

АННОТАЦИЯ

диссертации на соискание степени доктора философии (PhD) по
образовательной программе: 8D10101 – Общественное здравоохранение

Дуненовой Гаухар Абуевны

«Цифровой репозиторий и машинная диагностика иммуногистохимических
препаратов как интегральные элементы прецизионной медицины
при раке молочной железы»

Актуальность темы исследования: Рак молочной железы (РМЖ) является ведущим онкологическим заболеванием среди женщин во всём мире: в 2020 году зарегистрировано 2,3 млн новых случаев и более 680 тыс. смертей (Sung et al., 2021). Высокая заболеваемость (>80 на 100 000 женщин) зафиксирована в странах с высоким уровнем дохода (Австралия, Новая Зеландия, Европа, Северная Америка), тогда как минимальные показатели (<40 на 100 000) – в странах Центральной Америки, Африки и Южной и Центральной Азии (Heer et al., 2020). Несмотря на высокую заболеваемость, смертность в развитых странах остаётся низкой (12–15 на 100 000) благодаря эффективному скринингу и лечению, в то время как в странах с переходной экономикой она на 17% выше (Sung et al., 2021). 5-летняя выживаемость в странах с высоким доходом превышает 90%, тогда как в Индии она составляет 66%, а в Южной Африке – 40% (WHO). В Казахстане РМЖ занимает первое место среди онкозаболеваний у женщин с 5-летней общей выживаемостью на уровне 57,1% (Д.Р. Кайдарова et al., 2023). По данным Globocan (2022), зарегистрировано 4570 новых случаев и 1574 смертей, а стандартизованные показатели заболеваемости и смертности составили 36,9 и 12,3 на 100 000 соответственно (Ferlay et al., 2021).

Иммуногистохимическое (ИГХ) определение рецепторного статуса стало ключевым этапом диагностики (Francis et al., 2022), обозначив начало прецизионной медицины в лечении РМЖ. ИГХ позволяет классифицировать РМЖ на пять основных субтипов, включая HER2-положительные (Goddard et al., 2011; Gradishar et al., 2022; Hammond et al., 2010), что играет решающую роль в выборе лечения и прогнозировании выживаемости (Bradley et al., 2021; Mitri et al., 2012). Визуальный анализ ИГХ-изображений HER2 представляет собой сложный и трудоемкий процесс, требующий высокой квалификации и точности, особенно при гетерогенной окраске (Palm et al., 2023). Ситуация усугубляется нарастающим дефицитом патологов (Gross et al., 2023; Rozario et al., 2024): согласно данным литературы, лишь от 1,1% до 8,6% выпускников медицинских школ выбирают эту специализацию в качестве основной (Masuadi et al., 2021).

С развитием технологий цифровая патология, основанная на полнослайдовых изображениях (WSI) (Aeffner et al., 2018, 2019), и искусственный интеллект (ИИ) открыли новые возможности для автоматизации диагностики и повышения точности. Алгоритмы машинного обучения (МО) могут обрабатывать большие объемы данных, выявлять скрытые паттерны (Bhalla & Laganà, 2022; Yee, 2021) и улучшать воспроизводимость результатов, а точность цифрового анализа HER2 уже достигает 90% (Kabir et al., 2024; Yao et al., 2022).

В РК ИГХ исследования РМЖ введены в практику с 2012 года, но информация о фенотипах остаётся фрагментарной и не включена в централизованный онкологический регистр, что затрудняет комплексный эпидемиологический анализ. Внедрение цифровых технологий в социальные и экономические сферы является одним из приоритетов развития Казахстана, отражено в государственных программах в рамках снижения риска преждевременной смертности от онкологических заболеваний. Цифровые репозитории и машинная диагностика становятся важными элементами прецизионной медицины, повышая экономическую эффективность (Smetherman et al., 2022), и поддерживая национальные базы данных для принятия решений в здравоохранении (Lauricella & Pêgo-Fernandes, 2022).

