

УДК 004.9:602

DOI: 10.20998/2411-0558.2025.02.10

О. М. ДАЦОК, канд. техн. наук, доц. каф. БМІ ХНУРЕ, Харків,

В. О. ДЯЧЕНКО, ст. викл. каф. ЕОМ ХНУРЕ, Харків,

А. С. ГУК, ст. викл. каф. ЕОМ ХНУРЕ, Харків,

Л. Р. ФАТІЙ, магістр, каф. ЕОМ ХНУРЕ, Харків

ЕТАПИ ОБРОБКИ МЕДИЧНИХ ДАНИХ ДЛЯ РОЗВ'ЯЗАННЯ ЗАДАЧ АЛГОРИТМАМИ МАШИННОГО НАВЧАННЯ

У роботі висвітлено актуальність і перспективи застосування машинного навчання для обробки медичних даних. Розглянуто різноманіття джерел та форматів інформації, а також описано ключові алгоритми машинного навчання. Особливу увагу приділено попередній обробці та підготовці даних, що відіграє вирішальну роль у забезпеченні точності, надійності та етичності побудованих моделей. Результати роботи демонструють важливість мультидисциплінарного підходу, в якому поєднуються зусилля фахівців із медицини, технічних спеціалістів і дослідників із питань конфіденційності, з метою підвищення ефективності діагностики та лікування в сучасній охороні здоров'я. Л.: 1. Бібліогр.: 16 назв.

Ключові слова: машинне навчання, Big Data, медичні дані, валідація, кластеризація, нормалізація, аномальні значення, енергозбереження.

Постановка проблеми. Медичні дані сьогодні є однією з найважливіших складових сучасної охорони здоров'я. Завдяки їм лікарі отримують змогу оцінювати стан пацієнтів, прогнозувати перебіг захворювань і розробляти персоналізовані стратегії лікування. Зростання обсягу таких даних відбувається надзвичайно швидко: електронні медичні картки тепер містять не лише традиційні показники (наприклад, результати аналізів крові чи інформацію про проведені процедури), а й цілі масиви цифрових зображень, дані з зовнішніх пристроїв та навіть геномні дані. Водночас, із розвитком технологій зібрати велику кількість інформації стає простіше, проте з'являються нові завдання: як структурувати ці дані, як якісно їх аналізувати й які висновки можна робити для реальної клінічної практики.

Одним із ключових викликів є те, що "традиційні" статистичні методи часто не встигають опрацювати настільки різноманітні, великі та складні за структурою набори даних. Тут на допомогу приходять

методи машинного навчання [1], які здатні знаходити приховані закономірності, передбачати результати лікування та допомагати лікарям у прийнятті клінічних рішень [2]. Застосування алгоритмів штучного інтелекту відкриває низку можливостей: від автоматизації рутинних завдань, як-от розпізнавання патологій на рентгенівських знімках, до розробки персоналізованих планів лікування, що враховують індивідуальні особливості пацієнта. Однак поряд із цим виникають питання, пов'язані з безпекою та конфіденційністю медичних даних, необхідністю дотримання етичних норм і нормативно-правових вимог.

У загальному підсумку, використання машинного навчання дає змогу суттєво підвищити ефективність медичної допомоги, скоротити час на встановлення діагнозу та поліпшити результати лікування. Але для досягнення цих переваг потрібний комплексний підхід: треба правильно організовувати збирання та обробку даних, залучати профільних фахівців (інженерів, біоінженерів, медиків, юристів), а також безперервно працювати над удосконаленням технологій. Саме на перетині цих процесів і з'являються найперспективніші рішення, які здатні докорінно змінити обличчя сучасної медицини і відчутно покращити якість життя людей.

Аналіз літератури. Аналіз публікацій за даною темою демонструє, що протягом останніх років спостерігається постійне зростання інтересу до застосування алгоритмів штучного інтелекту в медицині. В статті [3] обговорюється потенціал Big Data для клінічної практики, а також наголошується на потребі у розробці відповідних інструментів, які могли б допомогти лікарям ефективно використовувати ці інформаційні масиви. Автори акцентують увагу на тому, що традиційні статистичні методи вже не завжди здатні впоратися з величезними обсягами та швидко змінними наборами даних, тоді як машинне навчання здатне пропонувати гнучкіші та точніші підходи до аналізу.

