

С.В. Підпригора, студент
Ю.В. Богоявленська, к.е.н., доц.
Державний університет «Житомирська політехніка»

Автоматизована система прогнозування виявлення цукрового діабету на основі медичних даних пацієнтів

Побудова автоматизованої системи прогнозування виявлення цукрового діабету є надзвичайно актуальним завданням, особливо в умовах стрімкого зростання кількості випадків цього захворювання в Україні та світі. Особливу актуальність проблема набула в Україні внаслідок воєнних дій, що призвели до погіршення загального стану здоров'я населення, збільшення кількості стресових ситуацій та порушення звичного способу життя. Цукровий діабет 2 типу, що становить основну частку захворювань на діабет, часто залишається непоміченим на початкових етапах, що значно ускладнює його подальше лікування і веде до серйозних ускладнень, таких як серцево-судинні захворювання, сліпота, ампутації кінцівок і ниркова недостатність. Рання та точна діагностика дозволяє ефективно управляти перебігом захворювання, значно покращує якість життя пацієнтів та знижує фінансове навантаження на систему охорони здоров'я. У зв'язку з постійним збільшенням обсягів медичних даних виникає необхідність використання сучасних інформаційних технологій, що здатні автоматизувати процеси аналізу та прогнозування захворювання. У запропонованій роботі розглядаються можливості та переваги впровадження алгоритмів машинного навчання (ML) та штучного інтелекту (AI) для аналізу медичних даних з метою виявлення цукрового діабету. Доведено, що застосування таких технологій суттєво підвищує точність прогнозування та діагностики захворювання, дозволяючи своєчасно виявляти пацієнтів групи ризику та забезпечувати необхідну медичну допомогу на ранніх стадіях. Окрім того, такі системи створюють нові можливості для монетизації рішень розробниками, підвищуючи економічну ефективність медичних закладів завдяки оптимізації витрат на лікування. Запропонований підхід демонструє значний потенціал для інтеграції у сучасну медичну практику, роблячи його важливим інструментом для подальшого вдосконалення сфери охорони здоров'я та сприяючи реалізації стратегічних завдань у галузі цифровізації медицини, розвитку превентивних стратегій та підвищення загального рівня громадського здоров'я населення України та інших країн.

Ключові слова: цукровий діабет; машинне навчання; штучний інтелект; рання діагностика; алгоритми AI/ML; хмарні сервіси; обробка даних; моделі монетизації; собівартість розробки.

Актуальність теми. Своєчасне виявлення цукрового діабету на основі медичних даних пацієнтів є надзвичайно актуальним через зростання захворюваності в усьому світі. Рання діагностика дозволяє своєчасно розпочинати лікування, щоб знижувати ризик ускладнень, таких як серцево-судинні захворювання, діабетична нефропатія та ретинопатія [4, 5]. Розробка автоматизованої системи прогнозування виявлення захворювання на основі використання медичних даних, про яку йтиметься у статті, сприятиме точнішому прогнозуванню та індивідуалізації підходів до лікування пацієнтів. Для цілей побудови системи пропонується обрати, зокрема, такі дані як рівень глюкози в крові, показники HbA1c, індекс маси тіла, а також інші показники, що ідентифікуватимуть фактори ризику для пацієнта. Її впровадження як підвищуватиме якість і тривалість активного життя пацієнтів, так і знижуватиме навантаження на систему охорону здоров'я через зниження розміру витрат на лікування завдяки вчасній діагностиці.

Аналіз останніх досліджень та публікацій, на які спираються автори. Вивченням методів автоматизованого прогнозування виявлення цукрового діабету займалися: Еліас Дрічас і Марія Тріпка з Університету Патри (Греція) [1], Гарлін Каур і Вінита Кумарі зі Школи інженерних наук і технологій ім. Джамія Хамдард (Індія) [2], М.Генок Дебернех і Інтаек Кім з Університету Мьонгжі (Республіка Корея) [3] та багато інших дослідників. Водночас в постковідний період не настільки оперативно переглядалося використання показників для цілей побудови автоматизованих систем в контексті зменшення навантаження на медичну систему завдяки побудові нових автоматизованих систем прогнозування.

Метою роботи є створення автоматизованої системи прогнозування виявлення цукрового діабету на ранніх стадіях на основі аналізу значних обсягів даних пацієнтів з допомогою алгоритмів машинного навчання (ML), штучного інтелекту (AI), а також пропонування оптимального алгоритму аналізу даних з прийнятною точністю для цілей мінімізації витрат і зменшення навантаження на медичну систему.

