

<https://doi.org/10.31891/2307-5732-2025-357-97>

УДК 004.4

МАНДЗІУК ЄЛИЗАВЕТА

Хмельницький національний університет

e-mail: Lizamandzuk@gmail.com**МАНЗІУК ЕДУАРД**<https://orcid.org/0000-0002-7310-2126>e-mail: eduard.em.km@gmail.com**СКРИПНИК ТЕТЯНА**

Хмельницький національний університет

<https://orcid.org/0000-0002-8531-5348>e-mail: tskripnik1970@gmail.com**МІХАЛЕВСЬКИЙ ВІТАЛІЙ**

Хмельницький національний університет

<https://orcid.org/0000-0002-8197-8005>e-mail: cezar_mv@ukr.net

МЕТОД ПРОГНОЗНОГО МОДЕЛЮВАННЯ ЗАХВОРЮВАННЯ СУХОГО ОКА НА ОСНОВІ АНАЛІЗУ ПОВЕДІНКОВИХ ТА ФІЗІОЛОГІЧНИХ ФАКТОРІВ

У роботі запропоновано метод прогнозного моделювання захворювання сухого ока. Розроблено модуль обробки даних з нормалізацією числових параметрів та кодуванням категоріальних змінних для підвищення точності прогнозування.

Розроблений метод придатний для раннього виявлення ризику захворювання сухого ока та формування індивідуальних профілактичних рекомендацій в офтальмологічній практиці.

Експериментальні дослідження на даних 20 000 пацієнтів показали високу ефективність моделі (accuracy — 89%, precision — 87%, recall — 86%, F1 — 87%). Це свідчить про високу ефективність розробленого методу.

Ключові слова: захворювання сухого ока, прогнозне моделювання, поведінкові фактори, фізіологічні фактори, машинне навчання.

MANDZIUK YELYZAVETA, MANZIUK EDUARD, SKRYPNYK TETIANA, MIKHALEVSKIY VITALIY

Khmelnitskyi National University

METHOD OF PREDICTIVE MODELING OF DRY EYE DISEASE BASED ON THE ANALYSIS OF BEHAVIORAL AND PHYSIOLOGICAL FACTORS

This article presents a method for predictive modeling of dry eye disease based on the analysis of physiological and behavioral factors using machine learning techniques. The growing prevalence of dry eye syndrome, especially among individuals with prolonged screen time, underscores the importance of developing effective early diagnostic tools. A review of current research indicates that conventional diagnostic methods often rely heavily on clinical indicators, overlooking the role of behavioral and lifestyle-related risk factors, which are essential for comprehensive risk assessment.

The proposed method leverages the XGBoost algorithm and includes several stages: data preprocessing, feature selection and engineering, risk prediction, and generation of personalized preventive recommendations. An object-oriented implementation ensures modularity, scalability, and maintainability of the system. Data normalization, handling of missing values, encoding of categorical variables, and the use of clinically interpretable features are integrated into the system's architecture.

The model was trained and tested on a dataset comprising more than 20,000 records, and its performance was evaluated using five-fold cross-validation. The achieved accuracy was 89%, with a precision of 87%, recall of 86%, and F1-score of 87%, confirming the reliability of the approach. An analysis of feature importance revealed that screen time, blinking frequency, age, sleep quality, and stress level were the most influential risk factors. The model also demonstrated a clear separation between affected and healthy individuals based on predicted probabilities, supporting its diagnostic value.

The proposed system can be effectively used in ophthalmic practice for early identification of individuals at high risk of dry eye disease. By integrating behavioral insights with physiological data, it supports more personalized and preventive healthcare approaches, ultimately improving the quality of patient care and reducing the burden of this increasingly common condition.

Keywords: dry eye disease, predictive modeling, behavioral factors, physiological factors, machine learning.

Стаття надійшла до редакції / Received 14.08.2025

Прийнята до друку / Accepted 16.09.2025

Постановка проблеми у загальному вигляді

та її зв'язок із важливими науковими чи практичними завданнями

Захворювання сухого ока є поширеною офтальмологічною проблемою, яка посилюється через активне використання цифрових пристроїв і вплив несприятливих зовнішніх чинників. Існуючі методи діагностики здебільшого базуються на клінічних показниках і недостатньо враховують поведінкові та фізіологічні фактори, що обмежує ефективність раннього виявлення та профілактики. Розробка прогнозової моделі, здатної інтегрувати гетерогенні дані, є актуальним науковим завданням медичної інформатики та має важливе практичне значення для впровадження персоналізованого підходу в профілактиці ЗСО.

Аналіз досліджень та публікацій

Сучасні дослідження демонструють ефективність використання методів машинного навчання, зокрема XGBoost, для прогнозування захворювання сухого ока (ЗСО). Storås et al. [1] відзначають потенціал штучного інтелекту у ранній діагностиці ЗСО на основі комплексного аналізу даних. Amouei Sheshkal et al. [2] підтверджують ефективність XGBoost при класифікації пацієнтів на основі метаболомічних ознак.