В условиях, когда в Казахстане ранее не проводились исследования, направленные на разработку и внедрение цифровых репозиторий для автоматической диагностики ИГХ-маркеров РМЖ с применением алгоритмов ИИ, данное исследование приобретает особую актуальность. Оно отвечает на потребность в инновационных подходах и закладывает основу для будущих разработок в области цифровой медицины, что определило выбор цели и задач данного исследования.

Цель исследования: Разработать и апробировать алгоритм машинной диагностики рака молочной железы на примере маркера Human Epidermal Growth Factor Receptor 2 (HER2) на основе прототипа цифрового репозитория иммуногистохимических изображений для совершенствования иммуногистохимической диагностики рака молочной железы в Республике Казахстан.

Задачи исследования

1. Проанализировать эпидемиологические показатели РМЖ в РК, показатели ИГХ диагностики в г.Алматы и Алматинской области и показатели лет жизни, скорректированных на нетрудоспособность (Disability-adjusted life years, DALY) в г.Алматы.
2. Изучить существующие в мире алгоритмы цифрового анализа HER2, их характеристики и критерии эффективности (систематический обзор).
3. Изучить организационную структуру действующей патоморфологической службы и определить условия для применения методов искусственного интеллекта (машинного обучения) в иммуногистохимической диагностике в Республике Казахстан.
4. Разработать алгоритм машинной диагностики HER2 рака молочной железы на основе создания прототипа цифрового репозитория и апробации путем тестирования на тренировочных и валидационных изображениях.

Материалы и методы исследования: Исследование проведено по следующей методологии

Дизайн исследования:

1. По первой задаче – эпидемиологическое исследование (фенотипы РМЖ, анализ и расчет бремени РМЖ).
2. По второй задаче – систематический обзор эффективности методов цифрового анализа маркера HER2 с использованием машинного обучения.
3. По третьей задаче – дескриптивное, аналитическое исследование, контент-анализ.
4. По четвертой задаче – экспериментальное исследование.

Объект исследования: Отчеты о заболеваемости РМЖ в РК за период 2012–2021 годы (форма №7 МЗ РК «Отчет о больных злокачественными заболеваниями», С50) (n=44 331). Данные ИГХ диагностики РМЖ: был проведен сбор данных из базы ЭРОБ за период 2020-2022 гг по г.Алматы и Алматинской области (n=2783, С50, С50.0-С50.9). Показатели ИГХ данных были извлечены из отчетов КГП на ПХВ «Алматинский онкологический центр» и ГКП на ПХВ «Алматинская региональная многопрофильная клиника», и объединены в общую базу данных пациентов РМЖ за 2020–2022 годы. Статистическая обработка материала производилась с помощью пакета IBM SPSS Statistics26. Статистические данные по показателю DALY для РМЖ в городе Алматы за период 2017–2021 годы. Для расчета YLL и YLD были использованы агрегированные данные о распространенности и смертности от РМЖ с разбивкой по коду заболевания, возрастным группам, полу и году смерти в городе Алматы, полученные по запросу от РЦЭЗ; таблица стандартной продолжительности жизни GBD 2017; весовые коэффициенты инвалидности для каждого состояния здоровья были получены из исследования GBD 2013. Систематический обзор существующих алгоритмов цифрового анализа HER2. НПА, регулирующие деятельность патоморфологической службы РК, включая телемедицину, текущая инфраструктура и нагрузка, вычислительные возможности для применения методов ИИ для цифрового анализа ИГХ изображений. Цифровые изображения HER2 маркера, собранные из патоморфологических лабораторий для разработки алгоритма МО.

Предмет исследования: Эпидемиологические показатели РМЖ: стандартизованные показатели заболеваемости и смертности, распределение по стадиям РМЖ в РК за период 2012-2021 гг; распространённость ИГХ фенотипов РМЖ в г.Алматы и в Алматинской области за 2020-2022 гг.; DALY РМЖ в г.Алматы за период 2017-2021 гг. Применяемые в мире алгоритмы цифрового анализа HER2 с использованием алгоритмов МО для оценки их точности, воспроизводимости и перспектив внедрения в клиническую практику. Текущие стандарты и практики морфологической и ИГХ диагностики РМЖ. Организационные и технические требования, необходимые для внедрения алгоритмов ИИ в патоморфологические лаборатории Казахстана, с учетом существующего национального портала патоморфологических

исследований. Оцифрованные изображения ИГХ маркера HER2 РМЖ (n=419). Прототип архитектуры. Алгоритм цифрового анализа изображений. Критерии точности.

