

Применение машинного обучения для повышения эффективности диагностики в медицине на основе метода радиотермометрии

М.В. Поляков¹, А.В. Хоперсков¹, Е.И. Борисовский¹

¹Волгоградский государственный университет, Университетский 100, Волгоград, Россия, 400062

Аннотация. В данной работе мы используем технологии машинного обучения для повышения эффективности медицинской диагностики на основе метода радиотермометрии. Оригинальность нашего подхода заключается в том, что для построения обучающих и тестовых наборов данных используются результаты компьютерного моделирования температурных полей в многокомпонентных биологических тканях. Мы исследуем границы применимости метода диагностики раковых заболеваний молочных желез по данным микроволновой радиотермометрии, в частности, определяем качество диагностики при различных размерах опухоли.

1. Введение

Ранняя диагностика рака молочной железы является важной проблемой современной медицины. Заболеваемость раком молочной железы растет во всем мире, являясь одним из наиболее распространенных видов рака. На данный момент нет эффективного средства для предотвращения рака молочной железы. Вероятность успешного лечения и полного выздоровления пациента полностью зависит от раннего выявления и диагностики. При раннем обнаружении рак молочной железы является излечимым заболеванием с вероятностью 97% [10].

Метод радиомикроволновой термометрии – неинвазивный метод исследования внутренней температуры в биологических тканях. Метод основан на измерении собственного излучения биотканей в радиомикроволновом диапазоне. Впервые возможность применения микроволновой радиотермометрии для диагностики рака молочной железы была показана в [6]. В какой-то степени можно считать, что теоретические основы метода микроволновой радиотермометрии в маммологии базируются на исследованиях французского ученого М. Gautherie [7]. Основываясь на клинических данных более чем 85000 пациенток, он убедительно показал, что термовыделение опухоли прямо пропорционально скорости ее роста. Таким образом, микроволновая радиотермометрия обладает уникальной возможностью обнаруживать в первую очередь быстро растущие опухоли. Помимо диагностики онкологических заболеваний молочных желез метод микроволновой термометрии применяется и в других областях медицины [1, 4].

В работе [8] утверждается, что минимальный размер раковой опухоли, обнаруживаемой при маммографии – 1.68 см в диаметре. Задача радиотермометрии заключается в обнаружении опухолей меньшего размера. Метод микроволновой термометрии может также

выявить раковые опухоли или ранние структурные изменения, которые не обнаруживаются и могут быть пропущены при использовании маммографии.

2. Постановка задачи

В работах [3, 11, 12] описаны математические модели и численные методы, используемые нами для построения выборки температурных данных. Была проведена верификация моделей [9], которая показала эффективность построения моделей молочных желез здоровых пациентов (без раковых патологий). Основной задачей нашего текущего исследования является определение порогового значения размера опухоли, которую возможно выявить методом радиомикроволновой термометрии. Для этого необходимо построить выборки данных компьютерного моделирования температур молочных желез, объем которых позволит проводить машинное обучение, а также бинарную классификацию тестовых данных (здоров-болен). При построении большого объема моделей необходимо учитывать тот факт, что наиболее часто злокачественные опухоли появляются в верхнем наружном квадранте.

3. Методы и инструменты

В качестве методов бинарной классификации данных компьютерного моделирования применяются: метод опорных векторов (SVM), метод ближайших соседей (KNN), наивный байесовский классификатор (НБК). Данные методы реализованы в библиотеке языка Python Scikit-learn. Помимо задачи классификации данная библиотека позволяет строить регрессии, проводить кластеризацию и т.д.

Обучающая выборка представляет собой набор данных температур в точках молочной железы (согласно методике обследования [2]) в микроволновом диапазоне (глубинная температура) и в инфракрасном диапазоне (кожная температура). Каждая модель имеет флаг «1» – болен или «0» – здоров. Были построены равные по объему выборки данных, содержащие по 80 моделей: без опухоли, с опухолью радиусов $R=0.5$ см, $R=0.75$ см и $R=1.0$ см (рисунок 1).

