

## **ПРИБОРЫ, СИСТЕМЫ И ИЗДЕЛИЯ МЕДИЦИНСКОГО НАЗНАЧЕНИЯ**

УДК 519.7

### **ФАКТОРНЫЙ АНАЛИЗ ВЛИЯНИЯ ПРИЗНАКОВ НА ТОЧНОСТЬ ДИАГНОСТИКИ РАКА МОЛОЧНОЙ ЖЕЛЕЗЫ ПО ДАННЫМ МИКРОВОЛНОВОЙ РАДИОТЕРМОМЕТРИИ<sup>1</sup>**

*Статья поступила в редакцию 30.12.2021, в окончательной варианте – 18.02.2022.*

**Гермашев Илья Васильевич**, Волгоградский государственный университет, 400062, Российская Федерация, г. Волгоград, пр. Университетский, 100,

доктор технических наук, профессор, ORCID: 0000-0001-5507-8508, e-mail: germashev@volsu.ru

**Дубовская Виктория Игоревна**, Волгоградский государственный университет, 400062, Российская Федерация, г. Волгоград, пр. Университетский, 100,

аспирант, ORCID: 0000-0002-4670-4682, e-mail: dubovskajav@volsu.ru

**Лосев Александр Георгиевич**, Волгоградский государственный университет, 400062, Российская Федерация, г. Волгоград, пр. Университетский, 100,

доктор физико-математических наук, профессор, ORCID: 0000-0002-1072-8375, e-mail: alexander.losev@volsu.ru

**Попов Илларион Евгеньевич**, Волгоградский государственный университет, 400062, Российская Федерация, г. Волгоград, пр. Университетский, 100,

магистрант, ORCID: 0000-0002-0997-8721, e-mail: popov.larion@volsu.ru

В работе представлена модель оценки влияния диагностических признаков онкологического заболевания на диагноз, поставленный интеллектуальной системой, в рамках конкретного метода классификации. За основу установления уровня влияния взяты факторные нагрузки признаков. Проведен сравнительный анализ факторных моделей, построенных методами главных компонент, максимального правдоподобия и наименьших квадратов.

**Ключевые слова:** факторный анализ, машинное обучение, консультативно-диагностические системы, рак молочной железы

### **FACTOR ANALYSIS OF THE INFLUENCE OF SIGNS ON THE ACCURACY OF BREAST CANCER DIAGNOSIS ACCORDING TO MICROWAVE RADIOTHERMOMETRY**

*The article was received by the editorial board on 30.12.2021, in the final version – 18.02.2022.*

**Germashev Ilya V.**, Volgograd State University, 100 Universitetskiy Prospect, Volgograd, 400062, Russian Federation,

Doct. Sci. (Engineering), Professor, ORCID: 0000-0001-5507-8508, e-mail: germashev@volsu.ru

**Dubovskaya Victoria I.**, Volgograd State University, 100 Universitetskiy Prospect, Volgograd, 400062, Russian Federation,

postgraduate student, ORCID: 0000-0002-4670-4682, e-mail: dubovskajav@volsu.ru

**Losev Alexander G.**, Volgograd State University, 100 Universitetskiy Prospect, Volgograd, 400062, Russian Federation,

Doct. Sci. (Physics and Mathematics), Professor, ORCID: 0000-0002-1072-8375, e-mail: alexander.losev@volsu.ru

**Popov Illarion E.**, Volgograd State University, 100 Universitetskiy Prospect, Volgograd, 400062, Russian Federation,

undergraduate student, ORCID: 0000-0002-0997-8721, e-mail: popov.larion@volsu.ru

The paper presents a model for assessing the impact of diagnostic signs of cancer on the diagnosis within a specific classification method. Factor loads of features are taken as the basis for determining the level of influence. A comparative analysis of factor models constructed by the methods of principal components, maximum likelihood and least squares is carried out.

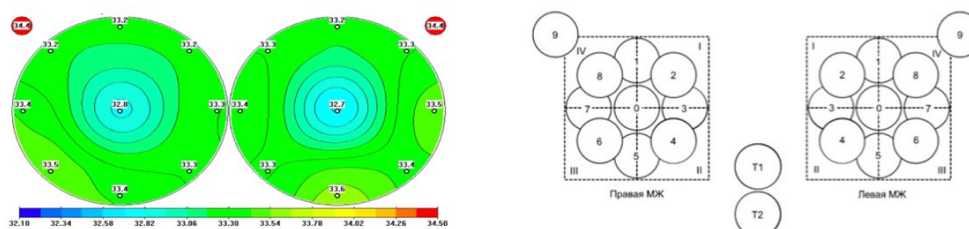
**Keywords:** factor analysis, machine learning, consultative and diagnostic systems, breast cancer

---

<sup>1</sup> Работа выполнена при финансовой поддержке РФФИ, проект № 19-01-00358 «Математические модели радиационных полей и анализа данных микроволновой радиотермометрии в ранней диагностике рака молочных желез».