Викладення основного матеріалу. З переліку всіх видів захворювань цукровий діабет є одним з найпоширеніших захворювань обміну речовин, що призводить до руйнування тканин і поступової загибелі пацієнтів. Зазвичай початок діабету 2 типу відбувається в середньому віці та старше, але в останні роки

збільшення випадків появи цього захворювання реєструється у дітей. Існує кілька факторів розвитку діабету, наприклад, генетична схильність, маса тіла, звички харчування та малорухливий спосіб життя. Вчасно не діагностований діабет призводить до дуже високого рівня цукру в крові (гіперглікемією), що зумовлює такі ускладнення, як ретинопатія, нефропатія, нейропатія, серцевий інсульт. Тому вирішення проблеми раннього виявлення цукрового діабету дуже важливе для покращення якості життя пацієнтів та збільшення тривалості їх життя. Для цілей дослідження імплементуватимуться методи машинного навчання та штучного інтелекту для діагностики, скринінгу та прогнозування цукрового діабету 2 типу [1]. Пріоритетним є створення та вдосконалення алгоритмів, що можуть ефективно аналізувати великі обсяги медичних даних та забезпечувати точне передбачення й виявлення діабету на ранніх стадіях.

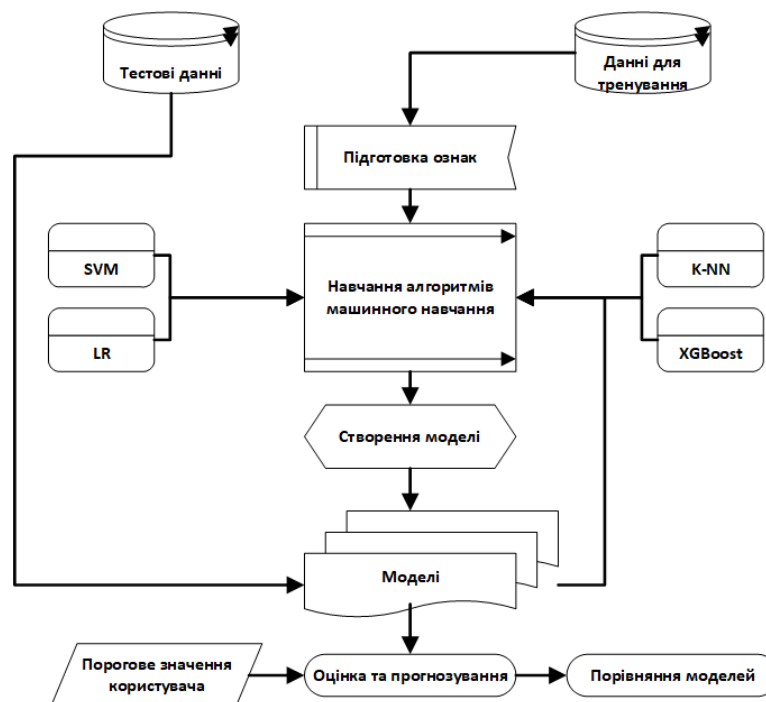
Для побудови системи мають генеруватися дані з результатів аналізу електронних медичних записів (EHR) з різних джерел, таких як національні бази даних, лікарні та медичні центри, з метою побудови моделей для класифікації, прогнозування ризиків та автоматизованого виявлення переддіабетичних станів. У межах цих досліджень застосовуються різні алгоритми машинного навчання, такі як логістична регресія, градієнтний бустинг, метод опорних векторів та інші [1, 2, 3].

Машинне навчання призначене для розробки алгоритмів і методів, що дозволяють комп'ютерам навчатися та здобувати можливість передбачення на основі минулого досвіду [2, 7, 8]. Ця гілка штучного інтелекту тісно пов'язана зі статистикою. Під навчанням мається на увазі те, що система здатна ідентифікувати та розуміти вхідні дані, щоб вона могла приймати рішення та здійснювати передбачення на його основі [9]. З метою побудови автоматизованої системи ML починається зі збору даних різними засобами та ресурсами [10]. Наступний крок – підготовка даних, тобто їх попередня обробка для фіксації проблеми, виявлення пов'язаних даних, і зменшення обсягу шляхом видалення несуттєвих даних (або вибір даних, що цікавлять) [2, 4, 10]. Оскільки обсяг даних, що використовується для навчання, великий, системі важко приймати рішення, тому алгоритми розроблені з використанням логіки, ймовірності, статистики, теорії керування тощо для аналізу даних і отримання знань з минулого досвіду [1, 9].

