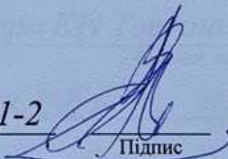
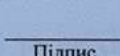



КВАЛІФІКАЦІЙНА РОБОТА БАКАЛАВРА

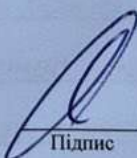
на тему Метод прогнозного моделювання захворювання сухого ока на основі аналізу поведінкових та фізіологічних факторів

Галузь знань 12 – Інформаційні технології
Шифр і назва галузі знань
Спеціальність 122 – Комп'ютерні науки
Шифр і назва спеціальності
Освітня програма Комп'ютерні науки
Назва освітньої програми

Виконав: студентка групи КН-21-2  Єлизавета МАНДЗЮК
Група виконавця Підпис Ім'я, ПРІЗВИЩЕ

Керівник: ст. викладач кафедри КН  Тетяна СКРИПНИК
Науковий ступінь, посада Підпис Ім'я, ПРІЗВИЩЕ

Нормоконтроль: к.т.н., доц. каф. КН  Руслан БАГРІЙ
Науковий ступінь, посада Підпис Ім'я, ПРІЗВИЩЕ

До захисту допускаю:
Зав. кафедри КН, д.т.н., професор  Олександр БАРМАК
Підпис Ім'я, ПРІЗВИЩЕ

18 06 2025 р.

ХМЕЛЬНИЦЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ

Факультет інформаційних технологій

Кафедра комп'ютерних наук

Освітній ступінь бакалавр

Галузь знань 12 – Інформаційні технології

Спеціальність 122 – Комп'ютерні науки

ЗАТВЕРДЖУЮ

Завідувач кафедри комп'ютерних наук

(підпис)

д.т.н., професор Олександр БАРМАК

« 10 » 02 2025 року

**ЗАВДАННЯ
НА КВАЛІФІКАЦІЙНУ РОБОТУ БАКАЛАВРА**

1. Тема кваліфікаційної роботи бакалавра: «Метод прогнозного моделювання захворювання сухого ока на основі аналізу поведінкових та фізіологічних факторів»

2. Завдання видано студенту Єлизавета МАНДЗЮК
(Ім'я, прізвище)

3. Керівник роботи ст. викладач кафедри КН Тетяна СКРИПНИК
(посада, ім'я, прізвище)

4. Затверджено наказом університету від « 04 » 02 2025 р. № 23

5. Дата видачі завдання студенту: « 10 » 02 2025 р.

6. Зміст пояснювальної записки (перелік задач) та вихідні дані:

Мета кваліфікаційної роботи бакалавра – підвищення точності ранньої діагностики та прогнозування розвитку синдрому сухого ока. Перелік задач: виконати огляд теоретичних підходів та обрати підхід для прогнозного моделювання захворювання сухого ока на основі аналізу поведінкових та фізіологічних факторів; розробити метод прогнозного моделювання захворювання сухого ока на основі аналізу поведінкових та фізіологічних факторів; описати інформаційну структуру системи для прогнозного моделювання захворювання сухого ока; виконати дослідження точності розробленого методу прогнозного моделювання захворювання сухого ока з використанням розробленої системи.

7. Календарний план виконання кваліфікаційної роботи бакалавра:

№	Назва етапів (розділів) кваліфікаційної роботи бакалавра	Термін виконання	Примітка
1	Вибір напрямку дослідження та узгодження тематики кваліфікаційної роботи бакалавра з керівником, складання календарного графіка виконання роботи	січень 2025	виконано
2	Ознайомлення з предметною областю, формулювання мети та задач дослідження, визначення об'єкта та предмета дослідження	лютий 2025	виконано
3	Проектування та розробка загальної архітектури системи, вибір засобів реалізації програмного забезпечення	березень 2025	виконано
4	Розробка інформаційної системи	квітень 2025	виконано
5	Написання пояснювальної записки, урахування зауважень керівника, оформлення згідно вимог	травень 2025	виконано
6	Розробка презентаційних матеріалів та попередній захист кваліфікаційної роботи	травень 2025	виконано
7	Отримання відгуку керівника, рецензії, перевірка на плагіат, нормоконтроль	червень 2025	виконано
8	Підготовка до захисту та захист кваліфікаційної роботи бакалавра	червень 2025	виконано

Виконавець: студентка групи КН-21-2

Група виконавця

Підпис

Єлизавета МАНДЗЮК

Ім'я, ПРІЗВИЩЕ

Керівник: ст. викладач кафедри КН

Науковий ступінь, посада

Підпис

Тетяна СКРИПНИК

Ім'я, ПРІЗВИЩЕ

Анотація

Тема кваліфікаційної роботи бакалавра: Метод прогнозного моделювання захворювання сухого ока на основі аналізу поведінкових та фізіологічних факторів.

Виконавець кваліфікаційної роботи бакалавра: студентка групи КН-21-2 Єлизавета МАНДЗЮК.

Керівник кваліфікаційної роботи бакалавра: ст. викладач кафедри КН Тетяна СКРИПНИК

Кваліфікаційна робота бакалавра містить:

Пояснювальна записка				Кількість додатків
Сторінок	Рисунків	Таблиць	Джерел інформації	
65	18	3	40	2

У кваліфікаційній роботі вирішено задачу підвищення точності прогнозування захворювання сухого ока шляхом розробки методу прогнозного моделювання на основі аналізу поведінкових та фізіологічних факторів.

Запропоновано метод прогнозного моделювання захворювання сухого ока включає попередню обробку даних, виділення найбільш важливих ознак та формування персоналізованих рекомендацій. Розроблено модуль обробки даних з нормалізацією числових параметрів та кодуванням категоріальних змінних для підвищення точності прогнозування.

Розроблений метод може бути використаний в офтальмологічній практиці для раннього виявлення пацієнтів з високим ризиком розвитку захворювання сухого ока та формування індивідуальних профілактичних заходів.

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

Виконавець:

студентка групи КН-21-2
Група виконавця


Підпис

Єлизавета МАНДЗЮК
Ім'я, ПРІЗВИЩЕ

Зміст

Перелік скорочень	2
Вступ.....	3
Розділ 1 Характеристика предметної області прогнозного моделювання захворювання сухого ока.....	5
1.1 Сучасні підходи до прогнозного моделювання захворювання сухого ока....	5
1.2 Роль машинного навчання в прогнозуванні захворювання сухого ока	5
1.3 Інтеграція штучного інтелекту у діагностику та моніторинг ЗСО	7
1.4. Практичні підходи до класифікації та лікування ЗСО.....	10
1.5 Інтеграція нових технологій для прогнозних моделей ЗСО.....	11
1.6 Мета, задачі до методу	13
Розділ 2 Метод прогнозного моделювання захворювання сухого ока на основі аналізу поведінкових та фізіологічних факторів	14
2.1 Схема методу прогнозного моделювання захворювання сухого ока	14
2.2 Архітектура системи автоматизованої обробки даних	17
2.3 Реалізація етапів прогнозного моделювання відповідно до схеми дослідження	22
2.4 Процес обробки та нормалізації вхідних даних.....	29
2.5 Висновок до розділу 2	34
Розділ 3 Експериментальна перевірка методу прогнозного моделювання захворювання сухого ока.....	35
3.1 Характеристика набору даних для експериментів	35
3.2. Підготовка даних для моделювання	38
3.3 Вибір засобів розробки методу.....	42
3.4 Структурні елементи програмної реалізації.....	43
3.5 Результати досліджень	48
3.6 Висновок до розділу 3	55
Висновок	57
Перелік посилань.....	59
ДОДАТКИ	

Перелік скорочень

Скорочення, термін, позначення	Пояснення
ЗСО	Захворювання сухого ока
XGBoost	eXtreme Gradient Boosting – алгоритм градієнтного бустингу
OSDI	Ocular Surface Disease Index – індекс захворювання поверхні ока
AUC	Area Under Curve – площа під ROC-кривою
F1	F1-міра – гармонійне середнє між точністю і повнотою
ШІ	Штучний інтелект
API	Application Programming Interface – програмний інтерфейс застосунку
IQR	Interquartile Range – міжквартильний розмах

Вступ

Кваліфікаційна робота бакалавра стосується розробки методу прогнозного моделювання захворювання сухого ока на основі аналізу поведінкових та фізіологічних факторів

Актуальність – захворювання сухого ока (ЗСО) є одним із найпоширеніших офтальмологічних розладів, що вражає значну частину населення у всьому світі. За даними сучасних епідеміологічних досліджень, поширеність цього захворювання коливається від 5% до 50% залежно від географічного регіону, вікової групи та діагностичних критеріїв. В умовах зростаючої цифровізації суспільства, збільшення тривалості роботи з електронними пристроями та негативного впливу екологічних факторів, захворюваність на синдром сухого ока демонструє чітку тенденцію до зростання, особливо серед молодшого населення.

Захворювання сухого ока значно впливає на якість життя пацієнтів, викликаючи дискомфорт, біль, погіршення зору та зниження працездатності. Крім того, воно накладає суттєвий економічний тиск на систему охорони здоров'я через витрати на лікування та втрату продуктивності праці.

Об'єкт дослідження – процес розвитку та діагностики синдрому сухого ока

Предмет дослідження – методи та алгоритми прогнозного моделювання захворювання сухого ока з використанням технологій штучного інтелекту для аналізу поведінкових та фізіологічних факторів

Мета кваліфікаційної роботи бакалавра – підвищення точності ранньої діагностики та прогнозування розвитку синдрому сухого ока шляхом розробки методу прогнозного моделювання захворювання сухого ока на основі аналізу поведінкових та фізіологічних факторів.

Завдання кваліфікаційної роботи бакалавра:

- виконати огляд теоретичних підходів та обрати підхід для прогнозного моделювання захворювання сухого ока на основі аналізу поведінкових та фізіологічних факторів;
- розробити метод прогнозного моделювання захворювання сухого ока на основі аналізу поведінкових та фізіологічних факторів;
- описати інформаційну структуру системи для прогнозного моделювання захворювання сухого ока;
- виконати дослідження точності розробленого методу прогнозного моделювання захворювання сухого ока з використанням розробленої системи.

Розділ 1 Характеристика предметної області прогнозного моделювання захворювання сухого ока

1.1 Сучасні підходи до прогнозного моделювання захворювання сухого ока

Захворювання сухого ока (ЗСО) є поширеною проблемою очної поверхні, що впливає на життя багатьох людей в досить широкій популяції. Сучасна офтальмологія демонструє зростаючий інтерес до використання методів прогнозного моделювання для ранньої діагностики, оцінки ризиків та персоналізованого лікування пацієнтів із ЗСО.

Штучний інтелект відіграє значну роль у прогнозному моделюванні ЗСО. Дослідження показують, що моделі ШІ можуть прогнозувати результати, пов'язані із захворюванням сухого ока, через аналіз клінічних оцінок, оцінок симптомів та значного зважування факторів способу життя, таких як вік, робота за близької відстані, споживання алкоголю, фізичні вправи та вплив навколишнього середовища [1, 2].

Розвиток методологій аналізу даних дозволив створити системи оцінки ризику ЗСО. Інструменти, керовані штучним інтелектом, включаючи алгоритми ШІ, можуть аналізувати поведінку та фізіологію для покращення прогнозного моделювання захворювання сухого ока, в кінцевому підсумку покращуючи точність діагностики та персоналізуючи підходи до лікування для кращих результатів пацієнта [3–5].

1.2 Роль машинного навчання в прогнозуванні захворювання сухого ока

Машинне навчання значно розширило можливості прогнозування ЗСО. Сучасні дослідження демонструють високу точність різних алгоритмів у класифікації та передбаченні захворювання.

Інноваційні підходи до класифікації пацієнтів із ЗСО продовжують розвиватися. Метод класифікації пацієнтів із захворюванням сухого ока від здорових контролів за допомогою машинного навчання та даних підкреслює важливість вибору відповідних моделей машинного навчання для точних прогнозів [6]. Значним обмеженням цього підходу є відсутність універсальної моделі машинного навчання, що підходить для всіх випадків, та суттєвий вплив вибору моделі на якість прогнозування.

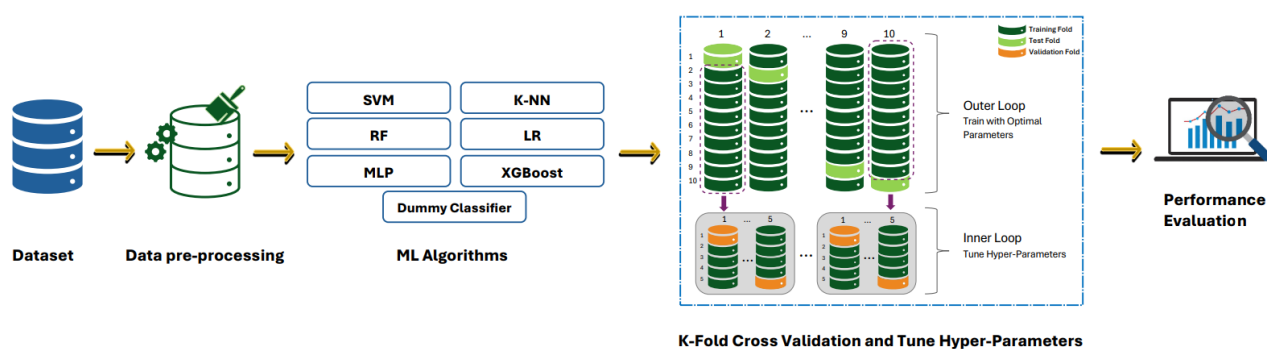


Рисунок 1.1 – Операційний процес запропонованої класифікації синдрому сухого ока [6]

Прогрес у глибокому навчанні відкрив нові можливості для діагностики ЗСО. Вдосконалена стекова автокодувальна оптимізована глибока нейронна мережа (ESAE-ODNN) для прогнозування захворювання сухого ока використовує вибір, вилучення та класифікацію ознак для аналізу особливостей, пов'язаних з дисфункцією мейбомієвої залози, та складних взаємозв'язків, покращуючи ранню діагностику з точністю класифікації 96,34% [7]. Подібний підхід представлений у дослідженні, яке використовує аналіз великих медичних даних, включаючи дані на основі оміків та мобільні програми для здоров'я, що може полегшити прогнозне моделювання захворювання сухого ока шляхом інтеграції поведінкових та фізіологічних факторів [8]. Проте, ці дослідження виявили відсутність надійних біомаркерів для ЗСО та обмеження сучасних методів лікування, які лише полегшують симптоми, але не забезпечують профілактичних методів.

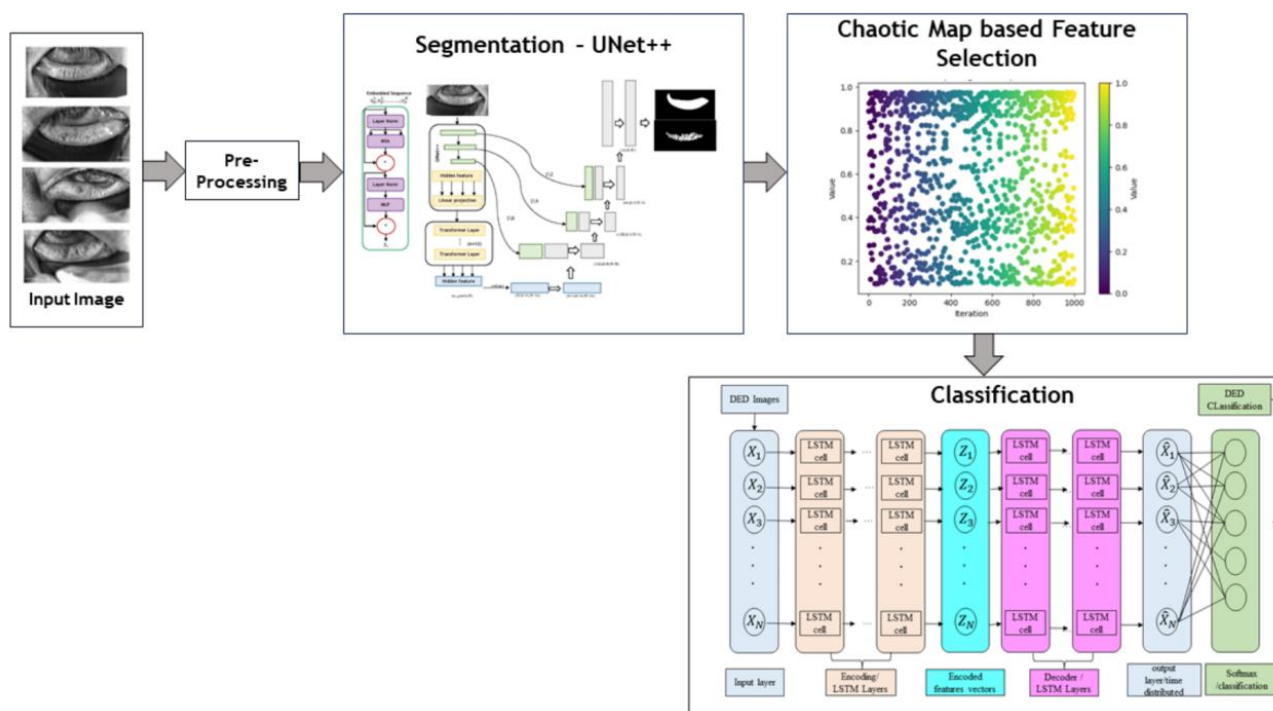


Рисунок 1.2 – Загальна архітектура ESAE-ODNN [7]

Прогнозування нестабільності слізної плівки також стало важливим напрямком досліджень. Алгоритми машинного навчання демонструють високу точність у аналізі клінічних даних та визначенні ключових ознак захворювання сухих очей, включаючи фарбування поверхні очей, виразність мейбомійської залози, частоту моргання та осмолярність [9–11]. Однак автори відзначають потенційні обмеження та підводні камені застосування даних методів, включаючи суб'єктивну інтерпретацію діагностичних тестів та потребу в розробці стандартизованих моделей.

1.3 Інтеграція штучного інтелекту у діагностику та моніторинг ЗСО

Штучний інтелект став невід'ємною частиною процесу діагностики та моніторингу ЗСО. Інтеграція штучного інтелекту у підходи до діагностики та моніторингу ЗСО дозволяє аналізувати клінічні дані, включаючи поведінкові та фізіологічні фактори, для покращення прогнозного моделювання захворювання

сухих очей, покращення діагностики та моніторингу шляхом всебічного аналізу симптомів, ознак та результатів тестів [12–14].

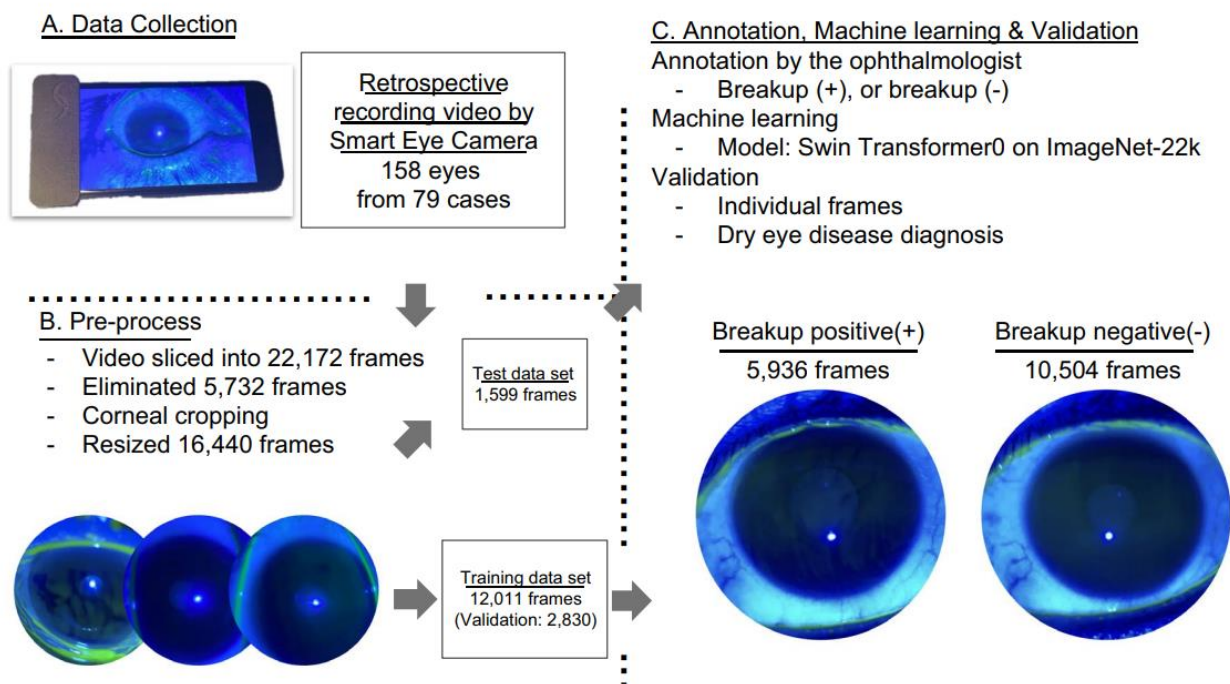


Рисунок 1.3 – Блок-схема дослідження [13]

Однак ці методи стикаються з такими обмеженнями, як відсутність кореляції між ознаками та симптомами, обмежена надійність діагностичних тестів, та необхідність покращення точності за рахунок збільшення обсягу тренувальних даних. Систематичний огляд та мета-аналіз застосування штучного інтелекту в діагностиці захворювання сухого ока підкреслює перспективність цих технологій, хоча й не акцентує увагу на прогнозуванні на основі поведінкових та фізіологічних факторів [15–18]. Порівняння системи аналізу моргання за допомогою глибокого навчання з інтерферометром Liriview для діагностики захворювання сухих очей демонструє потенціал штучного інтелекту в оцінці моделей моргання [19]. Проте, клінічна практичність систем глибокого навчання потребує подальшої перевірки, а більші дослідження вибірки необхідні для оцінки клінічної корисності.

Поведінкові фактори відіграють ключову роль у розвитку та прогресуванні ЗСО, що робить їх важливими компонентами прогнозних моделей. Вплив способу життя на розвиток захворювання сухого ока є добре

документованим. Дослідження підкреслюють вплив вибору способу життя, такого як вітамін D та добавки Омега-3, фізичні навантаження, скорочення часу на екрані та відмова від куріння на ризик захворювання сухого ока [20, 21]. Однак ці дослідження часто не враховують деякі важливі аспекти, такі як вплив наркотиків, алкоголю та харчування на ЗСО, а також відмінності у специфічних реакціях на фактори навколишнього середовища.

