

ВІДГУК

офіційного опонента

кандидата фізико-математичних наук, доцента

Панченка Тараса Володимировича,

завідувача кафедри теорії та технології програмування,

на дисертаційну роботу **Празднікової Маргарити Олександрівни**

на тему **«Моделі і методи машинного навчання в задачах прогнозування**

ускладнень серцево-судинних хвороб»,

подану на здобуття ступеня доктора філософії

у галузі знань 12 «Інформаційні технології»

за спеціальністю 122 «Комп'ютерні науки»

Актуальність теми дисертаційної роботи.

Проблематика прогнозування ускладнень серцево-судинних захворювань належить до тих напрямів сучасної медичної інформатики, у яких науковий результат безпосередньо пов'язаний із практичними потребами охорони здоров'я. Інфаркт міокарда, гострі порушення мозкового кровообігу та інші ускладнення серцево-судинної патології залишаються одними з найвагоміших чинників втрати працездатності, передчасної смертності й істотного навантаження на лікувальні заклади.

У клінічній практиці важливо не лише зафіксувати вже наявний патологічний стан, а й своєчасно виявити пацієнтів, у яких ризик небажаних подій є підвищеним у середньостроковій або довгостроковій перспективі. Саме така постановка задачі відповідає логіці превентивної медицини, персоналізованого спостереження та раціональної маршрутизації пацієнтів.

Традиційні шкали серцево-судинного ризику, що ґрунтуються на віці, статі, артеріальному тиску, рівні холестерину, наявності цукрового діабету, курінні та інших класичних факторах, зберігають значну клінічну цінність. Водночас вони не завжди достатньо повно відображають індивідуальну історію пацієнта, динаміку його звернень, супутні стани, повторювані скарги, призначення та зміст лікарських висновків.

Особливу складність і водночас наукову цінність становлять неструктуровані медичні тексти. У таких записах відображено фрагменти анамнезу, клінічні спостереження, уточнення діагнозів, результати консультацій різних спеціалістів і практичні рішення лікарів. Автоматизоване залучення цього

масиву інформації є нетривіальною задачею, оскільки реальні амбулаторні записи містять скорочення, варіативні формулювання, друкарські помилки, неповноту та різний ступінь стандартизації.

Дисертаційна робота Празднікової М. О. є актуальною саме тому, що в ній розглянуто можливість поєднання текстових лікарських записів із кодами МКХ, параметрами електрокардіографічних досліджень і показниками артеріального тиску для побудови моделей машинного навчання. Такий підхід відповідає сучасним тенденціям розвитку систем підтримки клінічних рішень і має очевидний прикладний потенціал для медичних інформаційних систем.

Важливо також, що дослідження виконано не на умовному або навчальному наборі даних, а на матеріалі реальних медичних записів. Це підвищує практичну значущість роботи, оскільки дозволяє оцінювати моделі в умовах, наближених до реальної експлуатації, де якість даних, неповнота інформації та неоднорідність джерел відіграють вирішальну роль.

Оцінка обґрунтованості наукових результатів, їх достовірності та новизни.

У дисертації сформульовано наукове завдання, що полягає у розробленні моделей і методів машинного навчання для прогнозування ризику ускладнень серцево-судинних захворювань на основі медичних даних пацієнтів. Мета роботи й поставлені завдання узгоджені між собою: від аналізу підходів до обробки текстів і формування ознак до експериментального порівняння моделей, оцінювання інформативності даних і побудови інтерпретованої шкали ризику.

Методична база дослідження є достатньо переконливою. Авторкою використано інструменти машинного навчання, математичної статистики, обробки природної мови, аналізу текстових даних та експериментального моделювання. У роботі розглянуто застосування наївного байєсівського підходу до текстової класифікації, нормалізацію й лематизацію текстів, ототожнення словоформ на основі відстані Левенштейна, залучення кодів Міжнародної класифікації хвороб, а також використання моделей XGBoost для аналізу параметрів ЕКГ і артеріального тиску.

