

## Алгоритм локализации опухоли при раке молочной железы по данным микроволновой термометрии

*В. А. Глазунов, А.Г. Лосев*

*Волгоградский государственный университет*

**Аннотация:** Данная работа выполнена в рамках исследований, посвященных разработке методов интеллектуального анализа медицинских термометрических данных, предназначенных для создания консультативной интеллектуальной системы диагностики рака молочной железы. Предложен новый подход к формированию признакового пространства на основе данных микроволновой радиотермометрии и построен алгоритм локализации злокачественных новообразований в молочной железе. В процессе работы было получено и протестировано около 200 признаков нового типа. Алгоритм локализации на их основе может достигать достоверности свыше 70% на тестовых выборках.

**Ключевые слова:** микроволновая термометрия, рак молочной железы, маммология, локализация опухоли, достоверность, интеллектуальный анализ данных, генетический алгоритм, алгоритм взвешенного голосования, термометрические диагностические признаки, кросс-валидация.

### Введение

Наиболее распространенной формой онкологических заболеваний среди женщин является рак молочной железы. По данным ВОЗ, каждый год в мире обнаруживается от 800 тыс. до 1 млн новых случаев этой патологии [1].

Для диагностики рака молочной железы применяются различные технологии, в том числе метод микроволновой термометрии. В его основе лежит измерение собственного электромагнитного излучения тканей человека в микроволновом диапазоне длин волн и инфракрасном диапазоне, что позволяет определять температуры на глубине нескольких сантиметров и на коже. Однако опыт применения данного подхода в медицинской практике обозначил ряд проблем. В частности, анализ термометрических данных, который лежит в основе диагностики, является достаточно сложной задачей и требует большого опыта и высокой квалификации врача-диагноста. Данное обстоятельство препятствует организации массового скрининга рака

молочной железы на основе метода микроволновой радиотермометрии и, как следствие, нивелирует основные его достоинства.

Подобные проблемы традиционно решаются созданием систем интерпретации и анализа медицинских данных. При этом в настоящее время наиболее перспективными в медицине считаются консультативные интеллектуальные системы, т.е. экспертные системы, содержащие механизм объяснения и обоснования предлагаемых решений на языке, понятном пользователю. Некоторыми авторами в данном направлении ведется достаточно активная работа [2 – 4], однако они, как правило, решают задачу определения самого наличия онкологических заболеваний молочных желез у человека. Вопрос обнаружения месторасположения злокачественного новообразования внутри молочной железы, т.е. локализации опухоли, до сего времени оставался неисследованным.

В рамках данной работы ставится задача автоматизированного определения локализации опухоли молочной железы по данным микроволновой термометрии. Это позволит дать врачу дополнительную информацию при постановке диагноза и укажет районы молочных желез, на которые стоит обратить дополнительное внимание при термометрическом исследовании.

Отдельно отметим, что метод микроволновой термометрии не имеет противопоказаний, с его помощью можно обнаруживать наличие заболевания молочной железы в 83 % случаев [5,6]. Сама возможность применения данного способа для диагностики рака молочной железы была доказана еще в середине 70-х годов XX века [7]. Однако технические проблемы создания радиотермометра были решены лишь в конце 20 века (см., например, [5]).

В стандартной методике измерения температур производятся в девяти точках молочной железы (см. рис. 1).

---

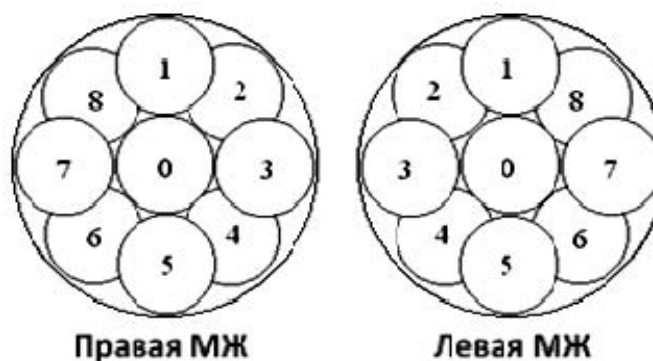


Рис. 1. – Расположение точек измерения на молочных железах

На самом деле метод микроволновой термометрии в состоянии выявить температурные аномалии внутри тела пациента, которые в большинстве случаев соответствуют патологическим нарушениям [8]. Результаты измерения позволяют строить картины распределения температур – термограммы, а также разрабатывать модели распределения температур в молочных железах [9].