Основные положения, выносимые на защиту

1. В Республике Казахстан за период 2012 по 2021 год заболеваемость РМЖ выросла, смертность снизилась. Стандартизированный показатель DALY в г.Алматы за период с 2017 по 2017 год снизился, что свидетельствует об улучшении выживаемости, несмотря на сохранявшееся значительное бремя заболевания. ИГХ диагностика РМЖ рутинно применяется в Алматы и Алматинской области для скрининга HER2, что создает потенциальные условия для формирования масштабных баз данных в цифровом репозитории ИГХ для эпидемиологического анализа, планирования ресурсов здравоохранения и улучшения алгоритмов цифровой диагностики.
2. Цифровой анализ HER2 является активно развивающимся и изучаемым методом в мире, демонстрирует высокую экспериментальную точность и потенциальную клиническую значимость.
3. Онкоморфологическая служба страны характеризуется значительными различиями в распределении нагрузки и ресурсов между регионами, а также нехваткой кадров. Имеется потенциал для внедрения технологий ИИ в ИГХ на базе национального портала патоморфологических исследований как дополнительного инструмента для патоморфологов.
4. Архитектура прототипа цифрового репозитория позволяет структурировать существующую практику использования национального портала патоморфологических исследований, обеспечивая развитие клинических, образовательных, научных направлений в патоморфологии онкологической службы РК. Алгоритм цифрового анализа HER2, разработанный на казахстанском датасете, имеет точность 75%.

Научная новизна: в результате проведенного исследования получены новые данные об эпидемиологических показателях РМЖ и распространенность ИГХ фенотипов РМЖ в г.Алматы и в Алматинской области в период 2020-2022 г. Впервые был рассчитан показатель DALY от РМЖ в г.Алматы за период 2017-2021гг (Авт.свид. №52072 от 02.12.2024г). Проведен систематический обзор существующих в мире алгоритмов цифрового анализа HER2, их характеристик, влияющих на эффективность с позиции клинической оценки. Впервые был создан датасет цифровых изображений ИГХ маркера HER2 РМЖ (n=419) (Авт.свид. №48772 от 01.08.2024г). Впервые был разработан прототип архитектуры (Авт.свид. №32121 от 30.01.2023г) и предложен алгоритм цифрового анализа изображений ИГХ маркера HER2, протестированный на собственном датасете и продемонстрировавший результаты точности 75%. Определены возможности применения методов ИИ в ИГХ на платформе Qazhisto.com.

Теоретическая значимость исследования: Полученные данные об эпидемиологии ИГХ фенотипов РМЖ позволяют разработать методические рекомендации для улучшения онкологического регистра, включая интеграцию ИГХ-диагностики с текущими данными пациентов. Использование DALY для оценки общего бремени РМЖ (смертность и заболеваемость) способствует оценке эффективности стратегий, оптимизации ресурсов и корректировке программ для снижения бремени РМЖ на национальном уровне. Систематический обзор мировых алгоритмов цифрового анализа HER2 формирует основу для разработки и организации компонентов нового алгоритма с клинической оценкой. Данные о структуре патоморфологической службы помогут разработать рекомендации по подготовке онкоморфологов для сложных исследований, а также могут стать научным обоснованием для внесения изменений в законодательство и инициирования разработки НПА, регулирующих применение искусственного интеллекта в медицине. Предложенная архитектура поддерживает создание крупных цифровых библиотек слайдов, доступных для патологов независимо от местоположения, что обогащает будущие исследования. Разработанный алгоритм цифрового анализа HER2 способствует совершенствованию системы накопления биомедицинских данных для дальнейших исследований.

