

МЕТОД ДЕРЕВ РІШЕНЬ У КОМП'ЮТЕРНІЙ ДІАГНОСТИЦІ

В роботі розглянуто питання прийняття рішень за допомогою дерев класифікації під час діагностики травм. Проведений аналіз використання методів роботи з архівними базами даних, накопичених у лікарнях у вигляді паперових картотек. Показано, що після комп'ютеризації даних пацієнтів для кращої класифікації і орієнтації по пацієнтам доцільно впроваджувати систему дерева рішень. Зазначений підхід дозволить визначити пацієнтів з однаковою хворобою, діагнозом чи симптомом. Після створення алгоритму дерева рішень необхідно здійснювати підбір програми, яка б задовільнила алфавіт ознак вхідних даних. Наведено приклад використання розробленого дерева рішень для визначення закритого перелому. Запропоновані рішення дозволять вирішувати складні завдання та приймати рішення, пов'язані з проектуванням, розробкою та управлінням програмними та апаратними системами комп'ютерної діагностики у медичній та інших сферах, пов'язаних з обробкою та класифікацією персоналії баз даних за визначеними параметрами, а також в умовах невизначеності.

В процесі побудовання дерева рішень для виведення норми і патології під час комп'ютерної діагностики закритого перелому запропоновано замість двопараметрового значення вводити трипараметрове. В такому разі кількість ознак різко збільшується, але при цьому й зростає можливість появи помилки першого і другого роду. Аналіз інформативності ознак показує, що найбільш інформативною незалежною змінною (ознакою образу) є зміна форми, одразу за нею – показник міри болю, і лише потім набряк і крепітація. Якщо в такій самій послідовності проводити класифікацію, тоді ентропійно-інформаційний аналіз дозволяє ранжувати ознаки образу за їх інформативністю. Задля подолання специфічності, при якій 92,3 % з діагностованих пацієнтів, віднесених класифікатором до категорії «патологія», насправді є хворими, та 7,7 % навпаки, є здоровими, – необхідно більш ретельно розглядати незалежні атрибути.

Ключові слова: комп'ютерна діагностика, база даних, прийняття рішень, дерево рішень, ентропійно-інформаційний аналіз

OMELCHENKO IHOR

Petro Mohyla Black Sea National University

A TREE DECISION METHOD IN COMPUTER DIAGNOSTICS

The paper considers the issue of decision-making using classification trees during the diagnosis of injuries. An analysis of the use of methods of working with archival databases accumulated in hospitals in the form of paper files was carried out. It is shown that after computerization of patient data, it is advisable to implement a decision tree system for better classification and patient orientation. This approach will allow identifying patients with the same disease, diagnosis or symptom. After creating a decision tree algorithm, it is necessary to select a program that would satisfy the alphabet of the input data. An example of using the developed decision tree to determine a closed fracture is given. The proposed solutions will make it possible to solve complex tasks and make decisions related to the design, development and management of software and hardware systems of computer diagnostics in medical and other fields related to the processing and classification of database personalities according to specified parameters, as well as in conditions of uncertainty. In the process of building a decision tree for deriving the norm and pathology during computer diagnostics of a closed fracture, it is proposed to enter a three-parameter value instead of a two-parameter value. In this case, the number of signs increases dramatically, but at the same time, the possibility of errors of the first and second kind also increases. The analysis of the informativeness of the signs shows that the most informative independent variable (sign of the image) is a change in shape, immediately after it is an indicator of the degree of pain, and only then swelling and crepitation. If the classification is carried out in the same sequence, then the entropy-informational analysis allows ranking the features of the image according to their informativeness. In order to overcome the so-called specificity, in which 92.3% of the diagnosed patients classified by the classifier as "pathology" are actually sick, and 7.7%, on the contrary, are healthy, it is necessary to consider independent attributes more carefully.

