

## Краткая информация о проекте

Наименование	AP22788239 «Использование искусственного интеллекта для раннего выявления рака молочной железы у женщин в Казахстане»
Актуальность	<p>Рак молочной железы остается одной из основных причин смертности от онкологических заболеваний. Важнейшей задачей является ранняя диагностика, поскольку именно она зачастую является ключом к улучшению результатов. Для решения этой важнейшей задачи используется множество методик искусственного интеллекта, предлагающих целый ряд методов диагностики рака молочной железы.</p> <p>В данной проектной работе представлена инновационная модель Deep Learning (DL), предназначенная для автоматизированной диагностики рака молочной железы.</p> <p>Исследование в рамках данного проекта вносит существенный вклад в предпринимаемые усилия по повышению эффективности раннего выявления и диагностики рака молочной железы, предлагая всестороннюю оценку методов DL и их потенциального влияния на улучшение результатов обследования пациентов и повышение качества медицинского обслуживания.</p>
Цель	Разработать наиболее эффективную модель глубокого обучения, которая позволит классифицировать МРТ-сканы и более точно выявлять рак молочной железы.
Задачи	<p>1. Сбор и предварительная обработка данных: Цель: Собрать и подготовить полный набор данных маммограмм, включающий здоровые и раковые образцы. Измеряемый показатель: Собрать более 1000 изображений маммограмм, сбалансированно представляющих различные стадии рака молочной железы. Обоснование и взаимосвязь: Данная задача способствует обучению модели ИИ, предоставляя ей необходимые данные для обучения, адаптации и повышения точности модели.</p> <p>2. Разработка диагностического инструмента на основе ИИ: Цель: Создать модель глубокого обучения, способную анализировать маммограммы на предмет выявления ранних признаков рака молочной железы. Измеряемый показатель: Достижение точности не менее 95% при выявлении раковых тканей на маммографических изображениях. Обоснование и взаимосвязь: Разработка этого важного инструмента, нацеленного на раннее выявление рака, взаимосвязана с задачами обучения и валидации, что обеспечивает технологическую готовность к завершению проекта.</p>

3. Обучение и валидация модели: Цель: Обучить модель глубокого обучения на собранных данных и проверить ее работоспособность. Измеряемый показатель: Достижение точности валидации более 95%, а также снижение количества ложноположительных и отрицательных результатов не менее чем на 50% по сравнению с существующими методами. Обоснование и взаимосвязь: Обучение является жизненно важным для способности модели точно диагностировать рак молочной железы. Валидация обеспечивает надежность и доверие к результатам работы инструмента.

4. Анализ эффективности средств вычислительного предсказания на основе ИИ в определении патогенности генетических вариантов в генах BRCA1 и BRCA2: Цель: Сравнить инструменты *in Silico* по основным параметрам, чтобы найти лучший метод ИИ для предсказания клинической значимости вариаций генов BRCA1 и BRCA2. Измеряемый показатель: В ходе исследования будет проведен анализ эффективности не менее пяти *in silico* инструментов, таких как PolyPhen2, SIFT, MutationAssessor, MutationTaster и Align GVGD. Обоснование и значимость: Цель исследования - выявить новые варианты генов BRCA1/BRCA2, что позволит продвинуть развитие персонализированной медицины в лечении и профилактике рака молочной железы.

5. Интеграция с системами здравоохранения: Цель: Интеграция инструмента ИИ в диагностические системы здравоохранения Казахстана. Измеряемый показатель: Внедрить инструмент не менее чем в 2 крупных больницах для тестирования, с целью достижения уровня удовлетворенности пользователей выше 80%. Обоснование и взаимосвязь: Интеграция связывает разработку модели ИИ с ее практическим применением, обеспечивая существенное влияние проекта на эффективность и доступность медицинских услуг.

6. Постоянное совершенствование и адаптация: Цель: Постоянное обновление модели на основе новых данных и обратной связи. Измеряемый показатель: Проводить обновления модели два раза в год, поддерживая или повышая уровень точности. Обоснование и взаимосвязь: Это гарантирует постоянную актуальность проекта и его адаптивность к меняющимся проблемам и достижениям в данной области.

