

ВІДГУК
офіційного опонента
доктора технічних наук **Березького Олега Миколайовича**
на дисертацію
Мельникової Наталії Іванівни
**«Моделі та методи підтримки персоналізованих рішень у медичних
системах»**,
подану на здобуття наукового ступеня доктора технічних наук
за спеціальністю
05.13.23 – системи та засоби штучного інтелекту

1. Актуальність теми дослідження

Штучний інтелект став рушієм четвертої промислової революції. Широке впровадження досягнень штучного інтелекту у різні сфери дає можливість суттєво покращити життя людей на планеті Земля. Помітною тенденцією в науці є процес гібридизації різних областей знань. Не є виключенням медицина. Тільки поєднання медицини та інформаційних технологій дає можливість автоматизувати процес постановки діагнозу, підвищити його швидкодію та достовірність.

Останнім часом набула розповсюдження персоналізована медицина, яка орієнтована на обробку мультимодальних медичних даних, що забезпечує керованість цього процесу. Персоналізація стає все більш важливою в науковому співтоваристві з багатьох причин, включаючи нові методи оптимізації та алгоритми, які допомагають скоротити медичні витрати та забезпечити якісне та ефективне медичне обслуговування. Дане дослідження спрямоване на вирішення проблеми формалізації стану хворого та формалізації динаміки процесів під час лікування чи реабілітації, що дозволить оптимізувати лікувальні процеси, проаналізувати індивідуальні характеристики пацієнта та прогнозувати можливі персоналізовані рішення щодо надання медичної допомоги на основі оцінки стану здоров'я пацієнта.

Проблема персоналізації даних вимагає чіткої стратегії визначення основних етапів обробки інформації, консолідації мультимодальних даних, порівняння структур даних, розроблення методів опрацювання історичних та теперішніх даних. При роботі з таким типом даних характерно урахування існуючих стандартів їхнього опрацювання, зберігання та передачі персоналізованих даних, які різняться за структурою та джерелами надходження. Ще однією з ключових характеристик персоналізованих медичних даних є їхня розмірність, а саме малі та великі набори даних, при опрацюванні яких спостерігається низька повторюваність результатів під час пошуку персоналізованих рішень та висока обчислювальна складність.

Існуючі методи штучного інтелекту вирішують лише окремі підзадачі: врахування невизначеності, мультимодальності, заповнення пропусків даних, забезпечення високої точності прогнозування під час опрацювання персоналізованих медичних даних. За цієї особливості та за одержаними результатами дослідження виникає потреба у розробленні нових підходів та методів щодо пошуку персоналізованих рішень для підвищення точності прогнозування цільових показників стану пацієнта, консолідації та аналізу медичних даних.

Дисертаційна робота Мельникової Наталії Іванівни присвячена вирішенню актуальної науково-прикладної проблеми розроблення нових чи удосконалення існуючих методів штучного інтелекту щодо оптимізації процесу обробки великих та малих мультимодальних медичних наборів даних для підтримки прийняття рішення.

2. Основні наукові результати досліджень та наукова новизна дисертації

Наукова новизна одержаних результатів полягає в тому, що обґрунтовано та розв'язано науково-прикладну задачу розроблення та удосконалення моделей, методів і засобів машинного навчання в задачах класифікації, кластеризації, прогнозування та візуалізації результатів опрацювання персональних даних для адаптації медичних рішень до пацієнта.

Наукові результати дисертації такі:

- 1) вперше розроблено модель відображення стану пацієнта в багатовимірному просторі умов, що за рахунок додаткових вимірів та параметрів забезпечило підвищення точності прогнозування;
- 2) вперше розроблено метод двоетапної обробки малих наборів даних на основі ієрархічного предиктора, який за рахунок кластеризації та прогнозування забезпечує підвищення точності опрацювання нових вхідних наборів даних;
- 3) вперше розроблено метод групування моделей машинного навчання, який, враховуючи основні ознаки та повторне навчання, забезпечує підвищення точності прогнозування даних;
- 4) вперше розроблено метод зменшення розмірності вхідних даних, який завдяки класифікаторам, асоціативним правилам та узагальненому рангу ознак на основі індексу Жакара, забезпечує підвищення точності вибору пріоритетних ознак на великих наборах даних;

