будем называть пару  $(f^q, X)$ , где  $X$  – «характерная» область множества значений  $f^q$ .

Алгоритм поиска характеристических признаков заключается в нахождении «максимальных» подобластей, которым принадлежат значения функции, соответствующие молочным железам одного класса, но не принадлежат значения функции, соответствующие молочным железам другого. Отметим также, что зачастую удобнее вместо всей «характерной» области множества значений  $f^q$  рассматривать ее связанные подмножества.

Для оценки эффективности алгоритмов классификации, в частности и диагностических методов в медицине, традиционно используются следующие показатели.

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

$$(15) Sens = \frac{TP}{TP + FN},$$

где  $TP$  – количество молочных желез класса «Больные», которые были классифицированы как «Больные»,  $FN$  – количество молочных желез класса «Больные», которые не были классифицированы как «Больные».

Специфичность (Specificity), вычисляется по формуле

$$(16) \text{Spec} = \frac{TN}{TN + FP},$$

где  $TN$  – количество молочных желез класса «Здоровые», которые не были классифицированы как «Больные»,  $FP$  – количество молочных желез класса «Здоровые», которые были классифицированы как «Больные».

Точность (Precision), вычисляется по формуле

$$(17) \text{Prec} = \frac{TP}{TP + FP}.$$

Достоверность (Accuracy), вычисляется по формуле

$$(18) \text{Acc} = \frac{TP + TN}{TP + FP + FN + TN}.$$

Также отметим, что в медицине важнейшими характеристиками диагностического метода считают чувствительность и специфичность. Для характеристических признаков, очевидно, выполнено  $\text{Spec} = 1$ , поэтому их эффективность оценивается чувствительностью и достоверностью.

Несколько характеристических признаков с лучшими показателями истинно-положительных результатов диагностики приведены в таблице 1.

Таблица 1. Характеристические признаки

Признак	Sens	Acc
$\nabla \left( t_0 - \frac{t_1 + t_2 + t_3}{3} \right), (-\infty; -1, 0)$	0,08	0,27
$t_{1, \text{np}} - t_{1, \text{л}}, (-\infty; -1, 0)$	0,07	0,26
$\sqrt{\sum_{i=0}^8 \frac{(t_{i, \text{np}} - t_{i, \text{л}})^2}{9}}, (1, 08; \infty)$	0,07	0,26

Во втором и третьем признаках используются значения температур в РТМ-диапазоне.

Очевидно, что характеристические признаки в состоянии обнаружить лишь достаточно серьезные аномалии поведения температурных полей. При использовании набора наиболее значимых характеристических признаков было обнаружено 43% молочных желез класса «Больные».

Далее, основываясь на методике поиска характеристических признаков, попробуем построить набор высокоинформативных признаков.

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 понимается количественный параметр, определяющий, насколько хорошо закономерность описывает различия между искомой и отделяемой группами. В качестве  $I(f^q, X)$  в данной работе использовались следующие характеристики.

Статистическая информативность, вычисляемая по формуле

$$(19) \quad ST(f^q, X) = -\ln \left( \frac{C_k^h C_{(n-k)}^s}{C_{k+(n-k)}^{h+s}} \right),$$

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

Эвристическая информативность, вычисляемая по формуле

$$(20) \quad HR(f^q, X) = \frac{\max \left\{ \frac{k}{h}, \frac{n-k}{s} \right\}}{\min \left\{ \frac{k}{h}, \frac{n-k}{s} \right\}}.$$

Энтропийная информативность, вычисляемая по формуле

$$(21) \quad EN(f^q, X) = H\left(\frac{k}{k+(n-k)}, \frac{(n-k)}{k+(n-k)}\right) - \frac{k+(n-k)-h-s}{k+(n-k)} H\left(\frac{k-h}{k+(n-k)-h-s}, \frac{(n-k)-s}{k+(n-k)-h-s}\right),$$

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

$$(22) \quad CI(f^q, X) = \sqrt[n]{I_1(f^q, X) I_2(f^q, X) \dots I_n(f^q, X)},$$

где  $I_1, I_2, \dots, I_n$  – другие информативности;

Среднее гармоническое информативностей, которое вычисляется по формуле

$$(23) \quad HM(f^q, X) = \frac{n}{\frac{1}{I_1(f^q, X)} + \dots + \frac{1}{I_n(f^q, X)}}.$$

Далее, пусть

$$(24) \quad I_s = \sup_{X_\alpha} I(f^q, X_\alpha),$$

где  $X_\alpha$  – 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 (или приближение с заданной точностью) величины  $I_s$ .