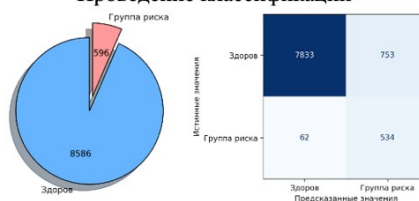
## Графическая аннотация (Graphical annotation)

## Проведение измерений



## Taking measurements

## Проведение классификации



## Classification

**Введение.** Данная работа посвящена улучшению качества консультативных систем в диагностике рака молочной железы методами машинного обучения. По статистике на 2020 г., рак молочной железы является самым распространённым типом рака среди всех онкологических заболеваний у женщин и одним из наиболее распространённых среди всех типов рака [1], при этом данная тенденция носит общий характер [2, 3]. Смертность пациентов с раком молочной железы напрямую зависит от размера опухоли на момент её обнаружения [4]. Чем меньше опухоль, тем выше шанс на выздоровление при соответствующем лечении. На 2020 г. смертность от данного вида рака составляла 30 %. Таким образом, актуальной является задача ранней диагностики рака молочной железы. Одним из перспективных способов решения данной задачи нам видится применение метода микроволновой радиотермометрии (РТМ).

РТМ-метод заключается в определении поверхностных (кожных) и глубинных температур тела за счёт измерения интенсивности его электромагнитного излучения. Такие измерения становятся возможными благодаря тому, что, как и любое нагретое тело, ткани человека излучают электромагнитные колебания по всему спектру. Так, поверхностные ткани излучают колебания в инфракрасном диапазоне, а глубинные температуры – длинноволновые, т. е. микроволновые колебания [5]. Исходя из этого производятся измерения температур в различных диапазонах. Перспективность метода заключается в его неинвазивности, безопасности и возможности ранней диагностики заболеваний, в частности рака молочной железы.

Известно, что образование злокачественной опухоли в молочной железе сказывается на биохимических процессах в железе, из-за чего начинает меняться температурное поле железы [6]. Так, например, одной из характеристических особенностей заболевания является повышенная температура в некоторой области, рядом с опухолью. Благодаря температурным аномалиям такого рода становится возможной диагностика по данным, полученным РТМ-методом. При этом возможной становится как диагностика врачом-диагностом по термограммам молочных желёз, так и автоматическая диагностика методами машинного обучения. По термограммам врач-диагност оценивает состояние температур пациента, наличие температурных аномалий и характерных паттернов для рака молочной железы. На рисунке 1 приведён пример термограммы кожных температур здорового пациента, по которому наблюдаются следующие закономерности: у пациента отсутствуют области с повышенной температурой, а также температурные поля обеих молочных желёз похожи друг на друга, что характерно для здоровых пациентов. Методами же машинного обучения исследуется возможность диагностики по числовым данным температур.

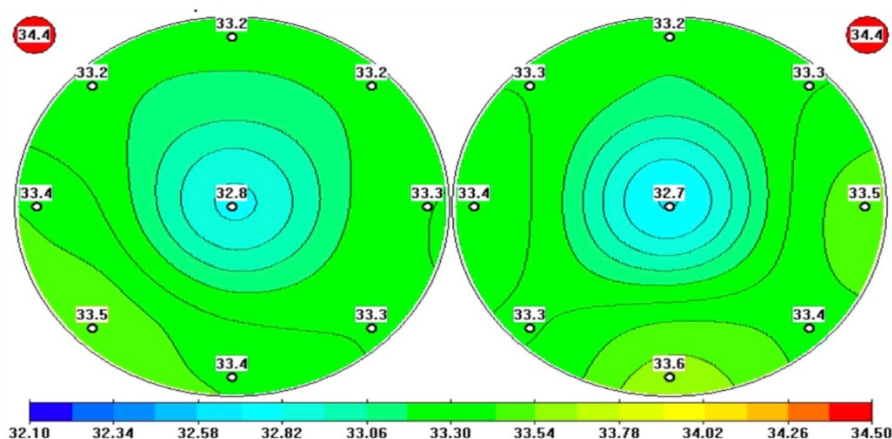


Рисунок 1 – Пример термограммы здорового пациента

Процесс построения интеллектуальных систем с использованием методов машинного обучения по результатам обследования сводится к четырём основным этапам [7]. Сначала строится дескриптивная модель, которая включает в себя гипотезы, характеризующие различные диагностические классы. В случае рака молочной железы одной из актуальных гипотез является наличие у пациентов с опухолью области с повышенной температурой. Затем строится математическая модель, которая заключается в математическом описании составленных гипотез. По математической модели проектируется признаковое пространство, используемое в алгоритмах классификации. Далее происходит построение алгоритмов машинного обучения, которое заключается в настройке и обучении алгоритмов классификации. Полученные алгоритмы классификации показывают высокую точность диагностики пациентов как с раком молочной железы [9], так и пациентов с подозрением на рак (группа риска) [8]. На последнем этапе используется механизм логического вывода результата.