Наступним кроком є тестування моделі для розрахунку точності та продуктивності системи. І, нарешті, оптимізація системи, тобто імпровізація моделі за допомогою нових правил або набору даних [2, 10]. Методи машинного навчання використовуються для класифікації, передбачення та розпізнавання образів.

Саме тому алгоритми машинного навчання доцільно використовувати для автоматичного аналізу великих обсягів медичних даних та діагностики різних захворювань.

Для цілей побудови автоматизованої системи прогнозування доцільно використовувати блок-схему оцінки роботи алгоритму прогнозування виявлення цукрового діабету на основі медичних даних пацієнтів (рис. 1) [1].



Джерело: удосконалено автором в частині використання ML-алгоритмів на основі [1]

Рис. 1. Блок-схема для оцінки роботи алгоритму прогнозування виявлення цукрового діабету на основі медичних даних пацієнтів

Цільовою аудиторією, на яку направлено результати використання автоматизованої системи, є медичні установи, лікарі, страхові компанії, пацієнти, державні органи охорони здоров'я, фармацевтичні компанії та фітнес-технологічні компанії. Медичні заклади та лікарі можуть використовувати систему для підвищення точності діагностики та персоналізованого лікування. Страхові компанії зможуть оцінювати ризики клієнтів і коригувати страхові плани. Пацієнти отримають інструмент для моніторингу свого здоров'я і профілактики діабету. Фармацевтичні компанії та урядові організації отримають доступ до даних для досліджень та оптимізації медичних витрат.

Витрати, пов'язані зі створенням (собівартість) та просуванням (збут) системи для прогнозування виявлення цукрового діабету, залежно від її рівня складності, можуть складатися з:

1. Витрат на розробку програмного забезпечення: заробітна плата програмістів, фахівців з машинного навчання (AI/ML), data-фахівців і data-інженерів та ін. Робота зі збору, очищення та підготовки даних для навчання моделей потребує пулу часу та інших видів ресурсів. Якщо використовується платна база даних або сторонні набори даних, це враховується. Приклад розрахунку:

$$B_p = K_p \times \$3000/\text{міс.} \times 3 \text{ місяці}, \quad (1)$$

де B_p – вартість розробки, K_p – кількість розробників.

Тестування програмного забезпечення, враховуючи автоматизоване тестування, ручне тестування та A/B-тестування:

$$B_m = K_m \times \$1500/\text{міс.} \times 3 \text{ місяці}, \quad (2)$$

де B_m – вартість тестування, K_m – кількість тестувальників.

Основні фахівці, які можуть бути потрібні для успішного виконання проєкту з розробки і підтримання системи:

1. Data Scientist (фахівець з даних) – 2–3 спеціаліста;
2. Backend Developer (розробник бекенду) – 1–2 спеціаліста;
3. Frontend Developer (розробник фронтенду) – 1 спеціаліст;
4. DevOps Engineer (фахівець з інфраструктури та розгортання) – 1 спеціаліст;
5. Machine Learning Engineer (інженер з машинного навчання) – 1–2 спеціаліста;
6. Database Administrator (адміністратор баз даних) – 1 спеціаліст;
7. UI/UX Designer (дизайнер інтерфейсів / користувачького досвіду) – 1 спеціаліст;
8. QA Engineer (тестувальник) – 2 спеціаліста;
9. Project Manager (менеджер проєкту) – 1 спеціаліст.

Для повної реалізації проєкту команда формуватиметься з 11–14 розробників та спеціалістів з різних галузей. З урахуванням того, що на розробку знадобиться 3–4 місяці, вартість робіт буде становити приблизно \$150 000,00;

2. Хостинг та інфраструктуру: вартість серверів, хмарних рішень (наприклад, AWS, Google Cloud) для розгортання моделей машинного навчання та обробки великих обсягів даних – це витрати на:

- хмарні сервери для роботи сервісу, зберігання даних, оновлення та підтримка безпеки – \$1500,00 / міс.;
- заробітна плата технічних фахівців, які займатимуться підтримкою продукту (гарячі лінії, служби підтримки клієнтів) – \$1000,00 / міс.