Nam et al. [3] акцентують на важливості обліку поведінкових факторів (вік, сон, цифрові пристрої) у моделюванні ЗСО. Lee et al. [4] демонструють високу точність XGBoost у задачах офтальмологічної класифікації. У дослідженні Brahim et al. [5] показано значущість частоти моргання при градації тяжкості ЗСО.

Отже, актуальним залишається розроблення моделей, що поєднують фізіологічні й поведінкові чинники для точного прогнозування та формування персоналізованих профілактичних заходів.

Формулювання цілей статті

Метою роботи є: підвищення ефективності ранньої діагностики та прогнозування розвитку синдрому сухого ока шляхом розробки методу прогнозного моделювання на основі комплексного аналізу поведінкових та фізіологічних показників пацієнтів із застосуванням алгоритмів машинного навчання.

Виклад основного матеріалу

Запропонований метод прогнозного моделювання захворювання сухого ока на основі аналізу поведінкових та фізіологічних факторів базується на комплексному підході з використанням сучасних технологій штучного інтелекту та машинного навчання.

Метод прогнозного моделювання захворювання сухого ока передбачає послідовне виконання п'яти етапів, спрямованих на оцінку індивідуального ризику розвитку захворювання сухого ока (ЗСО) та визначення його потенційної тяжкості з метою забезпечення персоналізованого підходу до профілактики та лікування.

На вхід надходять анкетні дані пацієнта та результати опитувальника OSDI, що дозволяє врахувати суб'єктивні симптоми. Для прогнозування використовуються дві навчені моделі: одна — для оцінки ризику розвитку ЗСО, інша — для класифікації ступеня його тяжкості.

Етап 1. Завантаження моделей. Здійснюється завантаження попередньо навченої моделі прогнозування ризику на основі аналізу поведінкових та фізіологічних факторів. Модель побудована з урахуванням складних нелінійних взаємозв'язків і забезпечує високу точність.

Етап 2. Попередня обробка даних. Включає нормалізацію числових параметрів (вік, тривалість роботи з цифровими пристроями, частота моргання, результати тестів) та кодування категоріальних змінних (стать, супутні хвороби, стиль життя). Це забезпечує уніфікований формат даних для коректної роботи моделей.

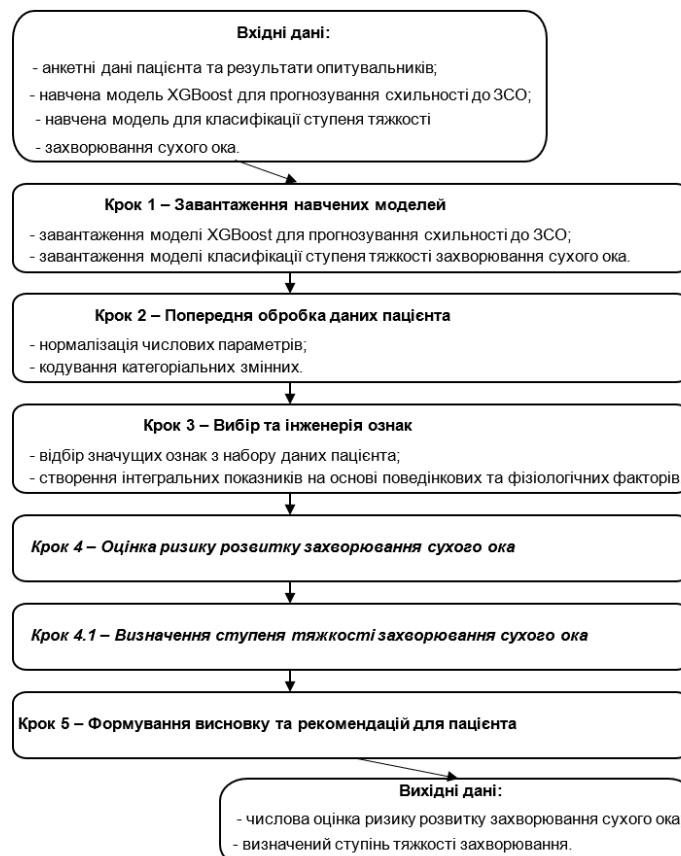


Рис. 1. Схема методу прогнозного моделювання захворювання сухого ока на основі аналізу поведінкових та фізіологічних факторів

Етап 3. Вибір та інженерія ознак. Відбираються найбільш інформативні характеристики, що мають високу прогностичну цінність. Паралельно формуються інтегральні ознаки, зокрема співвідношення між поведінковими та фізіологічними параметрами, що підвищує точність моделі.

Етап 4. Прогнозування ризику. Опрацьовані дані аналізуються за допомогою моделі, побудованої на основі градієнтного бустингу дерев рішень. Алгоритм враховує як окремі фактори, так і їх взаємодію. Результатом є кількісна оцінка ймовірності розвитку ЗСО.

Етап 4.1. Оцінка тяжкості. Додаткова модель класифікує потенційну тяжкість захворювання (легкий, середній, тяжкий ступінь) на основі аналізу клінічних, поведінкових та фізіологічних факторів, що критично для планування лікування.