Рассматриваемая концепция улучшения консультативных систем заключается в нескольких направлениях:

- 1) построение обоснования полученного диагноза в результате проведенной классификации на понятном специалисту языке;
- 2) повышение точности и адекватности моделей, лежащих в основе алгоритмов диагностики.

Вышеуказанные задачи могут быть решены в случае определения уровня влияния признаков на правильность/ошибочность поставленного классификатором диагноза. Для этого в ходе данной работы предполагается использование факторного анализа, в основе которого лежит выявление непосредственно ненаблюдаемых и неизмеряемых характеристик (общих факторов) посредством анализа вариаций наблюдаемых переменных и их корреляций.

**Материалы и методы.** Изначальная база данных представляет собой результаты обследований, проведённых медицинскими учреждениями. В процессе обследования температуры измерялись по схеме, показанной на рисунке 2.

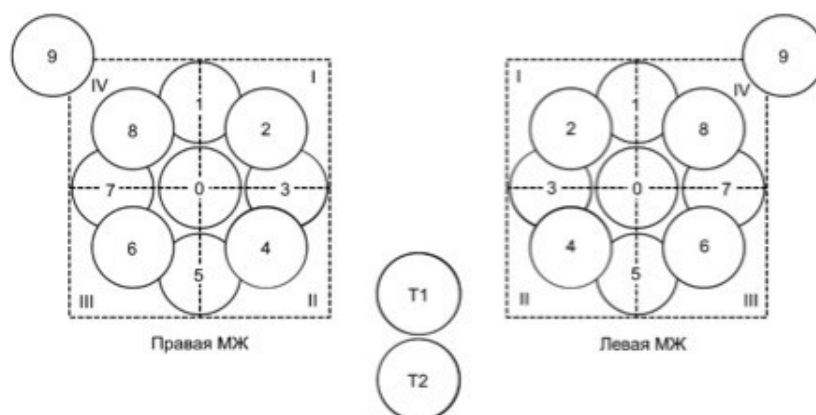


Рисунок 2 – Схема температурных измерений

Помимо температур в базу данных также записывается анамнез, который на данном этапе в алгоритмах классификации не используется, а также диагноз по каждой молочной железе пациента. Опишем данные подробнее:

- $ir_r^i = \{t_{0,ir,r}^i, \dots, t_{9,ir,r}^i\}$  – множество кожных температур правой молочной железы. Здесь и далее первый нижний индекс обозначает область, в которой измеряется температура. Значение индекса соответствует номеру области на рисунке 2.  $i$  – номер обследуемого пациента;
- $mw_r^i = \{t_{0,mw,r}^i, \dots, t_{9,mw,r}^i\}$  – множество глубинных температур правой молочной железы;
- $ir_l^i = \{t_{0,ir,l}^i, \dots, t_{9,ir,l}^i\}$  – множество кожных температур левой молочной железы пациента;
- $mw_l^i = \{t_{0,mw,l}^i, \dots, t_{9,mw,l}^i\}$  – множество глубинных температур левой молочной железы пациента;
- $ir_a^i = \{t_{0,ir,a}^i, t_{1,ir,a}^i\}$  – множество кожных температур опорной области (соответствует точкам T1 и T2 на рисунке 1);
- $mw_a^i = \{t_{0,mw,a}^i, t_{1,mw,a}^i\}$  – множество глубинных температур опорной области пациента;
- $y_r$  – диагноз правой молочной железы;
- $y_l$  – диагноз левой молочной железы.

Так как диагностика производится отдельно по каждой из молочных желез, база данных требует некоторых изменений. Она преобразовывалась таким образом, что каждая запись в ней соответствует описанию одной молочной железы, а не пациенту с обеими железами. Для этого каждая запись дублировалась, и в них вносились следующие изменения:

- оригинальная запись соответствует правой молочной железе;
- дублирующая запись соответствует левой молочной железе;
- в каждой записи остаётся только диагноз соответствующей молочной железы;
- индексы ' $r$ ' и ' $l$ ' заменяются на ' $c$ ' и ' $p$ ', где ' $c$ ' означает, что температуры принадлежат молочной железе записи, ' $p$ ' – температуры принадлежат парной молочной железе пациента.

Таким образом была получена база данных по молочным железам. В ней содержится 8586 записей о здоровых железах и 596 записей о железах, входящих в группу риска (рис. 3). В данную группу входят записи с температурными аномалиями, характерными для молочных желез со злокачественной опухолью.

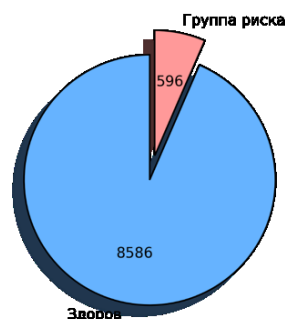


Рисунок 3 – Соотношение классов в базе данных

По полученным данным обследований пациентов специалистами была составлена следующая дескриптивная модель, характеризующая отличительные особенности пациентов из группы риска:

- температурные аномалии, возникающие из-за опухоли, делают значительными различия температурных полей парных молочных желез;
- повышенная разность температур сосков;
- неравномерное распределение температур в молочной железе, наличие «горячих» областей;
- повышенная разность глубинных и кожных температур молочной железы;
- и некоторые другие;

По составленной дескриптивной модели строилось её математическое описание и признаковое пространство. Для этого использовался набор норм и полунорм некоторых функциональных пространств. В результате были построены различные признаки, описывающие обозначенные выше особенности пациентов из группы риска [8].