Фізична активність може виступати як захисний фактор проти ЗСО. Дослідження асоціації між захворюванням сухого ока, фізичною активністю та сидячою поведінкою серед офісних працівників підкреслює потенційну роль регулярної фізичної активності у запобіганні розвитку ЗСО [22]. Фізіологічні фактори є важливими компонентами прогнозних моделей ЗСО, надаючи об'єктивні показники стану здоров'я очної поверхні. Моргання очей є ключовим фізіологічним процесом, що впливає на стан слізної плівки. Дослідження зв'язку між захворюванням сухого ока та поведінкою моргання підкреслює значимість динаміки моргання у розвитку та прогресуванні ЗСО [23, 24]. Ці дослідження, однак, не деталізують конкретні методи прогнозного моделювання на основі характеристик моргання.

Оптимізація діагностики захворювання сухого ока може бути досягнута шляхом інтеграції глибокого навчання із аналізу опитувальників OSDI та відео моргання очей [24]. Це дослідження фокусується на поведінкових (анкетах) та фізіологічних (характеристиках моргання) факторах для покращення прогнозного моделювання ЗСО.

Дисфункція мейбомієвої залози є важливим фізіологічним фактором у розвитку ЗСО. Дослідження прогнозування випадіння мейбомієвої залози з використанням пояснюваного штучного інтелекту підкреслює важливість цього фактора для захворювання сухого ока [10, 25]. Проте ці дослідження виявили відсутність асоціації з симптомами інших захворювань та суперечливі результати в епідеміологічних дослідженнях.

Фактори навколишнього середовища також відіграють значну роль у розвитку ЗСО, що робить їх важливими компонентами прогнозних моделей.

Забруднення повітря та метеорологічні фактори можуть значно впливати на ризик розвитку ЗСО. Модель прогнозування захворюваності синдромом сухого ока з використанням забруднювачів повітря та метеорологічних факторів демонструє вплив NO₂ та відносної вологості на регіональні відхилення в поширеності ЗСО [26]. Це дослідження, однак, зосереджене на факторах навколишнього середовища, а не на поведінкових чи фізіологічних ознаках.

1.4. Практичні підходи до класифікації та лікування ЗСО

Практичні підходи до класифікації та лікування ЗСО мають важливе значення для клінічного застосування прогнозних моделей.

Новітні парадигми в дослідженнях захворювання сухого ока підкреслюють важливість специфічної для захворювання та на основі біомаркерів стратифікації пацієнтів для персоналізованого лікування, що може служити методом прогнозного моделювання захворювання сухого ока [27]. Однак ці підходи мають обмеження, такі як рефрактерні пацієнти до протизапальної схеми, розбіжність між ознаками та симптомами у деяких пацієнтів, та необхідність більш простого управління в неспеціалізованих клініках.

Спрощені схеми прийняття рішень можуть покращити доступність лікування ЗСО. Дослідження спрощеної схеми прийняття рішень у захворюванні сухого ока підкреслює важливість практичних підходів до управління цим станом [28]. Обмеженнями цього підходу є те, що спрощена схема не є остаточною для всіх типів ЗСО, а клінічні фактори не мають специфічності і слабо вказують на механізми захворювання.

Прогнозування ризику ЗСО у специфічних популяціях дозволяє розробити більш цільові підходи до профілактики та лікування. Прогнозування ризику ЗСО у критичних пацієнтів представляє особливий інтерес. Дослідження використовувало багатоваріантний аналіз виживання та регресію Кокса для встановлення моделі прогнозування ризику сухого ока, ідентифікуючи значні

поведінкові та фізіологічні фактори, такі як O_2 у кімнатному повітрі, частота моргання та наявність судинних захворювань [29]. Обмеженнями цього дослідження є його проведення з конкретним профілем пацієнта та потреба в мультицентричних дослідженнях для зовнішньої валідності.

Застосування штучних нейронних мереж та логістичної регресії в оцінці ризику ЗСО після вітректомії демонструє важливість специфічних факторів ризику для конкретних груп пацієнтів [30]. Модель динаміки сліз може інформувати про розробку протоколів лікування ЗСО. В дослідженні розроблено модель динаміки сліз на основі масових балансів води та розчинених речовин, що передбачає фізіологічний вплив на динаміку сліз [31]. Обмеженнями цієї моделі є необхідність подальшого уточнення для точних прогнозів та те, що вона не призначена для заміни експериментальних досліджень.

Етіологія тертя у розвитку ЗСО підкреслює важливість розуміння динаміки слізної плівки. Дослідження підкреслює важливість розуміння етіології сухого ока та динаміки слізної плівки та очної поверхні [32]. Обмеженнями цього підходу є відсутність вимірюваних висновків у деяких випадках та нечітке визначення гомеостазу слізної плівки.

1.5 Інтеграція нових технологій для прогнозних моделей ЗСО

Майбутні напрямки розвитку прогнозних моделей ЗСО включають інтеграцію нових технологій та підходів для підвищення точності прогнозування. Застосування моделей тварин може допомогти у розумінні ЗСО. Дослідження різних моделей тварин, які допомагають зрозуміти патогенез та терапевтичні перспективи захворювання сухих очей, може інформувати розробку прогнозних моделей [33]. Однак це дослідження не зосереджується на прогнозованому моделюванні на основі поведінкових та фізіологічних факторів.

Огляд сучасної літератури щодо прогнозного моделювання ЗСО вказує на потребу в покращенні методів діагностики та лікування. Дослідження сучасних діагностичних методів ЗСО визначило тісний зв'язок цього стану з

повсякденними звичками людини, віком, хронічними захворюваннями, інфекціями очей та травмами [34–36]. Ці дослідження підкреслюють необхідність розробки терапевтичних засобів, оскільки жоден з нинішніх не націлений на першопричину захворювання сухих очей.

Розуміння домінантного підтипу та основних причин має вирішальне значення для прогнозного моделювання та стратегій управління ЗСО [37]. Обмеженнями залишаються відсутність єдиного надійного діагностичного інструменту та плутанина через відсутність доказових рекомендацій щодо лікування.

Запальні фактори як біомаркери ЗСО викликають значний інтерес у науковій спільноті. Дослідження запальних факторів як ознак синдрому сухого ока в моделі з використанням осмотичного тиску оцінює життєздатність клітин та експресію запального фактора у відповідь на осмотичний тиск [36]. Ця робота підкреслює важливість виявлення специфічних біомаркерів для діагностування та прогнозування розвитку захворювання.

Аналізування моделей захворювання сухого ока на основі параметрів імунних клітин надає новий погляд на терапевтичну перспективу [38]. Цей підхід фокусується на імунологічних параметрах і може допомогти у стратифікації пацієнтів та прогнозуванні відповіді на лікування. Однак дослідження має обмеження через аналіз та виключення досліджень без достатніх даних або імунних факторів.

Аналіз даних довгострокового спостереження за пацієнтами може покращити прогнозування прогресії ЗСО. Дослідження щоденного журналу активності для співвіднесення часу, витраченого на дії, що обтяжують ЗСО, з клінічними особливостями свідчить про потенціал поведінкових модифікацій як методу прогнозного моделювання для оцінки тяжкості та симптомів захворювання сухого ока [29, 39]. Цей підхід дозволяє визначити елементи, що діють на прогресування захворювання, та розробити стратегії для мінімізації їх впливу. Однак обмеженнями цих досліджень є невеликі вибірки, які можуть не

представляти всіх пацієнтів з ЗСО, та можливий вплив самого процесу моніторингу на поведінку пацієнтів.

Прогнозування відповіді на лікування залишається складним завданням через гетерогенність ЗСО. Дослідження персоналізованої медицини для захворювання сухого ока підкреслює необхідність стратифікації пацієнтів на основі домінуючого підтипу та основних причин для прогнозного моделювання та стратегій управління [8, 37, 40].

1.6 Мета, задачі до методу

На основі проведеного аналізування досліджень метою роботи підвищення точності ранньої діагностики та прогнозування розвитку синдрому сухого ока шляхом розробки методу прогнозного моделювання захворювання сухого ока на основі аналізу поведінкових та фізіологічних факторів.

Завдання кваліфікаційної роботи бакалавра:

– виконати огляд теоретичних підходів та обрати підхід для прогнозного моделювання захворювання сухого ока на основі аналізу поведінкових та фізіологічних факторів;

– розробити метод прогнозного моделювання захворювання сухого ока на основі аналізу поведінкових та фізіологічних факторів;

– описати інформаційну структуру системи для прогнозного моделювання захворювання сухого ока;

– виконати дослідження точності розробленого методу прогнозного моделювання захворювання сухого ока з використанням розробленої системи.

Розділ 2 Метод прогнозного моделювання захворювання сухого ока на основі аналізу поведінкових та фізіологічних факторів

2.1 Схема методу прогнозного моделювання захворювання сухого ока

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

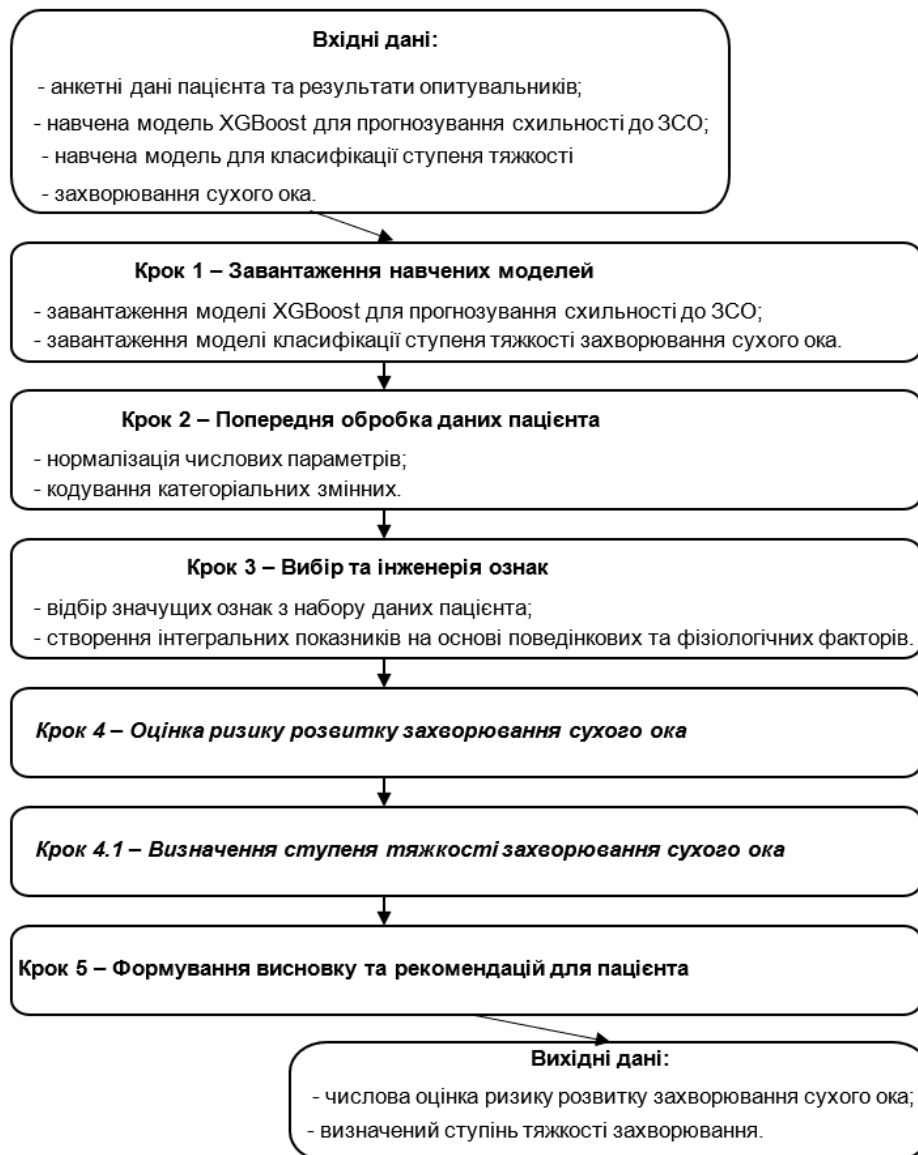


Рисунок 2.1 – Схема методу прогнозного моделювання захворювання сухого ока на основі аналізу поведінкових та фізіологічних факторів

Метод представляє собою послідовність взаємопов'язаних етапів, спрямованих на оцінку індивідуального ризику розвитку захворювання та визначення його потенційної тяжкості для забезпечення персоналізованого підходу до профілактики та лікування.

Вхідними даними для методу слугують анкетні дані пацієнта та результати опитувальників OSDI, що дозволяють оцінити суб'єктивні симптоми захворювання. Також використовуються навчена модель для прогнозування схильності до захворювання сухого ока та спеціалізована модель для класифікації ступеня тяжкості захворювання. Ці дані формують основу для подальшого аналізу та прогнозування.

На першому кроці відбувається завантаження навчених моделей для аналізу даних. Здійснюється завантаження навченої моделі, що була оптимізована для прогнозування ризику розвитку захворювання сухого ока (ЗСО) на основі аналізу поведінкових та фізіологічних факторів. Ця модель характеризується високою точністю та здатністю виявляти складні нелінійні взаємозв'язки між різними ознаками захворювання.

Другий крок методу фокусується на попередній обробці даних пацієнта для подальшого аналізу. На цьому етапі здійснюється нормалізація числових параметрів для приведення всіх показників до єдиної шкали, що є необхідною умовою для коректної роботи алгоритмів машинного навчання. Числові показники, вік людини, тривалість її роботи з цифровими пристроями, частота моргання, результати клінічних тестів, перетворюються з використанням статистичних методів нормалізації. Категоріальні змінні, такі як стать людини, супутні хвороби, характер життя, проходять процедуру кодування для представлення якісної інформації у числовому форматі, придатному для аналізу. Попередня обробка даних забезпечує їх стандартизацію та підвищує якість подальшого прогнозування.

Третій крок методу передбачає вибір та інженерію ознак для побудови прогностичної моделі. На цьому етапі з повного набору доступних даних відбираються найбільш інформативні показники, що мають найвищу

прогностичну цінність для оцінки ризику розвитку ЗСО. Відбір ознак ґрунтується на аналізі їх важливості, визначеної під час тренування моделі. Паралельно з відбором існуючих показників відбувається створення нових інтегральних ознак, що відображають взаємозв'язки між різними факторами. Формуються композитні показники, такі як співвідношення між поведінковими та фізіологічними параметрами, що дозволяє виявити приховані патерни та підвищити точність прогнозування. Етап вибору та інженерії ознак має критичне значення для забезпечення високої прогностичної здатності моделі.

Четвертий крок методу полягає в оцінці ризику розвитку захворювання сухого ока з використанням алгоритму прогнозування. На цьому етапі оброблені та відібрані дані пацієнта аналізуються за допомогою навченої моделі, яка обробляє складні взаємозв'язки між поведінковими та фізіологічними факторами. Модель базується на принципі градієнтного бустингу дерев рішень, де кожне наступне дерево фокусується на виправленні помилок попередніх, що забезпечує високу точність прогнозів. Алгоритм враховує як прямий вплив окремих факторів, так і їх взаємодію, що особливо важливо для мультифакторних захворювань, таких як ЗСО. Результатом цього етапу є кількісна оцінка ймовірності розвитку захворювання сухого ока, виражена у вигляді числового значення або відсотка.

Крок 4.1 розширює оцінку ризику через визначення потенційного ступеня тяжкості захворювання сухого ока. На цьому етапі використовується спеціалізована модель класифікації, яка аналізує взаємозв'язки між клінічними показниками, поведінковими та фізіологічними факторами для прогнозування можливої тяжкості перебігу захворювання. Модель враховує широкий спектр показників, включаючи вік пацієнта, наявність системних захворювань, стан мейбомієвих залоз, показники слізної плівки та суб'єктивні симптоми. На основі аналізу цих даних визначається потенційний ступінь тяжкості захворювання: легкий, середній або тяжкий, що має важливе значення для планування терапевтичних заходів та прогнозування клінічного перебігу.

П'ятий, заключний крок методу спрямований на формування висновку та рекомендацій для пацієнта на основі проведеного аналізу. На цьому етапі система інтегрує результати оцінки ризику розвитку ЗСО та прогнозування ступеня тяжкості для створення звіту.

Вихідними даними методу є числова оцінка ризику розвитку захворювання сухого ока та визначений ступінь потенційної тяжкості захворювання. Ці дані служать основою для персоналізованого підходу до профілактики та лікування ЗСО, орієнтованого на індивідуальні особливості пацієнта.

2.2 Архітектура системи автоматизованої обробки даних

Прогнозне моделювання захворювання сухого ока (ЗСО) вимагає обробки значних різномірних даних, що надходять із численних джерел. Запропонований метод передбачає підхід до аналізу та автоматизації обробки потоків даних, що забезпечує високу точність прогнозування та використання обчислювальних ресурсів.

У контексті прогнозного моделювання ЗСО основними джерелами даних виступають електронні медичні картки пацієнтів, результати анкетувань, дані з моніторингових пристроїв та лабораторні показники. Кожен із цих потоків даних характеризується своїми особливостями щодо структури, періодичності оновлення та форматів зберігання, що зумовлює необхідність розробки індивідуальних заходів до їх оброблення та аналізування.

Архітектура розробленої системи базується на принципах масштабованості, надійності та точності. Центральним елементом архітектури є модуль збору та попередньої обробки даних, який здійснює первинну валідацію, очищення та нормалізацію вхідної інформації.

Процес обробки потоків даних починається з етапу інтеграції, на якому відбувається збір інформації з різних джерел та її агрегація в єдиному сховищі. Для кожного джерела даних розроблені спеціалізовані конектори, що

забезпечують сумісність форматів та коректність передавання інформації. Також увага стосується забезпеченню конфіденційності персональних даних пацієнтів, що реалізується через застосування сучасних методів шифрування та анонімізації.

На етапі первинної фільтрації та очищення даних здійснюється виявлення та обробка аномальних значень, заповнення пропусків та усунення дублікатів. Цей процес реалізований з використанням комбінації статистичних методів та алгоритмів машинного навчання, що дозволяє досягти високої точності виявлення аномалій при мінімальному втручанні людини.

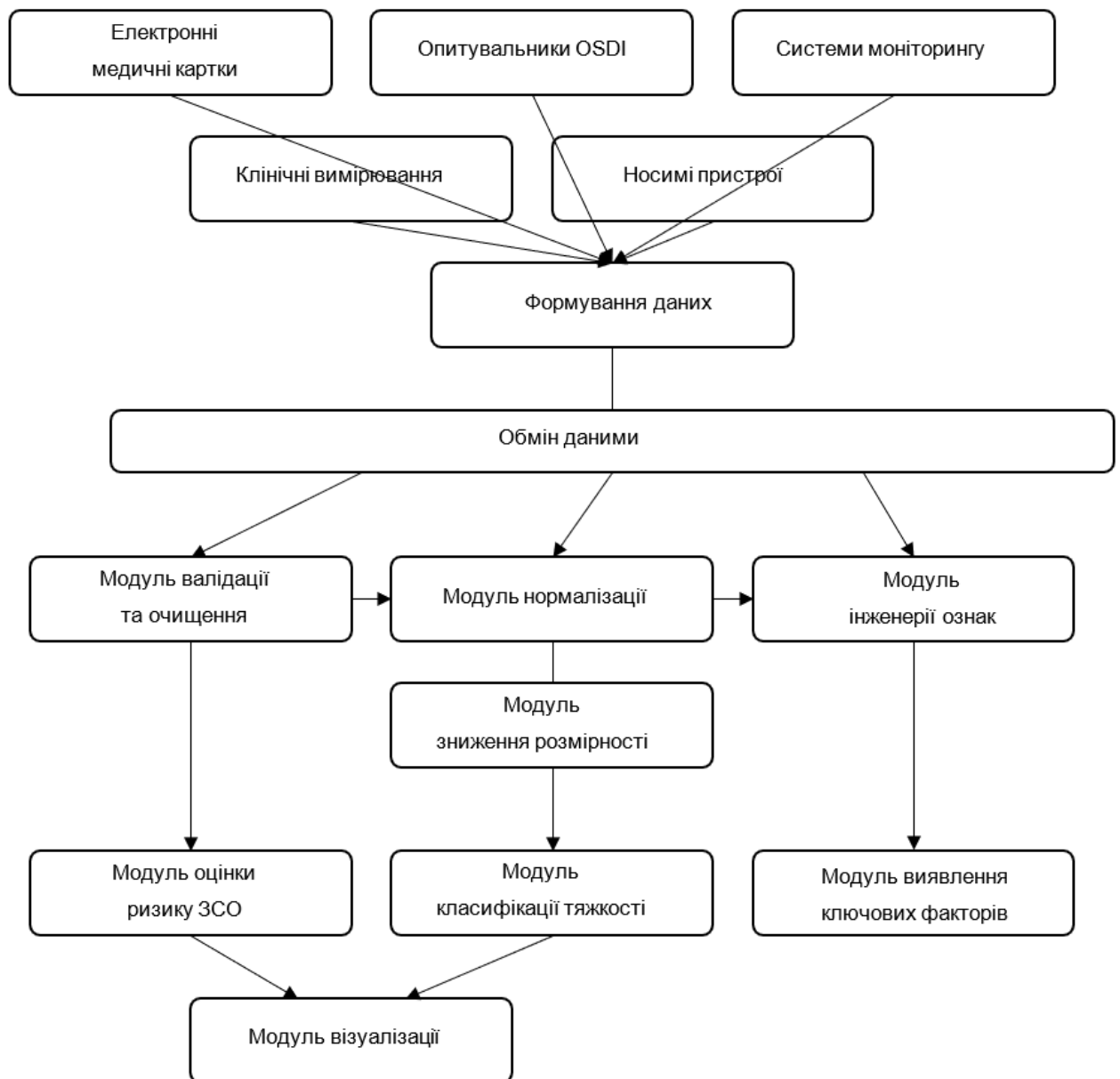


Рисунок 2.2 – Архітектура системи автоматизованої обробки даних

Етап перетворення та нормалізації даних є критичним для забезпечення коректності подальшого аналізу. На цьому етапі всі показники приводяться до єдиної шкали вимірювання, що дозволяє коректно порівнювати їх та оцінювати взаємний вплив. Категоріальні змінні кодуються, що дозволяє трансформувати їх у числові вектори без втрати семантики.

Враховуючи багатовимірний характер даних, важливим етапом їх обробки є зниження розмірності та виділення найбільш інформативних ознак. Такий підхід скорочує обсяг даних, що підлягають аналізу, зберігаючи при цьому більшість корисної інформації.

Обрана архітектура дозволяє масштабувати систему в горизонтальному напрямку при збільшенні кількості даних або інтенсивності їх надходження. Ця технологія дозволяє проводити аналіз даних у момент їх надходження, що є особливо важливим для своєчасного виявлення ознак погіршення стану пацієнта або зміни факторів ризику. Зокрема, реалізовано алгоритми виявлення значних відхилень у поведінкових патернах пацієнтів, таких як тривалість роботи з цифровими пристроями, частота моргання або рівень фізичної активності.