Подібний висновок міститься і в огляді [4]. Автори систематизують основні підходи машинного навчання такі як: регресія, класифікація, глибокі нейронні мережі, й показують, як ці методи можуть підвищити ефективність роботи лікарів у різних напрямках – від прогнозування лікувальних результатів до підтримки прийняття клінічних рішень у реальному часі. Також автори наголошують на важливості належної

валідації алгоритмів на медичних даних та необхідності врахування етичних аспектів при розробці моделей.

Одним із найвідоміших прикладів успішного застосування алгоритмів глибокого навчання є дослідження [5], де було продемонстровано можливість класифікувати види раку шкіри з точністю на рівні досвідчених дерматологів. У цій роботі застосовувалися глибокі згорткові нейронні мережі, які надзвичайно добре зарекомендували себе в аналізі зображень. Результати дослідження відкрили шлях до використання подібних алгоритмів для первинного скринінгу дерматологічних захворювань, що може суттєво пришвидшити встановлення діагнозу й зменшити кількість пропущених небезпечних патологій.

Подальший розвиток машинного навчання в медицині розглядається в [6] та [7]. В даній роботі йдеться про важливість інтеграції машинного навчання в рутинну медичну практику: від автоматизованого опрацювання електронних медичних записів до створення розумних систем підтримки прийняття рішень, які будуть здатні аналізувати історію хвороби та біомаркери пацієнта. Водночас відзначається, що швидке впровадження таких технологій вимагає високих обчислювальних ресурсів і чітких регуляторних норм, зокрема у сфері збереження конфіденційності даних.

Окремо варто відзначити колективну монографію [8], у якій висвітлюються сучасні підходи до проектування алгоритмів штучного інтелекту для охорони здоров'я. Це видання узагальнює великий досвід фахівців із різних країн у застосуванні нейронних мереж, методів обробки природної мови та інших інструментів машинного навчання для розв'язання реальних клінічних задач. Також у книзі значна увага приділена темі інтерпретації результатів та питанням довіри з боку лікарів і пацієнтів до "чорних скриньок" алгоритмів.

У контексті медичної візуалізації цікаві узагальнюючі матеріали пропонують в [9]. Автори аналізують переваги та недоліки глибоких нейронних мереж при обробці магнітно-резонансних знімків, порівнюючи їх з традиційними методами ілюстрації та машинного бачення. Згадується, що штучні нейронні мережі здатні не лише покращувати точність виявлення патологій, а й допомагають у сегментації органів, автоматизованому визначенні меж уражень тощо. При цьому

підкреслюється, що ефективність навчання напряду залежить від якості та репрезентативності наборів даних, оскільки навіть найкращий алгоритм не зможе адекватно працювати з зашумленими або неповними даними.

Роботи [10] та [11] присвячені питанню енергозбереження при обробці даних, які передаються в безпроводних сенсорних мережах і можуть бути використані для зчитування показників життєдіяльності людини. В роботах пропонується використання інструментів машинного навчання. За рахунок автоматизації класифікації вузлів датчиків та зчитування з них збільшується енергоефективність роботи мережі, що, в свою чергу, призведе до отримання більшого набору даних пацієнта.

Загалом, аналіз існуючої літератури показує, що машинне навчання в медицині вже перейшло зі сфери винятково академічних досліджень у практичну площину, однак залишається ще дуже багато проблем: брак єдиних стандартів обміну медичними даними, труднощі з пояснюваністю складних моделей, необхідність у значних обчислювальних потужностях та забезпеченні захисту інформації при передаванні даних. Проте прогресивний досвід провідних медичних закладів і дослідницьких груп доводить: вдале впровадження штучного інтелекту може радикально змінити підхід до діагностики, лікування і моніторингу стану здоров'я пацієнтів, забезпечуючи вищу точність та індивідуальний підхід.

Мета статті – аналіз ключових етапів та методів підготовки інформації, необхідних для досягнення точних і достовірних результатів у клінічній практиці із застосуванням методів машинного навчання.

Джерела та типи медичних даних.

Медичні дані можуть бути надзвичайно різноманітними за своєю природою й форматом зберігання, що істотно впливає на методи їхньої обробки та аналізу. Залежно від рівня впорядкованості інформації, виділяють структуровані, напівструктуровані та неструктуровані дані. Також важливим джерелом є медичні зображення й дані з зовнішніх пристроїв чи інших сенсорів. Розуміння специфіки кожного з цих типів дає змогу розробляти ефективні алгоритми машинного навчання і забезпечувати точніший аналіз клінічних показників.

