



УДК 004.852: 616-7

MACHINE LEARNING IN THE ANALYSIS OF MEDICAL AND BIOLOGICAL DATA: FROM THEORY TO PRACTICE

МАШИННЕ НАВЧАННЯ В АНАЛІЗІ МЕДИКО-БІОЛОГІЧНИХ ДАНИХ: ВІД ТЕОРІЇ ДО ПРАКТИКИ

Yastrebova O.S. / Ястребова О.С.*assistant / асистент*

ORCID: 0009-0004-0892-0925

Frych N.I. / Фрич Н.І.*c.m.s., as.prof. / к.м.н., доц.*

ORCID: 0000-0003-1235-7157

Shvets L.S. / Швець Л.С.*c.b.s., as.prof. / к.б.н., доц.*

ORCID: 0000-0002-1481-2160

Drohomyretska Z.Y. / Дрогомирецька З.І.*assistant / асистент*

ORCID: 0009-0001-1910-2256

Kovalchuk L.Ye. / Ковальчук Л.Є.*d.m.s., prof. / д.м.н., проф.*

ORCID: 0000-0003-2878-3441

Dovganuch N.V. / Довганич Н.В.*c.b.s., as.prof. / к.б.н., доц.*

ORCID: 0009-0000-6361-6791

Ivano-Frankivsk National Medical University, Department of Medical Biology and Medical Genetics, Ivano-Frankivsk, Halytska, 2, 76018

Івано-Франківський національний медичний університет, кафедра медичної біології і медичної генетики, Івано-Франківськ, Галицька, 2, 76018

Анотація. У статті розглянуто сучасні підходи до використання методів машинного навчання в аналізі медико-біологічних даних, що відкривають нові можливості для діагностики, прогнозування та персоналізованої медицини. Проаналізовано основні типи алгоритмів, які застосовуються для класифікації, кластеризації, регресійного аналізу та виявлення закономірностей у великих біомедичних масивах. Особливу увагу приділено проблемам якості даних, збалансованості вибірок та інтерпретації моделей, що залишається одним із ключових викликів у медичних дослідженнях. Наведено приклади практичного застосування машинного навчання для аналізу геномних, протеомних, клінічних та візуальних даних. Основна мета дослідження — продемонструвати ефективність поєднання біологічного контексту з аналітичними інструментами штучного інтелекту для отримання клінічно значущих результатів. Зроблено висновок, що інтеграція методів машинного навчання у біомедичні дослідження є важливим напрямом розвитку сучасної медицини, який сприяє переходу від описової до прогностичної моделі охорони здоров'я.

Ключові слова: машинне навчання; штучний інтелект; медико-біологічні дані; аналіз великих даних; персоналізована медицина; алгоритми класифікації; прогнозування захворювань; нейронні мережі; геноміка; цифрова медицина.

Вступ.

Сучасна медицина перебуває на етапі стрімкої цифрової трансформації,



коли обсяг медико-біологічних даних зростає експоненційно завдяки розвитку геноміки, протеоміки, медичної візуалізації та електронних систем охорони здоров'я. Аналіз таких великих і різномірних масивів інформації вимагає нових підходів, здатних виявляти складні закономірності, недоступні для традиційних статистичних методів. У цьому контексті машинне навчання — як галузь штучного інтелекту — стає одним із найперспективніших інструментів для обробки, інтерпретації та прогнозування біомедичних процесів.

Використання алгоритмів машинного навчання у біології та медицині відкриває широкі можливості: від виявлення молекулярних біомаркерів і моделювання перебігу захворювань до персоналізації лікування та автоматизації діагностичних процедур. Проте впровадження цих методів супроводжується низкою викликів — від забезпечення якості даних і репрезентативності вибірок до необхідності пояснення моделей і дотримання етичних принципів.

Отже, актуальність дослідження зумовлена потребою глибшого розуміння можливостей і обмежень машинного навчання у сфері медико-біологічного аналізу. Основний напрям статті полягає у систематизації теоретичних засад, розгляді практичних прикладів застосування алгоритмів машинного навчання та визначенні перспектив розвитку цього напрямку в контексті сучасної медицини.

