

СРАВНИТЕЛЬНЫЙ АНАЛИЗ ПРОГНОСТИЧЕСКОЙ ЗНАЧИМОСТИ ЛОКАЛИЗАЦИИ И РАЗМЕРНЫХ ХАРАКТЕРИСТИК ОПУХОЛИ ДЛЯ ДИАГНОСТИКИ РАКА МОЛОЧНОЙ ЖЕЛЕЗЫ НА ОСНОВЕ АЛГОРИТМОВ МАШИННОГО ОБУЧЕНИЯ

*Алтайский государственный медицинский университет, г. Барнаул
кафедра физики и информатики*

**Дорцвейлер Альберт Александрович, Карачев Илья Игоревич,
Тихонский Николай Дмитриевич, Трухачева Нина Васильевна**
E-mail: k-fizinf@asmu.ru

Научные руководители: Тихонский Н. Д., преподаватель; Трухачева Н. В.,
преподаватель

Ключевые слова: *рак молочной железы, опухоль, диагностика, нейросетевая модель, перцептрон, машинное обучение.*

Актуальность

Рак молочной железы занимает ведущее место среди всех онкологических заболеваний женщин. В последнее время данная патология получила распространение и среди лиц молодого возраста. Раннее выявление рака молочной железы критически важно для своевременного оказания медицинской помощи. Есть определенные клинические и морфологические признаки (возраст, менопауза, размеры и локализация образования, метастазы, в т. ч. в регионарные лимфатические узлы), позволяющие отнести пациента в группу риска возникновения рака молочной железы. Ввиду этого появляется необходимость создать инструмент, позволивший упростить и ускорить диагностику данного заболевания. Разработка нейросетевой модели, учитывающей клинические данные, необходимые для постановки предполагаемого диагноза, поможет

повысить эффективность и объективность первичного скрининга и заложить основу для интеллектуальных диагностических систем в онкологии.

Цель: разработать и оценить прогностическую эффективность модели на основе многослойного перцептрона (Multilayer Perceptron, MLP) для бинарной классификации новообразований молочной железы (доброкачественные/злокачественные) на основе анализа таких клинικο-морфологических признаков, как размер опухоли и ее анатомическая локализация, оценить значимость данных признаков для определения злокачественности новообразования для нейросетевой модели.

Материалы и методы

В исследовании использовался публичный датасет («Breast Cancer Prediction» с платформы Kaggle), содержащий клинико-морфологические данные пациентов с новообразованиями молочной железы. Признаки включали возраст – Age (0-100), наличие менопаузы – Menopause (0-нет, 1-есть), пораженные лимфатические узлы (Inv-Nodes), размер опухоли – Tumor size, cm (0-100), ее локализация – Breast (Right/Left) по квадрантам молочной железы – Breast Quadrant (Upper/Lower outer/inter). Целевой переменной является бинарный признак (Diagnosis Result) доброкачественное/злокачественное (Benign/ Malignant).

Для решения задачи бинарной классификации использовалась нейронная сеть типа многослойный перцептрон (MLP) со следующей архитектурой: входной слой, два скрытых слоя (64 и 32 нейрона) с функцией активации (Rectified Linear Unit, ReLU) и выходной слой. Категориальные признаки, такие как локализация опухоли, были преобразованы в числовой формат методом Label Encoding, а числовые признаки стандартизированы с помощью Z-масштабирования; выборка также была очищена от пропусков и артефактов. Датасет был разделен на обучающую (80%) и тестовую (20%) выборки с применением стратификации по целевому признаку, а оптимизация весовых коэффициентов сети проводилась алгоритмом Adam в течение 2000 эпох.

Прогностическая способность модели была оценена на тестовой выборке с использованием метрик точности (Accuracy), чувствительности (Sensitivity), специфичности (Specificity) и площади под Receiver Operating Characteristic (ROC)-кривой Area Under the Curve (AUC); результаты визуализированы с помощью матрицы ошибок и ROC-кривой (Receiver Operating Characteristic).

Результаты

По результатам валидации на тестовой выборке разработанная нейросетевая модель продемонстрировала высокую прогностическую способность, достигнув показателя общей точности классификации (Accuracy) 80,5%. Детальный анализ структуры ошибок, визуализированный с помощью матрицы неточностей, свидетельствует о высокой специфичности алгоритма, составившей 87,0%. Данный результат указывает на эффективность модели в исключении патологии у здоровых пациентов и низкую вероятность возникновения ошибок первого рода (ложноположительных результатов), что является важным фактором для снижения психологической нагрузки на пациентов при первичной диагностике.

