АННОТАЦИЯ

на диссертационную работу Сулейменовой Мадины Утегеновны на тему: «Машинное обучение в прогнозировании риска развития сердечно – сосудистых заболевании», представленную на соискание ученой степени PhD, по образовательной программе: 8D06114 – «Искусственный интеллект в медицине»

Актуальность темы исследования. Сердечно-сосудистые заболевания являются одной из основных причин смертности и инвалидности как в Казахстане, так и во всем мире, и оказывают влияние на продолжительность жизни. По данным ВОЗ, Казахстан относится к регионам с высоким риском развития сердечно-сосудистых заболеваний (ССЗ). В Казахстане в 2021 году в связи с ростом заболеваемости и смертности ожидаемая продолжительность жизни снизилась до 70,23 лет, тогда как в 2020 году она составляла 71,37 лет. За последние два года средняя продолжительность жизни сократилась на 2,95 года, что соответствует уровню 2013 года. Эти показатели свидетельствуют о необходимости выявления маркеров, играющих решающую роль в прогнозировании раннего сердечно-сосудистого старения и профилактике неблагоприятных исходов. Применение методов искусственного интеллекта (AI), таких как машинное обучение и моделирование, позволяет выявлять индивидуальные риски и прогнозировать вероятность развития ССЗ, что становится важной частью современной медицины, поскольку ускоряет анализ, прогнозирование и другие процессы.

Машинное обучение предлагает гибкую основу для прогнозирования, однако его клиническое внедрение требует строгой валидации и интерпретации. В нашем исследовании качество модели оценивалось не только по ROC-AUC, но и комплексно, с учетом внутренней валидации без утечки данных, оценки калибровки, клинической полезности на основе Decision Curve Analysis, а также интерпретации с использованием коэффициентов, permutation importance и SHAP. Существенная методологическая проблема возникает в тех случаях, когда development-выборка невелика: без тщательной проверки качество может казаться высоким, отбор признаков может быть нестабильным, а устойчивость модели может оставаться неопределенной.

Оценка риска развития ССЗ осложняется мультиморбидностью, гетерогенностью метаболических профилей и популяционно-специфическим базовым риском, что может влиять как на дискриминационную способность, так и, прежде всего, на калибровку вероятностного прогноза. Многие традиционные шкалы риска разработаны на смешанных по возрасту когортах и при применении у пожилых пациентов или в новых условиях могут демонстрировать ограниченную переносимость.

Таким образом, в повседневной практике полезность прогностического инструмента определяется не только его способностью стратифицировать пациентов по степени риска, но и тем, насколько хорошо предсказанные

вероятности откалиброваны для принятия пороговых решений и насколько поведение модели может быть интерпретировано с клинической точки зрения.

Исходя из вышеизложенного, были сформулированы цель и задачи данного исследования.

Цель диссертационной работы. Построение интеллектуальной системы прогнозирования риска развития сердечно-сосудистых заболеваний у лиц 65 лет и старше с применением воспроизводимого end-to-end процесса и моделей машинного обучения.

Задачи исследования. Для достижения поставленной цели необходимо решить следующие задачи:

1. Исследовать модели машинного обучения для определения прогностического риска развития заболеваний в медицине.
2. Сформировать development-выборку лиц 65+ с клиническими и биохимическими показателями с предобработкой данных в виде воспроизводимого pipeline без утечек данных.
3. Провести сравнительный анализ и обучить модели машинного обучения с использованием nested cross-validation и получить out-of-fold (OOF) вероятности.
4. Провести оценку качества моделей по показателям дискриминации и вероятностных прогнозов, проверить калибровку вероятностей, выполнить интерпретацию прогнозов и определить устойчивость значимости признаков (permutation importance).
5. Обосновать пороги принятия решений и границы категорий риска на основе Decision Curve Analysis (DCA) и сформировать шкалу Low/Intermediate/High на development-выборке.
6. Провести внешнюю валидацию разработанной шкалы на независимой когорте, оценить переносимость модели, выполнить вероятностную перекалибровку.