Для проведения факторного анализа, с целью установления уровня влияния признаков на поставленный диагноз, необходимо было проклассифицировать все записи из базы данных одним из алгоритмов машинного обучения. В качестве такого алгоритма была выбрана логистическая

регрессия. Для того чтобы проклассифицировать каждую запись, решено было использовать метод кросс-валидации. Метод заключается в последовательном обучении и тестировании алгоритма классификации на разных подвыборках, перемешивающихся друг с другом в обучающей выборке. Для этого вся база данных делилась на 10 равных и непересекающихся частей (максимальное различие в их размере – две записи). В каждую часть входит равное количество записей каждого класса. Далее 9 из 10 частей помещались в выборку, по которой обучалась логистическая регрессия. На оставшейся части, не вошедшей в обучающую выборку, проводилась классификация. Результаты классификации сохранялись в итоговую базу данных  $H$ . Таким образом, в ней, помимо реального класса, хранится и класс, поставленный логистической регрессией. После того как все записи из рассматриваемой части были проклассифицированы, алгоритм переобучался на новой обучающей выборке. Новая выборка составлялась таким образом, что одна из частей предыдущей выборки менялась местами с классифицируемой выборкой. Процесс классификации повторялся 10 раз, пока все записи из базы данных не были проклассифицированы. На рисунке 4 приведены результаты классификации в виде матрицы ошибок. Таким образом, логистическая регрессия проклассифицировала верно 90 % здоровых молочных желёз и 88 % желёз из группы риска.

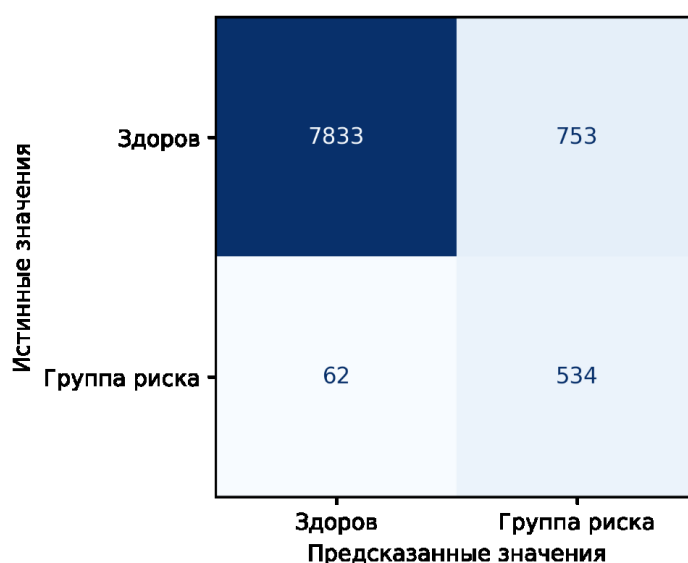


Рисунок 4 – Результаты классификации

На основе полученного результата классификации база  $H$  была разделена на две выборки:

1)  $H^0$ , состоящая из 815 элементов, представляющая собой данные молочных желёз, для которых алгоритм сделал ошибочную классификацию;

2)  $H^1$ , состоящая из 8367 элементов, представляющая собой данные молочных желёз, для которых алгоритм классификации определил верный диагноз.

Для исследования скрытой структуры вышеуказанных выборок без предположения о числе факторов и их нагрузках использовался разведочный факторный анализ.

Поставленная задача была реализована в трех основных этапах:

- подготовка соответствующей матрицы корреляций;
- выделение первоначальных факторов;
- получение окончательного решения с помощью вращения.

Для получения базовой модели использовался метод главных компонент. При детальном рассмотрении анализа главных компонент и, например, метода главных факторов, который является методом факторного анализа, можно отметить их сходство. Направления главных осей определяются путем нахождения собственных чисел и векторов корреляционной матрицы.

Уравнение нахождения собственных чисел и векторов в матричной записи имеет вид:

$$RV = \lambda V, \quad (1)$$

где  $R$  – исходная матрица корреляций;  $V$  – собственный вектор;  $\lambda$  – собственное число.

Решение основано на более простой форме в виде детерминанта матрицы:

$$Det(R - I\lambda) = 0. \quad (2)$$

При проведении факторного анализа используется характеристическое уравнение (2), как и в анализе главных компонент, но в методе главных факторов в качестве  $R$  выступает редуцированная

корреляционная матрица, у которой на главной диагонали вместо единиц расположены оценки общности. Для вычисления последних используют, например, квадрат множественного коэффициента корреляции между данной переменной и совокупностью остальных переменных, или максимальный по абсолютной величине коэффициент корреляции в строке матрицы. После расположения оценок общностей выделяются факторы аналогичным образом [10].

