

DOI: <https://doi.org/10.36910/6775-2524-0560-2023-53-46>

УДК 004:94:53:616-073

Капаціла Роман Ігорович¹, аспірант

<https://orcid.org/0009-0000-8618-1153>

Сверстюк Андрій Степанович², д.т.н., професор

<https://orcid.org/0000-0001-8644-0776>

¹Тернопільський національний технічний університет імені І. Пулюя, м. Тернопіль, Україна

²Тернопільський національний медичний університет імені І. Я. Горбачевського, м. Тернопіль, Україна

ЗАСТОСУВАННЯ ДЕРЕВ ПРИЙНЯТТЯ РІШЕНЬ ДЛЯ ПРОГНОЗУВАННЯ НАСТАННЯ СЕРЦЕВИХ ЗАХВОРЮВАНЬ

Капаціла Р.І., Сверстюк А.С. Застосування дерев прийняття рішень для прогнозування настання серцевих захворювань. У роботі запропоновано використання дерева прийняття рішень для прогнозування виникнення серцево-судинних захворювань. У якості вхідних даних для побудови дерева прийняття рішень було використано набір даних з 12-ти основних факторів, а саме: вік, стать, характер болю в грудях, кров'яний тиск в стані спокою, рівень холестерину в крові, рівень цукру в крові, тип результату електрокардіограми у стані спокою, максимальний серцевий ритм зафіксований при огляді, наявність стенокардії при фізичному навантаженні, депресія ST сегменту спричинена фізичними вправами порівняно з відпочинком, тип нахилу піку навантаження ST сегменту та факт виявлення хвороби серця в результаті діагностики. На основі вхідного набору даних побудована кореляційна матриця досліджуваних факторів. У вигляді таблиці наведено досліджувані фактори, англійські відповідники, а також діапазон їх допустимих значень. Наведено основні кроки алгоритму побудови дерева прийняття рішень. Отримана текстова інтерпретація дерева рішень із умовами переходу для вершин. Для оцінювання якості класифікації вхідних даних на основі дерева прийняття рішень наведено результат класифікації у вигляді матриці невідповідності. Проведена оцінка прецизійності та точності моделі. У якості співвідношення між чутливістю та специфічністю проведено ROC аналіз, за результатами якого встановлено площу під ROC кривою AUC = 0.9. Отримані значення AUC відповідають відмінній якості класифікації, що є передумовою для використання запропонованого дерева прийняття рішень для проектування інформаційно-діагностичної системи прогнозування ризику виникнення серцево-судинних захворювань серця.

Ключові слова: комп'ютерне моделювання, прогнозування, дерева прийняття рішень, діагностика серцевих захворювань, багатофакторний аналіз

Kapatsila R., Sverstiuk A. Application of decision trees for prediction of heart disease. The study proposes the utilization of a decision tree for predicting the onset of cardiovascular diseases. A dataset consisting of 12 key factors was used as input to build a decision tree. These factors include age, gender, chest pain nature, resting blood pressure, cholesterol level, blood sugar level, type of resting electrocardiogram result, maximum heart rate observed during examination, presence of angina during physical exertion, ST segment depression induced by exercise compared to rest, slope type of the ST segment at peak exercise, and the confirmation of heart disease through diagnostics. A correlation matrix of the investigated factors was constructed based on the input dataset. The research presents a table outlining the studied factors, their English counterparts, and the acceptable ranges of their values. The fundamental steps of the decision tree construction algorithm are outlined. The test interpretation of the decision tree with transition conditions for nodes is provided. To assess the classification quality of the input data using the decision tree, the classification results are presented in the form of a confusion matrix. Precision and accuracy of the model are evaluated. A Receiver Operating Characteristic (ROC) analysis is conducted to examine the relationship between sensitivity and specificity, resulting in an Area Under the Curve (AUC) value of 0.9. The obtained AUC values correspond to excellent classification quality, serving as a prerequisite for employing the proposed decision tree in the design of an information-diagnostic system for predicting the risk of cardiovascular diseases.

Keywords: computer modeling, forecasting, decision trees, diagnosis of heart diseases, multivariate analysis

Постановка проблеми.

Протягом останніх десятиліть спостерігаємо збільшення кількості серцево-судинних захворювань та поширеність серед різних вікових груп, що спричиняє серйозні виклики для збереження здоров'я населення у всьому світі. Для ефективного прогнозування настання серцевих захворювань необхідно враховувати множини впливових факторів, таких як вік, стать, анамнез, показники крові та інші. Одним із підходів до розв'язання цієї проблеми є застосування методів машинного навчання, зокрема, дерев прийняття рішень.

Проте, при використанні дерев прийняття рішень для прогнозування серцевих захворювань виникають певні проблемами. По-перше, дерева прийняття рішень можуть виявитися недостатньо ефективними при врахуванні слабких та складних взаємозв'язків між різними факторами, що впливають на ризик серцевих захворювань. По-друге, вони можуть бути схильні до перенавчання, що призводить до втрати узагальнюючої здатності моделі на нових даних. По-третє, важливо

вирішити питання вибору оптимальних гіперпараметрів для дерев, таких як глибина, щоб уникнути перетренування чи недооцінювання важливих зв'язків.

Отже, проблема полягає в розробці та оптимізації багатофакторних моделей прогнозування серцевих захворювань, які використовують дерева прийняття рішень, з урахуванням їхньої точності, ефективності та здатності узагальнення до нових даних.

Постановка завдання.

Мета даного дослідження полягає в розробці багатофакторної моделі прогнозування настання серцевих захворювань на основі ряду клінічних факторів, з використанням дерев прийняття рішень як інструменту машинного навчання.

Завдання включають:

1. Збір та обробка даних, а саме проведення аналізу даних та визначення їхню якість та придатність для моделювання.
2. Розробити багатофакторну модель, визначивши ключові фактори, що впливають на ризик серцевих захворювань.
3. Використовувати дерева прийняття рішень для побудови моделі та врахування взаємозв'язків між факторами.
4. Здійснити валідацію моделі для перевірки її точності та узагальнення на нових даних.
5. Використати метрики ефективності для оцінки точності та робастності моделі.

