

РЕЦЕНЗІЯ

офіційного рецензента

кандидата технічних наук, старшого наукового співробітника

Галелюки Ігоря Богдановича

на дисертаційну роботу **Празднікової Маргарити Олександрівни**

на тему **"Моделі і методи машинного навчання в задачах прогнозування ускладнень серцево-судинних хвороб",**

подану на здобуття наукового ступеня доктора філософії
за спеціальністю 122 – "Комп'ютерні науки"

Актуальність теми дисертації

Швидкий темп та спосіб життя сучасної людини часто супроводжуються негативними чинниками та звичками, які не тільки погіршують загальний стан здоров'я, але і зумовлюють виникнення різних хвороб, в тому числі захворювань серцево-судинної системи, які, як показує статистика, є однією з головних причин смертності як у світі, так і в Україні. Якщо взяти до уваги сучасні виклики та стреси, з якими зустрілися українці, то ризики виникнення та ускладнення таких захворювань критично зростають.

В таких умовах важливо не тільки оцінювати стан здоров'я громадян та діагностувати хвороби, але і прогнозувати ускладнення наявних захворювань. Використання нових підходів, зокрема методів штучного інтелекту та машинного навчання, дає можливість отримувати персоніфіковані прогнози для конкретного пацієнта не тільки на основі традиційних показників та даних аналізів, але і враховуючи записи лікаря протягом всього періоду спостереження. Саме такий підхід сприяє формуванню більш точних прогнозів, що у свою чергу зумовить розроблення своєчасних заходів щодо усунення причин ускладнень.

Саме тому, обрана автором тема, яка направлена на розробку нових підходів щодо прогнозування ускладнень серцево-судинних захворювань із застосуванням методів машинного навчання, є досить актуальною на даному етапі розвитку суспільства, науки і техніки.

Зв'язок дослідження з науковими програмами та темами

Дисертаційні дослідження проведено у відповідності та в рамках науково-дослідних робіт Інституту кібернетики імені В.М. Глушкова НАНУ. Дослідження відповідають пріоритетам, що визначені КМУ у постанові № 476, зокрема напрямку «Інформаційні технології у медицині». Здобувач брала активну участь як виконавець в таких науково-дослідних темах: «Розробити і дослідити методи розподілених обчислень для задач комп'ютерного моделювання і машинного навчання» (№ держреєстрації 0122U001515, 2022–2024 рр.), «Розробити моделі

та методи машинного навчання у прикладних задачах інформатизації» (№ держреєстрації 0125U001129, 2025–2027 рр.), «Розробити і дослідити моделі та методи проєктування застосувань з інтенсивною обробкою даних» (№ держреєстрації 0123U100822, 2023–2027 рр.)

Рівень обґрунтованості наукових результатів

Наукові результати, які отримано та представлено здобувачем в дисертації, мають достатнє теоретичне та експериментальне підґрунтя. Підходи та методи щодо аналізу існуючих методів представлення та обробки медичних даних й створення власних моделей прогнозування ризику серцево-судинних захворювань використано здобувачем у дослідженні у коректний та зрозумілий спосіб.

Огляд змісту дисертації та відповідності вимогам до оформлення

Дисертаційна робота відповідає вимогам до оформлення дисертацій доктора філософії та містить необхідні частини: анотацію, зміст, перелік умовних скорочень, вступ, 3 розділи, висновки, перелік джерел посилання та додаток.

В **першому** розділі автор проаналізувала методи машинного навчання, які зручні для застосування у задачах медичної класифікації та оцінювання ризиків. Крім того, детальну увагу приділено методам обробки та нормалізації текстових даних, узагальнення словоформ в розрізі застосування щодо медичних записів. На останок автор обґрунтувала вибір методів, необхідних для подальших досліджень та розробки власного підходу щодо використання медичних записів як інформативного джерела даних про пацієнта.

