АННОТАЦИЯ
к диссертации Еркос Айнур на соискание степени
доктора философии (PhD) по образовательной программе
8D06102 – Компьютерная инженерия
на тему: «Разработка метода и программного пакета для обработки больших медицинских данных»

Актуальность темы исследования. В условиях глобальной цифровизации систем здравоохранения объем медицинских данных, накапливаемых в процессе диагностики, лечения и долгосрочного мониторинга пациентов, экспоненциально растет. Этот массив данных, характеризующийся скоростью накопления, разнообразием структур и сложностью интерпретации, является классическим примером Big Data. Эффективное использование таких данных открывает беспрецедентные возможности для развития предиктивной медицины и создания систем поддержки медицинских решений. Эти возможности позволяют клиницистам, учитывая индивидуальные особенности пациентов, выявлять риски заранее и оптимизировать траектории лечения, в результате чего медицинская практика поднимается на новый уровень.

Одна из самых важных и ресурсоемких областей современной медицины — трансплантология, в частности трансплантация почки. Дефицит донорских органов и риск отторжения трансплантата требуют внедрения инструментов прогнозирования выживаемости с высокой точностью. Несмотря на наличие классических статистических подходов, они часто не учитывают нелинейные зависимости факторов риска в долгосрочной перспективе и их динамический характер. Ограничения этих подходов снижают качество решений в трансплантологии и усложняют работу врачей. Однако в этой области существуют сложные технологические барьеры на пути применения методов интеллектуального анализа данных (Data Mining).

Во-первых, медицинские выборки по своей природе несбалансированы: случаи с неблагоприятными исходами (отторжение органа, смерть) значительно реже, чем стабильные состояния. Это приводит к тому, что стандартные алгоритмы машинного обучения, показывая высокую общую точность, оказываются неэффективными в прогнозировании редких, но критически важных клинических событий. Существующие методы балансировки часто вводят шум или приводят к потере ценной информации.

Во-вторых, традиционные программные решения недостаточно гибки для обработки гетерогенных данных и не учитывают динамику изменения состояния пациента на разных этапах после операции. Они опираются на статические модели, что не позволяет полностью отразить временную эволюцию реальных клинических процессов.

Таким образом, разработка специального метода балансировки кластеров, эффективно работающего с шумом и сложной геометрией данных (например, метод NAUS), а также создание на его основе интегрированного программного пакета для многоэтапного прогнозирования является актуальной научно-технической задачей. Реализация таких инструментов не только повышает точность медицинских прогнозов в среднем на 16%, но и позволяет оптимизировать распределение донорских ресурсов, что имеет прямое социально-экономическое значение и способствует повышению качества жизни пациентов. Кроме того, эти нововведения служат основой для обновления клинических протоколов, укрепления междисциплинарного взаимодействия и повышения общей эффективности трансплантологии.

Цель диссертационной работы разработка метода балансировки данных и программного пакета на основе моделей машинного обучения для повышения точности прогнозирования клинических исходов почечной трансплантации.

Задачи исследования:

- Разработка метода балансировки классов для обработки медицинских данных, включая данные по ведущим нозологическим формам.
- Реализация пяти моделей машинного обучения, обученных на специфических когортах пациентов по длительности выживаемости трансплантата.
- Разработка программного пакета и интеграция в него обученных моделей для практического прогнозирования выживаемости органа после операции.

Методы исследования: апробированные методы интеллектуального анализа данных (машинное обучение, Data Mining), методы статистического анализа, а также проведение численных экспериментов.

Объект исследования

Большие медицинские данные в области нефрологии.

Предмет исследования

Алгоритм балансировки классов NAUS (Noise-Aware Undersampling with Subsampling) и программные средства для прогнозирования рисков на основе анализа медицинских показателей.

Научная новизна полученных результатов:

разработан метод балансировки классов, применимый к различным типам медицинских данных и обеспечивающий минимизацию ошибок прогнозирования, что подтверждено на примере обработки данных по ключевым нозологическим формам в трансплантологии.

Положения, выносимые на защиту:

Разработанный метод балансировки классов, применимый к различным типам медицинских данных, обеспечивает минимизацию

ошибок прогнозирования, что подтверждено ростом метрик точности прогностических моделей в среднем на 16% по сравнению с существующими аналогами.

Теоретическая и практическая значимость исследования

Теоретическая значимость работы состоит в исследовании методов обработки несбалансированных больших медицинских данных и разработке алгоритма NAUS, предлагающего многостадийный подход к подготовке данных с фильтрацией шума и радиальным недосэмплированием. Практическая значимость заключается в возможности применения разработанного программного пакета для прогнозирования рисков терминальной почечной недостаточности. Алгоритм NAUS позволяет эффективно обрабатывать сильно смещенные выборки, повышать точность моделей машинного обучения и снижать вычислительные затраты при работе с Big Data.

Достоверность научных результатов:

Достоверность обеспечивается использованием строгих методов интеллектуального анализа данных и статистического анализа. Результаты подтверждаются численными экспериментами по устойчивому балансу классов, а также успешным сопоставлением с существующими методами субдискретизации (undersampling).