Побудова та оцінка моделі.

Метод багатофакторного аналізу прогнозування настання серцевих захворювань з урахуванням найбільш поширених та широкодоступних медичних показників дозволяє створити математичну модель для оцінки та прогнозування можливого настання серцевих захворювань. Застосування даної моделі дає змогу використовувати обмежений масив інформації для попередження потенційних ризиків, які пов'язані із настанням серцево-судинних захворювань.

Оригінальна база даних включає в себе 76 атрибутів і ґрунтується на результати неінвазійних клінічних досліджень, проведених у ряді медичних установ, а саме: Клініка Клівленда, штат Огайо, США; Інститут Кардіології в Будапешті, Угорщина; Медичний центр адміністрації ветеранів у Лонг-Біч, Каліфорнія, США; та університетські лікарні Цюріха та Базеля, Швейцарія. Після комплексного аналізу всіх вхідних даних та виключення неконсистентних записів був сформований набір з 12 факторів. Зокрема, база даних з Клініки Клівленда вважається найбільш повною та є основою для даного дослідження [1]. Імена та номери соціального страхування пацієнтів з етичних міркувань були виключені з бази даних та замінені фіктивними значеннями. Для подальшого аналізу, побудови моделі та її оцінки використано агрегований набір даних з 12 основними факторами [1], який містить 746 унікальних записів, в яких середній вік пацієнтів становив 53 роки. У вхідній вибірці дані щодо статевого розподілу вказують на 79% чоловіків та 21% жінок. Приклад отримання діагностичної інформації з електрокардіосигналів наведено в роботі [2]

Наступним кроком проведено побудову матриці кореляції [3] для оцінювання можливості використання всіх факторів для подальшої побудови моделі. Як видно із матриці кореляції (рис.1) прямих залежностей між факторами аналізу не було виявлено.

Для побудови моделей прогнозування настання серцевих захворювань відібрано наступні фактори: вік, стать, характер болю в грудях, кров'яний тиск в стані спокою, рівень холестерину в крові, рівень цукру в крові, тип результату електрокардіограми у стані спокою, максимальний серцевий ритм зафіксований при огляді, наявність стенокардії при фізичному навантаженні, депресія ST сегменту спричинена фізичними вправами порівняно з відпочинком, тип нахилу піку навантаження ST сегменту та факт виявлення хвороби серця в результаті діагностики [4].

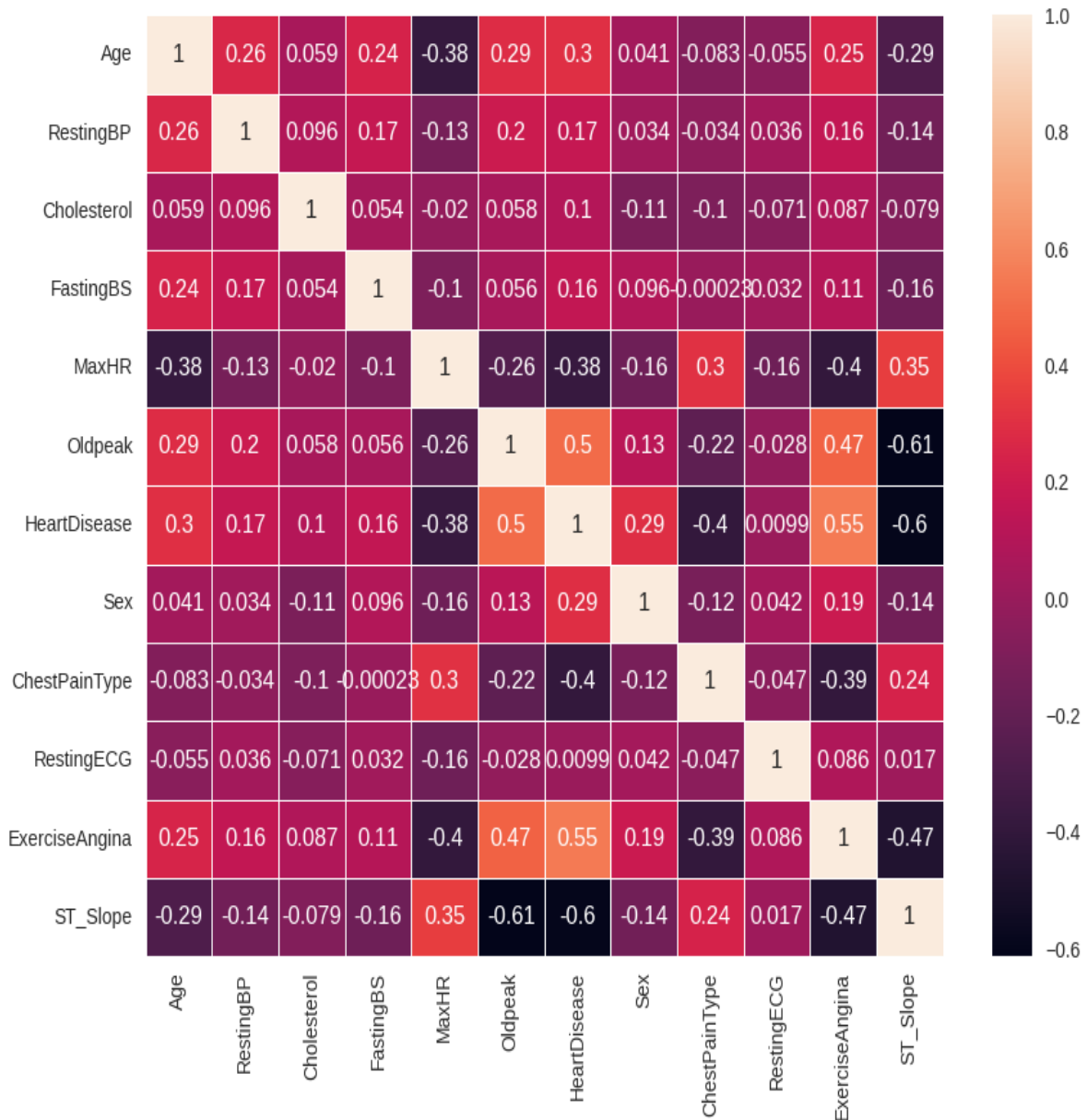


Рис. 1. Матриця кореляції досліджуваних факторів

Базуючись на методи оцінки адекватності факторів Кайзера-Мейера-Олкіна [5] проводимо аналіз обраного набору даних в результаті якого отримуємо значення 0.80 чого достатньо для подальшої побудови моделі.

