

РОЗВІДУВАЛЬНИЙ АНАЛІЗ ДАНИХ ДЛЯ ПОШУКУ ОЗНАК НЕВІДПОВІДНОСТІ ВТРУЧАНЬ ДІАГНОЗУ В МЕДИЧНІЙ ПРАКТИЦІ

¹Вінницький національний технічний університет

Анотація

Розглянуто проблему можливих зловживань у системі охорони здоров'я України, пов'язаних зі штучним завищенням обсягів медичних послуг для збільшення фінансування від Національної служби здоров'я. Основна увага зосереджена на виявленні невідповідностей між діагнозами та медичними втручаннями за допомогою розвідувального аналізу даних та великих мовних моделей. Результати свідчать, що близько 90% випадків потребують експертної перевірки. Запропонований підхід демонструє перспективність використання великих мовних моделей для автоматизованого виявлення підозрілих записів у медичних даних.

Ключові слова: медична статистика, великі мовні моделі, LLM.

Abstract

The study addresses the issue of potential abuses in Ukraine's healthcare system related to the artificial inflation of medical service volumes to increase funding from the National Health Service of Ukraine. The focus is on detecting inconsistencies between diagnoses and medical interventions using exploratory data analysis and large language models. The results indicate that approximately 90% of cases require expert review. The proposed approach demonstrates the potential of large language models for the automated detection of suspicious entries in medical data.

Keywords: medical statistics, large language models, LLM.

Вступ

Медичні експерти та міжнародні звіти стверджують, що частина медичних закладів в Україні штучно завищують обсяги наданих послуг, щоб отримувати більше коштів від Національної служби здоров'я (НСЗУ). Через модель фінансування «гроші за пролікований випадок» медики створюють фіктивні записи та штучно перекодовують випадки, щоб забезпечити більші виплати [1]. Це підтверджується, наприклад, аномальним зростанням стаціонарного лікування у 2023 році, попри значний відтік населення через війну. Попри намагання НСЗУ впроваджувати нові, ефективніші методи оцінки реальних медичних потреб, ця тенденція поки нікуди не поділась. Це підтверджує і новий документ ВООЗ та Світового банку опублікований вже у 2024 [2].

У даному дослідженні здійснено розвідувальний аналіз даних (EDA) на основі відкритого набору даних, що був опублікований на платформі Kaggle. Основною метою є знайти ознаки невідповідності втручань діагнозу в медичній практиці із використанням великих мовних моделей.

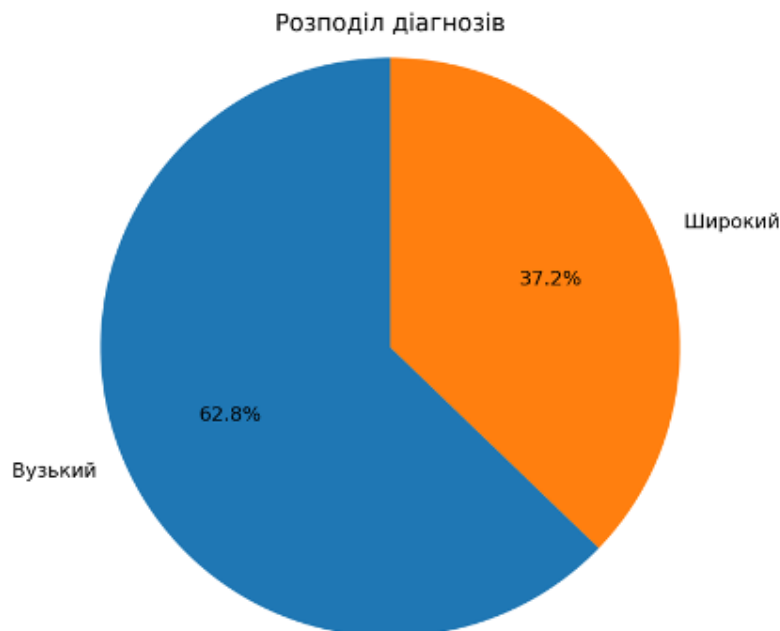
Результати дослідження

У дослідженні використано медичний датасет, що містить інформацію про близько 280 000 пацієнтів за тримісячний період. З кожного запису про пацієнта були використані наступні значення:

- patient_id – номер пацієнта для ідентифікації
- patient_gender – стать;
- age_group – вікова група;
- diagnosis_icd_10 – код діагнозу в системі International Statistical Classification of Diseases and Related Health Problems, 10th Revision (ICD-10);
- interventions_ambulatory_achi – коди виконаних медичних втручань в системі Australian Classification of Health Interventions (ACHI) на амбулаторному рівні;
- interventions_inpatient_achi – коди виконаних медичних втручань в системі ACHI на стаціонарному рівні;
- completed_interventions_achi – коди завершених медичних втручань в системі ACHI;
- not_completed_interventions_achi – коди незавершених медичних втручань в системі ACHI;

На фрагменті датасету що наведений на Рисунку 1 можна побачити, що поле `diagnosis_icd_10` часто містить діапазон значень. Це може вказувати як на кілька суміжних діагнозів у пацієнта, так і на надто широкий запис. Надто широкий чи неточний запис діагнозу привертає особливу увагу, бо може приховувати за собою призначення додаткових невинуватених чи зайвих скерувань.