Тобто вартість витрати на інфраструктуру \$30000,00 / рік;

3. Витрати на просування (збут) продукту будуть складатись з витрат на:

- аналіз цільового ринку (лікарі, медичні установи, страхові компанії, пацієнти);
- рекламу для кожного сегменту, у т. ч. у соцмережах (Facebook, Google Ads);
- SEO-оптимізацію сайту, контент-маркетинг (статті, вебінари, блоги);
- канали розподілу, дилерські та інші мережі, якот на співпрацю з медичними організаціями, страховими компаніями та клініками, що можуть використовувати або пропонувати сервіс;
- промоушен через участь у медичних конференціях, наукових симпозіумах та виставках, щоб залучити потенційних клієнтів і партнерів.

Вартість витрат на просування (збут) \$20000.00 / рік.

До складу витрат потрібно врахувати і витрати на сертифікацію програмного забезпечення у відповідних регулюючих органах (наприклад, FDA, CE для Європи), що може вимагати часу та значних коштів. Також витрати на юридичні консультації щодо захисту даних (GDPR, HIPAA) та захисту інтелектуальної власності (патенти на методи діагностики, алгоритми) – \$10000,00 / рік.

Інші можливі витрати можуть складати:

- на навчання медичних працівників чи технічних команд у лікарнях та клініках для інтеграції рішення в їхню практику;
- на інтеграцію сервісу з іншими медичними системами (електронні медичні записи, телемедицина тощо);
- залежно від обсягу фінансових операцій і географічного охоплення сервісу, річні витрати на банківське обслуговування можуть бути близько 3–5 % від суми транзакцій [6].

Ці витрати можуть складати близько \$20000,00 / рік.

Отже, розрахунок витрат на створення та просування сервісу для прогнозування виявлення цукрового діабету та перший рік експлуатації цієї системи буде становити приблизно \$230000,00 – \$250000,00.

Мінімальна кількість клієнтів для виходу на точку беззбитковості залежить від ціни. Якщо, наприклад, вартість підписки або послуги становить \$1000,00 на рік, тоді 250 клієнтів достатньо для покриття всіх витрат, що не так багато, навіть для локального ринку України.

Технології штучного інтелекту (AI) та машинного навчання (ML), застосовані для діагностики та прогнозування цукрового діабету, мають великий потенціал для монетизації. Варіанти *моделей монетизації* (рішень), які можуть бути використані для генерації прибутку від таких рішень залежно від поставлених цілей:

1. Створення спеціалізованого програмного забезпечення для медичних установ, яке використовує AI/ML для автоматизації діагностики та прогнозування захворювань, – одне з найпоширеніших рішень. Медичні заклади можуть купувати ліцензії для використання цього ПЗ або підписуватися на SaaS-моделі (програмне забезпечення як послуга), що забезпечує постійний дохід для розробників;

2. Передплата на використання хмарних платформ, що надають послуги з аналізу медичних даних на основі AI. Такі рішення можуть бути доступні як для великих лікарень, так і для приватних клінік, що прагнуть отримати доступ до новітніх технологій діагностики без необхідності розгортати власну інфраструктуру;

3. AI-системи, інтегровані в платформи EHR для автоматизованого аналізу даних пацієнтів і прогнозування діабету. Розробники таких рішень отримують дохід від угод з постачальниками EHR-систем або стягнення плати за кожен оброблений запит або обчислення прогнозу для пацієнта;

4. AI-інструменти, що здатні прогнозувати розвиток діабету, цінні для фармацевтичних компаній та страхових компаній. Зокрема, фармацевтичним компаніям доцільно використовувати такі технології для вибору цільових груп пацієнтів для клінічних випробувань, за згодою, або для просування своїх продуктів; страховим компаніям – для оцінки ризиків пацієнтів і розробки індивідуальних страхових планів, ґрунтуючись на прогнозах AI;

5. AI/ML-алгоритми, у поєднанні з телемедициною, можуть використовуватися для онлайн-діагностики діабету або переддіабетичних станів. Платформа може стягувати плату за кожну консультацію або використовувати модель передплати для надання послуг пацієнтам на постійній основі. Це дозволяє пацієнтам отримувати своєчасну допомогу та рекомендації віддалено;

6. Розробники можуть укласти партнерства з медичними та фармацевтичними компаніями для реклами своїх продуктів або послуг на платформах, що використовують AI для діагностики діабету. Це може бути модель спільного використання доходів або продаж рекламних місць на платформі.

Отже, використання алгоритмів машинного навчання (ML) та штучного інтелекту (AI) для аналізу медичних записів дозволить підвищувати точність діагностики, що робитиме такі системи важливими інструментами в боротьбі з діабетом, а обрані моделі монетизації забезпечать прибуток розробникам, зменшивши навантаження на бюджети і медичну систему.

