Одеський національний університет імені І. І. Мечникова
Факультет математики, фізики та інформаційних технологій
Кафедра оптимального керування і економічної кібернетики

Кваліфікаційна робота

на здобуття ступеня вищої освіти «бакалавр»

«Система оцінки ризиків розвитку вітамін-D-дефіцитних станів на основі методів штучного інтелекту»

«AI-based system for assessment of the risk of vitamin D deficiency conditions»

Виконала: здобувачка денної форми навчання
спеціальності 113 Прикладна математика
Освітня програма «Прикладна математика»

Корхова Аріна Сергіївна

Керівник канд. фіз.-мат. наук, доц. Таірова М. С.
(науковий ступінь, вчене звання, прізвище та ініціали, підпис)
Рецензент канд. фіз.-мат. наук, доц. Стехун А. О.
(науковий ступінь, вчене звання, прізвище та ініціали)

Рекомендовано до захисту:
Протокол засідання кафедри
№ ___ від _____2024 р.

Захищено на засіданні ЕК №_____
протокол № ____ від _____ 2024 р.
Оцінка_____/___ ___/_____
(за національною шкалою, шкалою ECTS, бали)

Завідувач кафедри

_____- _____
(підпис) (прізвище, ініціали)

Голова ЕК

_____ _____
(підпис) (прізвище, ініціали)

Одеса – 2024

ЗМІСТ

ВСТУП

Актуальність. Дефіцит вітаміну D на сьогоднішній день вважається європейською та глобальною пандемією. Найбільш гострий дефіцит вражає країни з низьким і середнім рівнем доходу, де дефіцит вітаміну D зустрічається у 50-66% дорослих і 90-99% немовлят, тоді як у США - до 37% дорослі і до 46% темношкірих немовлят страждають на цей стан. Епідеміологічні дослідження виконані на території України встановили, що більшість населення має дефіцит вітаміну D - 81,8 %, недостатність вітаміну D відмічається у 13,6 % населення і лише 4,6 % жителів мають рівень 25(OH)D у сироватці крові в межах норми.

Вітамін D відноситься до жиророзчинних вітамінів і чинить плейотропний ефект в багатьох тканинах організму включаючи кістково-м'язову систему та адипоцити. Численні епідеміологічні, експериментальні та клінічні дослідження вказують на наявність кореляційних зв'язків між рівнем вітаміну D у сироватці крові та ризиком розвитку дисліпідемій, порушенням вуглеводного обміну та захворювань серцево-судинної системи. Ці хвороби створюють значне навантаження на громадське здоров'я підвищуючи витрати на систему охорони здоров'я, сприяють росту рівня захворюваності та смертності. Доведено зв'язок рівнів вітаміну D з ліпопротеїдами високої щільності (ЛПВЩ), ліпопротеїдами низької щільності (ЛПНЩ), ліпопротеїдами дуже низької щільності (ЛПДНЩ), тригліцеридами (ТГ) та коефіцієнтом атерогенності (КА).

Окремо встановлено, що причиною низької концентрації 25-гідроксивітаміну D в сироватці крові є надлишкова вага та ожиріння [1]. В літературі описані кореляційні зв'язки індексу маси тіла (ІМТ) та рівнів 25(OH)D [2, 3].

За даними статистики найбільша кількість населення з ожирінням зафіксована на Тихоокеанських островах в державі Науру – 88,5%. Понад 80% населення мають зайву вагу в Палау, на Маршаллових островах. У Кувейті 73% населення мають зайву вагу, а в США – 67,9%. В Європейських країнах

кількість пацієнтів з ожирінням відрізняється: у Великій Британії зайву вагу мають 63,75% населення, Іспанії – 61,6%, Чехії – 62,3%, Литві – 57,8%, Польщі – 58,3%, Естонії – 55,8%. За даними Всесвітньої організації охорони здоров'я на 2016 рік – 58,4% українців старше 18 років мали зайву вагу.

Машинне навчання, область штучного інтелекту, дозволяє комп’ютерам вивчати дані, щоб робити прогнози. Використання штучного інтелекту в медицині привернуло увагу вчених через його потенціал для трансформації системи прогнозування, діагностики та лікування різних патологічних станів. В останні роки з’являється все більше публікацій, присвячених використанню алгоритмів машинного навчання для дослідження ризиків розвитку дефіциту вітаміну D та його корекції [5, 6]. Моделі машинного навчання здатні ідентифікувати закономірності та зв’язки, які можуть бути неочевидними за допомогою звичайних статистичних методів, завдяки використанню великомасштабних наборів даних і розширених аналітичних інструментів.

У цій роботі досліджується застосування алгоритмів машинного навчання для прогнозування дефіциту вітаміну D на основі широкого спектру джерел даних. Моделі машинного навчання здатні ідентифікувати закономірності та зв’язки, які можуть бути неочевидними за допомогою звичайних статистичних методів завдяки використанню великомасштабних наборів даних і передових аналітичних інструментів.

Сьогодні більшість науковців світу визнають необхідність ранньої діагностики та профілактики вітамін-D-дефіцитних станів, особливо в групах ризику [4, 5, 6]. Таким чином, актуальним стає своєчасний скринінг дефіциту та недостатності вітаміну D, оскільки корекція статусу 25(OH)D є більш легким завданням, аніж лікування захворювань пов’язаних з його низьким рівнем. Через це кількість лабораторних запитів на визначення 25(OH)D сироватки крові зростає, що додатково спричиняє збільшення витрат на систему охорони здоров’я [7, 8].

Створення математичної моделі за допомогою, якої можна визначити рівень вітаміну D сприятиме зменшенню витрат та дозволить проводити скринінг статусу вітаміну D більш масово.

Мета. Створення математичної моделі, яка дозволяє спрогнозувати вплив кожного із лабораторних і антропометричних показників на рівень 25(OH)D сироватки крові та у подальшому дозволить розробити систему ранньої діагностики та профілактики вітамін D дефіцитних станів.

Предмет дослідження. Рівень вітаміну D серед дорослого населення жителів Півдня України.

Об’єкт дослідження. Моделі машинного навчання для визначення рівня 25(OH)D.

Методи дослідження. Методи розвідувального, кореляційного аналізу, методи машинного навчання та штучного інтелекту.

РОЗДІЛ 1
МАТЕРІАЛИ ТА МЕТОДИ

У ході дослідження було обстежено 928 жителів (жінок - 507; чоловіків - 421) Півдня України (Херсонська, Миколаївська та Одеська область) у віці від 19 до 82 років (середній вік жінок — 47.7±15.3 років, чоловіків — 46.7±15.5 років).

Для визначення ступеня ожиріння ІМТ визначали за формулою маса тіла/зріст (кг/м2) згідно з рекомендаціями Міжнародної групи з питань ожиріння ВООЗ (WHO, 1997). Критерієм абдомінального ожиріння вважали співвідношення ОТ/ОС більше 0,8. Визначення рівня вітаміну 25(OH)D total (оцінка загального рівня 25(OH)D2 та 25(OH)D3) відбувалась за допомогою автоматичного імунохімічного аналізатора Architech i2000sr (ТОВ «СМАРТЛАБ»). Ліпідний спектр крові визначали за загальноприйнятими показниками (ЗХС, ТГ, ЛПВЩ, ЛПДНЩ, ЛПНЩ, КА). Концентрацію ЗХС, ТГ, ХС ЛПВЩ визначали ферментативно-колометричним методом на автоматичному біохімічному аналізаторі Cobas 6000; Roche Diagnostics.

Дослідження проводилось протягом календарного року, що дало змогу оцінити коливання рівня 25(OH)D у різні місяці з різною тривалістю інсоляції. Попередньо серед всіх досліджуваних було проведено анкетування, що дозволило виключити з дослідження пацієнтів з аутоімунними захворюваннями; онкологічною патологією; хронічними захворюваннями печінки та нирок; вагітних та жінок, які перебувають на грудному вигодовуванні; прийом лікарських засобів, що впливають на метаболізм (глюкокортикоїди, замісна гормональна терапія, антиконвульсанти, та ін.), а також препаратів, що містять вітамін D.

На наступному етапі дослідження проводилось лабораторне визначення рівня 25(OH)D сироватки крові та показників ліпідного обміну (загальний холестерин (ЗХ), тригліцериди (ТГ), ліпопротеїни низької щільності (ЛПНЩ), ліпопротеїни дуже низької щільності (ЛПДНЩ), ліпопротеїни високої

щільності (ЛПВЩ), коефіцієнт атерогенності (КА)) – виконувалися за загальноприйнятими методиками.

Рівень 25(OH)D у сироватці крові оцінювали відповідно до рекомендацій Комітету з рекомендацій з ендокринної практики (Endocrine Practice Guidelines Committee) та (Institute of Medicine) (Таблиця 1.1).