Таблиця 1. Досліджувані фактори ризику виникнення серцево-судинних захворювань серця.

Назва факторів	Англomовні позначення досліджуваних факторів у вхідному наборі даних	Діапазон допустимих значень у вхідному наборі даних
Вік	Age	0 - 99
Кров'яний тиск в стані спокою	RestingBP	0 - 200
Рівень холестерину	Cholesterol	0 - 603

Рівень цукру	FastingBS	0 - 1
Максимальний серцевий ритм зафіксований при огляді	MaxHR	60 - 202
Депресія ST сегменту спричинена фізичними вправами порівняно з відпочинком	Oldpeak	0 – 6.2
Стать	Sex	0, 1
Тип болю в грудях	ChestPainType	0, 1, 2, 3
Тип результату електрокардіограми у стані спокою	RestingECG	0, 1, 2
Наявність стенокардії при фізичному навантаженні	ExerciseAngina	1, 0
Тип нахилу піку навантаження ST сегменту	ST_Slope	0, 1, 2

Серед 12 проаналізованих початкових факторів та побудованої матриці кореляції відібрано 11 факторів ризику [6] для передбачення останнього фактору який індикує наявність серцевих захворювань. Для побудови моделі буде використано фактори, що найбільше впливають на появу серцевих захворювань: вік, кров'яний тиск в стані спокою, рівень холестерину в крові, рівень цукру в крові, максимальний серцевий ритм зафіксований при огляді, депресія ST сегменту спричинена фізичними вправами порівняно з відпочинком, стать, тип болю в грудях, тип результату електрокардіограми у стані спокою, наявність стенокардії при фізичному навантаженні, тип нахилу піку навантаження ST сегменту. Детальна інформація та допустимі значення проаналізованих факторів ризику розвитку серцевих захворювань, що були використані для побудови моделі, наведено в Таблиці 1.

На основі отриманих результатів багатофакторного аналізу прогнозування ризику виникнення серцевих захворювань отримано дерево прийняття рішень для прогнозування ризику виникнення серцевих захворювань [7], яке наведено на рис. 2.