Связь темы с планами научно-исследовательских программ

Диссертационная работа выполнена в рамках грантового финансирования научных исследований:

- «Персональное прогнозирование выживаемости трансплантата у реципиентов трансплантата почки перед трансплантацией с применением методов машинного обучения» (№ гос. регистрации 0122РК00958, 2022-2024 гг.).
- «Разработка алгоритмического механизма для интеллектуально-программного модуля, использующего искусственный интеллект для обработки и анализа данных, связанных с трансплантацией почки» (№ гос. регистрации 0124РК00417, НАО «КазНУ имени аль-Фараби», 2024-2026 гг.).

Структура и содержание диссертационной работы

Диссертация состоит из введения, четырех основных глав, заключения и списка использованных источников.

Во введении обосновывается актуальность темы, чётко определяются цель и задачи исследования, определяются объект и предмет исследования, раскрывается научная новизна и практическое значение работы, а также предлагаются положения, выносимые на защиту. В этом разделе особенно подчёркивается необходимость современных медицинских исследований, поскольку объём больших данных быстро растёт и их эффективный анализ улучшает клинические решения.

В первой главе проведён комплексный аналитический обзор современных методов обработки больших медицинских данных. В биомедицине исследованы проблемы классового дисбаланса и гетерогенности данных. Проанализированы существующие программные решения для мониторинга состояния пациентов после трансплантации почки и выявлены ограничения классических статистических методов по сравнению с алгоритмами машинного обучения. В ходе обзора всесторонне рассмотрены сильные и слабые стороны методов, что укрепило основу исследования.

Во второй главе предложено теоретическое обоснование и реализация метода выравнивания данных NAUS (Noise-Aware Undersampling with Subsampling). Предложены результаты верификации метода на контрольных медицинских наборах данных, подтверждающие его преимущество в сохранении информации в выборке. Этот подход, учитывая шум, ориентирован на повышение качества данных, поэтому усиливает устойчивость моделей.

В третьей главе оописан процесс разработки пяти прогностических моделей. Проанализированы алгоритмы для различных временных горизонтов выживания трансплантата. Предложена кросс-валидация и оценка моделей по метрикам AUC и F1-score, которая показала высокую прогностическую точность в условиях динамически изменяющихся показателей пациента. Модели предназначены для предварительного выявления риска трансплантации, что облегчает работу врачей.

В четвертой главе предложена архитектура и программная реализация разработанного пакета. Приведены результаты экспериментальных испытаний системы в условиях, близких к клинической практике, и продемонстрирована эффективность интеграции моделей в процесс принятия врачебных решений. Пакет благодаря модульной структуре легко адаптируется и готов к широкому использованию в больницах.

Публикации по результатам исследования

По теме диссертации опубликованы 7 печатных работ, в том числе 2 в изданиях, рекомендованных КОКСНВО МНВО РК, 3 работы в рецензируемых журналах, входящих в международную базу цитирования «Web of Science» и/или «SCOPUS», и 2 работы в сборниках международных конференций.

- Salybekov A., Yerkos A., Kunikeyev A.D., Buribayev Z., Mochida Y., Ishioka K., Hidaka S., Kobayashi S. Predictive Modeling of Graft Failure Risk in Deceased Donor Kidney Transplants: Leveraging Machine Learning for Improved Outcomes and Data-Driven Insights: TH-PO036 // Journal of the American Society of Nephrology. - 2023. - Т. 34. - №. 11S. - С. 99 (ASN Kidney Week, November 02, 2023)

- Imanbek R., Buribayev Z., Yerkos A., Shaikalamova S., Abdirazak Z., Dauletiya D. HRBwSD: Hybrid radial based with standard deviation // 2024 IEEE 4th International Conference on Smart Information Systems and Technologies (SIST). - IEEE, 2024. - С. 487-491.
- Buribayev Z., Shaikalamova S., Yerkos A., Imanbek R. EKMGS: A HYBRID CLASS BALANCING METHOD FOR MEDICAL DATA PROCESSING //Scientific Journal of Astana IT University. - 2024. - С. 5-16.
- Buribayev Z., Yerkos A., Zhetpisbay Z. NEW APPROACH TO ADDRESSING CLASS IMBALANCE IN MEDICAL DATASETS CONSIDERING SPECIFICS //Scientific Journal of Astana IT University. - 2025. - С. 78-87.
- Salybekov A. A., Yerkos A., Sedlmayr M., Wolfien M. Ethics and algorithms to navigate AI's emerging role in organ transplantation //Journal of Clinical Medicine. - 2025. - Т. 14. - №. 8. - С. 2775 (CiteScore: 5.2 / Impact Factor: 2.9; Q1)
- Salybekov A.A., Wolfien, M., Yerkos, A., Buribayev, Z., Hidaka, S., & Kobayashi, S. Phase-Specific Kidney Graft Failure Prediction with Machine Learning Model //Frontiers in Artificial Intelligence. - Т. 8. - С. 1682639 (CiteScore: 7.3 / Impact Factor: 4.7; Q2)
- Buribayev Z., Yerkos A., Zhetpisbay Z., Wolfien M. Noise-Aware Undersampling for imbalanced medical data (NAUS) //Informatics in Medicine Unlocked. - 2026. - С. 101731 (CiteScore: 8.7 / Impact Factor: 4.3; Q2)