Таблиця 1.1

Рівні 25(OH)D сироватки крові згідно рекомендацій Комітету ендокринологів із створення настанов із клінічної практики (Endocrine Practice Guidelines Committee) та Інституту медицини (Institute of Medicine)

Значення	Рівень 25(OH)D сироватки крові
Дефіцит	<20 нг/мл (<50 нмоль/л)
Недостатність	21-29 нг/мл (51-74 нмоль/л)
Достатній рівень	≥30 нг/мл (≥75 нмоль/л)
Токсичний рівень	>150 нг/мл (>375 нмоль/л)

Аналіз отриманих даних проводили за допомогою спеціалізованих відкритих бібліотек для машинного навчання, наукових обчислень, аналізу та візуалізації даних Scikit-learn [26], NumPy [9], SciPy [10], Pandas [11], Matplotlib [12], Seaborn [27]. Для статистичної обробки результатів дослідження використовували первинні описові статистики, тестування на підпорядкованість даних нормальному розподілу, а також методи кореляційного та регресійного аналізу. В якості прогнозних моделей були використані моделі багатофакторної лінійної регресії та багатофакторної логістичної регресії. Адекватність моделей лінійної регресії оцінювалася за допомогою скоригованих коефіцієнтів детермінації. Впливовість факторів моделі оцінювалася за допомогою стандартизованих коефіцієнтів регресії та коефіцієнтів еластичності. Для оцінки якості моделей логістичної регресії використовувалися метрики на основі матриці помилок та ROC-аналіз. Статистично значущими вважались результати при $p<0,001$. Коефіцієнт Джині та scoring='accuracy' — метрики якості, що використовувалися для оцінки передбачуваної моделі дерева рішень.

РОЗДІЛ 2
РОЗВІДУВАЛЬНИЙ АНАЛІЗ

Учасники дослідження мали рівні сироватки крові 25(OH)D в діапазоні від 4,31 нг/мл до 89,19 нг/мл (середній рівень у жінок становив 26,5±13,3 нг/мл, у чоловіків — 26,8±11,7 нг/мл). У досліджуваній групі поширеність дефіциту, недостатності та достатнього рівня вітаміну D становила відповідно 33,6%, 33% та 33,4%. Чоловіки (65,5%) і жінки (64,3%) становили майже однакову кількість пацієнтів з рівнем 25-гідроксивітаміну D ≤ 30 нг/мл (Таблиця 2.1).

Таблиця 2.1

Рівень 25(OH)D у жінок та чоловіків

Рівень 25(OH)D	Чоловіки		Жінки	
	Абс.	%	Абс.	%
Дефіцит	136	32,30	176	34,71
Недостатність	140	33,25	150	29,59
Достатній рівень	145	34,44	181	35,70

2.1. Аналіз за показниками ліпідного обміну

Індекс маси тіла у групі дослідження коливався від 16.1 кг/м^2 до 41.3 кг/м^2, при цьому середній ІМТ у чоловіків становив 25.2±4.0 кг/м^2, а у жінок — 25.6±4.5 кг/м^2. Кореляційний аналіз зв'язку між індексом маси тіла та рівнем 25(OH)D у сироватці крові показав наявність статистично значущого зворотного зв’язку (коефіцієнт Спірмена r=-0.181, p<0.001) (Рис. 2.1). Це вказує на те, що рівень 25(OH)D залежить від ІМТ і свідчить про те, що люди з надмірною вагою та ожирінням більш схильні до станів дефіциту вітаміну D.

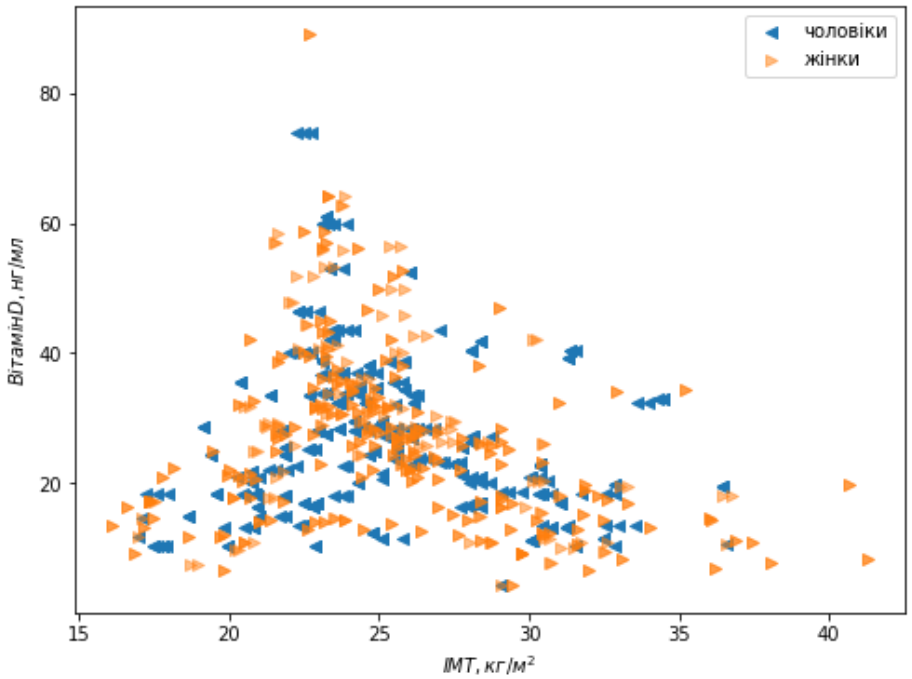

Рис. 2.1. Кореляція індекса маси тіла та рівня 25(OH)D у чоловіків та жінок

Дані обстеження показують, що середній рівень ІМТ у групі дослідження був найвищим у пацієнтів з дефіцитом вітаміну D і найнижчим — у пацієнтів з достатнім рівнем 25(OH)D (Рис. 2.2).

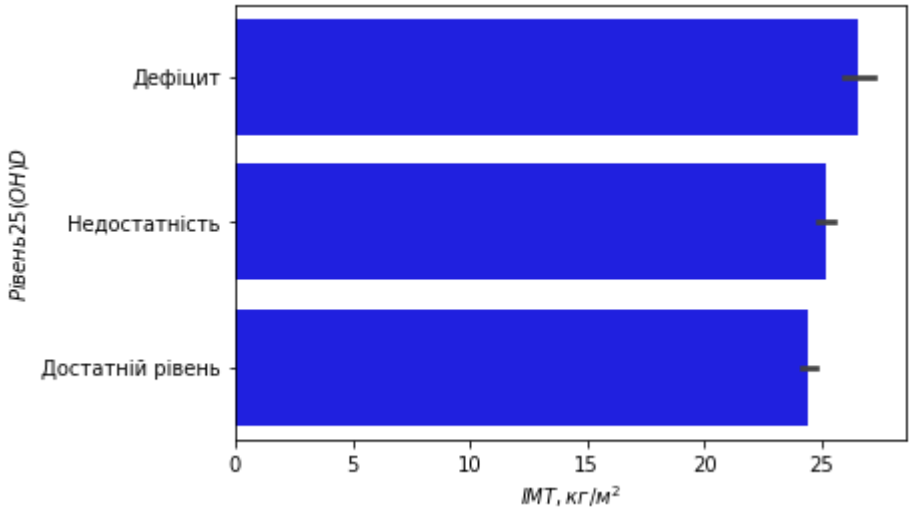

Рис. 2.2. Середній індекс маси тіла в залежності від рівня 25(OH)D

Статистичне порівняння середнього індексу маси тіла у пацієнтів з дефіцитом, недостатністю та достатнім рівнем вітаміну D за допомогою t-тесту Ст'юдента показало, що ці групи достовірно відрізняються одна від одної (Таблиця 2.2),

що підтверджує зроблені раніше висновки щодо сприйнятливості до вітамін D дефіцитних станів у людей з ожирінням та надлишковою вагою.

Таблиця 2.2

Статистичне порівняння індексу маси тіла в залежності від рівня 25(OH)D

Рівень вітаміну D та середній показник ІМТ, M±m, кг/м^2		p-value
Дефіцит, 26.6±6.1	Недостатність, 25.2±3.0	<0.001
Дефіцит, 26.6±6.1	Достатній рівень, 24.5±2.6	<0.001
Недостатність, 25.2±3.0	Достатній рівень, 24.5±2.6	0.002

2.2. Аналіз за антропометричними показниками

Крім того, було проведено аналіз антропометричних даних, щоб з'ясувати вплив цих показників на рівень 25(OH)D у сироватці крові. Зокрема, оцінювали співвідношення окружності талії до окружності стегон (ОТ/ОС).

За даними дослідження, індекс ОТ/ОС та рівень вітаміну D мають білтш слабкий зв'язок, ніж показник індексу маси тіла (коефіцієнт кореляції Спірмена становить r=-0.09, p=0.006) (Рис. 2.3). Це свідчить про те, що, хоча залежність не така виражена, як для ІМТ, пацієнти з вищими показниками ОТ/ОС більш вразливі до захворювань, пов'язаних з дефіцитом вітаміну D.