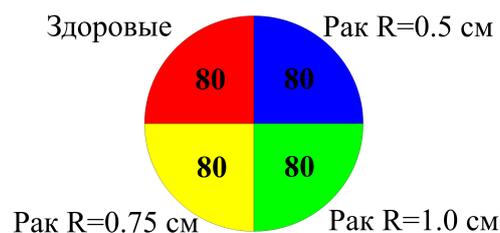


Рисунок 1. Структура исходного набора данных.

Был сделан срез с кожными и глубинными температурами точек $0, \dots, 8$ и мы имеем комбинированный набор данных

$$X = \begin{bmatrix} T_0^1 & T_1^1 & \dots & T_{18}^1 \\ T_0^2 & T_1^2 & \dots & T_{18}^2 \\ \dots & \dots & \dots & \dots \\ T_0^{240} & T_1^{240} & \dots & T_{18}^{240} \end{bmatrix}, Y = \begin{bmatrix} y_1 \\ y_2 \\ \dots \\ y_{240} \end{bmatrix}, \quad (1)$$

где T_0^i, \dots, T_8^i – внутренние температуры в точках $0, \dots, 8$, T_9^i, \dots, T_{18}^i – температуры кожи в точках $0, \dots, 8$ и $y_i \in \{\text{Healthy}, R=0.5, 0.75, 1.0\}$ – это метка i -ой модели.

В обучающей и тестовой выборках соотношение больных и здоровых принято равным, для сохранения однородности данных. На первой стадии был проведен статистический анализ данных, который показал общие отличия между выборками. Особую сложность вызывает проведение качественного анализа большого объема данных. Статистический анализ подтверждает значимые различия в некоторых точках молочной железы (рисунок 2).

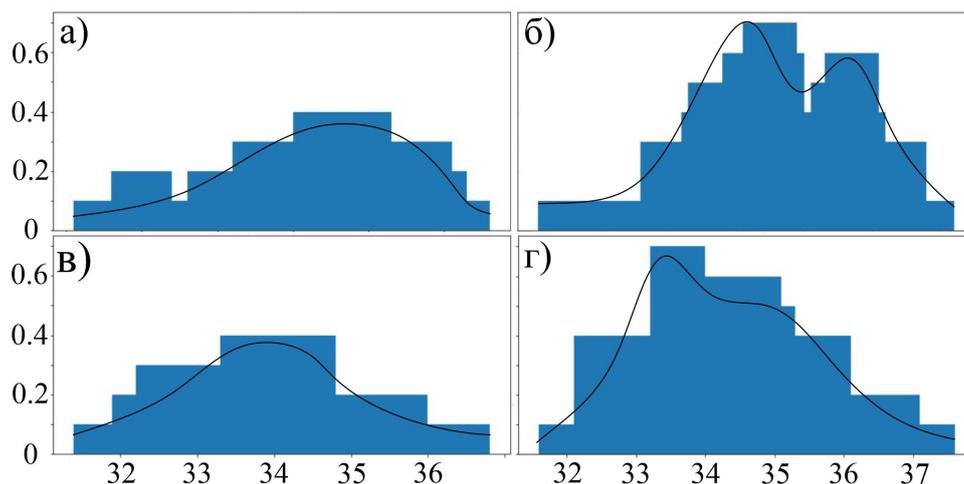


Рисунок 2. Частотное распределение данных внутренней температуры для а) точки «0» модели без опухоли, б) точки «0» модели с опухолью $R = 0.75$ см, в) точки «3» модели без опухоли, г) точки «3» модели с опухолью $R = 0.75$ см.

