АННОТАЦИЯ
диссертационной работы Тасмурзаева Нурдәулета на тему:
«Разработка интеллектуальной системы для выявления ишемической болезни сердца на основе Интернета Медицинских Вещей (IoMT) и машинного обучения»,
представленной на соискание степени доктора философии (PhD)
по образовательной программе «8D07116 – Интеллектуальные системы управления»

Актуальность исследования. Ишемическая болезнь сердца (ИБС) остается ведущей причиной смертности и инвалидности во всем мире, включая Республику Казахстан. По данным Всемирной организации здравоохранения, сердечно-сосудистые заболевания ежегодно уносят жизни более 17 миллионов человек. Ранняя и точная диагностика ИБС является критически важной предпосылкой для эффективного лечения и предотвращения осложнений. В то же время клиническая практика сталкивается с постоянными системными ограничениями: нехваткой квалифицированных специалистов, высокой стоимостью диагностических процедур и неравным доступом к медицинской помощи, особенно в отдаленных и ресурсоограниченных регионах. В этом контексте разработка интеллектуальных систем на основе машинного обучения для поддержки принятия клинических решений имеет первостепенное значение.

Традиционные методы машинного обучения, включая логистическую регрессию и ансамблевые алгоритмы, такие как XGBoost, CatBoost и RandomForest, широко используются для классификации ИБС на основе клинических данных. Однако эти методы имеют существенные ограничения. Во-первых, клинические наборы данных обычно характеризуются значительным дисбалансом классов. Во-вторых, традиционные модели рассматривают признаки как независимые входные переменные, что препятствует адекватному моделированию сложных взаимодействий между демографическими показателями, лабораторными данными, результатами ЭКГ и данными эхокардиографии. В-третьих, ограниченный размер доступных клинических когорт повышает риск переобучения при применении высокопараметрических архитектур.

Трансформерные архитектуры, основанные на механизме self-attention, открывают фундаментально новые возможности для анализа структурированных медицинских данных, позволяя модели явно учитывать взаимосвязи между признаками. Тем не менее, при небольших размерах наборов данных стандартные архитектуры трансформеров склонны к переобучению, а игнорирование клинической структуры входных данных приводит к потере клинически значимой информации.

Отдельная проблема касается конфиденциальности и безопасности медицинских данных. Передача клинических записей в облачные

вычислительные среды создает риски нарушения конфиденциальности пациентов. Следовательно, практическое развертывание интеллектуальных медицинских систем требует полностью локальной обработки данных на устройствах (on-device), что, в свою очередь, требует компактных моделей, адаптированных к ресурсным ограничениям встроенных платформ периферийных вычислений (edge computing).

Таким образом, разработка интеллектуальной системы, сочетающей иерархическую архитектуру трансформера, учитывающую клиническую структуру признаков, надежное обучение на несбалансированных и малых наборах данных, а также стратегию эффективного сжатия модели и периферийного развертывания с прозрачностью решений, представляет собой важную научную задачу как теоретического, так и прикладного значения.

Целью диссертационной работы является разработка интеллектуальной системы на основе гибридной трансформерной архитектуры и модели для Edge-устройств, в целях повышения точности и эффективности выявления ишемической болезни сердца.

Для достижения этой цели в работе решаются следующие задачи:

1) Разработка гибридной трансформерной архитектуры (HG-Former), включающей алгоритм ADASYN для устранения дисбаланса и оптимизацию обучения методом адаптивной минимизации резкости (ASAM) для повышения точности выявления ИБС;

2) Разработка модели для переноса и эффективного развертывания модели на Edge-устройствах, включающей дистилляцию знаний и SHAP.

Методы исследования: методы машинного обучения и глубокого обучения для классификации табличных данных, методы периферийного развертывания.

Объектом исследования является процесс интеллектуального анализа клинических медицинских данных для выявления ишемической болезни сердца с использованием методов машинного обучения, объяснимого искусственного интеллекта и Edge-вычислений.

Предметом исследования являются модели, методы и алгоритмы машинного обучения для выявления ИБС, включая гибридную трансформерную архитектуру HG-Former (с оптимизатором ASAM) и алгоритмы сжатия моделей для развёртывания на Edge-устройствах.

Научная новизна полученных результатов заключается в следующем:

− предложен алгоритм выявления ишемической болезни сердца на основе новой гибридной архитектуры, объединяющей иерархическую трансформерную архитектуру и метод Adaptive Sharpness-Aware Minimization, что обеспечило эффективное обучение модели в условиях ограниченного набора данных;

− предложена модель выявления ишемической болезни сердца для Edge-устройств, основанная на совместном применении дистилляции знаний и SHAP-ориентированного отбора признаков. Это позволило значительно сократить

количество используемых признаков при сохранении высокой точности классификации.