### **Экспертная база данных и модели распределения опухолей**

В данной работе использовалась экспертная база данных, полученная на основе информации онкологических центров применяющих при диагностике метод микроволновой термометрии. При обследованиях применялся комплекс – РТМ-01-РЭС. База содержит результаты измерения внутренних и кожных температур, возраст, анамнез, диагноз, некоторую информацию о местоположении опухоли (злокачественной или доброкачественной) и номер датчика, который применялся при измерениях.

База содержит информацию о 726 молочных железах, из которых 198 поражены злокачественными опухолями, 215 относятся к группе риска, 148 – к здоровым молочным железам здоровых пациенток, а 165 – к здоровым молочным железам больных пациенток. В данном исследовании использовались только данные о молочных железах больных пациенток, поскольку задача определения наличия патологий в значительной мере решена в других работах ([2 – 4]).

Данные в экспертной базе данных получены с помощью различных температурных датчиков (помехозащищенном (№ 2) и совмещенном (№ 6)), которые имеют некоторые статистические отличия в своих результатах. По этой причине база делилась по типу датчика, примененного для измерений, а обучение и тестирование разрабатываемых алгоритмов производилось отдельно на этих частях. Отметим, что датчиком №2 были получены измерения для 107 молочных желез, которые имеют злокачественные опухоли, а датчиком №6 – для 91 молочной железы.

При описании места расположения опухоли медики чаще всего используют привязку к квадрантам. Каждая молочная железа делится на четыре квадранта так, как это показано на рисунке.



Рис. 2. – Расположение квадрантов на молочных железах

Приведем некоторые примеры текстовых описаний расположения опухоли, которые есть в базе:

- «В верхне-наружном квадранте опухоль до 2,5 см.»
- «В верхне-внутреннем квадранте опухоль размерами 2,5\*2,0 см.»

С другой стороны, достаточно часто при описании месторасположения опухоли указывается, что опухоль расположена на границе соседних квадрантов. Например:

- «На границе нижних квадрантов опухоль до 4 см.»

- «На границе наружных квадрантов узел без четких контуров»

Следовательно, к четырем, указанным выше, вариантам расположения опухоли добавляется еще четыре. Также иногда указывается, что опухоль расположена в центре молочной железы, например:

- «В центральном отделе опухоль до 3 см.»

Это добавляет еще один вариант локализации.

Таким образом, одним из возможных вариантов индексации положения опухоли внутри молочной железы является соотнесение их к одной из указанных выше девяти областей. Отметим, что в данном варианте количество и расположение областей соответствуют точкам измерения температур в стандартной методике микроволновой радиотермометрии. По этим причинам области нумеруются так же, как соответствующие точки измерения. Область  $d_0$  содержит точку  $t_0$ ,  $d_1 - t_1$ , и так далее до  $d_8$ .

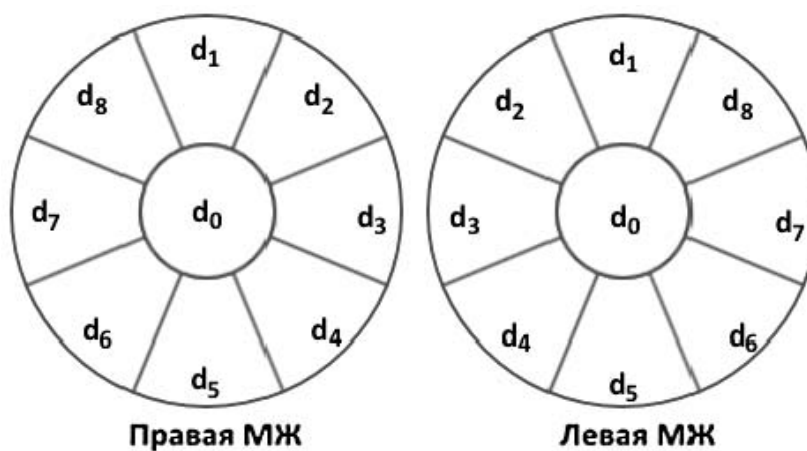


Рис. 3. – Схема расположения областей

В случаях, когда потребуется явное указание, к какой молочной железе имеет отношение область, индекс будет дополняться буквами R и L для правой и левой молочной железы соответственно. Пример:  $d_{R1}$ .