Основна частина.

Машинне навчання (МН) є підрозділом штучного інтелекту, що ґрунтується на здатності комп'ютерних систем автоматично виявляти закономірності в даних та приймати рішення без прямого програмування [1]. Залежно від типу завдання, виділяють навчання з учителем, без учителя, напівконтрольоване навчання та навчання з підкріпленням [3]. У медико-біологічних дослідженнях найбільш поширеними є алгоритми класифікації (логістична регресія, дерева рішень, метод опорних векторів, нейронні мережі) та регресійного аналізу (лінійна, поліноміальна, LASSO-регресія) [1; 4].

Особливу роль відіграють глибокі нейронні мережі, здатні аналізувати складні багатовимірні дані, зокрема медичні зображення та омічні профілі [4; 6]. Поєднання МН із методами статистики та біоінформатики дозволяє створювати



моделі, що не лише класифікують або прогнозують результати, а й надають нові біологічні інсайти [2].

Медико-біологічні дані мають складну структуру і включають різні типи інформації: геномні послідовності, експресію генів, протеомні профілі, клінічні показники, результати лабораторних досліджень, електронні медичні записи, а також медичні зображення [5]. Їхня різноманітність створює як нові можливості, так і виклики для дослідників [6]. Перед застосуванням алгоритмів машинного навчання необхідно здійснити попередню обробку: нормалізацію, усунення пропущених значень, масштабування та кодування категоріальних змінних [8]. Особливу увагу слід приділяти збалансованості вибірки, оскільки нерівномірне представлення класів може призвести до спотворення результатів моделі [3].

Важливою частиною процесу є також відбір ознак (feature selection), що дозволяє зменшити вимірність даних і покращити інтерпретацію результатів. Для цього застосовують такі підходи, як аналіз головних компонент (PCA), методи регуляризації або алгоритми на основі інформаційної ентропії [1; 4].

Методи МН успішно застосовується у низці напрямів сучасної біомедицини.

У геноміці – для прогнозування патогенності мутацій, ідентифікації генетичних асоціацій із захворюваннями, побудови моделей ризику розвитку спадкових патологій [2; 6].

У медичній діагностиці – для автоматизованої інтерпретації рентгенівських, КТ та МРТ-зображень, що значно підвищує точність і швидкість встановлення діагнозів [3; 4].

У фармакології – для виявлення нових мішеней дії лікарських засобів і моделювання взаємодій «лікарська речовина – білок».

У персоналізованій медицині – для створення предиктивних моделей лікувальної відповіді, підбору оптимальної терапії та оцінки прогнозу перебігу хвороби [7].

Попри значний потенціал машинного навчання, його застосування у медицині пов'язане з низкою проблем. Серед основних — обмежена доступність



якісних і репрезентативних медичних даних, складність забезпечення конфіденційності пацієнтів, ризик перенавчання моделей та обмежена інтерпретованість складних алгоритмів [5; 8]. Перспективи розвитку галузі пов'язані з інтеграцією багаторівневих біологічних даних (multi-omics), створенням прозорих та етично обґрунтованих моделей штучного інтелекту, а також удосконаленням підходів до пояснення результатів [2; 4]. Очікується, що подальше поєднання біологічних знань і обчислювальної аналітики сприятиме формуванню нової парадигми медицини — від реактивної до превентивної та персоналізованої [1].

Висновки.

Застосування методів машинного навчання в аналізі медико-біологічних даних демонструє значний потенціал для розвитку сучасної медицини, зокрема у напрямках ранньої діагностики, прогнозування перебігу захворювань і персоналізованої терапії. Проведений аналіз показує, що машинне навчання дозволяє ефективно працювати з великими обсягами різномірних даних, виявляючи складні нелінійні взаємозв'язки між біологічними параметрами, які не можуть бути описані традиційними статистичними підходами.