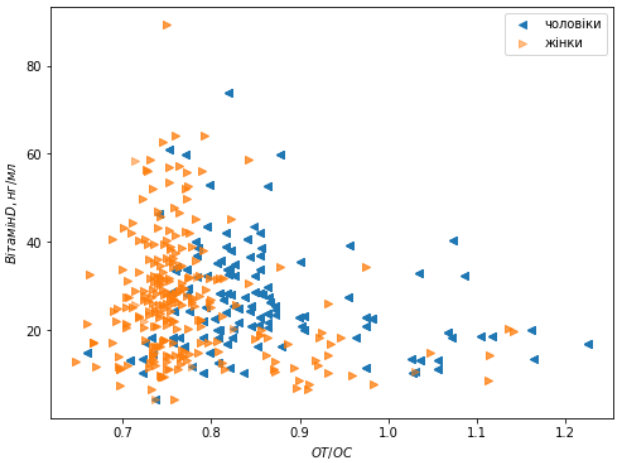

Рис. 2.3. Кореляція індекса ОТ/ОС та рівня 25(OH)D у чоловіків та жінок

Щоб отримати більше доказів сприйнятливості пацієнтів із надмірною вагою до дефіциту вітаміну D, також було проведено статистичне порівняння середніх показників ОТ/ОС для пацієнтів із різними рівнями вітаміну D. На рисунку 2.4 показано середні рівні та стандартні відхилення індексу ОТ/ОС.

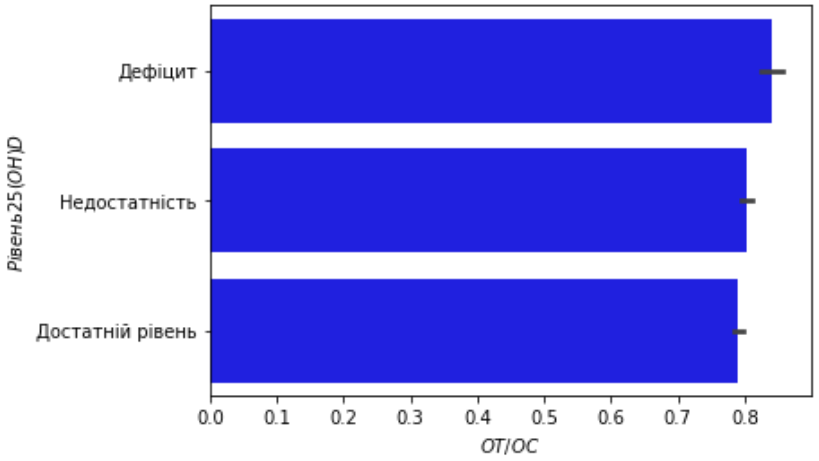

Рис. 2.4. Середній індекс ОТ/ОС в залежності від рівня 25(OH)D

Статистичне дослідження t-критерію Ст'юдента груп пацієнтів із різними рівнями вітаміну D показало, що середні індекси ОТ/ОС у цих групах достовірно відрізнялися один від одного (Таблиця 2.3). Таким чином, можна зробити статистично підтверджений висновок, що люди з надмірною вагою або ожирінням схильні до дефіциту вітаміну D.

Таблиця 2.3

Статистичне порівняння індексу ОТ/ОС в залежності від рівня 25(OH)D

Рівень вітаміну D та середній індекс ОТ/ОС, M±m		p-value
Дефіцит, 0.84±0.14	Недостатність, 0.8±0.08	<0.001
Дефіцит, 0.84±0.14	Достатній рівень, 0.79±0.07	<0.001
Недостатність, 0.8±0.08	Достатній рівень, 0.79±0.07	0.034

Комплексне статистичне дослідження з використанням коефіцієнтів еластичності дозволяє кількісно оцінити залежність рівня 25(OH)D сироватки крові від ІМТ та ОТ/ОС. Таким чином, кількість вітаміну D падає в середньому на 0,603 нг/мл ($p < 0,001$) при збільшенні ІМТ на 1 кг/м^2 і на 0,163 нг/мл ($p < 0,001$) при збільшенні окружності талії на 1 см. Таким чином, ступінь ожиріння є одним із факторів, пов'язаних з недостатністю вітаміну D.

Після дослідження індексу маси тіла встановлено, що 36 хворих (чоловіки: 11; жінки: 25) мали дефіцит маси тіла, 453 пацієнти (чоловіки: 225; жінки: 228) мали норму, 284 пацієнти (чоловіки: 113; жінки: 171) мали надмірну вагу, 127 пацієнтів (чоловіки: 65; жінки: 62) мали ожиріння першого ступеня, 24 пацієнти (чоловіки: 7; жінки: 17) мали ожиріння другого ступеня та 4 пацієнти (чоловіки: 0 жінки: 4) мали ожиріння третього ступеня.

Середнє та стандартне відхилення рівня 25(OH)D сироватки крові для кожної групи було розраховано відносно індексу маси тіла (Рис. 2.5).

Для кожної з груп відносно індексу маси тіла були розраховані середні та стандартні відхилення рівня 25(OH)D сироватки крові (Рис. 2.5). Виходячи з розрахунків, виявилося, що пацієнти з недостатньою вагою та ожирінням мали дефіцит і недостатність вітаміну D (середній рівень вітаміну D у цих групах був менше 20 нг/мл). Також можна спостерігати тенденцію до зниження середнього рівня вітаміну D із зростанням ІМТ.

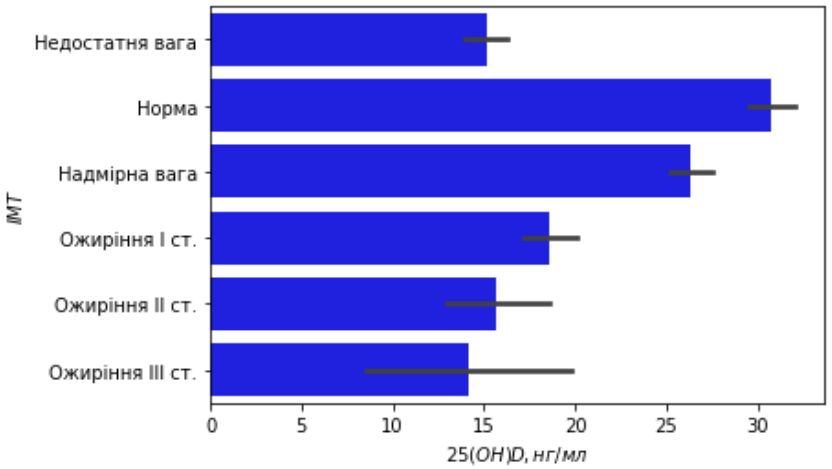

Рис. 2.5. Середній рівень 25(OH)D в залежності від ІМТ (за групами)

У таблиці 2.4 показано, що у групах із недостатньою вагою та ожирінням була найбільша частка пацієнтів із рівнем вітаміну D менше 30 нг/мл, тоді як у групі з нормальною вагою таких пацієнтів було найменше.

Таблиця 2.4

Статистичне порівняння груп пацієнтів різного стана ожиріння в залежності від рівня 25(OH)D

Рівень 25(OH)D	Оцінка індексу маси тіла					
	Недостатня вага	Норма	Надмірна	Ожиріння I ступеню	Ожиріння II ступеню	Ожиріння III ступеню
Дефіцит	88.9%	23.2	21.5	69.3	91.7	100.0
Недостатність	11.1	25.2	52.8	17.3	0.0	0.0
Достатній	0.0	51.7	25.7	13.4	8.3	0.0

За виключенням груп із ожирінням I, II та III ступенів порівняння груп за індексом маси тіла виявило статистично значущі відмінності середніх рівнів 25(OH)D у сироватці крові в усіх групах (Таблиця 2.5). У результаті було підтверджено результати попередніх етапів дослідження, які підтверджують теорію про те, що рівень вітаміну D залежить від ІМТ.

Таблиця 2.5

Середній рівень 25-гідроксивітаміну D сироватки крові в залежності від індексу маси тіла

ІМТ та середній рівень 25(OH)D, M±m, нг/мл		p-value
Недостатня вага, 15.1±3.7	Норма, 30.7±13.6	<0.001
Норма, 30.7±13.6	Надмірна вага, 26.3±10.1	<0.001
Надмірна вага, 26.3±10.1	Ожиріння I ступеня, 18.6±8.4	<0.001
Ожиріння I ступеня, 18.6±8.4	Ожиріння II ступеня, 15.7±7.3	0.09
Ожиріння II ступеня, 15.7±7.3	Ожиріння III ступеня, 14.1±6.5	0.68
Недостатня вага, 15.1±3.7	Ожиріння I ступеня, 18.6±8.4	<0.001

РОЗДІЛ 3
МАТЕМАТИЧНЕ МОДЕЛЮВАННЯ ТА ШТУЧНИЙ ІНТЕЛЕКТ

3.1. Математичне моделювання залежності рівня 25(OH)D сироватки крові від антропометричних та лабораторних показників

Було побудовано моделі багатофакторної лінійної регресії із різними комбінаціями чинників, за допомогою яких визначався вплив певних факторів на формування значення досліджуваного показника:

$$y = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \ldots + \alpha_n x_n, \quad (3.1)$$

де y — значення показника, що моделюється; α_0, …, α_n — коефіцієнти рівняння лінійної множинної регресії; x_1, …, x_n — фактори, які потенційно впливають на значення показника y. Знак коефіцієнта α_i вказує на напрямок взаємозв'язку чинника з індексом i та досліджуваного показника, тобто схильність до збільшення або зменшення значення показника, що моделюється, за рахунок збільшення обраного чинника [13].

