TECHNISCHE UNIVERSITÄT DRESDEN

Universitätsklinikum Carl Gustav Carus
DIE DRESDNER.

Center for Medical Informatics

University Hospital Carl Gustav Carus - Center for Medical Informatics · 01304 Dresden Germany

Dr.-Ing. Markus Wolfien
Telefon: 0351 458-877723
Telefax: 0351 458-5738
markus.wolfien@ukdd.de
markus.wolfien@tu-dresden.de

Joint institution of the Faculty of Medicine at the Technical University of Dresden and University Hospital Carl Gustav Carus Dresden

Business Division

Information Technology

Data Integration Center

Professorship for Medical Informatics at the Institute for Medical Informatics and Biometry

Department of Information Technology

Postal address
Zentrum für Medizinische Informatik
01304 Dresden

Zentrum für Medizinische Informatik

PhD Review for Ms. Ainur Yerkos

This report is presented to confirm my strong support of Ms. Ainur Yerkos and to document her progress and achievements under my international supervision from 2022 to the present. Throughout this period, Ms. Yerkos has demonstrated a consistently high level of scientific rigor, independence, and intellectual curiosity, coupled with outstanding reliability and an exceptional capacity to work across disciplines (Specialty: 8D06102 - Computer Engineering). Our collaboration has been characterized by regular scientific exchange and a productive international research workflow, resulting in multiple peer-reviewed publications and substantial methodological and applied contributions in machine learning for biomedical data.

Overview and Objectives of the Doctoral Work

Ms. Yerkos' doctoral work, entitled "Development of a method and a software package for big medical data processing" is situated at the interface of machine learning, bioinformatics, statistics, and medical decision support. Her research addresses a central challenge in real-world biomedical data science: developing reliable algorithms that remain robust under conditions that frequently occur in clinical practice, including heterogeneous data sources, time-dependent risk patterns, missingness, and severe class imbalance driven by rare events. The overarching objective of her work is to develop and validate methodological approaches that improve the reliability and interpretability of predictive models in clinically relevant settings, with a particular focus on transplantation medicine as an exemplary high-impact domain.

The interdisciplinary nature of this doctoral trajectory is demanding, because it requires technical mastery of machine learning methods as well as domain understanding sufficient to frame problems correctly, interpret results

responsibly, and communicate implications for clinical processes. Ms. Yerkos has managed this challenge very successfully. She has continuously demonstrated the ability to translate clinical questions into technically sound modeling strategies and, conversely, to communicate methodological constraints and opportunities back into the medical context in a meaningful and pragmatic way.

Scientific Contributions and Key Achievements

A major applied contribution of Ms. Yerkos' work is reflected in the peer-reviewed article on *phase-specific kidney graft failure prediction*, which proposes a time-dynamic approach to post-transplant surveillance. Instead of treating graft failure prediction as a static, single-timepoint problem, the study develops models tailored to defined post-transplant intervals (0–3 months, 3–9 months, 9–15 months, 15–39 months, and 39–72 months). This "*phase-specific*" framing is practically important, because it aligns prediction with clinically meaningful time windows and enables differentiated follow-up strategies along the patient journey. The reported predictive performance illustrates the value of this approach, particularly in the mid-term period, where the models achieved the highest accuracy (for example, ROC AUC approximately 0.92 in the 9–15-month window, accompanied by strong classification quality reflected by an F1 score around 0.85). These results support the central conclusion that phase-specific modeling can provide more robust decision support for identifying high-risk patients and tailoring follow-up protocols, thereby improving post-transplant care strategies.

A second major contribution is Ms. Yerkos' work on methodological innovation for imbalanced medical datasets, formalized in the open-access publication introducing *Noise-Aware Undersampling with Subsampling (NAUS)*. Severe class imbalance is a recurring obstacle in clinical machine learning, particularly when the event of interest is rare but clinically critical. NAUS addresses this by combining clustering, noise removal, and Tomek-link identification to construct refined subsamples and systematically remove redundant and noisy observations. The method was evaluated on multiple disease-related datasets, including chronic kidney disease, liver disease, and heart disease, and compared against established oversampling approaches, such as SMOTE, ADASYN, and LoRAS as well as classical undersampling baselines. By explicitly integrating "noise awareness" into the undersampling process and validating the approach across multiple classifiers (including Random Forest and LightGBM), the work contributes a practical and reusable strategy to enhance the reliability of predictive modeling under real-world imbalance conditions. This line of research is directly aligned with the needs of biomedical informatics, where robust

handling of imbalanced outcomes often determines whether models are clinically usable.