Однако при использовании метода главных компонент происходит поиск факторов, которые объясняют всю изменчивость исходного набора признаков. Далее факторы, которые объясняют наименьшую долю изменчивости, отбрасываются, так как считаются несущественными. При этом факторы представляют собой линейные комбинации исходных переменных. Таким образом, предположений о структуре изучаемых признаков нет.

В основе же факторного анализа лежит гипотеза о том, что изменчивость – результат влияния определенного числа факторов-причин. При этом учитывается и наличие специфических факторов, что позволяет предотвратить потерю информации.

Поэтому с целью приближения модели к реальным данным предлагается использовать такие методы факторного анализа, как метод максимального правдоподобия и наименьших квадратов.

**Результаты исследований.** Для проведения факторного анализа было отобрано 62 диагностических признака. Основанием для выбора данных признаков являлись значения их информативности. В данном контексте термин информативность рассматривается как количественный параметр, определяющий, насколько хорошо закономерность описывает различия между искомой и отделяемой группами.

Рассмотрим исходную матрицу корреляций  $r_{ij}$  между переменными-признаками для выборки  $H^0$ , фрагмент, которой представлен в таблице 1.

Таблица 1 – Фрагмент матрицы корреляций Пирсона

	MG001	MG002	MG003	MG004	MG005	MG006	MG007	MG008
MG001	1	0,977	0,797	0,669	0,036	-0,092	0,350	-0,077
MG002	0,977	1	0,894	0,753	0,036	-0,095	0,394	-0,082
MG003	0,797	0,894	1	0,806	0,034	-0,094	0,430	-0,090
MG004	0,669	0,753	0,806	1	0,025	-0,061	0,477	-0,057
MG005	0,036	0,036	0,034	0,025	1	-0,576	0,308	-0,423
MG006	-0,092	-0,095	-0,094	-0,061	-0,576	1	0,433	0,738
MG007	0,350	0,394	0,430	0,477	-0,308	0,433	1	0,318
MG008	-0,077	-0,082	-0,090	-0,057	-0,423	0,738	0,318	1

Данная матрица не является положительно определенной, что создает затруднения при проведении факторного анализа предпочтительными методами. Одна из возможных причин – определитель матрицы стремится к 0 либо является отрицательным, вследствие линейной зависимости между переменными. Несмотря на то, что факторный анализ нацелен на поиск подобных зависимостей, в случае когда такие зависимости перестают быть вероятностными, становятся жестко детерминированными, многомерные методы анализа могут давать сбой.

Как можно заметить, согласно таблице Чеддока, сила корреляционной связи между многими переменными характеризуется как высокая (значения коэффициентов в таблице 1 больше 0,7).

Для устранения вышеуказанной проблемы на этапе подготовки исходной матрицы предлагается оставить те признаки, для которых  $r_{ij} \leq 0,7$ .

Таким образом, для проведения факторного анализа было отобрано 25 признаков. Фрагмент списка диагностических признаков приведен в таблице 2.

Таблица 2 – Фрагмент списка высокоинформативных диагностических признаков

Название признака	Математическое описание признака	Название признака	Математическое описание признака
MG001	$\ T_c^{i,mw} - T_p^{i,mw}\ _1$	MG025	$\max_{t \in T_c^{i,g}t} - \min_{t \in T_p^{i,g}t}$
MG005	$s(T_c^{i,mw}) - s(T_p^{i,mw})$	MG026	$T_{0,c}^{i,g} - T_{0,p}^{i,g}$
MG006	$\overline{T_c^{i,mw}} - \overline{T_p^{i,mw}}$	MG027	$T_{9,c}^{i,g} - T_{9,p}^{i,g}$
MG007	$\max_{t \in T_c^{i,mw}t} - \min_{t \in T_p^{i,mw}t}$	MG032	$\min_{t \in T_a^{i,g}}  t $
MG008	$T_{0,c}^{i,mw} - T_{0,p}^{i,mw}$	MG036	$\max_{t \in T^{i,mw}}  \overline{T^{i,mw}} - t $
MG010	$\ T_c^{i,ir} - T_p^{i,ir}\ _1$	MG048	$\max_{t \in T^{i,ir} \setminus T_0^{i,ir}}  T_0^{i,ir} - t $
MG014	$s(T_c^{i,ir}) - s(T_p^{i,ir})$	MG053	$\max_{t \in T^{i,ir}}  T_{2,a}^{i,ir} - t $
MG019	$\ T_c^{i,g} - T_p^{i,g}\ _1$	MG054	$\max_{t \in T^{i,g}}  \overline{T^{i,g}} - t $
MG024	$\overline{T_c^{i,g}} - \overline{T_p^{i,g}}$	MG061	$\max_{t \in T^{i,g}}  T_{1,a}^{i,g} - t $
MG062	$\max_{t \in T^{i,g}}  T_{2,a}^{i,g} - t $		

Метод максимального правдоподобия и наименьших квадратов основан на гипотезе, что наблюдаемые данные представляют собой выборку, соответствующую  $k$ -факторной модели. Поэтому для определения числа факторов за основу взята методика, в которой число факторов определено в ходе использования метода главных компонент при помощи критерия Кайзера и Кеттелла. Предполагается, что число факторов определяется числом компонент, собственные значения которых больше 1 (табл. 3).