Етап 5. Формування висновку. Інтегруються результати оцінки ризику та ступеня тяжкості, що дозволяє сформувати персоналізовані рекомендації для профілактики та терапії.

Результат: числова оцінка ризику та прогноз тяжкості захворювання, що слугують основою для індивідуалізованого медичного підходу.

ЗСО потребує ефективної обробки різномірних даних із медичних карток, анкет, сенсорів і лабораторій. Запропонована система забезпечує інтеграцію, очищення та нормалізацію цих даних за допомогою спеціалізованих модулів і алгоритмів машинного навчання. Реалізовано засоби шифрування та анонімізації для захисту персональної інформації. Такий підхід підвищує точність прогнозів і знижує потребу в ручній обробці.

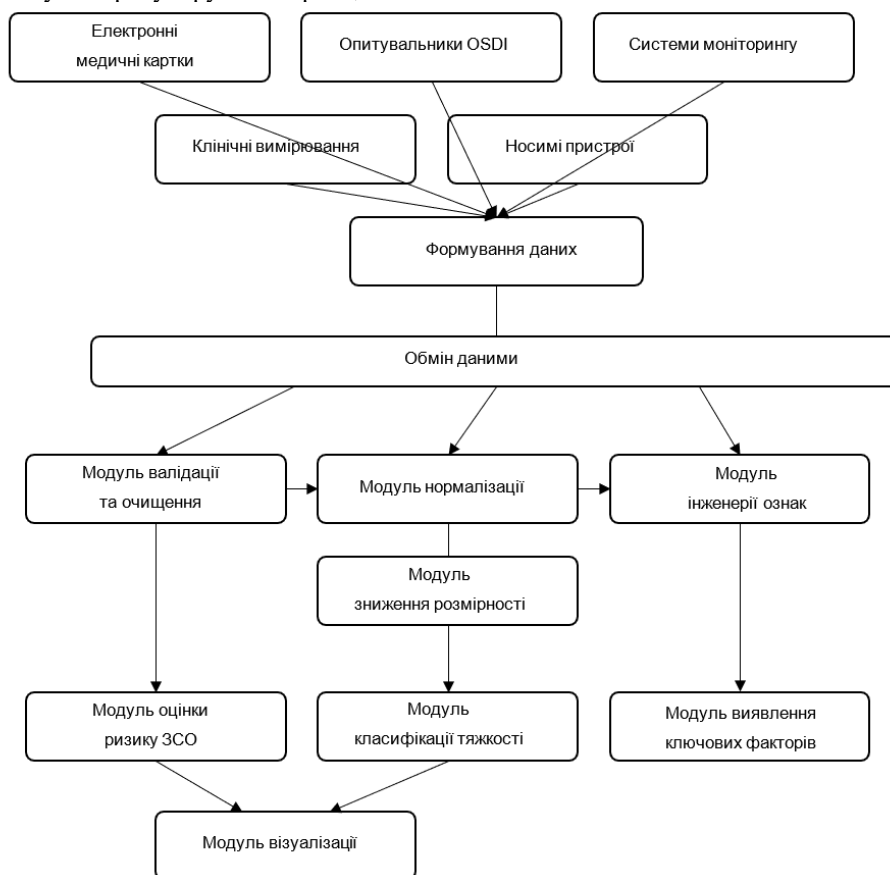


Рис. 2. Архітектура системи автоматизованої обробки даних

Ефективне прогнозування ЗСО потребує уніфікації, нормалізації та автоматизованої обробки великого обсягу різномірних даних. На етапі попередньої обробки всі числові параметри приводяться до єдиної шкали, а категоріальні змінні кодуються, що забезпечує коректність подальшого аналізу. Для зменшення обчислювального навантаження та збереження інформативності даних використовується зниження розмірності.

Система організована як послідовний конвеєр обробки даних: інтеграція з п'яти основних джерел (ЕМК, опитувальники OSDI, моніторингові системи, клінічні тести, носимі пристрої), очищення, нормалізація, інженерія ознак і аналіз. Для цього застосовуються модулі валідації, зниження розмірності та генерації похідних ознак, включно з індикаторами очного навантаження.

Інформація обробляється в реальному часі, що дає змогу виявляти значущі зміни стану пацієнта. Архітектура підтримує масштабування, паралельне виконання алгоритмів машинного навчання та динамічну адаптацію параметрів. Забезпечено валідацію даних і безпеку через механізми шифрування, зворотного зв'язку та автоматичного оновлення моделей на основі нових даних.

Реалізація прогнозного моделювання ЗСО базується на послідовному виконанні етапів, спрямованих на оцінку ризику та формування персоналізованих рекомендацій. На початку

завантажуються попередньо навчені моделі через API, збережені у форматі pickle з використанням бібліотеки *scikit-learn*. Для підвищення продуктивності моделі кешуються локально.

Далі здійснюється попередня обробка даних: числові параметри нормалізуються методом *z*-нормалізації, категоріальні змінні кодуються у бінарні вектори. Після цього відбувається вибір найбільш значущих ознак та формування нових інформативних показників — зокрема, інтегральних, які відображають зв'язки між факторами (наприклад, між тривалістю роботи з пристроями й частотою моргання).