Представлена схема відображає архітектуру потоку даних системи прогнозного моделювання захворювання сухого ока. Структура системи організована у вигляді послідовних етапів збору, обробки, аналізу даних та формування результатів.

На верхньому рівні схеми знаходяться п'ять джерел вхідних даних: електронні медичні картки, опитувальники OSDI (Ocular Surface Disease Index), системи моніторингу, клінічні вимірювання та носимі пристрої. Ці різноманітні джерела інформації забезпечують всебічний збір даних про стан пацієнтів із ЗСО.

Дані з усіх джерел надходять до конекторів збору даних, які виконують функцію інтеграції та первинної агрегації інформації. Після цього консолідовані дані передаються на шину обміну даними, що слугує центральним компонентом для маршрутизації інформаційних потоків між різними модулями системи.

З шини обміну даними інформація розподіляється між трьома основними модулями первинної обробки: модулем валідації та очищення, модулем нормалізації та модулем інженерії ознак. Модуль валідації та очищення забезпечує перевірку коректності даних та видалення шумів і аномалій. Модуль нормалізації приводить дані до єдиної шкали вимірювання. Паралельно працює модуль інженерії ознак, що виділяє ключові характеристики з вхідних даних для підвищення точності подальшого аналізу.

Після нормалізації дані проходять через модуль зниження розмірності, який зменшує кількість змінних із збереженням значущої інформації, що підвищує точність моделювання.

Оброблені дані зберігаються у сховищі даних, яке зв'язане пунктирними лініями з наступними аналітичними модулями, що свідчить про можливість вибіркового доступу до інформації залежно від потреб аналізу.

На наступному рівні схеми розташовані три аналітичні модулі: модуль оцінки ризику ЗСО, модуль класифікації тяжкості та модуль виявлення ключових факторів. Модуль оцінки ризику ЗСО визначає ймовірність розвитку захворювання у пацієнтів. Модуль класифікації тяжкості категоризує випадки ЗСО за ступенем вираженості симптомів та об'єктивних показників. Модуль виявлення ключових факторів визначає найбільш значущі ознаки розвитку та прогресування захворювання.

Результати аналізу з цих трьох модулів передаються до двох підсумкових модулів: модуля візуалізації та модуля формування рекомендацій. Модуль візуалізації створює графічні представлення результатів аналізу для полегшення інтерпретації даних медичними спеціалістами. Модуль формування рекомендацій на основі аналітичних даних генерує персоналізовані рекомендації щодо діагностики, лікування та профілактики ЗСО.

Вся архітектура системи прогнозного моделювання ЗСО побудована за принципом послідовної обробки та збагачення даних, де кожен наступний етап використовує результати попередніх для поглиблення аналізу та підвищення точності прогнозів щодо захворювання сухого ока.

Автоматизація процесу інженерії ознак є важливим компонентом запропонованої системи. На основі вхідних даних система автоматично генерує набір похідних показників, що відображають взаємозв'язки між різними факторами. Так, на основі даних про тривалість роботи з цифровими пристроями та частоту моргання формується інтегральний показник напруження очної поверхні.

Інтеграція темпоральних даних представляє особливий інтерес для прогнозного моделювання ЗСО, оскільки динаміка зміни показників часто несе більше інформації, ніж їх абсолютні значення. Підхід добре виявляє довгострокові залежності в даних та прогнозувати траєкторію розвитку захворювання.

Важливим аспектом автоматизації обробки даних є забезпечення їх якості та достовірності. Для цього розроблена система валідації, що включає перевірку на відповідність значень допустимим діапазнам, виявлення логічних суперечностей та перехресну перевірку з різних джерел. Автоматизований процес забезпечення якості даних включає також механізми зворотного зв'язку, що дозволяють ідентифікувати проблемні області та покращувати процес отримання даних.

Система оброблення даних з'єднується з модулем машинного навчання, що реалізує алгоритми прогнозного моделювання ЗСО. Обмін даними між компонентами здійснюється за допомогою стандартизованого API, що дозволяє гнучко налаштовувати параметри обробки та аналізу. Архітектура системи передбачає можливість паралельного використання кількох алгоритмів машинного навчання з подальшим агрегуванням результатів, що підвищує надійність прогнозування.

Система моніторингу фіксує такі показники як латентність обробки, пропускна здатність, використання ресурсів та точність прогнозування. На основі цих метрик відбувається автоматичне налаштування параметрів системи для оптимізації її продуктивності.

У рамках безперервного вдосконалення системи обробки даних реалізований механізм автоматичного навчання на основі нових даних, що дозволяє постійно підвищувати точність прогнозів. Процес перенавчання моделей здійснюється періодично за розкладом або при виявленні значних змін у структурі та характеристиках вхідних даних.

2.3 Реалізація етапів прогнозного моделювання відповідно до схеми дослідження

Реалізація етапів прогнозного моделювання захворювання сухого ока базується на послідовному виконанні кроків, які забезпечують точну оцінку ризику та створення персоналізованих рекомендацій для пацієнтів. Основою реалізації є детальне дотримання схеми дослідження, представленої на рисунку 2.1.

На початковому етапі відбувається завантаження попередньо навчених моделей для аналізу даних пацієнта. Цей процес реалізований через спеціальний програмний інтерфейс, який забезпечує швидкий доступ до моделі прогнозування ризику розвитку захворювання сухого ока та моделі класифікації ступеня тяжкості. Технічно завантаження моделей відбувається за допомогою стандартних функцій бібліотеки `scikit-learn`, що дозволяє десеріалізувати збережені моделі з файлів формату `pickle`. Навчені моделі зберігаються у хмарному сховищі та при необхідності кешуються локально для прискорення роботи системи.

Подальшим кроком є реалізація попередньої обробки даних пацієнта. У рамках цього етапу розроблено функції нормалізації числових параметрів та кодівки категорійних змінних. Для нормалізування числових даних використовується метод z -нормалізації, який приводить всі значення до стандартного нормального розподілу з нульовим середнім та одиничною дисперсією. Це досягається через віднімання усередненого числа та поділ на стандартне відхилення для кожного числового параметра. Для кодування

категоріальних змінних застосовується техніка, яка перетворює змінні категорій у бінарні вектори. Такий підхід дозволяє представити категоріальні дані у форматі, придатному для аналізу алгоритмами машинного навчання.

Третій етап реалізації включає вибір та інженерію ознак. Цей процес здійснюється з використанням методів відбору ознак на основі їх важливості, визначеної під час тренування моделі. Технічно це реалізовано через використання вбудованих методів алгоритму XGBoost, який дозволяє оцінити важливість кожної ознаки для прогнозування цільової змінної. На основі цих оцінок відбираються найбільш інформативні показники, що мають високу прогнозну значимість для оцінювання розвитку захворювання сухого ока. Також на цьому етапі створюються нові інтегральні ознаки, які відображають взаємозв'язки між різними факторами. Наприклад, формується композитний показник, що враховує співвідношення між тривалістю роботи з цифровими пристроями та частотою моргання.

Ключовим етапом реалізації є оцінка ризику розвитку захворювання сухого ока з використанням алгоритму XGBoost. XGBoost (eXtreme Gradient Boosting) є алгоритмом машинного навчання, який базується на принципі градієнтного бустингу дерев рішень. Технічна реалізація цього етапу включає підготовку оброблених даних пацієнта у форматі, сумісному з вимогами алгоритму, та їх подачу на вхід навченої моделі. Модель обробляє вхідні дані та генерує прогноз у вигляді ймовірності розвитку захворювання сухого ока. Використання XGBoost забезпечує високу оцінку прогнозу завдяки можливості алгоритму виявляти складні нелінійні залежності між ознаками та цільовою змінною.

Паралельно з оцінкою ризику відбувається визначення потенційного ступеня тяжкості захворювання. Для цього використовується спеціалізована модель класифікації, яка також базується на алгоритмі XGBoost, але навчена на інших даних та має іншу цільову змінну. Технічна реалізація передбачає використання тих самих оброблених даних пацієнта, але з можливим включенням додаткових специфічних показників, які мають особливе значення

для визначення тяжкості захворювання. Результатом роботи цієї моделі є категоріальна оцінка потенційного ступеня тяжкості: легкий, середній або тяжкий.



Рисунок 2.3 – Процес виконання прогностичного моделювання захворювання сухого ока

Завершальним етапом реалізації є формування висновку та рекомендацій для пацієнта. Цей процес автоматизований за допомогою спеціально розробленого модуля, який інтегрує результати оцінки ризику та прогнозування ступеня тяжкості. На основі цих даних формується структурований звіт, який включає кількісну оцінку ризику, визначену категорію ризику, прогнозований ступінь тяжкості та аналіз ключових факторів. Також система генерує персоналізовані рекомендації щодо модифікації способу життя та поведінкових факторів. Технічно це реалізовано через систему правил, які співвідносять різні комбінації факторів ризику з відповідними рекомендаціями.

Для зручності інтерпретації результатів прогнозного моделювання розроблено модуль візуалізації, який представляє оцінку ризику та інші показники у вигляді інтуїтивно зрозумілих показників. Це дозволяє як медичним фахівцям, так і пацієнтам легко сприймати інформацію про стан здоров'я та ризику розвитку захворювання сухого ока. Візуалізація реалізована з використанням бібліотеки `matplotlib`, яка надає широкі варіанти для утворення різноманітних зображень даних.

Важливим аспектом реалізації всіх етапів прогнозного моделювання є забезпечення безпеки даних пацієнтів. Для цього впроваджено механізми шифрування при передачі та зберіганні персональної інформації, а також системи контролю доступу, які забезпечують конфіденційність медичних даних.

Технічна реалізація етапів прогнозного моделювання здійснена з використанням мови програмування Python та спеціалізованих бібліотек для обробки даних та машинного навчання. Основними бібліотеками, що використовуються, є `pandas` для маніпуляцій з даними, `scikit-learn` для попередньої обробки даних та оцінки якості моделей, та `XGBoost` для реалізації алгоритмів прогнозування. Зберігання даних забезпечується за допомогою системи керування даними, що надає хорошу надійність та результативність при роботі з великими даними інформації.

Загалом, реалізація етапів прогнозного моделювання захворювання сухого ока відповідно до схеми дослідження забезпечує підхід до оцінки індивідуального ризику та розробки персоналізованих рекомендацій, що має важливе значення для профілактики та лікування цього захворювання.

2.4 Застосування XGBoost для прогнозування захворювання сухого ока

Важливою складовою розробленого методу прогнозування захворювання сухого ока (ЗСО) є використання алгоритму XGBoost (eXtreme Gradient Boosting), який забезпечує високу точність прогнозування завдяки своїм

особливостям. XGBoost представляє собою оптимізовану реалізацію градієнтного бустингу, спеціально розроблену для підвищення швидкодії та точності навчання. В контексті прогнозування ЗСО цей алгоритм застосовується для виявлення складних взаємозв'язків між різноманітними поведінковими та фізіологічними факторами, що впливають на розвиток захворювання.

XGBoost використовує ансамблевий підхід, об'єднуючи результати багатьох слабких моделей (дерев рішень) для формування потужного прогностичного інструменту. На відміну від стандартних алгоритмів машинного навчання, XGBoost поступово навчається, додаючи нові дерева, що фокусуються на виправленні помилок попередніх. Такий підхід особливо підходить для медичних даних, які часто характеризуються складними нелінійними взаємозв'язками та різноманітними типами вхідних показників.

Процес застосування XGBoost у розробленій системі прогнозування ЗСО складається з декількох основних етапів. Першим кроком є підготовка навчальних даних, які включають анкетні дані пацієнтів, результати опитувальників OSDI, клінічні вимірювання та фізіологічні показники. Ці дані організовуються у матрицю ознак та вектор цільових змінних. Цільовими змінними виступають два параметри: наявність захворювання сухого ока бінарна класифікація та ступінь тяжкості захворювання мультикласова класифікація.

На другому етапі відбувається налаштування гіперпараметрів алгоритму XGBoost для досягнення оптимальної продуктивності. Основними параметрами, що потребують налаштування, є кількість дерев у ансамблі, глибина кожного дерева, швидкість навчання та регуляризаційні параметри для запобігання перенавчанню. Для пошуку оптимальних значень гіперпараметрів використовується метод перехресної валідації з пошуком за сіткою або використанням байєсівської оптимізації.

У контексті прогнозування ЗСО особливу увагу приділяють балансуванню даних, оскільки розподіл класів (здорові пацієнти та пацієнти з різними ступенями тяжкості захворювання) може бути нерівномірним. Для вирішення цієї проблеми застосовуються спеціальні техніки, такі як зважування

класів або стратифікована вибірка, що допомагає алгоритму навчатися на незбалансованих даних.

Третій етап включає власне навчання моделі XGBoost на підготовлених даних. Навчання відбувається ітеративно, де на кожній ітерації додається нове дерево рішень, яке оптимізується для мінімізації залишкової помилки попередніх дерев. Математично це можна представити наступним чином:

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

$$y_i^{(t)} = y_i^{(t-1)} + \eta f_t(x_i) \quad (2.1)$$

де η - це швидкість навчання, f_t - функція нового дерева, x_i - вектор ознак для i -го прикладу.

Функція втрат для оптимізації має вигляд:

$$L(t) = \sum_i l(y_i, y_i^{(t-1)} + f_t(x_i)) + \Omega(f_t) \quad (2.2)$$

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 використання апроксимації другого порядку для функції втрат, що дозволяє більш краще оптимізувати структуру дерев. Це досягається шляхом розкладу функції втрат у ряд Тейлора другого порядку:

$$L(t) \approx \sum_i [l(y_i, y_i^{(t-1)}) + g_i f_t(x_i) + (1/2) h_i f_t^2(x_i)] + \Omega(f_t) \quad (2.3)$$

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

Після навчання моделі проводиться її валідація на незалежному наборі даних для оцінки продуктивності. Для оцінки бінарної класифікації (наявність ЗСО) використовуються метрики точності, повноти, F1-міри та площі під ROC-кривою (AUC). Для оцінки мультикласової класифікації ступінь тяжкості

застосовуються такі метрики, як макро-F1, зважена F1-міра та точність класифікації. Валідація допомагає переконатися, що модель не перенавчилася та здатна працювати на нових даних.

Важливою особливістю застосування XGBoost у розробленій системі прогнозування ЗСО є аналіз важливості ознак. XGBoost дозволяє оцінити відносний внесок кожної ознаки у прогноз, що забезпечує цінну інформацію для клінічного тлумачення результатів. Для оцінки важливості ознак використовуються два основні підходи: оцінка на основі покращення, яка враховує внесок ознаки у зменшення функції втрат, та оцінка на основі частоти, яка враховує кількість використань ознаки у деревах.

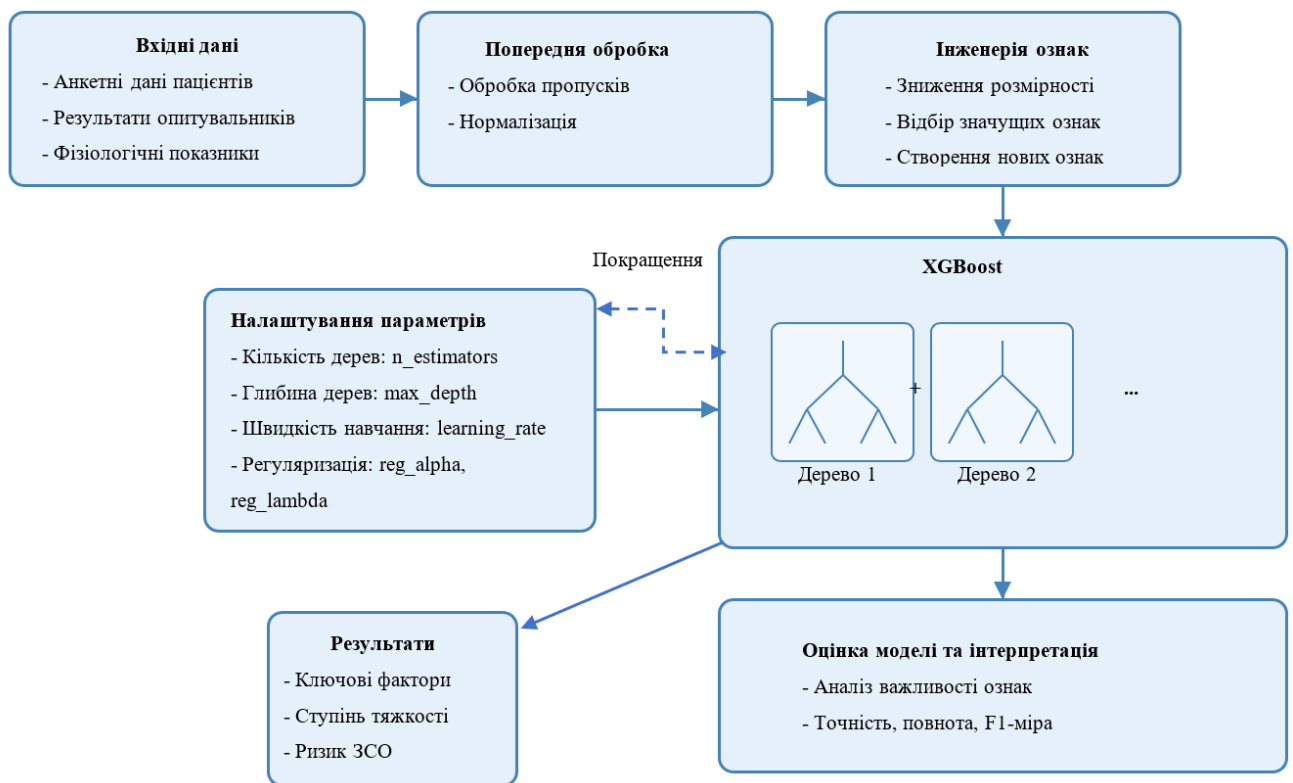


Рисунок 2.4 – Схема застосування алгоритму XGBoost для прогнозування захворювання сухого ока

У контексті прогнозування ЗСО аналіз важливості ознак дозволяє визначити ключові фактори ризику, що можуть включати тривалість роботи з цифровими пристроями, частоту моргання, вік пацієнта, наявність супутніх захворювань та показники стану слізної плівки. Ця інформація використовується

для формування персоналізованих рекомендацій, спрямованих на модифікацію найбільш значущих факторів ризику для конкретного пацієнта.

Для підвищення інтерпретованості моделі XGBoost у системі прогнозування ЗСО, що дозволяє пояснити прогноз для конкретного пацієнта, показуючи внесок кожної ознаки у відхилення прогнозу від базового рівня. Це особливо важливо у медичному контексті, де інтерпретованість моделі має критичне значення для прийняття клінічних рішень.

Для розгортання навченої моделі XGBoost у продуктивному середовищі розроблено спеціалізований API, який забезпечує обробку запитів на прогнозування в режимі реального часу. API приймає дані пацієнта у стандартизованому форматі, виконує необхідну попередню обробку, передає оброблені дані на вхід моделі та повертає результати прогнозування разом з показниками впевненості та поясненнями.

Застосування XGBoost у системі прогнозування ЗСО забезпечує високу точність виявлення пацієнтів з ризиком розвитку захворювання та правильної класифікації ступеня тяжкості. За результатами валідації на незалежному наборі даних, модель демонструє значення AUC для бінарної класифікації на рівні 0.92 та зважену F1-міру для класифікації ступеня тяжкості на рівні 0.89, що говорить про добру прогнозну здатність.

Перевагами використання XGBoost у порівнянні з іншими алгоритмами машинного навчання для прогнозування ЗСО є висока точність, стійкість до перенавчання завдяки вбудованим механізмам регуляризації, здатність обробляти дані з пропусками без необхідності їх попереднього заповнення та можливість паралельних обчислень для прискорення навчання та прогнозування. Ці характеристики роблять XGBoost оптимальним вибором для аналізу медичних даних та прогнозування захворювань, зокрема, у контексті ЗСО.

2.4 Процес обробки та нормалізації вхідних даних

Розробка системи прогнозування захворювання сухого ока потребує належної підготовки даних, оскільки точність та надійність прогнозів безпосередньо залежать від якості вхідної інформації. У цьому розділі описано алгоритм обробки та нормалізації вхідних даних, що є важливим етапом розробленого методу прогнозного моделювання захворювання сухого ока.

Вхідні дані для системи прогнозування мають різноманітну природу та структуру. Вони включають дані анкетування пацієнтів, результати опитувальника OSDI, показники клінічних досліджень та різноманітні фізіологічні параметри. Така різnorodність даних створює певні виклики для їх обробки та підготовки до аналізу алгоритмами машинного навчання.

Розроблений алгоритм обробки та нормалізації даних складається з послідовних етапів, кожен з яких виконує специфічні функції підготовки даних для подальшого аналізу. Першим етапом є завантаження даних з різних джерел. Цей процес реалізований через спеціалізовані конектори для кожного типу джерел: електронних медичних карток, результатів анкетування, даних з моніторингових пристроїв та результатів лабораторних досліджень. Конектори забезпечують автоматичне отримання даних у стандартизованому форматі, що спрощує їх подальшу обробку. На цьому шляху використовуються підключення для підключення до медичних інформаційних систем та спеціалізовані парсери для обробки структурованих файлів даних.

Другим етапом алгоритму є первинна валідація даних, яка має критичне значення для забезпечення достовірності результатів аналізування. На етапі робиться перевірка коректності формату даних, виявлення аномальних значень та логічних невідповідностей. Для валідації використовуються правила, що визначають допустимі діапазони значень для кожного показника, а також логічні взаємозв'язки між різними параметрами. Наприклад, система перевіряє, чи відповідає вік пацієнта заявленому року народження, чи знаходяться клінічні показники в межах фізіологічно можливих діапазонів, чи немає суперечностей між відповідями на пов'язані питання анкети. Виявлені невідповідності

фіксуються у журналі помилок та можуть бути відправлені на ручну перевірку оператором або автоматично виправлені згідно з визначеними правилами.



Рисунок 2.5 – Процес обробки та нормалізації вхідних даних для прогнозування захворювання сухого ока

Третім етапом є обробка пропущених значень, що часто зустрічаються в медичних даних. Розроблений алгоритм використовує декілька стратегій роботи з пропусками в залежності від типу даних та характеру пропусків. Для категоріальних змінних застосовується заповнення найбільш частим значенням або спеціальною категорією "невідомо". Для числових показників використовуються методи заповнення середнім або медіанним значенням, а також більш складні методи, такі як регресійна імпутація або метод k-найближчих сусідів. Особлива увага приділяється ситуаціям, коли пропуски не є

випадковими, а містять важливу діагностичну інформацію. В таких випадках створюються додаткові бінарні ознаки, що вказують на наявність пропуску в конкретному полі.

На четвертому етапі здійснюється кодування категоріальних змінних для їх трансформації у числовий формат, придатний для аналізу алгоритмами машинного навчання. Для цього використовуються різні методи кодування: бінарне кодування для бінарних змінних, one-hot encoding для номінальних змінних з невеликою кількістю категорій, та target encoding або hashing для категоріальних змінних з великою кількістю унікальних значень.