Позитивною рисою роботи є те, що авторка не обмежилася одним алгоритмом або одним джерелом даних. У дослідженні послідовно порівнюються окремі групи ознак та їх поєднання, що дозволяє оцінити внесок текстових даних, кодів діагнозів і структурованих клінічних параметрів у підсумкову якість прогнозування. Така логіка експерименту робить висновки роботи більш аргументованими.

Достовірність отриманих результатів підтверджується проведенням серії чисельних експериментів, використанням крос-валідації та застосуванням загальноприйнятих метрик якості класифікації, зокрема точності, чутливості, специфічності, F1-міри та площі під ROC-кривою. Наведені в дисертації результати для прогнозування інфаркту міокарда та цереброваскулярних ускладнень демонструють працездатність запропонованого підходу.

Наукова новизна дисертаційної роботи, на мою думку, полягає у таких основних положеннях:

1. Запропоновано підхід до використання сукупності текстів лікарських записів за весь період амбулаторного спостереження як джерела ознак для прогнозування ризику ускладнень неінфекційних захворювань.
2. Розроблено модель машинного навчання для середньо- та довгострокового прогнозування ризику інфаркту міокарда, у якій текстові ознаки поєднано з кодами МКХ і параметрами електрокардіографічних досліджень.
3. Побудовано модель прогнозування ризику уражень судин головного мозку, що інтегрує інформацію з медичних текстів, діагностичних кодів і показників артеріального тиску.
4. Запропоновано спосіб переходу від результату класифікації до кількісної індивідуальної оцінки ризику з подальшим формуванням чотирирівневої скорингової шкали.
5. Показано можливість практичної інтеграції розроблених моделей у захищений контур медичної інформаційної системи з режимами індивідуального та групового аналізу.

Окремо слід відзначити виважений вибір моделей для аналізу текстових даних. Для медичних застосувань важливо не лише досягти високих числових метрик, а й забезпечити прийнятну інтерпретованість результатів. Тому використання порівняно простіших моделей, які дають змогу аналізувати внесок окремих лексичних маркерів, є виправданим і практично корисним.

У цілому наукові результати дисертації є обґрунтованими, логічно пов'язаними з поставленою метою та підтвердженими експериментальними даними. Робота свідчить про здатність здобувачки самостійно формулювати наукові задачі, добирати адекватні методи їх розв'язання та доводити отримані результати до програмної реалізації.

Практичне значення отриманих результатів.

Практична цінність дисертаційної роботи полягає у можливості використання запропонованих моделей як допоміжного інструменту підтримки

прийняття лікарських рішень. Йдеться не про заміну клінічного судження лікаря, а про додатковий аналітичний механізм, який може допомогти виявляти пацієнтів із підвищеним ризиком інфаркту міокарда або гострих порушень мозкового кровообігу.

Розроблені моделі можуть бути корисними для формування груп підвищеного ризику, планування профілактичних оглядів, уточнення маршрутизації пацієнтів, організації диспансерного спостереження та підтримки управлінських рішень у медичних закладах. Особливо важливо, що в роботі передбачено режими як індивідуального, так і групового аналізу.

Суттєвим практичним результатом є програмна реалізація відповідних моделей і модуля аналізу. Впровадження результатів у Державній науковій установі «Центр інноваційних технологій охорони здоров'я» Державного управління справами та проведення тестової експлуатації на реальних медичних даних підтверджують прикладний характер дисертації.

Позитивної оцінки заслуговує також оприлюднення початкового програмного коду розробленого програмного забезпечення. За умов обмеженого доступу до первинних медичних даних відкритість програмної частини підвищує прозорість роботи, полегшує перевірку окремих етапів дослідження та створює передумови для подальшого розвитку запропонованих підходів.

Перспективними напрямками практичного використання результатів є розширення підходу на інші класи неінфекційних захворювань, застосування моделей у профілактичній медицині, підтримка медичного аудиту, аналіз популяційних ризиків і розвиток інтелектуальних компонентів медичних інформаційних систем.

Оцінка змісту дисертації, її завершеності та дотримання принципів академічної доброчесності.