Keywords: computer diagnostics, database, decision making, decision tree, entropy and information analysis

Постановка проблеми

Теорія прийняття рішень відіграє важливу роль у комп'ютерній інженерії, тому що в цій галузі часто доводиться вирішувати складні завдання та приймати рішення, пов'язані з проектуванням, розробкою та управлінням програмними та апаратними системами. До ключових аспектів теорії прийняття рішень, які застосовуються в комп'ютерній інженерії, можна віднести такі:

- аналіз вимог: перш ніж розпочати проектування чи розробку програми чи системи, необхідно провести аналіз вимог. Це включає збір інформації про цілі і завдання системи, а також обмеження і очікування користувачів. Аналіз вимог допомагає визначити, які рішення будуть найефективнішими для досягнення поставленої мети;

- моделювання та симуляція: для оцінки різних альтернативних рішень часто використовуються моделювання та симуляція. Це дозволяє інженерам створювати віртуальні моделі системи та проводити експерименти, щоб визначити, які рішення будуть найбільш ефективними та оптимальними;

- аналіз ризиків: прийняття рішень у комп'ютерній інженерії також включає оцінку ризиків. Інженери повинні аналізувати потенційні ризики, пов'язані з вибором конкретних рішень, та розробляти стратегії щодо їх управління. Це може включати заходи щодо зниження ризиків або планування резервних

варіантів;

- однією з основних цілей у комп'ютерній інженерії є оптимізація продуктивності та ефективності систем. Інженери застосовують методи оптимізації для вибору найкращих рішень з урахуванням обмежень, таких як бюджет, час та інші ресурси;

- прийняття рішень за умов невизначеності: в інженерії часто існує невизначеність і невідомі чинники. Теорія прийняття рішень включає методи роботи з невизначеністю, такі як статистичний аналіз, теорія ймовірності та багато інших, щоб приймати рішення за недостатньої інформації;

- інструменти підтримки прийняття рішень: У комп'ютерній інженерії також використовуються спеціальні інструменти для підтримки прийняття рішень, такі як системи управління проектами, аналітичні інструменти та методи штучного інтелекту, включаючи машинне навчання;

- цикли зворотного зв'язку: Після прийняття рішення важливо оцінювати його результати та проводити зворотний зв'язок. Якщо рішення не відповідає очікуванням чи цілям, воно може вимагати корекції чи перегляду.

В державних лікарнях крани є архівна база даних у вигляді паперових картотек, які не зручно зберігати, переглядати та аналізувати, та з розвитком телемедицини настала необхідність комп'ютеризувати дані пацієнтів. Це значно пришвидшить обробку даних по пацієнтам, але для кращої класифікації і орієнтації по пацієнтам необхідно впроваджувати систему дерева рішень, оскільки, з нею можна буде визначити пацієнтів з однаковою хворобою, діагнозом чи симптомом.

Аналіз останніх джерел

Дерева класифікації не обмежені використанням тільки одновимірних розгалужень по предикторним змінним. Якщо безперервні предиктори виміряні хоча б за інтервальною шкалою, то дерева класифікації можуть використовувати розгалуження по лінійним комбінаціям, подібно тому, як це робиться в лінійному дискримінантному аналізі. При цьому розгалуження за лінійними комбінаціями, які застосовуються для побудови дерев класифікації, мають ряд важливих відмінностей від своїх аналогів з дискримінантного аналізу. У лінійному дискримінантному аналізі максимальну кількість лінійних дискримінантних функцій одного мінімуму з числа предикторних змінних і числа класів залежної змінної мінус один. При рекурсивному підході, який використовується в модулі дерева класифікації, яке не має цих обмежень. Наприклад, для десяти предикторних змінних і всього двох класів залежної змінної може на використовувати десять послідовних розгалужень по лінійним комбінаціям. Це вигідно відрізняється від єдиного розгалуження по лінійній комбінації, пропонуваного в даному випадку традиційним нерекурсивним лінійним дискримінантним аналізом. При цьому значна частина інформації, що міститься в предикторних змінних, може залишитися невикористаною.