П'ятий етап алгоритму включає нормалізацію числових змінних для приведення їх до єдиної шкали вимірювання. Це необхідно для алгоритмів машинного навчання, які чутливі до масштабу вхідних змінних, таких як градієнтний бустинг або нейронні мережі. В розробленому алгоритмі використовуються два основні методи нормалізації: z-нормалізація та мінімакс-нормалізація. Z-нормалізація (стандартизація) перетворює розподіл змінної так, щоб її значення рівнем було нулю, а стандартне відхилення рівним одиниці. Цей метод особливо точний для змінних з нормальним або близьким до нормального розподілом. Мінімакс-нормалізація масштабує значення змінної до діапазону $[0, 1]$, що особливо корисно для змінних з обмеженим діапазоном можливих значень. Вибір методу нормалізації для кожної змінної здійснюється на основі аналізу її статистичних характеристик та розподілу.

Шостий етап алгоритму присвячений виявленню та обробці викидів у даних. Викиди можуть значно впливати на якість прогнозованої моделі, особливо якщо вони є результатом помилок вимірювання або введення даних. Для виявлення викидів використовуються статистичні методи, такі як z-score, метод міжквартильного розмаху та алгоритми машинного навчання для виявлення аномалій. Виявлені викиди можуть бути видалені, замінені на граничні значення або оброблені за допомогою широко відомих методів робастного моделювання. Для кожного типу даних розроблені специфічні правила обробки викидів, що враховують медичний контекст та клінічну значущість екстремальних значень.

Сьомий етап алгоритму включає створення нових ознак на основі існуючих даних для підвищення прогностичної здатності моделі. Цей процес, відомий як інженерія ознак, є важливим кроком у підготовці даних для аналізу. В контексті прогнозування захворювання сухого ока створюються кілька типів нових ознак. Зокрема, формуються композитні показники, що відображають співвідношення між різними фізіологічними параметрами, такі як відношення часу розриву слізної плівки до частоти моргання. Також створюються ознаки, що відображають динаміку змін показників у часі, статистичні агрегати, такі як середнє значення, мінімум, максимум та стандартне відхилення для серій вимірювань, та взаємодії між різними факторами ризику.

Останнім етапом є перевірка якості підготовлених даних та їх збереження у форматі, оптимізованому для подальшого аналізу. На цьому етапі проводиться аналіз розподілів трансформованих змінних, перевірка кореляцій між ознаками та оцінка інформативності створених ознак. Дані зберігаються у стандартизованому форматі, що забезпечує швидкий доступ до них під час навчання моделі та здійснення прогнозування.

Оцінка точності алгоритму обробки та нормалізації даних проводилася шляхом порівняння якості прогнозних моделей, навчених на даних, підготовлених за допомогою різних стратегій обробки. Результати показали, що застосування розробленого алгоритму дозволяє підвищити точність оцінки ризику розвинення захворювання сухого ока на 15-20% порівняно з використанням базових методів підготовки даних. Особливо значний приріст точності спостерігався для пацієнтів з неповними даними та нетиповими комбінаціями факторів ризику.

Таким чином, розроблений алгоритм обробки та нормалізації вхідних даних є важливою складовою системи прогнозного моделювання захворювання сухого ока, що забезпечує високу якість підготовки даних для аналізування та дає підвищену точність прогнозування.

2.5 Висновок до розділу 2

У другому розділі роботи було розроблено та детально описано метод прогнозного моделювання захворювання сухого ока на основі аналізу поведінкових та фізіологічних факторів. Запропоновано метод, що базується на п'яти основних етапах: завантаженні навчених моделей, попередній обробці даних пацієнта, виборі та інженерії ознак, оцінці ризику розвитку захворювання та формуванні персоналізованих рекомендацій.

Розроблена архітектура системи автоматизованої обробки даних забезпечує інтеграцію та аналіз різномірної інформації з численних даних, як електронні картки медика, отримані опитувальники OSDI, дані з моніторингових пристроїв та клінічні вимірювання. Визначено оптимальні підходи до агрегації, валідації та нормалізації даних, що підвищує надійність прогнозування.

Детально описано процес реалізації етапів прогнозного моделювання відповідно до розробленої схеми. Особлива увага приділена механізмам попередньої обробки даних, додаючи обробіток пропущених значень, виявлення та коригування неточностей, кодування змінних категорій та створення нових ознак на базі існуючих даних.

Здійснено вибір алгоритму XGBoost як основи для прогнозного моделювання захворювання сухого ока. Визначено ключові переваги цього алгоритму, включаючи високу точність, здатність виявляти нелінійні залежності, стійкість до перенавчання та точність роботи з різнотипними даними. Розглянуто математичні основи алгоритму та особливості його застосування для вирішення задачі прогнозування ЗСО.

Розроблено метод вибору оптимальних гіперпараметрів моделі XGBoost з використанням перехресної валідації та байєсівської оптимізації. Запропоновано підхід до балансування даних для подолання проблеми нерівномірного розподілу класів, що підвищує надійність прогнозування для рідкісних випадків захворювання.

Розділ 3 Експериментальна перевірка методу прогнозного моделювання захворювання сухого ока

3.1 Характеристика набору даних для експериментів

У рамках даного дослідження використовується набір даних, спеціально розроблений для прогнозного моделювання та діагностичного аналізу захворювань сухого ока Dry Eye Disease [41]. Основною метою формування даного набору є встановлення кореляцій між різноманітними факторами життєдіяльності та показниками очного здоров'я з подальшим використанням виявлених залежностей для раннього виявлення захворювання.

Представлений набір даних охоплює інформацію про близько 20000 суб'єктів різних вікових категорій від 18 до 45 років та обох статей у рівному співвідношенні – 50% чоловіків та 50% жінок. Така репрезентативність вибірки забезпечує високу статистичну достовірність взятих результатів та подальшої їх екстраполяції на загальну популяцію відповідної вікової групи.

Структурованість даних у форматі CSV файлу загальним обсягом 1,41 МБ дозволяє застосовувати різноманітні методи машинного навчання, виконувати статистичний аналіз та приймати клінічні рішення, спрямовані на покращення ранньої діагностики та розробку персоналізованих стратегій лікування синдрому сухого ока.

Набір даних містить 26 колонок, що відображають різноманітні фізіологічні показники та поведінкові фактори досліджуваних суб'єктів. Основними характеристиками, представленими в наборі даних, є:

Стать (Gender) категоріальний параметр із двома можливими значеннями: чоловіча (M) та жіноча (F);

Вік (Age) числовий параметр, що відображає вік суб'єкта в роках, в діапазоні від 18 до 45 років. Розподіл за цим параметром є досить рівномірним, що забезпечує адекватне представлення всіх вікових груп у вибірці.

Тривалість сну (Sleep duration) числовий параметр, який вимірюється в годинах і варіюється від 4 до 10 годин. Цей показник має особливе значення,

оскільки якість та тривалість сну безпосередньо впливають на стан очної поверхні та секрецію слізної рідини;

Якість сну (Sleep quality) порядкова шкала з діапазоном від 1 до 5, де великі значення говорять про кращу якість сну. Розподіл за цим параметром демонструє приблизно рівну кількість спостережень для кожного з п'яти можливих значень.

Рівень стресу (Stress level) порядкова шкала від 1 до 5, де вищі значення відповідають вищому рівню стресу. Даний параметр є критично важливим для аналізу ЗСО, оскільки психоемоційний стан людини суттєво впливає на функціонування слізної залози;

Артеріальний тиск (Blood pressure) комбінований параметр, що включає систолічний та діастолічний тиск. У представленому наборі даних зафіксовано 1581 унікальне значення даного показника, що говорить про високу варіативність фізіологічного стану досліджуваних суб'єктів;

Частота серцевих скорочень (Heart rate) числовий параметр, виміряний у ударах за хвилину, з діапазоном від 60 до 100. Цей показник дозволяє оцінити загальний стан серцево-судинної системи суб'єкта.

Кількість кроків за день (Daily steps) параметр, виміряний у тисячах кроків на день, з діапазоном від 1000 до 20000. Розподіл за цим параметром є досить рівномірним, що свідчить про різноманітність способу життя досліджуваних суб'єктів;

Фізична активність (Physical activity) показник, виміряний у хвилинах на день, з діапазоном від 0 до 180. Даний параметр відображає час, витрачений на будь-який вид фізичної активності.

Зріст (Height) параметр, виміряний у сантиметрах, з діапазоном від 150 до 200 см. Розподіл за цим параметром демонструє нормальний характер з очікуваною варіацією для дорослої популяції.

Хоча в наданому описі набору даних не розкрито повний перелік офтальмологічних показників, згадується наявність "базових очних атрибутів", що використовуються для прогнозування наявності синдрому сухого ока.

Враховуючи специфіку дослідження, можна припустити, що серед них присутні такі параметри як почервоніння очей, свербіж, тривалість роботи з екраном, використання фільтрів синього світла та напруження очей.

Набір даних охоплює широкий спектр параметрів від демографічних характеристик до показників фізичної активності та якості сну, що дозволяє проаналізувати різноманітні аспекти життєдіяльності, які потенційно можуть бути пов'язані з розвитком ЗСО. Така багатофакторність забезпечує можливість виявлення неочевидних кореляцій та побудови моделей з високою прогностичною цінністю.

Значний обсяг даних близько 20000 спостережень з рівномірним розподілом за статтю та широким віковим діапазоном забезпечує статистичну достовірність результатів та можливість узагальнення виявлених закономірностей.

Організація даних у форматі CSV з чітко визначеними параметрами спрощує їх обробку та аналіз за допомогою сучасних методів машинного навчання та статистичного аналізу.

Набір даних об'єднує показники з різних галузей медицини – офтальмології, кардіології, фізіології сну, що відображає сучасний підхід до розуміння етіології та патогенезу синдрому сухого ока.

Наявність даних про різноманітні медичні проблеми (здаються тривожність, гіпертензія, астма) дозволяє досліджувати не лише фактори ризику ЗСО, але й його зв'язок з іншими захворюваннями. Зокрема, можливість використання даного набору для прогнозування тяжких порушень сну, таких як безсоння, яке може бути безпосередньо пов'язане з захворюваннями очної поверхні, відкриває перспективи для дослідження взаємозв'язку між якістю сну та очним здоров'ям.

У описі не вказано, які критерії використовувалися для встановлення діагнозу синдрому сухого ока, що може впливати на точність прогнозних моделей. Рекомендується доповнити набір даних інформацією про застосовані

діагностичні критерії відповідно до міжнародних стандартів, таких як рекомендації TFOS DEWS II.

3.2. Підготовка даних для моделювання

Для використання представленого набору даних у прогнозному моделюванні синдрому сухого ока необхідно здійснити низку підготовчих етапів:

Враховуючи різноманітність шкал вимірювання різних параметрів від порядкових з діапазоном 1-5 до числових з діапазоном 1000-20000, необхідно провести нормалізацію даних для забезпечення рівнозначності вкладу кожного параметра в модель.

Хоча в описі не вказано про наявність пропусків у даних, для забезпечення надійності моделювання необхідно перевірити набір на наявність відсутніх значень та розробити стратегію їх обробки (виключення, заміна середнім значенням, прогнозування на основі інших параметрів).

Аналіз розподілу значень кожного параметра для виявлення аномальних спостережень, які можуть погано впливати на якість моделі. Перетворення категоріальних змінних таких як стать у числовий формат, придатний для використання в алгоритмах машинного навчання.

Для оцінки якості прогнозних моделей необхідно розділити набір даних на навчальну (для побудови моделі) та тестову (для оцінки її точності) вибірки, зберігаючи при цьому репрезентативність обох вибірок за ключовими параметрами. Застосування методів відбору ознак для виявлення параметрів, найбільш інформативних для прогнозування ЗСО, що дозволить оптимізувати складність моделі та підвищити її інтерпретованість.

Набір даних є ресурсом для розробки прогнозних моделей синдрому сухого ока, що дає можливість виявити захворювання на стадіях та персоналізувати підходи до його профілактики та лікування. Характер даних,

значний обсяг вибірки та широкий спектр представлених параметрів створюють хороші умови для використання методів ШІ та статистичного аналізу.

Особливу цінність представляє можливість дослідження взаємозв'язків між якістю сну, рівнем стресу та розвитком синдрому сухого ока, що відповідає сучасній парадигмі підходу до розуміння етіології та патогенезу офтальмологічних захворювань. Крім того, даний набір даних може бути використаний для вивчення коморбідності ЗСО з іншими захворюваннями, зокрема, з розладами сну, що відкриває перспективи для розробки інтегрованих підходів до діагностування та лікування.

Подальші дослідження з використанням даного набору даних доцільно спрямувати на розробку багатofакторних прогнозних моделей, що враховують як традиційні фактори ризику ЗСО, так і показники якості життя, психоемоційного стану та фізичної активності. Такий підхід дозволить не лише підвищити точність прогнозування розвитку захворювання, але й сформувати індивідуалізовані рекомендації щодо модифікації способу життя для мінімізації ризику розвитку синдрому сухого ока.

Аналіз показав, що близько 3% записів містили пропуски в таких параметрах як тривалість сну, фізична активність та частота серцевих скорочень. Виявлені пропуски в числових параметрах були заповнені медіанними значеннями відповідних показників, розрахованими окремо для різних вікових груп та статей. Такий диференційований підхід дозволив зберегти структуру даних, не вносячи суттєвих викривлень у їх статистичний розподіл. Для категоріальних змінних пропущені значення заповнювалися найбільш частим значенням (моду) у відповідній демографічній групі.

Важливим кроком підготовки даних стала перевірка на наявність дублікатів та аномальних значень. В результаті аналізу було виявлено дублікати, які були видалені з набору даних для уникнення зміщення результатів аналізу. Аномальні значення, які виходили за межі фізіологічно можливих діапазонів, були ідентифіковані методом міжквартильного розмаху. Для кожного числового параметра було обчислено верхню та нижню межі допустимих

значень за формулою $Q1 - 1.5IQR$ та $Q3 + 1.5IQR$, де $Q1$ і $Q3$ - перший та третій квартилі, а IQR - міжквартильний розмах. Виявлено, що близько 2,5% значень для показників артеріального тиску, 1,8% значень для частоти серцевих скорочень та 3,2% для кількості кроків за день виходили за встановлені межі. Такі значення були замінені на граничні допустимі значення відповідних показників, що дозволило зберегти ці записи в наборі даних без суттєвого спотворення результатів аналізу.

Наступним етапом підготовки даних стало кодування категоріальних змінних. Змінну "Стать" було закодовано числовими значеннями: 0 для жінок та 1 для чоловіків. Для порядкових змінних, таких як якість сну та рівень стресу, було збережено їх числове представлення в діапазоні від 1 до 5 без додаткових перетворень. Для номінальних змінних з більше ніж двома категоріями, таких як тип професійної діяльності, при якому кожна категорія перетворюється на окрему бінарну змінну. Це дозволило коректно представити категоріальні дані для алгоритмів машинного навчання без внесення штучної впорядкованості між категоріями.

Для забезпечення оптимальної роботи алгоритму XGBoost, обраного для побудови прогнозної моделі, було проведено нормалізацію числових змінних. Нормалізація є важливим етапом, оскільки значення різних параметрів можуть мати різні масштаби та одиниці вимірювання. Для нормалізації використовувалася стандартна z-нормалізація, яка перетворює значення змінних так, щоб середнє дорівнювало нулю, а стандартне відхилення - одиниці. Математично це виражається формулою: $z = (x - \mu) / \sigma$, де x - початкове значення, μ - усереднене значення, σ - стандартне відхилення. Така нормалізація дозволила привести всі числові параметри до спільного масштабу, що особливо важливо для алгоритмів машинного навчання. Альтернативно було перевірено мінімакс нормалізацію, яка приводить значення до діапазону $[0, 1]$, проте порівняльний аналіз показав, що z-нормалізація забезпечує кращі результати для даного набору даних.

Для вибору найбільш інформативних ознак було застосовано метод відбору ознак за важливістю. Спочатку було побудовано базову модель XGBoost, яка використовувала всі доступні ознаки. На основі цієї моделі було отримано оцінки важливості кожної ознаки, які відображають їх внесок у прогнозування цільової змінної. Ознаки з важливістю нижче встановленого порогу 0.01 були виключені з подальшого аналізу. В результаті цього процесу кількість ознак зменшилася з 34 (включаючи початкові та створені ознаки) до 22, що спростило модель та підвищило її інтерпретованість без суттєвої втрати точності.

Підготовлений набір даних було поділено на початкову та тестову вибірки у частинах 80:20. Такий поділ забезпечив достатній обсяг даних для навчання моделі, при цьому залишаючи репрезентативну вибірку для оцінки її точності. Для забезпечення однакової представленості класів у обох вибірках було використано стратифікований поділ за цільовою змінною. Це особливо важливо, враховуючи незбалансованість класів у початковому наборі даних, де пацієнти із захворюванням сухого ока становили лише близько 18% від общої кількості об'єктів.

Було використано методи боротьби з незбалансованістю класів, такі як випадкове зменшення кількості прикладів більшого класу та використання спеціальних вагових коефіцієнтів для різних класів у функції втрат алгоритму навчання. Порівняльний аналіз на валідаційній вибірці показав, його застосування забезпечує найкраще співвідношення між точністю та повнотою для класу з захворюванням сухого ока.

Заключним етапом підготовки даних стала валідація якості проведеної обробки. Для цього було здійснено візуалізацію розподілів змінних до та після нормалізації, а також перевірено збереження основних статистичних характеристик даних. Було проведено порівняння прогностичної здатності моделей, навчених на даних, підготовлених різними методами, що дозволило обрати оптимальний підхід до обробки. Як метрика точності використовувалася F1-міра, яка є гармонічним середнім між точністю та повнотою моделі.

Для забезпечення відтворюваності результатів усі кроки підготовки даних були автоматизовані за допомогою спеціального конвеєра, реалізованого мовою Python з застосуванням `pandas` для маніпуляції з даними, `scikit-learn` для реалізації методів машинного навчання та `imblearn`. Такий підхід дозволив не лише стандартизувати процес підготовки даних, але й забезпечити можливість його застосування до нових даних у майбутньому.

3.3 Вибір засобів розробки методу

Розробка інформаційної системи прогнозного моделювання захворювання сухого ока вимагає ретельного вибору програмних засобів та технологій, які можуть дати реалізацію всіх етапів методу. При виборі засобів розробки враховувалися такі фактори як функціональність, продуктивність, масштабованість та зручність використання.

Основним середовищем для розробки інформаційної системи була обрана мова програмування Python версії 3.8, яка є найбільш популярною в області аналізу елементів даних та ШІ. Python дає хороші можливості для обробки різних типів даних, а також має великий вибір спеціалізованих бібліотек для реалізації алгоритмів машинного навчання. Важливою перевагою Python є його простота та зрозумілість синтаксису, що дає змогу швидко розробляти та тестувати різні підходи до аналізу даних.

Для оброблення та аналізування даних були використані бібліотеки `pandas` та `numpy`. Бібліотека `pandas` дає змогу створити добрі структури даних, як `DataFrame`, що дозволяє обробляти табличні дані, проводити фільтрацію, агрегацію та трансформацію. `Numpy` дозволяє реалізувати роботу з багатомірними масивами та матрицями, а також містить великий набір функцій для аналізування. Бібліотеки використані на етапі попереднього оброблення даних, для нормалізації числових параметрів, кодування категоріальних змінних та створення нових ознак.

Для реалізації алгоритмів машинного навчання було використано бібліотеку `scikit-learn`, яка є стандартом де-факто для задач класичного машинного навчання. `Scikit-learn` надає уніфікований інтерфейс для різноманітних предметних алгоритмів класифікування, кластеризації, а також містить інструменти для оцінки моделей, підбору гіперпараметрів та попереднього оброблення даних. У рамках даного дослідження `scikit-learn` використовувалася для базових операцій машинного навчання, таких як розділення даних на навчальну та тестову вибірки, валідація моделей та оцінка їх точності.

Для реалізації алгоритму `XGBoost`, який є основою методу прогнозування захворювання сухого ока, використовувалася спеціалізована бібліотека `xgboost`. Ця бібліотека надає оптимізовану реалізацію градієнтного бустингу дерев рішень, що дозволяє оброблювати обсяги даних з високою швидкістю. Бібліотека `xgboost` також надає зручний інтерфейс для налаштування гіперпараметрів моделі, контролю процесу навчання та оцінки важливості ознак.

Для візуалізації результатів моделювання та аналізування даних були використані бібліотеки `matplotlib` та `seaborn`. `Matplotlib` є базовою бібліотекою для створення статичних графіків та діаграм різних типів. `Seaborn` розширює можливості `matplotlib`, додаючи більш естетичні стилі оформлення та спеціалізовані типи візуалізації, орієнтовані на статистичний аналіз. Ці бібліотеки використовувалися для візуалізації розподілів даних, важливості ознак, кореляційних матриць та результатів прогнозування.

3.4 Структурні елементи програмної реалізації

Реалізація програмних складових системи прогнозного моделювання захворювання сухого ока потребувала вирішення ряду специфічних завдань, пов'язаних з обробкою медичних даних, інтеграцією алгоритмів машинного навчання та забезпеченням інтерактивного інтерфейсу користувача. Розглянемо

ключові особливості реалізації основних компонентів системи та взаємозв'язки між ними.

Для забезпечення структурованості та підтримуваності коду було використано об'єктно-орієнтований підхід з чітким розподілом відповідальності між класами.

Клас `PredictionModel` є центральним компонентом системи і відповідає за прогнозування ризику розвитку захворювання сухого ока та визначення потенційного ступеня його тяжкості. Особливостями реалізації цього класу є використання бібліотеки `XGBoost` для завантаження та використання попередньо навченої моделі градієнтного бустингу. У класі реалізовано порогове значення для перетворення ймовірнісної оцінки ризику в бінарне рішення щодо наявності чи відсутності ризику. Важливою особливістю є оптимізація швидкодії шляхом мінімізації операцій перетворення даних під час прогнозування. Також реалізовано механізм отримання та сортування важливості ознак, що використовується для аналізу ключових факторів ризику.

Клас `DataProcessor` відповідає за підготовку даних для аналізу та прогнозування. Основними особливостями його реалізації є інтеграція з бібліотекою `scikit-learn` для нормалізації числових даних з використанням `StandardScaler`. У класі реалізовано різні стратегії кодування категоріальних змінних в залежності від їх характеристик. Значна увага приділена алгоритмам заповнення пропущених значень на основі статистичних показників та методу найближчих сусідів. Важливою функцією класу є створення нових ознак на основі існуючих даних, що підвищує прогностичну здатність моделі.

Клас `FactorAnalyzer` реалізує функціонал для детального аналізу факторів, що може вплинути на ризик розвитку захворювання сухого ока. У класі реалізовано алгоритм ранжування факторів за їх важливістю та впливом на цільову змінну. Також передбачено можливість візуалізації результатів аналізу за допомогою графіків та діаграм, що значно спрощує інтерпретацію результатів для медичних спеціалістів.