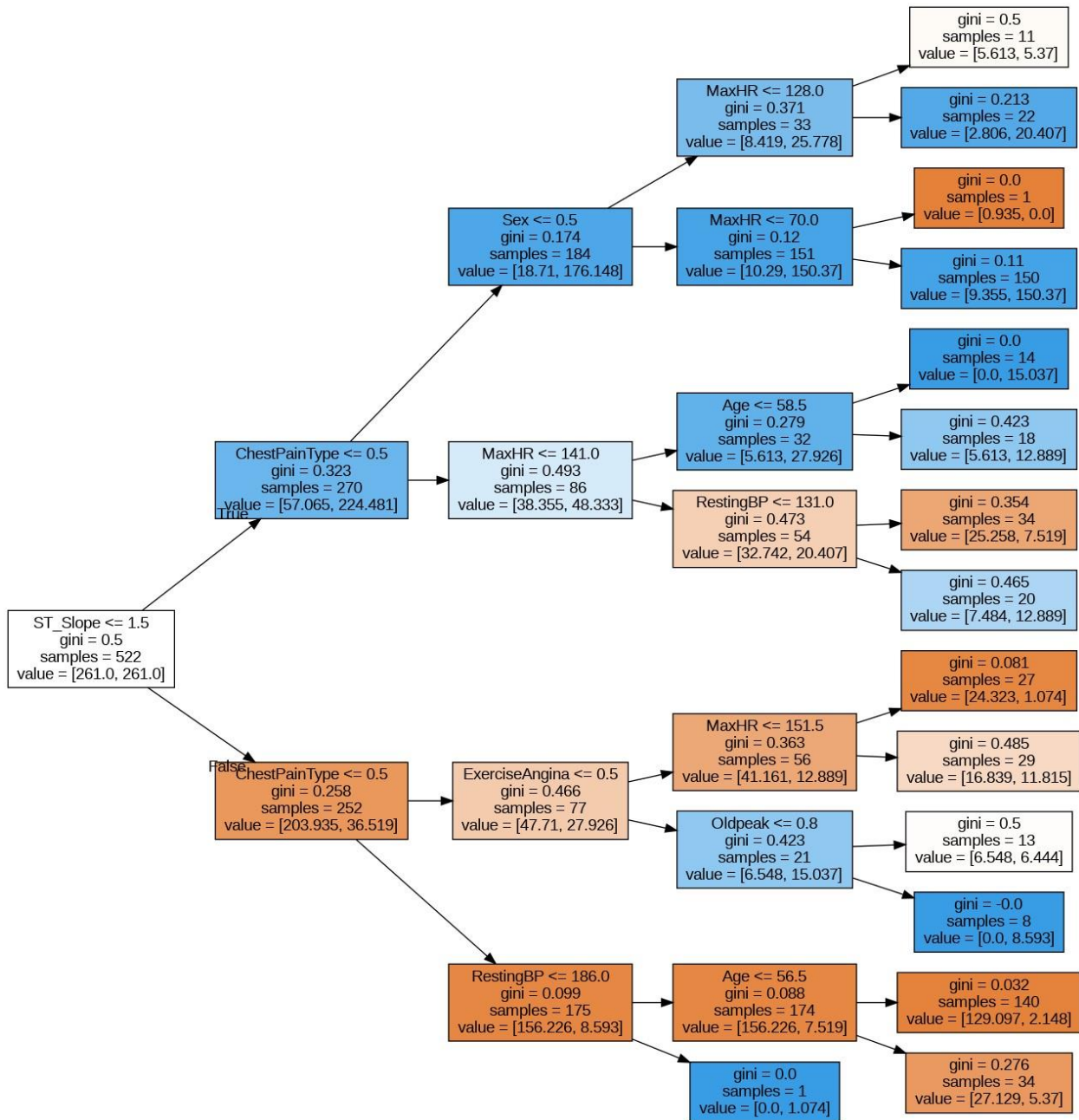


Рис.2. Дерево прийняття рішень ризику виникнення серцево-судинних захворювань серця

Для побудови дерева прийняття рішень використано алгоритм CART (Classification and Regression Trees) [8]. Більш детальна репрезентація дерева прийняття рішень наведена на рис. 3 із приведенням умов переходів для кожної із вершин.

```

|--- ST_Slope <= 1.50
| |--- Sex <= 0.50
| | |--- Oldpeak <= 0.50
| | | |--- FastingBS <= 128.00
| | | | |--- weights: [5.61, 5.37] class: 0.0
| | | | |--- FastingBS > 128.00
| | | | |--- weights: [2.81, 20.41] class: 1.0
| | | |--- Oldpeak > 0.50
| | | |--- FastingBS <= 70.00
| | | |--- weights: [0.94, 0.00] class: 0.0
    
```

```

| | | |--- FastingBS > 70.00
| | | | |--- weights: [9.35, 150.37] class: 1.0
| | | |--- Sex > 0.50
| | | |--- FastingBS <= 141.00
| | | |--- Age <= 58.50
| | | | |--- weights: [0.00, 15.04] class: 1.0
| | | |--- Age > 58.50
| | | | |--- weights: [5.61, 12.89] class: 1.0
| | | |--- FastingBS > 141.00
| | | |--- RestingBP <= 131.00
| | | | |--- weights: [25.26, 7.52] class: 0.0
| | | |--- RestingBP > 131.00
| | | | |--- weights: [7.48, 12.89] class: 1.0
|--- ST_Slope > 1.50
| |--- Sex <= 0.50
| | |--- ExerciseAngina <= 0.50
| | | |--- FastingBS <= 151.50
| | | | |--- weights: [24.32, 1.07] class: 0.0
| | | |--- FastingBS > 151.50
| | | | |--- weights: [16.84, 11.81] class: 0.0
| | | |--- ExerciseAngina > 0.50
| | | |--- MaxHR <= 0.80
| | | | |--- weights: [6.55, 6.44] class: 0.0
| | | |--- MaxHR > 0.80
| | | | |--- weights: [0.00, 8.59] class: 1.0
| |--- Sex > 0.50
| | |--- RestingBP <= 186.00
| | | |--- Age <= 56.50
| | | | |--- weights: [129.10, 2.15] class: 0.0
| | | |--- Age > 56.50
| | | | |--- weights: [27.13, 5.37] class: 0.0
| | |--- RestingBP > 186.00
| | | |--- weights: [0.00, 1.07] class: 1.0

```

Рис.3 – Текстова інтерпретація дерева рішень із умовами переходу між вершинами

Алгоритм CART – це алгоритм для побудови дерев прийняття рішень, який може використовуватися як для задач класифікації, так і для регресії. CART був розроблений Лео Брейманом, Джеромом Фрідманом, Чарльзом Дж. Стоуном та Ріхардом Олшеном [9, 10].

Даний алгоритм можна розбити на наступні кроки:

1. Вибір атрибута для розділення (Splitting Attribute):

○ Визначається критерій вибору, такий як коефіцієнт Gini для класифікації або середньоквадратична помилка для регресії.

○ Вибирається атрибут і поріг, що максимізують чистоту вузлів або мінімізують помилковість.

2. Розділення вузла:

○ Дані розділяються на дві частини відповідно до вибраного атрибута та порогу.

○ Формуються два нових вузли дерева.

3. Рекурсивний процес:

○ Процес рекурсивно повторюється для кожного нового вузла, розглядаючи його як точку вибору нового атрибута для поділу.

4. Зупинка рекурсії (Stopping Criteria):

○ Рекурсія зупиняється, коли досягнуті умови зупинки, такі як максимальна глибина дерева, мінімальна кількість прикладів у вузлі або інші визначені умови.

5. Прогноз для листа:

○ Кожен листовий вузол призначається певному значенню класу.

Далі детальніше розглянемо сам процес побудови дерева у рамках даного дослідження. Для побудови дерева рішенням критерієм вибору буде обчислення індексу Gini [11]. Індекс Gini у нашому контексті можна обрахувати наступним чином:

$$Gini(T) = 1 - \sum_{j=1}^n p_j^2, \quad (1)$$

де T – поточний вузол, p_j – імовірність класу j у вузлі T , n – кількість класів.

Вибір атрибута та порогу для подальшого розділення визначається шляхом максимізації/мінімізації вагових значень індексу Gini для двох нових вузлів, що виникають в результаті розділу:

$$Gini_split = \frac{N_L}{N} Gini(T_L) + \frac{N_R}{N} Gini(T_R), \quad (2)$$

де N_L і N_R – кількість прикладів у лівому та правому піддереві відповідно, N – загальна кількість прикладів у вузлі, $Gini(T_L)$ і $Gini(T_R)$ – індекс Gini для лівого та правого піддереву відповідно.

Цей процес рекурсивно повторюється, і дерево формується шляхом оптимізації індексу Gini на кожному кроці. Такий підхід дозволяє побудувати дерево, яке робить оптимальні розділи для класифікації.

