

СРАВНИТЕЛЬНЫЙ АНАЛИЗ МЕТОДИК ТЕМПЕРАТУРНЫХ ИЗМЕРЕНИЙ В ДИАГНОСТИКЕ РАКА МОЛОЧНОЙ ЖЕЛЕЗЫ¹

Попов И. Е.², Поляков М. В.³, Петренко А. Ю.⁴

(Волгоградский государственный университет, Институт
математики и информационных технологий, Волгоград)

Диагностика рака молочной железы методом микроволновой радиотермометрии достигла высокой точности за счет построения высокоинформативного признакового пространства, а также подбора алгоритма машинного обучения, являющегося наиболее эффективным при классификации рака. Основной особенностью метода микроволной термометрии является измерение температур в заданных областях обследуемого органа. Перспективным направлением исследования видится добавление точек измерений в схему обследования. Дополнительные точки измерения могут добавить необходимую для алгоритмов машинного обучения информацию, по которой они смогут более точно определять температурные аномалии, характерные для рака молочной железы. В данной работе сравниваются стандартная и расширенная схемы температурных измерений. Данные генерировались с помощью математических моделей температурных полей молочных желез. Сравнение производится по фиксированному признаковому пространству, каждый признак которого считается по заданному множеству температур. Это позволяет применять один и тот же признак к обеим схемам измерений. Для сравнения было построено 4 различных комбинаций признаков. Проведены вычислительные эксперименты. В результате расширенная схема температурных измерений показала себя более эффективной, чем стандартная.

Ключевые слова: микроволновая радиотермометрия, машинное обучение, рак молочной железы, диагностика.

1. Введение

За последние годы была получена высокая точность обнаружения опухолей в диагностике рака молочной железы методом микроволновой радиотермометрии [1, 4]. Было сформировано оптимальное признаковое пространство, отображающее основные характеристические особенности различных диагностических

¹ Работа выполнена в рамках проекта РФФИ №19-01-00358.

² Илларион Евгеньевич Попов, магистрант (popov.larion@volsu.ru).

³ Максим Валентинович Поляков, аспирант (m.v.polyakov@volsu.ru).

⁴ Александр Юрьевич Петренко, магистрант (a.petrenko@volsu.ru).

классов. Выявлен наилучший алгоритм классификации в диагностике рака молочной железы [2]. Одним из потенциальных путей развития метода нам видится расширение схемы температурных измерений.

Метод микроволновой радиотермометрии заключается в последовательном измерении температур по заданной схеме [7]. Измерения проводятся в двух диапазонах: инфракрасном и микроволновом. Инфракрасные излучения фиксируют температуру кожи, микроволновые – температуру на глубине в несколько сантиметров. Таким образом врачу-диагносту и системе диагностики становится известной информация о кожных и глубинных температурах. Стандартная схема измерений представлена на рис. 1. Здесь точка 0 соответствует соску, остальные соответствуют внешнему радиусу молочной железы.

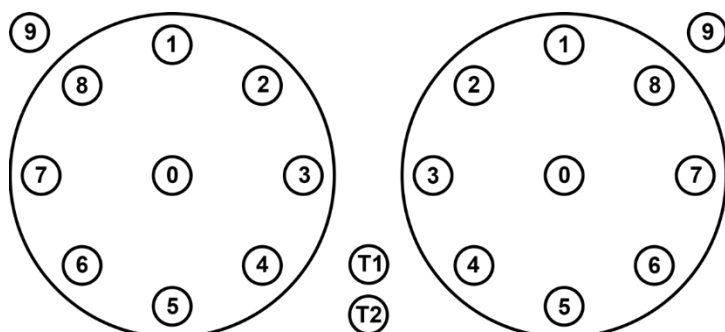


Рис 1. Стандартная схема измерений

Далее для диагностики по полученным температурам проектируется признаковое пространство, основанное на характеристических особенностях диагностических классов [3]. Однако нам видится возможным увеличение точности диагностики за счет добавления точек измерений. Одной из причин является то, что основной характеристической особенностью рака молочной железы является наличие «горячей» зоны. И большее количество точек измерений позволит точнее определять их наличие. Другой причиной увеличения числа точек является потенциальная

возможность расширения признакового пространства за счет обнаружения новых характеристических закономерностей, которые заключаются в отношении добавленных и оригинальных точек.