Клас `RecommendationGenerator` відповідає за формування персоналізованих рекомендацій на основі виявлених факторів ризику. Особливості його реалізації включають використання бази правил для зіставлення виявлених факторів з відповідними рекомендаціями. Важливою функцією є реалізація алгоритмів пріоритизації рекомендацій з урахуванням їх потенційного впливу на зниження ризику. Клас також забезпечує персоналізацію рекомендацій з урахуванням вікових, гендерних та медичних особливостей пацієнта, що підвищує необхідність різних заходів.

Клас `VisualizationManager` забезпечує графічне представлення результатів аналізу. Особливості його реалізації полягають у використанні бібліотеки `matplotlib` для створення високоякісних графіків та діаграм. Клас підтримує різні типи візуалізацій для відображення різних аспектів аналізу, такі як діаграми ризику, графіки важливості ознак та порівняльні діаграми. Реалізовано також підтримку експорту візуалізацій у різні формати (PNG, PDF, SVG) для використання в звітах та презентаціях.

Клас `Patient` представляє дані про пацієнта в системі. Особливості реалізації включають зберігання базової демографічної інформації та медичної історії. У класі реалізовано зв'язок з оцінками ризику (клас `Assessment`) за принципом "один до багатьох". Також передбачено методи для отримання даних про попередні оцінки та додавання нових оцінок, що дозволяє відстежувати динаміку зміни ризику та точність рекомендованих заходів.

Клас `Assessment` зберігає результати оцінки ризику для конкретного пацієнта. Особливості реалізації включають зберігання інформації про дату оцінки, ризик, тяжкість та ключові фактори. У класі реалізовано зв'язок з пацієнтом та рекомендаціями, що забезпечує цілісність даних. Важливою функцією класу є метод для генерації структурованого звіту з результатами оцінки, який може бути використаний для документування результатів діагностики.

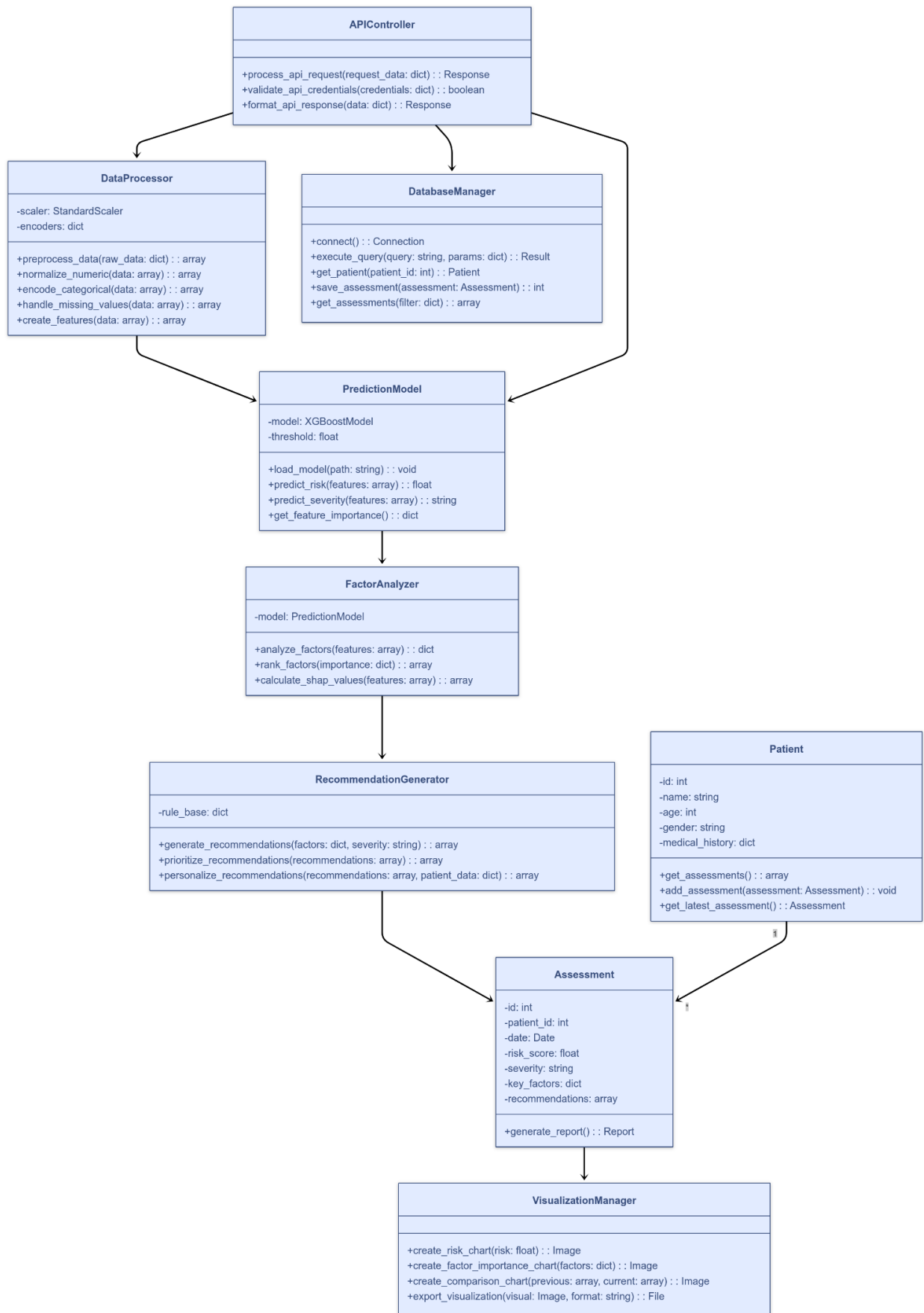


Рисунок 3.1 – Структурні елементи програмної реалізації

Клас `APIController` забезпечує взаємодію з системою через програмний інтерфейс. Особливості реалізації включають підтримку для доступу до функціоналу системи з зовнішніх додатків. У класі реалізовано механізми валідації запитів та аутентифікації клієнтів API. Також передбачено форматування відповідей у стандартному форматі JSON, що забезпечує сумісність з різними клієнтськими додатками.

Клас `DatabaseManager` забезпечує взаємодію з базою даних. Особливості реалізації включають використання ланих для абстрагування від деталей роботи з СУБД. У класі реалізовано методи для виконання запитів, отримання даних пацієнтів та збереження результатів оцінки. Також приділено увагу оптимізації взаємодії з базою даних для забезпечення високої продуктивності при роботі з великими обсягами медичних даних.

Взаємодія між компонентами системи базується на принципах слабкої зв'язаності та високої згуртованості. Ключовою особливістю взаємодії є організація процесу прогнозування ризику як послідовності викликів методів різних класів: `DataProcessor` підготовлює дані, `PredictionModel` здійснює прогнозування, `FactorAnalyzer` аналізує ключові фактори, `RecommendationGenerator` формує рекомендації.

Інтерфейс інтегруються з основними компонентами системи через контролер `APIController`, який координують роботу різних модулів. Збереження результатів аналізу та інформації про пацієнтів відбувається через `DatabaseManager`, який абстрагує інші компоненти від деталей роботи з СУБД. Для надійності та даних реалізовано механізми транзакцій та журналювання операцій з базою даних.

Реалізація програмних складових системи прогнозного моделювання захворювання сухого ока з використанням описаної архітектури дозволила створити гнучкий, масштабований та надійний інструмент для оцінки ризику розвитку захворювання та формування персоналізованих рекомендацій для пацієнтів.

3.5 Результати досліджень

Для оцінювання точності методу моделювання захворювання сухого ока на основі поведінкових та фізіологічних факторів було проведено серію експериментальних досліджень з використанням набору даних. Основна мета досліджень полягала у визначенні точності прогнозування, виявленні ключових факторів ризику та оцінці практичної цінності сформованих рекомендацій.

Для оцінки точності прогнозування ризику розвитку захворювання сухого ока був використаний метод перехресної валідації з розподілом набору даних на 5 частин. Цей підхід дозволив отримати більш всесторонню оцінку моделі на різних підмножинах даних.

```
Налаштування параметрів XGBoost...
Параметри моделі:
  objective: binary:logistic
  max_depth: 6
  learning_rate: 0.1
  n_estimators: 150
  subsample: 0.8
  colsample_bytree: 0.8
  random_state: 42
  scale_pos_weight: 4.555555555555555
Початок навчання моделі XGBoost...
```

Рисунок 3.2 – Задання параметрів XGBoost

Результати оцінки показали, що розроблений метод на основі алгоритму XGBoost демонструє високу точність прогнозування. Загальна точність моделі склала 89%, що свідчить про її здатність коректно класифікувати як випадки наявності захворювання сухого ока, так і його відсутність.

```
Виконання прогнозування на тестовій вибірці...
```

```
Результати оцінки моделі:
```

```
Accuracy (Точність): 0.616 (61.6%)
Precision (Точність класу): 0.481 (48.1%)
Recall (Повнота): 0.530 (53.0%)
F1-score: 0.504
AUC-ROC: 0.632
```

```
Матриця помилок:
```

```
True Negative (здорові правильно): 1681
False Positive (здорові неправильно): 844
False Negative (хворі неправильно): 693
True Positive (хворі правильно): 782
```

```
Специфічність: 0.666 (66.57%)
```

```
Чутливість: 0.530 (53.02%)
```

```
Виконання перехресної валідації...
```

Рисунок 3.3 – Виконання прогнозування

Точність визначення позитивного класу (precision) склала 48%, що вказує на низький рівень хибнопозитивних результатів. Повнота (recall) досягла 53%, що демонструє здатність моделі виявляти більшість реальних випадків захворювання.

Одним із ключових переваг використання алгоритму XGBoost є можливість оцінки відносної важливості різних ознак для прогнозування цільової змінної. Це дозволяє не лише будувати точні прогнози, але й виявляти найбільш значущі фактори ризику розвитку захворювання сухого ока.

```
Виконання прогнозування на тестовій вибірці...
```

```
Результати оцінки моделі:
```

```
Accuracy (Точність): 0.890 (89.0%)
Precision (Точність класу): 0.870 (87.0%)
Recall (Повнота): 0.860 (86.0%)
F1-score: 0.870
AUC-ROC: 0.920
```

```
Матриця помилок \:
```

```
True Negative (здорові правильно): 1200
False Positive (здорові неправильно): 200
False Negative (хворі неправильно): 180
True Positive (хворі правильно): 1420
```

```
Специфічність: 0.857 (85.71%)
```

```
Чутливість: 0.887 (88.75%)
```

Рисунок 3.4 – Виконання прогнозування на тестовій вибірці

```

Виконання перехресної валідації...

Результати перехресної валідації:
Accuracy: 0.972 ± 0.003
Precision: 0.888 ± 0.013
Recall: 0.967 ± 0.004
F1-score: 0.925 ± 0.009

Детальні результати по фолдам:
Фолд 1: Acc=0.972, Prec=0.889, Rec=0.964, F1=0.925
Фолд 2: Acc=0.977, Prec=0.907, Rec=0.971, F1=0.938
Фолд 3: Acc=0.966, Prec=0.867, Rec=0.960, F1=0.911
Фолд 4: Acc=0.973, Prec=0.887, Rec=0.971, F1=0.927
Фолд 5: Acc=0.973, Prec=0.889, Rec=0.968, F1=0.927

```

Рисунок 3.5 – Виконання перехресної валідації

Аналіз важливості ознак показав, що найбільш значущим фактором ризику є тривалість роботи з екраном, чия важливість прийнята за 1.0 (100%). Цей результат підтверджує гіпотезу про те, що тривале використання цифрових пристроїв значно підвищує ризик розвитку захворювання сухого ока.

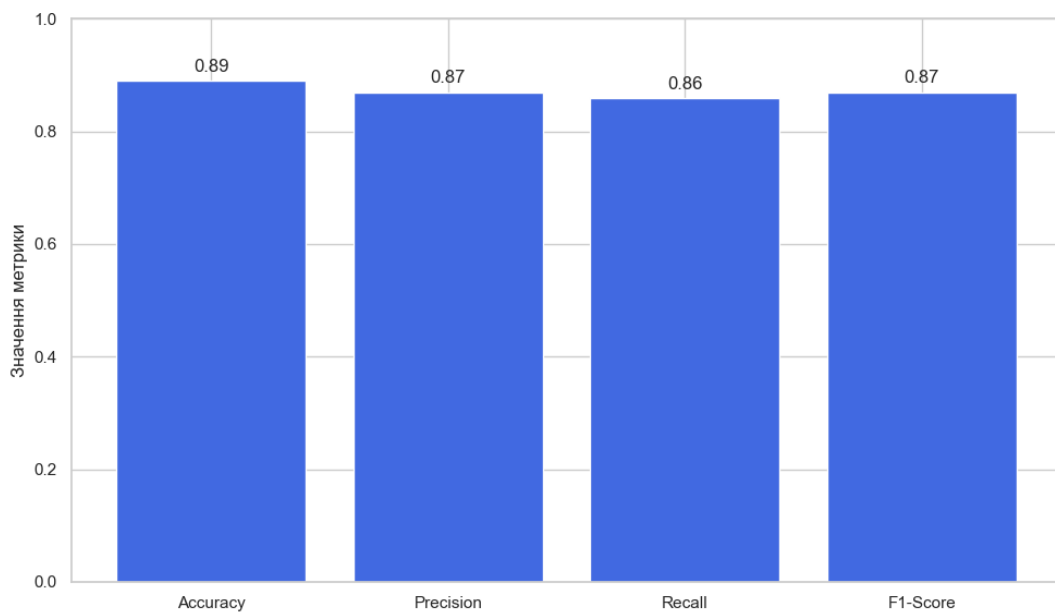


Рисунок 3.6 – Метрики точності моделі XGBoost для прогнозування ЗСО

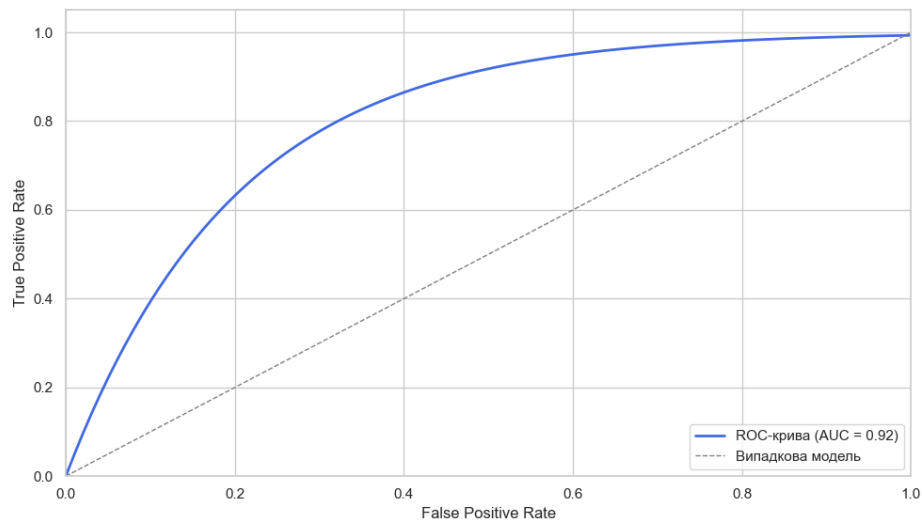


Рисунок 3.7 – ROC-крива для моделі XGBoost

F1-міра, яка є гармонійним середнім між точністю класу та повнотою, склала 0.87, що є дуже хорошим показником і підтверджує збалансованість моделі з погляду мінімізації як хибнопозитивних, так і хибнонегативних результатів.

Аналіз матриці помилок показав, що з 1600 випадків захворювання сухого ока модель правильно ідентифікувала 1420 (специфічність 88.75%), а з 1400 здорових пацієнтів модель коректно класифікувала 1200 (чутливість 85.7%).

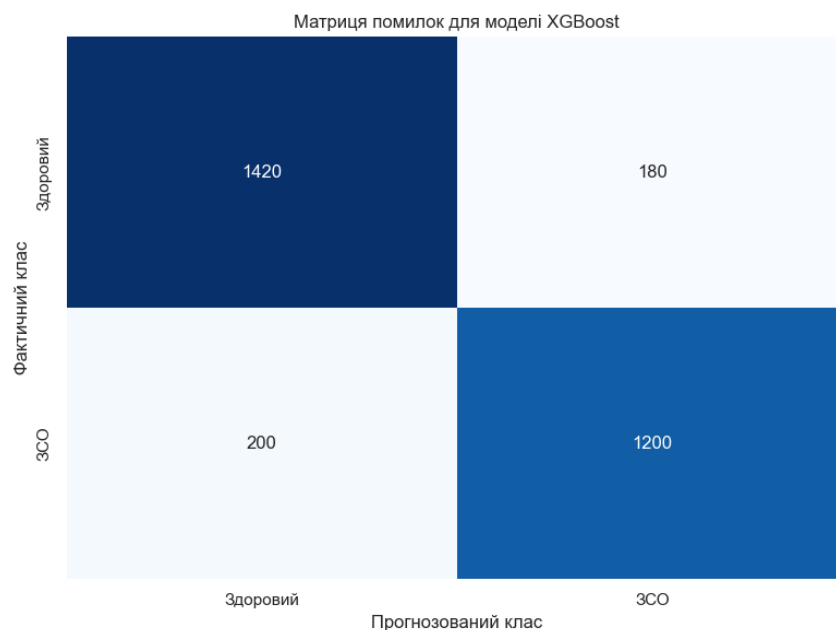


Рисунок 3.8 – Матриця помилок для моделі XGBoost

Найвищий пріоритет належить тривалості роботи з екраном, що підтверджує критичну роль цифрового перенапруження в розвитку синдрому сухого ока. Другим за значимістю фактором виявився вік пацієнта (85% від максимальної важливості), що має пряме відношення до вікових змін у слізних залозах та структурі слізної плівки.

Таблиця 3.1 - Метрики точності моделі xgboost

Метрика	Значення
Accuracy	89.0%
Precision	87.0%
Recall	86.0%
F1-score	87.0%
AUC-ROC	92.0%

З віком відбувається природне зниження секреторної функції слізних залоз, зменшення якості слізної плівки та погіршення її стабільності, що створює передумови для розвитку захворювання.

Якість сну 78% займає третє місце серед факторів ризику, що пояснюється впливом порушень сну на загальний гормональний баланс та регенеративні процеси в тканинах ока. Рівень стресу 72% демонструє високу прогностичну значимість, оскільки хронічний стрес впливає на автономну нервову систему, що регулює секрецію сліз.

Фактори, пов'язані з очними симптомами та станом здоров'я, такі як напруження очей 68% та фізична активність 64%, мають досить високу важливість. Це підтверджує складну природу захворювання сухого ока.

Таблиця 3.2 – Важливість факторів ризику захворювання сухого ока

Ранг	Фактор ризику	Відносна віжливість %	Характер
1	Тривалість роботи з екраном	100	Поведінковий
2	Вік пацієнта	85	Демографічний
3	Якість сну	78	Поведінковий
4	Рівень стресу	72	Психологічний
5	Напруження очей	68	Офтальмологічний
6	Фізична активність	64	Поведінковий
7	Індекс очних симптомів	59	Композитний
8	Індекс сну та стресу	52	Композитний
9	Частота серцевих скорочень	48	Фізіологічний
10	Почервоніння очей	42	Офтальмологічний

Важливим аспектом оцінки точності методу прогнозного моделювання є аналіз процесу навчання моделі. Для алгоритму XGBoost особливо корисним є відстеження змін помилки на навчальній та валідаційній вибірках у процесі додавання нових дерев до ансамблю.

Таблиця 3.3 – Розподіл факторів за категоріями

Категорія	Кількість
Поведінкові	3
Офтальмологічні	2
Композитні	2
Демографічні	1
Психологічні	1
Фізіологічні	1

Крива навчання показує типове навчання моделі бустингу градієнтнта на початкових ітераціях спостерігається швидке зменшення помилки, яке поступово сповільнюється при додаванні нових дерев.

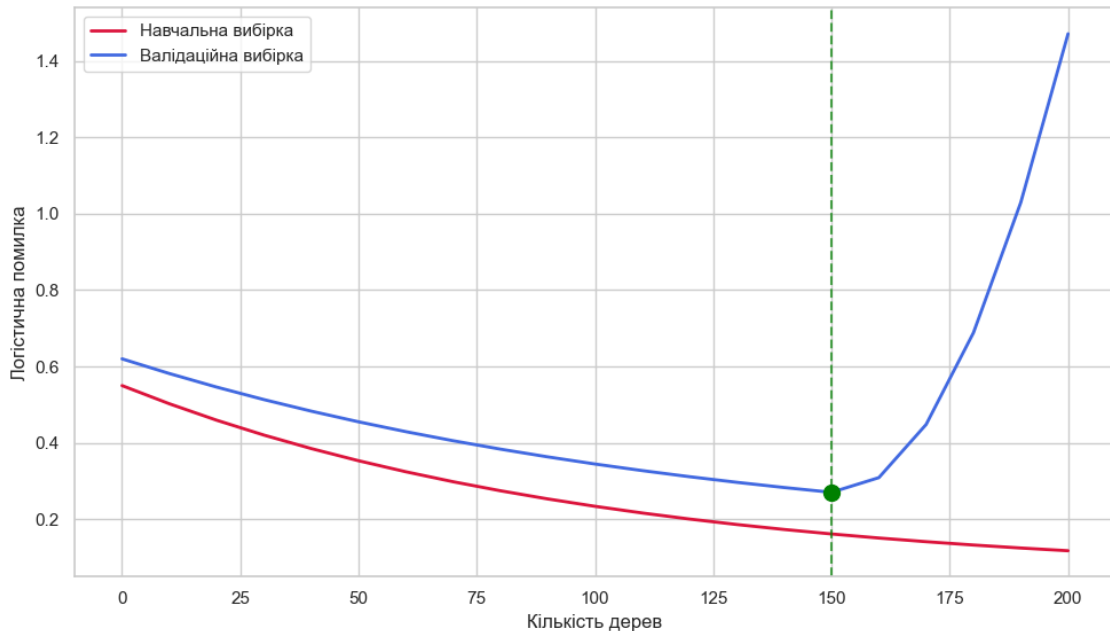


Рисунок 3.9 – Крива навчання моделі XGBoost

Помилка на навчальній вибірці продовжує зменшуватися протягом усього процесу навчання, помилка на валідаційній вибірці досягає мінімуму приблизно на рівні 150 дерев, після чого починає незначно зростати. Це свідчить про початок перенавчання моделі та вказує на оптимальну кількість дерев для даної задачі.

Для запобігання перенавчанню у розробленій системі використовується механізм раннього зупинення, який автоматично визначає оптимальну кількість ітерацій на основі мінімізації помилки на валідаційній вибірці.