На кожному етапі побудови дерева вибирається оптимальний роздільний атрибут, який максимізує індекс Gini або мінімізує помилковість. Для цього можна використовувати методи оптимізації, такі як пошук по всіх можливим роздільних атрибутів та їх порогів. За допомогою рекурсивного процесу повторюємо процес вибору роздільного атрибута та його порогу для кожного нового вузла, аж до виконання умови зупинки.

Рекурсія зупиняється при досягненні умови зупинки, такої як максимальна глибина дерева, яка у даному випадку рівна 4.

Мета алгоритму CART – мінімізувати цей індекс Gini під час кожного розділу для отримання більш точного та чистого дерева класифікації.

Після побудови моделі на основі дерева прийняття рішень наступним кроком буде аналіз точності та чутливості моделі [12]. Чутливість вимірює частку фактичних позитивних випадків, які модель правильно ідентифікувала. Цей показник має вирішальне значення в охороні здоров'я, де нездатність визначити стан (наприклад, рак або хворобу серця) може мати серйозні або навіть летальні наслідки. Точність вимірює частку позитивних ідентифікацій, які були насправді правильними. У сфері охорони здоров'я висока точність означає, що поставлений діагноз є більш надійним. Нижча точність може призвести до хибно позитивних результатів, що може призвести до непотрібного лікування чи тестів, викликаючи стрес і додаткові витрати на лікування.

З метою оцінки якості моделі необхідно отримати точність класифікації моделі. Засобами Python [13] обраховуємо точність та отримуємо значення в 0.85, що відповідає 85% точності прогнозування на основі вхідних параметрів. Матриця невідповідності наведена на рис. 4.

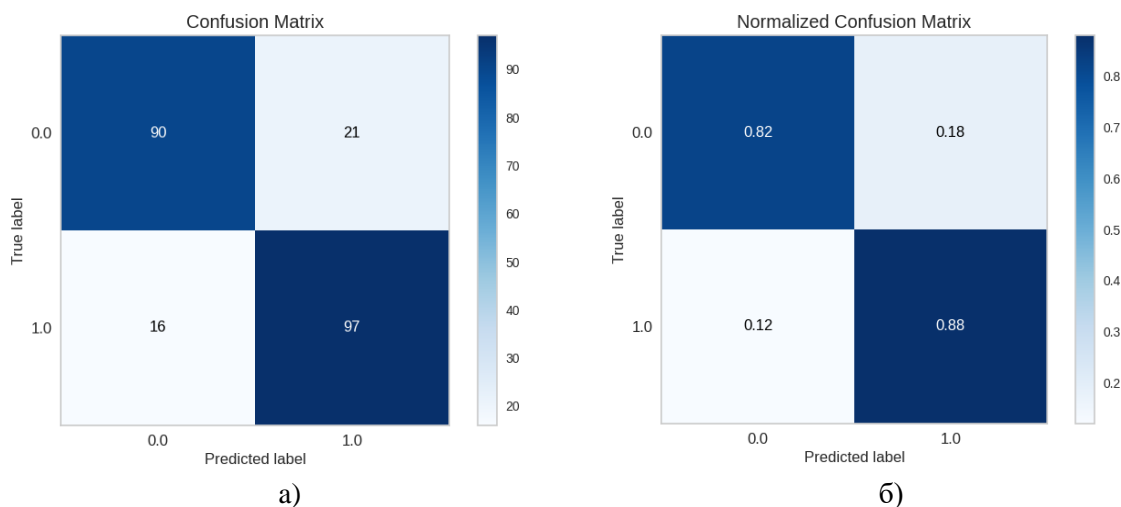


Рис.4. а) Матриця невідповідності; б) – Нормалізована матриця невідповідності

Для перевірки нашої моделі можна використати F-міру [14]: це показник, який поєднує прецизійність і чутливість в одне значення, що дозволяє легше оцінити продуктивність моделі. Це гармонійне середнє значення чутливості та прецизійності, спрямоване на збалансування двох показників. F-міра коливається від 0 до 1, де 1 вказує на ідеальну прецизійність і чутливість, а 0 означає відсутність чутливості та прецизійності. F-міра особливо корисна в ситуаціях, коли розподіл класів є незбалансованим, або коли кількість хибно позитивних та хибно негативних результати значно відрізняються, що часто буває в охороні здоров'я. Наприклад, відсутність діагнозу (хибно негативний результат) може мати серйозніші наслідки, ніж неправильне встановлення діагнозу здоровому пацієнту (хибно позитивний результат).

На рис. 5 приведено значення для прецизійності, чутливості та F-міри. Як видно із результату середнім значенням є 0.85, що означає що у 85% випадків ми отримаємо коректний результат при аналізі вхідних даних.

	precision	recall	f1-score	support
Normal (0)	0.87	0.82	0.84	111
Heart Disease (1)	0.83	0.88	0.85	113
accuracy			0.85	224
macro avg	0.85	0.85	0.85	224
weighted avg	0.85	0.85	0.85	224

Рис.5. Оцінка прецизійності та чутливості моделі

Ще одна вкрай важлива метрика для оцінки точності моделі є специфічність, яка є дуже важливим показником завдань класифікації у сфері охорони здоров'я [15], особливо коли потрібно мінімізувати кількість помилкових спрацювань за результатами математичного моделювання кардіосигналів [16-18]. У медичному світі це може означати неправильне діагностування пацієнта як хворого, хоча він насправді здоровий, що призведе до непотрібного стресу, додаткових тестів і, можливо, навіть лікування. Специфічність вимірює частку фактично негативних (без захворювання), які правильно визначені як такі. Іншими словами, це спосіб виміряти, наскільки добре модель правильно ідентифікує тих, хто не має захворювання. Висока специфічність означає, що тест добре допомагає уникнути хибно позитивних результатів, що робить його цінним у медичному закладі, де ціна помилкового діагнозу може бути високою. Це особливо важливо для скринінгових тестів, гарантуючи, що лише ті, хто ймовірно має захворювання, будуть направлені на більш детальне обстеження. Для обрахунку специфічності можна використати наступну формулу:

$$\text{Специфічність} = \frac{\text{True Negative (TN)}}{\text{True Negative (TN)} + \text{False Positive (FP)}} \quad (3)$$

де True Negative (TN) – кількість випадків, коли модель правильно визначила відсутність класу; False Positive (FP) – кількість випадків, коли модель помилково визначила наявність класу.