In addition to original research, Ms. Yerkos has also contributed to interdisciplinary scholarship at the interface of medical practice, algorithmic development, and responsible innovation. The joint review article "*Ethics and Algorithms to Navigate AI's Emerging Role in Organ Transplantation*" summarizes and critically contextualizes how machine learning, deep learning, and generative AI can support transplantation workflows, including outcome prediction, organ allocation and donor-recipient matching, precision pathology, and personalization of immunosuppressive regimens. Importantly, the review also addresses ethical implications, highlighting the need for transparency, accountability, data privacy and security, and mitigation of bias to ensure fair and equitable deployment of AI tools in transplantation. Ms. Yerkos' involvement here demonstrates that her competencies extend beyond algorithmic implementation to include the broader clinical and societal framing that is increasingly expected from researchers working in high-stakes medical AI.

Observations on Research Skills and Scientific Maturity

Across our supervision period, Ms. Yerkos consistently displayed strong intrinsic motivation and a proactive research attitude. She regularly contributed original ideas, independently proposed methodological improvements, and showed excellent responsiveness to feedback. A particularly positive aspect is her ability to incorporate comments efficiently and thoughtfully, often improving both methodological clarity and scientific narrative quality in revised versions of analyses and manuscripts.

Her international research stay in Dresden provided further direct evidence of her maturity as an independent researcher. She integrated rapidly into a new environment, established a productive work routine, contributed meaningfully to collaborative writing, and, at the same time, advanced independent methodological work with NAUS. The stay also highlighted her strong interdisciplinary communication skills, as she actively engaged in discussions bridging computational methodology and clinical transplantation practice.

Conclusion and Recommendation

In summary, Ms. Ainur Yerkos has demonstrated excellent technical progress and strong personal development since 2022. Her work is characterized by methodological depth, applied relevance, and an unusual strength in navigating interdisciplinary challenges between machine learning and clinical medicine. Her publications provide tangible evidence of her ability to contribute both novel

methods for imbalanced clinical data and clinically meaningful predictive modeling approaches for transplantation outcomes, complemented by thoughtful engagement with the ethical dimensions of medical AI.

Based on her achievements, her scientific maturity, and her proven ability to work effectively in international collaborations, I strongly support Ms. Yerkos and recommend her doctoral work for the next steps toward completion.

Yours sincerely,

Medizinische Fakultät Carl Gustav Carus
der Technischen Universität Dresden
Professur für Medizinische Informatik am IMB
Fetscherstr. 74 • 01307 Dresden
Tel.: 0351/ 458 24 37 Fax: 0351/ 458 57 38

Dr.-Ing. Markus Wolfien

International Scientific Supervisor
Group Leader for Data Science
TUD Dresden University of Technology Dresden
Center for Scalable Data Analytics and Artificial Intelligence (ScaDS.AI)
Institute for Medical Informatics and Biometry
University Hospital Carl Gustav Carus Dresden

ДРЕЗДЕНСКИЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ
Центр медицинской информатики
Университетская клиника имени Карла Густава Каруса
Центр медицинской информатики, 01304, Дрезден, Германия

Др.-инж. Маркус Вольфиен
Телефон: 0351 458-877723
Телефакс: 0351 458-5738
markus.wolfien@ukdd.de
markus.wolfien@tu-dresden.de

Отзыв на PhD работу Еркос Айнұр

Этот отчет представлен для подтверждения моей решительной поддержки Еркос Айнұр и документирования её прогресса и достижений под моим международным руководством с 2022 года по настоящее время. На протяжении этого периода Еркос Айнұр демонстрировала неизменно высокий уровень научной строгости, независимости и интеллектуального любопытства в сочетании с выдающейся надежностью и исключительной способностью работать на стыке различных дисциплин (специальность: 8D06102 «Компьютерная инженерия»). Наше сотрудничество характеризовалось регулярным научным обменом и продуктивным международным исследовательским процессом, результатом которого стали многочисленные рецензируемые публикации, а также существенный методологический и прикладной вклад в области машинного обучения для биомедицинских данных.