Висновки та перспективи подальших досліджень. За результатами проведеного дослідження можна зробити висновки, що створення автоматизованої системи прогнозування виявлення цукрового діабету на ранніх стадіях на основі аналізу значних обсягів даних пацієнтів з допомогою алгоритмів машинного навчання, штучного інтелекту є доцільним, фінансово й економічно обґрунтованим, а запропонований алгоритм аналізу даних з прийнятною точністю сприяє мінімізації витрат і зменшенню навантаження на медичну систему.

Подальші дослідження в галузі машинного навчання і штучного інтелекту для діагностики цукрового діабету мають значний потенціал для покращення медичної практики. Перспективи подальших досліджень:

- більш точне прогнозування, глибша індивідуалізація лікування та зменшення навантаження на медичну систему;

- створення персоналізованих моделей, які враховуватимуть індивідуальні особливості пацієнтів – генетичні фактори, спосіб життя та супутні захворювання, що забезпечить ефективніше виявлення захворювання на ранніх стадіях і планування лікування;

- використання нових алгоритмів та методів, таких як глибоке навчання та ансамблеві моделі, може підвищити точність діагностики та прогнозування, що важливо для зменшення кількості хибнопозитивних та хибнонегативних діагнозів; також перспективним є створення інструментів, доступних для використання в регіонах із обмеженим доступом до медичних послуг;

- подальша інтеграція різних джерел даних, таких як медичні записи, дані з сенсорів (наприклад, фітнес-трекери), геномні дані та інші, що дозволить побудувати більш повну картину здоров'я пацієнта та покращити прогнози.

Удосконалення методів AI/ML сприятиме автоматизації рутинних завдань у медицині, таких як скринінг пацієнтів та моніторинг стану здоров'я. Це дозволить лікарям більше зосереджуватись на

складних випадках, а пацієнтам – отримувати своєчасні рекомендації щодо профілактики або лікування діабету. Глибше використання AI/ML дозволить підвищити якість медичних послуг, забезпечуючи монетизацію розробникам, і зменшити кількість ускладнень від цього захворювання.

Список використаної літератури:

1. *Dritsas E.* Data-Driven Machine-Learning Methods for Diabetes Risk Prediction / *E.Dritsas, M.Trigka* // *Sensors*. – 2022. – № 22, 5304. DOI: 10.3390/s22145304.
2. *Kaur H.* Predictive modelling and analytics for diabetes using a machine learning approach / *H.Kaur, V.Kumari*. – New Delhi, India, Department of Computer Science and Engineering, School of Engineering Sciences and Technology, Jamia Hamdard, 2018 [Electronic resource]. – Access mode : <https://www.emerald.com/insight/content>.
3. *Deberne H.M.* Prediction of Type 2 Diabetes Based on Machine Learning Algorithm / *H.M. Deberne, I.Kim* // *Int. J. Environ. Res. Public Health*. – 2021. – № 18, 3317. DOI: 10.3390/ijerph18063317.
4. *Dr. Varma S.* Diabetes Prediction using Machine Learning Techniques / *S.Dr. Varma, S.Mitushi* // *International Journal of Engineering Research & Technology (IJERT)*. – 2020 [Electronic resource]. – Access mode : <http://www.ijert.org>.
5. [Electronic resource]. – Access mode : https://uk.wikipedia.org/wiki/%D0%A6%D1%83%D0%BA%D1%80%D0%BE%D0%B2%D0%B8%D0%B9_%D0%B4%D1%96%D0%B0%D0%B1%D0%B5%D1%82.
6. *Богоявленська Ю.В.* Проектний аналіз : Навчальний посібник / *Ю.В. Богоявленська*. – К. : Кондор, 2004.
7. [Electronic resource]. – Access mode : <https://bimedis.net/latest-news/browse/778/suchasni-tehnologiyi-u-medicini-shtuchnij-intelekt>.
8. [Electronic resource]. – Access mode : <https://uk.shaip.com/blog/real-world-applications-of-machine-learning-in-healthcare/>.
9. [Electronic resource]. – Access mode : <https://www.it.ua/knowledge-base/technology-innovation/machine-learning>.
10. [Electronic resource]. – Access mode : <https://kharkov-future.com.ua/uk/articles-yak-rozroblyayetsya-mashynne-navchannya-klyuchovi-etapy>.