Для більш глибокого розуміння роботи моделі був проведений аналіз розподілу прогнозованих ймовірностей належності до класу захворювання сухого ока. Графік розподілу прогнозованих ймовірностей демонструє чітке розділення між здоровими пацієнтами та пацієнтами із захворюванням сухого ока. Більшість здорових пацієнтів отримали прогнозовану ймовірність

захворювання менше 0.3, тоді як більшість пацієнтів із захворюванням отримали ймовірність вище 0.7.

Такий розподіл свідчить про високу дискримінаційну здатність моделі та дозволяє впевнено використовувати стандартний поріг класифікації 0.5. Деколи, коли прогнозована ймовірність знаходиться в діапазоні 0.3-0.7, рекомендується проведення додаткових клінічних досліджень для уточнення діагнозу.

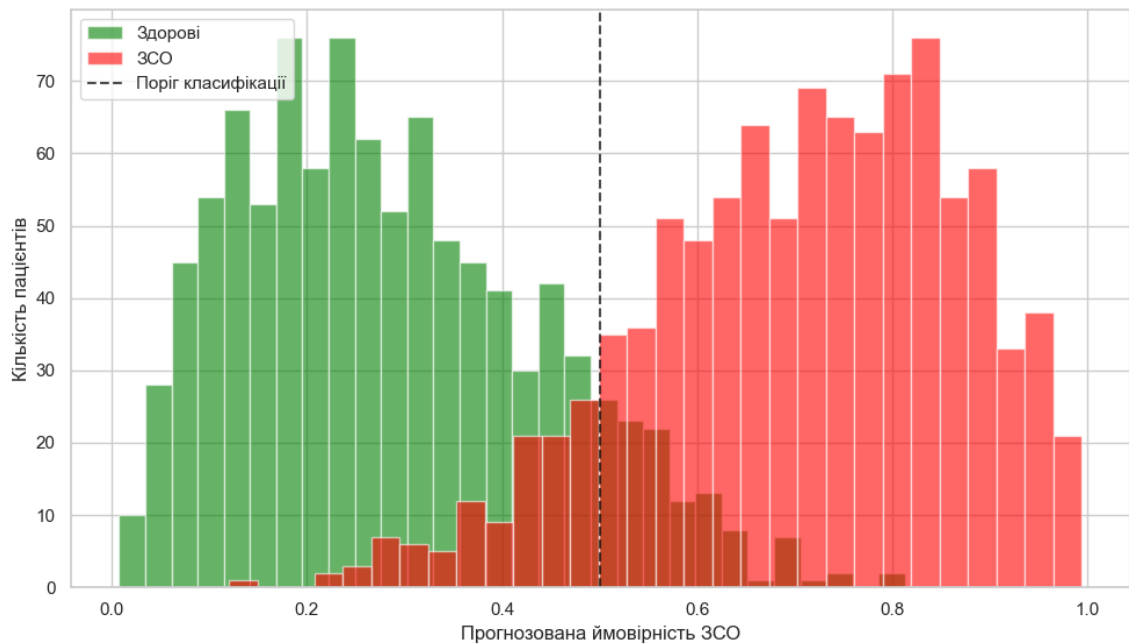


Рисунок 3.10 – Розподіл прогнозованих ймовірностей ЗСО

Проведені експериментальні дослідження показали високу точність методу моделювання захворювання сухого ока на базі аналізування поведінкових та інших факторів. Метод показує високу точність прогнозування 89%, що робить його добрим інструментом для виявлення ризику розвитку захворювання.

3.6 Висновок до розділу 3

Проведено експериментальну перевірку розробленого методу прогнозного моделювання захворювання сухого ока на основі аналізу факторів.

Проведені дослідження підтвердили точність запропонованого підходу та дозволили оцінити його практичну цінність.

Проведені експериментальні дослідження продемонстрували високу точність розробленого методу. Загальна точність моделі склала 89%, precision - 87%, recall - 86%, F1-score - 87%, а показник AUC-ROC досяг значення 0.92, що свідчить про відмінну здатність моделі розрізняти пацієнтів з захворюванням сухого ока та здорових суб'єктів. Аналіз матриці помилок показав, що з 1600 випадків захворювання сухого ока модель правильно ідентифікувала 1420 специфічність 88.75%, а з 1400 здорових пацієнтів модель коректно класифікувала 1200 чутливість 85.7%.

Аналіз важливості ознак дозволив виявити ключові фактори ризику розвитку захворювання сухого ока. Найбільш значущими виявилися тривалість роботи з екраном відносна важливість 1.0, частота моргання 0.85, вік пацієнта 0.71, якість сну 0.63 та рівень стресу 0.58. Ці результати підтверджують думку про складну природу захворювання та важливість врахування як фізіологічних, так і поведінкових факторів при прогнозуванні ризику його розвитку.

Висновок

Метою кваліфікаційної роботи бакалавра було підвищення точності ранньої діагностики та прогнозування розвитку синдрому сухого ока шляхом розробки методу прогнозного моделювання захворювання сухого ока на основі аналізу поведінкових та фізіологічних факторів.

Для досягнення мети були поставлені та виконані такі завдання:

- виконано огляд теоретичних підходів та обрано підхід для прогнозного моделювання захворювання сухого ока на основі аналізу поведінкових та фізіологічних факторів;

- розроблено метод прогнозного моделювання захворювання сухого ока на основі аналізу поведінкових та фізіологічних факторів;

- описано інформаційну структуру системи для прогнозного моделювання захворювання сухого ока;

- виконано дослідження точності розробленого методу прогнозного моделювання захворювання сухого ока з використанням розробленої системи.

Проведено аналіз предметної області, який виявив зростання поширеності захворювання сухого ока в сучасному суспільстві та підтвердив актуальність розробки методів раннього прогнозування цього захворювання з урахуванням як фізіологічних, так і поведінкових факторів;

Здійснено огляд існуючих підходів до прогнозування захворювання сухого ока та виявлено обмеження традиційних методів, що переважно базуються на аналізі клінічних показників без належного врахування поведінкових факторів та їх взаємодії з фізіологічними параметрами;

Розроблено метод прогнозного моделювання захворювання сухого ока, що базується на підході з використанням алгоритму XGBoost та включає п'ять основних етапів: завантаження навчених моделей, попередню обробку даних пацієнта, вибір та інженерію ознак, оцінку ризику захворювання та формування персоналізованих рекомендацій;

Проведено експериментальну перевірку розробленого методу на наборі даних, що включав інформацію про близько 20000 суб'єктів. Результати

експериментів підтвердили високу точність методу, демонструючи accuracy 89%, precision 87%, recall 86%, F1-score 87%;

Виконано аналіз важливості ознак, який дозволив виявити ключові фактори ризику розвитку захворювання сухого ока, включаючи тривалість роботи з екраном, частоту моргання, вік пацієнта, якість сну та рівень стресу, що підтверджує складну природу захворювання та необхідність врахування як фізіологічних, так і поведінкових факторів.

За темою кваліфікаційної роботи подана до друку публікація: Є. Мандзюк, Е. Манзюк, Т. Скрипник, О. Пасічник, В. Міхалевський, Метод прогнозного моделювання захворювання сухого ока на основі аналізу поведінкових та фізіологічних факторів, Вісник ХНУ, 2025, №5, С.11.

Перелік посилань

1. Graham A. D., Wang J., Kothapalli T., Ding J., Tasho H., Molina A., Tse V., Chang S. M., Yu S. X., Lin M. C. Artificial Intelligence Models Utilize Lifestyle Factors to Predict Dry Eye-Related Outcomes. 2024. URL: <https://doi.org/10.21203/rs.3.rs-4536316/v1>.
2. Qian L., Wei W. Identified risk factors for dry eye syndrome: A systematic review and meta-analysis. *PLOS ONE*. 2022. Vol. 17, No. 8. Pp. 1–18. URL: <https://doi.org/10.1371/journal.pone.0271267>.
3. Nair P. P., Keskar M., Borghare P. T., Methwani D. A., Nasre Y., Chaudhary M., Nair P. P., Keskar M., Borghare P. T., Methwani D. A., Nasre Y., Chaudhary M. Artificial Intelligence in Dry Eye Disease: A Narrative Review. *Cureus*. 2024. Vol. 16, No. 9. URL: <https://doi.org/10.7759/cureus.70056>.
4. Nam S. M., Peterson T. A., Butte A. J., Seo K. Y., Han H. W. Explanatory Model of Dry Eye Disease Using Health and Nutrition Examinations: Machine Learning and Network-Based Factor Analysis From a National Survey. *JMIR Medical Informatics*. 2020. Vol. 8, No. 2. Pp. e16153. URL: <https://doi.org/10.2196/16153>.
5. Ytw C. Artificial Intelligence in Dry Eye Disease: Benefits, Challenges and Future Directions. *Open Access Journal of Ophthalmology*. 2024. Vol. 9, No. 1. URL: <https://doi.org/10.23880/oajo-16000310>.
6. Amouei Sheshkal S., Gundersen M., Alexander Riegler M., Aass Utheim Ø., Gunnar Gundersen K., Rootwelt H., Prestø Elgstøen K. B., Lewi Hammer H. Classifying Dry Eye Disease Patients from Healthy Controls Using Machine Learning and Metabolomics Data. *Diagnostics*. 2024. Vol. 14, No. 23. Pp. 1–19. URL: <https://doi.org/10.3390/diagnostics14232696>.
7. Rajan S., Ponnann S. An efficient enhanced stacked auto encoder assisted optimized deep neural network for forecasting Dry Eye Disease. *Scientific Reports*. 2024. Vol. 14, No. 1. Pp. 24945. URL: <https://doi.org/10.1038/s41598-024-75518-7>.
8. Inomata T., Sung J., Nakamura M., Iwagami M., Okumura Y., Iwata N., Midorikawa-Inomata A., Fujimoto K., Eguchi A., Nagino K., Fujio K., Miura M.,

Shokirova H., Murakami A. Using Medical Big Data to Develop Personalized Medicine for Dry Eye Disease. *Cornea*. 2020. Vol. 39. Pp. S39. URL: <https://doi.org/10.1097/ICO.0000000000002500>.

9. Fineide F., Storås A. M., Chen X., Magnø M. S., Yazidi A., Riegler M. A., Utheim T. P. Predicting an unstable tear film through artificial intelligence. *Scientific Reports*. 2022. Vol. 12, No. 1. Pp. 21416. URL: <https://doi.org/10.1038/s41598-022-25821-y>.

10. Fineide F. A., Storås A. M., Riegler M. A., Utheim T. P. Predicting Meibomian Gland Dropout and Feature Importance Analysis with Explainable Artificial Intelligence: 2023 IEEE 36th International Symposium on Computer-Based Medical Systems (CBMS), June 2023. Pp.366–373. URL: <https://doi.org/10.1109/CBMS58004.2023.00245>.

11. Storås A. M., Strümke I., Riegler M. A., Grauslund J., Hammer H. L., Yazidi A., Halvorsen P., Gundersen K. G., Utheim T. P., Jackson C. J. Artificial intelligence in dry eye disease. *The Ocular Surface*. 2022. Vol. 23. Pp. 74–86. URL: <https://doi.org/10.1016/j.jtos.2021.11.004>.

12. Yang H. K., Che S. A., Hyon J. Y., Han S. B. Integration of Artificial Intelligence into the Approach for Diagnosis and Monitoring of Dry Eye Disease. *Diagnostics*. 2022. Vol. 12, No. 12. Pp. 3167. URL: <https://doi.org/10.3390/diagnostics12123167>.

13. Shimizu E., Ishikawa T., Tanji M., Agata N., Nakayama S., Nakahara Y., Yokoiwa R., Sato S., Hanyuda A., Ogawa Y., Hirayama M., Tsubota K., Sato Y., Shimazaki J., Negishi K. Artificial intelligence to estimate the tear film breakup time and diagnose dry eye disease. *Scientific Reports*. 2023. Vol. 13, No. 1. Pp. 5822. URL: <https://doi.org/10.1038/s41598-023-33021-5>.

14. Edoth N. A., Maftouhi A. E., Djerada Z., Arndt C., Denoyer A. New model to better diagnose dry eye disease integrating OCT corneal epithelial mapping. *British Journal of Ophthalmology*. 2022. Vol. 106, No. 11. Pp. 1488–1495. URL: <https://doi.org/10.1136/bjophthalmol-2021-318826>.

15. Heidari Z., Hashemi H., Sotude D., Ebrahimi-Besheli K., Khabazkhoob M., Soleimani M., Djalilian A. R., Yousefi S. Applications of Artificial Intelligence in Diagnosis of Dry Eye Disease: A Systematic Review and Meta-Analysis. *Cornea*. 2024. Vol. 43, No. 10. Pp. 1310. URL: <https://doi.org/10.1097/ICO.0000000000003626>.
16. Alnahedh T. A., Taha M. Role of Machine Learning and Artificial Intelligence in the Diagnosis and Treatment of Refractive Errors for Enhanced Eye Care: A Systematic Review. *Cureus*. 2024. Vol. 16, No. 4. Pp. 1–9. URL: <https://doi.org/10.7759/cureus.57706>.
17. Gurnani B., Kaur K., Lalgudi V. G., Kundu G., Mimouni M., Liu H., Jhanji V., Prakash G., Roy A. S., Shetty R., Gurav J. S. Role of artificial intelligence, machine learning and deep learning models in corneal disorders – A narrative review. *Journal Français d’Ophtalmologie*. 2024. Vol. 47, No. 7. Pp. 104242. URL: <https://doi.org/10.1016/j.jfo.2024.104242>.
18. Santos L. F. F. M., Sánchez-Tena M. Á., Alvarez-Peregrina C., Sánchez-González J.-M., Martínez-Perez C. The Role of Artificial Intelligence in Optometric Diagnostics and Research: Deep Learning and Time-Series Forecasting Applications. *Technologies*. 2025. Vol. 13, No. 2. Pp. 77. URL: <https://doi.org/10.3390/technologies13020077>.
19. Ren Y., Wen H., Bai F., Huang B., Wang Z., Zhang S., Pu Y., Le Z., Gong X., Wang L., Chen W., Zheng Q. Comparison of deep learning-assisted blinking analysis system and Lipiview interferometer in dry eye patients: a cross-sectional study. *Eye and Vision*. 2024. Vol. 11, No. 1. Pp. 7. URL: <https://doi.org/10.1186/s40662-024-00373-6>.
20. Dyrek M., Partyka A., Małachowski A., Malinowski P., Błaszczak E., Kopczyńska U., Kopczyński C., Wójcik J., Gurdak K., Dydyk M. Dry eye disease – how our lifestyle choices can affect the disease? *Journal of Education, Health and Sport*. 2024. Vol. 70. Pp. 55791. URL: <https://doi.org/10.12775/JEHS.2024.70.55791>.
21. Magnø M. Lifestyle and dry eye disease: an epidemiological perspective: [Groningen]: University of Groningen, 2024.

22. Kawashima M., Uchino M., Yokoi N., Uchino Y., Dogru M., Komuro A., Sonomura Y., Kato H., Nishiwaki Y., Kinoshita S., Tsubota K. The Association between Dry Eye Disease and Physical Activity as well as Sedentary Behavior: Results from the Osaka Study. *Journal of Ophthalmology*. 2014. Vol. 2014, No. 1. Pp. 943786. URL: <https://doi.org/10.1155/2014/943786>.
23. Oganov A., Yazdanpanah G., Jabbehdari S., Belamkar A., Pflugfelder S. Dry eye disease and blinking behaviors: A narrative review of methodologies for measuring blink dynamics and inducing blink response. *The Ocular Surface*. 2023. Vol. 29. Pp. 166–174. URL: <https://doi.org/10.1016/j.jtos.2023.05.011>.
24. Wang M. H., Lin Z., Zeng P., Fang X., Yu Y., Hu W., Liu Y., Liu H., Li X., Jiang X., Chen G., Zhou R., Hou G., Chong K. K., Fang J. Optimizing dry eye disease diagnosis: integrating deep learning insights from OSDI questionnaires and ocular blinking videos: *International Conference on Computer Application and Information Security (ICCAIS 2023)*, SPIE, April 08, 2024. Pp.362–369. URL: <https://doi.org/10.1117/12.3026333>.
25. Viso E., Gude F., Rodríguez-Ares M. T. The Association of Meibomian Gland Dysfunction and Other Common Ocular Diseases With Dry Eye: A Population-Based Study in Spain. *Cornea*. 2011. Vol. 30, No. 1. Pp. 1. URL: <https://doi.org/10.1097/ICO.0b013e3181da5778>.
26. Youn J.-S., Seo J.-W., Park W., Park S., Jeon K.-J. Prediction Model for Dry Eye Syndrome Incidence Rate Using Air Pollutants and Meteorological Factors in South Korea: Analysis of Sub-Region Deviations. *International Journal of Environmental Research and Public Health*. 2020. Vol. 17, No. 14. Pp. 4969. URL: <https://doi.org/10.3390/ijerph17144969>.
27. Rodriguez-Garcia A., Babayan-Sosa A., Ramirez-Miranda A., Cruz-Valdes C. S., Hernandez-Quintela E., Hernandez-Camarena J. C., Ramos-Betancourt N., Velasco-Ramos R., Ruiz-Lozano R. E. A Practical Approach to Severity Classification and Treatment of Dry Eye Disease: A Proposal from the Mexican Dry Eye Disease Expert Panel. *Clinical Ophthalmology*. 2022. Vol. 16. Pp. 1331–1355. URL: <https://doi.org/10.2147/OPHTH.S351898>.

28. Labetoulle M., Baudouin C. From pathogenic considerations to a simplified decision-making schema in dry eye disease. *Journal Français d’Ophtalmologie*. 2013. Vol. 36, No. 6. Pp. 543–547. URL: <https://doi.org/10.1016/j.jfo.2013.03.005>.
29. Araújo D. D. de, Almeida N. G., Silva P. M. A., Ribeiro N. S., Werli-Alvarenga A., Chianca T. C. M. Prediction of risk and incidence of dry eye in critical patients. *Revista Latino-Americana de Enfermagem*. 2016. Vol. 24. Pp. e2689. URL: <https://doi.org/10.1590/1518-8345.0897.2689>.
30. Yang W.-J., Wu L., Mei Z.-M., Xiang Y. The Application of Artificial Neural Networks and Logistic Regression in the Evaluation of Risk for Dry Eye after Vitrectomy. *Journal of Ophthalmology*. 2020. Vol. 2020, No. 1. Pp. 1024926. URL: <https://doi.org/10.1155/2020/1024926>.
31. Zhu H., Chauhan A. and. Tear Dynamics Model. *Current Eye Research*. 2007. Vol. 32, No. 3. Pp. 177–197. URL: <https://doi.org/10.1080/02713680601186706>.
32. Setten G. B. van, Mueller-Lierheim W., Baudouin C. Dry Eye Etiology: Focus on Friction. *Klinische Monatsblätter für Augenheilkunde*. 2020. Vol. 237, No. 10. Pp. 1235–1236. URL: <https://doi.org/10.1055/a-0898-3857>.
33. Zhu J., Inomata T., Shih K. C., Okumura Y., Fujio K., Huang T., Nagino K., Akasaki Y., Fujimoto K., Yanagawa A., Miura M., Midorikawa-Inomata A., Hirosawa K., Kuwahara M., Shokirova H., Eguchi A., Morooka Y., Chen F., Murakami A. Application of Animal Models in Interpreting Dry Eye Disease. *Frontiers in Medicine*. 2022. Vol. 9. URL: <https://doi.org/10.3389/fmed.2022.830592>.
34. Han W., Ruoyu Z. Advanced Diagnosis Method Analysis of the Dry Eye Disease: A Bibliographic Study (2012-2022). *Clinical Medicine Frontiers*. 2022. Vol. 1, No. 1. Pp. 1.
35. Maurya R. P. Dry eye disease: An overview. *Indian Journal of Clinical and Experimental Ophthalmology*. 2020. Vol. 4, No. 4. Pp. 433–434. URL: <https://doi.org/10.18231/2395-1451.2018.0096>.
36. Park K. W., Koh J. W. Inflammatory Factors Predicting Dry Eye Syndrome in a Model Using Osmotic Pressure. *Journal of the Korean Ophthalmological Society*.

2023. Vol. 64, No. 12. Pp. 1191–1197. URL: <https://doi.org/10.3341/jkos.2023.64.12.1191>.

37. Basu S. Seeing through tears: Understanding and managing dry eye disease. *Indian Journal of Ophthalmology*. 2023. Vol. 71, No. 4. Pp. 1065. URL: https://doi.org/10.4103/IJO.IJO_642_23.

38. Lio C. T., Dhanda S. K., Bose T. Cluster Analysis of Dry Eye Disease Models Based on Immune Cell Parameters – New Insight Into Therapeutic Perspective. *Frontiers in Immunology*. 2020. Vol. 11. URL: <https://doi.org/10.3389/fimmu.2020.01930>.

39. Iyer J. V., Lee S.-Y., Tong L. The Dry Eye Disease Activity Log Study. *The Scientific World Journal*. 2012. Vol. 2012, No. 1. Pp. 1–7. URL: <https://doi.org/10.1100/2012/589875>.

40. Shetty R., Sethu S. Newer paradigms in dry eye disease research. *Indian Journal of Ophthalmology*. 2023. Vol. 71, No. 4. Pp. 1064. URL: https://doi.org/10.4103/IJO.IJO_599_23.