До структурованих даних відносяться ті, які зазвичай зберігаються у вигляді таблиць чи баз даних із жорстко визначеними полями й

атрибутами. У медицині це можуть бути електронні медичні картки та різні реєстри, де чітко прописано такі параметри, як вік, стать, діагноз, результати аналізів тощо. Завдяки такій впорядкованості структуровані дані зручні у попередній обробці й легко інтегруються з багатьма алгоритмами машинного навчання. Проте не всі показники важливі для діагностики чи лікування, можуть опинитися в цих базах; крім того, часто зустрічаються пропущені значення та невідповідності у форматах зберігання навіть у межах одного типу даних.

Неструктуровані дані становлять найбільш складний для автоматизованого опрацювання пласт інформації, адже до них належать текстові нотатки лікарів, епікризи, висновки спеціалістів, листи пацієнтів та інші документи у вільному форматі. Такі тексти містять важливі клінічні деталі, проте відсутність чіткої структури ускладнює автоматичний аналіз. Для розуміння та витягання релевантної інформації часто застосовують методи обробки природної мови, зокрема моделі на основі глибоких нейронних мереж, що дає змогу розпізнавати медичні терміни, класифікувати текст за темами чи навіть виявляти ключові взаємозв'язки між симптомами й діагнозами.

Напівструктуровані дані включають різноманітні звіти, які містять як текст, так і певні шаблони, структуровані блоки чи формалізовані теги. Сюди можна віднести, наприклад, PDF-файли з результатами аналізів, у яких є чіткі таблиці з показниками та розшифрування у вигляді вільного тексту. Для роботи з ними використовують комбіновані методи: вилучення таблиць та розмічених фрагментів і обробку природної мови для аналізу текстової складової. Важливою проблемою лишається неоднорідність форматів, адже кожна клініка чи лабораторія може мати власні, індивідуальні форми звітів.

Окрему категорію становлять медичні зображення, зокрема рентгенівські знімки, результати магнітно-резонансної та комп'ютерної томографії, ультразвукові зображення, а також гістологічні зразки. Обробка таких даних потребує спеціалізованих методів комп'ютерного зору: від традиційних алгоритмів до глибоких згорткових нейронних мереж, які демонструють вражаючі результати в розпізнаванні патологій і сегментації органів.

У сучасній практиці невпинно зростає роль даних із портативних пристроїв та IoT-обладнання [12]. Це можуть бути фітнес-трекери,

розумні годинники чи портативні монітори, які збирають інформацію про частоту серцебиття, рівень фізичної активності, сон, артеріальний тиск тощо. Ці дані бувають дуже неоднорідними, оскільки різні виробники використовують власні стандарти збирання та зберігання інформації. Проте їхній аналіз відкриває можливості для більш точного моніторингу здоров'я пацієнта у реальному часі та попереджувальних інтервенцій [13].

Остання, але не менш важлива категорія – результати лабораторних та діагностичних досліджень. Здебільшого це кількісні або якісні показники, які можуть бути представлені в табличному форматі, однак для правильного трактування нерідко потребують урахування клінічного контексту, адже однакові показники можуть мати різне значення для різних груп пацієнтів чи при різних супутніх хворобах.

Отже, кожен тип медичних даних має свої особливості, які впливають на вибір методів машинного навчання та інструментів попередньої обробки. У медичних проєктах часто стикаються з потребою інтегрувати кілька різних джерел інформації. В якості прикладу можна навести поєднання структурованих даних з електронних карток, текстових нотатків лікарів і зображень сканів комп'ютерної томографії. Саме такий мультимедійний підхід може дати найповніше уявлення про стан пацієнта й забезпечити прийняття більш обґрунтованих клінічних рішень.

Використання методів машинного навчання у сучасній медицині.

У сфері охорони здоров'я процес застосування машинного навчання починається з формулювання конкретної клінічної задачі: діагностики захворювання, прогнозування ускладнень, аналіз ризиків, автоматизації виявлення патологічних утворень тощо. Далі відбувається збирання та підготовка даних, після чого обирають і навчають модель, здатну розпізнати ці дані у вхідній інформації. Завершальним етапом є оцінка точності й узагальнювальної здатності моделі, а також її впровадження у реальну клінічну практику з урахуванням питань безпеки й етики.

