

ВІДГУК

офіційного опонента — кандидата технічних наук, професора
Попова Антона Олександровича
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 дисертації.

Серцево-судинні та цереброваскулярні захворювання залишаються однією з провідних причин передчасної смертності, інвалідизації населення та зростання навантаження на систему охорони здоров'я. Особливого значення в сучасній клінічній практиці набуває не лише виявлення вже сформованої патології, а й середньо- та довгострокове прогнозування ризику гострих ускладнень, насамперед інфаркту міокарда та гострих порушень мозкового кровообігу. Така постановка задачі безпосередньо пов'язана з профілактичною медициною, персоналізованим спостереженням пацієнтів і раціональним використанням ресурсів медичних закладів.

Традиційні шкали ризику та класичні фактори — вік, стать, артеріальний тиск, наявність цукрового діабету, рівень холестерину, куріння, ожиріння тощо — безумовно залишаються клінічно значущими, однак не завжди забезпечують достатній рівень індивідуалізованої точності. Водночас сучасні медичні інформаційні системи накопичують значні масиви даних, які не обмежуються структурованими числовими показниками. Особливо цінним, але складним для автоматизованого аналізу джерелом інформації є тексти лікарських записів, висновків та історій звернень пацієнтів, у яких у вільній формі зафіксовано анамнестичні відомості, супутні захворювання, скарги, призначення та клінічні інтерпретації лікарів різних спеціальностей.

У цьому контексті дисертаційна робота Празднікової М. О. є актуальною, оскільки спрямована на розроблення моделей машинного навчання, здатних

залучати неструктуровані медичні тексти в поєднанні з кодами МКХ, параметрами електрокардіографічних досліджень і показниками артеріального тиску. Такий підхід відповідає сучасному напрямку розвитку інтелектуальних систем підтримки клінічних рішень та має практичне значення для раннього виявлення груп підвищеного ризику, планування профілактичних заходів і оптимізації маршрутизації пацієнтів.

Актуальність роботи додатково посилюється тим, що дослідження виконано не на абстрактному демонстраційному наборі даних, а на основі реальних амбулаторних медичних записів із типовими для клінічної практики особливостями: неоднорідністю формулювань, скороченнями, орфографічними помилками, неповнотою, шумом і різним рівнем стандартизації. Саме такі властивості даних значною мірою визначають складність переходу від експериментальних моделей машинного навчання до їх застосування в реальних інформаційних системах медичних установ.

Оцінка обґрунтованості наукових результатів дисертації, їх достовірності та новизни.

У дисертаційній роботі виконано комплексне дослідження моделей і методів машинного навчання для прогнозування ризику ускладнень серцево-судинних хвороб на основі поєднання текстових лікарських записів із додатковими клінічними даними. Робота має чітко визначену мету — створення методів і моделей машинного навчання для середньострокового та довгострокового прогнозування ризику ускладнень серцево-судинних захворювань на основі медичних даних пацієнтів. Поставлені завдання логічно відповідають цій меті: від аналізу методів обробки текстів і формування ознак до побудови моделей, аналізу інформативності ознак, розроблення шкали ризику та оцінювання якості прогнозування.

Обґрунтованість отриманих результатів забезпечується використанням сучасного апарату машинного навчання, математичної статистики, аналізу текстових даних та експериментальної перевірки моделей. У роботі розглянуто найвні байєсівські класифікатори для текстових ознак, методи нормалізації та

лематизації, ототожнення словоформ за відстанню Левенштейна, використання кодів МКХ, а також моделі на основі XGBoost для аналізу показників ЕКГ і артеріального тиску. Застосування різних комбінацій попередньої обробки та різних джерел даних дало змогу авторці не обмежуватися одним експериментальним сценарієм, а порівняти кілька варіантів побудови моделей.

Достовірність результатів підтверджується значною серією чисельних експериментів, використанням крос-валідації, а також оцінюванням моделей за загальноприйнятими метриками, зокрема точністю, чутливістю, специфічністю, F1-мірою та площею під ROC-кривою. Наведені в роботі результати для задачі прогнозування інфаркту міокарда на рівні близько 93,6 % точності та для задачі оцінювання ризику цереброваскулярних хвороб на рівні 91,6 % свідчать про працездатність обраного підходу й перспективність використання неструктурованих клінічних текстів у таких задачах.