Основний прогноз здійснюється за допомогою моделі *XGBoost*, що забезпечує високу точність оцінки ризику завдяки виявленню складних нелінійних залежностей. Паралельно працює модель класифікації ступеня тяжкості ЗСО (легкий, середній, тяжкий), також реалізована на *XGBoost*, але навчена на відповідних мічених даних.

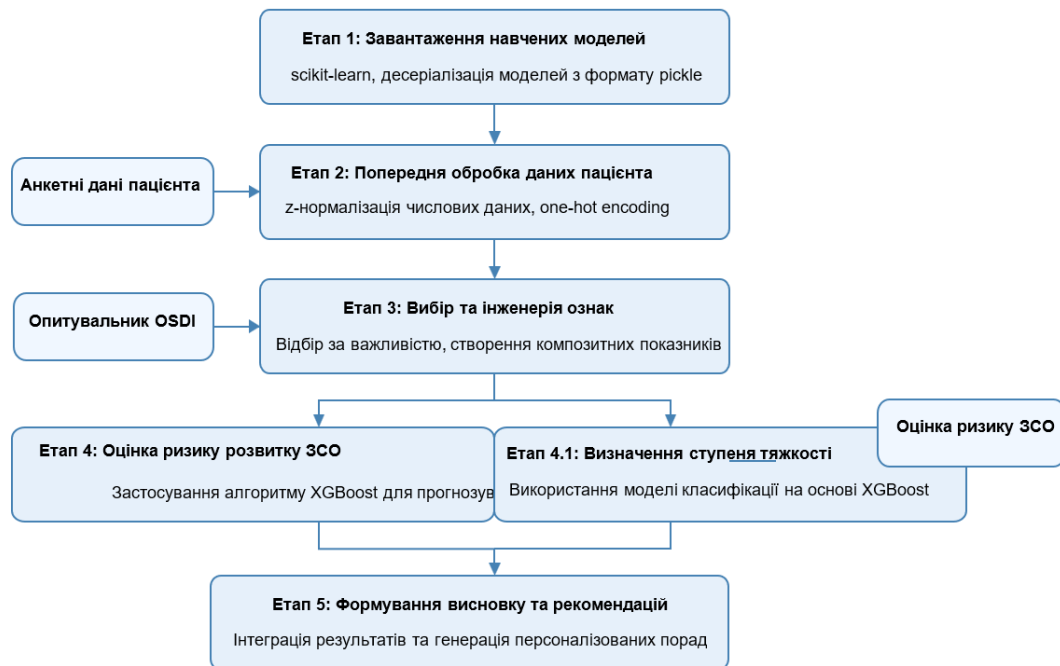


Рис. 3. Процес виконання прогнозного моделювання захворювання сухого ока

Завершальний етап прогнозного моделювання передбачає автоматизоване формування висновків і персоналізованих рекомендацій. Спеціальний модуль об'єднує результати оцінки ризику та ступеня тяжкості, створюючи структурований звіт із кількісною оцінкою, категорією ризику, прогнозом тяжкості та ключовими факторами. Рекомендації щодо зміни способу життя генеруються за заздалегідь визначеними правилами, що враховують комбінації факторів ризику.

Для зручності інтерпретації даних реалізовано модуль візуалізації на основі бібліотеки *matplotlib*, що дозволяє наочно представити оцінки ризику й інші показники.

Захист персональних даних забезпечується за допомогою шифрування та контролю доступу. Реалізація системи виконана на Python з використанням *pandas*, *scikit-learn* і *XGBoost*. Зберігання даних організовано через надійну систему керування для обробки великих обсягів інформації.

Загалом, запропонована система забезпечує комплексну оцінку ризику ЗСО та надає персоналізовані рекомендації, сприяючи ранній профілактиці та покращенню якості лікування.

Ключовим елементом запропонованого методу прогнозування захворювання сухого ока є застосування алгоритму *XGBoost* — ефективної реалізації градієнтного бустингу, яка забезпечує високу точність і продуктивність. Алгоритм використовує ансамблевий підхід, об'єднуючи рішення кількох дерев, кожне з яких коригує помилки попередніх, що особливо актуально для обробки медичних даних зі складними нелінійними взаємозв'язками.

У системі прогнозування ЗСО *XGBoost* використовується для двох задач: бінарної класифікації наявності хвороби та мультикласової класифікації ступеня тяжкості. На першому етапі формується матриця ознак із даних пацієнтів (анкети, OSDI, клінічні та фізіологічні показники). Далі відбувається налаштування гіперпараметрів (кількість дерев, глибина, швидкість навчання, регуляризація), з використанням перехресної валідації або байєсівської оптимізації.