Метод	Основні завдання	Переваги	Недоліки
Логістична регресія	бінарна класифікація	легка інтерпретація	порівняно проста модель
	оцінка ризиків	простота реалізації	може не враховувати складні залежності
Дерева рішень	класифікація,	інтерпретованість правил	схильність до перенавчання
	регресія	підходить для різних типів даних	потрібне ретельне налаштування
Випадковий ліс	класифікація,	висока точність	менш інтерпретовані, ніж окремі дерева
	регресія,	стійкість до шумів і пропусків у даних	більші вимоги до обчислювальних ресурсів
Метод опорних векторів	класифікація (бінарна, багатокласова)	добре працює з високою розмірністю	чутливість до параметрів ядра
	регресія	можна вибирати різні ядра	не завжди підходить для дуже великих вибірок
Глибокі нейронні мережі	розпізнавання зображень,	висока ефективність на великих вибірках	потреба у великих наборах даних
	обробка текстів, аналіз часових рядів	автоматичне виявлення ознак	складне пояснення
Кластеризація	пошук підгруп у вибірці,	дає змогу знаходити приховані структури	важко інтерпретувати кластери
	виявлення аномалій	працює без розмічених даних	результати чутливі до гіперпараметрів
Гібридні методи	розв'язання задач з обмеженою кількістю міток	можна залучати велику кількість нерозмічених даних	методи складніші у налаштуванні
		зменшує витрати на розмітку	ризик підлаштування до даних з шумом без розмітки
Навчання з підкріпленням	роботизовані операції	можливість навчатися на власному досвіді	важко створити безпечне середовище навчання
	динамічне планування лікування	підходить для складних систем із зворотним зв'язком	великі вимоги до обчислень і налаштування нагород

Рис. 1. Аналіз використання методів машинного навчання у сфері охорони здоров'я

Методи з вчителем найчастіше застосовують для розв'язання задач класифікації та регресії. Методи без вчителя дають змогу виявляти підгрупи пацієнтів зі спільними ознаками чи аномальні відхилення в показниках, про які раніше не було відомо. Глибоке навчання відкрило нові можливості в аналізі зображень і текстів. Крім того, гібридне навчання дозволяє ефективно працювати з великими масивами даних, коли розмічених прикладів небагато. В роботі проведений аналіз існуючих методів машинного навчання в медицині. Аналіз проведено за критеріями виконання основних завдань, переваг використання та недоліків.

Загальна структура процесу застосування машинного навчання в медицині представлена на рис. 2. Вона умовно відображає послідовність етапів – від збирання даних до впровадження в клінічну практику.

Етап збору та попередньої обробки даних охоплює вибірку різних джерел. На етапі формування та очищення вибірки відбувається видалення дублікатів, корекція аномалій, заповнення пропусків.

Вибір методу машинного навчання залежить від поставленої задачі) і формату даних. На етапі навчання та оптимізації моделі налаштовуються гіперпараметри та регуляризація для запобігання перенавчанню.



Рис. 2. Процес застосування машинного навчання у сфері охорони здоров'я

Етап валідації та тестування дає змогу оцінити узагальнювальну здатність моделі, використовуючи тестову вибірку або перехресну перевірку. Далі здійснюється оцінка точності показників, а на останньому етапі модель інтегрується у клінічні процеси, а також відбувається постійний моніторинг її продуктивності та коригування за необхідності. Зворотні зв'язки на схемі (див. рис. 2) забезпечують коректний підхід для налаштування параметрів, враховуючи отримані результати.

Попередня обробка та підготовка медичних даних для застосування методів машинного навчання.

На етапі обробки та підготовки медичних даних формується якісний, чистий і зручний для аналізу набір даних, який у подальшому дозволить алгоритмам досягти високих показників точності. У медичній сфері складність підготовки даних зумовлена їх різноманітністю.

Зазвичай етап попередньої обробки охоплює кілька ключових фаз. Збір та інтеграція даних може об'єднувати інформацію з електронних медичних карток, лабораторних систем, баз зображень та ін.

Наступним завданням є очищення даних, що передбачає виявлення дублікатів записів, виправлення або вилучення некоректних значень, а також аналіз та обробку пропущених значень. У медичному контексті

відсутність інформації може бути досить поширеним явищем: пацієнти не завжди здають повний перелік аналізів, або дані можуть бути заархівованими у різних форматах без централізованого оновлення. Водночас пропуски не завжди слід видаляти: їх можна заповнювати за допомогою статистичних або інтелектуальних методів.