Першим етапом побудови моделі є визначення показників, які доцільно до неї включити. Для цього спочатку проводився кореляційний аналіз показників, що вивчалися. Попередньо перевірялася нормальність законів розподілу змінних. Було встановлено, що основний досліджуваний показник — рівень 25(OH)D сироватки крові — має відхилення від нормального розподілу (статистика тесту Д'Агостіно-Пірсона дорівнює 154,1; p-value < 0,01), тому в якості показника ступеня тісноти статистичного зв'язку було обрано коефіцієнт кореляції Спірмена. Враховуючи результати кореляційного аналізу, до математичних моделей включалися лише ті показники, в яких $p<0.001$, тобто зв'язок був підтверджений (Таблиця 3.1).

Таблиця 3.1

Кореляційні зв’язки між рівнем 25-гідроксивітміном D та іншими показниками

Показник	r	p
Вік	-0.469	<0.001
ІМТ	-0.181	<0.001
КА	-0.171	<0.001
ТГ	-0.161	<0.001
ЛПДНЩ	-0.158	<0.001
ЛПВЩ	0.151	<0.001
ЛПНЩ	-0.102	0.002
Загальний холестерин	-0.09	0.006
ОТ/ОС	-0.09	0.006
Стать	-0.01	0.753

Отже, для аналізу розглядаються такі показники: вік, індекс маси тіла (ІМТ), коефіцієнт атерогенності (КА), тригліцериди (ТГ), ліпопротеїди дуже низької щільності (ЛПДНЩ), ліпопротеїди високої щільності (ЛПВЩ). А такі показники як ліпопротеїди низької щільності (ЛПНЩ), загальний холестерин, співвідношення об'єму талії та об'єму стегон та стать надалі у моделях не розглядалися ($p>0.001$).

Наступним етапом було проведення математичного моделювання, яке дозволяє оцінити вплив чинників на значення рівня 25(OH)D і дає можливість спрогнозувати це значення на основі мінімальної кількості факторів.

В якості показника адекватності регресійної моделі зазвичай використовують коефіцієнт детермінації R^2, який, втім, має недолік: при

включенні додаткового чинника до рівняння моделі коефіцієнт R^2 може тільки збільшитись, що може призводити до отримання хибних результатів моделювання. Тому для більш якісної оцінки адекватності моделей в даній роботі були використані такі показники, як: коефіцієнт детермінації, скоригований за Тейлом R^2_T та коефіцієнт детермінації, скоригований за Аземією R^2_A. Вони коригують коефіцієнт детермінації з урахуванням кількості факторів, які входять до різних моделей, тобто зменшують вплив залежності значення коефіцієнта детермінації від кількості чинників [13].

У таблиці 3.2 представлені стандартизовані коефіцієнти та показники адекватності багатофакторних лінійних моделей, побудованих на чинниках, що впливають на значення рівня 25(OH)D.

Таблиця 3.2

Моделі оцінки значення рівня 25(OH)D за даними лабораторного дослідження (стандартизовані коефіцієнти)

Модель (y)	R^2_T	R^2_A	Фактори впливу (x), коефіцієнти (α)					
			вік, α_1	ІМТ, α_2	КА, α_3	ТГ, α_4	ЛПДНЩ, α_5	ЛПВЩ, α_6
1	0.223	0.217	-0.013	-0.055	-0.012	-0.145	0.182	0.118
2	0.224	0.219	-0.013	-0.055	-0.011	-0.126	-	0.075
3	0.224	0.219	-0.014	-0.054	-0.019	-0.021	-	-
4	**0.224**	**0.220**	**-0.013**	**-0.055**	**-0.02**	**-**	**-**	**0.09**
5	0.224	0.220	-0.014	-0.054	-0.025	-	-	-
6	0.219	0.216	-0.014	-0.058	-	-	-	-
7	0.199	0.197	-0.014	-	-	-	-	-

До таблиці 3.2 були включені моделі з найбільшими значеннями показників адекватності та різною кількістю чинників. Виходячи з показника R^2_T можемо побачити, що найкращими є моделі 2, 3, 4 та 5, а за показником R^2_A — моделі 4 та 5. Для подальшого аналізу оберемо модель 4, оскільки вона включає чинник ЛПВЩ, який безпосередньо впливає на коефіцієнт атерогенності та ризик розвитку атеросклерозу. За даними кореляційного аналізу пацієнти з більш високим рівнем ЛПВЩ мали більш високий рівень 25(OH)D сироватки крові.

Зроблені розрахунки показують, що для чотирифакторної моделі 4 формула розрахунку значення 25(OH)D є наступною:

$$25(OH)D = 53.622 - 0.342 \cdot вік - 0.407 \cdot ІМТ - 0.631 \cdot КА + 1.004 \cdot ЛПВЩ, \quad (3.2)$$

для якої $R^2_T = 0.224$ та $R^2_A = 0.220$. Зі значень коефіцієнтів попередньо можемо зробити висновок, що найбільш впливовими чинниками моделі є ЛПВЩ (значення стандартизованого коефіцієнта дорівнює 0.09), ІМТ (значення стандартизованого коефіцієнта дорівнює -0.055) та КА (значення стандартизованого коефіцієнта дорівнює -0.02).

Для оцінки статистичної значущості моделей використовувася F-критерій Фішера. Для моделі 4 спостережуване значення F-статистики дорівнює 55.39, у той час як критичне значення дорівнює 0.12. Це свідчить про те, що модель 4 є статистично значущою та може бути використаною для прогнозування рівня 25(OH)D.

На рисунку 3.1 представлено порівняння прогнозних значень рівня 25(OH)D за моделлю 4 та їх реальних значень, отриманих під час лабораторного обстеження.

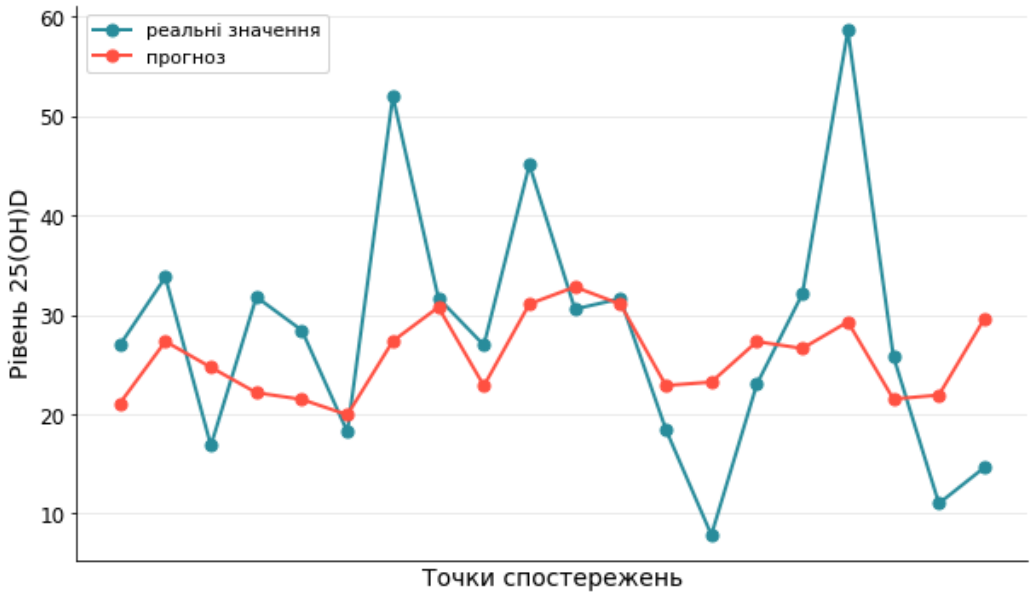

Рис. 3.1. Порівняння прогнозних значень рівня 25(OH)D за моделлю 4 та їх реальних значень, отриманих під час лабораторного обстеження

Тепер розрахуємо часткові коефіцієнти детермінації для моделі 4

$$25(\text{OH})\text{D} = 53.622 - 0.342 \cdot \text{вік} - 0.407 \cdot \text{ІМТ} - 0.631 \cdot \text{КА} + 1.004 \cdot \text{ЛПВЩ},$$

$$R^2_T = 0.224,\ R^2_A = 0.220. \qquad (3.3)$$

Частковий коефіцієнт детермінації показує, на яку величину зменшиться коефіцієнт детермінації, якщо певний регресор виключити з моделі. Тобто можна зробити висновок, що чим більший частковий коефіцієнт детермінації, тим більш впливовим є у моделі відповідний регресор [13].

$$\Delta R^2_1 = 0.000109,\ \Delta R^2_2 = 0.000062,\ \Delta R^2_3 = 0.000005,\ \Delta R^2_6 = 0.000279.$$

Зважаючи на величини ΔR^2_i, можна зробити висновок, що найбільш впливовим фактором у моделі 4 є рівень ЛПВЩ, потім вік та ІМТ.