Практическая значимость работы: Создан маркированный датасет цифровых изображений ИГХ маркера HER2 РМЖ (акт внедрения на базе кафедры эпидемиологии, биостатистики и

доказательной медицины факультета медицины и здравоохранения КазНУ) который позволяет разрабатывать и внедрять алгоритмы цифрового анализа цифровых изображений, развивать научную коллаборацию по продвижению научных и клинических исследований в области онкологии и информационных технологий.

Созданный датасет цифровых изображений ИГХ маркера HER2 РМЖ на платформе Qazhisto может использоваться в образовательных целях для обучения студентов и специалистов-патологоанатомов, и в качестве вспомогательного инструмента для формирования «второго мнения» (second opinion) для действующих онкоморфологов.

Предложенная архитектура в настоящее время служит основой для развития и потенциальной реализации возможностей платформы Qazhisto (акт внедрения на базе ЦМИ АОЦ).

Разработаны методические рекомендации по проведению телеконсультаций биообразцов опухолей на республиканском уровне, которые содержат рекомендации по организационной структуре и рабочему процессу телепатологии с использованием платформы Qazhisto (акты внедрения на базе ЦМИ КазНИИОиР, ЦМИ АОЦ, и 6 региональных онкоцентров).

Полученные научные данные могут быть предложены как основа для применения методов искусственного интеллекта в иммуногистохимическом анализе в структуре действующей платформы Qazhisto патоморфологической службы Республики Казахстан.

Предложенные рекомендации по использованию алгоритма позволят расширить рациональное использование ресурсов платформы Qazhisto и накапливать данные.

Личный вклад диссертанта: Включает разработку теоретической и методологической программы исследования, участие на всех этапах работы, создание базы данных и датасета, статистическую обработку данных, написание разделов диссертации, интерпретацию данных и разработку практических рекомендаций. Обсуждение задач и результатов проводилось совместно с консультантами. Диссертационная работа выполнена на инициативной основе.

Выводы

1. Распространенность РМЖ в РК за период 2012–2021 гг. увеличилась с 314,4 в 2012 г. до 444,3 в 2021 г. на 100 тыс. женского населения (средний темп прироста - 3,92%). За период 2020-2022 г. в г.Алматы и Алматинской области фенотип Luminal B, HER2 negative был наиболее распространённым (42,4%). Доля HER2 enriched, который потенциально подходит для регистрации в цифровом репозиторий ИГХ изображений HER2 РМЖ, составила 16,7% случаев. Фенотип, ассоциированный с наибольшим риском смерти, был представлен тройным негативным РМЖ, составляющий 11,9% случаев и увеличивающий риск смерти в 7 раз ($p=0,001$). Эта группа также имеет наибольший процент женщин моложе 45 лет (24,9%). 58% всех случаев заболевания были зарегистрированы на стадии II. Преобладавшей гистологической формой рака была NST, которая составила 95,9%. Отмечалось снижение стандартизированного DALY за изучаемый период с 638,9 до 489,5, с некоторым увеличением YLD (6,7% vs 6,0%) в динамике.
2. В мировой практике точность алгоритмов цифрового анализа ИГХ HER2 на искусственно созданном с использованием вычислительных алгоритмов датасете на базе общедоступного ресурса Warwick достигла 98,8%.
3. В РК в 2021 году наблюдались значительные региональные различия в нагрузке патоморфологической службы: объем ИГХ исследований варьировал от 283 в ООД Талдыкорган до 9050 в ЦМИ АОЦ. Нагрузка на штатные ставки колебались от 54 до 2262,5 исследований на единицу, а на физические лица — от 81 до 2262,5. Средняя укомплектованность кадрами составляла 64%. Решением проблемы может быть целенаправленный учет резидентов и специалистов в области патоморфологии, который может стать основой оптимизации планирования и регионального распределения кадровых ресурсов. Средой, позволяющей создать оптимальные условия для обучения специалистов и сбора централизованной информации об имеющихся кадровых ресурсах в области патоморфологии, может служить существующая платформа Qazhisto, позволяющая использовать в диагностических и образовательных целях алгоритмы машинного обучения.
4. Экспериментальное исследование цифрового анализа HER2 с использованием МО показало среднюю точность модели (Accuracy) 75%, с точностью Precision 1.0 для класса HER2 «3+».