Перевести текстовые описания из существующей базы в модель по областям  $d_0 - d_8$  можно различными способами. Можно считать, что опухоль располагается строго в одной из областей  $d_i$ . В таком случае, например, описанию «на границе нижних квадрантов опухоль до 4 см.» будет

соответствовать области  $d_5$ . Такая модель данных в дальнейшем называется строгой.

Однако одним из основных вариантов использования результатов данного исследования является модификация существующей методики измерений температур с целью повышения качества диагностики рака молочной железы. Предполагается увеличение количества точек измерения температур в тех областях молочной железы, где наиболее вероятно нахождение опухоли. Для этого область предполагаемого расположения опухоли не должна иметь размеры области  $d_i$ , наоборот, она должна быть значительно больше.

Поэтому в рамках данной работы также рассматривалась нестрогая модель распределения опухолей по областям  $d_i$ . В ее рамках расположение опухоли в пределах молочной железы считается более обширным, чем в строгой, и дополнительно занимает две соседние области. В этом случае описанию «на границе нижних квадрантов опухоль до 4 см.» будет соответствовать уже набор областей:  $d_4, d_5, d_6$ .

Область  $d_0$  считается особой, поскольку соседствует со всеми другими областями, и поэтому рассматривается также как в строгой модели данных.

Подводя итог сказанному выше, следует заметить, что в работе применяется четыре различных набора исходных данных. Температурные данные делятся на две группы по использованному для измерений типу датчика, и далее на основе каждой группы создается две новые – по модели распределения опухоли по областям. Таким образом, в работе были применены следующие наборы исходных данных:

- датчик №2, строгая модель распределения опухолей;
  - датчик №6, строгая модель распределения опухолей;
  - датчик №2, нестрогая модель распределения опухолей;
  - датчик №6, нестрогая модель распределения опухолей;
-

Обучение и проверка результатов производились независимо на всех перечисленных наборах. В работе применялась кросс-валидация по методу Монте-Карло, в процессе которой исходный набор данных случайным образом делился на обучающую и тестовую выборку в соотношении 2 к 1. При делении учитывалось существующая статистика распределения опухолей по областям. Для каждого из четырех наборов было получено по 10 обучающих и тестовых выборок.

### **Диагностические признаки для локализации опухоли**

В предшествующих исследованиях, на основе существующего клинического опыта применения микроволновой радиотермометрии последних двух десятилетий и анализа математических моделей был выявлен ряд качественных термометрических признаков рака молочной железы (см., например, [2, 3, 4, 5, 8]). Например, у пациенток с патологией отмечается наличие повышенной величины термоасимметрии между одноименными точками молочных желез; повышенный разброс температур между отдельными точками в пораженной молочной железе; повышенная разница температур сосков; повышенная температура соска в пораженной молочной железе по сравнению со средней температурой молочной железы; значительный модуль разности кожной и глубинной температур и некоторые другие. Также (см., например, [2 – 4]) были предложены математические модели описания данных эффектов и описаны их количественные характеристики.

В указанных работах для определения наличия патологий используются, так называемые, термометрические диагностические признаки. В частности, в работе [3] приводится следующее определение признака:

«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  $f^q$ . Под «информативной» областью понимается такая область, которая максимизирует значение  $V$ .

Построенные на данной основе признаковые пространства до сего времени использовались для определения одного из двух состояний, например, больна пациентка или здорова. В данный момент наработанный опыт позволяет приступить к задаче локализации патологии, для решения которой требуется существенная модификация описанного выше подхода.

При создании признакового пространства, предназначенного для решения задачи локализации злокачественного новообразования, за основу можно брать функции из полученных ранее признаков [3, 4]. Однако для задачи локализации подойдут лишь те из них, которые могут показывать повышение температуры в конкретном наборе областей. Иными словами, например, признаки, показывающие общую неравномерность распределения температур по всей молочной железе, не являются подходящими.