Розглянувши ситуацію, коли є багато категорій, але мало предикторів, необхідно припустити, що потрібно розсортувати монети різних номіналів, маючи тільки дані їх розмірів (товщина та діаметр). У звичайному лінійному дискримінантному аналізі можна отримати максимум дві дискримінантні функції, і монети можуть бути успішно розсортовані тільки в тому випадку, якщо вони розрізняються не більше ніж двома параметрами, які представлені у вигляді лінійних комбінацій товщини та діаметру монети. Навпаки, в підході, який використовується в модулі дерева класифікації немає обмежень в кількості розгалужень по лінійним комбінаціям, які можна виконати.

Апарат розгалуження по лінійним комбінаціям, реалізований в модулі дерева класифікації, може бути використаний також як метод аналізу при побудові дерев класифікації з одновимірним розгалуженням. Насправді одномірне розгалуження - окремий випадок розгалуження по лінійній комбінації. Можна уявити таке розгалуження по лінійній комбінації, при якому вагові коефіцієнти при всіх предикторних змінних, крім якоїсь однієї, рівні нулю. Оскільки значення комбінації фактично залежить від значень тільки однієї предикторної змінної (коефіцієнт при якій відмінний від нуля), отримане в результаті цього розгалуження буде одновимірним.

Реалізовані в модулі дерева класифікації методи дискримінантного одновимірного розгалуження по категоріальним і порядковим предикторам і дискримінантного багатовимірного розгалуження по лінійним комбінаціям порядкових предикторів є адаптацію відповідних алгоритмів пакета QUEST (Quick, Unbiased, Efficient Statistical Trees). QUEST – це програма дерев класифікації, розроблена Loh і Shih (1997), в якій використовуються поліпшені варіанти методу рекурсивного квадратичного дискримінантного аналізу і яка містить ряд нових засобів для підвищення надійності та ефективності дерев класифікації, які вона буде.

Алгоритми пакета QUEST досить складні, проте в модулі Дерева класифікації є опція Тип розгалуження, що надає користувачеві інший, концептуально більш простий підхід. Реалізований тут алгоритм одновимірного розгалуження за методом CART є адаптацією алгоритмів пакета CART, див Breiman та ін (1984). CART (Classification And Regression Trees) - це програма дерев класифікації, яка при побудові дерева здійснює повний перебір всіх можливих варіантів одновимірного розгалуження.

Опції аналізу QUEST і CART природно доповнюють один одного. У випадках, коли є багато предикторних змінних з великим числом рівнів, пошук методом CART може виявитися досить тривалим. Крім того, цей метод має схильність вибирати для розгалуження ті предикторні змінні, у яких більше рівнів. Однак оскільки тут виконується повний перебір варіантів, є гарантія, що буде знайдений варіант розгалуження, що дає найкращу класифікацію (по відношенню до навчальної вибірки).

Метод QUEST – швидкий. Його перевага в швидкості перед методом CART стає особливо помітною, коли предикторні змінні мають десятки рівнів, метод QUEST необхідна була 1 секунда часу процесора, а CART – 30,5 годин). Відсутність у методу QUEST зміщення у виборі змінних для розгалуження також є його суттєвою перевагою у випадках, коли одні предикторні змінні мають мало рівнів, а інші – багато (предиктори з багатьма рівнями часто породжують "методи тикю", які добре узгоджуються з даними, але дають погану точність прогнозу. Нарешті, метод QUEST не жертвує точністю прогнозу заради швидкості обчислень (Lim, Loh, & Shih, 1997). Поєднання опцій QUEST і CART дозволяє повністю використовувати всю гнучкість апарату дерев класифікації.

Незважаючи на наявність ґрунтовних розробок, недостатньо дослідженим залишається питання прийняття рішень за допомогою дерев класифікації під час діагностики травм.