За змістом дисертаційна робота Празднікової М. О. відповідає галузі знань 12 «Інформаційні технології» та спеціальності 122 «Комп'ютерні науки». Тематика дослідження безпосередньо пов'язана з машинним навчанням, інтелектуальним аналізом даних, обробкою природної мови, медичною інформатикою та побудовою прикладних програмних систем.

Дисертація є завершеною науковою працею. Її структура є логічною: теоретичний аналіз і постановка задачі послідовно переходять до побудови моделей, експериментального дослідження, аналізу результатів і практичного впровадження. Виклад матеріалу в цілому послідовний, аргументований і достатній для розуміння як методичної, так і прикладної складової роботи.

Робота складається з анотацій українською та англійською мовами, переліку умовних позначень, вступу, трьох розділів основної частини, загальних висновків, списку використаних джерел і додатку. Загальний обсяг дисертаційного дослідження становить 132 сторінки друкованого тексту, обсяг основного тексту — 102 сторінки. Робота містить 26 рисунків, 15 таблиць, 42 формули та список використаних джерел із 100 найменувань.

У вступі належним чином обґрунтовано актуальність теми, сформульовано мету й завдання дослідження, визначено об'єкт, предмет і методи, наведено наукову новизну та практичне значення, а також подано відомості про апробацію, публікації й особистий внесок здобувачки.

Перший розділ має оглядово-методичний характер. У ньому розглянуто методи машинного навчання для класифікації та прогнозування, зокрема наївний байєсівський класифікатор, логістичну регресію, дерева рішень, випадковий ліс і XGBoost. Також описано базові підходи до попередньої обробки текстів, нормалізації, лематизації, використання відстані Левенштейна та інструментів обробки природної мови.

Другий розділ становить основну науково-експериментальну частину роботи. У ньому описано підготовку набору даних, формування текстових ознак, побудову моделей прогнозування інфаркту міокарда та уражень судин головного мозку, використання великих мовних моделей для відбору медичних термінів, а також інтеграцію текстових і структурованих клінічних даних.

У третьому розділі розкрито прикладний аспект дисертації: запропоновано підхід до формування індивідуальної оцінки ризику, описано концепцію кінцевої науково-технічної продукції, архітектуру програмного модуля, захищений контур обробки даних, сценарії використання та результати тестової експлуатації.

Оформлення дисертації загалом відповідає вимогам до наукових кваліфікаційних робіт. Аналіз змісту дисертації, переліку публікацій і зазначеного особистого внеску не дає підстав для сумнівів щодо самостійності основних результатів. Використані джерела наведено належним чином, а запозичені ідеї та результати інших авторів супроводжуються відповідними посиланнями.

Мова та стиль викладення результатів.

Дисертаційну роботу написано українською мовою з дотриманням загальних вимог наукового стилю. Авторка коректно використовує термінологію комп'ютерних наук, машинного навчання, аналізу даних, обробки природної

мови та медичної інформатики. Математичні формулювання, таблиці й рисунки переважно застосовано доречно; вони підтримують основний виклад і сприяють розкриттю змісту дослідження.

Структура тексту є зрозумілою, а послідовність подання матеріалу відповідає логіці проведеного дослідження. Разом із тим у роботі трапляються окремі редакційні неточності: нерівномірне використання українських і англійських скорочень, поодинокі друкарські помилки та різний ступінь деталізації під час опису експериментів для окремих джерел даних. Ці зауваження не мають принципового характеру й не впливають на зміст отриманих наукових результатів.

Оприлюднення результатів дисертаційної роботи.

Основні результати дисертаційної роботи оприлюднено у трьох статтях у наукових виданнях України категорії «Б», зокрема у виданнях «Кібернетика та комп'ютерні технології» та «Проблеми керування та інформатики». Тематика цих публікацій безпосередньо відповідає змісту дисертації та охоплює прогнозування ризику інфаркту міокарда, моделі оцінювання ризику серцево-судинних захворювань і прогнозування цереброваскулярних хвороб.

