Еркос Айнұрдың 8D06102 – Компьютерлік инженерия білім беру бағдарламасы бойынша философия докторы (PhD) дәрежесін алу үшін ұсынылған «Үлкен медициналық деректерді өңдеуге арналған әдіс пен бағдарламалық пакетті әзірлеу» тақырыбындағы диссертациялық жұмысына

АҢДАТПА

Зерттеу тақырыбының өзектілігі. Денсаулық сақтау жүйелерін жаһандық цифрландыру жағдайында диагностика, емдеу және пациенттерді ұзақ мерзімді бақылау процесінде жинақталатын медициналық деректер көлемі экспоненциалды түрде артып келеді. Жиналу жылдамдығымен, құрылымдарының әртүрлілігімен және интерпретациялау күрделілігімен сипатталатын бұл деректер жиынтығы Big Data-ның классикалық үлгісі болып табылады. Мұндай деректерді тиімді пайдалану предиктивті медицинаны дамытуға және дәрігерлік шешімдерді қолдау жүйелерін құруға теңдессіз мүмкіндіктер ашады. Бұл мүмкіндіктер клиницистерге жеке пациенттердің ерекшеліктерін ескере отырып, алдын ала қауіптерді анықтауға және емдеу траекторияларын жетілдіруге жол ашады, нәтижесінде медициналық практика жаңа деңгейге көтеріледі.

Қазіргі медицинаның ең маңызды және ресурс қажет ететін салаларының бірі - трансплантология, атап айтқанда бүйрек трансплантациясы. Донорлық мүшелердің тапшылығы және трансплантаттың қабылданбау қаупі өміршеңдікті болжаудың жоғары дәлдіктегі құралдарын енгізуді талап етеді. Классикалық статистикалық тәсілдердің болуына қарамастан, олар көбінесе ұзақ мерзімді кезеңдегі тәуекел факторларының бейсызық тәуелділіктерін және динамикалық сипатын ескере алмайды. Бұл тәсілдердің шектеулері трансплантология саласындағы шешімдердің сапасын төмендетіп, дәрігерлердің жұмысын қиындатады. Дегенмен, бұл салада деректерді интеллектуалды талдау (Data Mining) әдістерін қолдану жолында күрделі технологиялық кедергілер бар.

Біріншіден, медициналық таңдамалар өз табиғаты бойынша теңгерімсіз болып келеді: жағымсыз нәтижелері бар жағдайлар (мүшенің қабылданбауы, өлім) тұрақты жағдайлар санынан айтарлықтай аз. Бұл стандартты машиналық оқыту алгоритмдері жоғары жалпы дәлдікті көрсеткенімен, сирек кездесетін маңызды клиникалық оқиғаларды болжауда тиімсіз болуына әкеледі. Қолданыстағы теңгерімдеу әдістері көбінесе шу енгізеді немесе құнды ақпараттың жоғалуына әкеледі.

Екіншіден, дәстүрлі бағдарламалық шешімдер гетерогенді деректерді өңдеу үшін жеткілікті икемділікке ие емес және операциядан кейінгі әртүрлі кезеңдердегі пациент жағдайының өзгеру динамикасын

ескермейді. Олар көбінесе статикалық модельдерге сүйенеді, бұл реалды клиникалық процестердің уақытша эволюциясын толық бейнелеуге мүмкіндік бермейді.

Осылайша, шумен және деректердің күрделі геометриясымен тиімді жұмыс істей алатын кластарды теңгерімдеудің арнайы әдісін (мысалы, NAUS әдісі) әзірлеу, сондай-ақ оның негізінде көп сатылы болжауға арналған интеграцияланған бағдарламалық пакетті жасау өзекті ғылыми-техникалық міндет болып табылады. Мұндай құралдарды іске асыру медициналық болжамдардың дәлдігін орта есеппен 16%-ға арттырып қана қоймай, донорлық ресурстарды бөлуді оңтайландыруға мүмкіндік береді, бұл тікелей әлеуметтік-экономикалық маңызға ие және пациенттердің өмір сүру сапасын арттыруға ықпал етеді. Сондай-ақ, бұл жаңалықтар клиникалық протоколдарды жаңғыртуға, командааралық ынтымақтастықты күшейтуге және трансплантологияның жалпы тиімділігін жоғарылатуға негіз болады.

Диссертациялық жұмыстың мақсаты бүйрек трансплантациясының клиникалық нәтижелерін болжау дәлдігін арттыру үшін деректерді теңгерімдеу әдісін және машиналық оқыту модельдеріне негізделген бағдарламалық пакетті әзірлеу.

Зерттеу міндеттері:

- Медициналық деректерді өңдеу үшін кластарды теңгерімдеу әдісін әзірлеу.
- Трансплантаттың өміршеңдік ұзақтығы бойынша пациенттердің ерекше когорталарында оқытылған бес машиналық оқыту моделін іске асыру.
- Бағдарламалық пакетті әзірлеу және оған операциядан кейінгі мүшенің өміршеңдігін болжау үшін оқытылған модельдерді интеграциялау.

Зерттеу объектісі

Нефрология саласындағы үлкен медициналық деректер.

Зерттеу пәні

NAUS (Noise-Aware Undersampling with Subsampling) кластарды теңгерімдеу алгоритмі және медициналық көрсеткіштер негізінде тәуекелдерді болжауға арналған бағдарламалық құралдар.

Алынған нәтижелердің ғылыми жаңалығы:

Медициналық деректердің әртүрлі түрлеріне қолданылатын және болжау қателіктерін азайтуды қамтамасыз ететін кластарды теңгерімдеу әдісі әзірленді.