References:

1. Dritsas, E. and Trigka, M. (2022), «Data-Driven Machine-Learning Methods for Diabetes Risk Prediction», *Sensors*, No. 22, 5304, doi: 10.3390/s22145304.
2. Kaur, N. and Kumari, V. (2018), «Predictive modelling and analytics for diabetes using a machine learning approach», Department of Computer Science and Engineering, School of Engineering Sciences and Technology, Jamia Hamdard, New Delhi, India, [Online], available at: <https://www.emerald.com/insight/content/>
3. Deberne, H.M. and Kim, I. (2021), «Prediction of Type 2 Diabetes Based on Machine Learning Algorithm», *Int. J. Environ. Res. Public Health*, No. 18, 3317, doi: 10.3390/ijerph18063317.
4. Dr. Varma, S. and Mitushi, S. (2020), «Diabetes Prediction using Machine Learning Techniques», *International Journal of Engineering Research & Technology (IJERT)*, [Online], available at: <http://www.ijert.org>
5. [Online], available at: https://uk.wikipedia.org/wiki/%D0%A6%D1%83%D0%BA%D1%80%D0%BE%D0%B2%D0%B8%D0%B9_%D0%B4%D1%96%D0%B0%D0%B1%D0%B5%D1%82
6. Bohoiavlenska, Yu.V. (2004), *Proektnyi analiz*, Navchalnyi posibnyk, Kondor, K.
7. [Online], available at: <https://bimedis.net/latest-news/browse/778/suchasni-tehnologiyi-u-medicini-shtuchnij-intelekt>
8. [Online], available at: <https://uk.shaip.com/blog/real-world-applications-of-machine-learning-in-healthcare/>
9. [Online], available at: <https://www.it.ua/knowledge-base/technology-innovation/machine-learning>
10. [Online], available at: <https://kharkov-future.com.ua/uk/articles-yak-rozroblyayetsya-mashynne-navchannya-klyuchovi-etapy>

Підпригора Сергій Вікторович – магістрант Державного університету «Житомирська політехніка», гр. АТ-28м.

Наукові інтереси:

- машинне навчання (ML);
- штучний інтелект (AI);
- робототехніка;
- обробка великих даних (Big Data);
- розробка й оптимізація систем обліку та логістики.

Богоявленська Юлія Вячеславівна – кандидат економічних наук, доцент, докторант кафедри фінансів та цифрової економіки Державного університету «Житомирська політехніка».

<https://orcid.org/0000-0003-4101-7127>.

Наукові інтереси:

- проектний менеджмент;
- міжнародні наукові, інноваційні, експертні проекти;
- цифрова економіка;
- фінанси.

Pidopryhora S.V., Bogoyavlenska Yu.V.

Automated diabetes detection prediction system based on patients' medical data

The development of an automated system for predicting diabetes detection is an extremely relevant task, especially given the rapid increase in the number of diabetes cases in Ukraine and worldwide. The issue has become particularly burning in Ukraine due to military actions, which have deteriorated the overall health condition of the population, increased stress levels, and disrupted the normal way of life. Type 2 diabetes, which accounts for the majority of diabetes cases, often remains unnoticed in the early stages, significantly complicating subsequent treatment and leading to severe complications such as cardiovascular diseases, blindness, limb amputations, and kidney failure. Early and accurate diagnosis allows effective disease management, significantly improves patients' quality of life, and reduces the financial burden on the healthcare system. Given the continuous growth of medical data volumes, there is a clear need for modern information technologies capable of automating disease analysis and prediction processes. This paper examines the potential and benefits of implementing machine learning (ML) and artificial intelligence (AI) algorithms for medical data analysis aimed at diabetes detection. It has been proven that the application of such technologies significantly enhances the accuracy of disease prediction and diagnosis, enabling timely identification of at-risk patients and providing necessary medical care at early stages. Furthermore, these systems create new opportunities for developers to monetize solutions, improving the economic efficiency of medical institutions by optimizing treatment costs. The proposed approach demonstrates substantial potential for integration into modern medical practice, making it an essential tool for further improving the healthcare sector. It contributes to the strategic objectives of medicine digitalization, the development of preventive strategies, and the enhancement of overall public health levels in Ukraine and other countries.

Keywords: diabetes mellitus; machine learning; artificial intelligence; early diagnosis; ML/AI algorithms; cloud services; data processing; models of monetization; development costs.

Стаття надійшла до редакції 15.04.2025.