Таблица 3 – Объясненная совокупная дисперсия

Компонент	Извлечение суммы квадратов нагрузок			Ротация суммы квадратов нагрузок		
	Всего	% дисперсии	Суммарный %	Всего	% дисперсии	Суммарный %
1	4,911	21,351	21,351	3,940	17,131	17,131
2	2,436	10,593	31,944	2,242	9,748	26,879
3	2,252	9,792	41,736	2,217	9,638	36,517
4	2,097	9,119	50,855	2,099	9,127	45,644
5	1,475	6,414	57,269	2,000	8,698	54,342
6	1,458	6,338	63,608	1,649	7,169	61,511
7	1,273	5,534	69,141	1,497	6,507	68,018
8	1,106	4,809	73,950	1,364	5,932	73,950

Таким образом, на основании данных из таблицы 3, число факторов  $k$  принимаем равным 8, и эти факторы описывают 73,9 % совокупной дисперсии признаков.

Далее был проведен факторный анализ методом максимального правдоподобия, в результате которого было извлечено 8 факторов за 15 итераций. Вращение по методу варимакс с нормализацией Кайзера сошлось за 6 итераций. Полученная матрица факторных нагрузок  $c^0$  приведена в таблице 4.

Таблица 4 – Факторная матрица для выборки  $H^0$  после вращения

Признак	1	2	3	4	5	6	7	8
MG065	0,845	0,021	0,174	-0,060	0,182	-0,031	0,020	0,216
MG061	0,842	0,074	-0,086	-0,051	0,120	0,083	-0,027	-0,106
MG025	0,834	0,014	-0,017	0,010	0,128	0,015	0,293	-0,090
MG064	0,741	0,034	0,071	-0,074	0,138	-0,013	0,021	0,220
MG019	0,553	-0,025	0,127	0,346	-0,009	-0,112	-0,043	-0,019
MG048	0,542	0,342	0,090	0,138	0,034	0,539	0,054	-0,078
MG005	0,015	0,920	0,016	0,020	0,001	-0,031	-0,022	0,033
MG014	0,122	0,818	0,012	-0,018	-0,002	0,063	0,012	-0,024
MG008	0,036	-0,450	0,008	-0,086	-0,282	0,214	-0,053	0,015
MG063	-0,233	0,031	-0,961	0,037	-0,134	-0,001	-0,025	-0,030

Аналогичным образом была получена матрица факторных нагрузок с использованием метода наименьших квадратов (извлечено 8 факторов, потребовалось 17 итераций, метод вращения – варимакс с нормализацией Кайзера, вращение сошлось за 7 итераций).

Таблица 5 – Факторная матрица для выборки  $H^0$  после вращения (метод наименьших квадратов)

Признак	1	2	3	4	5	6	7	8
MG065	0,856	0,020	0,177	-0,081	0,177	-0,036	0,022	0,169
MG025	0,834	0,014	-0,019	0,009	0,163	0,063	0,288	-0,082
MG061	0,819	0,089	-0,100	-0,047	0,139	0,131	-0,025	-0,046
MG064	0,725	0,043	0,058	-0,079	0,131	0,002	0,017	0,210
MG019	0,566	-0,026	0,130	0,245	-0,038	-0,127	-0,035	-0,048
MG048	0,527	0,358	0,096	0,141	0,046	0,512	0,012	0,005
MG005	0,017	0,935	0,020	0,009	-0,006	-0,052	-0,030	0,013
MG014	0,115	0,791	-0,018	-0,039	0,049	0,076	-0,025	-0,032
MG008	0,011	-0,428	-0,050	-0,109	-0,129	0,193	-0,065	-0,017
MG063	-0,252	0,029	-0,955	0,055	-0,117	0,049	0,012	-0,028

Полученные факторные нагрузки для выборки  $H^0$  можно интерпретировать как уровень влияния признака на ошибочность результата классификации в рамках конкретного метода машинного обучения.

В процессе проведения факторного анализа также рассчитывались и анализировались следующие показатели:

1) критерий сферичности Бартлетта – так как уровень значимости для каждого из примененного метода меньше 0,05, то можно сделать вывод о приемлемости проведения факторного анализа;

2) КМО (мера адекватности выборки Кайзера – Майера – Олкина) – значения для методов примерно одинаковы и  $\approx 0.65$ , таким образом можно говорить об адекватности проведенного анализа и его применимости к данной выборке.