<p>Ожидаемые и достигнутые результаты</p>	<p>Ожидаемые результаты:</p> <ol style="list-style-type: none"><li>1) Успешное накопление и предварительная обработка массива данных, содержащего более 1000 маммографических изображений, представляющих сбалансированное разнообразие стадий рака молочной железы. Этот обширный набор данных будет способствовать эффективному обучению ИИ-модели, обеспечивая ее адаптивность и точность в диагностике рака молочной железы.</li><li>2) Будет успешно разработан диагностический инструмент на основе ИИ, который будет соответствовать или превышать заданный критерий точности в 95% при выявлении раковых тканей на маммографических изображениях.</li><li>3) Успешное обучение и валидация модели ИИ, достижение точности валидации более 95% и значительное снижение ложно положительных и отрицательных результатов не менее чем на 50% по сравнению с существующими методами. Это позволит подтвердить способность модели к точной диагностике рака молочной железы и повысит ее надежность и достоверность в медицинской диагностике.</li><li>4) Успешное завершение сравнительного анализа не менее пяти программных пакетов <i>in silico</i>, в результате которого будет определен наиболее эффективный инструмент вычислительного прогнозирования на основе ИИ для оценки клинической значимости генетических вариаций в генах BRCA1 и BRCA2. Предполагается, что этот анализ позволит выявить ранее не идентифицированные генетические варианты, ассоциированные с раком молочной железы, что даст ценные сведения о новых генетических маркерах и будет способствовать развитию персонализированных подходов в лечении рака молочной железы.</li><li>5) Эффективная интеграция ИИ-диагностики в систему здравоохранения Казахстана, в частности, как минимум в двух крупных больницах для первоначальных испытаний, с целью достижения уровня удовлетворенности пользователей более 80%. Такая интеграция позволит преодолеть разрыв между разработкой модели ИИ и ее применением в реальных условиях и продемонстрирует ощутимое влияние проекта на повышение качества медицинских услуг.</li><li>6) Регулярное обновление модели ИИ раз в два года на основе новых данных и отзывов пользователей, обеспечивающее поддержание или повышение уровня точности модели. Такая стратегия постоянного совершенствования и адаптации позволит проекту оставаться актуальным и эффективным, динамично</li></ol>
---	---

	<p>реагируя на изменяющиеся проблемы и достижения в данной области.</p> <p><b>Достигнутые результаты:</b></p> <p><b>За 2024 год</b></p> <p>- Проведены сбор и предварительная обработка данных: собран и подготовлен обширный массив данных маммограмм, включающий как здоровые, так и раковые примеры. Накопленный и предварительно обработанный набор данных содержит 1109 маммографических изображений, представляющих сбалансированное разнообразие стадий рака молочной железы. Проведена тщательная предобработка данных, где удалены выбросы и аномальные значения. Получен чистый, сбалансированный набор данных, готовый к эффективному обучению ИИ-модели, обеспечивая ее адаптивность и точность в диагностике рака молочной железы.</p> <p>- Разработан диагностический инструмент на основе искусственного интеллекта: создана DL-модель, способная анализировать маммограммы на предмет выявления ранних признаков рака молочной железы. В результате разработанный диагностический инструмент на основе искусственного интеллекта соответствует или превышает заданный критерий точности в 95% при выявлении раковых тканей на маммографических изображениях.</p> <p><b>За первое полугодие 2025 года</b></p> <p>- Проведено обучение и валидация ИИ-модели на реальных клинических данных. В ходе работы выполнена комплексная подготовка реального датасета к обучению и валидации. Модель продемонстрировала точность более 95% на валидационных данных. Более того, было проанализировано как интерпретирует созданная модель реальный датасет и какова уверенность модели в предсказаниях.</p>
<p>Имена и фамилии членов исследовательской группы с их идентификаторами (Scopus Author ID, Researcher ID, ORCID, при наличии) и ссылками на соответствующие профили</p>	<ol style="list-style-type: none"> <li>1. Маткерим Базаргул, PhD: h-Index – 4, Scopus: 55898795100, ORCID ID: 0000-0002-5336-687</li> <li>2. Нуртас Марат, PhD: h-Index – 7, Scopus: 57189710532, Web of Science: AAW-7412-2020, ORCID ID: 0000-0003-4351-0185</li> <li>3. Алтайбек Айжан Алтайбеккызы, PhD: h-Index – 5, Scopus: 57218957117, ORCID ID: 0000-0001-8431-7950</li> <li>4. Есен Гани Аманжолұлы, магистр по направлению «Data Science», докторант</li> <li>5. Адиат Лашын, магистр по направлению «Data Science»</li> <li>6. Мурат Назерке, студентка 3 курса специальности «Computer Science»</li> </ol>

Список публикаций со ссылками на них	1) Gani Esen, Aizhan Altaibek, Jandos Amankulov, Bazargul Matkerimd, Marat Nurtas. Enhancing Breast Cancer Detection with Dimensionality Reduction Techniques: A Study Using PCA and LDA on Wisconsin Breast Cancer Data. Proceedings of the 14th International Conference on Current and Future Trends of Information and Communication Technologies in Healthcare (ICTH 2024), October 28-30, 2024, Leuven, Belgium. (Scopus, Q2, percentile: 69) Computer proceedings. <a href="https://doi.org/10.1016/j.procs.2024.11.128">https://doi.org/10.1016/j.procs.2024.11.128</a> <a href="https://www.sciencedirect.com/science/article/pii/S1877050924033623">https://www.sciencedirect.com/science/article/pii/S1877050924033623</a>
Информация о патентах	Результаты не подлежат патентованию в патентных ведомствах.