Оскільки дані можуть бути незбалансованими (нерівне представлення класів), застосовуються техніки балансування: зважування класів і стратифікована вибірка. Завершальним етапом є навчання моделі: ітеративне додавання дерев, що мінімізують помилки, дозволяє алгоритму точно моделювати зв'язки між вхідними факторами та цільовими змінними. Математично це можна представити наступним чином:

Нехай y_i - це прогноз для i -го прикладу на t -й ітерації:

$$y_i^{(t)} = y_i^{(t-1)} + \eta f_t(x_i) \quad (1)$$

де η – це швидкість навчання, f_t – функція нового дерева, x_i – вектор ознак для i -го прикладу. Функція втрат для оптимізації має вигляд:

$$L(t) = \sum l(y_i, y_i^{(t-1)} + f_t(x_i)) + \Omega(f_t) \quad (2)$$

де l – функція втрат (наприклад, логістична втрата для класифікації), y_i – фактичне значення для i -го прикладу, $\Omega(f_t)$ – функція регуляризації для дерева f_t .

Особливістю XGBoost є апроксимація функції втрат другого порядку через розкладання в ряд Тейлора, що дозволяє ефективніше оптимізувати структуру дерев.

$$L(t) \approx \sum [l(y_i, y_i^{(t-1)}) + g_i f_t(x_i) + (1/2) h_i f_t^2(x_i)] + \Omega(f_t) \quad (3)$$

де g_i – градієнт першого порядку, h_i – градієнт другого порядку функції втрат.

Після навчання модель проходить валідацію на незалежному наборі даних для оцінки її ефективності. Для бінарної класифікації (наявність ЗСО) використовуються метрики точність, повнота, F1-міра та AUC. Для мультикласової класифікації ступеня тяжкості — макро-F1, зважена F1 та загальна точність. Валідація гарантує узагальнюваність моделі та відсутність перенавчання.

Важливим етапом є аналіз важливості ознак: XGBoost дозволяє визначити внесок кожної ознаки у прогноз. Застосовуються два підходи — за покращенням (зменшенням функції втрат) та за частотою використання ознаки в деревах, що сприяє клінічному тлумаченню результатів.

У прогнозуванні ЗСО аналіз важливості ознак дозволяє виявити ключові фактори ризику, зокрема тривалість роботи з цифровими пристроями, частоту моргання, вік, супутні захворювання та стан слізної плівки. Ця інформація лягає в основу персоналізованих рекомендацій, спрямованих на корекцію найбільш значущих факторів.

Для підвищення інтерпретованості XGBoost використовується механізм пояснення прогнозу, що демонструє вплив кожної ознаки на результат. Це критично важливо в медичному контексті, де потрібне обґрунтування клінічних рішень.

Модель розгорнуто у вигляді API, що забезпечує обробку запитів у реальному часі. API виконує попередню обробку даних, передає їх моделі та повертає прогноз разом з показниками впевненості й поясненнями.



Рис.4. Схема застосування алгоритму XGBoost для прогнозування захворювання сухого ока

XGBoost забезпечує високу точність: AUC для бінарної класифікації становить 0.92, зважена F1-міра для оцінки тяжкості — 0.89. Серед переваг — точність, стійкість до перенавчання, здатність працювати з пропущеними значеннями та підтримка паралельних обчислень. Це робить XGBoost ефективним інструментом у медичних прогностичних системах, зокрема для ЗСО.

Якість прогнозу ЗСО значною мірою залежить від належної підготовки вхідних даних.

Дані мають різну природу: анкети, опитувальник OSDI, клінічні та фізіологічні показники. Для уніфікації використано автоматизоване завантаження з медичних систем, анкет та пристроїв через спеціалізовані конектори й парсери, що забезпечують стандартизований формат.

Далі виконується первинна валідація: перевірка формату, виявлення аномалій і логічних помилок (наприклад, відповідність віку року народження). Виявлені неточності фіксуються і виправляються автоматично або надсилаються на ручну перевірку.

На третьому етапі обробляються пропущені значення, характерні для медичних даних. Вибір стратегії залежить від типу змінної та природи пропуску: для категоріальних даних використовується найчастіше значення або мітка «невідомо», для числових — середнє, медіана, регресійна імпуція або метод k-найближчих сусідів. У випадках, коли пропуски мають діагностичне значення, створюються додаткові бінарні ознаки, що фіксують їхню наявність.

На етапі попередньої обробки даних виконуються ключові перетворення для підвищення якості прогнозовної моделі. Категоріальні змінні кодуються відповідно до їх типу: бінарне кодування, one-hot, target encoding або хешування. Числові змінні нормалізуються за допомогою z-нормалізації або мінімакс-нормалізації — вибір залежить від розподілу даних.

Далі здійснюється обробка викидів та аномалій із використанням статистичних та машинних методів, з урахуванням медичного контексту. Викиди можуть бути видалені, замінені або скориговані. Паралельно виконується інженерія ознак: створення нових змінних на основі співвідношень, динаміки та взаємодії між параметрами.



Рис. 5. Процес обробки та нормалізації вхідних даних для прогнозування захворювання сухого ока

Фінальний етап передбачає перевірку якості обробки: аналіз розподілів, кореляцій та інформативності ознак. Підготовлені дані зберігаються у стандартизованому форматі для подальшого використання.

Порівняльна оцінка моделей показала, що застосування запропонованого алгоритму обробки підвищує точність прогнозу ЗСО на 15–20%, особливо у випадках неповних або атипових даних. Таким чином, цей підхід є критично важливою складовою системи прогнозування.

