

Построение настраиваемого медиком классификатора онкологических заболеваний молочных желез методом комбинированной термометрии

А.В. Зенович, А.Ю. Петренко

Волгоградский государственный университет, Волгоград

Аннотация: Рассматривается задача диагностики онкологических заболеваний молочных желез методом комбинированной термометрии. Предложен алгоритм формирования наборов двумерных обобщенных признаков диагностики, на базе которых строятся классификаторы методом логистической регрессии. Каждый двумерный признак формируется на основе двух одномерных методом сужающихся эллипсов. При этом путем вариации полуосей эллипсов можно управлять чувствительностью и специфичностью, фиксируя один из этих параметров и добиваясь максимального значения второго. Вычислительные эксперименты показали, что, варьируя полуоси эллипсов, можно добиться чувствительность и специфичности около 80%. Построенные наборы позволяют создать настраиваемый медиком классификатор, позволяющий врачу задавать границу, ниже которой не должна опускаться специфичность классификатора, и добиваться при этом максимальной чувствительности.

Ключевые слова: комбинированная термометрия, алгоритмы диагностики, анализ данных, чувствительность алгоритма, специфичность алгоритма, логистическая регрессия, настраиваемый классификатор, скрининг молочной железы, маммология.

Введение

В настоящее время активно создаются медицинские приборы нового поколения, предназначенные для постановки диагноза методом построения карт теплового излучения человеческого организма. При многих заболеваниях в организме возникают локальные температурные аномалии. Выявление данных аномалий несомненно поможет медику поставить правильный диагноз. Одним из бурно развивающихся методов выявления аномалий является микроволновая радиотермометрия, заключающаяся в измерении внутренних и поверхностных температур тканей по интенсивности их теплового излучения, соответственно, в микроволновом и инфракрасном диапазоне [1-4].

За последние десятилетия метод микроволновой радиотермометрии получил достаточно широкое применение в маммологии (см. [3-6]), флебологии, хирургии, гинекологии и многих других областях медицины. В

России прогресс в данной области диагностики связан с созданием приборов, позволяющих комбинировать измерение внутренних (глубинных) температур тканей по их естественному тепловому излучению в микроволновом диапазоне и измерение температуры кожных покровов по инфракрасному тепловому излучению. Одним из таких приборов является радиотермометр РТМ-01-РЭС (см. Рис. 1),



Рис 1. - Радиотермометр РТМ-01-РЭС

Диагностический комплекс, построенный на базе данного прибора, позволяет строить карты температурных полей человека и выявлять температурные аномалии, позволяющие квалифицированному врачу поставить диагноз. Ранее проводились серии вычислительных экспериментов по изучению распространения радиационных и тепловых полей в биотканях [7-8].

Существующая методика диагностики рака молочной железы по данным микроволновой радиотермометрии (см. [3]) подразумевает измерение 44

поверхностных и глубинных температур в соответствии со схемой на Рис. 2. Измерения проводятся в 10 точках на каждой железе (см. рис. 2), в аксиллярной области и в двух контрольных точках Т1 и Т2.

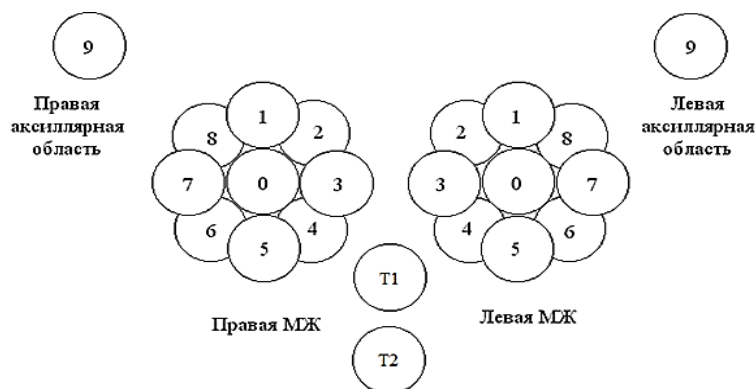


Рис 2. - Схема обследования молочной железы.

Достоинствами данного метода является его неинвазивный характер и абсолютная безвредность для организма. Хотелось бы применять данный метод при ежегодных профилактических медицинских осмотрах с целью выявления опухолей на ранней стадии. Сдерживающим фактором является то, что для постановки диагноза необходим врач – маммолог высокой квалификации, активно применяющий данный метод, а таких врачей считанные единицы. Выход из данной ситуации видится в создании консультативной информационной системы с элементами искусственного интеллекта, которая наряду с постановкой предварительного диагноза содержит механизм обоснования предлагаемого решения на языке, понятном медику. Работа над созданием данной системы активно ведется в ВолГУ (см. [5], [6], [9]). Подход к созданию такой системы основан на формализации знаний врачей и нахождении соответствующих высокоинформативных диагностических признаков (см. [6], [9]).