Наукова новизна результатів дисертаційного дослідження, на мій погляд, полягає насамперед у таких положеннях:

1. Запропоновано підхід, у межах якого сукупність текстів лікарських записів за весь період амбулаторного спостереження використовується як основне джерело ознак для оцінювання ризику ускладнень неінфекційних захворювань, що дає змогу враховувати анамнез, супутні діагнози, призначення та клінічні спостереження незалежно від спеціалізації окремого лікаря;
2. Побудовано модель машинного навчання для середньо- та довгострокового прогнозування ризику інфаркту міокарда, яка поєднує вектори слів із реальних лікарських записів, коди Міжнародної класифікації хвороб та параметри електрокардіографічних досліджень;
3. Побудовано модель машинного навчання для прогнозування ризику уражень судин головного мозку, у якій текстові ознаки й коди діагнозів поєднано з показниками артеріального тиску;

4. Запропоновано спосіб переходу від класифікаційного результату до кількісної індивідуальної оцінки ризику з подальшим формуванням інтерпретованої чотирирівневої скорингової шкали;
5. Показано можливість практичної інтеграції побудованих моделей у захищений контур медичної інформаційної системи з режимами індивідуального та групового аналізу.

Окремо слід відзначити, що в роботі обґрунтовано використання відносно простих і достатньо інтерпретованих моделей для текстової класифікації. Для медичних застосувань це є істотною перевагою, оскільки надмірно складні нейромережеві моделі можуть забезпечувати формально високі метрики, але залишатися непрозорими для лікаря і складними для впровадження в регульованих середовищах. Обраний підхід дає можливість аналізувати внесок окремих лексичних маркерів і формувати пояснення, ближчі до клінічної логіки.

Отже, у дисертаційній роботі поставлене наукове завдання виконано повністю, а здобувачка продемонструвала володіння методологією наукового дослідження, експериментального моделювання, аналізу медичних даних і програмної реалізації прикладних систем машинного навчання.

Практичне значення отриманих результатів.

Практичне значення результатів дисертаційної роботи полягає в тому, що розроблені моделі можуть використовуватися як додатковий інструмент підтримки прийняття лікарських рішень під час виявлення пацієнтів із підвищеним ризиком інфаркту міокарда та гострих порушень мозкового кровообігу. Йдеться не про заміну клінічного рішення, а про автоматизовану аналітичну підказку, що може привернути увагу лікаря до пацієнтів, які потребують додаткового спостереження, уточнення діагностики або корекції профілактичної стратегії.

Суттєвим практичним результатом є програмна реалізація моделей і модуля для індивідуального та групового аналізу. Здобуті результати впроваджено в Державній науковій установі «Центр інноваційних технологій охорони здоров'я» Державного управління справами, де проведено тестову експлуатацію на реальних медичних даних у захищеному середовищі. Наявність акта впровадження

підтверджує, що робота не обмежилася теоретичною постановкою задачі та демонстраційними експериментами.

Окремо слід відзначити публікацію початкового програмного коду розробленого програмного забезпечення у відкритому доступі. Це відповідає сучасним принципам відкритої науки, полегшує перевірку й подальший розвиток результатів, а також створює передумови для незалежного відтворення програмної частини дослідження. За умов обмеженого доступу до первинних медичних даних відкритість програмних засобів є особливо важливою складовою наукової доброчесності та практичної корисності роботи.

Потенційними напрямками використання отриманих результатів є клінічна профілактична медицина, планування ресурсів медичних закладів, формування груп диспансерного спостереження, аналіз популяційних ризиків, а в перспективі — адаптація підходу до інших неінфекційних захворювань і задач медичного страхування.

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

За своїм змістом дисертаційна робота Празднікової М. О. відповідає Стандарту вищої освіти зі спеціальності 122 «Комп'ютерні науки» в галузі знань 12 «Інформаційні технології». Тематика роботи безпосередньо належить до методів машинного навчання, інтелектуального аналізу даних, обробки природної мови, побудови програмних моделей і прикладних інформаційних систем.

Дисертаційна робота є завершеною науковою працею. Вона має логічну структуру, у якій теоретичний огляд і постановка задачі послідовно переходять до побудови моделей, експериментального дослідження та практичного впровадження. Робота виконана українською мовою з дотриманням вимог наукового стилю. Виклад загалом є послідовним, зрозумілим і достатньо аргументованим. Термінологія відповідає сучасним напрямкам комп'ютерних наук, машинного навчання, аналізу даних і медичної інформатики.