Объект и предметы исследования: Табличные мультимодальные данные (клинические и биохимические) лиц 65 лет и старше г. Алматы. Модели машинного обучения для прогнозирования риска развития сердечно-сосудистых заболеваний на основе клинических и биохимических данных

Методы исследования: выполнены сбор и предобработка мультимодальных данных, корреляционный анализ и построение прогностических моделей (Elastic Net Logistic Regression, Random Forest) внутренняя и внешняя валидация (n=695) с оптимизацией гиперпараметров; оценка качества (ROC-AUC, PR-AUC, Brier, LogLoss, калибровка); обоснование порогов решений (DCA); интерпретация факторов (OR, permutation importance, SHAP).

Научная новизна:

- Предложен воспроизводимый end-to-end процесс для построения вероятностной интерпретируемой шкалы риска на табличных мультимодальных данных, объединяющий: стандартизированную предобработку без утечек данных, честную оценку прогностической способности, проверку корректности вероятностных прогнозов.

- Впервые предложена вероятностная шкала риска ССЗ для лиц 65+, построенная на сочетании клинических и биохимических маркеров, для практической стратификации пациентов и поддержку принятия решений. Инструмент выдаёт вероятность риска p и категорию (Low/Intermediate/High) для маршрутизации пациента.

Положение выносимое на защиту:

1. Предложенный воспроизводимый end-to-end процесс машинного обучения на малой выборке прогнозирует риск развития ССЗ на 93%.
2. Прогностическая способность разработанной модели и качество вероятностного прогноза (Brier/LogLoss) подтверждены внешней валидацией на независимой когорте (n=695), с учётом различий базового риска популяций (ROC-AUC≈0.835) . Интерпретируемая вероятностная шкала CARDIO RISK KZ применима для оценки риска сердечно-сосудистых заболеваний у лиц 65 лет и старше.

Практическая и научно-техническая значимость:

Разработанная интерпретируемая вероятностная шкала с применением моделей машинного обучения позволяет оценить возможные риски развития сердечно-сосудистых заболеваний для групп пациентов 65 лет и старше жителей г. Алматы со сходными клиническими и биохимическими данными. Использование данной шкалы позволит принимать эффективные профилактические и лечебные решения с учётом персонализированного подхода.

Степень достоверности и апробация результатов. По теме диссертации опубликованы 3 статьи в журналах, индексируемых в базе Scopus, 2 статьи в научных изданиях, рекомендованных Комитетом по обеспечению качества в сфере науки и высшего образования Министерства науки и высшего образования Республики Казахстан, а также получены 2 свидетельства о внесении сведений в Государственный реестр прав на объекты, охраняемые авторским правом.

Статьи в журналах, индексируемых в базе Scopus:

Suleimenova, M., Abzaliyev, K., Abzaliyeva, S., & Nassyrova, N. (2026). An Interpretable and Externally Validated Model for Cardiovascular Disease Risk Assessment in Older Adults. Applied Sciences, 16(8), 3903. https://doi.org/10.3390/app16083903

Abzaliyev, K., Suleimenova, M., Chen, S., Mansurova, M., Abzaliyeva, S., Manapova, A., Kurmanova, A., Bugibayeva, A., Sundetova, D., Bitemirova, R., Baizhigitova, N., Abdykassymova, M., & Sagalbayeva, U. Predicting Cardiovascular Aging Risk Based on Clinical Data Through the Integration of Mathematical Modeling and Machine Learning. Applied Sciences, 2025, 15(9), 5077. https://doi.org/10.3390/app15095077 (Scopus: Q2)

Abzaliyev, K., Suleimenova, M., Abzaliyeva, S., Mansurova, M., Shomanov, A., Bugibayeva, A., Tolemisova, A., Kurmanova, A., & Nassyrova, N. Multidimensional Visualization and AI-Driven Prediction Using Clinical and Biochemical Biomarkers in Premature Cardiovascular Aging. Biomedicines, 2025, 13(10), 2482. https://doi.org/10.3390/biomedicines13102482 (Scopus: Q1)

Статьи в журналах из перечня научных изданий, рекомендованных Комитетом по обеспечению качества в сфере науки и высшего образования Министерства науки и высшего образования Республики Казахстан для публикации основных результатов научной работы:

Suleimenova, M., Chen Siming, Shomanov, A., Abzaliyev, K., Kurmanova, A., Abzaliyeva, S., Abdykassymova, M., Sagalbayeva, U., Bitemirova, R., Sundetova, D., Bugibayeva, A. AI in Identifying Aging Markers. Фтизиопульмонология, №01(43), 2024, с. 130–139.https://www.doi.org/10.26212/2227-1937.2024.50.73.018

Сулейменова М.У., Мансурова М.Е. Искусственный интеллект как инструмент исследования признаков старения. Вестник НИА РК, №2, 2024.