Качественные медицинские признаки и их одномерные математические модели.

Как уже отмечалось, каждому пациенту ставится в соответствие вектор из 44 температур (t_0, \dots, t_{43}). Физический смысл каждой компоненты данного вектора показан в Таблице 1.

Таблица № 1.

Физический смысл компонентов вектора температур.

Обозначение	Точка, в которой измеряется температура	РТМ или ИК температура
t_0	Левый сосок	РТМ
t_{20}	Правый сосок	РТМ
t_{10}	Левый сосок	ИК
t_{30}	Правый сосок	ИК
t_i , где $i=1, \dots, 9$.	Область с номером i на Рис. 2 для левой железы	РТМ
t_{i+20} , где $i=1, \dots, 9$.	Область с номером i на Рис. 2 для правой железы	РТМ
t_{i+10} , где $i=1, \dots, 9$	Область с номером i на Рис. 2 для левой железы	ИК
t_{i+30} , где $i=1, \dots, 9$	Область с номером i на Рис. 2 для правой железы	ИК
t_{40}, t_{41}	Точки T1 и T2	РТМ
t_{42}, t_{43}	Точки T1 и T2	ИК

Можно проводить диагностику непосредственно по температурам, но это приводит к не очень хорошим результатам [9]. Гораздо продуктивней оказывается подход, связанный с использованием информации, выявленной специалистами - маммологами. В процессе проведения обследования

врачами [3] были выявлены качественные признаки рака молочной железы. К таким признакам относится термоасимметрия, повышенная температура соска по сравнению с остальными точками, повышенная разница кожной и глубинной температуры (внутренний градиент) и некоторые другие. Для каждого из них предложены одномерные математические признаки [6],[9], состоящие из моделирующей функции и промежутка, на котором изменяются значения данной функции. Если значение функции для конкретного пациента попадает в промежуток, признак считается выполненным. Как показано в [9], признаки можно делить на характеристические и высокоинформативные. Если для конкретного пациента выполнен характеристический признак, система считает его больным. Если выполняется высокоинформативный признак, то пациент не считается автоматически больным, но выполнение каждого такого признака повышает вероятность того, что пациент болен. Важнейшим признаком высокоинформативного признака является его информативность, показывающая, насколько хорошо признак отделяет больных пациентов от здоровых. Различают ([9]) статистическую, эвристическую, энтропийную и некоторые другие виды информативности.

Промежуток для каждого признака подбирается с помощью вычислительных экспериментов на обучающей выборке с целью увеличения информативности признака [9]. Очевидно, что для каждого качественного признака можно построить достаточно большой набор моделирующих функций. Например, при анализе повышенной термоасимметрии можно брать разность температур между одноименными точками, средние арифметические разностей температур между различными областями, разницу температур сосков, дисперсию разностей температур между одноименными точками молочной железы и т.п. В данный момент участниками проекта используется более тысячи одномерных признаков, разработана технология построения

минимизирующих наборов таких признаков, на их основе построены и проанализированы различные алгоритмы классификации [9]. В данной работе используются 6 моделирующих функций (см. Таблицу 2) соответствующих одномерных признаков.

Таблица № 2

Обобщенные одномерные признаки, используемые при построении двумерных

Название	Моделирующая функция	Признак
Thermoassimetry_RTМ	$\max_{j=0,\dots,8} \epsilon_j - \epsilon_{j+20} $	Термоассиметрия
Thermoassimetry_IR	$\max_{j=10,\dots,18} \epsilon_j - \epsilon_{j+20} $	
Increased_Temperature_t0_RTМ	$\max_{j,k} \epsilon_j - \epsilon_k , j = 0,20; k = j + 1, \dots, j + 8$	Повышенная разность температур соска и отдельных точек железы
Increased_Temperature_t0_IR	$\max_{j,k} \epsilon_j - \epsilon_k , j = 10,30; k = j + 1, \dots, j + 8$	
Inner_Gradient	$\max_{j,k} \epsilon_{j+k} - \epsilon_{j+k+10} , j = 0,20; k = 0, \dots, 8$	Внутренний градиент
Second_Derivative	$\max_{j,k} \left (\epsilon_j - \epsilon_{j+10}) - \frac{(\epsilon_{j+k} - \epsilon_{j+k+10}) + (\epsilon_{j+k+1} - \epsilon_{j+k+10+1})}{3} \right $ $\left (\epsilon_j - \epsilon_{j+10}) - \frac{(\epsilon_{j+7} - \epsilon_{j+7+10}) + (\epsilon_{j+8} - \epsilon_{j+10+8})}{3} \right - (\epsilon_{j+1} - \dots)$	Признаки, базирующиеся на физико-математич