Обзор и цели докторской работы

Докторская работа Еркос Айнұр под названием «Разработка метода и программного пакета для обработки больших медицинских данных» находится на стыке машинного обучения, биоинформатики, статистики и медицинской поддержки принятия решений. Её исследование направлено на решение центральной проблемы современной биомедицинской науки о данных: разработку надежных алгоритмов, сохраняющих устойчивость в условиях, часто встречающихся в клинической практике, включая гетерогенные источники данных, временные паттерны рисков, пропуски в данных и выраженный дисбаланс классов, вызванный редкими событиями. Основная цель её работы — разработка и валидация методологических подходов, повышающих надежность и интерпретируемость прогностических моделей в клинически значимых условиях, с особым акцентом на медицину трансплантации как пример высокозначимой области.

Междисциплинарный характер этой траектории обучения требует больших усилий, так как он предполагает как техническое мастерство в методах машинного обучения, так и понимание предметной области, достаточное для правильной постановки задач, ответственной интерпретации результатов и

анализа их влияния на клинические процессы. Ерқос Айнұр успешно справилась с этой задачей. Она постоянно демонстрировала способность переводить клинические вопросы в технически обоснованные стратегии моделирования и, в свою очередь, содержательно и прагматично передавать информацию о методологических ограничениях и возможностях обратно в медицинский контекст.

Научный вклад и ключевые достижения

Важный прикладной вклад работы Ерқос Айнұр отражен в рецензируемой статье о «phase-specific kidney graft failure prediction», в которой предлагается подход с учетом временной динамики для наблюдения за пациентами после трансплантации. Вместо того чтобы рассматривать прогнозирование отказа трансплантата как статичную задачу, решаемую в один момент времени, в исследовании разрабатываются модели, адаптированные к определенным интервалам после операции (0–3 месяца, 3–9 месяцев, 9–15 месяцев, 15–39 месяцев и 39–72 месяца). Такая «фазоспецифичная» структура важна на практике, так как она согласует прогнозы с клинически значимыми временными окнами и позволяет применять дифференцированные стратегии последующего наблюдения на пути пациента.

Представленные показатели прогностической эффективности иллюстрируют ценность этого подхода, особенно в среднесрочном периоде, где модели достигли наивысшей точности (например, ROC AUC приблизительно 0,92 в окне 9–15 месяцев при высоком качестве классификации, отраженном показателем F1 около 0,85). Эти результаты подтверждают центральный вывод о том, что фазоспецифичное моделирование может обеспечить более надежную поддержку принятия решений для выявления пациентов высокого риска и адаптации протоколов наблюдения, тем самым улучшая стратегии ухода после трансплантации.

Вторым крупным вкладом является работа Ерқос Айнұр над методологическими инновациями для несбалансированных медицинских наборов данных, оформленная в публикации открытого доступа, представляющей «Noise-Aware Undersampling with Subsampling (NAUS)». Выраженный дисбаланс классов является повторяющимся препятствием в клиническом машинном обучении, особенно когда интересующее событие редко, но критически важно с клинической точки зрения. NAUS решает эту проблему путем сочетания кластеризации, удаления шума и идентификации связей Томека для построения уточненных подвыборок и систематического удаления избыточных и зашумленных наблюдений.