Набор областей, для которых признак показывает повышение температуры, далее называется набором  $N$ .

Приведем пример получения признака локализации. Построение нового признака начинается с выбора функции  $f$ . Возьмем в качестве примера функцию:

$$f = t_0 - t_1$$

Различные значения данной функции могут показывать повышение температуры в различных точках. Если значение  $f$  меньше нуля, то температура в точке  $t_1$  повышена, а если оно больше нуля, то температура повышена в точке  $t_0$ . Это значит, что в данном случае на ее основе можно получить два признака локализации. В первом варианте в набор  $N$  нужно добавить область  $d_1$ , а во втором –  $d_0$ . Нахождение оптимального множества  $X$  будет описано ниже.

---



Перейдем к определению понятия признака, которое будет использоваться в данном исследовании.

Признаком локализации будем называть четверку  $(f, Y, X, N)$ , состоящую из функции от температур  $f$ , критерия качества  $Y$ , множества  $X$  и набора областей молочной железы  $N$ . Где:

- $X$  является подмножеством области значений  $f$ ;
- к  $N$  относятся области  $d_i$ , для которых выполнение признака увеличивает вероятность наличия опухоли.

Если значение функции  $f$  принадлежит множеству  $X$  для конкретных температурных данных, то признак считается выполненным.

Если для признака заданы  $f, N$  и произвольное множество  $X$ , то для него на выборке температурных данных могут быть найдены следующие величины:

Чувствительность (Sensitivity), или полнота (Recall), которая вычисляется по формуле:

$$Sensitivity = \frac{TP}{TP + FP}$$

где  $TP$  – количество верных положительных результатов (действительно обнаружили опухоль),  $FP$  – количество ошибочных положительных результатов (опухоль обнаружена там, где ее нет).

Специфичность (Specificity), которая вычисляется по формуле:

$$Specificity = \frac{TN}{TN + FN}$$

где  $TN$  – количество верных отрицательных результатов (опухоль действительно отсутствует),  $FN$  – количество ошибочных отрицательных результатов (опухоль не обнаружена там, где она есть).

Каждый признак локализации, по сути, решает задачу классификации молочных желез. Для оценивания качества классификации часто

---

используются разные варианты F-мер, например, в работе [3] применялось среднее гармоническое между чувствительностью и специфичностью. Этот вариант достаточно хорошо себя зарекомендовал, поэтому он рассматривался в качестве критерия  $Y$ :

$$Y = \frac{2 * Sensitivity * Specificity}{Sensitivity + Specificity} \quad (1)$$

Кроме этого также рассматривался вариант со средним геометрическим:

$$Y = \sqrt{Sensitivity * Specificity} \quad (2)$$

Практика показала, что выбор варианта не оказывает существенного воздействия на работу алгоритма локализации. Для дальнейшего использования был выбран вариант (1).

Оптимальное множество  $X$  для признака находится в процессе максимизации величины  $Y$  при фиксированных  $f$  и  $N$ .

В таблице 1 приводятся некоторые примеры полученных признаков на данных датчика № 6 и строгой модели распределения опухолей по областям.

Таблица № 1

#### Примеры признаков локализации

Функция $f$	Критерий $Y$	Множество $X$	Набор $N$
$t_{pmm\ cp} - (t_{pmm\ 6} + t_{pmm\ 7} + t_{pmm\ 8}) / 3$	0,51	$(-\infty; -0,49)$	$d_6, d_7, d_8$
$(t_{pmm\ 8} + t_{pmm\ 1}) / 2 - (t_{uk\ 8} + t_{uk\ 1}) / 2$	0,49	$(1,46; +\infty)$	$d_1, d_8$
$t_{uk\ 8} - t_{uk\ 4}$	0,48	$(0,14; +\infty)$	$d_8$
$t_{pmm\ 0} - (t_{pmm\ 7} + t_{pmm\ 8}) / 2$	0,46	$(-\infty; -0,64)$	$d_7, d_8$

#### Алгоритм локализации

В качестве основной характеристики успешности работы алгоритма локализации применялась достоверность (Ассигасу), вычисляемая по следующей формуле:

$$Accuracy = \frac{TN + TP}{TN + FN + TP + FP}$$

где  $TN$  – количество верных отрицательных результатов (опухоль действительно отсутствует),  $TP$  – количество верных положительных результатов (действительно обнаружили опухоль),  $FN$  – количество ошибочных отрицательных результатов (опухоль не обнаружена там, где она есть),  $FP$  – количество ошибочных положительных результатов (опухоль обнаружена там, где ее в действительности нет).