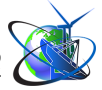
Водночас ефективність моделей безпосередньо залежить від якості та збалансованості вхідних даних, правильного вибору алгоритмів, а також від наявності біологічно обґрунтованої інтерпретації результатів. Для впровадження таких систем у клінічну практику важливим є питання прозорості, верифікації та етичної відповідальності використання штучного інтелекту у медицині.

Інтеграція машинного навчання з біоінформатикою, геномікою, протеомікою та цифровою медициною сприятиме створенню комплексних аналітичних платформ, здатних не лише підтримувати прийняття клінічних рішень, а й формувати нові знання про механізми захворювань. У перспективі це забезпечить перехід від описового до прогностичного рівня біомедичних досліджень і сприятиме формуванню інтелектуальної медицини майбутнього, орієнтованої на пацієнта, доказовість і технологічну ефективність.



Література:

1. Rahmani, A. M., Yousefpoor, E., Yousefpoor, M. S., Mehmood, Z., Haider, A., Hosseinzadeh, M., & Ali Naqvi, R. (2021). *Machine Learning (ML) in Medicine: Review, Applications, and Challenges*. Mathematics, 9(22), Article 2970. <https://doi.org/10.3390/math9222970> MDPI+1
2. Cai, Z., Poulos, R. C., Liu, J., & Zhong, Q. (2022). *Machine learning for multi-omics data integration in cancer*. iScience, 25(2), Article 103798. <https://doi.org/10.1016/j.isci.2022.103798> PubMed
3. Sim, J. Z. T., Lee, S., Goh, J. Y., Wang, S., & Poh, K. K. (2021). *Machine learning in medicine: what clinicians should know*. BMJ Health & Care Informatics, 28(1), e10071847. <https://doi.org/10.1136/bmjhci-2021-10071847> PMC
4. Chakraborty, C., Pan, D., & Ghosh, S. (2024). *From machine learning to deep learning: Advances of the next generation in medicine*. Patterns, 05(06), 100934. <https://doi.org/10.1016/j.patter.2024.100934> ScienceDirect
5. Kolasa, K., Tambor, M., & Sagan, A. (2024). *Systematic reviews of machine learning in healthcare: A scoping review*. Journal of Evidence-Based Medicine, 17(1), 3–20. <https://doi.org/10.1080/14737167.2023.2279107> Taylor & Francis Online
6. Wang, S., Wang, S., & Wang, Z. (2023). *A survey on multi-omics-based cancer diagnosis using machine learning with the potential application in gastrointestinal cancer*. Frontiers in Medicine, 9, Article 1109365. <https://doi.org/10.3389/fmed.2022.1109365> Frontiers
7. Peng, J., Zhang, W., Li, H., Zhao, Y., & Zhu, L. (2021). *Machine Learning Techniques for Personalized Medicine: Applications in Autoimmune Rheumatic Diseases*. Frontiers in Pharmacology, 12, 720694. <https://doi.org/10.3389/fphar.2021.720694> Frontiers
8. Cho, H., Kim, J., Lee, Y., & Lee, J. (2024). *Machine Learning and Health Science Research: Tutorial*. Journal of Medical Internet Research, 26(1), e50890. <https://doi.org/10.2196/50890> jmir.org



Abstract. *The article discusses modern approaches to the use of machine learning methods in the analysis of biomedical data, which opens new opportunities for diagnostics, prediction, and personalized medicine. The main types of algorithms applied for classification, clustering, regression analysis, and pattern recognition in large biomedical datasets are analyzed. Special attention is given to issues of data quality, sample balance, and model interpretability, which remain key challenges in medical research. Examples of practical applications of machine learning for the analysis of genomic, proteomic, clinical, and imaging data are provided. The main aim of the study is to demonstrate the effectiveness of combining biological context with artificial intelligence analytical tools to obtain clinically significant results. It is concluded that the integration of machine learning methods into biomedical research is a crucial direction in the development of modern medicine, promoting the transition from a descriptive to a predictive model of healthcare.*

Keywords: *machine learning; artificial intelligence; biomedical data; big data analysis; personalized medicine; classification algorithms; disease prediction; neural networks; genomics; digital medicine.*

Статтю надіслано: 14.10.2025 р.

© Ястребова О.С.