

УДК 004.03

Л. В. Волинець, Н. А. Гарматюк, В. А. Готович к.т.н.

(Тернопільський національний технічний університет імені Івана Пулюя, Україна)

## ВЕЛИКІ ЗА ОБСЯГОМ НАБОРИ БІОМЕДИЧНИХ ДАНИХ ТА МАШИННЕ НАВЧАННЯ

L. V. Volynets, N. A. Harmatiuk, V. A. Hotovych Ph.D.

### BIG BIOMEDICAL DATA SETS AND MACHINE LEARNING

На даний час великі за обсягом набори та колекції медичних даних продукуються різними за своєю природою способами. При цьому активно використовуються засоби візуалізації, які досягли значного прогресу, прискоривши процеси формування, зберігання та опрацювання біомедичних зображень. Експоненційне зростання обсягів даних формує потребу розроблення нових методів аналітичного опрацювання великих даних, що базуються на методах машинного навчання та штучного інтелекту.

Великі за обсягом дані характеризуються різноманітністю, обсягом, і швидкістю, достовірністю, цілісністю тощо. У сфері охорони здоров'я, набори та колекції даних отримані від постачальників медичних послуг, здебільшого містять пов'язані з лікуванням пацієнтів дані [1] (див. рис. 1).



Рисунок 1. Класифікація джерел великих за обсягом медичних даних

Розвиток інформаційних і комунікаційних технологій призвів до збільшення обсягів збирання медичних даних за допомогою автоматизованих інформаційних джерел, зокрема, медичних приладів, високопродуктивних пристроїв потокового спостереження та сенсорного IoT-обладнання. Зібрані великі за обсягом набори та колекції медичних даних використовуються для різнотипових задач, у тому числі для пошуку ліків, діагностування, прогнозування процесів перебігу та уточнення медичних діагнозів тощо. При цьому вони відіграють ключову роль [2] у наукових дослідженнях в галузі охорони здоров'я, зокрема, при спостереженні соціальних комунікацій пацієнтів, управлінні та плануванні медичних послуг, медичній промисловості тощо. Значна частина великих з обсягом медичних даних формується на основі зображень, що отримані за допомогою ехографії, комп'ютерної томографії, МРТ, Мамографії, і

т. п. Тому, зважаючи на експоненційні обсяги зростання наборів та колекцій медичних даних, формується потреба розробки швидкого та ефективних засобів аналітичного опрацювання біомедичних зображень [3]. Адже на основі виконаної належним чином класифікації зображень, можна досягти кращої діагностики захворювань.

Класифікаційні алгоритми можуть бути адаптовані до груп біомедичних зображень. Тому, класифікація є одним з ключових етапів у схемах біомедичних досліджень [4]. Машинне навчання – це набір методів штучного інтелекту, який формує здатність інформаційних-систем знаходити власні вирішення задач шляхом виявлення закономірностей у наборах даних. Алгоритми машинного навчання дають інформаційним-системам змогу виявляти закономірності та здійснювати класифікацію наявного набору даних та допомагають у процесах прийняття рішень. При використанні методів машинного навчання видобуті знання продукуються на основі сформованого досвіду інформаційних систем, а математичні та статистичні моделі використовуються для навчання на наборах даних.

Впродовж останнього періоду часу відбулося суттєве вдосконалення методів машинного навчання та глибинного навчання для класифікації біомедичних зображень [5]. На даний час існує два основних підходи до впровадження методів машинного навчання – це символний та субсимвольний [6]. Процеси використовуються для організації баз даних та знань біомедичних зображень для предіагностики різнотипових захворювань та при проведенні різноманітних досліджень. Автор [7] запропонував метод самонавчання програмно-алгоритмічних засобів на основі перемаркування. Для управління медичними даними було виділено множину ознак з вищими показниками результатів класифікації. Відповідно до експертних знань предметної області, надійні немарковані дані медичних записів були відібрані для розширення навчальної вибірки. Потім продуктивність класифікатора з трьома моделями навчання була оптимізована за допомогою керованої моделі навчання для формування трьох основних класів.

На даний час, великі за обсягом набори та колекції медичних даних, відіграють ключову роль при оперативному реагуванні на спалахи інфекційних захворювань, тому в процесі подальших досліджень доцільно провести розширений аналіз методів та засобів класифікації біомедичних зображень.

### Література

1. Tchito Tchanga, C., Mih, T. A., Tchagna Kouanou, A., Fozin Fonzin, T., Kuetche Fogang, P., Mezatio, B. A., & Tchiotsop, D. (2021). Biomedical image classification in a big data architecture using machine learning algorithms. *Journal of Healthcare Engineering*, 2021, 1-11.
2. Lopez, D., & Durai, M. A. (Eds.). (2017). *HCI Challenges and Privacy Preservation in Big Data Security*. IGI Global.
3. Belle, A., Thiagarajan, R., Soroushmehr, S. M., Navidi, F., Beard, D. A., & Najarian, K. (2015). Big data analytics in healthcare. *BioMed research international*, 2015.
4. Reddy, C. K., & Sun, J. (2013). Big data analytics for healthcare. In *Tutorial presentation at the SIAM International Conference on Data Mining, Austin, TX. achc.org.co* (Vol. 4).
5. Pitropakis, N., Panaousis, E., Giannetsos, T., Anastasiadis, E., & Loukas, G. (2019). A taxonomy and survey of attacks against machine learning. *Computer Science Review*, 34, 100199.
6. Kouanou, A. T., Tchiotsop, D., Kengne, R., Zephirin, D. T., Armele, N. M. A., & Tchinda, R. (2018). An optimal big data workflow for biomedical image analysis. *Informatics in Medicine Unlocked*, 11, 68-74.
7. Wang, L., Qian, Q., Zhang, Q., Wang, J., Cheng, W., & Yan, W. (2022). Classification model on big data in medical diagnosis based on semi-supervised learning. *The Computer Journal*, 65(2), 177-191.