Дисертація 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, список використаних джерел із 100 найменувань.

У вступі обґрунтовано актуальність теми, сформульовано мету й завдання дослідження, визначено об'єкт, предмет і методи дослідження, наведено наукову новизну, практичне значення, зв'язок роботи з науково-дослідними темами, відомості про апробацію, публікації та особистий внесок здобувачки.

У першому розділі наведено теоретичні та методичні основи дослідження. Розглянуто методи машинного навчання для класифікації та прогнозування, включаючи наївний байєсівський підхід, логістичну регресію, дерева рішень, випадковий ліс, XGBoost, а також базові інструменти обробки текстів: попередню обробку, нормалізацію, лематизацію, відстань Левенштейна та засоби обробки природної мови. Проведений огляд дає змогу визначити місце роботи серед сучасних досліджень із прогнозування серцево-судинних станів за різними типами даних.

У другому розділі описано розроблені моделі прогнозування ускладнень серцево-судинних захворювань. Розділ містить опис набору даних, підготовку текстових записів, побудову моделей прогнозування інфаркту міокарда та ураження судин головного мозку, використання великих мовних моделей для відбору медичних термінів, моделі за даними ЕКГ і артеріального тиску, а також інтегровані мультимодальні моделі. Саме цей розділ є центральним з погляду наукових результатів і демонструє послідовне просування від окремих джерел даних до їх об'єднання.

У третьому розділі розглянуто практичне застосування результатів: перехід від класифікації до індивідуальної оцінки ризику, концепцію кінцевої науково-технічної продукції, інтеграцію в медичну інформаційну систему, архітектуру та захищений контур обробки даних, режими експлуатації, сценарії використання, впровадження в ДНУ «ЦІТОЗ» ДУС та результати тестової

експлуатації. Наявність цього розділу є сильною стороною дисертації, оскільки підтверджує прикладну спрямованість роботи.

Дисертаційна робота оформлена відповідно до вимог наказу Міністерства освіти і науки України від 12 січня 2017 р. № 40 «Про затвердження вимог до оформлення дисертації».

Аналіз наданих матеріалів дисертаційної роботи, переліку публікацій і зазначеного особистого внеску не дає підстав для сумнівів у самостійності отриманих результатів. Використані ідеї, методи та результати інших авторів супроводжуються посиланнями на відповідні джерела. Внесок здобувачки у працях, виконаних у співавторстві, окреслено достатньо конкретно: розроблення основних положень, програмного забезпечення, попереднє опрацювання ЕКГ-даних і текстових записів, формування навчальних вибірок та інтеграція моделей в інформаційне середовище медичної установи.

Мова та стиль викладення результатів.

Дисертаційна робота написана українською мовою. Стиль викладення матеріалу є загалом чітким і зрозумілим, структура роботи логічна. Авторка коректно використовує сучасну термінологію машинного навчання, аналізу даних, обробки природної мови та медичної інформатики. Математичні формулювання, таблиці й рисунки в цілому застосовано доцільно; вони підтримують основний виклад і сприяють розкриттю змісту дослідження.

Водночас у тексті трапляються окремі мовні й редакційні неточності, зокрема нерівномірне використання українських і англійських скорочень, поодинокі друкарські помилки та дещо різний рівень деталізації під час опису експериментів для різних джерел даних. Зазначені недоліки мають редакційний характер і не впливають на сутність отриманих наукових результатів.

Оприлюднення результатів дисертаційної роботи.

Основні результати дисертаційної роботи висвітлено у трьох статтях у наукових виданнях України категорії «Б»: у журналі «Кібернетика та комп'ютерні технології» за 2024 і 2025 роки та у журналі «Проблеми керування та

інформатики» за 2026 рік. Тематика цих статей безпосередньо відповідає основним науковим результатам дисертації: прогнозуванню ризику інфаркту міокарда за сукупністю текстів лікарських висновків, моделям прогнозування ризику серцево-судинних захворювань і моделі прогнозування цереброваскулярних хвороб.

Результати дисертації апробовано на шести міжнародних наукових і науково-практичних конференціях, зокрема на конференціях, присвячених інноваційним технологіям в охороні здоров'я, пацієнт-орієнтованому підходу, медицині в умовах воєнного часу, а також на конференціях із ширшої тематики машинного навчання та комп'ютерної обробки даних. Частину матеріалів опубліковано у виданнях, що індексуються в наукометричній базі Scopus, у формі конференційних публікацій.