Положения, выносимые на защиту:

− разработанный алгоритм выявления ишемической болезни сердца на основе гибридной трансформерной архитектуры HG-Former и метода Adaptive Sharpness-Aware Minimization, обученный на наборе данных Z-Alizadeh Sani, обеспечивает высокую эффективность классификации ИБС с точностью 96,72%, что превосходит существующие альтернативные методы в среднем на 10%;

− разработанная модель для выявления ишемической болезни сердца на Edge-устройствах, с применением методов дистилляции знаний и SHAP-ориентированного отбора признаков, достигает точности классификации 91,80%, что в среднем на 5-8% выше по сравнению с базовыми компактными моделями. При этом количество используемых признаков сокращено с 56 до 16.

Теоретическая и практическая значимость исследования. Разработана интеллектуальная система для выявления ишемической болезни сердца, объединяющая высокоточную гибридную трансформерную архитектуру HG-Former, устойчивое обучение на малых и несбалансированных медицинских данных, и возможность локального развертывания модели на Edge-устройствах. Предложенный подход повышает точность и надежность классификации ИБС, снижает риск переобучения, обеспечивает прозрачность принятия решений за счет SHAP-анализа и attention-механизмов, а также позволяет выполнять обработку медицинских данных без передачи в облако, что повышает конфиденциальность, доступность и практическую применимость интеллектуальных медицинских систем. Практическая применимость разработанной системы подтверждается её внедрением в интеллектуальную диагностическую платформу ZhurekAI в АО «Научно-исследовательский институт кардиологии и внутренних болезней».

Достоверность научных результатов обеспечивается применением передовых методов глубокого обучения, включая гибридные архитектуры трансформеров (HG-Former) и периферийные вычисления (Edge ML). Результаты подтверждаются вычислительными экспериментами на реальных клинических данных, внешней валидацией на независимой когорте из базы данных MIMIC-IV, успешным применением методов дистилляции знаний и анализа SHAP, а также успешным сравнением предложенных моделей с существующими методами классификации ишемической болезни сердца.

Связь темы с планами научно-исследовательских программ. Диссертационная работа выполнена в рамках научных исследовательских проектов Министерства науки и высшего образования Республики Казахстан (источник финансирования: Комитет науки Министерства науки и высшего образования Республики Казахстан):

− грантовое финансирование (ГФ) AP26103523 «Разработка интеллектуальной системы мониторинга и диагностики нарушений сна с

использованием технологии блокчейн и IoMT (Internet of Medical Things)» на 2025-2027 гг.

Структура и содержание диссертационной работы. Диссертация состоит из введения, четырех основных глав, заключения и списка использованных источников.

Во введении обоснована актуальность темы, четко определены цель и задачи исследования, выявлены объект и предмет исследования, раскрыты научная новизна и практическая значимость работы, а также изложены основные положения, выносимые на защиту.

В первой главе представлен обзор литературы, охватывающий классические методы машинного обучения для диагностики ИБС, современные методы искусственного интеллекта в кардиологии, математические основы архитектур трансформеров, алгоритмы оптимизации и сжатия моделей, а также критический анализ существующих гибридных и комбинированных подходов.

Во второй главе изложены математические модели предложенной системы, описаны наборы данных и методы их предобработки, основные технические компоненты архитектуры HG-Former, процедуры обучения и оптимизации, а также представлены результаты предварительных сравнительных экспериментов путем последовательного удаления (ablation) архитектурных компонентов.

В третьей главе описана гибридная архитектура HG-Former, включая группировку клинических признаков и механизм токенизации, внутригрупповые и межгрупповые энкодеры трансформера, интеграцию алгоритма ASAM, результаты экспериментов на наборе данных Z-Alizadeh Sani, анализ SHAP и интерпретацию механизма внимания. Полная конфигурация HG-Former, обученная совместно с оптимизатором ASAM, достигла на отложенной тестовой выборке AUC 0,9755, точности 0,9672, F1-меры 0,9762, Macro Recall 0,9767 и PR-AUC 0,9919, существенно превзойдя десять классических базовых моделей машинного обучения, включая CatBoost, XGBoost, RandomForest и FT-Transformer, а также пять оптимизированных роевыми алгоритмами стекинг-ансамблей. Внешняя валидация на независимой когорте из 800 пациентов, извлеченной из базы данных MIMIC-IV, показала AUC 0,9620 и точность 0,9038.

В четвертой главе представлен оптимизированный алгоритм развертывания на периферийных устройствах, включающий дистилляцию знаний в компактную модель-студент ProtoNN, отбор признаков, ориентированный на SHAP, квантование и прунинг, развертывание TFLite на Raspberry Pi Zero 2 W. Совместное применение дистилляции знаний и отбора признаков на основе SHAP позволило сократить пространство входных признаков с 56 до 16 клинически информативных переменных, одновременно сжав модель до размера 4,35 КБ в формате TFLite. Модель-студент достигла классификационной точности 91,80% и AUC 0,9651 при снижении задержки вывода до 0,0740 мс на один прогноз – примерно в 400 раз быстрее исходной модели-учителя.