Для описаної моделі дане значення 0.82, що відповідає 82% специфічності.

Щоб описати більш точно співвідношення істинних спрацювань до хибно позитивних спрацювань для запропонованої моделі побудовано графік ROC на рис. 6 [17, 18]. Для кращої оцінки продуктивності моделі класифікації обраховуємо площу під кривою ROC (AUC) [19, 20]. Для запропонованої моделі AUC становить 0.90, що є високим значенням, а чим вище AUC, тим краще модель розрізняє класи.

Подальше використання описаної математичної моделі, що враховує можливі фактори ризику появи серцевих захворювань, забезпечує можливість простого та швидкого аналізу вхідних даних пацієнта для швидкого реагування на потенційне виникнення серцевих захворювань. Даний підхід дозволяє простіше та швидше виявляти серцеві захворювання та застосовувати ефективні методи лікування та попереджувати більш серйозні наслідки захворювання потенційних пацієнтів.

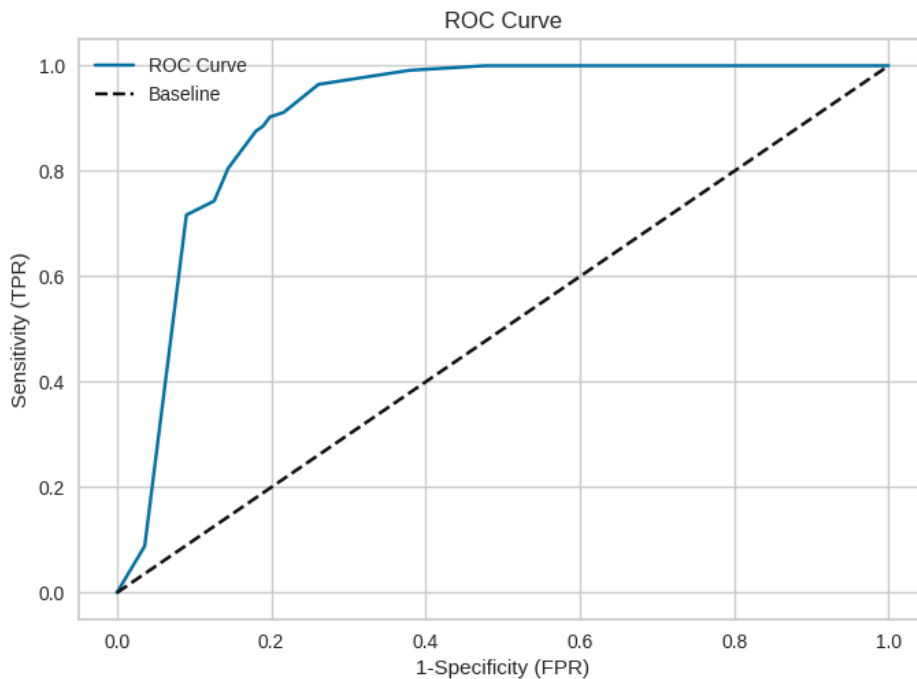


Рис.6. ROC крива

Подібні підходи також досліджуються в сучасних наукових роботах [21], пов'язаних із прогнозуванням та аналізом ряду базових показників для серцевої діагностики. Згідно із останніми дослідженнями доволі вагомим фактором виникнення серцевих хвороб залишаються рівень холестерину в крові людини, проте іншими вагомими факторами є рівень цукру в крові. Зазвичай першими симптомами будуть біль у грудях, підвищений кров'яний тиск та підвищена частота серцебиття особливо в результаті невеликих навантажень. Додатковими факторами, які не були включені в дану модель, є наявність ожиріння у пацієнта [22]. Також деякі дослідження вказують на залежність між рівнем стресу та ішемічними захворюваннями [23]. Хоча дані фактори і не були включені в дану модель, їхній вплив може бути врахований при подальшій деталізації та покращенні запропонованої моделі.

Наявні методи кардіодіагностики дозволяють отримати доволі багато інформації з використанням порівняно дешевих та простих методів діагностики. Також багато коли використовується метод анкетування [24] з метою отримання додаткової інформації для аналізу факторів ризику. Беручи до уваги дані факти можна зробити припущення про потенційну перспективність розвитку моделювання настання серцево-судинних захворювань шляхом аналізу більшої кількості вхідних факторів, що безпосередньо впливають на виникнення захворювань.

Базуючись на отриманих результатах, включаючи вплив медико-біологічних факторів [25] у пацієнтів із потенційними або наявними захворюваннями серця можливим буде використання запропонованої математичної моделі для проектування інформаційно-діагностичної системи оцінювання та прогнозування настання серцево-судинних захворювань. У наступних дослідженнях необхідно провести ширший аналіз можливих факторів та опрацювати можливість підвищення результатів передбачення та класифікації шляхом зміни порогових значень ідентифікації можливості потенційного настання захворювань серця.

Висновки та перспективи подальших дослідження. Запропонована математична модель, яка враховує фактори ризику настання серцево-судинних захворювань та дає змогу завчасно передбачити настання потенційних ускладнень або можливості виникнення захворювання в цілому. Також дана модель дасть змогу зменшити кількість звернень пов'язаних із хворобами серця за рахунок подальшої профілактики та зменшення впливу факторів ризику на організм пацієнта. Базуючись на результатах ROC аналізу, а саме опираючись на показник AUC, який становить 0.9, можна зробити висновок про те, що запропонована модель є перспективною для використання при проектуванні інформаційно-діагностичної системи прогнозування ризику виникнення серцево-судинних захворювань серця.