Для  $j$ -х молекулярных признаков  $(f_j^q, V, X)$  считается выполненным, если  $f_j^q \in X$ .

Некоторые примеры высокоинформативных признаков приведены в таблице 2. В данном случае при поиске использовался показатель комбинированной информативности статистической и эвристической информативностей.

Во втором признаке используются значения температур в ИК-диапазоне, в четвертом и пятом – в РТМ-диапазоне.

Таблица 2. Высокоинформативные признаки

Признак	Sens	Spec	Acc
$t_{1,np} - t_{1,l}, (-\infty, -0,9)$	0,14	0,99	0,31
$\nabla\left(t_0 - \frac{t_1 + t_2 + t_3}{3}\right), (-\infty, -0,89)$	0,13	0,99	0,30
$\nabla(t_0 - t_9), (-\infty, -1,49)$	0,11	0,99	0,29
$t_0 - t_6, (1,4, \infty)$	0,1	0,99	0,28
$t_{cp} - \frac{t_4 + t_5 + t_6 + t_7}{4}, (0,3, \infty)$	0,1	0,99	0,28

Отметим, что применяя более сложные конструкции из данных признаков, например, на основе генетических алгоритмов [3], можно получить признаки, обладающие большей информативностью, специфичностью и чувствительностью. Однако резко осложняется возможность их применения в блоке обоснования предлагаемого диагностического решения.

## 5. Алгоритм классификации

Полученные наборы характеристических и высокоинформативных признаков в первую очередь предназначены для обозначения найденных аномалий поведения температурных полей и обоснования предполагаемого диагноза в консультативной интеллектуальной системе. Предлагаемая их структура вполне позволяет обнаруживать и описывать особенности поведения температурных полей молочных желез на языке, понятном пользователю системы, т.е. врачу-диагносту. Однако вполне обоснованным выглядит их дальнейшее использование в различных алгоритмах классификации. Для иллюстрации проверим данную возможность на простейших алгоритмах, в которых найденные признаки используются непосредственно (в отличие от более мощных и эффективных алгоритмов типа нейронных сетей или генетических алгоритмов, использующих достаточно сложные конструкции из базовых элементов [3, 9]).

Опишем кратко используемый алгоритм классификации.

На первом этапе проводится проверка выполнения характеристических признаков. При этом если выполнен хотя бы один из них, то молочная железа относится к классу «Больные».

На втором этапе проверяется выполнение высокоинформативных признаков для неклассифицированных на первом этапе молочных желез. При этом для каждой молочной железы считается сумма информативностей выполненных признаков. Если найденная сумма выше некоторого критического значения (будем называть его критерием классификации), то молочная железа относится к классу «Больные», в противном случае – к классу «Здоровые».

Критерий классификации можно определять различными способами. Опишем один из возможных вариантов. Введем следующие обозначения:  $H_m$  – среднее арифметическое множества сумм информативностей выполненных признаков для каждой молочной железы класса «Здоровые»;  $S_m$  – среднее арифметическое множества сумм информативностей выполненных признаков для каждой молочной железы класса «Больные»;  $H_d$  – стандартное отклонение множества сумм информативностей выполненных признаков для каждой молочной железы класса «Здоровые»;  $S_d$  – стандартное отклонение множества сумм информативностей выполненных признаков для каждой молочной железы класса «Больные».

Набору высокоинформативных признаков будем ставить в соответствие критерий классификации:

$$(25) S_0 = \frac{x_1 + x_2}{2},$$

где

$$(26) x_1 = [\min\{H_m, H_d\}, \max\{H_m, H_d\}],$$

$$(27) x_2 = [\min\{S_m, S_d\}, \max\{S_m, S_d\}].$$

Другими словами,  $S_0$  – точка пересечения интервалов, концами которых являются среднее арифметическое и стандартное отклонение сумм информативностей выполненных признаков для каждой молочной железы из определенного класса.

Сформированные наборы высокоинформативных признаков не являются оптимальными для данного алгоритма классификации, поскольку в них содержится большое количество

«родственных» признаков, обозначающих одну и ту же аномалию температурного поля. Поиск оптимального набора высокоинформативных признаков является актуальной проблемой. В данном случае использовалось достаточно простое эвристическое решение, состоящее из двух этапов. На этапе фильтрации из набора исключаются малоинформативные признаки, информативность которых меньше 3.