Қорғауға шығарылатын тұжырым:

Әзірленген кластарды теңгерімдеу әдісі болжау қателіктерін барынша азайтуды қамтамасыз етеді, бұл қолданыстағы аналогтармен салыстырғанда дәлдік метрикаларының орта есеппен 16%-ға өсуімен расталады.

Зерттеудің теоретикалық және практикалық маңыздылығы

Теоретикалық маңыздылығы теңгерімсіз үлкен медициналық деректерді өңдеу әдістерін зерттеуден және шуды сүзгілеу мен радиалды субдискретизацияны ұсынатын NAUS алгоритмін әзірлеуден тұрады. Практикалық маңыздылығы бүйрек жеткіліксіздігінің қаупін болжауға арналған бағдарламалық пакетті қолдану мүмкіндігінде.

Ғылыми нәтижелердің сенімділігі:

Дәлдік қатаң интеллектуалдық деректер талдау әдістері мен статистикалық талдау әдістерін қолдану арқылы қамтамасыз етіледі. Нәтижелер класс балансының тұрақтылығы бойынша сандық тәжірибелермен расталады, сондай-ақ субдискретизация (undersampling) әдістерімен сәтті салыстыру арқылы расталады.

Тақырыптың ғылыми-зерттеу бағдарламалары жоспарларымен байланысы

Диссертациялық жұмыс ғылыми зерттеулерді гранттық қаржыландыру аясында орындалды:

- «Машиналық оқыту арқылы операция дейінгі бүйрек трансплантатының өмір сүруін жеке болжалдау әдісі» (мемлекеттік тіркеу № 0122РК00958, 2022-2024 жж.).
- «Бүйрек трансплантациясына қатысты деректерді өңдеу және талдау үшін жасанды интеллект пайдаланатын интеллектуалды бағдарламалық модуль үшін алгоритмдік механизмді әзірлеу» (мемлекеттік тіркеу № 0124РК00417, әл-Фараби атындағы ҚазҰУ, 2024-2026 жж.).

Диссертациялық жұмыстың құрылымы мен мазмұны

Диссертация кіріспе, төрт негізгі тарау, қорытынды және пайдаланылған дереккөздер тізімінен тұрады.

Кіріспеде тақырыптың өзектілігі негізделеді, зерттеудің мақсаты мен міндеттері айқындалады, зерттеу объектісі мен пәні анықталады, жұмыстың ғылыми жаңалығы мен практикалық маңызы ашылады, сондай-ақ қорғауға шығарылатын ережелер ұсынылады. Бұл бөлімде заманауи медициналық зерттеулердің қажеттілігі ерекше атап өтіледі, себебі үлкен деректер көлемі жылдам өсіп келеді және оларды тиімді талдау клиникалық шешімдерді жақсартады.

Бірінші тарауда үлкен медициналық деректерді өңдеудің заманауи әдістерінің кешенді аналитикалық шолуы жүргізілді. Биомедицинада класс теңсіздігі мен деректердің гетерогенділігі мәселелері зерттелді. Бүйрек трансплантациясынан кейінгі науқастардың жағдайын бақылауға арналған қолданыстағы бағдарламалық шешімдер талданды және классикалық статистикалық әдістердің машиналық оқыту алгоритмдерімен салыстырғандағы шектеулері анықталды. Шолу барысында әдістердің күшті және әлсіз жақтары жан-жақты қарастырылды, бұл зерттеудің негізін нығайтты.

Екінші тарауда деректерді теңестіру әдісі NAUS (Noise-Aware

Undersampling with Subsampling) теориялық негіздемесі мен жүзеге асырылуы ұсынылды. Әдістің бақылау медициналық деректер жиынтықтарындағы верификация нәтижелері ұсынылды, ол үлгідегі ақпаратты сақтаудағы артықшылығын растайды. Бұл тәсіл шуды ескере отырып, деректердің сапасын арттыруға бағытталған, сондықтан модельдердің тұрақтылығын күшейтеді.

Үшінші тарауда бес болжамдық модельдерді әзірлеу процесі сипатталды. Трансплантанттың тірі қалуының әртүрлі уақыт горизонттары үшін алгоритмдер талданды. Кросс-валидация әдісі және AUC пен F1-score метрикалары бойынша модельдерді бағалау ұсынылды, ол науқастың динамикалық өзгеретін көрсеткіштер жағдайында жоғары болжамдық дәлдікті көрсетті. Модельдер трансплантация қаупін алдын ала анықтауға арналған, бұл дәрігерлердің жұмысын жеңілдетеді.

Төртінші тарауда әзірленген пакеттің архитектурасы мен бағдарламалық жүзеге асырылуы ұсынылды. Жүйенің клиникалық практикаға жақын жағдайларда эксперименттік сынақ нәтижелері келтірілді және модельдерді дәрігерлік шешім қабылдау процесіне интеграциялаудың тиімділігі көрсетілді. Пакет модульдік құрылымы арқасында оңай бейімделеді және ауруханаларда кеңінен қолданылуға дайын.

Зерттеу нәтижелері бойынша жарияланымдар

Диссертация тақырыбы бойынша 7 баспа жұмысы жарияланды, оның ішінде 2 жұмыс ҒЖБССҚК ҚР ұсынған басылымдарда, 3 жұмыс «Web of Science» және/немесе «SCOPUS» халықаралық дәйексөз базасына кіретін рецензияланған журналдарда және 2 жұмыс халықаралық конференциялар жинақтарында жарияланды.

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)