	$(t_j - t_{j+10}) - \frac{(c_{j+8} - c_{j+8+10}) + (c_{j+12} - c_{j+12+10})}{3} - (c_j)$	<p>еских моделях поведения температу рных полей</p>
--	--	---

Двумерные признаки. Алгоритм создания двумерных признаков методом сужающихся эллипсов.

Двумерный признак [5] характеризует взаимное влияние двух качественных медицинских признаков друг на друга. Обозначим $x_j = f_j^{q_1}$, $y_j = f_j^{q_2}$ значения моделирующих функций двух одномерных признаков на j -м векторе обучающей выборки. Рассмотрим множество точек $P = \{(x_1, y_1), \dots, (x_k, y_k)\}$, характеризующее здоровых пациентов и множество $Q = \{(x_{k+1}, y_{k+1}), \dots, (x_n, y_n)\}$, отвечающее за больных. Выберем область X координатной плоскости, которая содержит все точки из P . При этом желательно, чтобы как можно больше точек множества Q находилось вне области X . Пациента из тестирующей выборки будем считать больным, если ему соответствует точка, находящаяся вне области X . Таким образом мы построили двумерный характеристический признак. В качестве областей ранее [5] предлагались выпуклые многоугольники и различные виды эллипсов. В данной работе в качестве X строился с помощью известного алгоритма Хачияна [10] эллипс минимальной площади, содержащий все точки из P .

Комбинируя шесть функций из таблицы 2, получаем 15 двумерных характеристических признаков. Чтобы получить высокоинформативные признаки, будем варьировать (уменьшать и увеличивать) полуоси эллипса Хачияна. В качестве показателей, характеризующих качество признака,

будем рассматривать его чувствительность $Sens$ (долю верно диагностированных больных) и специфичность $Spec$ (долю верно диагностированных здоровых). Под эффективностью G будем понимать среднее геометрическое указанных величин.

Проводились вычислительные эксперименты с целью построения двумерных вычислительных признаков с максимальной эффективностью. Алгоритм построения наилучшего эллипса состоит в следующем:

- 1) Алгоритмом Хачияна строим эллипс минимальной площади, содержащий все точки из P .
- 2) Для каждой точки находим сумму расстояние от неё до всех остальных, исключаем точку с наибольшей суммой.
- 3) Перестраиваем эллипс по оставшимся точкам.
- 4) Методом равномерного поиска, изменяя коэффициенты сжатия полуосей от 1 до 0.8, ищем эллипс с наибольшей эффективностью и сохраняем его. Если эффективность сжатой области выше, чем у не сжатой, построенной в пункте 3, сохраняем его, в противном случае сохраняем не сжатый эллипс.
- 5) Если не достигнуто минимальное число точек, по которому можно построить эллипс, из оставшихся точек исключаем точку с наибольшей суммой расстояний и переходим к пункту 4.
- 6) Из сохранённых в пункте 5 эллипсов выбираем эллипс с наибольшей эффективностью.

Данным алгоритмом построены 15 двумерных высокоинформативных признаков. В таблице №3 приводятся результаты вычислительных экспериментов, позволяющих найти чувствительность, специфичность и эффективность данных признаков на тестовой выборке.