У таблиці 3.3 представлені показники впливовості факторів отриманих моделей.

Таблиця 3.3

Показники впливовості факторів отриманих моделей

Модель	R^2	R^2_T	R^2_A
1 (вік, ІМТ, КА, ТГ, ЛПДНЩ, ЛПВЩ)	0.229	0.223	0.217
2 (вік, ІМТ, КА, ТГ, ЛПВЩ)	0.229	0.224	0.219
3 (вік, ІМТ, КА, ТГ)	0.227	0.224	0.219
4 (вік, ІМТ, КА, ЛПВЩ)	**0.228**	**0.224**	**0.219**
5 (вік, ІМТ, КА)	0.227	0.224	0.220
6 (вік, ІМТ)	0.221	0.219	0.216
7 (вік)	0.200	0.199	0.197

У таблиці 3.3 видно, що додавання шостого фактору до моделі зменшує коефіцієнти детермінації скориговані за Тейлом та за Амемією, а додавання четвертого та п’ятого факторів мало незначний вплив. Отже, можна зробити висновок, що їх додавання є надлишковим.

Для визначення міри впливу фактору на залежну змінну без урахування одиниць їх виміру були використані коефіцієнти еластичності. Коефіцієнт еластичності показує, на скільки відсотків зміниться рівень вітаміну D, якщо при інших рівних умовах певний чинник збільшити на один відсоток [13]. У таблиці 3.4 представлені коефіцієнти еластичності розраховані для моделей 1-7.

Таблиця 3.4

Коефіцієнти еластичності для отриманих моделей

Модель	вік, ε_1	ІМТ, ε_2	КА, ε_3	ТГ, ε_4	ЛПДНЩ, ε_5	ЛПВЩ, ε_6
1	-0.007	-0.087	-0.020	-0.223	0.078	0.139
2	-0.007	-0.088	-0.019	-0.039	-	0.032
3	-0.007	-0.086	-0.032	-0.032	-	-
4	**-0.007**	**-0.088**	**-0.034**	**-**	**-**	**0.138**
5	-0.007	-0.086	-0.043	-	-	-
6	-0.007	-0.092	-	-	-	-
7	-0.008	-	-	-	-	-

Як видно з таблиці 3.4, коефіцієнт еластичності ε_6 моделі 4 показує, що якщо показник ЛПВЩ збільшити на 1%, то рівень 25(OH)D збільшиться на 0.138%. Аналогічні висновки отримано і для інших коефіцієнтів еластичності.

3.2. Математичне моделювання залежності ризику вітамін D дефіцитних станів від антропометричних та лабораторних показників

З практичної точки зору, доцільніше прогнозувати не рівень вітаміну D, а його статус, тобто ризик дефіциту або недостатності. З цією метою отриманий раніше показник 25(OH)D був перетворений у бінарну змінну відповідно до його значень. Згідно з класифікацією, розробленою Міжнародним Інститутом Медицини та Комітетом ендокринологів зі створення настанов із клінічної практики та Методичних рекомендацій з лікування та профілактики дефіциту вітаміну D у населення країн Центральної Європи, концентрації <30 нг/мл визначалися як недостатній рівень вітаміну D [14].

Для попередньої оцінки рівня впливу досліджуваних антропометричних та лабораторних показників на ризик недостатності вітаміну D було розраховано бісеріальні коефіцієнти кореляції, які демонструють тісноту та напрямок зв’язку між чинником та ризиком дефіциту вітаміну D. Їх значення та рівні достовірності наведені у таблиці 3.5.

Таблиця 3.5

Кореляційні зв’язки між ризиком недостатності 25-гідроксивітаміну D та іншими показниками

Показник	r	p
Вік	0.417	<0.001
ІМТ	0.150	<0.001
КА	0.128	<0.001
ТГ	0.075	0.02
ЛПДНЩ	0.071	0.03
ЛПВЩ	-0.117	<0.001
ЛПНЩ	0.091	0.006
Загальний холестерин	0.031	0.34
ОТ/ОС	0.144	<0.001

Отже, можна зробити висновок, що найбільш впливовими чинниками є вік, індекс маси тіла, коефіцієнт атерогенності, ліпопротеїди високої щільності та співвідношення окружності талії до окружності стегон.

Далі було побудовано математичні моделі за допомогою алгоритму класифікації на основі логістичної регресії для прогнозування ризику дефіциту рівня 25(OH)D сироватки крові. У рамках логістичного регресійного аналізу

ймовірність недостатнього рівня 25(OH)D, виражена через логістичну регресію, може бути подана у вигляді наступного рівняння:

$$OR = P / (1 - P), P = \frac{1}{1+e^{-y}}, \quad (3.4)$$

де $log(OR) = y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n$, P – ймовірність недостатнього рівня вітаміну D.

У таблиці 3.6 дано значення відношення шансів (OR) і 95% довірчий інтервал (CI) моделей множинної логістичної регресії для моделей 1-6. 95% довірчий інтервал – це діапазон значень, в межах якого справжнє значення сукупності буде лежати з 95% ймовірністю [15].

Таблиця 3.6

Значення відношення шансів (OR) і 95% довірчий інтервал (CI) моделей множинної логістичної регресії

Модель	Чинник, OR (95% CI)					
	Вік	ІМТ	КА	ТГ	ЛПДНЩ	ЛПВЩ
1	0.056 (0.045, 0.067)	0 (-0.030, 0.030)	-0.118 (-0.274, 0.038)	3.940 (-1.816, 9.696)	-8.543 (-21.110, 4.024)	-0.995 (-1.340, -0.649)
2	0.057 (0.046, 0.067)	0.001 (-0.029, 0.031)	-0.128 (-0.282, 0.027)	0.031 (-0.083, 0.145)	-	-1.015 (-1.360, -0.670)
3	0.052 (0.041, 0.062)	-0.063 (-0.084, -0.041)	0.017 (-0.137, 0.154)	0.022 (-0.097, 0.119)	-	-
4	0.057 (0.046, 0.067)	-0.001 (-0.030, 0.029)	-0.102 (-0.226, 0.022)	-	-	-1.009 (-1.352, -0.665)
5	0.052 (0.041, 0.062)	-0.063 (-0.084, -0.042)	0.017 (-0.104, 0.137)	-	-	-
6	0.052 (0.041, 0.062)	-0.061 (-0.079, -0.043)	-	-	-	-

У таблиці 3.7 представлені коефіцієнти логістичної регресії при відповідних показниках для отриманих раніше моделей 1-6.

Таблиця 3.7

Коефіцієнти логістичної регресії показників моделей 1-6

Модель	Коефіцієнти (β)					
	вік	ІМТ	КА	ТГ	ЛПДНЩ	ЛПВЩ
1	0.064	0.036	-0.059	0.261	-0.593	-0.664
2	0.064	0.037	-0.062	-0.009	-	-0.665
3	0.065	0.043	0.067	-0.041	-	-
4	**0.064**	**0.037**	**-0.069**	**-**	**-**	**-0.668**
5	0.065	0.044	0.035	-	-	-
6	0.065	0.045	-	-	-	-

Для оцінки якості моделі використовується таблиця спряженості, також відома як матриця невідповідностей, що будується на основі результатів класифікації (Таблиця 3.8). Рядки цієї матриці містять справжні значення класів, а стовпці – класи, передбачені моделлю. В результаті побудови моделі отримають чотири можливі типи результатів: TP (True Positives) – правильно передбачені позитивні об’єкти класу (істинно позитивні випадки); TN (True Negatives) – правильно передбачені негативні об’єкти класу (істинно негативні випадки); FN (False Negatives) – позитивні об’єкти класу, передбачені як негативні (хибнонегативні приклади); FP (False Positives) – негативні об’єкти, передбачені як позитивні (хибнопозитивні випадки) [16]. В рамках даного дослідження клас 0 асоціюється із достатнім рівнем вітаміну D, клас 1 — із недостатнім рівнем або дефіцитом вітаміну D.

Таблиця 3.8

Матриця помилок на основі результатів класифікації

		Передбачені	
		Негативно (0)	Позитивно (1)
Справжні	Негативно (0)	TN	FP
	Позитивно (1)	FN	TP

На основі матриці помилок були обчислені такі показники якості отриманих моделей:

1. Доля правильних відповідей (Accuracy)
2. Точність (Precision)
3. Чутливість (Sensitivity)
4. Специфічність (Specificity)
5. F1 Score

Показники 2-5 рекомендуються для використання при дисбалансі класів, що мало місце для даного дослідження (33,4 % пацієнтів мали достатній рівень вітаміну D, а 66,4 % — недостатній або дефіцит) [17].