4. Обсуждение результатов

Эксперименты показали существенную зависимость эффективности диагностики по данным радиомикроволновой термометрии от размера опухоли. Метод опорных векторов дал наилучший результат по отношению к другим методам машинного обучения (таблица 1), что говорит о том, что данный метод лучше применим к такого рода задачам и к такой структуре обучающей выборки. Для набора данных с опухолью $R=0.5$ см выигрыш метода SVM по отношению к НБК равен 10%, что для задачи медицинской диагностики является существенным. В качестве меры эффективности для сравнения методов использовалось отношение верно классифицированных моделей к объему тестовой выборки, т.е. $E = \frac{\omega}{F}$, где ω – количество верно распознанных моделей, F – объем тестовой выборки.

Таблица 1. Эффективность методов классификации данных.

	НБК	KNN	SVM
$R=0.5$ см	0.475	0.525	0.575
$R=0.75$ см	0.7	0.675	0.725
$R=1$ см	0.74	0.75	0.79

Для определения зависимости эффективности диагностики от размера опухоли нами использовалась бинарная классификация «Здоровые» и «Рак», Связано это с тем, что наибольший интерес представляет корректное выявление именно злокачественных опухолей.

Мерой эффективности медицинской диагностики принято считать среднее геометрическое чувствительности и специфичности

$$G = \sqrt{L \cdot S}, \quad (2)$$

где $L = \frac{TP}{TP + FN}$, $S = \frac{TN}{TN + FP}$, TP – доля верно классифицированных желез класса «Рак», FN – доля неверно классифицированных желез класса «Рак», TN – доля верно классифицированных желез класса «Здоровые», FP – доля неверно классифицированных желез класса «Здоровые».

Таблица 2. Чувствительность, специфичность и эффективность метода SVM для различных выборок данных.

	L	S	G
$R=0.5$ см	0.68	0.49	0.577
$R=0.75$ см	0.78	0.725	0.75
$R=1$ см	0.82	0.76	0.79

Рассчитанные показатели чувствительности и специфичности (таблица 2) достаточно высоки. При этом следует отметить ограниченное только температурами признаковое пространство и достаточно небольшие обучающие выборки.

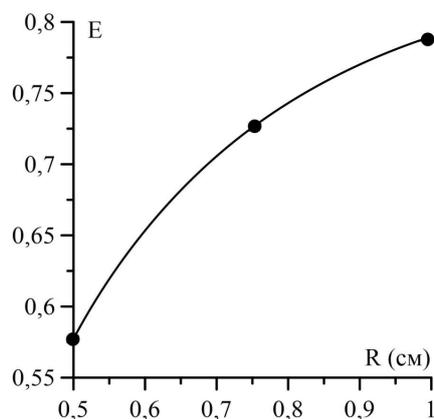


Рисунок 3. Зависимость эффективности метода диагностики по данным радиомикроволновой термометрии от размера опухоли.

Эффективность диагностики возрастает с увеличением радиуса опухоли (рисунок 3). Даже для малых опухолей радиуса $R=0.5$ см с вероятностью 57.5% возможно правильное определение класса. Можно ожидать успешное приложение метода радиомикроволновой термометрии и для опухолей меньшего размера, с расширением признакового пространства, увеличением объема выборки и применением эвристик.

5. Благодарности

Работа выполнена в рамках ГЗ Минобрнауки «Разработка параллельных алгоритмов и создание программного обеспечения для моделирования нелинейной динамики сплошных сред» (проект 0633-2020-0003). М.В. Поляков благодарит РФФИ за финансовую поддержку в рамках научного проекта №19-37-90142.