Практические рекомендации: *Министерству Здравоохранения Республики Казахстан:* Рассмотреть вопрос создания референсного центра (регистра) для централизованного сбора и анализа данных по РМЖ, включая цифровые ИГХ изображения, для улучшения планирования ресурсов, эпидемиологического мониторинга и скрининга. Включить показатели DALY каждые 5 лет для оценки бремени заболевания и распределения ресурсов. Внедрить платформу Qazhisto для хранения и анализа данных в профильные учреждения. Увеличить число патоморфологов через реформирование программы резидентуры, используя цифровой репозиторий в обучении как образовательную платформу для подготовки специалистов в области цифровой патологии. Рассмотреть возможность выделения в отдельную категорию в отчетных формах резидентов и специалистов-патоморфологов для более эффективного учета их численности и контроля движения. Создать экспертную группу для установления стандартов нагрузки патологоанатома. Инициировать создание специализированной экспертной группы с целью разработки научно-обоснованных рекомендаций по использованию технологий ИИ в медицинской практике.

Менеджерам здравоохранения (руководителям патоморфологических лабораторий онкоморфологической службы): Ввести регулярное обучение для медицинского персонала и патоморфологов современным методам цифровой патологии, организовав вебинары и семинары с привлечением локальных и международных экспертов.

Для МОН и образовательных учреждений: Создать междисциплинарную группу для развития совместных исследований с международными центрами, что позволит адаптировать и внедрить передовые технологии цифровой патологии в местную практику. Содействовать развитию грантовых программ для стимулирования исследований и разработок в области цифровой патологии и ИИ.

Публикации по теме диссертации: По теме диссертации работы опубликовано 4 работы, из них 2 статьи в журналах, индексируемых в базе данных Scopus: Cancers (Basel) (Scopus, 5-year IF: 4.9, ISSN 2072-6694, CiteScore 8,0, Rank#81/404, 79 percentile in category «Oncology»); Scientific Reports (Scopus, 5-year IF: 4.3, ISSN:2045-2322, CiteScore 7,5, Rank#14/171, 92 percentile in category «Multidisciplinary»), 1 статья в журнале, рекомендованном Комитетом по контролю в сфере образования и науки МОН РК, 1 методические рекомендации (ISBN 978-601-7548-27-8).

Апробация диссертации на конференциях: Междисциплинарная конференция КазНУ «Автоматизация и Машинное Обучение в Медицине и Организации Здравоохранения», 05.02.2021, доклад «Цифровой репозиторий и машинная диагностика иммуногистохимических препаратов как интегральные элементы Прецизионной Медицины при раке молочной железы». Форум науки «Интеллектуальный потенциал независимого Казахстана: 30 лет становления и развития», секция молодых ученых «ИИ, большие данные & медицина: вызовы и перспективы», 09.09.2021 года, доклад «Машинная диагностика ИГХ изображений. Перспективы развития».

Инновационные патенты, авторские свидетельства: Авт.свид. №32121 «Архитектура обеспечения цифрового анализа в иммуногистохимической диагностике рака молочной железы» 30.01.2023. Авт.свид. №48772 «HER2 Breast Cancer Digital Image Dataset (ADEL Dataset). Датасет цифровых изображений HER2 рака молочной железы» 01.08.2024. Авт.свид. №52072 «Адаптированная методология изучения бремени онкологических заболеваний в Республике Казахстан» 02.12.2024.

По результатам диссертационной работы разработаны: Методические рекомендации по проведению телеконсультаций биообразцов опухолей на республиканском уровне, с рекомендациями по организационной структуре и рабочему процессу телепатологии с использованием платформы Qazhisto. Акты внедрения результатов научно-исследовательской работы в деятельность организаций практического здравоохранения.

Объем и структура диссертации: Диссертация изложена на 178 страницах, состоит из введения, 3 разделов, заключения, выводов, практических рекомендаций и приложений. Работа иллюстрирована 22 таблицами и 27 рисунками. Список литературы содержит 41 отечественных и 235 зарубежных источников.