Таким чином, наукові результати, описані в дисертаційній роботі, достатньо повно висвітлені у публікаціях здобувачки та пройшли апробацію в професійній науковій і медичній спільноті. Публікації відповідають темі дисертації, а заявлений особистий внесок у працях зі співавторами узгоджується з основними результатами, що виносяться на захист.

Недоліки та зауваження до дисертаційної роботи.

Разом із позитивною оцінкою дисертаційної роботи доцільно зазначити окремі недоліки та дискусійні положення:

1. У роботі виконано значний обсяг досліджень у напрямі інженерії ознак для вилучених числових параметрів ЕКГ, зокрема запропоновано низку похідних ритм-залежних та відносних часових інтервалів. Водночас, вибір саме таких математичних комбінацій та відношень (наприклад, PRratio чи TQRSratio) подано переважно як евристичний підхід. Робота виграла б, якби авторка навела ширше теоретичне або клінічне обґрунтування доцільності впровадження саме цих похідних показників з погляду електрофізіології міокарда.
2. Зіткнувшись із суттєвим і природним для медичної практики дисбалансом класів у вибірці ЕКГ-даних, авторка застосувала процедуру випадкового

зменшення вибірки мажоритарного класу. Хоча це дозволило уникнути упередженості моделей і штучної генерації неіснуючих медичних значень (як у SMOTE), такий крок призвів до втрати значної частини доступної інформації. Доцільно було б обговорити в межах дискусії можливість використання алгоритмічних ансамблевих методів балансування (на кшталт Balanced Random Forest або специфічних штрафних функцій втрат) без радикального зменшення обсягу аналізованих записів.

3. Під час побудови моделей середньо- та довгострокового прогнозування (особливо мультимодальних варіантів у Розділі 2.7) у вибірку включалися всі лікарські записи пацієнта, зроблені до першої фіксації гострої події. Проте в роботі чітко не формалізовано так зване "вікно прогнозу" (наприклад, ризик виникнення події протягом наступних 6 місяців, 1 року чи 5 років). Зважаючи на динаміку змін таких показників, як артеріальний тиск, часова стандартизація прогностичного горизонту додала б моделям вищої практичної цінності для практикуючих лікарів.
4. У розділі 2.3.3 здобувачка описує оригінальний та перспективний підхід до фільтрації шуму в текстових ознаках за допомогою великої мовної моделі GPT-3.5 Turbo. Проте, як показують результати експериментів, модель, що базувалася виключно на відібраних ШІ "сухих" медичних термінах, продемонструвала дещо нижчу точність, ніж моделі на основі повного частотного словника з лематизацією. Авторка справедливо зауважує, що це пов'язано з втратою інформативних немедичних лексем (епітетів, показників ступеня тощо). Було б доцільно в роботі детальніше окреслити шляхи компромісного поєднання автоматичного термінологічного аналізу з контекстними маркерами природної мови для збереження прогностичної сили.

Висловлені зауваження мають переважно рекомендаційний і дискусійний характер. Вони не є визначальними, не заперечують наукову новизну та практичну значущість отриманих результатів і не впливають на загальну позитивну оцінку дисертаційної роботи Празднікової М. О.

Висновок про дисертаційну роботу.

Вважаю, що дисертаційна робота здобувачки ступеня доктора філософії Празднікової М. О. на тему «Моделі і методи машинного навчання в задачах прогнозування ускладнень серцево-судинних хвороб» виконана на високому науковому рівні, не порушує принципів академічної доброчесності та є завершеним науковим дослідженням, сукупність теоретичних і практичних результатів якого розв’язує наукове завдання, що має істотне значення для галузі інформаційних технологій, зокрема для комп’ютерних наук, машинного навчання, інтелектуального аналізу медичних даних і систем підтримки прийняття клінічних рішень.

Дисертаційна робота за актуальністю, практичною цінністю та науковою новизною повністю відповідає вимогам чинного законодавства України, що передбачені пунктами 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 М. О. заслуговує на присудження ступеня доктора філософії у галузі знань 12 «Інформаційні технології» за спеціальністю 122 «Комп’ютерні науки».

Офіційний опонент

кандидат технічних наук,
професор кафедри електронної інженерії
факультету електроніки
Національного технічного університету
України
«Київський політехнічний інститут
імені Ігоря Сікорського»

_____ Антон ПОПОВ

«_____» _____ 2026 р.