Ще одним завданням є виявлення та обробка аномальних значень: у великих медичних вибірках можуть траплятися записи з екстремальними показниками. Якщо такий запис зумовлений помилкою, він потенційно спотворить підсумковий аналіз. Однак у деяких випадках екстремальні значення можуть бути справжньою клінічною аномалією, сигналізуючи про критичний стан, що потребує додаткової уваги лікарів. Тому важливо розрізняти технічні та реальні аномалії.

Масштабування та нормалізація ознак також часто є важливим етапом. Для багатьох алгоритмів надзвичайно великі або надзвичайно малі значення показників можуть спричинити уповільнення чи нестабільність навчання. Використовують стиснення діапазону даних, зазвичай до $[0; 1]$. Доречність конкретного методу залежить від характеру медичних показників і вимог щодо інтерпретації результатів.

Наступним важливим завданням є вибір і створення ознак. Методи, спрямовані на відбір найінформативніших змінних дозволяють зменшити розмірність проблеми та підвищити точність і швидкість роботи алгоритмів. Водночас фахова експертиза лікарів і медичних дослідників може бути незамінною для виділення обчислювально простих, але дуже важливих закономірностей.

Нарешті, у медичних проєктах особливу роль відіграють питання анонімізації й захисту персональних даних. При обробці персональних відомостей пацієнтів необхідно строго дотримуватися законодавчих вимог та етичних принципів, аби гарантувати, що конфіденційна інформація не буде розкрита стороннім особам і не буде використана поза межами первинної згоди. Застосовують різноманітні техніки деідентифікації, а також протоколи безпечного передавання і зберігання даних.

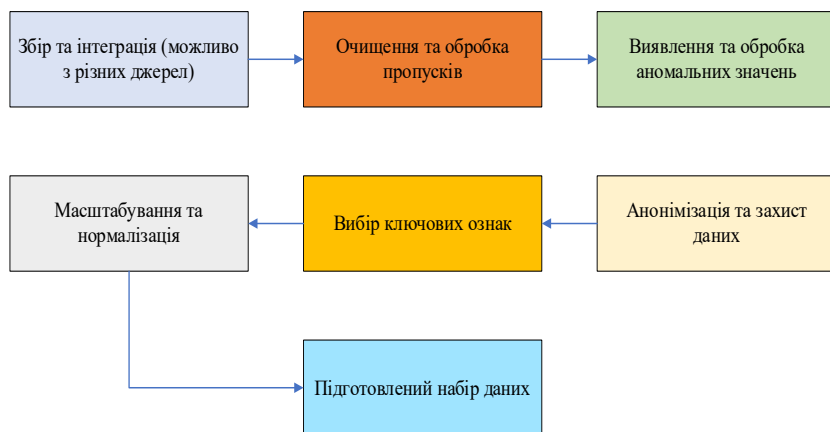


Рис. 3 Послідовність основних кроків у попередній обробці та підготовці медичних даних

У комплексі ці кроки (рис. 3) закладають фундамент для наступних процедур машинного навчання: від вибору відповідної моделі до валідації результатів і безпечного розгортання систем у реальному медичному середовищі. При правильній реалізації попередньої обробки можна уникнути багатьох проблем, зменшити ймовірність похибок і забезпечити більш високу точність алгоритмів, що безпосередньо впливає на якість діагностики й лікування.

Висновки. Машинне навчання в медицині не є просто набором універсальних алгоритмів. Його успішне застосування потребує глибокого розуміння природи медичних даних, специфіки їхнього збирання, структурування та очищення, а також спільної роботи лікарів, інженерів, дослідників і фахівців із біоетики. Правильна обробка даних – включно з виявленням та усуненням аномальних записів, масштабуванням показників, забезпеченням конфіденційності та безпеки – є фундаментом, що дозволяє уникнути суттєвих похибок і максимально розкрити потенціал алгоритмів штучного інтелекту. Такий комплексний підхід дає змогу не лише підвищити точність та ефективність аналізу, а й істотно поліпшити якість і швидкість надання медичної допомоги.