- 5) удосконалено метод заповнення пропусків даних, який, використовуючи асоціативні та продукційні правила, забезпечує підвищення стійкості моделі до помилок даних;
- 6) удосконалено метод пошуку зміни стану пацієнта, який, використовуючи простір умов та аналіз характеристик стану, забезпечує підвищення точності підбору схеми лікування;
- 7) удосконалено метод консолідації мультимодальних даних, який за рахунок попереднього визначення структури та семантики даних забезпечує агрегування даних з різною структурою;
- 8) одержала подальшого розвитку теорія підтримки прийняття персоналізованих рішень, яка ґрунтується на розробленні моделі відображення стану пацієнта в багатовимірному просторі умов та розроблених методах двоетапної обробки малих наборів даних, групуванні моделей машинного навчання, зменшенні розмірності вхідних даних, заповненні пропусків даних, пошуку зміни стану пацієнта та консолідації мультимодальних даних, що забезпечило адаптацію прийняття персоналізованих медичних рішень.

3. Ступінь обґрунтованості наукових положень, висновків і рекомендацій, сформульованих у дисертації, їх достовірність і новизна

Дисертаційна робота Мельникової Н. І. виконана на високому науковому рівні, є завершеною науково-дослідницькою роботою, матеріал подано в логічній послідовності, поставлені задачі в межах визначеної науково-прикладної проблеми глибоко опрацьовані.

Аналіз змісту розділів, використаного методологічного та програмно-алгоритмічного інструментарію та способів його застосування дає підстави зробити висновок про належну обґрунтованість винесених дисертантом на захист основних наукових результатів.

Наукові положення, висновки та рекомендації, сформульовані у дисертації, є обґрунтованими за рахунок проведеного теоретичного аналізу, використання відомостей і положень, отриманих з науково-технічної літератури, а також вони підтверджені науковими та практичними результатами за рахунок відповідних матеріалів про впровадження дисертаційних досліджень в межах національних та міжнародних проектів, а також за рахунок апробації у медичних закладах охорони здоров'я.

Оцінюючи дисертаційну роботу в цілому, слід відзначити новизну розроблених та впроваджених результатів, що дало змогу вирішити

актуальну наукову проблему розроблення нових та удосконалення існуючих методів штучного інтелекту щодо оптимізації процесу обробки персоналізованих мультимодальних медичних наборів даних для підтримки прийняття лікарських рішень.

Розроблені моделі, методи і засоби підтримки персоналізованих рішень у медичних системах дають змогу підвищити прогнозовану точність та візуалізацію результатів опрацювання персональних даних щодо оцінки стану хворого, та забезпечити якісний процес прийняття персоналізованих медичних рішень.

4. Практичне значення результатів дисертаційної роботи

Практична цінність роботи така:

- 1) розроблено ієрархічний предиктор, який включає двоетапну обробку малих наборів даних, що забезпечує покращення стійкості моделі до нових вхідних даних у модулі системи підтримки прийняття рішень у лікуванні орфанних хворих «Реєстр первинних імунодефіцитів»;
- 2) розроблено алгоритми видобування PPD та імпутації відсутніх медичних персоналізованих даних, що забезпечують аналіз мультимодальних даних при паралельній реалізації в розподілених базах даних модуля системи оброблення та зберігання персоналізованих військових медичних даних;
- 3) розроблено інформаційну систему підтримки прийняття рішень у лікуванні орфанних хворих «Реєстр первинних імунодефіцитів»;
- 4) розроблено програмний модуль для підтримки прийняття рішень у визначенні схеми лікування пацієнтів з COVID-19.

5. Рекомендації щодо використання результатів дисертації

Розроблені моделі, методи та алгоритми можуть бути використані при розробці нових чи вдосконаленні вже існуючих медичних систем підтримки прийняття лікарських рішень, а також при проектуванні інтелектуальних систем, що базуються на опрацюванні мультимодальних персоналізованих даних та даних малої чи великої розмірності для забезпечення підвищення точності пошуку рішень у різних галузях.

6. Повнота викладення наукових положень, висновків і рекомендацій в опублікованих працях

Наведені в дисертації положення та результати в повній мірі опубліковані у 56 наукових публікаціях, із них 21 стаття, де 2 статті

опубліковано в журналах з Q1 та 3 статті опубліковано в журналах з Q2, 7 статей у фахових іноземних виданнях, 7 статей у наукових фахових виданнях України, 5 монографій, дві з яких – одноосібні, а також 35 публікацій у матеріалах конференцій, 26 з яких індексуються у наукометричних базах Scopus та Web of Science.