Результати дослідження також апробовано на шести міжнародних наукових і науково-практичних конференціях. Частина конференційних матеріалів опублікована у виданнях, що індексуються в наукометричній базі Scopus. Це свідчить про належне представлення результатів у професійному науковому середовищі та про їх відповідність тематиці дисертаційного дослідження.

Отже, результати дисертації достатньо повно висвітлено у публікаціях здобувачки, а наведений особистий внесок у працях зі співавторами узгоджується з основними положеннями, що виносяться на захист.

Недоліки та зауваження до дисертаційної роботи.

Оцінюючи дисертаційну роботу загалом позитивно, вважаю за необхідне висловити такі зауваження та побажання:

1. Викликає подив формулювання «точність біля 80%» (стор. 21), адже мова йде про параметр точності моделі.
2. В роботі не згадані праці Thierry Namon та його дослідницької групи Університету Париж XIII та члена дослідницької лабораторії LIMSI-CNRS, який досяг значних результатів в аналізі медичних текстів методами штучного інтелекту.

3. Відсутнє порівняння з роботами та результатами Владислава Маланіна, який досліджував дуже близьку тематику. Аналіз його робіт міг би посилити результати дисертантки.
4. Не достатньо прозоро описано вилучення «підозрілих» пацієнтів з вибірки (підрозділ 2.2). Можлива ситуація вилучення важливих ознак або маркерів під виглядом «підозрілих».
5. Не розкриті питання захисту чутливих даних пацієнтів: дотримання HIPAA, GDPR та інших застосовних стандартів, вимог та обмежень. Зокрема, при роботі з GPT (стор.64-68), відомо, що моделі від OpenAI використовують передані дані для подальшого навчання (що неприпустимо для чутливої медичної інформації пацієнтів без їхньої письмової згоди).
6. Не достатньо розкрито питання захисту даних при передачі (підрозділ 3.3): які протоколи та стандарти застосовуються, як гарантується захист від можливих кібератак (Man-in-the-middle, підміна DNS, ключів, тощо).
7. Отримані результати впровадження (підрозділ 3.6) виглядають як недостатньо надійні, отже потребують додаткових обґрунтувань чи пояснень, чому такий рівень точності вважається прийнятним.
8. Має місце ряд граматичних та технічних помилок, як-от: моїної (замість мовної), термітів (замість термінів) на стор. 9, 25, та ін.

Наведені зауваження мають переважно дискусійний і рекомендаційний характер. Вони не знижують наукової новизни, практичного значення та загальної позитивної оцінки дисертаційної роботи Празднікової М. О.

Висновок про дисертаційну роботу.

Вважаю, що дисертаційна робота Празднікової Маргарити Олександрівни на тему «Моделі і методи машинного навчання в задачах прогнозування ускладнень серцево-судинних хвороб» є завершеним самостійним науковим дослідженням, виконаним на належному науковому рівні. Сукупність отриманих у роботі результатів розв'язує актуальне науково-прикладне завдання у галузі інформаційних технологій, пов'язане з розробленням і застосуванням моделей машинного навчання для аналізу медичних даних та підтримки клінічних рішень.

Дисертаційна робота за актуальністю, науковою новизною, обґрунтованістю результатів і практичною цінністю відповідає вимогам чинного законодавства України, зокрема пунктам 6–9 «Порядку присудження ступеня доктора філософії та скасування рішення разової спеціалізованої вченої ради закладу вищої освіти, наукової установи про присудження ступеня доктора

філософії», затвердженого постановою Кабінету Міністрів України від 12 січня 2022 р. № 44.

Здобувачка Празднікова Маргарита Олександрівна заслуговує на присудження ступеня доктора філософії у галузі знань 12 «Інформаційні технології» за спеціальністю 122 «Комп'ютерні науки».

Офіційний опонент

кандидат фізико-математичних наук,
доцент, завідувач кафедри теорії
та технології програмування
факультету комп'ютерних наук
та кібернетики
Київського національного університету
імені Тараса Шевченка

Тарас ПАНЧЕНКО

«_____» _____ 2026 р.