Список літератури:

1. Flach, P. A. Machine Learning: The Art and Science of Algorithms that Makes Sense of Data. Cambridge: Cambridge University Press, 2012, 291 p.
2. Perova, I., Bodyanskiy Y., Adaptive human machine interaction approach for feature selection-extraction task in medical data mining // International Journal of Computing, 17 (2), 2018, pp. 113-119.

3. Obermeyer, Z., Emanuel, EJ. Predicting the Future—Big Data, Machine Learning, and Clinical Medicine // *New England Journal of Medicine*, 375(13), 2016, pp. 1216-1219.
4. Rajkomar, A., Dean, J., Kohane, I. Machine Learning in Medicine // *New England Journal of Medicine*, 380(14), 2019, pp. 1347–1358.
5. Esteva, A., Kuprel, B., Novoa, RA., Ko, J., Swetter, SM., Blau, HM., Thrun, S Dermatologist-level classification of skin cancer with deep neural networks // *Nature*, 542(7639), 2017, pp. 115–118.
6. Topol. EJ. High-performance medicine: the convergence of human and artificial intelligence // *Nature Medicine*, 2019, pp. 44–56.
7. Beam, AL., Kohane, IS. Big Data and Machine Learning in Health Care // *JAMA*. 320(11), 2018, pp. 1101–1102.
8. Bohr, A., Memarzadeh, K., *Artificial Intelligence in Healthcare*. Academic Press. 2020.
9. Lundervold, AS., Lundervold, A. An overview of deep learning in medical imaging focusing on MRI. // *Zeitschrift für Medizinische Physik*, 29(2), 2019, pp.102–127.
10. Дяченко В.О. Інтелектуальні підходи енергозбереження у безпроводних сенсорних комп'ютерних мережах // *Системи управління, навігації та зв'язку*. Збірник наукових праць. Полтава, 2020. 4 (62). с. 114-118.
11. Дяченко В.О., Міхаль О.П. Перспективність застосування вдосконаленого класичного алгоритму карт Кохонена у розподілених енергокритичних сенсорних мережах // *Системи управління, навігації та зв'язку*. 36. *Наук. праць*. 4 (74). Полтава, 2023. с. 75-79.
12. Perova, I., P. Zhernova, O. Datsok, Y. Bodyanskiy, and O. Velychko. Recognition of Human Primitive Motions for the Fitness Trackers, *Advances in Intelligent Systems and Computing*, Vol. 1020. 2020.
13. Velychko, O., Datsok, O., Perova, I. Essential R Peak Detector Based on the Polynomial Fitting, *Lecture Notes on Data Engineering and Communications Technologies*, 77, 2022, pp. 148–163.

References:

1. Flach, P. A. (2012), *Machine Learning: The Art and Science of Algorithms that Makes Sense of Data*. Cambridge: Cambridge University Press, 291 p.
2. Perova, I., Bodyanskiy Y. (2018), “Adaptive human machine interaction approach for feature selection-extraction task in medical data mining”, *International Journal of Computing*, 17 (2), pp. 113-119.
3. Obermeyer, Z., Emanuel, EJ. (2016). “Predicting the Future—Big Data, Machine Learning, and Clinical Medicine”, *New England Journal of Medicine*, 375(13), pp. 1216-1219.
4. Rajkomar, A., Dean, J., Kohane, I. (2019), “Machine Learning in Medicine”, *New England Journal of Medicine*, 380(14), pp. 1347–1358.
5. Esteva, A., Kuprel, B., Novoa, RA., Ko, J., Swetter, SM., Blau, HM., Thrun, S. (2017), “Dermatologist-level classification of skin cancer with deep neural networks”, *Nature*, 542(7639), pp. 115–118.
6. Topol. EJ. (2019), “High-performance medicine: the convergence of human and artificial intelligence”, *Nature Medicine*, pp. 44–56.
7. Beam, AL., Kohane, IS. (2018), “Big Data and Machine Learning in Health Care”. *JAMA*. 320(11), pp. 1101–1102.
8. Bohr, A., Memarzadeh, K. (Ed.). (2020), *Artificial Intelligence in Healthcare*. Academic Press.