Таблица № 3

Двумерные признаки, построенные методом сужающихся эллипсов

Названия пары одномерных признаков	Sens	Spec	G
Thermoassimetry_RTМ Thermoassimetry_IR	0,74	0,89	0,81
Thermoassimetry_RTМ Increased_Temperature_t0_RTМ	0,93	0,64	0,77
Thermoassimetry_RTМ Increased_Temperature_t0_IR	0,63	0,92	0,75
Thermoassimetry_RTМ Inner_Gradient	0,74	0,83	0,79
Thermoassimetry_RTМ Second_Derivative	0,89	0,83	0,86
Thermoassimetry_IR Increased_Temperature_t0_RTМ	0,89	0,69	0,79
Thermoassimetry_IR Increased_Temperature_t0_IR	0,56	0,89	0,70
Thermoassimetry_IR Inner_Gradient	0,59	0,92	0,74
Thermoassimetry_IR Second_Derivative	0,78	0,89	0,83
Increased_Temperature_t0_RTМ Increased_Temperature_t0_IR	0,70	0,97	0,83
Increased_Temperature_t0_RTМ Inner_Gradient	0,93	0,56	0,72
Increased_Temperature_t0_RTМ Second_Derivative	0,89	0,94	0,92
Increased_Temperature_t0_IR Inner_Gradient	0,48	0,86	0,64
Increased_Temperature_t0_IR Second_Derivative	0,78	0,78	0,78
Inner_Gradient Second_Derivative	0,81	0,86	0,84

Видно, что лучшие признаки обеспечивают эффективность порядка 80%.

Эллипсы рассеивания.

Рассмотрим статистический подход, позволяющий построить другое семейство двумерных высокоинформативных признаков на основе эллипсов рассеивания [11]. Будем считать x_1, \dots, x_k и y_1, \dots, y_k значениями двух нормально распределенных случайных величин X и Y . Обозначим m_x, m_y математические ожидания случайных величин X и Y , σ_x, σ_y - их среднеквадратичные отклонения, r - коэффициент корреляции X и Y . Построим однопараметрическое семейство эллипсов рассеивания, задаваемых уравнением (см [11])

$$\frac{(x - m_x)^2}{\sigma_x^2} + \frac{2r(x - m_x)(y - m_y)}{\sigma_x \sigma_y} + \frac{(y - m_y)^2}{\sigma_y^2} = c^2.$$

Варьируя параметр c , найдем эллипс, обеспечивающий максимальную эффективность.

В таблице 4 приведены результаты вычислительных экспериментов, позволяющие оценить чувствительность, специфичность и эффективность двумерных признаков данного семейства на тестовой выборке.

Таблица № 4

Двумерные признаки, построенные методом эллипсов рассеяния.

Названия пары одномерных признаков	Sens	Spec	G
Thermoassimetry_RTM Thermoassimetry_IR	0,77	0,75	0,76
Thermoassimetry_RTM Increased_Temperature_t0_RTM	0,79	0,73	0,76
Thermoassimetry_RTM Increased_Temperature_t0_IR	0,80	0,69	0,75
Thermoassimetry_RTM Inner_Gradient	0,73	0,80	0,76
Thermoassimetry_RTM Second_Derivative	0,82	0,80	0,81
Thermoassimetry_IR Increased Temperature_t0_RTM	0,74	0,72	0,73
Thermoassimetry_IR Increased_Temperature_t0_IR	0,74	0,64	0,69
Thermoassimetry_IR Inner_Gradient	0,71	0,81	0,76
Thermoassimetry_IR Second_Derivative	0,77	0,87	0,82
Increased_Temperature_t0_RTM Increased_Temperature_t0_IR	0,81	0,70	0,76
Increased_Temperature_t0_RTM Inner_Gradient	0,72	0,74	0,73
Increased_Temperature_t0_RTM Second_Derivative	0,69	0,88	0,78
Increased_Temperature_t0_IR Inner_Gradient	0,65	0,75	0,70
Increased_Temperature_t0_IR Second_Derivative	0,80	0,75	0,77
Inner_Gradient Second_Derivative	0,67	0,84	0,75

Анализируя таблицы 3 и 4, нетрудно заметить, что по сравнению с эллипсами рассеяния метод сужающихся эллипсов дает почти для всех пар одномерных признаков выигрыш в эффективности в 3-5%.

Вычислительные эксперименты с целью построения настраиваемого классификатора.

Эффективность классификатора зависит от двух параметров – чувствительности и специфичности, при этом увеличение одного из них зачастую приводит к уменьшению другого. Во многих ситуациях система должна дать медику возможность задавать пороговое значение одного из этих параметров (например, специфичности) и подобрать набор признаков, обеспечивающий максимальную эффективность другого. С этой целью была проведена серия вычислительных экспериментов, продемонстрировавших возможность создания такого настраиваемого медиком классификатора.