Свидетельство о внесении сведений в Государственный реестр прав на объект, охраняемый авторским правом, на программу для ЭВМ под названием:

«CARDIO RISK KZ: Интеллектуальная система для расчёта индивидуального риска преждевременного сердечно-сосудистого старения». № 64683, 2025.

«Программа для вероятностной оценки риска сердечно-сосудистых заболеваний у лиц 65 лет и старше». № 68445, 2026.

Личный вклад исследователя. В соответствии с поставленными целями и задачами диссертационной работы все задачи были выполнены. На основе клинических и биохимических данных лиц старше 65 лет был разработан воспроизводимый pipeline, а также без утечки данных построены модели Elastic Net логистической регрессии и Random Forest. Модели были оценены с использованием nested cross-validation, рассчитаны показатели ROC-AUC, PR-AUC, Brier, LogLoss и калибровки. На основе Decision Curve Analysis были определены пороги риска (t_low = 0.23, t_high = 0.40) и сформирована шкала Low / Intermediate / High. Разработанная шкала была проверена на независимой внешней когорте (n = 695), а результаты реализованы в виде программного калькулятора.

Объем и структура работы. Диссертация написана на казахском языке. Работа состоит из введения, 4 глав и заключения. Общий объем диссертации составляет 86 страниц, включает 24 рисунка, 10 таблиц и 53 использованных источника.

Во введении обоснована актуальность работы, показана необходимость разработки клинически применимых и интерпретируемых моделей машинного обучения для прогнозирования риска сердечно-сосудистых заболеваний у лиц 65 лет и старше. Кроме того, изложены цель, задачи и методы диссертационной работы, научная новизна полученных результатов, теоретическая и практическая значимость работы, а также приведен список публикаций.

В первой главе представлен обзор литературы по современному состоянию исследований сердечно-сосудистых заболеваний, сердечно-сосудистого старения и факторов риска у лиц старше 65 лет. Также проведен анализ традиционных шкал оценки сердечно-сосудистого риска и методов

машинного обучения, показаны их преимущества и ограничения. Обоснована актуальность разработки новых интерпретируемых и клинически удобных подходов к оценке риска для пожилой популяции.

Во второй главе описаны материалы и методы исследования. Рассмотрены подготовка данных development cohort (n=100) и external validation cohort (n=695), предварительная обработка, отбор клинических и лабораторных показателей, корреляционный анализ и методы многомерной визуализации. Кроме того, изложены этапы построения моделей машинного обучения, nested cross-validation, получения out-of-fold прогнозов, применения методов интерпретации, определения порогов риска на основе Decision Curve Analysis, проведения внешней валидации и разработки цифрового прототипа.

В третьей главе представлены результаты построения и оценки моделей прогнозирования риска сердечно-сосудистых заболеваний. Сравнивались логистическая регрессия, Random Forest и другие подходы, после чего для клинического применения была выбрана финальная модель как наиболее интерпретируемая и устойчивая. Качество моделей оценивалось по показателям ROC-AUC, PR-AUC, Brier score и LogLoss, анализировались результаты калибровки, а с помощью DCA были обоснованы пороги риска t_low = 0.23 и t_high = 0.40. Кроме того, представлены результаты независимой внешней валидации и предложена интеллектуальная система CARDIO RISK KZ для оценки риска сердечно-сосудистых заболеваний у лиц старше 65 лет.

В четвертой главе проведено обсуждение результатов исследования. Обоснованы научная и практическая значимость разработанной модели прогнозирования риска сердечно-сосудистых заболеваний, а также объяснен выбор клинической логистической регрессии в качестве финальной модели. В качестве основных предикторов показана роль наследственности, индекса массы тела, физической активности и сахарного диабета, а также описаны возможности применения разработанной шкалы CARDIO RISK KZ в клинической практике, ограничения исследования и перспективы дальнейшего развития.

В заключении приведены основные результаты и выводы диссертационной работы.