В заключении изложены основные результаты и выводы диссертации.

Публикации. Результаты диссертационного исследования опубликованы в 6 работах. Среди них 5 статей в международных научных журналах, включенных в базы Web of Science и Scopus, 1 статья в журналах, рекомендованных Комитетом по обеспечению качества в сфере науки и высшего образования МНВО РК, а также 2 авторских свидетельства и 1 патент на полезную модель по результатам практической реализации диссертационного исследования.

Научные публикации:

1. Tasmurzayev, N.; Amangeldy, B.; Imankulov, T.; Imanbek, B.; Postolache, O.A.; Konysbekova, A. A Wearable IoT-Based Measurement System for Real-Time Cardiovascular Risk Prediction Using Heart Rate Variability. Eng 2025, 6, 259. https://doi.org/10.3390/eng6100259 (IF: 0.4, Scopus Percentile: 58%, Q2).

2. Tasmurzayev, N.; Amangeldy, B.; Imanbek, B.; Baigarayeva, Z.; Imankulov, T.; Dikhanbayeva, G.; Amangeldi, I.; Sharipova, S. Digital Cardiovascular Twins, AI Agents, and Sensor Data: A Narrative Review from System Architecture to Proactive Heart Health. Sensors 2025, 25, 5272. https://doi.org/10.3390/s25175272 (IF: 3.5, Scopus Percentile: 88%, Q1).

3. Tasmurzayev, N.; Imanbek, B.; Boltaboyeva, A.; Dikhanbayeva, G.; Zhussupbekov, S.; Saparbayeva, Q.; Amirkhanova, G. Explainable AI for Coronary Artery Disease Stratification Using Routine Clinical Data. Algorithms 2025, 18, 693. https://doi.org/10.3390/a18110693 (IF: 2.3, Scopus Percentile: 78%, Q1).

4. Tasmurzayev, N.; Baigarayeva, Z.; Amangeldy, B.; Imanbek, B.; Kurmanbek, S.; Dikhanbayeva, G.; Amirkhanova, G. Interpretable Machine Learning for Coronary Artery Disease Risk Stratification: A SHAP-Based Analysis. Algorithms 2025, 18, 697. https://doi.org/10.3390/a18110697 (IF: 2.3, Scopus Percentile: 78%, Q1).

5. Tasmurzayev, N.; Amangeldy, B.; Baigarayeva, Z.; Boltaboyeva, A.; Imanbek, B.; Maeda-Nishino, N.; Zhussupbekov, S.; Baidauletova, A. Enhancing Cardiovascular Disease Classification with Routine Blood Tests Using an Explainable AI Approach. Algorithms 2025, 18, 708. https://doi.org/10.3390/a18110708 (IF: 2.3, Scopus Percentile: 78%, Q1).

6. Tasmurzayev, N.; Turmakhanbet, D.; Kakharov, A.; Aitkazin, M.; Baidauletova, A.; Kozhamberdiyeva, M. CVD Prediction from HRV Derived from Wearable PPG. JPCSIT, vol. 3, no. 4, Dec. 2025. https://doi.org/10.26577/jpcsit202542.

7. Авторское свидетельство: Турмаханбет Д. Н., Тасмурзаев Н. М., Мауленбеков М. А., Ибашова А. Б., Сапарбаева К. Ж., Иманбек Б. Т., Болтабоева А. К., Айтказин М. К., Искакова А. М., Кудьярова Ж. Б., Амангельды Б. С., Байгараева Ж. Е., Жолдыбаева Ж. К., Байдаулетова А. И. Метод определения уровня ментального стресса на основе анализа вариабельности сердечного ритма и амплитуды пульсовой волны PPG. Свидетельство о государственной регистрации прав на объект авторского права № 62846 от 9 октября 2025 года.

8. Авторское свидетельство: Турмаханбет Д. Н., Тасмурзаев Н. М., Мауленбеков М. А., Ибашова А. Б., Сапарбаева К. Ж., Иманбек Б. Т., Болтабоева А. К., Айтказин М. К., Искакова А. М., Кудьярова Ж. Б., Амангельды Б. С., Байгараева Ж. Е., Жолдыбаева Ж. К., Байдаулетова А. И. Программный комплекс для оценки сосудистого возраста и ригидности артерий на основе анализа фотоплетизмограммы. Свидетельство о государственной регистрации прав на объект авторского права № 63527 от 29 октября 2025 года.

9. Патент на полезную модель: Иманбек Б. Т., Болтабоева А. К., Амангельды Б. С., Тасмурзаев Н. М., Нурланов Ю. Ю., Копенов А. Б., Ибрашева З. С., Омирбекова Ж. Ж., Мансурова М. Я., Байгараева Ж. Е., Амирханова Г. А. Система мониторинга параметров сердечно-сосудистой системы. Патент на полезную модель № 10282 от 7 ноября 2025 года.