В качестве алгоритма локализации использовался алгоритм взвешенного голосования признаков [10]. Локализация производится независимо для каждой молочной железы, поэтому доступно 9 вариантов локализации из областей  $d_0 - d_8$ .

В качестве веса  $i$ -го диагностического признака использовалась следующая величина:

$$k_i * \frac{Y_i}{NCount_i}$$

Здесь  $Y_i$  – критерий качества  $i$ -го признака,  $NCount_i$  – количество областей в наборе  $N_i$  признака, а  $k_i$  – дополнительный коэффициент.

Деление на  $NCount_i$  применялось для учета того, что признак может повышать вероятность нахождения опухоли сразу в нескольких областях  $d_j$ . Коэффициенты  $k_i$  добавлены для более точного учета особенностей признаков в рамках набора.

Поскольку максимальное количество признаков в наборе может достигать до 200, для задачи подбора коэффициентов  $k$  применялся эвристический алгоритм, а именно – стандартная версия генетического алгоритма [11]. Точные характеристики алгоритма подбирались в процессе вычислительных экспериментов.

Также с целью повышения достоверности алгоритма применялась минимизация наборов признаков. Для этой цели применялась модифицированная версия генетического алгоритма, который использовался в задаче подбора коэффициентов  $k_i$ .

При формировании оптимального набора признаков проводился вычислительный эксперимент, направленный на поиск наиболее подходящего количества признаков в наборе. Сравнению подвергались наборы с численностью от 20 до 150 признаков, наилучшие результаты показали наборы из 25 – 35 признаков.

### Результаты

Тестирование алгоритма осуществлялось на четырех наборах данных, которые делятся по использованному типу датчика и модели распределения опухолей по областям.

В таблице приведены подробные результаты тестирования, которые являются усредненными по всем обучающим и тестовым выборкам, использованным для каждого из четырех наборов данных. В числовых ячейках приводится значение достоверности (Accuracy).

Таблица № 2

#### Результаты тестирования

Тип датчика	База	Модель распределения опухолей по областям	
		Строгая	Нестрогая
№ 2	Обучающая	38,86 %	78,02 %
	Тестовая	32,15 %	73,63 %
№ 6	Обучающая	36,95 %	64,61 %
	Тестовая	31,44 %	61,75 %

При применении строгой модели распределения опухолей по областям достоверность является относительно невысокой – около 30 – 40%. В то же

время при использовании нестрогой модели эта характеристика может достигать 60 – 70%, что уже может быть подходящим для практического использования.

### **Заключение**

В процессе работы был разработан новый тип термометрического диагностического признака – признак локализации. Было получено и протестировано около 200 таких признаков.

На основе этих признаков был разработан алгоритм локализации опухоли, который при использовании нестрогой модели расположения опухоли по областям может достигать достоверности свыше 70% на тестовых выборках.

Работа выполнена при финансовой поддержке РФФИ, проект № 19-01-00358.

### **Литература**

1. Пак Д. Д., Усов Ф. Н., Фетисова Е. Ю., Волченко А. А., Ефанов В. В. Современные подходы к лечению больных с карциномой *in situ* молочной железы // Онкология. 2013. № 4. С. 34-39.
2. Зенович А. В., Петренко А. Ю. Построение настраиваемого медиком классификатора онкологических заболеваний молочных желез методом комбинированной термометрии. //Инженерный вестник Дона, 2019, № 5. URL: [ivdon.ru/ru/magazine/archive/N5y2019/5974](http://ivdon.ru/ru/magazine/archive/N5y2019/5974).
3. Лосев А. Г., Левшинский В. В. Интеллектуальный анализ термометрических данных в диагностике молочных желез // УБС. 2017. №70. С. 133-135.
4. Мазепа Е. А., Гришина О. В., Левшинский В. В., Сулейманова Х. М. Об унификации метода анализа данных микроволновой радиотермометрии