У той час, як більшість наведених досліджень зосереджені на створенні вибірки з травматологічної діагностики та відборі ознак по захворюванням, існує необхідність створення бази даних по травматологічним захворюванням, створенню алгоритму дерева рішень та підбору програм, яка б задовільнила алфавіт ознак вхідних даних.

Метою роботи є: дослідження методів створення дерев рішень у комп'ютерній діагностиці, побудова графічного образу та опис його процедурної реалізації.

Виклад основного матеріалу

Всі чотири незалежних змінних з таблиці 1 визначаються двома дискретними атрибутами. Цільову функцію кодуємо символами двозначної логіки (так або ні). Спробуємо побудувати дерево рішення за такими правилами:

- у вузлах дерева, які не є його листям, знаходяться незалежні змінні за яким розрізняються випадки;
- у листях дерева знаходяться значення цільової функції (залежної змінної);
- на ребрах дерева знаходяться атрибути, які характеризують незалежні змінні.

Для побудови дерева рішень введемо поняття ентропії . Припустимо, що є множина A , яка складається з n елементів, частина з яких ($m < n$) характеризуються деякою властивістю S . Тоді ентропія множини відносно властивості обчислюється цією формулою:

$$H(A, s) = - \frac{m \cdot \log_2\left(\frac{m}{n}\right)}{n} - \frac{(n - m) \cdot \log_2\left(\frac{n - m}{n}\right)}{n} \quad (1)$$

Обчислимо зокрема ентропію цільової функції по властивості $S = \text{"так"}$:

> **H:=unapply(-m*log[2](m/n)/n-(n-m)*log[2]((n-m)/n),n,m);**

$$H := (n, m) \rightarrow - \frac{m \ln\left(\frac{m}{n}\right)}{\ln(2) n} - \frac{(n - m) \ln\left(\frac{n - m}{n}\right)}{\ln(2) n}$$

> **H_As:=evalf(H(16,8));**

$$H_{As} := 1.$$

Припустимо тепер, що множина A (цільова функція, частина елементів якої мають властивість-ознаку "так") класифікується певною незалежною змінною, яка характеризується атрибутом Q , котрий може приймати q різних дискретних значень. Тоді зміна ентропії (приріст інформації, англ. information gain) можна характеризувати так:

$$gain(A, Q) = H(A, s) - \sum_{k=1}^q \frac{n_k}{n} H(A_k, s) \quad (2),$$

де n_k – кількість елементів множини A , які мають атрибут k ; $H(A_k, s)$ – ентропія підмножини $A_k \subseteq A$, які мають атрибут k відносно властивості S .

Обчислимо за формулою (2) зміну ентропії, якщо класифікувати незалежною змінною "с – набряк, пухлина", характеризується двома ознаками наявна пухлина чи ні: пухлина наявна: $n_1 = 7$ (перша колонка таблиці), $m_1 = 4$ пухлина відсутня: $n_2 = 9, m_2 = 4$

> **gain_Ac:=evalf(H_As-(7*H(7,4)+9*H(9,4))/16);# приріст інформації за умови класифікації пухлин**

$$gain_{Ac} := 0.0114824072$$

Приріст інформації відносно невеликий і наявність(відсутність пухлини), схоже, не найкраща ознака для класифікації випадків таблиці. Обчислимо також зміни ентропії (приріст інформації) для інших незалежних змінних (ознак), користуючись даними таблиці:

> **gain_form:=evalf(H_As-(8*H(8,6)+8*H(8,2))/16); # Зміна форми**

$$gain_{form} := 0.1887218755$$

> **gain_Ah:=evalf(H_As-(9*H(9,7)+7*H(7,1))/16); # Видно кістку**

$$gain_{Ah} := 0.3112781245$$

```
> gain_Aw:=evalf(H_As-(9*H(9,6)+7*H(7,2))/16); # крепірація
gain_Aw := 0.1058433444
```

Аналіз інформативності ознак показує, що найбільш інформативною змінною (ознакою образу) є видимість кістки, за нею – зміна форми кісточки і лише потім наявність пухлин чи криперація (майже рівною мірою). Таким чином ентропійно-інформаційний аналіз дозволяє ранжувати ознаки образу за їх інформативністю.