Для дальнейшей работы понадобилось использование метрики качества классификации. Чаще всего используются различные варианты  $F$ -мер, в частности – среднее гармоническое между точностью ( $Prec$ ) и полнотой, иначе называемой чувствительностью ( $Sens$  в (15)):

$$(28) F_1 = 2 \frac{Prec \cdot Sens}{Prec + Sens}.$$

Учитывая, что в медицинской диагностике наиболее значимыми являются параметры  $Sens$  и  $Spec$ , было решено в качестве метрики качества классификации использовать среднее гармоническое между  $Sens$  и  $Spec$ :

$$(29) D = 2 \frac{Sens \cdot Spec}{Sens + Spec}.$$

Отметим, что аналогичный подход применяется не впервые [17]. Ниже будет приведена сравнительная таблица эффективности диагностики с наборами признаков, сформированными при использовании разных метрик.

Итак, на втором этапе применялась «жадная» схема формирования набора признаков. А именно, в результирующий набор, который изначально был пуст, поочередно помещалось по одному признаку из исходного набора. С каждым новым признаком выполнялась диагностика и вычислялась метрика эффективности классификации  $D$ . Если полученное значение  $D$  было выше предыдущего, то признак запоминался и перемещался в результирующий набор в конце итерации. Результирующий набор сформирован, если в конце итерации не было добавлено нового признака.

В процессе исследования были проведены вычислительные эксперименты с различным выбором информативности, различными наборами высокоинформативных признаков, различными

методами формирования обучающей и тестовой выборок. Затем полученный алгоритм классификации применялся к тестовой выборке. В таблице 3 приведены результаты расчетов для случая, когда обучающая выборка составляла 62%, а тестовая – 38% от всей имеющейся базы данных.

Таблица 3. Результаты работы алгоритма классификации

Информативность	Выборка	<i>Sens</i>	<i>Spec</i>	<i>Acc</i>	<i>D</i>	<i>F</i> <sub>1</sub>
Среднее гармоническое статистической, энтропийной, эвристической	Обучающая	0,87	0,86	0,87	0,86	0,92
	Тестовая	0,84	0,59	0,79	0,69	0,86
Среднее гармоническое статистической, эвристической	Обучающая	0,84	0,78	0,83	0,81	0,89
	Тестовая	0,79	0,52	0,73	0,62	0,82
Эвристическая	Обучающая	0,75	0,79	0,76	0,77	0,83
	Тестовая	0,73	0,53	0,69	0,62	0,79
Среднее гармоническое энтропийной, эвристической	Обучающая	0,87	0,77	0,85	0,82	0,9
	Тестовая	0,82	0,43	0,74	0,57	0,84
Комбинированная статистической, энтропийной	Обучающая	0,77	0,67	0,75	0,72	0,83
	Тестовая	0,79	0,45	0,72	0,57	0,82

Также на базе найденных высокоинформативных признаков была проведена классификация с помощью логистической регрессии. Результаты приведены в таблице 4.

Таким образом, в обоих случаях на тестовой выборке получалась чувствительность в районе 80%, а специфичность, соответственно, в районе 60%. Лучшие критерии эффективности алгоритма получались в районе 65% (при проверке по всей экспертной базе, естественно, результаты существенно улучшаются и становятся выше 90%). Полученные результаты оказались достаточно стабильными, т.е. изменение состава обучающей и тестовой выборок практически не меняло получаемый критерий эффективности алгоритма.

Таблица 4. Результаты работы алгоритма логистической регрессии

Информативность	Выборка	<i>Sens</i>	<i>Spec</i>	<i>Acc</i>	<i>D</i>	$F_1$
Эвристическая	Обучающая	0,82	0,81	0,82	0,81	0,88
	Тестовая	0,73	0,6	0,71	0,66	0,8
Среднее гармоническое статистической, эвристической	Обучающая	0,66	0,83	0,7	0,74	0,78
	Тестовая	0,65	0,59	0,64	0,62	0,74
Комбинированная статистической, эвристической	Обучающая	0,66	0,81	0,69	0,73	0,77
	Тестовая	0,64	0,59	0,63	0,61	0,74
Комбинированная энтропийной, эвристической	Обучающая	0,69	0,89	0,73	0,77	0,8
	Тестовая	0,68	0,55	0,65	0,61	0,76
Среднее гармоническое энтропийной, эвристической	Обучающая	0,83	0,82	0,83	0,83	0,89
	Тестовая	0,73	0,52	0,68	0,6	0,78

Как было отмечено выше, в качестве метрики качества классификации используется целый ряд параметров: различные варианты  $F$ -мер, коэффициент корреляции Мэтьюса, а также самые различные усреднения множественных метрик. При этом наиболее популярным является использование  $F_1$ , т.е. среднего гармонического между точностью и чувствительностью. Обоснованность использования в нашем случае среднего гармонического между чувствительностью и специфичностью показывают приведенные в таблице 5 расчеты. Из них видно, что использование  $F_1$  в большинстве случаев слишком сильно минимизирует специфичность.