У наукових працях повною мірою подано всі розділи рецензованої дисертації.

Зміст реферату повністю відображає основні положення дисертації.

7. Оцінка змісту дисертаційної роботи, її завершеність.

Дисертаційна робота складається із вступу та шести розділів, висновків, списку використаних джерел з 302 найменуваннями та додатку, 116 рисунків та 32 таблиці. Повний обсяг дисертації 293 сторінки, у тому числі 239 сторінок основного тексту.

У вступі проведено аналіз проблеми опрацювання мультимодальних персоналізованих медичних даних і попередньої обробки даних та аналізу якості моделі даних, обґрунтовується актуальність теми, сформульована мета та основні задачі досліджень, подається короткий зміст роботи.

У першому розділі введено поняття персоналізації даних орієнтованих на пацієнта, що є вхідним набором процесу обробки медичних даних для пошуку персоналізованих рішень, що є основою забезпечення якісного та ефективного медичного обслуговування хворого. Проаналізовано існуючі практичні рішення, що відображають основні підходи до використання медичних даних та урахування персоналізації. Проведено порівняльний аналіз класичних методів, визначено обмеження існуючих методів опрацювання та аналізу великих та малих наборах мультимодальних медичних даних. Сформульовано науково-прикладну проблему та задачі дослідження.

У другому розділі введено модель простору станів пацієнта, яку представлено у вигляді гіперкуба, як відображення функціонального відношення загального стану пацієнта, розроблено метод пошуку шаблонів, що дозволяє зменшити трудомісткість і використовувати паралельний та розподілений режим для розрахунку, розроблено метод консолідації мультимодальних даних для забезпечення агрегації даних з різною структурою.

У третьому розділі проведено аналіз цільових даних та виявлені взаємозалежності між ними, що мають великий вплив на процеси опрацювання даних в системах діагностування та прогнозування станів пацієнта, розроблено метод двоетапної розробки на основі ієрархічного предиктора, що дає змогу підвищити точність процесу узагальнення

результатів на малих наборах даних, розроблено метод групування моделей машинного навчання, на основі стекінгової моделі, що базується на використанні випадкового лісу і на деформації метаознак і повторному навчанні на розширеному наборі даних, та забезпечує підвищення точності прогнозування даних та паралельної обробки даних як малої, так і великої розмірності.

У четвертому розділі проаналізовано існуючі методи відбору важливих ознак: фільтри, вбудовані алгоритми. Розроблено гібридний ансамбль вибору пріоритетних ознак на великих наборах даних, який складається з класифікаторів, асоціативних правил та узагальненого рангу ознак на основі індексу Жакара, що дозволяє уникнути кореляції ознак та збільшує узагальнення моделі. Розроблено метод зменшення розмірності вхідних даних, що дає змогу уникнути кореляції результатів слабких предикторів і збільшує узагальнення моделі.

У п'ятому розділі визначені типи аномалій та шляхи їхнього пошуку, що дає можливість диференціації підходів щодо аналізу персоналізованих медичних даних та визначені особливості їхнього застосування і основні проблеми, розроблено метод заповнення пропусків даних на основі ймовірнісних продукційних залежностей, що збільшує стійкість моделі до помилок в даних та забезпечення аналізу мультимодальних даних при паралельній реалізації в розподілених базах даних, що ґрунтується на застосуванні двох алгоритмів видобування PPD та імпутації відсутніх медичних персоналізованих даних.

У шостому розділі запропоновано архітектуру системи підтримки прийняття медичних рішень щодо прогнозування станів пацієнта на підставі опрацювання та аналізу персоналізованих медичних даних; подано результати розробки програмного модулю щодо прогнозування динаміки поширення COVID-19, що дозволяє аналізувати дані пацієнтів і на їх основі давати результат про наявність або відсутність COVID у пацієнта; подано результати імплементації системи для виявлення помилок, що дає змогу експериментувати із різними параметрами для моделі на одних і тих же даних.

У додатку наведені акти впровадження результатів дисертаційної роботи.

Загалом дисертаційна робота написана на високому професійному рівні. Результати досліджень обґрунтовані, подані структуровано та послідовно. Реферат відповідає змісту дисертаційної роботи.