Другий розділ присвячено створенню та експериментальному дослідженню моделей прогнозування та оцінки ризику серцево-судинних захворювань на основі текстів лікарських висновків. Автор присвятила увагу застосуванню мультимодальних вхідних даних для розроблених моделей, зокрема об'єднанню числових даних ЕКГ та артеріального тиску з текстовими записами. Як результат, оцінено показники виявлення хворих при застосуванні сукупностей різних даних.

В **третьому** розділі здійснено практичну перевірку розроблених моделей та формалізовано сценарії та механізми застосування моделей для підтримки прийняття рішень у клінічній практиці. Описано потенційні приклади інтеграції розробленого модуля прогнозування у медичні системи для підсилення та доповнення клінічного висновку. Наведено результати пілотного впровадження розробленого модуля у Державній науковій установі "Центр інноваційних технологій охорони здоров'я" Державного управління справами, включно порівняння з експертною оцінкою лікарів. Усе це зробило можливим дисертанту сформулювати перспективи розвитку, масштабування та застосування запропонованих методів та технологій.

Наукова новизна результатів дисертації

Результати, отримані здобувачем у дисертаційній роботі, мають достатній рівень наукової новизни.

Здобувачем запропоновано підхід на основі мультимодальності вхідних даних, зокрема сукупності текстових даних з лікарських записів та числових показників ЕКГ та артеріального тиску, для забезпечення більш точного середньо- та довгострокового прогнозування ризиків ускладнення серцево-судинних захворювань.

Здобувачем вперше запропоновано та побудовано модель на базі методів машинного навчання для прогнозування ризику інфаркту міокарда, у якій вхідні дані є мультимодальними та являють собою сукупність показників електрокардіограми та текстових лікарських записів, що дало змогу досягнути точності прогнозу у 93,6%.

Здобувачем вперше запропоновано та побудовано модель на базі методів машинного навчання для прогнозування ризику ураження судин головного мозку, у якій вхідні дані є мультимодальними та являють собою сукупність показників артеріального тиску та текстових лікарських записів, що дало змогу досягнути точності прогнозу у 91,6%.

Для інтерпретації отриманих результатів прогнозування здобувачем запропоновано метод перерахунку логарифмів ймовірнісних оцінок ризику у вербальну шкалу, що робить можливим зіставлення з клінічно зрозумілими ознаками.

Практична значимість роботи

Практичне значення отриманих результатів полягає у доведенні запропонованих моделей до практично реалізованого програмного засобу, який впроваджено у Державній науковій установі "Центр інноваційних технологій охорони здоров'я" Державного управління справами та випробувано у захищеному середовищі на реальних медичних даних. Отримана довідка про впровадження також свідчить про практичну цінність роботи.

Повнота викладення результатів дослідження в публікаціях

Результати роботи викладено у 3 наукових публікаціях у фахових журналах категорії "Б". Крім того, отримані результати було апробовано автором на 4 конференціях.

Опубліковані праці охоплюють основні результати дисертаційної роботи, відображають зміст її основних розділів та свідчать про належний рівень апробації проведених досліджень.

Зауваження та рекомендації

1) У першому розділі бажано б проаналізувати моделі, системи та шкали оцінювання ризиків серцево-судинних захворювань, які на даний момент широко

використовують або які перебувають на етапах розробки, а також навести їх переваги і недоліки.

2) Доцільно було б порівняти розроблені в дисертації моделі з існуючими на ринку системами та шкалами оцінювання ризиків серцево-судинних захворювань, серед яких є SCORE2, Framingham, QRISK3, ASCVD, GRACE та ряд подібних.

3) В анотації, вступі і т.д. вказано, що в роботі представлено дослідження, присвячене середньо- та довгостроковому прогнозуванню, хоча в дисертації не визначено чітко ці строки та не вказано результати, які чітко прив'язано до цих строків.

4) У загальних висновках слід було порівняти отриману точність прогнозів з точність прогнозів аналогічних моделей чи систем.