9. Lundervold, AS., Lundervold, A. (2019), "An overview of deep learning in medical imaging focusing on MRI". *Zeitschrift für Medizinische Physik*, 29(2), pp.102–127.
10. Дяченко, В.О. (2020) "Інтелектуальні підходи енергозбереження у безпроводних сенсорних комп'ютерних мережах". Системи управління, навігації та зв'язку. Збірник наукових праць. Полтава, 4 (62). с. 114-118.
11. Дяченко, В.О., Міхаль, О.П. (2023) "Перспективність застосування вдосконаленого класичного алгоритму карт Кохонена у розподілених енергокритичних сенсорних мережах". Системи управління, навігації та зв'язку. Збірник наукових праць. 4 (74). Полтава, с. 75-79.
12. Perova, I., P. Zhernova, O. Datsok, Y. Bodyanskiy, and O. Velychko. (2020). "Recognition of Human Primitive Motions for the Fitness Trackers, *Advances in Intelligent Systems and Computing*, Vol. 1020.
13. Velychko, O., Datsok, O., Perova, I. (2022), "Essential R Peak Detector Based on the Polynomial Fitting", *Lecture Notes on Data Engineering and Communications Technologies*, 77, pp. 148–163.

Статтю представив д-р тухн. наук, проф. Національного технічного університету "Харківський політехнічний інститут" В.І. Носков.

Надійшла (received) 30.03.2025

Datsok Oleh, Cand.Sc.Tech, Associate Professor
Kharkiv National University of Radio Electronics
Nauky Ave. 14, Kharkiv, Ukraine, 61166
Tel: (057) 7021-364, e-mail: oleh.datsok@nure.ua
ORCID ID: 0000-0003-4489-3819

Diachenko Vladyslav, Lecturer
Kharkiv National University of Radio Electronics
Nauky Ave. 14, Kharkiv, Ukraine, 61166
Tel: (057) 7021-354, e-mail: vladyslav.diachenko@nure.ua
ORCID ID: 0000-0003-2725-8784

Huk Artem, Lecturer
Kharkiv National University of Radio Electronics
Nauky Ave. 14, Kharkiv, Ukraine, 61166
Tel: (057) 7021-354, e-mail: artem.huk@nure.ua
ORCID ID: 0009-0004-2011-1901

Fatii Lidiia, Master student
Kharkiv National University of Radio Electronics
Nauky Ave. 14, Kharkiv, Ukraine, 61166
Tel: (057) 7021-354, e-mail: lidiia.fatii@nure.ua
ORCID ID: 0009-0007-0558-4809

УДК 004.9:602

Етапи обробки медичних даних для розв'язання задач алгоритмами машинного навчання / Дацок О.М., Дяченко В.О., Гук А.С., Фатій Л.Р. // Вісник НТУ "ХПІ". Серія: Інформатика та моделювання. – Харків: НТУ "ХПІ". – 2025. – № 2 (14). – С. 144 – 156.

У роботі висвітлено актуальність і перспективи застосування машинного навчання для обробки медичних даних. Розглянуто різноманіття джерел та форматів інформації, а також описано ключові алгоритми машинного навчання. Окрему увагу приділено попередній обробці та підготовці даних, що відіграє вирішальну роль у забезпеченні точності, надійності й етичності побудованих моделей. Результати роботи демонструють важливість мультидисциплінарного підходу, в якому поєднуються зусилля фахівців із медицини, технічних спеціалістів і дослідників із питань конфіденційності, з метою підвищення ефективності діагностики та лікування в сучасній охороні здоров'я. Іл.: 3. Бібліогр.: 13 назв.

Ключові слова: машинне навчання, Big Data, медичні дані, валідація, кластеризація, нормалізація, аномальні значення, енергозбереження.

UDC 004.9:602

Stages of medical data processing for solving problems with machine learning algorithms / Datsok O., Diachenko V., Huk A., Fatii L. // Herald of the National Technical University "KhPI". Series of "Informatics and Modeling". – Kharkov: NTU "KhPI". – 2025. – № 2 (14). – P. 144 – 156.

The paper highlights the relevance and prospects of machine learning for medical data processing. The article discusses the variety of sources and formats of information and describes key machine learning algorithms. Particular attention is paid to data pre-processing and preparation, which plays a crucial role in ensuring the accuracy, reliability, and ethics of the built models. The results of the work demonstrate the importance of a multidisciplinary approach that combines the efforts of medical professionals, technical specialists, and privacy researchers to improve the efficiency of diagnosis and treatment in modern healthcare. Il.: 3. Bibl.: 13 titles.

Keywords: machine learning, Big Data, medical data, validation, clustering, normalization, anomalous values, energy saving.