Таблица 5. Эффективность диагностики на тестовой выборке при использовании наборов признаков, полученных с разными метриками качества классификации

Информативность	Метрика	<i>Sens</i>	<i>Spec</i>	<i>Acc</i>	<i>D</i>	$F_1$
Среднее гармоническое статистической, энтропийной, эвристической	$F_1$	0,88	0,24	0,75	0,38	0,85
	<i>D</i>	0,78	0,59	0,74	0,67	0,83
Среднее гармоническое статистической, эвристической	$F_1$	0,91	0,29	0,78	0,44	0,87
	<i>D</i>	0,73	0,52	0,69	0,61	0,79
Эвристическая	$F_1$	0,64	0,5	0,61	0,56	0,73
	<i>D</i>	0,62	0,53	0,6	0,57	0,71
Среднее гармоническое энтропийной, эвристической	$F_1$	0,9	0,24	0,76	0,38	0,86
	<i>D</i>	0,8	0,43	0,72	0,56	0,82

## 6. Заключение

В работе предложен новый метод формирования информационных признаков на базе количественного описания качественных признаков, получаемых с помощью анализа медицинских знаний и физико-математических моделей температурных полей молочных желез. Предлагаемый 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 качественно новых диагностических признаков.

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 [3] позволяет получить чувствительность и специфичность в районе 80%. Более того, в разрабатываемой консультационной системе предполагается применять алгоритмы классификации на базе нейронных сетей, генетических алгоритмов, нечеткой логики и т.д. Однако обоснование получаемых результатов будет про-

изводиться на основе полученных характеристических и высокоинформативных признаков.

Отдельно заметим, что реальная чувствительность и специфичность консультационной системы должна выявляться путем опытной эксплуатации, т.е. более интересным является вопрос, насколько полученные признаки улучшат диагностику, осуществляемую специалистами.

Отметим, что предложенный метод нахождения диагностических признаков может быть применен в других областях медицины, использующих данные микроволновой радиотермометрии.

Работа выполнена при финансовой поддержке РФФИ (проект № 15-47-02475-р\_поволжье\_a).

### Литература

1. ВАЙСБЛАТ А.В., ВЕСНИН С.Г., КОНКИН М.А. и др. *Использование микроволновой радиотермометрии в диагностике рака молочной железы* [Электронный ресурс]. – URL: <http://www.resltd.ru/rus/literature/cancer.htm>.
2. ВЕСНИН С.Г., КАПЛАН М.А., АВАКЯН Р.С. *Современная микроволновая радиотермометрия молочных желез // Опухоли женской репродуктивной системы*. – 2008. – №3. – С. 28–33.
3. ЗЕНОВИЧ А.В., ГЛАЗУНОВ В.А., ОПАРИН А.С., ПРИМАЧЕНКО Ф.Г. *Алгоритмы принятия решений в консультативной интеллектуальной системе диагностики молочных желез // Вестник Волгоградского государственного университета*. – Серия 1: Математика. Физика. – 2016. – №6. – С. 129–142.
4. КОБРИНСКИЙ Б.А. *Системы искусственного интеллекта в медицине: состояние, проблемы и перспективы // Новости искусственного интеллекта*. – 1995. – №2. – С. 65–79.
5. КОБРИНСКИЙ Б.А. *Консультативные интеллектуальные медицинские системы: классификация, принципы построения, эффективность // Врач и информационные технологии*. – 2008. – №2. – С. 38–47.