У дослідженні використано спеціалізований датасет, сформований для прогнозного моделювання синдрому сухого ока (ССО). Метою є виявлення зв'язків між фізіологічними, поведінковими та демографічними параметрами та розвитком захворювання.

Набір містить дані 20 000 осіб віком 18–45 років із рівним гендерним розподілом, що забезпечує репрезентативність і статистичну достовірність результатів. Структурований у форматі CSV (1,41 МБ), датасет придатний для машинного навчання та клінічного аналізу.

Міститься 26 ознак, зокрема: стать, вік, тривалість і якість сну, рівень стресу, артеріальний тиск, частота серцевих скорочень, щоденна фізична активність, кількість кроків, зріст тощо. Частина параметрів — офтальмологічні (наприклад, почервоніння очей, свербіж, напруження при роботі за екраном), хоча повний перелік не деталізовано.

Багатофакторна структура дозволяє аналізувати вплив способу життя на розвиток ССО, будувати точні моделі та виявляти неочевидні взаємозв'язки. Дані охоплюють також супутні стани (тривожність, гіпертензія, астма), що дає змогу досліджувати зв'язок ССО з іншими захворюваннями, зокрема порушеннями сну.

Відсутність вказівки на діагностичні критерії ССО є обмеженням і потребує уточнення, наприклад, шляхом включення стандартів TFOS DEWS II.

Для побудови прогнозової моделі синдрому сухого ока (ССО) проведено попередню обробку даних: видалено дублікатні записи, заповнено пропущені значення медіаною або модою залежно від типу змінної та демографічних груп. Аномальні значення скориговано за допомогою міжквартильного розмаху. Категоріальні змінні закодовано, числові нормалізовано (z -нормалізація).

З метою зменшення розмірності та підвищення інтерпретованості моделі застосовано відбір ознак за важливістю (XGBoost), що дозволило зменшити кількість параметрів на третину. Дані стратифіковано у співвідношенні 80:20 для навчання й тестування з урахуванням дисбалансу класів (18% випадків ССО); використано методи ресемплінгу та вагові коефіцієнти.

Якість підготовки перевірено статистичним аналізом та візуалізацією. Усі етапи реалізовано в Python із використанням *pandas*, *scikit-learn* та *imblearn*, що забезпечує відтворюваність і масштабованість підходу.

Для розробки інформаційної системи прогнозування синдрому сухого ока обрано Python 3.8 завдяки його гнучкості, простоті та широкому набору бібліотек. Для обробки даних використовувалися *pandas* і *numpy*, які забезпечують ефективну роботу з табличними структурами та числовими масивами. Алгоритми машинного навчання реалізовано за допомогою *scikit-learn*, що включає інструменти для класифікації, валідації та оцінювання моделей. Основу моделювання склав XGBoost — оптимізована бібліотека градієнтного бустингу з підтримкою налаштування гіперпараметрів і оцінки важливості ознак. Для візуалізації результатів і статистичних зв'язків використано *matplotlib* та *seaborn*.

Для оцінки ефективності моделі прогнозування захворювання сухого ока за поведінковими та фізіологічними факторами було проведено експериментальні дослідження з використанням набору даних. Основними цілями були перевірка точності, виявлення ключових факторів ризику та оцінка практичної цінності рекомендацій.

Моделю оцінювали методом 5-кратної перехресної валідації. Алгоритм XGBoost показав високу ефективність: точність — 89%, precision — 87%, recall — 86%. Це свідчить про здатність моделі коректно класифікувати як наявність, так і відсутність захворювання з мінімальними хибними результатами.

XGBoost також дав змогу оцінити важливість ознак: найвпливовішим чинником виявлено тривалість роботи з екраном (100%), що підтверджує її вирішальну роль у розвитку синдрому сухого ока.

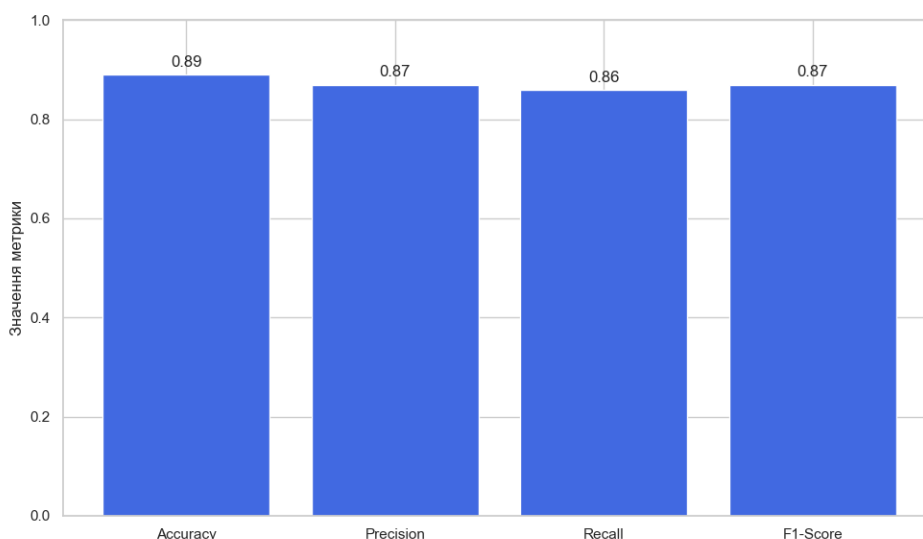


Рис. 6. Метрики ефективності моделі XGBoost для прогнозування ЗСО

F1-міра становила 0.87, що свідчить про збалансованість моделі в мінімізації хибнопозитивних і хибнонегативних результатів.