У наступних дослідженнях необхідно використати інші методи прогнозування настання серцевих захворювань, серед яких доцільно виділити багатофакторний регресійний аналіз, РСА-метод, нейромережі.

Список бібліографічного опису

1. Janosi Andras, Steinbrunn William, Pfisterer Matthias, Detrano Robert. 1988. Heart Disease. UCI Machine Learning Repository. doi: 10.24432/C52P4X.
2. Мазуренко В. П. Статистика: навч.-метод. посіб. / В. П. Мазуренко. – К.: ВПЦ «Київський університет», 2006.
3. Сверстюк А. С. Обґрунтування та верифікація математичної моделі синхронно зареєстрованих кардіосигналів з використанням вектора циклічних ритмічно пов'язаних випадкових процесів. Вимірвальна та обчислювальна техніка в технологічних процесах. 2009. № 1. С. 143–147.
4. Sofogianni A, Stalikas N, Antza C, Tziomalos K. Cardiovascular Risk Prediction Models and Scores in the Era of Personalized Medicine. *J Pers Med*. 2022. doi: 10.3390/jpm12071180.
5. Kaiser, H.F. An index of factorial simplicity. *Psychometrika* 39, 31–36 (1974). doi: 10.1007/bf02291575
6. Mensah, George A., Gregory A. Roth, and Valentin Fuster. "The global burden of cardiovascular diseases and risk factors: 2020 and beyond." *Journal of the American College of Cardiology* 74.20 (2019): 2529-2532.
7. Хом'юк І.В., Хом'юк В.В., Краєвський В.О. Теорія імовірностей та математична статистика. Навчальний посібник. / Вінниця: ВНТУ, 2009. – 189 с.
8. Geron A. Applied machine-learning using Scikit-learn and TensorFlow. 2018. Vol. 1.
9. Breiman, L. ; Friedman, Jerome H. ; Olshen, Richard A. ; Stone, C. J.: Classification and Regression Trees. In : , 1984
10. David James. 2018. Introduction to Machine Learning with Python: A Guide for Beginners in Data Science (1st. ed.). CreateSpace Independent Publishing Platform, North Charleston, SC, USA.
11. Tangirala, Suryakanthi. "Evaluating the impact of GINI index and information gain on classification using decision tree classifier algorithm." *International Journal of Advanced Computer Science and Applications* 11.2 (2020): 612-619.
12. Altman DG, Bland JM. Diagnostic tests. 1: Sensitivity and specificity. *BMJ*. 1994 Jun 11;308(6943):1552. doi: 10.1136/bmj.308.6943.1552.
13. Unpingco, Jose. Python for Probability, Statistics, and Machine Learning. 2019.
14. Leon Derczynski. Complementarity, F-score, and NLP Evaluation. In Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16), Portorož, Slovenia. European Language Resources Association (ELRA). 2016.
15. Ghiasi, Mohammad M., Sohrab Zendejboudi, and Ali Asghar Mohsenipour. "Decision tree-based diagnosis of coronary artery disease: CART model." *Computer methods and programs in biomedicine* 192 (2020): 105400.
16. Лупенко С., Зозуля А., Сверстюк А., Стадник Н. Математичне моделювання та методи опрацювання сигналів серця на базі циклічних випадкових процесів та векторів. *Sciences and Education a New Dimension. Natural and Technical Sciences*, VI (20). Issue 172. P. 47-54.
17. Лупенко С.А., Сверстюк А.С., Луцик Н.С., Стадник Н.Б., Зозуля А.М. Умовний циклічний випадковий процес як математична модель коливних сигналів та процесів із подвійною стохастичністю. Поліграфія і видавнича справа. 2016. № 1(71). С.147-159.
18. Lupenko S., Lytvynenko Ia., Sverstiuk A., Shelestovskyi B., Horkunenko A. Software for statistical processing and modeling of synchronously registered cardio signals. The Fourth International Workshop on Computer Modeling and Intelligent Systems (CMIS-2021). 27 of April 2021, Zaporizhzhia, Ukraine. p. 194 - 205.
19. Hu, Nan. "Using Receiver Operating Characteristic (ROC) Analysis to Evaluate Information-Based Decision-Making." *Advanced Methodologies and Technologies in Business Operations and Management*, edited by Mehdi Khosrow-Pour, D.B.A., IGI Global, 2019, pp. 764-776. doi: 10.4018/978-1-5225-7362-3.ch057
20. Carrington A.M. et al., "Deep ROC Analysis and AUC as Balanced Average Accuracy, for Improved Classifier Selection, Audit and Explanation," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 1, pp. 329-341, 1 Jan. 2023, doi: 10.1109/TPAMI.2022.3145392.
21. Townsend N., Nichols M., Scarborough P., Rayner M. Cardiovascular disease in Europe—Epidemiological update 2015. *Eur. Heart J*. 2015;36:2696–2705. doi: 10.1093/eurheartj/ehv428.
22. Powell-Wiley, Tiffany M., et al. "Obesity and cardiovascular disease: a scientific statement from the American Heart Association." *Circulation* 143.21 (2021): e984-e1010.
23. Szyller, Jakub, and Iwona Bil-Lula. "Heat shock proteins in oxidative stress and ischemia/reperfusion injury and benefits from physical exercises: a review to the current knowledge." *Oxidative Medicine and Cellular Longevity* 2021 (2021): 1-12.
24. Raffle, Angela E., and JA Muir Gray. Screening: evidence and practice. Oxford University Press, USA, 2019.
25. Марценюк В.П., Кравець Н.О., Сверстюк А.С. Інформаційна система медико-біологічних досліджень: проєкт на основі Web-технологій. Український журнал телемедицини та медичної телематики. 2003. №1(1). С. 57-60.