Таким образом в рамках данной работы рассматривается расширенная схема измерений, представленная на рис. 2. Такая схема выбрана из-за того, что, как правило, «горячая» зона затрагивает сосок и область внешнего радиуса, а точки внутреннего радиуса могут дать дополнительную информацию о размере и форме «горячей» зоны.

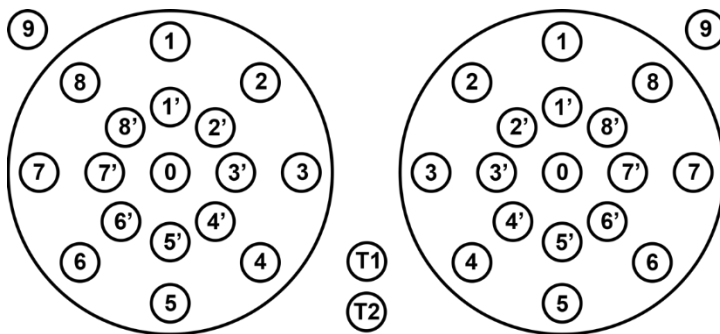


Рис 2. Расширенная схема измерений.

Для сравнения использовались молочные железы, построенные на основе математических моделей из работ [5, 6]. Модели построены на основе уравнений теплопроводности для биологических тканей с учетом анатомии молочной железы и валидировались по натурным данным. Таким образом температурные поля полученных моделей ведут себя схоже с температурными полями молочных желез реальных пациентов. В тоже время модель позволяет получать информацию о температуре молочной железы в любой её области.

2. Сравнительный анализ.

База данных содержит 50 здоровых молочных желез и столько же желез с опухолью. По температурным данным строилось признаковое пространство, а затем проводилась классификация. В качестве результатов классификации использовались следующие показатели: специфичность – доля верно диагностированных здоровых пациентов, чувствительность – доля верно диагностированных пациентов с раком молочной железы и эффективность – среднегеометрическое по двум предыдущим показателям. Для того чтобы результаты были показательными, признаковое пространство для расширенной схемы температурных измерений выбиралось подобное оригинальной схеме. Рассмотрим это подробнее.

2.1. ПРОЕКТИРОВАНИЕ ПРИЗНАКОВОГО ПРОСТРАНСТВА

В качестве основы для сравнения использовалось признаковое пространство из работы [3]. Оно было выбрано по нескольким причинам. Во-первых, оно эффективно при диагностике рака молочной железы, а во-вторых, описанный в статье метод конструирования признаков позволяет использовать их при различных схемах измерений. Так как математические модели строятся по одной молочной железе, без описания парного органа, то признаковое пространство проектировалось на основе следующих гипотез:

- повышенный разброс температур в молочной железе;
- повышенное соотношение кожных и глубинных температур в молочной железе.

В таблице

Таблица 1 приведено полное признаковое пространство по оригинальным температурам.

Здесь $T^{i,ir}$ – множество кожных температур i -й молочной железы; $T^{i,mw}$ – множество глубинных температур и $T^{i,g}$ – множество внутренних градиентов – разница между глубинной и кожной температурами:

$$T^{i,g} = T^{i,mw} - T^{i,ir}.$$

Таблица 1. Признаковое пространство

№	Значение	№	Значение
1	$\max_{t \in T^{i,g}} t $	2	$\min_{t \in T^{i,g}} t $
3	$\min_{t \in T^{i,g}} t$	4	$\ T^{i,g}\ _1$
5	$\ T^{i,g}\ _2$	6	$\max_{t \in T^{i,mw}} T^{i,mw} - t $
7	$s(T^{i,mw})$	8	$\max_{t \in T^{i,mw}} t - \min_{t \in T^{i,mw}} t$
9	$\max_{t \in T^{i,mw} \setminus T_0^{i,mw}} T_0^{i,mw} - t $	10	$T_0^{i,mw} - \bar{t}$
11	$\max_{t \in T^{i,ir}} T^{i,ir} - t $	12	$s(T^{i,ir})$
13	$\max_{t \in T^{i,ir}} t - \min_{t \in T^{i,ir}} t$	14	$\max_{t \in T^{i,ir} \setminus T_0^{i,ir}} T_0^{i,ir} - t $
15	$T_0^{i,ir} - \overline{T^{i,g}}$	16	$\max_{t \in T^{i,g}} T^{i,g} - t $
17	$s(T^{i,g})$	18	$\max_{t \in T^{i,g}} t - \min_{t \in T^{i,g}} t$
19	$\max_{t \in T^{i,g} \setminus T_0^{i,g}} T_0^{i,g} - t $	20	$T_0^{i,g} - \overline{T^{i,g}}$

Функция $s(t)$ – среднеквадратичное отклонений, $\|t\|_1$ – манхэттенская норма, \bar{t} – среднееарифметическое по заданному множеству t .