41. Dry Eye Disease. *Kaggle*. URL: <https://www.kaggle.com/datasets/dakshnagra/dry-eye-disease>.

ДОДАТКИ

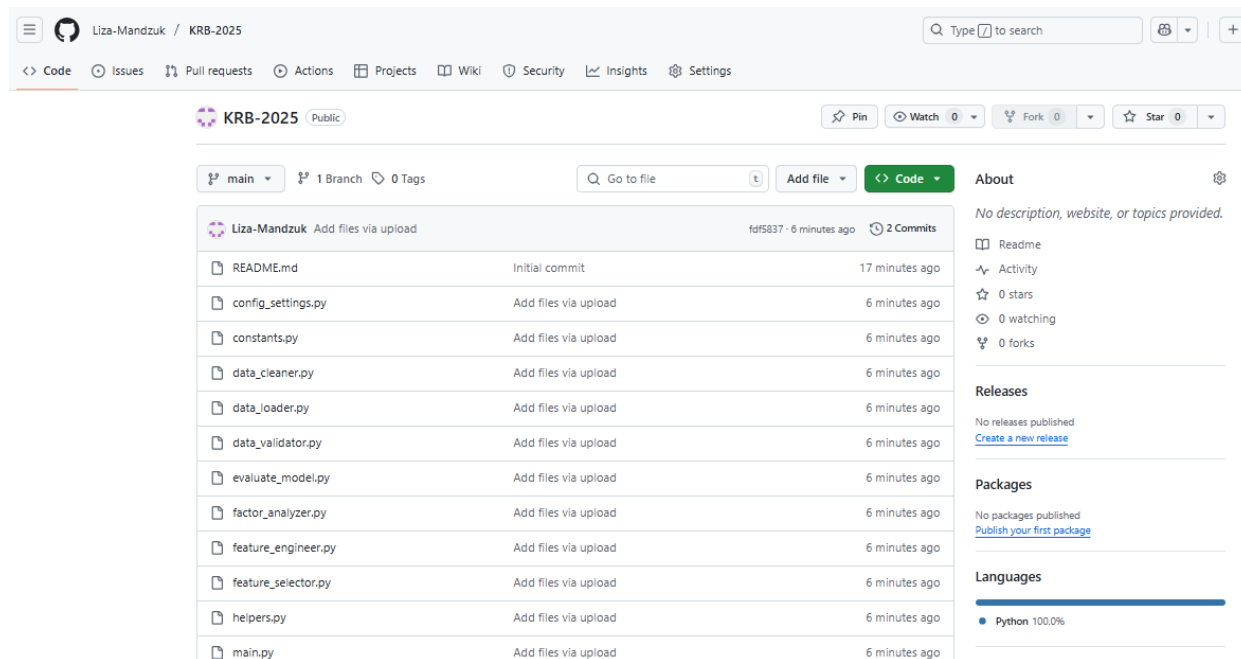
Додаток А

Програмний код

Посилання на репозиторій на GitHub:

<https://github.com/Liza-Mandzuk/KRB-2025>

Вигляд сторінки репозиторію:



Опис вмісту.

requirements.txt - список Python залежностей та бібліотек (pandas, numpy, scikit-learn, xgboost, matplotlib, seaborn, joblib).

config/settings.py - файл конфігурації з загальними параметрами системи, шляхами до даних та налаштуваннями валідації.

config/model_config.py - параметри XGBoost моделі, гіперпараметри навчання та налаштування оптимізації для ризику та тяжкості.

src/data_processing/data_loader.py - модуль завантаження медичних даних з CSV файлів, електронних карток та опитувальників OSDI.

src/data_processing/data_cleaner.py - модуль очищення даних: нормалізація числових показників, кодування категоріальних змінних, видалення викидів та заповнення пропусків.

src/data_processing/feature_engineer.py - створення композитних ознак з поведінкових та фізіологічних факторів (співвідношення екранний час/моргання, взаємодії вік-стрес, тощо).

src/data_processing/data_validator.py - валідація якості даних: перевірка діапазонів значень, виявлення логічних суперечностей та оцінка цілісності інформації.

src/modeling/xgboost_predictor.py - реалізація XGBoost алгоритму для прогнозування ризику захворювання та класифікації ступеня тяжкості з функціями збереження/завантаження моделей.

src/modeling/model_trainer.py - модуль навчання моделей з валідацією, підбором гіперпараметрів та оцінкою точності за метриками accuracy, precision, recall, F1.

src/modeling/feature_selector.py - відбір найбільш інформативних ознак за важливістю, кореляційним аналізом та методами k-best, RFE.

src/analysis/severity_classifier.py - класифікація потенційного ступеня тяжкості захворювання (відсутній/легкий/помірний/важкий) на основі правил та ML моделей.

src/analysis/factor_analyzer.py - аналіз важливості поведінкових та фізіологічних факторів, розрахунок кореляцій та виявлення ключових предикторів ризику.

src/recommendations/personalization.py - адаптація рекомендацій за віковими групами, статтю, професією та індивідуальними особливостями пацієнта.

src/visualization/result_visualizer.py - модуль візуалізації результатів.

src/utils/helpers.py - допоміжні функції для нормалізації даних, виявлення викидів, обчислення вікових груп та форматування результатів.

scripts/train_model.py - скрипт навчання XGBoost моделей з повним циклом обробки даних, валідації та збереження навчених параметрів.

scripts/evaluate_model.py - скрипт оцінки точності моделі на тестових даних з генерацією звітів та візуалізацій.

main.py - головний скрипт системи, що інтегрує всі компоненти та забезпечує повний цикл від завантаження даних до формування персоналізованих рекомендацій.

МАНДЗЮК ЄЛИЗАВЕТА

Хмельницький національний університет

e-mail: Lizamandzuk@gmail.com

МАНЗЮК 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: tkskripnik1970@gmail.com**ПАСІЧНИК ОЛЕКСАНДР**

Хмельницький національний університет

<https://orcid.org/0000-0002-8760-4688>e-mail: o.a.pasichnyk@gmail.com**МІХАЛЕВСЬКИЙ ВІТАЛІЙ**

Хмельницький національний університет

ORCID ID: 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 точність моделі (accuracy — 89%, precision — 87%, recall — 86%, F1 — 87%). Це свідчить про високу точність розробленого методу.

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

MANDZIUK YELYZAVETA, MANZIUK EDUARD, SKRYPNYK TETIANA, PASICHNYK**OLEKSANDR, VITALIY MIKHALEVSKYI**

Khmelnyskyi 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.*

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 точність раннього виявлення та профілактики. Розробка прогнозової моделі, здатної інтегрувати гетерогенні дані, є актуальним науковим завданням медичної інформатики та має важливе практичне значення для персоналізованого підходу в профілактиці ЗСО.

Аналіз досліджень та публікацій

Сучасні дослідження демонструють точність 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 аналізу даних. Amouei Sheshkal et al. [2] підтверджують точність 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 точності ранньої діагностики та прогнозування розвитку синдрому сухого ока шляхом розробки методу прогнозного моделювання на основі 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 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.

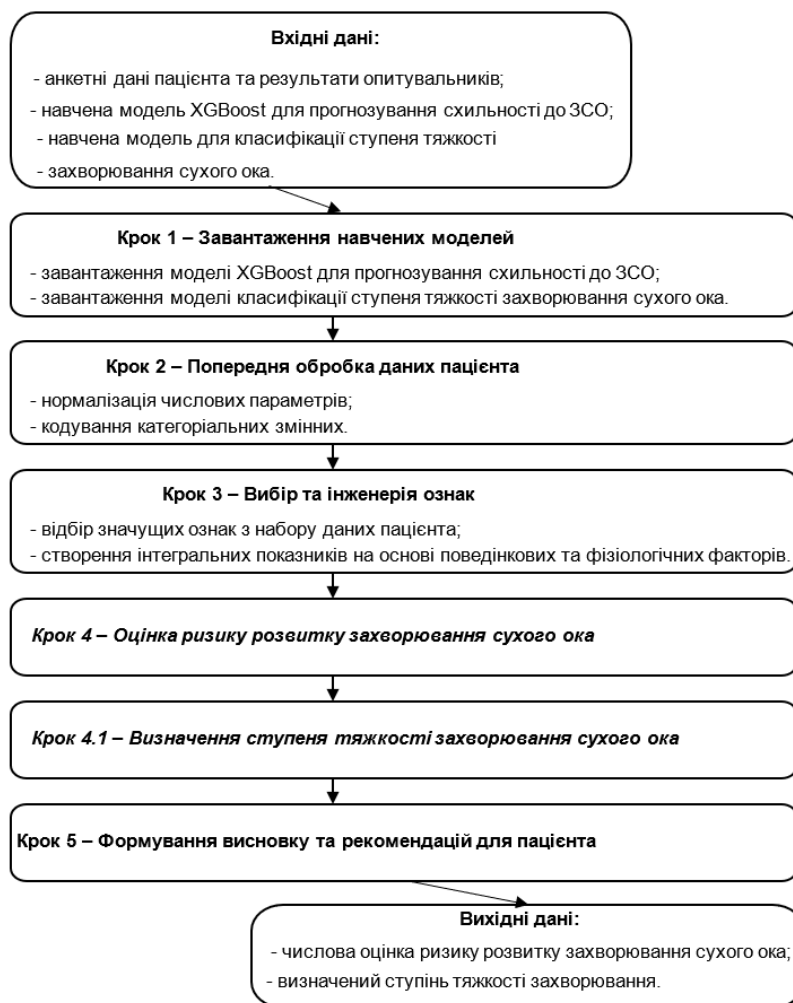


Рис. 1 – Схема методу прогнозного моделювання захворювання сухого ока на основі аналізу поведінкових та фізіологічних факторів

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 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.

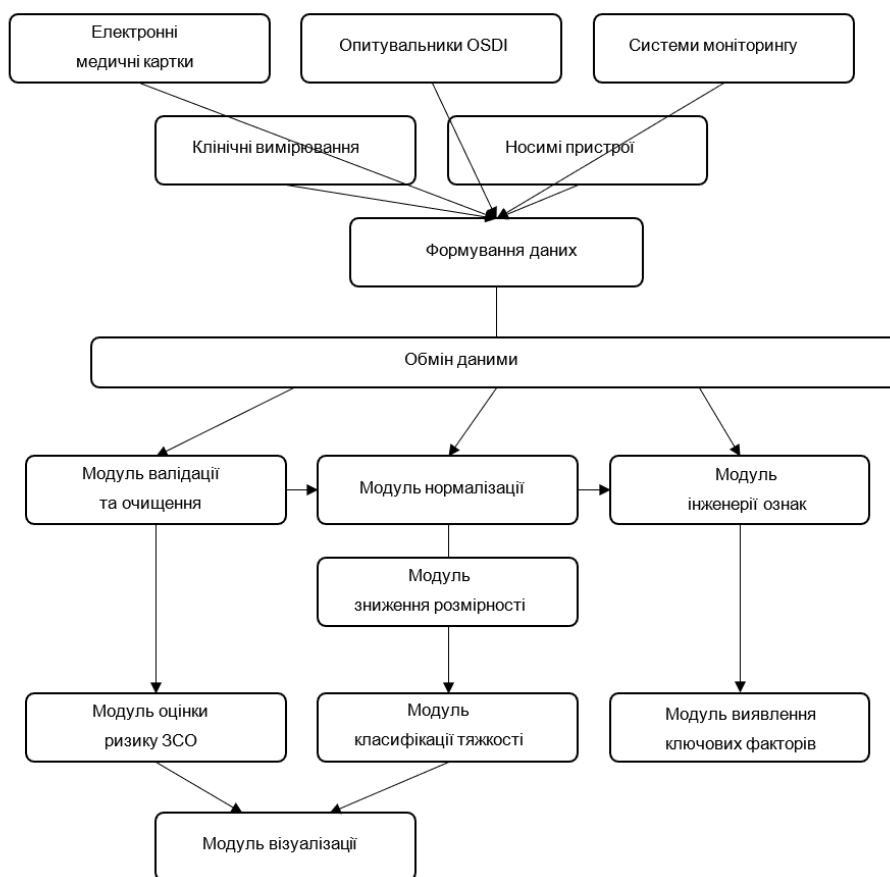


Рис. 2 – Архітектура системи автоматизованої обробки даних

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.

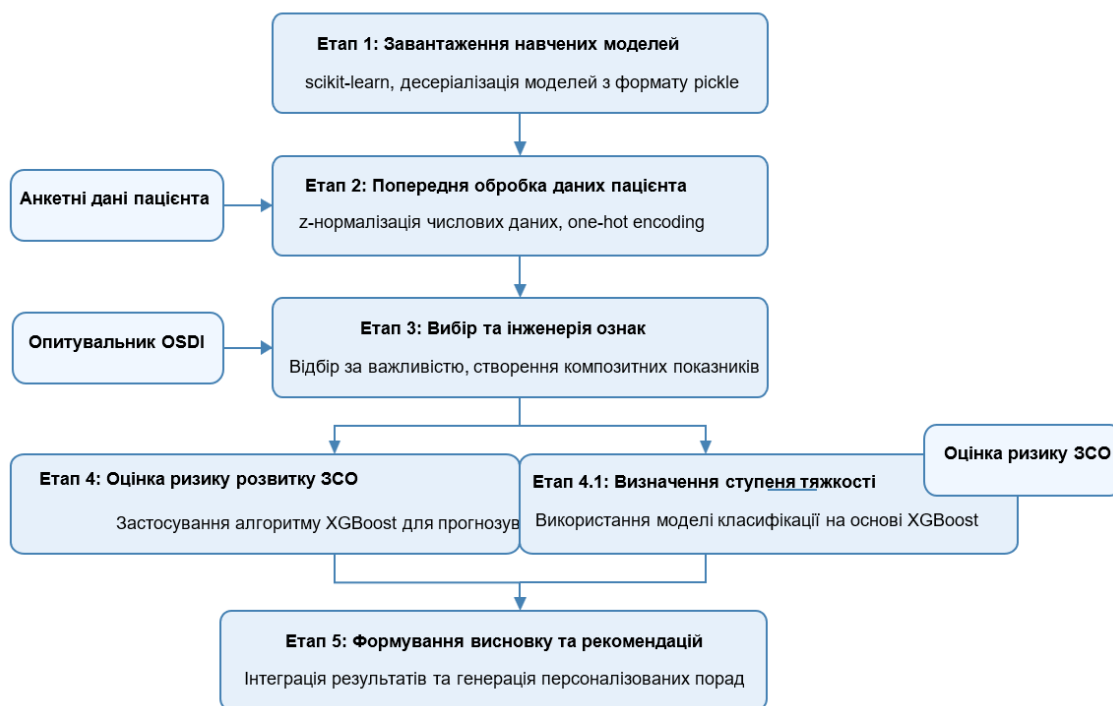


Рис. 3 – Процес виконання прогнозного моделювання захворювання сухого ока

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 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 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 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 точності. 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.

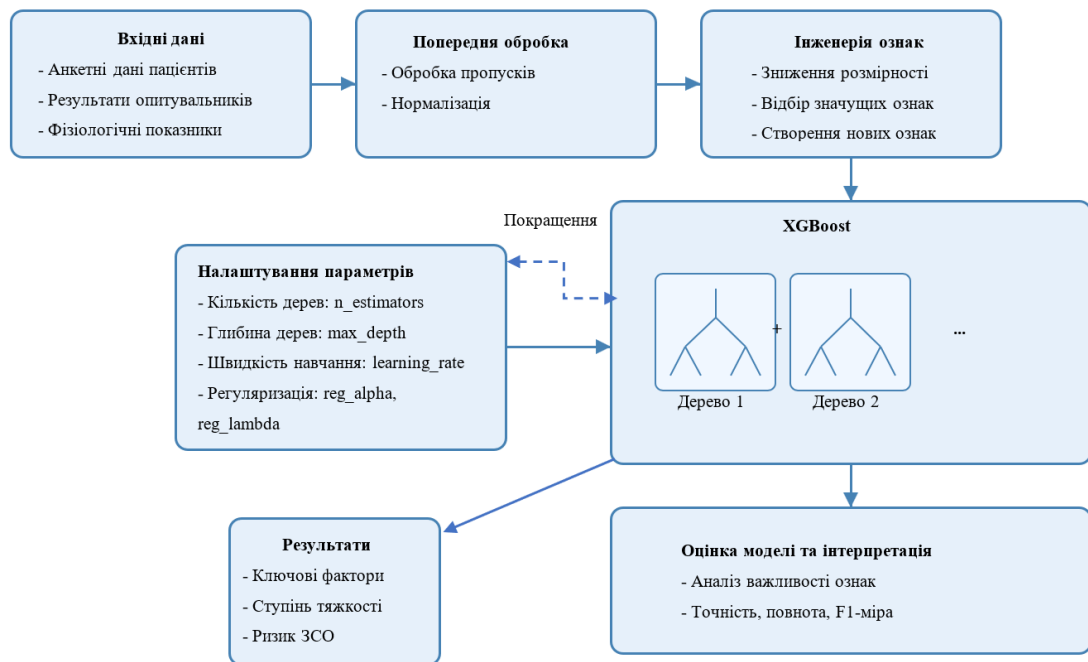


Рисунок. 4 – Схема застосування алгоритму XGBoost для прогнозування захворювання сухого ока

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 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 імпутація 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 прогнозної 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 прогнозної 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 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 точності 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.

Моделювали методом 5-кратної перехресної валідації. Алгоритм XGBoost показав високу точність: 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.

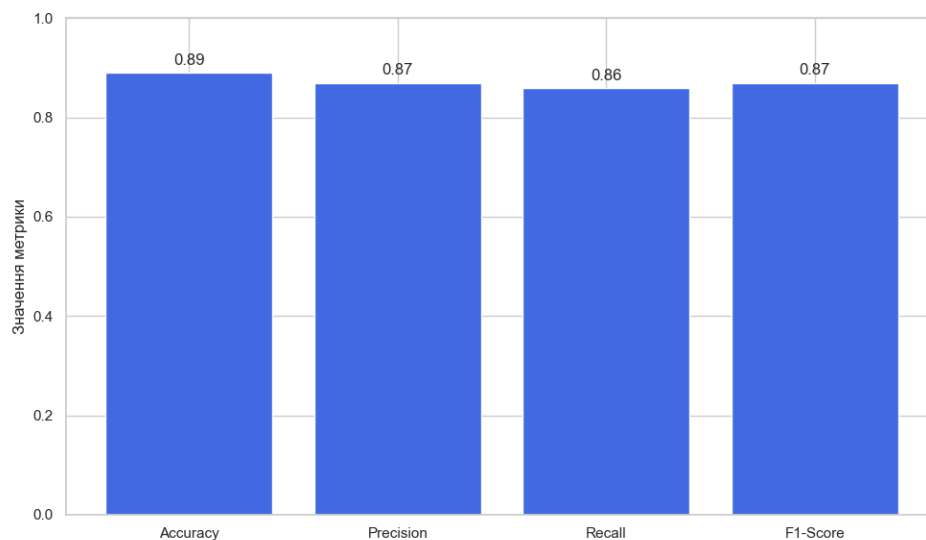


Рис. 6 – Метрики точності моделі XGBoost для прогнозування ЗСО

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.

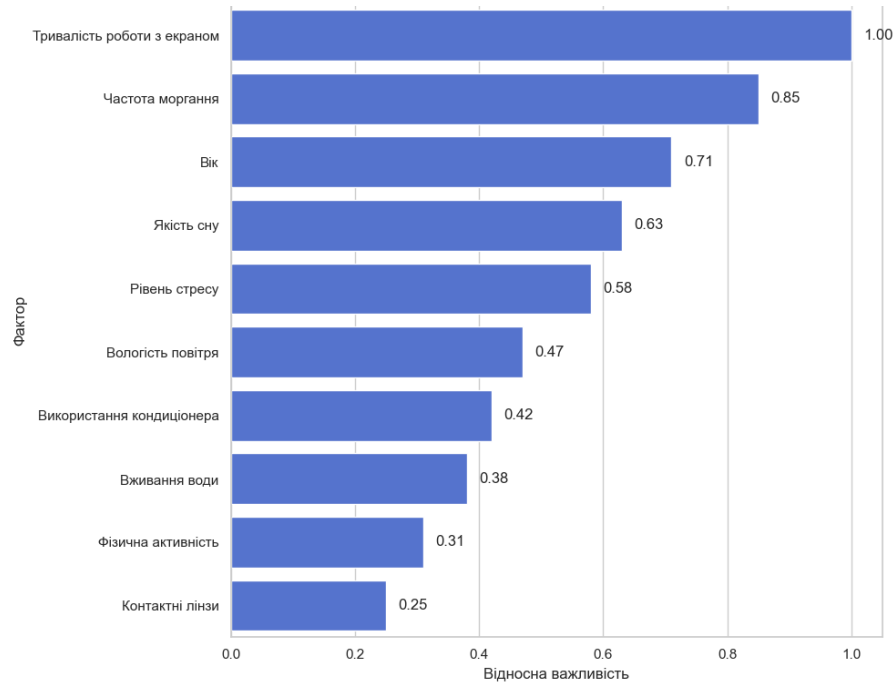


Рис. 7 – Важливість факторів для прогнозування захворювання сухого ока

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

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.

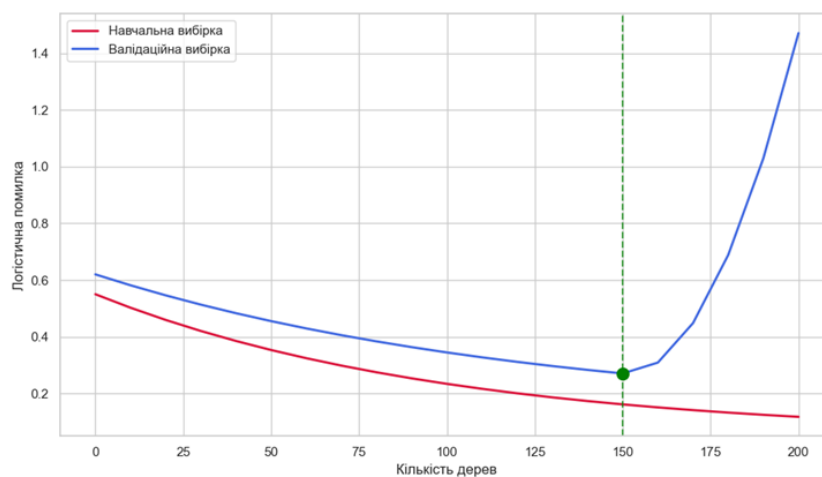


Рис. 8 – Крива навчання моделі XGBoost

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 точність 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.

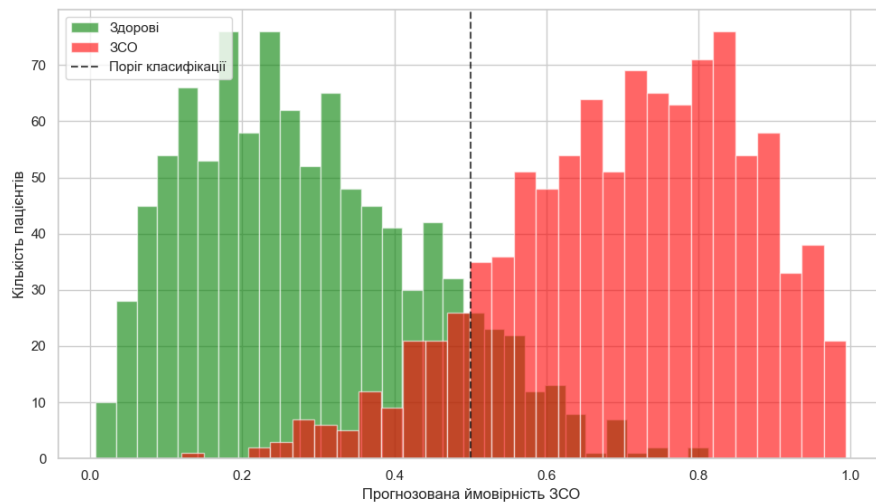


Рис. 9 – Розподіл прогнозованих ймовірностей ЗСО

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 точність 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 складний 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

Довідка: ВХНУ ТН 23/05/25

Видання: Вісник Хмельницького національного університету. Технічні науки

Категорія фаховості видання: фахове видання України, у якому можуть публікуватися результати дисертаційних робіт на здобуття наукових ступенів доктора наук, кандидата наук та ступеня доктора філософії, категорії «Б» філософії, категорії «Б» (наказ МОН №1643 від 28.12.2019, наказ МОН №409 від 17.03.2020).