6. Литература

- [1] Гудков, А.Г. Использование метода многоканальной микроволновой радиометрии для функциональной диагностики головного мозга / А.Г. Гудков, В.Ю. Леушин, И.А. Сидоров, С.Г. Веснин, И.О. Порохов, М.К. Седанкин, С.В. Агасиева, С.В. Чижиков, Е.Н. Горлачева, М.И. Лазаренко, В.Д. Шашурин // *Медицинская техника*. – 2019. – Т. 2, № 314. – С. 22-25.
- [2] Лосев, А.Г. О некоторых характерных признаках в диагностике патологии молочных желез по данным микроволновой радиотермометрии / А.Г. Лосев, Е.А. Мазепа, Т.В. Замечник // *Современные проблемы науки и образования*. – 2014. – Т. 6. – С. 254-261.
- [3] Поляков, М.В. Математическое моделирование пространственного распределения радиационного поля в биоткани: определение яркостной температуры для диагностики / М.В. Поляков, А.В. Хоперсков // *Вестник Волгоградского государственного университета. Серия 1: Математика. Физика*. – 2016. – Т. 5, № 36. – С. 73-84.
- [4] Седанкин, М.К. Моделирование собственного теплового излучения почки в микроволновом диапазоне / М.К. Седанкин, В.Ю. Леушин, А.Г. Гудков, С.Г. Веснин, Д.А. Хромов, И.О. Порохов, И.А. Сидоров, С.В. Агасиева, Е.Н. Горлачева // *Медицинская техника*. – 2019. – Т. 1, № 313. – С. 44-47.
- [5] Седанкин, М.К. Антенны-аппликаторы для медицинских микроволновых радиотермографов / М.К. Седанкин, В.Ю. Леушин, А.Г. Гудков, С.Г. Веснин, И.А. Сидоров, С.В. Агасиева, Л.М. Овчинников, Н.А. Ветрова // *Медицинская техника*. – 2018. – Т. 4, № 310. – С. 13-15.
- [6] Barrett, A.H. Subcutaneous Temperature: A method of Noninvasive Sensing / A.H. Barrett, P.C. Myers // *Science*. – 1975. – Vol. 190. – P. 669-671.
- [7] Gautherie, M. Temperature and Blood Flow Patterns in Breast Cancer During Natural Evolution and Following Radiotherapy / M. Gautherie // *Biomedical Thermology*. – 1982. – P. 21-64.
- [8] Keyserlingk, J.R. Functional infrared imaging of the breast / J.R. Keyserlingk, P.D. Ahlgren, E. Yu, N. Belliveau, M. Yassa // *IEEE Eng Med. Biol. Mag.* – 2000. – Vol. 19(3). – P. 30-41.
- [9] Levshinskii, V. Verification and Validation of Computer Models for Diagnosing Breast Cancer Based on Machine Learning for Medical Data Analysis / V. Levshinskii, M. Polyakov, A. Losev, A. Khoperskov // *Communications in Computer and Information Science*. – 2019. – Vol. 1084. – P. 447-460.
- [10] Ng, E.Y.K. A review of thermography as promising non-invasive detection modality for breast tumor / E.Y.K. Ng // *Int J Therm Sci*. – 2009. – Vol. 48(5). – P. 849-859.
- [11] Polyakov, M.V. Numerical Modeling of the Internal Temperature in the Mammary Gland / M.V. Polyakov, A.V. Khoperskov, T.V. Zamechnic // *Lecture Notes in Computer Science*. – 2017. – Vol. 10594. – P. 128-135.
- [12] Polyakov, M.V. Simulation Modeling of radiation field in biological tissue of mammary gland / M.V. Polyakov, A.V. Khoperskov, A.V. Svetlov // *Journal of Computational and Engineering Mathematics*. – 2017. – Vol. 4(3). – P. 3-10.

The use of machine learning to improve the effectiveness of diagnostics in medicine based on the method of radiothermometry

M.V. Polyakov¹, A.V. Khoperskov¹, E.I. Borisovskii¹

¹Volgograd State University, University 100, Volgograd, Russia, 400062

Abstract. We use machine learning technologies to increase the effectiveness of medical diagnostics in this work. The originality of our approach lies in the fact that computer models are used to build training and test data sets. They allow the calculation of temperature fields in multicomponent biological tissues. We investigate the applicability of the method for the diagnosis of breast cancer according to microwave radiothermometry. In this case, we determine the quality of diagnosis for various sizes of the tumor.