Для оригинальных данных эти множества состоят из температур по точкам 0, 1, ..., 8. В случае добавления точек измерения, эти множества состоят из температур по точкам 0, 1, 1', 2, 2', ..., 8, 8'.

Для их сравнения были выбраны четыре признаковых пространства:

- 1) признаки по оригинальному множеству температур;
- 2) признаки по расширенному множеству температур;
- 3) наилучшие признаки по двум предыдущим множествам;
- 4) наилучшие признаки в каждой паре.

В таблице Таблица 2 приведён состав каждого признакового пространства. Пояснения по таблице: в ячейке записано значение «о» – в признаковое пространство входит признак по оригинальному множеству температур, «р» – признак по расширенному

множеству, «о, р» – оба варианта признака или «–» – не входит в признаковое пространство.

Таблица 2. Признаковые пространства

Признак	Номер признакового пространства			
	1	2	3	4
1	о	р	о, р	о
2	о	р	о, р	о
3	о	р	о, р	р
4	о	р	о, р	р
5	о	р	о, р	р
6	о	р	–	о
7	о	р	–	р
8	о	р	–	р
9	о	р	–	р
10	о	р	–	р
11	о	р	–	о
12	о	р	–	о
13	о	р	–	о
14	о	р	–	о
15	о	р	р	р
16	о	р	о, р	о
17	о	р	о, р	о
18	о	р	о, р	о
19	о	р	о, р	о
20	о	р	р	р

Опишем подробнее последние две группы признаковых пространств. Помимо прямого сравнения признаков по различным схемам температурных измерений нас также интересует вопрос их совместного использования, повышается ли от этого эффективность. Но простое объединение данных признаковых пространств даёт в два раза больше признаков, что нам видится некорректным при сравнении. Поэтому решено было сократить их количество вдвое. При этом рассматривались два варианта сокращений. Первый вариант – выбор половины наилучших признаков

среди всего множества. Второй вариант – попарное сравнение одинаковых признаков из двух множеств и выбор наилучшего из них. Критерием оценивания был выбран f -критерий дисперсионного анализа.

Таким образом, в третьем признаковом пространстве находится 9 оригинальных признаков и 11 признаков по расширенному множеству температур. В четвертом – 10 оригинальных признаков и 10 признаков по расширенному множеству.

2.2. ВЫЧИСЛИТЕЛЬНЫЕ ЭКСПЕРИМЕНТЫ

В качестве алгоритма классификации использовался байесовский наивный классификатор. Для получения усредненных результатов вычислительные эксперименты проводились множество раз методом повторной перекрестной проверки. Метод случайным образом разделяет базу данных на N равных частей с сохранением баланса классов в каждой. Одна часть помещается в тестовую выборку, остальные – в обучающую. По обучающей выборке происходит настройка алгоритма классификации, а по тестовой выборке осуществляется классификация. Затем другая часть помещается в тестовую выборку и процесс повторяется. После того как все части побывали в тестовой выборке, база данных снова разбивается на N равных частей случайным образом. И так повторяется M раз. В конце получается $M \times N$ результатов классификаций, которые усредняются. В таблице Таблица 3 приведены результаты таких вычислительных экспериментов.

Таблица 3. Результаты классификации

Группа признаков	Специфичность		Чувствительность		Эффективность	
	Средн.	stdev	Средн.	stdev	Средн.	stdev
1	0,702	0,141	0,781	0,119	0,733	0,087
2	0,727	0,125	0,783	0,122	0,749	0,082
3	0,698	0,132	0,800	0,110	0,742	0,087
4	0,732	0,134	0,791	0,112	0,755	0,088

Здесь «средн.» – среднее арифметическое метрики по всем результатам классификации, «stdev» – среднеквадратичное отклонение по всем результатам.

Как видно, классификация по признаковому пространству с расширенным множеством температур показывает увеличение специфичность на 2% и незначительное увеличение чувствительности. Наилучшую эффективность показывает четвертое признаковое пространство. По сравнению с первым специфичность увеличилась на 3%, чувствительность – на 1%.