- // Вестник ВолГУ. Серия 1: Математика. Физика. 2017. №6 (43). С. 38-50.
5. Веснин, С. Г., Каплан М. А., Авакян Р. С. Современная микроволновая радиотермометрия молочных желез // Опухоли женской репродуктивной системы. 2008. № 3. С. 28-35.
  6. Сдвижков А.М., Веснин С.Г., Карташева А.Ф. О месте радиотермометрии в маммологической практике//Актуальные проблемы маммологии, М., 2000. С. 28-40.
  7. Barrett A.H., Myers P.C. Subcutaneous Temperature: A method of Noninvasive Sensing // Science. 1975. Vol.190. pp. 669-671.
  8. Vesnin S., Turnbull A., Goryanin I., Modern Microwave Thermometry for Breast Cancer. // URL: [researchgate.net/publication/320735583\\_Modern\\_Microwave\\_Thermometry\\_for\\_Breast\\_Cancer](https://researchgate.net/publication/320735583_Modern_Microwave_Thermometry_for_Breast_Cancer).
  9. Поляков М.В., Хоперсков А.В. Вычислительные эксперименты для исследования радиационных и тепловых полей в биотканях. // Инженерный вестник Дона, 2017, № 2. URL: [ivdon.ru/ru/magazine/archive/n2y2017/4239](http://ivdon.ru/ru/magazine/archive/n2y2017/4239).
  10. Вьюгин В.В. Математические основы теории машинного обучения и прогнозирования. М.: 2013. 387 с.
  11. Карпенко А.П. Современные алгоритмы поисковой оптимизации. Алгоритмы, вдохновленные природой. М.: Издательство МГТУ им. Н.Э Баумана, 2014, С. 65-95.

### References

1. Pak D. D., Usov F. N., Fetisova E. YU., Volchenko A. A., Efanov V. V. Onkologiya. 2013. № 4. pp. 34-39.
2. Zenovich A. V., Petrenko A. YU. Inzenernyj vestnik Dona, 2019, № 5. URL: [ivdon.ru/ru/magazine/archive/N5y2019/5974](http://ivdon.ru/ru/magazine/archive/N5y2019/5974).

3. Losev A. G., Levshinskiy V. V. UBS. 2017. №70. pp. 133-135.
4. Mazepa E. A., Grishina O. V., Levshinskiy V. V., Suleymanova KH. M. Vestnik VolGU. Seriya 1: Matematika. Fizika. 2017. №6 (43). pp. 38-50.
5. Vesnin, S. G., Kaplan M. A., Avakyan R. S. Opukholi zhenskoj reproduktivnoy sistemy. 2008. № 3. pp. 28-35.
6. Sdvizhkov A.M., Vesnin S.G., Kartasheva A.F. Aktual'nyye problemy mammologii, M., 2000. pp. 28-40.
7. Barrett A.H., Myers P.C. Subcutaneous Temperature: A method of Noninvasive Sensing. Science. 1975. Vol.190. pp. 669-671.
8. Vesnin S., Turnbull A., Goryanin I., Modern Microwave Thermometry for Breast Cancer. URL: [researchgate.net/publication/320735583\\_Modern\\_Microwave\\_Thermometry\\_for\\_Breast\\_Cancer](https://www.researchgate.net/publication/320735583_Modern_Microwave_Thermometry_for_Breast_Cancer).
9. Polyakov M.V., KHoperskov A.V. Inzenernyj vestnik Dona, 2017, № 2. URL: [ivdon.ru/ru/magazine/archive/n2y2017/4239](http://ivdon.ru/ru/magazine/archive/n2y2017/4239).
10. V'yugin V.V. Matematicheskiye osnovy teorii mashinnogo obucheniya i prognozirovaniya [Mathematical foundations of the theory of machine learning and forecasting]. M.: 2013. 387 p.
11. Karpenko A.P. Sovremennyye algoritmy poiskovoy optimizatsii. Algoritmy vdokhnovlennyye prirodoy. [Modern search engine optimization algorithms. Algorithms inspired by nature]. M.: Izdatel'stvo MGTU im. N.E Baumana, 2014, pp. 65-95.