У таблиці 3.9 наведені вищезгадані метрики для оцінки ризику дефіциту рівня 25(OH)D за допомогою моделей 1-6.

Таблиця 3.9

Метрики для оцінки ризику дефіциту рівня 25(OH)D моделей

Модель	Доля правильних відповідей (Accuracy)	Точність (Precision)	Чутливість	Специфічність	F1 Score
1	0.750	0.791	0.878	0.441	0.832
2	0.746	0.787	0.878	0.426	0.830
3	0.737	0.791	0.854	0.456	0.821
4	**0.746**	**0.787**	**0.878**	**0.426**	**0.830**
5	0.741	0.792	0.860	0.456	0.825
6	0.741	0.789	0.866	0.441	0.826

За даними таблиці 3.9 найбільш точною моделлю за всіма критеріями є модель 1, яка включає шість показників. Проте модель 4, яка включає лише чотири показники, має дуже близькі показники точності. Отже, використання моделі 4 із меншою кількістю лабораторних показників видається більш доцільним.

Крім того, для оцінки якості моделей застосовувався ROC-аналіз з використанням ROC-кривих (Receiver Operator Characteristic). ROC-крива показує залежність кількості правильно класифікованих позитивних випадків від кількості неправильно класифікованих негативних випадків, тобто цей графік описує взаємозв'язок між чутливістю моделі та її специфічністю [18].

На рисунку 3.2 наведено ROC-криві моделей 1-6.

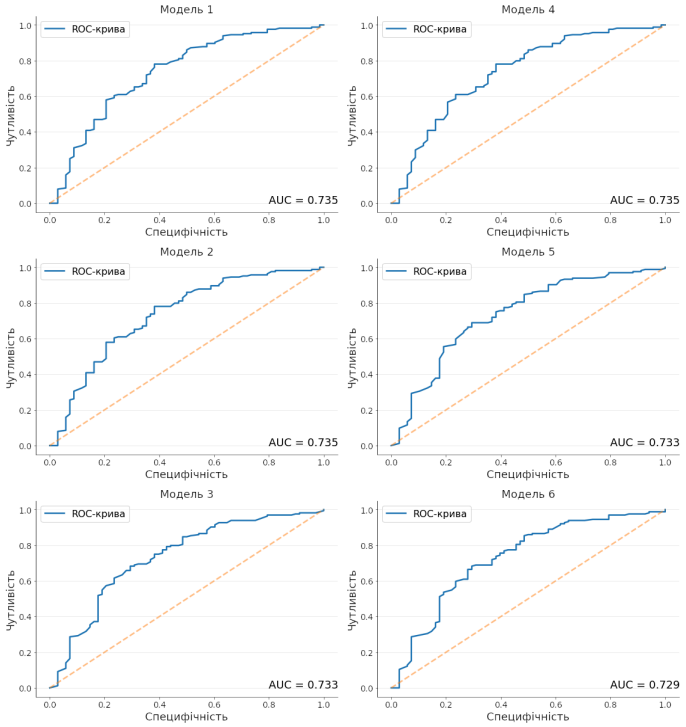

Рис. 3.2. ROC-криві логістичної регресії для моделей 1-6

Якість моделі класифікації визначається площею фігури під ROC-кривою, яка називається AUC (Area Under Curve). Чим вище значення AUC, тим вища прогностична цінність моделі. Значення AUC побудованих моделей знаходяться в інервалі від 0.729 до 0.735, отже, вони можуть вважатися достатньо ефективними. Зауважимо, що AUC моделі 4 співпадає з AUC моделі 1, тобто за цим показником вони мають однакову якість, що є черговим аргументом на користь використання моделі 4 на практиці.

3.3. Побудова класифікаційної моделі за допомогою алгоритмів конструювання дерев рішень для прогнозування дефіциту вітаміну D

Дерево рішень — це ієрархічна модель навчання під наглядом, за допомогою якої локальний регіон ідентифікується в послідовності рекурсивних поділів на меншу кількість кроків. Дерево рішень складається з внутрішніх вузлів рішень і кінцевих листів. Кожен вузол прийняття рішення m реалізує тестову функцію $f_m(x)$ з дискретними результатами, що позначають гілки. За наявності вхідних даних у кожному вузлі застосовується тест і одна з гілок береться залежно від результату. Цей процес починається в корені і повторюється рекурсивно, доки не буде досягнуто листкового вузла, після чого значення, записане в листі, становить вихід. Дерево рішень також є непараметричною моделлю в тому сенсі, що не припускається жодної параметричної форми для щільності класів, і структура дерева не є фіксованою апріорі, але дерево росте, гілки та листя додаються під час навчання залежно від складності проблеми в даних [23].

Одне з найскладніших завдань для алгоритмів побудови дерева рішень – вирішити, коли зупинити процес вирощування дерева. Більшість методів побудови дерева створюють дуже складні моделі, які переповнюють навчальні дані. Переобладнані дерева не тільки мають слабкі прогнозні можливості щодо нових раніше невидимих даних, але також можуть бути надзвичайно складними для інтерпретації, що є ключовим бар'єром для прийняття цих моделей у практичних застосуваннях. Загальним підходом для уникнення переобладнання в деревах рішень є рання зупинка, яка змушує алгоритм побудови зупинятися до того, як дерево стане надто складним. Популярні критерії зупинки включають обмеження максимальної глибини дерева, вимогу мінімальної кількості точок вибірки на листових вузлах або обчислення приросту точності, отриманого новими вузлами [24].

Гіперпараметр max_depth контролює загальну складність дерева рішень. Цей гіперпараметр дозволяє отримати компроміс між недостатньо підігнаним і

надто пристосованим деревом рішень [25]. У цьому дослідженні будівництво почалося з порівняння дерев із глибиною від 3 до 7, щоб зрозуміти вплив параметра та знайти найбільш оптимальне дерево.

У розділі 3.2 використовувалась логістична регресія для прогнозування дефіциту вітаміну D за допомогою демографічних, клінічних і лабораторних даних. Модель логістичної регресії використовувала бінарну систему класифікації для прогнозування дефіциту вітаміну D на основі таких факторів, як вік, стать, ІМТ та результати лабораторних досліджень. Хоча логістична регресія дала корисну інформацію про взаємозв'язок між змінними та ймовірністю дефіциту вітаміну D, вона була обмежена в своїй здатності фіксувати нелінійні кореляції та взаємодії між ними.

На противагу цьому, дерева рішень забезпечують гнучкий та інтуїтивно зрозумілий метод представлення складних взаємодій даних. Дерева рішень ділять область функцій на підмножини на основі основних правил прийняття рішень, що призводить до зрозумілих маршрутів прийняття рішень. Дерева рішень, рекурсивно розділяючи дані на основі найважливіших характеристик, можуть фіксувати складні межі рішень і зв'язки предикторів.

На рисунку 4.1 зображено візуалізацію оптимальної моделі дерева рішень для прогнозування дефіциту вітаміну D на основі різних демографічних, клінічних і лабораторних параметрів; з максимальною глибиною 5 дерево рішень може генерувати до 5 рівнів розбиття, кожен з яких відображає правило прийняття рішення на основі окремої функції.

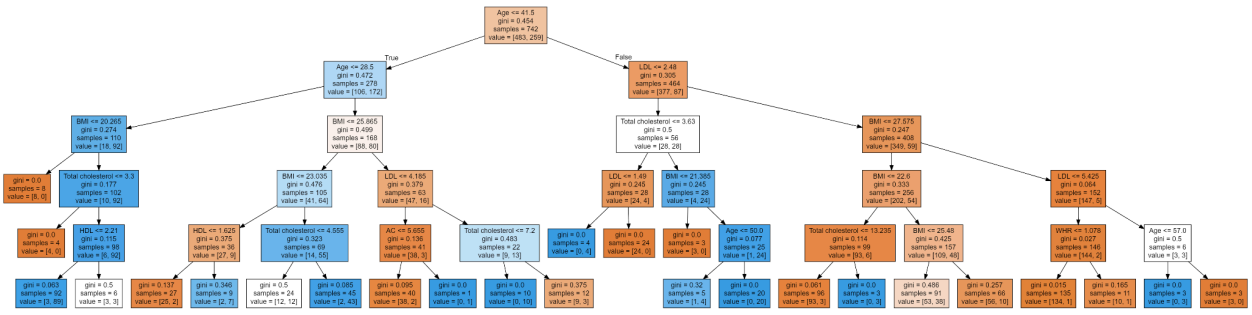

Рис. 4.1. Візуалізація моделі дерева рішень

На рисунку 4.2 показано результати аналізу методу дерева рішень щодо важливості ознак у виявленні потенційного дефіциту вітаміну D.