Чувствительность модели (Sensitivity), отражающая полноту выявления злокачественных новообразований, при использовании стандартного порога классификации 0,5 составила 72,2%. Интегральная оценка разделяющей способности классификатора была проведена методом ROC-анализа. Полученная ROC-кривая демонстрирует быстрый рост доли истинно положительных результатов относительно ложноположительных, а рассчитанный показатель площади под кривой (AUC) составил 0,9251. Значение AUC, превышающее 0,9, интерпретируется как отличное качество классификации и подтверждает гипотезу о высокой информативности комплекса признаков, включающего размер опухоли и анатомическую локализацию, для оценки риска малигнизации. Полученные метрики свидетельствуют о том, что модель успешно формирует пространство признаков, в котором классы доброкачественных

и злокачественных новообразований линейно разделимы с высокой степенью достоверности.

Заключение

В ходе настоящего исследования была разработана и протестирована нейросетевая модель для прогнозирования риска злокачественности новообразований молочной железы, основанная на анализе клинических данных и морфометрических параметров. Полученные результаты экспериментально подтвердили гипотезу о высокой информативности таких признаков, как размер опухоли и ее локализация в квадрантах молочной железы, для задач автоматизированной диагностики. Достигнутый показатель качества классификации ROC-AUC на уровне 0,9251 свидетельствует о том, что архитектура многослойного перцептрона способна эффективно выявлять нелинейные закономерности в медицинских данных даже при ограниченном объеме выборки. Модель продемонстрировала высокую надежность в исключении патологии (специфичность 87,0%), что делает ее перспективным инструментом для первичного скрининга и снижения нагрузки на профильных специалистов. Вместе с тем выявленная потребность в повышении чувствительности алгоритма определяет направление для дальнейших исследований, которые должны быть сосредоточены на расширении обучающего датасета и калибровке пороговых значений классификатора, а также вычисления доверительных интервалов для параметра AUC методом Bootstrap при $n > 100000$. Разработанный программный прототип может служить технологической основой для создания интеллектуальных систем поддержки принятия врачебных решений в клинической онкологии.

Список литературы:

1. Mehrparvar F. Breast Cancer Prediction Dataset [Электронный ресурс]. Kaggle. 2023. URL: <https://www.kaggle.com/datasets/fatemehmehrparvar/breast-cancer-prediction/data> (дата обращения: 31.01.2026).

2. Yamaguchi T., Koyama Y., Inoue K. et al. Development of a deep learning-based automated diagnostic system (DLADS) for classifying mammographic lesions – a first large-scale multi-institutional clinical trial in Japan. *Breast Cancer*. 2025; 32: 1115–1124. <https://doi.org/10.1007/s12282-025-01741-3>.

3. Мирзабеков М.К., Богомолов О.А., Школьник М.И., Трухачева Н.В., Тихонский Н.Д. Сравнительный анализ показателей выживаемости пациентов с почечно-клеточным раком и опухолевым тромбозом почечной и нижней полой вены I–II уровня, подвергшихся хирургическому лечению открытым и лапароскопическим доступом. *Пермский медицинский журнал*. 2025; 42(4): 105–114. <https://doi.org/10.17816/pmj424105-114>.

4. Мирзабеков М.К., Тихонский Н.Д., Школьник М.И., Богомолов О.А., Трухачева Н.В. Прогнозирование безрецидивной выживаемости больных с почечно-клеточным раком и опухолевым тромбозом почечной и нижней полой вены I–II уровней с использованием расширенной модели Кокса и методов машинного обучения. *Наука и инновации в медицине*. 2025; 10(3): 237–242. <https://doi.org/10.35693/SIM686422>.

Как цитировать:

Дорцвейлер А. А., Карачев И. И., Тихонский Н. Д., Трухачева Н. В. Сравнительный анализ прогностической значимости локализации и размерных характеристик опухоли для диагностики рака молочной железы на основе алгоритмов машинного обучения. *Scientist (Russia)*. 2026; 2 (32): 263-267.