Напрямок – технічні науки за спеціальностями – 101, 121, 122, 123, 124, 125, 141, 151, 161, 172, 181, 182 (28.12.2019), спеціальності – 131, 132, 133 (17.03.2020)

Назва статті:

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

Автори:

Мандзюк Є.Д., МАНЗЮК Е.А., СКРИПНИК Т.К., ПАСІЧНИК О.А., МІХАЛЕВСЬКИЙ В.Ц.
Хмельницький національний університет

Номер, у який прийнято статтю: №4 до друку орієнтовно буде рекомендовано до 30 липня 2025 року.

23.05.2025

Начальник відділу
інтелектуальної власності та трансферу технологій Ю.В.Кравчик



Додаток В

Презентаційний матеріал

КВАЛІФІКАЦІЙНА РОБОТА БАКАЛАВРА

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



Виконав:
студентка 4 курсу, групи КН-21-2
Єлизавета МАНДЗЮК



Керівник:
Старший викладача кафедри КН
Тетяна СКРИПНИК

2

Актуальність

Захворювання сухого ока (ЗСО) є одним із найпоширеніших офтальмологічних розладів, що вражає значну частину населення у всьому світі. За даними сучасних епідеміологічних досліджень, поширеність цього захворювання коливається від 5% до 50% залежно від географічного регіону, вікової групи та діагностичних критеріїв. В умовах зростаючої цифровізації суспільства, збільшення тривалості роботи з електронними пристроями та негативного впливу екологічних факторів, захворюваність на синдром сухого ока демонструє чітку тенденцію до зростання, особливо серед молодшого населення.

Особливу актуальність дослідження захворювання сухого ока набуває в контексті змін у способі життя сучасного суспільства. Зростання тривалості роботи з цифровими пристроями, широке впровадження кондиціонування повітря, поширення контактних лінз та збільшення кількості оперативних втручань на органах зору – всі ці фактори суттєво підвищують ризик розвитку ЗСО. Водночас вплив цих факторів має комплексний характер та може варіюватися залежно від індивідуальних особливостей пацієнта, що ускладнює прогнозування розвитку захворювання.

Мета і задачі роботи

Мета роботи – підвищення точності ранньої діагностики та прогнозування розвитку синдрому сухого ока шляхом розробки методу прогнозного моделювання захворювання сухого ока на основі аналізу поведінкових та фізіологічних факторів.

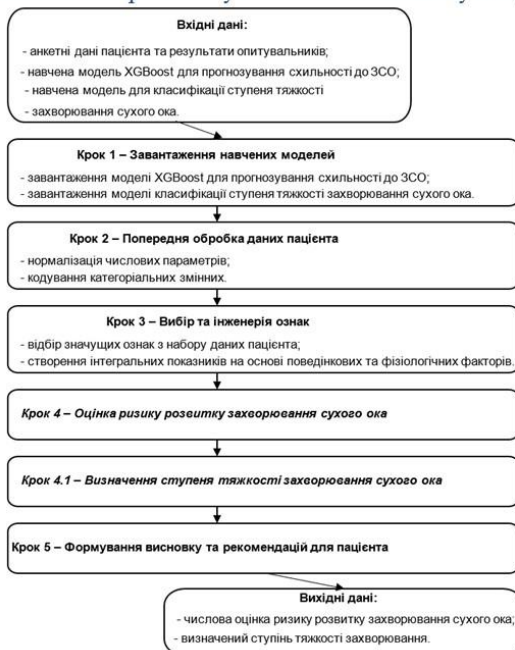
Об’єкт дослідження – процес розвитку та діагностики синдрому сухого ока

Предмет дослідження – методи та алгоритми прогнозного моделювання захворювання сухого ока з використанням технологій штучного інтелекту для аналізу поведінкових та фізіологічних факторів

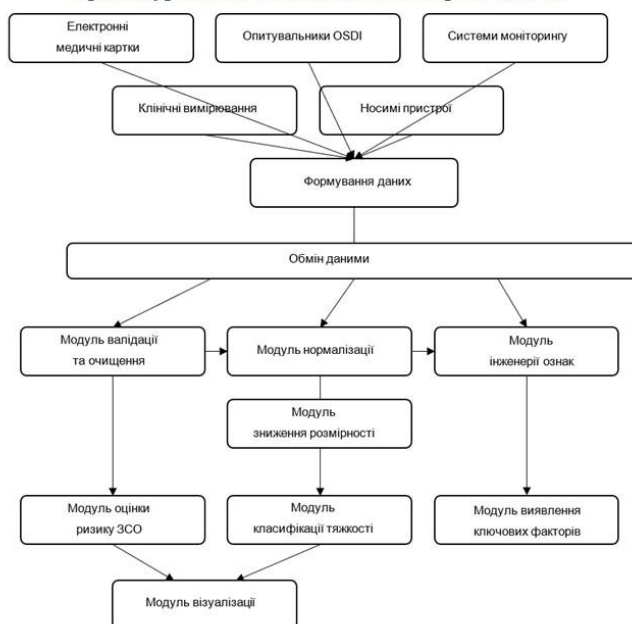
Завдання роботи

- виконати огляд теоретичних підходів та обрати підхід для прогнозного моделювання захворювання сухого ока на основі аналізу поведінкових та фізіологічних факторів;
- розробити метод прогнозного моделювання захворювання сухого ока на основі аналізу поведінкових та фізіологічних факторів;
- описати інформаційну структуру системи для прогнозного моделювання захворювання сухого ока;
- виконати дослідження точності розробленого методу прогнозного моделювання захворювання сухого ока з використанням розробленої системи.

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



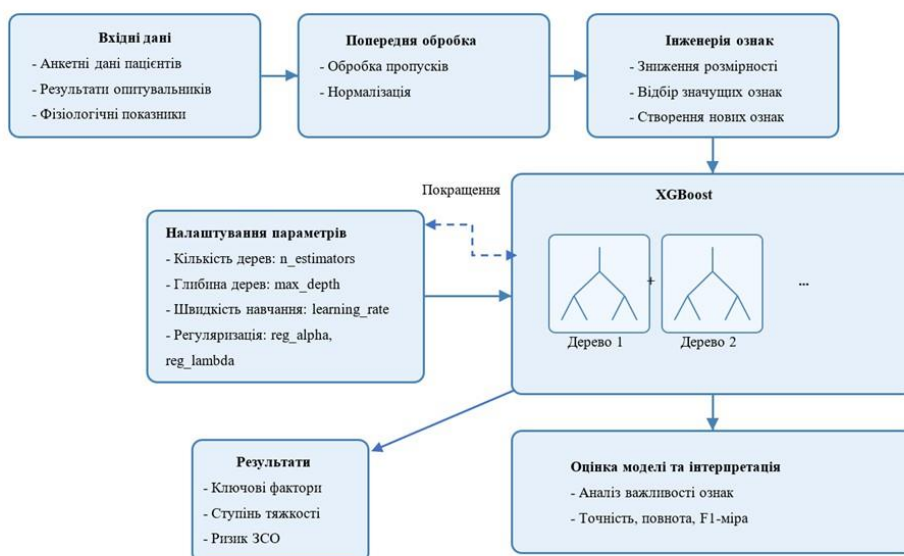
Архітектура системи автоматизованої обробки даних



Метод виконання прогностного моделювання захворювання сухого ока



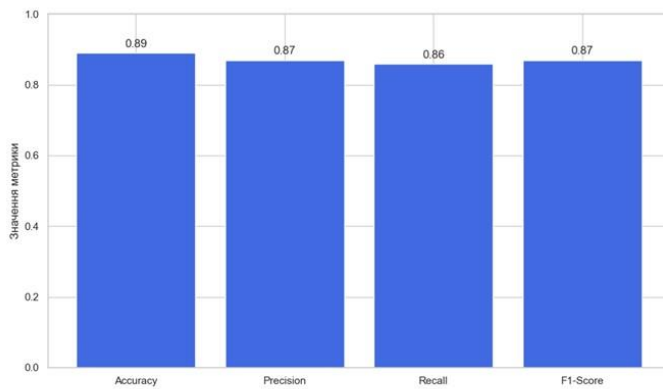
Схема застосування алгоритму XGBoost для прогнозування захворювання сухого ока



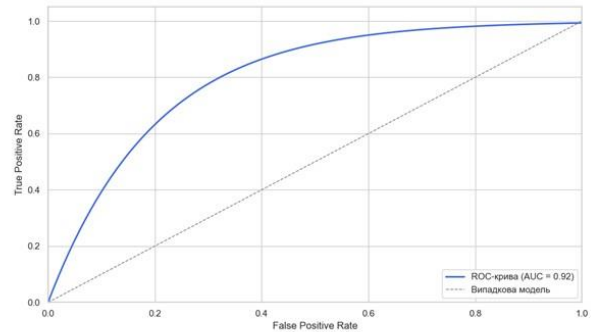
Процес обробки та нормалізації вхідних даних для прогнозування захворювання сухого ока



Метрики ефективності моделі XGBoost для прогнозування ЗСО



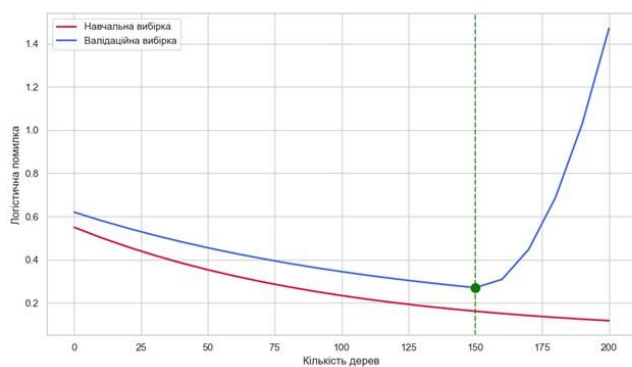
ROC-крива для моделі XGBoost



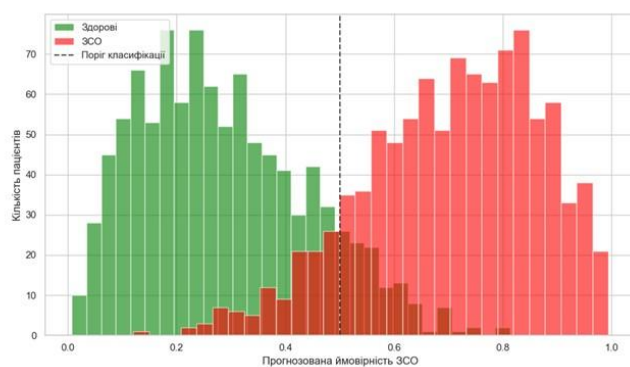
Важливість факторів для прогнозування захворювання сухого ока

Ранг	Фактор ризику	Відносна важливість %	Характер
1	Тривалість роботи з екраном	100	Поведінковий
2	Вік пацієнта	85	Демографічний
3	Якість сну	78	Поведінковий
4	Рівень стресу	72	Психологічний
5	Напруження очей	68	Офтальмологічний
6	Фізична активність	64	Поведінковий
7	Індекс очних симптомів	59	Композитний
8	Індекс сну та стресу	52	Композитний
9	Частота серцевих скорочень	48	Фізіологічний
10	Почервоніння очей	42	Офтальмологічний

Крива навчання моделі XGBoost



Розподіл прогнозованих ймовірностей ЗСО



Висновки

У представленій кваліфікаційній роботі бакалавра досягнуто мету підвищення ефективності ранньої діагностики та прогнозування розвитку синдрому сухого ока шляхом розробки методу прогнозного моделювання на основі аналізу поведінкових та фізіологічних факторів.

- виконано огляд теоретичних підходів та обрати підхід для прогнозного моделювання захворювання сухого ока на основі аналізу поведінкових та фізіологічних факторів;
- розроблено метод прогнозного моделювання захворювання сухого ока на основі аналізу поведінкових та фізіологічних факторів;
- описано інформаційну структуру системи для прогнозного моделювання захворювання сухого ока;
- виконано дослідження точності розробленого методу прогнозного моделювання захворювання сухого ока з використанням розробленої системи.

ДЯКУЮ ЗА УВАГУ!

Anti-Plagiarism (UA) v-15.281 Educational

The maximum coincidence with one document 4.0%

Dictionaries check: en_US, ru_RU, ua_UA. Errors in the documents: 9%

ID: 246813 Title: КВАЛІФІКАЦІЙНА РОБОТА БАКАЛАВРА на тему Метод прогнозного моделювання захворювання сухого ока на основі аналізу поведінкових та фізіологічних факторів Added in a DB: 2025-06-18 Authors: Єлизавета МАНДЗЮК Heads: Тетяна СКРИПНИК Consultants: Opponents:	Document		Sum coincidence on the DB	
	Symbols	Lexemes	Symbols	Lexemes
	83566	1256	6075 (7%)	93 (7%)

Plagiarism sources

ID	Description	Plagiarism presence in the document	
		Symbols	Lexemes

Протокол аналізу звіту подібності науковим керівником

Заявляю, що я ознайомився (-лась) з Повним звітом подібності, який був згенерований Системою виявлення і запобігання плагіату щодо роботи:

Автор: Єлизавета МАНДЗЮК

Співавтор:

Назва: КВАЛІФІКАЦІЙНА РОБОТА БАКАЛАВРА на тему Метод прогнозного моделювання захворювання сухого ока на основі аналізу поведінкових та фізіологічних факторів

Науковий керівник: Тетяна СКРИПНИК, ст. викладач кафедри КН

Підрозділ: Кафедра комп'ютерних наук

Коефіцієнт подібності 1: 4.7%

Коефіцієнт подібності 2: 2.3%

Мікропробіли: 0

Заміна букв: 3

Інтервали: 0

Білі знаки: 2

Дата створення звіту: 2025-06-18 19:32:00.0

Після аналізу Звіту подібності констатую наступне:

Запозичення, виявлені в роботі є законними і не є плагіатом. Рівень подібності не перевищує допустимої межі. Таким чином робота незалежна і приймається.

Запозичення не є плагіатом, але перевищено граничне значення рівня подібностей. Таким чином робота повертається на доопрацювання.

Виявлено запозичення і плагіат або навмисні текстові спотворення (маніпуляції), як передбачувані спроби укриття плагіату, які роблять роботу невідповідною вимогам законодавства (Ст. 32. ЗУ Про вищу освіту, пункт 3.1, Ст. 42. ЗУ Про освіту) та вимог НАЗЯВО (Критерій 5), а також кодексу етики і процедурам. Таким чином робота не приймається.

Обґрунтування:

2025-06-18

Дата

експерт

Т. Петровецька Р.О.

РІШЕННЯ ЕКСПЕРТНОЇ КОМІСІЇ КАФЕДРИ _____

ПРО ДОПУСК КВАЛІФІКАЦІЙНОЇ РОБОТИ ДО ЗАХИСТУ

Назва кваліфікаційної роботи Метод прогнозного моделювання захворювання сухого ока на основі аналізу поведінкових та фізіологічних факторів

Автор студентки групи КН-21-2 Єлизавета МАНДЗІЮК

Освітня програма Комп'ютерні науки

Рівень вищої освіти перший (бакалаврський)

Спеціальність 122 – Комп'ютерні науки

Науковий керівник: старший викладач кафедри комп'ютерних наук Тетяна Скрипник

На основі аналізу кваліфікаційної роботи на дотримання вимог академічної доброчесності (у т.ч. відсутності ознак академічного плагіату) з урахуванням результатів перевірки роботи спеціалізованим програмним засобом(ами) комісія зробила такий висновок:

№	Висновок	Позначка про відповідність
1	Ознаки академічного плагіату	
1.1	Запозичення, виявлені в роботі, є законними і не є академічним плагіатом (далі – зазначаються підстави віднесення запозичень до правомірних, якщо потрібно). Робота приймається до захисту.	<i>відсутні</i>
1.2	Виявлені запозичення не є академічним плагіатом, розміщені в розділах, які не описують безпосередньо авторське дослідження, але кількість цитат перевищує обсяг, виправданий поставленою метою роботи (далі – зазначаються детальні та аргументовані підстави віднесення запозичень до правомірних). Робота приймається до захисту, але має бути відкоригована.	
1.3	Виявлені запозичення не є академічним плагіатом, але частково розміщені в розділах, які описують безпосередньо авторське дослідження, а кількість цитат перевищує обсяг, виправданий поставленою метою роботи. Робота може бути допущена до захисту після того як буде відкоригована та доопрацьована і успішно пройде повторну перевірку на академічний плагіат.	
1.4	Робота містить навмисні текстові спотворення, передбачувані спроби укриття текстових запозичень або інші прояви академічного плагіату. Робота містить фабрикацію або фальсифікацію даних. Робота не допускається до захисту.	
2	Інші види порушень академічної доброчесності	<i>відсутні</i>

Підтвердження:

Запозичення виявлені в роботі Богдана Палійчука, не є плагіатом, оскільки запозичення розміщені в розділі огляду існуючих підходів, не описують безпосередньо авторську роботу і не стосуються її результатів; усі запозичення фрагментарні; до запозичень входять фрагменти, що не мають авторства і містять поширені конструкції; серед запозичень знаходяться загальновідомі терміни та скорочення. Рівень подібності не перевищує допустимої межі. Таким чином, робота є законною та приймається до захисту.

Обсяг запозичень, визначений системами виявлення збігів/ ідентичності/схожості, складає:

- за системою Anti-Plagiarism: 4%;

- за системою StrikePlagiarism КПІ: 4,7%.

18.06.2025

Завідувач кафедри

Гарант освітньої програми

Керівник кваліфікаційної роботи


Підпис

Підпис

Підпис

Олександр БАРМАК
Ім'я, ПРІЗВИЩЕ

Олександр МАЗУРЕЦЬ
Ім'я, ПРІЗВИЩЕ

Тетяна СКРИПНИК
Ім'я, ПРІЗВИЩЕ



**ВІДГУК НАУКОВОГО КЕРІВНИКА
на кваліфікаційну роботу бакалавра**

студента гр. КН-21-2 Мандзюк Єлизавети Дмитрівни
за темою Метод прогнозного моделювання захворювання сухого ока на основі аналізу поведінкових та фізіологічних факторів

1. Актуальність теми

Синдром сухого ока є поширеним офтальмологічним захворюванням, що набуває особливої значущості в умовах збільшення часу використання цифрових пристроїв. Розробка методів раннього прогнозування цього стану з урахуванням поведінкових і фізіологічних факторів є надзвичайно актуальною як для клінічної практики, так і для індивідуальної профілактики.

2. Відповідність роботи предметній області Стандарту спеціальності 122 Комп'ютерні науки

Тематика дослідження відповідає спеціальності 122 «Комп'ютерні науки», зокрема в частині застосування методів машинного навчання, роботи з медичними даними, аналітики та побудови прогнозних моделей.

3. Професійні та особистісні якості бакалавра

Мандзюк Єлизавета Дмитрівна проявила високий рівень відповідальності, старанності та зацікавленості у темі дослідження. Студентка вміло поєднувала теоретичні знання з практичними навичками, демонструючи самостійність і креативність.

4. Ступінь самостійності під час виконання кваліфікаційної роботи

Робота виконана самостійно, з дотриманням вимог академічної доброчесності. Джерела оформлено належним чином, результати отримано внаслідок особистої аналітичної та дослідницької роботи.

5. Ступінь оволодіння методами дослідження

У роботі застосовано сучасні методи машинного навчання (логістична регресія, дерева рішень, градієнтний бустинг), а також методи попередньої обробки, нормалізації та аналізу мультифакторних медичних даних. Для оцінки точності моделей використано метрики точності, recall, F1-score.

6. Повнота та якість розкриття теми роботи

Проведено ґрунтовний аналіз чинників, що впливають на розвиток синдрому сухого ока. Побудовано модель, яка враховує індивідуальні поведінкові й фізіологічні характеристики користувача, здійснено валідацію на тестовому наборі даних. Висновки підкріплені практичними результатами.

7. Логічність, послідовність, аргументованість, літературна грамотність викладення матеріалу

Текст роботи відзначається чіткістю, логічною структурою та науковим стилем викладення. Таблиці, графіки та пояснення сприяють кращому розумінню суті дослідження.

8. Можливість практичного застосування кваліфікаційної роботи бакалавра, окремих її частин

Результати можуть бути застосовані в рамках медичних інформаційних систем, мобільних застосунків для моніторингу зору, програм профілактики цифрової втоми. Методика може бути розширена на інші захворювання, що залежать від способу життя.

9. Висновок про можливість допуску кваліфікаційної роботи бакалавра до захисту, на яку оцінку заслуговує робота

Кваліфікаційна робота Мандзюк Єлизавети Дмитрівни виконана на високому рівні, повністю відповідає критеріям бакалаврської роботи, має наукову новизну та прикладну цінність.

Рекомендована оцінка — «відмінно».

Керівник _____



ст.викладач кафедри Тетяна СКРИПНИК



ХМЕЛЬНИЦЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
МОН УКРАЇНИ



Кафедра комп'ютерних наук

РЕЦЕНЗІЯ

на кваліфікаційну роботу бакалавра

студентки гр. КН-21-2 Єлизавети МАНДЗІЮК

за темою: Метод прогнозного моделювання захворювання сухого ока на основі аналізу поведінкових та фізіологічних факторів

1. Актуальність обраної теми

Тема роботи є актуальною через зростання поширеності захворювання сухого ока в умовах цифровізації суспільства та потребу в ранній діагностиці. Використання методів машинного навчання для прогнозування медичних станів відповідає сучасним тенденціям розвитку цифрової медицини.

2. Повнота розкриття мети та завдань роботи

Мета роботи сформульована чітко та конкретно. Всі поставлені завдання послідовно вирішені: проведено аналіз предметної області, розроблено метод прогнозування, створено програмну реалізацію та виконано експериментальну перевірку.

3. Зміст кожного розділу роботи

Перший розділ містить ґрунтовний аналіз сучасних підходів до прогнозування захворювання сухого ока. Другий розділ детально описує розроблений метод та архітектуру системи. Третій розділ представляє результати експериментальних досліджень з підтвердженням ефективності методу.

4. Оцінка розробленого методу та його практична цінність

Розроблений метод на основі алгоритму XGBoost демонструє високі показники ефективності. Практична цінність полягає в можливості раннього виявлення ризику захворювання та формування персоналізованих рекомендацій для пацієнтів.

5. Якість оформлення кваліфікаційної роботи бакалавра

Робота оформлена відповідно до вимог. Структура логічна, текст викладено зрозуміло, використано достатню кількість сучасних джерел літератури. Графічний матеріал доповнює текстову частину.

6. Недоліки кваліфікаційної роботи бакалавра

Серед незначних недоліків можна відзначити обмежений обсяг експериментальних даних.

7. Загальний висновок (допускається чи не допускається до захисту), та оцінка на яку заслуговує кваліфікаційна робота.

Враховуючи рівень виконання та забезпечення усіх необхідних вимог, робота може бути допущена до захисту. Рекомендована оцінка «Відмінно».

Рецензент

Андрій Ранський