References

1. Janosi Andras, Steinbrunn William, Pfisterer Matthias, Detrano Robert. 1988. Heart Disease. UCI Machine Learning Repository. doi: 10.24432/C52P4X.

2. Mazurenko V. P. Statystyka: navch.-metod. posib. / V. P. Mazurenko. – K.: VPTs «Kyivskyy universytet», 2006.
3. Sverstiuk A.S. Obhruntuvannia ta veryfikatsiia matematychnoi modeli synkronno zareiestrovanykh kardiosyhnaliv z vykorystanniam vektora tsyklichnykh rytmichno pov'iazanykh vypadkovykh protsesiv. Vymiriuvalna ta obchysliuvalna tekhnika v tekhnolohichnykh protsesakh. 2009. № 1. S. 143–147.
4. Sofogianni A, Stalikas N, Antza C, Tziomalos K. Cardiovascular Risk Prediction Models and Scores in the Era of Personalized Medicine. *J Pers Med.* 2022. doi: 10.3390/jpm12071180.
5. Kaiser, H.F. An index of factorial simplicity. *Psychometrika* 39, 31–36 (1974). doi: 10.1007/bf02291575
6. Mensah, George A., Gregory A. Roth, and Valentin Fuster. "The global burden of cardiovascular diseases and risk factors: 2020 and beyond." *Journal of the American College of Cardiology* 74.20 (2019): 2529-2532.
7. Khom'yuk I.V., Khom'yuk V.V., Krayevs'kyi V.O. Teoriya imovirnostey ta matematychna statystyka. Navchal'nyy posibnyk. / Vinnytsya: VNTU, 2009. – 189 s.
8. Geron A. Applied machine-learning using Scikit-learn and TensorFlow. 2018. Vol. 1.
9. Breiman, L. ; Friedman, Jerome H. ; Olshen, Richard A. ; Stone, C. J.: Classification and Regression Trees. In: , 1984
10. David James. 2018. Introduction to Machine Learning with Python: A Guide for Beginners in Data Science (1st. ed.). CreateSpace Independent Publishing Platform, North Charleston, SC, USA.
11. Tangirala, Suryakanthi. "Evaluating the impact of GINI index and information gain on classification using decision tree classifier algorithm." *International Journal of Advanced Computer Science and Applications* 11.2 (2020): 612-619.
12. Altman DG, Bland JM. Diagnostic tests. 1: Sensitivity and specificity. *BMJ.* 1994 Jun 11;308(6943):1552. doi: 10.1136/bmj.308.6943.1552.
13. Unpingco, Jose. Python for Probability, Statistics, and Machine Learning. 2019.
14. Leon Derczynski. Complementarity, F-score, and NLP Evaluation. In Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16), Portorož, Slovenia. European Language Resources Association (ELRA). 2016.
15. Ghiasi, Mohammad M., Sohrab Zendejboudi, and Ali Asghar Mohsenipour. "Decision tree-based diagnosis of coronary artery disease: CART model." *Computer methods and programs in biomedicine* 192 (2020): 105400.
16. Lupenko S., Zozulia A., Sverstiuk A., Stadnyk N. Matematychni modeliuvannia ta metody opratsiuvannia syhnaliv sertsia na bazi tsyklichnykh vypadkovykh protsesiv ta vektoriv. *Sciences and Education a New Dimension. Natural and Technical Sciences.* 2018. VI (20). Issue 172. P. 47-54.
17. Lupenko S.A., Sverstiuk A.S., Lutsyk N.S., Stadnyk N.B., Zozulia A.M. Umovnyi tsyklichnyi vypadkovyi protses yak matematychna model kolyvnykh syhnaliv ta protsesiv iz podviinoiu stokhastychnosti. Polihrafiia i vydavnycha sprava. 2016. № 1(71). S.147-159.
18. Lupenko S., Lytvynenko Ia., Sverstiuk A., Shelestovskyi B., Horkunenko A. Software for statistical processing and modeling of synchronously registered cardio signals. The Fourth International Workshop on Computer Modeling and Intelligent Systems (CMIS-2021). 27 of April 2021, Zaporizhzhia, Ukraine. p. 194 - 205.
19. Hu, Nan. "Using Receiver Operating Characteristic (ROC) Analysis to Evaluate Information-Based Decision-Making." *Advanced Methodologies and Technologies in Business Operations and Management*, edited by Mehdi Khosrow-Pour, D.B.A., IGI Global, 2019, pp. 764-776. doi: 10.4018/978-1-5225-7362-3.ch057
20. M. Carrington et al., "Deep ROC Analysis and AUC as Balanced Average Accuracy, for Improved Classifier Selection, Audit and Explanation," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 1, pp. 329-341, 1 Jan. 2023, doi: 10.1109/TPAMI.2022.3145392.
21. Townsend N., Nichols M., Scarborough P., Rayner M. Cardiovascular disease in Europe—Epidemiological update 2015. *Eur. Heart J.* 2015;36:2696–2705. doi: 10.1093/eurheartj/ehv428.
22. Powell-Wiley, Tiffany M., et al. "Obesity and cardiovascular disease: a scientific statement from the American Heart Association." *Circulation* 143.21 (2021): e984-e1010.
23. Szyller, Jakub, and Iwona Bil-Lula. "Heat shock proteins in oxidative stress and ischemia/reperfusion injury and benefits from physical exercises: a review to the current knowledge." *Oxidative Medicine and Cellular Longevity* 2021 (2021): 1-12.
24. Raffle, Angela E., and JA Muir Gray. Screening: evidence and practice. Oxford University Press, USA, 2019.
25. Martseniuk V.P., Kravets N.O., Sverstiuk A.S. Informatsiina systema medyko-biologichnykh doslidzen: proekt na osnovi Web-tekhnolohii. *Ukrainskyi zhurnal teledytsyny ta medychnoi telematyky.* 2003. 1(1). S. 57-60.