Таким образом, и один, и другой метод могут быть применены для решения поставленной задачи. Полученные факторные модели являются адекватными и применимыми на практике.

Для проведения дальнейшего анализа использовались факторные нагрузки, полученные с помощью метода максимального правдоподобия. Важно отметить, что переменные, имеющие нулевой вклад в ошибочность решения ( $c_i^0 = 0$ ), одни и те же в результатах, полученных обоими методами. К ним относятся, например, асимметрии внутренних температур в манхэттенской метрике.

Такой же алгоритм получения факторных нагрузок  $c^1$  был использован и для выборки  $H^1$  (табл. 6).



Таблица 6 – Факторная матрица после вращения (выборка  $H^1$ )

Признак	1	2	3	4	5	6	7	8
MG065	0,832	0,120	0,040	0,072	-0,024	-0,024	0,127	0,004
MG061	0,777	0,129	0,076	0,156	0,303	-0,013	-0,011	-0,019
MG025	0,770	0,170	0,072	-0,002	0,144	-0,030	0,003	-0,052
MG064	0,747	0,094	0,029	0,085	0,056	-0,022	0,033	0,143
MG024	0,116	-0,031	0,002	-0,017	-0,020	0,019	-0,017	0,007
MG010	0,158	0,982	0,089	-0,011	-0,026	0,014	-0,023	-0,008
MG001	-0,025	0,768	0,266	-0,007	0,029	0,028	-0,064	0,085
MG019	0,413	0,512	-0,010	-0,017	-0,041	-0,011	0,038	-0,047

Полученные факторные нагрузки для выборки  $H^1$  можно интерпретировать как уровень влияния признака на правильность поставленного диагноза.

Значения  $c^0$  и  $c^1$  позволяют оценить уровень доверия к признаку при постановке диагноза.

На их основе стало возможным разделение признаков на три различные группы:

- 1) признаки, имеющие высокое значение  $c_{ij}^1$ ;
- 2) признаки, имеющие высокое значение  $c_{ij}^0$ ;
- 3) признаки, имеющие высокое значение  $c_{ij}^0$  и  $c_{ij}^1$ .

**Выводы.** Основной вклад в установление правильного диагноза имеют признаки первой группы, для которых уровень влияния  $c_{ij}^0 = 0$ . Таким образом, данные признаки наиболее точно отмечают тепловые изменения в органе. В данной группе преобладают признаки, касающиеся внутренних температур. Следовательно, в обосновании результатов их необходимо поместить в начале иерархии признаков. К таковым признакам относятся: асимметрия внутренних температур по области в манхэттенской метрике, в сосках, в аксиллярной области, асимметрия внутренних градиентов в тех же точках. Проведя классификацию молочных желез, основанную только на этих признаках, получим результат (специфичность 0,88, чувствительность 0,85, эффективность 0,86), который является сопоставимым с представленным в разделе 2, при котором использовались все 62 признака.

Преобладающее количество признаков, для которых уровень влияния  $c_{ij}^1 = 0$ , относится к группе функционалов, моделирующих поведение температурных полей в ИК-диапазоне. Например, среднеквадратическое отклонение и колебание температур молочной железы, отклонение температуры соска, аксиллярной области, отклонение кожных температур в опорных точках. Следовательно, логично поместить данные признаки ниже по иерархии, так как они могут носить не основной, а уточняющий характер.

Для тех признаков, у которых уровень влияния на правильность и ошибочность одинаково высок, предлагается использование нечеткого вывода Мамдани для получения более однозначного решения.

Использование методов максимального правдоподобия и наименьших квадратов является более предпочтительным для данного исследования, так как увеличивается адекватность модели и ее соответствие реальным данным. При этом более точно удастся идентифицировать признаки, имеющие высокий уровень влияния на правильность результата. Если сравнивать предлагаемые здесь методы максимального правдоподобия и наименьших квадратов с методом главных компонент [11], то можно сделать следующие выводы. Количество признаков в первой группе признаков, которые имеют сильное влияние на правильность поставленного диагноза, увеличилось на 6 признаков, во второй группе признаков, имеющих сильное влияние на ошибочность результата классификации, количество осталось неизменным и равно 9.

#### Библиографический список

1. Sung, H. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries / H. Sung, J. Ferlay, R. L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal, F. Bray // CA Cancer J. Clin. – 2021. – Vol. 71. – P. 209–249. – DOI: 10.3322/caac.21660.
2. Bray, F. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries / F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre and A. Jemal // CA: A Cancer Journal for Clinicians. – 2018. – Vol. 68. – P. 394–424. – DOI: 10.3322/caac.21492.
3. Siegel, R. L. Cancer statistics / R. L. Siegel, K. D. Miller and A. Jemal // CA: A Cancer Journal for Clinicians. – 2016. – Vol. 66. – P. 7–30. – DOI: 10.3322/caac.21332.