5) Не зовсім однозначно та зрозуміло описано вибір та підготовку вхідних даних для роботи алгоритму MultinomialNB, оскільки в п.п. 2.3.2 (стор. 58–60) в першому абзаці вказано, що алгоритм MultinomialNB використовує "саме кількість появи слів, а не частотність", в другому абзаці вказано, що "MultinomialNB працює з частотами появи слів", в третьому абзаці – "оцінюється на основі частот появи слів", а у формулі 2.17 вже згадується про кількість появи слів.

6) У табл. 2.13 сума прогнозів для кожного з фактів не дає 100 %.

7) У висновку до розділу 2 (стор. 93) твердження щодо бази даних – "використано фрагмент ... бази Ескулап..., що містить мільйони лікарських записів та сотні тисяч пацієнтів", – є недоречним. У висновку правильно було б вказати саме розмір вибірки з цієї бази, який використано у дисертаційному дослідженні.

8) У формулі 3.4 (стор. 98) не описано коефіцієнти А і В, їх доцільність, розрахунок чи джерело їх значень та вплив на кінцеву оцінку.

9) На рис. 3.2 (стор. 101) автору слід було чітко вказати ті підсистеми, в основі роботи яких лежать результати дисертаційного дослідження. Крім того, терміни і назви з рисунку мають бути однаковими з тими, що в тексті опису рисунку. Наприклад, на рис. написано "Цифрувальник ЕКГ", а в описі "Підсистема автоматичного цифрування кардіограм". Лише залишається здогадуватися, чи це те ж саме, чи одне з них складова частина іншого чи навпаки.

10) Табл. 32 (стор. 106) містить помилку. Частка співпадіння для помірної розбіжності має становити 20,65%, а не 19,3%.

11) Якщо всюди по тексту дисертації використано термін "Матриця помилок", то й у назві табл. 3.2 також доцільно використати цей термін, а не "Матриця плутанини".

12) Скорочення, які зустрічаються тільки один раз по тексту дисертації, не потрібно включати у Перелік умовних позначень, наприклад, LSTM, CNN та інші.

13) На програмний засіб доцільно було б отримати Свідоцтво про реєстрацію авторського права.

Наведені зауваження та рекомендації ніяким чином не впливають на загальну позитивну оцінку дисертаційної роботи.

Рекомендації щодо впровадження результатів дослідження

Отримані результати може бути застосовано у медичних закладах різних форм власності як додаткове джерело оцінювання індивідуальних ризиків для пацієнтів з метою планування обстеження, лікування чи профілактики. Крім того, розроблені методи та моделі може бути застосовано страховими компаніями для кількісної оцінки ризиків для окремої людини, оптимізації страхових тарифів та контролю витрат з метою обґрунтування рішень при медичному страхуванні.

Розроблені моделі та програмний засіб доцільно інтегрувати у існуючі медичні системи, які присутні на вітчизняному ринку.

Висновок

Дисертаційна робота Празднікової Маргарити Олександрівни на тему "Моделі і методи машинного навчання в задачах прогнозування ускладнень серцево-судинних хвороб" є завершеною науково-дослідною роботою, виконаною на належному науковому рівні.

Зважаючи на вище сказане, можна стверджувати, що дисертаційна робота **Празднікової Маргарити** за актуальністю, науковою новизною, обґрунтованістю наукових положень, теоретичним та практичним значенням отриманих результатів повністю відповідає вимогам "Порядку присудження ступеня доктора філософії та скасування рішення разової спеціалізованої вченої ради закладу вищої освіти, наукової установи про присудження ступеня доктора філософії", затвердженого Постановою КМУ №44 від 12 січня 2022 року, а її автор заслуговує присудження їй ступеня **доктора філософії** за спеціальністю **122 – Комп'ютерні науки**.

Офіційний рецензент

кандидат технічних наук, старший науковий співробітник,
провідний науковий співробітник
відділу перетворювачів форми інформації
Інституту кібернетики імені В.М. Глушкова
Національної академії наук України



Ігор ГАЛЕЛЮКА