6. КОРЕНЕВСКИЙ Н.А., ЛУКАШОВ М.И., АРТЕМЕНКО М.В., АГАРКОВ Н.М. *Синтез гибридных нечетких решающих правил для классификации клинических вариантов течения генитального герпеса на основе моделей системных взаимосвязей // Фундаментальные исследования. – 2014. – №10. – С. 901–907.*
7. ЛОСЕВ А.Г., МАЗЕПА Е.А., ЗАМЕЧНИК Т.В. *О некоторых характерных признаках в диагностике патологии молочных желез по данным микроволновой радиотермометрии // Современные проблемы науки и образования. – 2014. – №6. – С. 254.*
8. ЛОСЕВ А.Г., МАЗЕПА Е.А., СУЛЕЙМАНОВА Х.М. *О взаимосвязи некоторых признаков РТМ-диагностики заболеваний молочных желез // Вестник Волгоградского государственного университета. – Серия 1: Математика. Физика. – 2015. – №4(29). – С. 35–44.*
9. ЛОСЕВ А.Г., ХОПЕРСКОВ А.В., АСТАХОВ А.С., СУЛЕЙМАНОВА Х.М. *Проблемы измерения и моделирования тепловых и радиационных полей в биотканях: анализ данных микроволновой радиотермометрии // Вестник Волгоградского государственного университета. – Серия 1: Математика. Физика. – 2015. – №6. – С. 31–71.*
10. ЯСНИЦКИЙ Л.Н. *Введение в искусственный интеллект: учебное пособие для студентов вузов. – М.: Academia. – 2005. – 176 с.*
11. BARRETT A.H., MYERS P.C. *Subcutaneous Temperature: A method of Noninvasive Sensing // Science. – 1975. – Vol. 190. – P. 669–671.*
12. BARRETT A.H., MYERS P.C., SADOWSKY N.L. *Microwave Thermography in the Detection of Breast Cancer // Am. J. Roengenol. – 1980. – №34. – P. 365–368.*
13. BURKE H.B. *Artificial neural networks for cancer research: outcome prediction // Seminars in Surgical Oncology. – 1994. – Vol. 10. – №1. – P. 73–79.*
14. GAUTHERIE M. *Temperature and Blood Flow Patterns in Breast Cancer During Natural Evolution and Following Radiotherapy // Biomedical Thermology. – 1982. – Vol. 107. – P. 21–64.*

15. HAMLIN S.K., STRAUSS P.Z. *Basic Concepts of Hemorheology in Microvascular Hemodynamics* // Critical care nursing clinics of North America. – 2014. – Vol. 26. – P. 337–344.
16. JAIN L.C., MARTIN N.M. *Fusion of Neural Networks, Fuzzy Systems and Genetic Algorithms: Industrial Applications* // CRC Press. – 1998. – 368 p.
17. KENNEDY K., MAC NAMEE B., DELANY S. *Learning without default: a study of one-class classification and low-default portfolio problem* // Proc. of the 20<sup>th</sup> Irish conference on Artificial intelligence and cognitive science (AICS'09), August 19-21, 2009, Dublin, Ireland. – 2009. – P. 174–187.
18. KONONENKO I., BRATKO I., KUKAR M. *Application of machine learning to medical diagnosis* // Machine Learning, Data Mining and Knowledge Discovery: Methods and Applications. – 1998. – P. 389–408.
19. KONONENKO I. *Machine Learning for Medical Diagnosis: History, State of the Art and Perspective* // Artificial Intelligence in Medicine. – 2001. – Vol. 23(1). – P. 89–109.
20. LESMO L., SAITTA L., TORASSO P. *Learning of Fuzzy Production Rules for Medical Diagnoses* // Approximate Reasoning in Decision Analysis. – 1982. – P. 249–260.
21. RODRIGUES D.B., MACCARINI P.F., SALAH S., COLEBECK E., TOPSAKAL E., PEREIRA P.J., LIMA-OVIEIRA P., STAUFFER P.R. *Numerical 3D modeling of heat transfer in human tissues for microwave radiometry monitoring of brown fat metabolism* // Proc. of SPIE 8584 (SPIE BIOS), February 2-7, 2013, San Francisco, California, United States. – 2013. – DOI: 10.1117/12.2004931.

## **THE THERMOMETRY DATA MINING IN THE DIAGNOSTICS OF MAMMARY GLANDS**

**Alexander Losev**, Volgograd State University, Volgograd, Doctor of Science, professor (allosev59@gmail.com).

**Vladislav Levshinsky**, Volgograd State University, Volgograd, student (vladi.lev.email@gmail.com).

*Abstract: The paper deals with the microwave thermometry data and proposes a method for forming informative features, based on qualitative descriptions of medical knowledge about the behavior of the temperature fields of mammary glands. There are some already known qualitative characteristics of breast cancer, which were form a basis for quantitative features, e.g. feature 'a large thermal asymmetry between mammary glands' can be described by temperature differences between corresponding points of right and left mammary glands. If the value of such difference is sufficiently large, then it may be an indication of pathology. After preprocessing, which involves weighting and cleaning, informative features may be applied in various classification algorithms, such as logistic regression, which yields about seventy percents of accuracy on a test sample, or the more complex ones that yield better accuracy: neural networks, genetic algorithm, and fuzzy classification. The significance of proposed features consists of the fact that they were formed from qualitative characteristics and each of them has a qualitative description, therefore they are of interest for further study and can be applied in diagnosis-advisory systems.*

**Keywords:** data mining, microwave radiothermometry, intelligent advisory systems.

*Статья представлена к публикации  
членом редакционной коллегии А.И. Михальским.*

*Поступила в редакцию 01.11.2016.*

*Опубликована 30.11.2017.*