4. Michaelson, J. S. Predicting the survival of patients with breast carcinoma using tumor size / J. S. Michaelson, M. Silverstein, J. Wyatt, G. Weber, R. Moore, E. Halpern, D. B. Kopans, K. Hughes // *Cancer*. – 2002, Aug 15. – Vol. 95 (4). – P. 713–723. – DOI: 10.1002/cncr.10742.
5. Vesnin, S. Modern Microwave Thermometry for Breast Cancer / S. Vesnin, A. K. Turnbull, J. M. Dixon, I. Goryanin // *Journal of Molecular Imaging & Dynamics*. – 2017. – Vol. 7, iss. 2. – DOI: 10.4172/2155-9937.1000136.
6. Chanmugam, A. Thermal analysis of cancerous breast model / A. Chanmugam, R. Hatwar, C. Herman // *Int. Mech. Eng. Congress Expo*. – 2012. – P. 134–143. – DOI: 10.1115/IMECE2012-88244.
7. Losev, A. G. Intellectual analysis of microwave radiothermometry data in the diagnosis of breast cancer / A. G. Losev, V. V. Levshinsky // *Mathematical physics and computer modeling*. – 2017. – Vol. 20, № 5. – P. 49–62. DOI: 0.15688/mpcm.jvolsu.2017.5.6
8. Levshinskii, V. V. Mathematical models for analyzing and interpreting microwave radiometry data in medical diagnosis / V. V. Levshinskii // *Journal of Computational and Engineering Mathematics*. – 2021. – Vol. 8, № 1. – P. 3–14. – DOI: 10.14529/jcem210101.
9. Losev, A. G. Neural Networks in Diagnosis of Breast Cancer/ A. G. Losev, D. A. Medvedev, A. V. Svetlov // “Smart Technologies” for Society, State and Economy. ISC 2020. Lecture Notes in Networks and Systems. – Springer, Cham., 2021. – Vol. 155. – DOI: 10.1007/978-3-030-59126-7\_25.
10. Окунь, Ян. Факторный анализ / Ян. Окунь ; пер. с польск. Г. З. Давидовича ; науч. ред. В. М. Жуковская. – Москва : Статистика, 1974. – 199 с.
11. Гермашев, И. В. Модель иерархии признаков в диагностике рака молочной железы по данным микроволновой радиотермометрии / И. В. Гермашев, В. И. Дубовская, А. Г. Лосев // *Математические методы в технологиях и технике*. – 2021. – № 6. – С. 78–83.

#### References

1. Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., Bray, F. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.*, 2021, vol. 71, pp. 209–249. DOI: 10.3322/caac.21660.
2. Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A. and Jemal, A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*, 2018, vol. 68, pp. 394–424. DOI: 10.3322/caac.21492.
3. Siegel, R. L., Miller, K. D. and Jemal, A. Cancer statistics, 2016. *CA: A Cancer Journal for Clinicians*, 2016, vol. 66, pp. 7–30. DOI: 10.3322/caac.21332.
4. Michaelson, J. S., Silverstein, M., Wyatt, J., Weber, G., Moore, R., Halpern, E., Kopans, D. B., Hughes, K. Predicting the survival of patients with breast carcinoma using tumor size. *Cancer*, 2002, Aug 15, vol. 95 (4), pp. 713–23. DOI: 10.1002/cncr.10742.
5. Vesnin, S., Turnbull, A. K., Dixon, J. M., Goryanin, I. Modern Microwave Thermometry for Breast Cancer. *Journal of Molecular Imaging & Dynamics*, 2017, vol. 7, iss. 2. DOI: 10.4172/2155-9937.1000136.
6. Chanmugam, A., Hatwar, R., Herman, C. Thermal analysis of cancerous breast model. *Int. Mech. Eng. Congress Expo*, 2012, pp. 134–143. DOI: 10.1115/IMECE2012-88244.
7. Losev, A. G., Levshinsky, V. V. Intellectual analysis of microwave radiothermometry data in the diagnosis of breast cancer. *Mathematical physics and computer modeling*, 2017, vol. 20, no. 5, pp. 49–62. DOI: 0.15688/mpcm.jvolsu.2017.5.6.
8. 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, pp. 3–14. DOI: 10.14529/jcem210101.
9. Losev, A. G., Medvedev, D. A., Svetlov, A. V. Neural Networks in Diagnosis of Breast Cancer. *Smart Technologies for Society, State and Economy. ISC 2020. Lecture Notes in Networks and Systems*. Springer, Cham., 2021, vol. 155. DOI: 10.1007/978-3-030-59126-7\_25.
10. Okun, Yan. *Faktornyy analiz* [Factor analysis]. Moscow, Statistika Publ., 1974. 199 p.
11. Germashev, I. V., Dubovskaya, V. I., Losev, A. G. Model ierarkhii priznakov v diagnostike raka molochnoy zhelezy po dannym mikrovolnovoy radiotermometrii [A model of the hierarchy of signs in the diagnosis of breast cancer according to microwave radiothermometry]. *Matematicheskiye metody v tekhnologiyakh i tekhnike* [Mathematical methods in technology and engineering], 2021, no. 6, pp. 78–83.