Для этой цели данные 15 признаков, полученных методом сужающихся эллипсов, использовались для построения классификатора стандартным методом линейной регрессии. При этом варьировались полуоси полученных эллипсов, в результате было построено однопараметрическое семейство из трехсот классификаторов. Чтобы построить классификатор, отвечающий значению параметра k , полуоси эллипсов Хачияна в каждом из наших 15 признаков умножаем на k , после чего методом линейной регрессии строим классификатор и оцениваем его параметры на тестовой выборке. При оценивании использовалась кросс – валидация (данные делились на n частей, на $n-1$ -й части проводилось обучение, оставшаяся часть использовалась в качестве тестовой). Значение параметра k варьировалось от 0.01 до 3 с шагом 0.01. На рис. 3 изображен график зависимости чувствительности от специфичности для такого набора классификаторов.

2. Gonzalez F.J. Thermal simulation of breast tumors // Revista Mexicana de fisica, 2007. Vol. 53. pp.323-326.
 3. Веснин С.Г., Каплан М.А., Авакян Р.С. Современная микроволновая радиотермометрия молочных желез // Опухоли женской репродуктивной системы. 2008. № 3. С. 28-35.
 4. Кобринский Б.А. Консультативные интеллектуальные медицинские системы: классификация, принципы построения, эффективность // Врач и информационные технологии. 2008. № 2. С. 38-47.
 5. Зенович А.В., Батурин Н.А., Медведев Д.А., Петренко А.Ю. Алгоритмы формирования двумерных признаков диагностики заболеваний молочных желез методами комбинированной радиотермометрии // Математическая физика и компьютерное моделирование. 2018. Т. 21. № 4. С. 44-56.
 6. Глазунов В.А., Зенович А.В., Опарин А.С., Примаченко Ф.Г. Алгоритмы принятия решений в консультативной интеллектуальной системе диагностики молочных желез // Вестник Волгоградского ун-та. Сер. 1. Мат. Физ. 2016. № 6 (37). С. 141-154.
 7. 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
 8. Малышев И.В. Методы микроволновой регистрации и локации биологических дисперсных сред. // Инженерный вестник Дона, 2015, №4. URL: ivdon.ru/ru/magazine/archive/n4y2015/3485
 9. Лосев А.Г., Зенович А.В., Бочкарев О.А., Левшинский В.В. Интеллектуальный анализ многомерных термометрических данных в медицинской диагностике // Вестник Волгоградского ун-та. Сер. 1. Мат. Физ. 2016. № 5 (36). С. 150-161.
-

10. Khachiyan, L. G. Rounding of Polytopes in the Real Number Model of Computation // Mathematics of Operations Research, 1996, vol. 21(2), pp. 307–320.
11. Вентцель Е.С. Теория вероятностей: Учеб. для вузов. 6-е изд. стер. Москва: Высшая школа, 1999. 576 с.

References

1. Barrett A.H., Myers P.C., Sadowsky N.L. American Journal of Roentgenology, 1980. Vol. 34 (2). pp.365-368.
 2. Gonzalez F.J. Revista Mexicana de fisica, 2007. Vol. 53. pp.323-326.
 3. Vesnin S.G., Kaplan M.A., Avakyan R.S. Opukholi zhenskoy reproduktivnoy sistemy. 2008. No3. pp. 28-35.
 4. Kobrinskiy B.A. Vrach i informatsionnye tekhnologii. 2008. No2. pp. 38-47.
 5. Zenovich A.V., Baturin N.A., Medvedev D.A., Petrenko A.Yu. Matematicheskaya fizika i komp'yuternoe modelirovanie. 2018. V. 21. №4. pp. 44-56.
 6. Glazunov V.A., Zenovich A.V., Oparin A.S., Primachenko F.G. Vestnik Volgogradskogo gosudarstvennogo universiteta. Seriya 1: Matematika. Fizika. 2016. № 6 (37). pp. 141-154.
 7. Polyakov M.V., Khoperskov A.V. Inženernyj vestnik Dona (Rus), 2017, №2. URL: ivdon.ru/ru/magazine/archive/n2y2017/4239
 8. Malyshev I.V. Inženernyj vestnik Dona (Rus), 2015, №4. URL: ivdon.ru/ru/magazine/archive/n4y2015/3485
 9. Losev A.G., Zenovich A.V., Bochkarev O.A. Vestnik Volgogradskogo gosudarstvennogo universiteta. Seriya 1: Matematika. Fizika. 2016. No 5 (36). pp. 150-161.
 10. Khachiyan, L. G. Mathematics of Operations Research, 1996, vol. 21(2), pp. 307–320.
-



11. Venttsel' E.S. Teoriya veroyatnostey: Ucheb. dlya vuzov. [Probability Theory: Textbook for universities] 6-e izd. ster. Moskva: Vysshaya shkola, 1999. 576 p.