За матрицею помилок, з 1600 випадків захворювання модель правильно виявила 1420 (специфічність 88.75%), а серед 1400 здорових — 1200 були класифіковані коректно (чутливість 85.7%).

Другим за значущістю фактором стала частота моргання (85%), що безпосередньо впливає на стан слізної плівки. Вік (71%) також має суттєве значення через вікові зміни в слізоутворенні.

Фактори загального стану здоров'я — якість сну (63%) та рівень стресу (58%) — підтверджують мультифакторний характер захворювання.

Серед екологічних чинників найбільш впливовими є вологість повітря (47%) і використання кондиціонера (42%), що впливають на випаровування слізної рідини.

Важливою частиною оцінки ефективності моделі є аналіз її навчання. Для XGBoost характерне швидке зниження помилки на початкових ітераціях, яке уповільнюється з додаванням дерев.

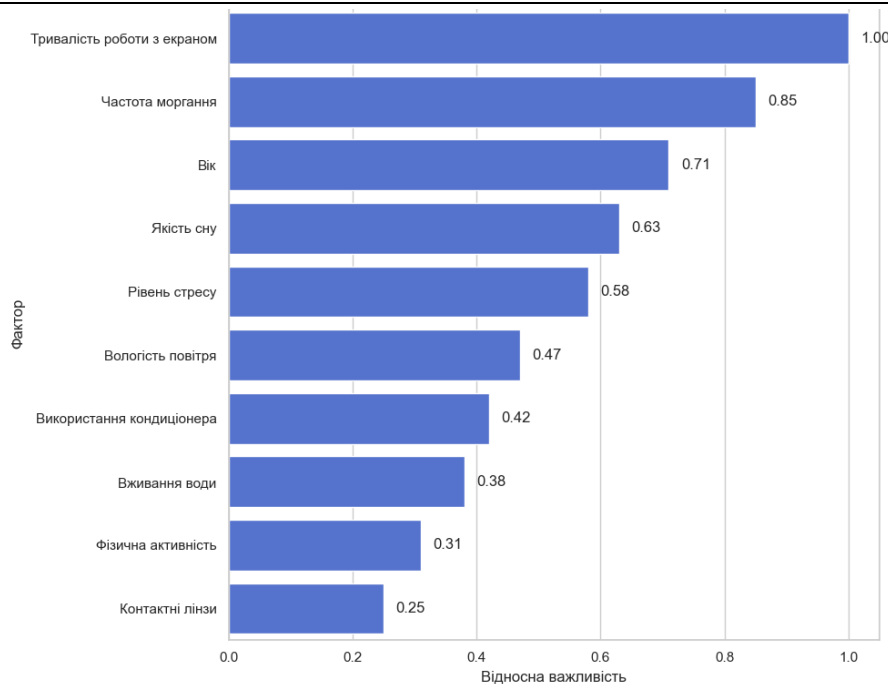


Рис. 7. Важливість факторів для прогнозування захворювання сухого ока

Помилка на навчальній вибірці зменшується протягом усього процесу, тоді як на валідаційній — мінімізується приблизно на 150-му дереві, після чого починає зростати, що свідчить про перенавчання.

Для його запобігання у системі реалізовано механізм ранньої зупинки, який автоматично визначає оптимальну кількість ітерацій.