Метод был протестирован на нескольких наборах данных, связанных с заболеваниями, включая хроническую болезнь почек, заболевания печени и сердца, и сопоставлен с устоявшимися подходами к избыточной выборке (oversampling), такими как SMOTE, ADASYN и LORAS, а также с

классическими базовыми методами недостаточной выборки (undersampling). Путем явного интегрирования «noise awareness» (учета шума) в процесс недостаточной выборки и валидации подхода на нескольких классификаторах (включая Random Forest и LightGBM), работа вносит вклад в виде практичной и пригодной для повторного использования стратегии повышения надежности прогностического моделирования в условиях реального дисбаланса. Эта линия исследований напрямую соответствует потребностям биомедицинской информатики, где устойчивая обработка несбалансированных результатов часто определяет, будут ли модели пригодны для клинического использования.

В дополнение к оригинальным исследованиям, Ерқос Айнұр также внесла вклад в междисциплинарную науку на стыке медицинской практики, разработки алгоритмов и ответственных инноваций. Совместная обзорная статья «Ethics and Algorithms to Navigate Al's Emerging Role in Organ Transplantation» обобщает и критически контекстуализирует то, как машинное обучение, глубокое обучение и генеративный ИИ могут поддерживать рабочие процессы трансплантации, включая прогнозирование исходов, распределение органов и подбор донора и реципиента, точную патологию и персонализацию режимов иммуносупрессии. Важно отметить, что в обзоре также рассматриваются этические последствия, подчеркивая необходимость прозрачности, подотчетности, конфиденциальности и безопасности данных, а также смягчения предвзятости для обеспечения справедливого и равноправного внедрения инструментов ИИ в трансплантологию. Участие Ерқос Айнұр в этой работе демонстрирует, что её компетенции выходят за рамки алгоритмической реализации и включают более широкое клиническое и социальное понимание, которое все чаще ожидается от исследователей, работающих в области высокорискового медицинского ИИ.

Наблюдения за исследовательскими навыками и научной зрелостью

На протяжении всего периода нашего руководства Ерқос Айнұр неизменно демонстрировала сильную внутреннюю мотивацию и проактивный исследовательский подход. Она регулярно предлагала оригинальные идеи, независимо выдвигала методологические улучшения и проявляла отличную реакцию на обратную связь. Особенно позитивным аспектом является её способность эффективно и вдумчиво учитывать замечания, что часто приводило к улучшению как методологической ясности, так и качества научного изложения в пересмотренных версиях анализов и рукописей.

Её международная исследовательская стажировка в Дрездене стала дополнительным прямым доказательством её зрелости как независимого исследователя. Она быстро интегрировалась в новую среду, установила продуктивный рабочий график, внесла значимый вклад в совместное написание работ и одновременно продвинула независимую методологическую работу по NAUS. Стажировка также высветила её сильные навыки междисциплинарной коммуникации, так как она активно участвовала в дискуссиях, связывающих вычислительную методологию и клиническую практику трансплантации.

Заключение и рекомендации

Подводя итог, Ерқос Айнұр продемонстрировала отличный технический прогресс и сильное личностное развитие с 2022 года. Её работа характеризуется методологической глубиной, прикладной значимостью и необычайной способностью преодолевать междисциплинарные барьеры между машинным обучением и клинической медициной. Её публикации являются наглядным свидетельством её способности предлагать как новые методы для несбалансированных клинических данных, так и клинически значимые подходы к прогностическому моделированию исходов трансплантации, дополненные вдумчивым анализом этических аспектов медицинского ИИ.

Основываясь на её достижениях, научной зрелости и проверенной способности эффективно работать в рамках международного сотрудничества, я решительно поддерживаю Ерқос Айнұр и рекомендую её докторскую работу для перехода к следующим этапам завершения.

Искренне Ваш,
Др.-инж. Маркус Вольфиен

Медицинский факультет имени Карла Густава Каруса
Технического университета Дрездена
Профессор медицинской информатики в IMB
(подпись) Фетчерштрассе 74, 01307 Дрезден
Тел.: 0351 458 2437

Международный научный руководитель
Руководитель группы по анализу данных (Data Science)
Дрезденский технический университет (TUD)
Центр масштабируемой аналитики данных и искусственного интеллекта (ScaDS.AI)
Институт медицинской информатики и биометрии
Университетская клиника имени Карла Густава Каруса, Дрезден