Рис. 1. Фрагмент датасету



Для розвідувального аналізу було вирішено використати великі мовні моделі, зокрема модель gpt-4o-mini від компанії OpenAI.

Результати виконання вказують на те, що приблизно 90% записів з групи потребують уваги відповідного спеціаліста з різних причин. Графічне представлення результат обробки записів можна побачити на Рисунку 3.

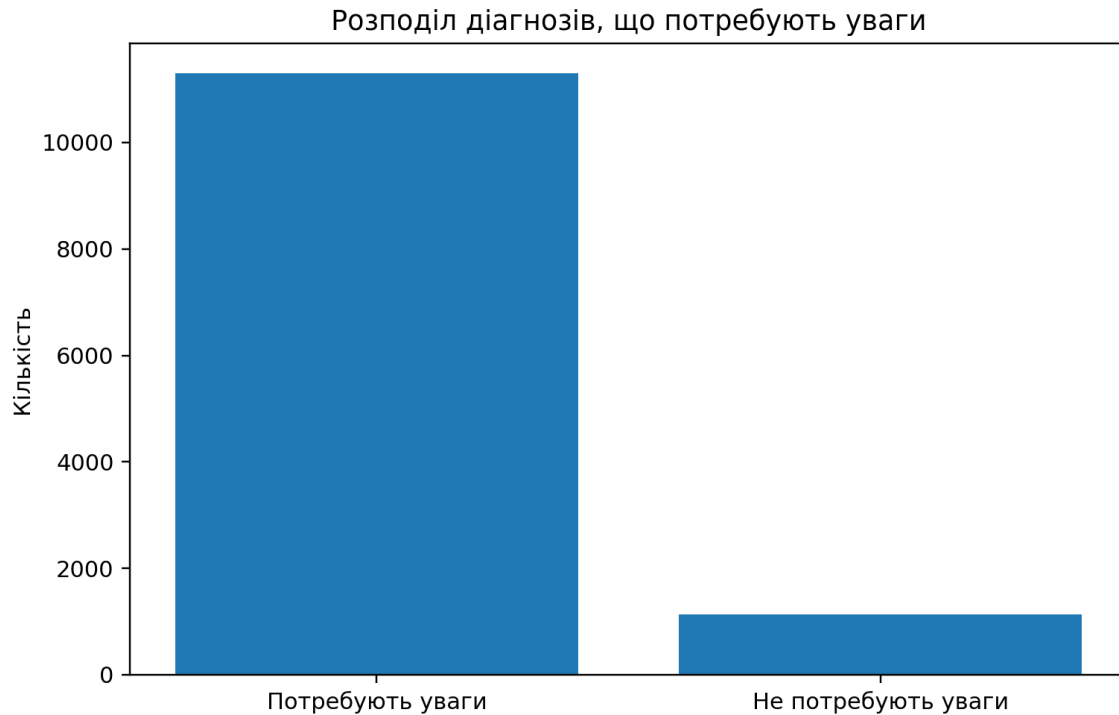


Рис. 3. Розподіл діагнозів, що потребують уваги

Висновки

Проведене дослідження демонструє, що значна частина записів у вибраній із датасету групі містить потенційні невідповідності між діагнозами та втручаннями, що може вказувати на можливі зловживання в певних закладах охорони здоров'я, оскільки близько 90% проаналізованих записів потребують уваги фахівців. Використання великих мовних моделей разом зі структурованим виводом і функційним викликом, дозволило автоматизувати перевірку медичних записів і підвищити точність виявлення подібних записів. Отримані результати підкріплюють припущення про можливі спроби штучного завищення кількості послуг, що надаються, з метою отримання більшого фінансування.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Бреденкамп, Керін, Еліна Далс, Олена Дорошенко, Юрій Джигир, Ярно Хабіхт, Лорейн Хоукінс, Олександр Кацага, Катерина Майнзюк, Христина Пак, та Ольга Зусь. 2022. Реформа фінансування системи охорони здоров'я в Україні: прогрес та майбутні напрями. Міжнародний розвиток у фокусі. Вашингтон, округ Колумбія: Світовий банк. doi:10.1596/978-1-4648-1908-7.
2. Фінансування системи охорони здоров'я в Україні: реформа, стійкість і відновлення. Копенгаген: Європейське регіональне бюро ВООЗ; 2024.

Бобко Богдан Володимирович — аспірант кафедри системного аналізу та інформаційних технологій, Вінницький національний технічний університет, Вінниця, e-mail: bobko.bogdan@gmail.com

Жуков Сергій Олександрович — кандидат технічних наук, доцент кафедри системного аналізу та інформаційних технологій, Вінницький національний технічний університет, Вінниця, e-mail: sazhukov@gmail.com

Bobko Bohdan V. – Postgraduate student of the Department of System Analysis and Information Technologies, Vinnytsia National Technical University, Vinnytsia, e-mail: bobko.bogdan@gmail.com

Zhukov Serhii O. – Candidate of technical sciences, associate professor of the department of System Analysis and Information Technologies, Vinnytsia National Technical University, Vinnytsia, e-mail: sazhukov@gmail.com