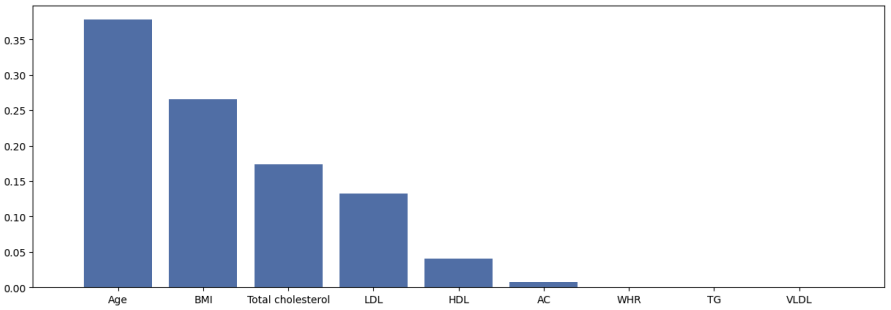

Рис. 4.2. Важливість ознак

Результати визначення важливості ознак показали, що найбільш впливовими вважаються такі антропометричні параметри, як вік та індекс маси тіла (ІМТ). Рівень ліпопротеїнів високої щільності (ЛПВЩ) є найважливішим лабораторним показником.

Якщо значення кореляції за абсолютною величиною перевищує 0.5, то це свідчить про наявність достатньо тісного взаємозв'язку між ознаками (за умови p-value < 0.05). Детальна матриця теплової карти для кореляції ознак між собою показана на рисунку 4.3.

Рис. 4.3. Важливість функцій з використанням теплової карти кореляційної матриці

Метрики для визначення ризику дефіциту 25(OH)D наведено у таблиці 4.1.

Таблиця 4.1

Оцінювальні метрики для оцінки ризику дефіциту 25(OH)D

Метрика оцінки	Значення
Доля правильних відповідей (Accuracy)	0.914
Точність (Precision)	0.901
Чутливість (Sensitivity/Recall)	0.859
Специфічність (Specificity)	0.946
F1 Score	0.879

Для оцінки якості моделей застосовувався ROC-аналіз з використанням ROC-кривих. На рисунку 4.4 показана ROC-крива оптимальної моделі.

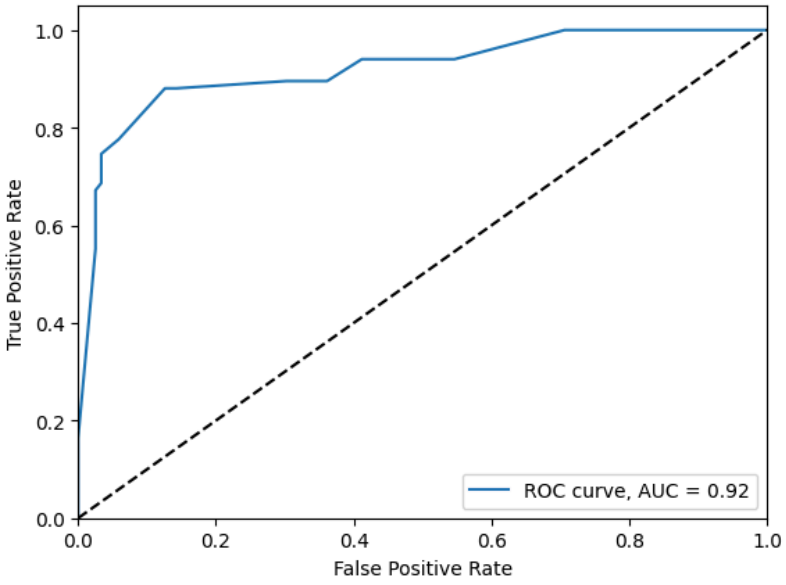

Рис. 4.4. ROC-крива оптимальної моделі

Таким чином, ми побудували моделі дерева рішень, логістичної регресії та лінійної регресії. Результати показали, що модель дерева рішень мала

найвищу точність у класифікаційних завданнях, але всі три моделі успішно ідентифікували важливі показники статусу вітаміну D. Це підкреслює потенційне використання методів машинного навчання в цілеспрямованих дієтичних втручаннях і персоналізованих оцінках ризиків для здоров’я, демонструючи їхню ефективність у прогнозуванні дефіциту вітаміну D.

ВИСНОВКИ

1. Враховуючи наявність статистично значимих зв’язків між 25(OH)D та ЛПВЩ, ЛПНЩ, ЛПДНЩ, КА та ТГ можна дійти висновку, що зниження рівня вітаміну D, може призводити до достовірного підвищення ризику порушення ліпідного обміну і як наслідок розвитку ішемічної хвороби серця та загальної смертності.
2. Було проведено статистичний аналіз ступеня залежності рівня вітаміна D від антропометричних показників, таких як індекс маси тіла (ІМТ) та співвідношення окружності талії до окружності стегон (ОТ/ОС). Дане дослідження продемонструвало більш високу негативну кореляцію між рівнем 25(OH)D та ІМТ, ніж індексом ОТ/ОС. Гіпотеза зв’язку вітаміна D із вказаними антропометричними показниками підтвердилася, зокрема, й при статистичному порівнянні груп за індексом маси тіла. Варто зазначити, що серед осіб обох статей незалежно від віку зі збільшенням ІМТ та окружності талії спостерігається зниження рівня 25(OH)D. Зниження вмісту вітаміну D до рівня, коли можна говорити про його дефіцит, найчастіше відзначалося серед пацієнтів з недостатньою вагою та у тих, хто мав ІМТ ≥ 30.
3. Проведено аналіз рівня впливу антропометричних та лабораторних показників на рівень вітаміну D. За результатами проведеного дослідження найбільш вагомими чинниками виявились ліпопротеїди високої щільності (ЛПВЩ), індекс маси тіла (ІМТ) та вік.
4. Побудовано математичні моделі залежності рівня 25(OH)D сироватки крові від певних показників (вік, ІМТ, ЛПВЩ, ЛПДНЩ, КА і ТГ). Найбільш високі показники адекватності мала модель 4, до якої були включені вік, індекс маси тіла (ІМТ), коефіцієнт атерогенності (КА), та ліпопротеїди високої щільності (ЛПВЩ). Додавання інших досліджуваних чинників не сприяло покращенню якості обраної моделі. З іншого боку, показники адекватності отриманих моделей були на рівні

0.19-0.224, що свідчить про недостатньо високу якість прогнозу за цими моделями.

5. Проведено кореляційний аналіз та математичне моделювання залежності ризику дефіциту вітаміну D від антропометричних та лабораторних показників. Найбільш впливовими чинниками виявилися вік, індекс маси тіла, коефіцієнт атерогенності, ліпопротеїди високої щільності та співвідношення окружності талії до окружності стегон.
6. За допомогою моделей множинної логістичної регресії було оцінено ризик вітамін D дефіцитних станів у мешканців південних регіонів України. Оцінка якості моделей проводилася за п'ятьма показниками на основі матриці помилок, а також ROC-аналізу. Найбільш оптимальним за даними проведеного дослідження є використання моделі 4, яка включає вік, ІМТ, КА та ЛПВЩ.
7. Методи машинного навчання більш ефективні для прогнозування недоліків, ніж традиційні статистичні методи. Модель дерева рішень виявилася значно кращою, ніж логістична регресія з попереднього дослідження, з точки зору точності прогнозування. Модель дерева рішень мала AUC-ROC 92%, що представляло 17% підвищення точності в порівнянні з результатами логістичної регресії.
8. Використання моделей машинного навчання в клінічній практиці демонструє потенціал для покращення результатів лікування пацієнтів і зниження тягаря розладів, пов’язаних із вітаміном D. Правильно визначивши пацієнтів із ризиком дефіциту, постачальники медичних послуг можуть застосовувати спеціальні методи лікування, такі як добавки, зміна дієти та способу життя, щоб покращити статус вітаміну D і запобігти наслідкам для здоров’я.

Таким чином, математичне моделювання дозволяє оцінити рівень впливу відібраних антропометричних та лабораторних показників на рівень 25(OH)D сироватки крові та спрогнозувати ризик вітамін D дефіцитних станів. Можна зробити висновок, що, незважаючи на те, що лабораторний контроль 25(OH)D є

найбільш інформативним показником забезпеченості вітаміну D в організмі, використання математичної моделі дає змогу виявити осіб з групи ризику для подальшої діагностики та зменшити кількість лабораторних досліджень серед людей, які за даною оцінкою мають достатній рівень вітаміну D. Зрештою, дослідження підкреслює потенціал машинного навчання для прогнозування недостатності вітаміну D і розробки індивідуального лікування. Використовуючи потужність методів, керованих даними, медичні працівники можуть отримати інструменти та знання, необхідні для покращення здоров'я та благополуччя пацієнтів.