8. Зауваження до дисертаційної роботи

Зауваження до дисертаційної роботи такі:

1. Необхідно більш детально було б описати статистичну вибірку. Доцільно (вже навіть у вступі) вказати кількість хворих, дані яких використано у дослідженні та, відповідно, розподіл за групами. Також вартувало б вказати їх діагнози відповідно до міжнародної класифікації хвороб або хоча би які групи захворювань підлягали обстеженню (серцево-судинна система, ендокринні захворювання і т. д.).

2. Розроблені у дисертаційному дослідженні методи та ансамблі моделей прогнозування даних для оцінки стану хворого доведено до практичної реалізації інформаційної системи підтримки прийняття рішень для лікування хворих з орфанними хворобами. Орфанні хвороби – це вроджені або набуті захворювання, які трапляються вкрай рідко – рідше ніж один випадок на 2000 населення країни. Тому, застосування комплексного використання інформаційних методик при цьому має бути обґрунтоване, адже кількість таких пацієнтів невелика. Наприклад, це може бути доцільно при чисельних обстеженнях хворого на орфанну патологію та отриманні значних за обсягом цифрових даних.

3. У першому розділі не проведено аналіз і порівняння існуючих систем підтримки прийняття рішень персоналізованих медичних систем.

4. Також у дисертаційному дослідженні необхідно було б використати (чи розробити власні) критерії ефективності застосованих підходів. Необхідно вказати, що саме конкретно покращилося у якості чи ефективності профілактики (або діагностики чи лікування) зазначених захворювань після застосування розроблених методів і моделей.

5. У роботі введено поняття простору станів пацієнта як багатовимірної системи і відповідно інформаційної моделі стану пацієнта (п. 2.4, С. 100-121). Разом з тим, незрозуміло в який спосіб в них враховується мультимодальність даних, які необхідно опрацювати.

6. Вказано, що стекінгова модель на основі алгоритмів машинного навчання забезпечує підвищення точності прогнозування даних та паралельної обробки даних як малої, так і великої розмірності (підрозділ 3.4, С. 165-167). Проте недостатньо обґрунтовано вибір стекінгової моделі для розв'язання такого типу задач на малих вибірках даних, хоча відомо, що вона має найменшу точність.

7. Із розробленого методу імпутації відсутніх даних (розд. 5, С. 196-201) незрозуміло, як оцінюється стійкість моделі та підвищення її стійкості до помилок в даних.

8. При описі методів доцільно було б дотримуватись такої схеми: постановка задачі, математичний опис, комп'ютерні експерименти і порівняння з методами-конкурентами.

9. Як в дисертаційній роботі, так і в рефераті зустрічаються термінологічні, синтаксичні і редакційні неточності та помилки.

9. Загальні висновки

Отже, дисертація Мельникової Наталії Іванівни на тему «Моделі та методи підтримки персоналізованих рішень у медичних системах» є завершеним науковим дослідженням, в якому розв'язано актуальну наукову-прикладну проблему розроблення та удосконалення моделей, методів і засобів машинного навчання в задачах класифікації, кластеризації, прогнозування та візуалізації результатів опрацювання персональних даних для адаптації медичних рішень до пацієнта.

Дисертаційна робота відповідає паспорту спеціальності 05.13.23 – системи та засоби штучного інтелекту (у частині формули спеціальності та низки напрямків досліджень) та визначеним вимогам підготовки докторських дисертацій.

Основні результати роботи повністю викладені в опублікованих працях, що пройшли належну апробацію у наукових науково-дослідних темах, конференціях та семінарах.

Дисертаційна робота виконана на високому науковому рівні. За актуальністю розв'язаних задач, обсягом досліджень, науковим рівнем і практичною цінністю отриманих результатів відповідає вимогам п. 7 та 9 Порядку присудження та позбавлення наукового ступеня доктора наук, затвердженого постановою Кабінету Міністрів України, а її автор, Мельникова Наталія Іванівна, заслуговує на присудження їй наукового ступеня доктора технічних наук за спеціальністю 05.13.23 – системи та засоби штучного інтелекту.

Офіційний опонент,
професор кафедри комп'ютерної інженерії,
Західноукраїнського національного університету,
д.т.н., професор

Олег БЕРЕЗЬКИЙ