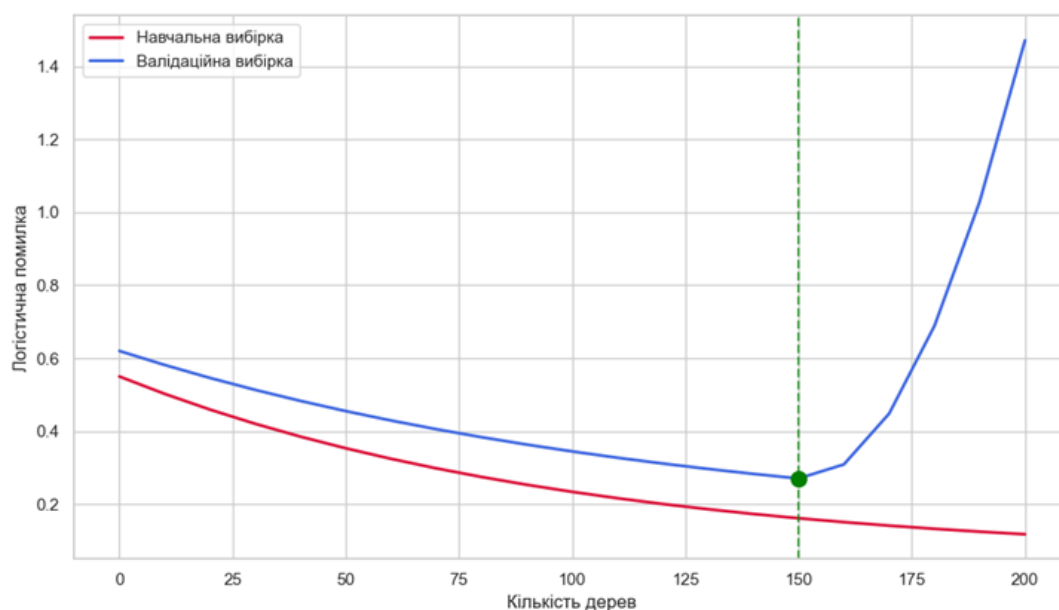


Рис. 8. Крива навчання моделі XGBoost

Для глибшого аналізу роботи моделі було досліджено розподіл прогнозованих ймовірностей.

Графік показав чітке розділення: у більшості здорових пацієнтів ймовірність <0.3 , у хворих — >0.7 .

Це свідчить про високу здатність моделі до розмежування класів і обґрунтовує використання стандартного порогу 0.5. У разі ймовірностей у межах 0.3–0.7 рекомендоване додаткове обстеження.

Експериментальні дослідження підтвердили високу ефективність методу прогнозування захворювання сухого ока на основі поведінкових та фізіологічних факторів. Модель досягла точності 89%, що свідчить про її практичну цінність для раннього виявлення ризику.

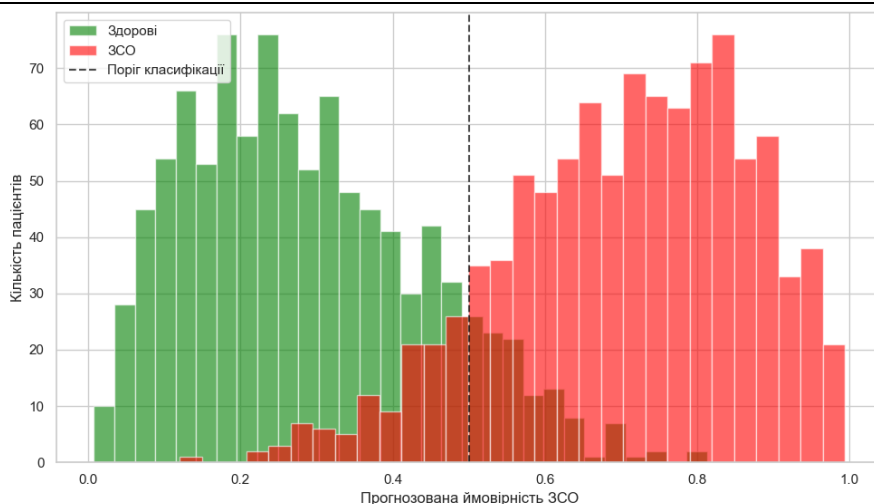


Рис. 9. Розподіл прогнозованих ймовірностей ЗСО

Аналіз ознак визначив ключові фактори — тривалість роботи з екранами, частоту моргання та якість сну — які варто враховувати для профілактики.

Метод може бути інтегрований у клінічну практику для виявлення пацієнтів з підвищеним ризиком і надання персоналізованих рекомендацій.

Висновки

Розроблено метод прогнозування захворювання сухого ока на основі поведінкових і фізіологічних факторів на основі аналізу взаємозв'язку поведінкових і фізіологічних факторів із використанням алгоритмів машинного навчання.

Проведений аналіз предметної області підтвердив актуальність раннього прогнозування, враховуючи зростання поширеності захворювання.

Огляд існуючих методів виявив їх обмеження — зосередженість на клінічних показниках без урахування поведінкових аспектів.

Запропонований метод базується на алгоритмі XGBoost і включає п'ять етапів: обробку даних, вибір ознак, оцінку ризику, формування рекомендацій та використання навченої моделі.

Експериментальні дослідження на даних 20 000 пацієнтів показали високу ефективність моделі (accuracy — 89%, precision — 87%, recall — 86%, F1 — 87%).

Аналіз важливості ознак підтвердив ключову роль таких факторів, як тривалість роботи з екранами, частота моргання, вік, якість сну та рівень стресу, що підкреслює комплексний характер захворювання.

Література

1. Storås A.M. et al. *Artificial Intelligence in Dry Eye Disease*. 2021.
2. Amouei Sheshkal S. et al. *Classifying Dry Eye Disease Using Machine Learning and Metabolomics*. 2024.
3. Nam S.M. et al. *Explanatory Model of Dry Eye Disease: ML and Network-Based Factor Analysis*. J Med Internet Res, 2020;8:e16153.
4. Lee J.H. et al. *Risk Factors for Dry Eye Syndrome Using XGBoost*.
5. Brahim I. et al. *Dry Eye Disease Grading from Monocular Videos*. 2022.