3. Заключение

В результате расширенная схема температурных измерений повышает эффективность диагностики относительно оригинальной схемы. Поэтому возникает новая область исследований: построение оптимальной схему температурных измерений, изучение полученных таким образом более точных температурных карт и выявление новых закономерностей в данных. Составленная в данной работе схема измерений видится нам потенциально полезной, кроме того, что она более эффективная, еще и в том, что возникает возможность определения размера и формы «горячих» зон, а также в рассмотрении отношения температур внутреннего и внешнего радиусов.

Также необходимо подтвердить полученные результаты на натуральных данных, полученных в процессе измерения температур у здоровых и больных пациентов, так как в таком случае станет доступна информация о температурах парной молочной железы. Признаки, построенные по гипотезам об их соотношении, являются наиболее эффективными, из-за чего точность диагностики может дополнительно повыситься при применении расширенной схемы измерений.

Литература

1. ЗАМЕЧНИК Т.В., ЛОСЕВ А.Г., ЛЕВШИНСКИЙ В.В. *Результаты оптимизации диагностических признаков рака*

- молочной железы, выявляемые методом микроволновой радиотермометрии // Медицинский вестник Северного Кавказа. – 2019. – №14(1.1). – С. 48–52.
2. ПОПОВ И.Е. *Сравнительный анализ алгоритмов классификации в диагностике рака молочной. Выпускная квалификационная работа.* – Волгоград: б.н., 2020. – 29 с.
 3. LEVSHINSKII V.V. *Mathematical models for analyzing and interpreting microwave radiometry data in medical diagnosis* // Journal of Computational and Engineering Mathematics. – 2021. – Vol. 8. – No. 1. – P. 3–14.
 4. LOSEV A.G., MEDVEDEV D.A., SVETLOV A.V. *Neural Networks in Diagnosis of Breast Cancer* // In: "Smart Technologies" for Society, State and Economy. – P. 220–227.
 5. POLYAKOV M.V., KHOPERSKOV A.V., ZAMECHNIC T.V. *Numerical modeling of the internal temperature in the mammary gland* // Lecture Notes in Computer Science. – 2017. – Vol. 10594 LNCS. – P. 128–135.
 6. POLYAKOV M., LEVSHINSKII V., KHOPERSKOV A. *Modeling of brightness temperature in biological tissue* // Journal of Physics: Conference Series. – 2019. – Vol.1368(4). – 042057.
 7. ZHAROV V.P., VESNIN S.G., SUEN J.Y., HARMS S.E.M.D. *Photothermal/microwave radiometry for imaging and temperature feedback* // Proc. SPIE 4618, Biomedical Optoacoustics III, (10 June 2002).

COMPARATIVE ANALYSIS OF TEMPERATURE MEASUREMENT METHODOLOGY IN BREAST CANCER DIAGNOSIS

Illarion Popov, VolSU, Institute of Mathematics and Information Technology, undergraduate (popov.larion@volsu.ru).

Maksim Polyakov, VolSU, Institute of Mathematics and Information Technology, graduate student (m.v.polyakov@volsu.ru).

Alexander Petrenko, VolSU, Institute of Mathematics and Information Technology, undergraduate (a.petrenko@volsu.ru).

Abstract: Diagnosis of breast cancer by microwave thermometry has achieved high accuracy due to the construction of a highly informative feature space, as well as the

selection of a machine learning algorithm, which is the most effective in the classification of cancer. The main feature of the microwave thermometry method is the measurement of temperatures in specified areas of the examined organ. Adding measurement points in the survey scheme is seen as a promising direction of research. Additional measurement points can add information that machine learning algorithms need to more accurately determine temperature abnormalities that are characteristic of breast cancer. This paper compares standard and extended temperature measurement schemes. The data were generated using mathematical models of the temperature fields of the mammary glands. Comparison is made according to a fixed feature space, each feature of which is counted according to a given set of temperatures. This allows the same characteristic to be applied to both measurement schemes. For comparison, 4 different combinations of features were built. Computational experiments were carried out on them. As a result, the extended temperature measurement scheme proved to be more efficient than the standard one.

Keywords: microwave thermometry, machine learning, breast cancer, diagnostics

УДК 519.68

ББК 32.973.202-018.2