За результатами проведеного дослідження була опублікована стаття [19] та підготовлено до публікації нова стаття "Features of the use of artificial intelligence for the purpose of screening vitamin D deficiency in the adult population". Також результати роботи були апробовані на науково-практичних конференціях "Інформаційні технології і автоматизація" (Одеса, ОНТУ, 2023) [21], "Стан, досягнення та перспективи інформаційних систем і технологій" (Одеса, ОНТУ, 2024) [22] та опубліковані у вигляді тез. Результати розділів 2 та 3 доповідалися на науково-практичній конференції з міжнародною участю "Сучасні теоретичні та практичні аспекти клінічної медицини" (Одеса, ОНМедУ, 2023) [20]

У майбутньому можливо розробити веб-додаток, який дозволить користувачам вводити подробиці про свій раціон і спосіб життя для отримання індивідуальних оцінок стану вітаміну D. За допомогою даних, наданих користувачами, ця програма може використовувати складні математичні моделі, включаючи дерева рішень і лінійну регресію, щоб забезпечити точні прогнози рівня вітаміну D у сироватці крові. Програма також може надавати користувачам персоналізовані пропозиції щодо їжі та способу життя, щоб допомогти їм досягти або підтримувати ідеальний рівень вітаміну D.

Розширення вибору моделей для включення нових даних і складних методів машинного навчання, таких як ансамблі моделей (розширення, випадкові ліси), може значно підвищити прогнозну точність і стійкість оцінок.

Веб-додаток міг би надати більш точні прогнози та врахувати більший спектр факторів і взаємодій, які впливають на рівень вітаміну D, шляхом об’єднання різних моделей.

Подібні моделі можна використовувати для прогнозування ймовірності пов’язаних проблем зі здоров’ям, таких як ризик переломів, на додаток до недостатності вітаміну D. Це дозволить користувачам вживати профілактичних заходів проти переломів та інших проблем, пов’язаних із низьким рівнем вітаміну D.

Крім того, використовуючи ці моделі прогнозування, програма може представити повну картину здоров’я кісток людини, надаючи інформацію про ризик переломів і статус вітаміну D. Як пацієнти, так і медичні працівники можуть виявити цю можливість подвійного прогнозування дуже корисною для раннього втручання та індивідуальних програм лікування, щоб зменшити небезпеки, пов’язані з недостатністю вітаміну D і проблемами зі здоров’ям кісток.

СПИСОК ЛІТЕРАТУРИ

1. Cimmino G, Morello A, Conte S, et al. Vitamin D inhibits Tissue Factor and CAMs expression in oxidized low-density lipoproteins-treated human endothelial cells by modulating NF-κB pathway. Eur J Pharmacol. 2020;885:173422. doi: https://doi.org/10.1016/j.ejphar.2020.173422
2. Pan′kiv V.I., Povoroznyuk V.V., Pan′kiv I.V., Boyko V.I., Hluhovs′ka S.V. Stan zabezpechennya vitaminom D naselennya Zakhidnoho rehionu Ukrayiny. International Journal of Endocrinology. 2019;15(3):268-271. doi: https://doi.org/10.22141/2224-0721.15.3.2019.172115
3. Park C.Y., Han S.N. The Role of Vitamin D in Adipose Tissue Biology: Adipocyte Differentiation, Energy Metabolism, and Inflammation. J Lipid Atheroscler. 2021;10(2):130-144. doi: https://doi.org/10.12997/jla.2021.10.2.130
4. Shanyhin A.V. Znachennia ratsionu kharchuvannia ta rivnia insoliatsii v zabezpechenosti vitaminom D. Suchasni aspekty profilaktyky. Health of Society. 2022;11(1):16–22. doi: https://doi.org/10.22141/2306-2436.11.1.2022.288
5. Прутіян Т. Л. Оцінка якості життя в жінок у постменопаузі з артеріальною гіпертензією, ожирінням та остеопорозом / Т. Л. Прутіян, О. О. Добровольська // Сучасні аспекти модернізації науки: стан, проблеми, тенденції: XXIII Міжнар. наук.-практ. конф., 07 серпня 2022, м. Дікірх (Люксембург), дистанційно) : матер. – Дікірх, 2022. – С. 105–110.
6. Setayesh L, Amini A, Bagheri R, Moradi N, Yarizadeh H, Asbaghi O, Casazza K, Yekaninejad MS, Wong A, Suzuki K, Mirzaei K. Elevated Plasma Concentrations of Vitamin D-Binding Protein Are Associated with Lower High-Density Lipoprotein and Higher Fat Mass Index in Overweight and Obese Women. *Nutrients*. 2021; 13(9):3223. doi: https://doi.org/10.3390/nu13093223

7. Holick MF. The vitamin D deficiency pandemic: Approaches for diagnosis, treatment and prevention. Rev Endocr Metab Disord. 2017;18(2):153-165. doi: https://doi.org/10.1007/s11154-017-9424-1
8. Sizar, O.; Khare, S.; Goyal, A.; Bansal, P.; Givler, A. Vitamin D Deficiency; Updated 21 July 2021; StatPearls Publishing: Treasure Island, FL, USA, 2021. Available online: https://www.ncbi.nlm.nih.gov/books/NBK532266/ (accessed on 19 February 2023).
9. NumPy — 2024 — https://numpy.org/
10. SciPy — 2024 — https://scipy.org/
11. Pandas — 2024 — https://pandas.pydata.org/
12. Matplotlib — 2024 — https://matplotlib.org/
13. Яровий, А.Т. Економетрія : навч.-метод. посіб. для студ. мат. та екон. спец. / А. Т. Яровий, Є. М. Страхов. – Одеса : Освіта України, 2017. – 129 с.
14. Holick, M.F., Binkley, N.C., Bischoff-Ferrari, H.A., Gordon, C.M., Hanley, D.A, Heaney, R.P., Murad, M.H., & Weaver, C.M. (2011). Evaluation, treatment, and prevention of vitamin D deficiency. An Endocrine Society Clinical Practice Guideline. J. Clin. Endocrinol. Metab., 96 (7), 1911-1930.
15. Lisa Sullivan, PhD — Confidence Intervals — Boston University — https://sphweb.bumc.bu.edu/otlt/mph-modules/bs/bs704_confidence_intervals/bs704_confidence_intervals_print.html
16. Ting, K.M. (2011). Confusion Matrix. In: Sammut, C., Webb, G.I. (eds) Encyclopedia of Machine Learning. Springer, Boston, MA. https://doi.org/10.1007/978-0-387-30164-8_157
17. Sarang Narkhede — Understanding Confusion Matrix — 2018 — https://towardsdatascience.com/understanding-confusion-matrix-a9ad42dcfd62
18. Hajian-Tilaki K. Receiver Operating Characteristic (ROC) Curve Analysis for Medical Diagnostic Test Evaluation. Caspian J Intern Med. 2013 Spring;4(2):627-35. PMID: 24009950; PMCID: PMC3755824.

19. Shanyhin, Anton, Babienko, Volodymyr, Strakhov, Yevhen, Korkhova, Arina. Mathematical modeling of the dependence of the risk of vitamin D deficiency on anthropometric and laboratory parameters. *Journal of Education, Health and Sport* [online]. 28 April 2023, T. 13, nr 4, s. 356–366.
20. Шанигін А.В., Страхов Є.М., Корхова А.С., Коломійченко Ю.В. Система оцінки ризиків розвитку вітамін-D-дефіцитних станів. Сучасні теоретичні та практичні аспекти клінічної медицини для здобувачів освіти другого (магістерського) рівня : наук.-практ. конф. з міжнар. участю, присвячена 95-річчю з дня народження Л. В. Прокопової. Одеса, 27–28 квітня 2023 року. Одеса: ОНМедУ, 2023.
21. Корхова А. С., Страхов Є. М. Моделювання оцінки ризику виникнення станів дефіциту вітаміну D. // Інформаційні технології і автоматизація – 2023. Матеріали XVI міжнародної науково-практичної конференції. Одеса, 19-20 жовтня 2023 р. Одеса, Видавництво ОНТУ, 2023 р. С. 428-429.
22. Strakhov Ye., Korkhova A. Features of using artificial intelligence for screening vitamin D deficiency in adults // Стан, досягнення та перспективи інформаційних систем і технологій — 2024. Матеріали XXIV Всеукраїнської науково-технічної конференції молодих вчених, аспірантів та студентів. Одеса, 18-19 квітня 2024 р. Одеса, Видавництво ОНТУ, 2024 р. С. 474-475.
23. Alpaydin, Ethem. Introduction to machine learning / Ethem Alpaydin—3rd ed., 2014, s. 213-214
24. García Leiva, R., Fernández Anta, A., Mancuso, V., & Casari, P. (2019). A Novel Hyperparameter-Free Approach to Decision Tree Construction That Avoids Overfitting by Design. IEEE Access, 7, 99978-99987.
25. Loïc Estève, Guillaume Lemaitre, Olivier Grisel, Gael Varoquaux, Arturo Amor, Lilian, и др. INRIA/scikit-learn-mooc: Third MOOC session. Zenodo; 2022.
26. Scikit-learn — 2024 — https://scikit-learn.org/stable/

27. Seaborn — 2024 — https://seaborn.pydata.org/
28. Korkhova A.S., Strakhov Ye.M. — 2024 — Resource access mode: https://github.com/arina-korkhova/vitaminD