Іншими словами, вектор ознак варто створювати у такій послідовності $X = (Kist, Form, Puh, Krep)$ і в такій самій послідовності проводити класифікацію. Таким чином, ентропійно-інформаційний аналіз дозволяє ранжувати ознаки образу за їх інформативністю. Тепер можна побудувати дерево рішень, послідовно класифікуючи випадки (спостереження) за ознаками згідно з їх рангом (спочатку по більш інформативним, потім – по менш інформативним).

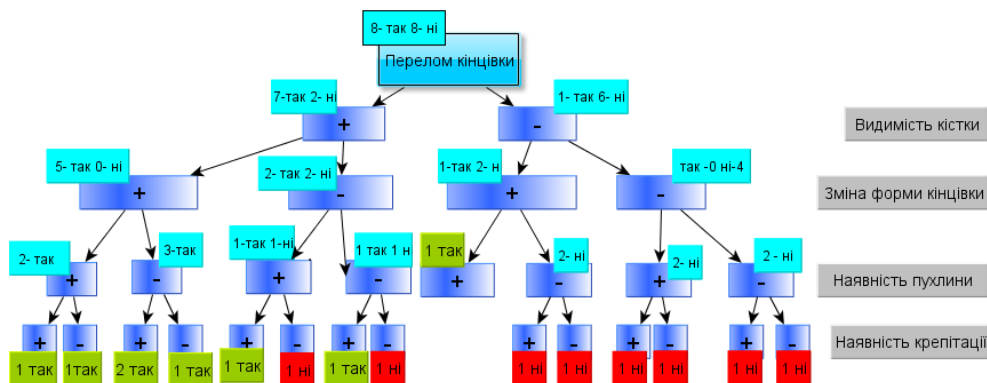


Рис. 1. Дерево прийняття рішень по переломам

Розглянемо вектор-ознаку $x = (0,0,0,0)$. Вже по перших двох ознаках (незалежно від решти) дерево рішень прогнозує, що перелом відсутній. Для опису правила прийняття рішення після побудови дерева рішень доволі легко формалізуються у такій простенькій процедурі із застосуванням умовних виразів. ситуації у вигляді процедури `xvor`, яка характеризується 4 ознаками з таблиці 2:

```
> restart:
```

```
xvor:=proc (ki,f,p,kr)
if ki=1 and f=1 then print('patologia')
elif ki=1 and p=1 then print('patologia')
elif f=1 and p=1 and kr=1 then print('patologia')
elif f=1 and p=1 and kr=1 then print('patologia')
elif ki=0 and p=1 and f=1 then print('patologia')
else print('norma') end if end proc:
```

Далі прогнозування (діагностика) можуть відбуватися автоматично. Закодований вектор ознак подаємо на вхід процедури, яка надає прогноз:

```
> xvor(1,1,1,1);# Вектори ( кістка видима, присутня зміна форми кінцівки, наявна пухлина)
xvor(1,1,0,0);
xvor(1,0,1,1);
xvor(0,0,0,1);
xvor(0,0,1,1);
```

```
patologia
patologia
patologia
norma
norma
```

Розрахунки і побудова дерева рішення для закритого перелому

Для більш наглядного прикладу змінимо завдання – розробимо дерево рішень для виведення норми і патології для закритого перелому. Для цього змінимо незалежні зміни та введемо тривекторну ознаку «сила болю» замість «видимість кістки», оскільки видимість кістки непотрібна для визначення закритого перелому.

Отже, для створення дерева рішень створимо таблицю ознак по таким критеріям «Сила болю», «Набряк, пухлина», «Зміна форми кінцівки» і «Крепітація».

Оскільки в таблицю ввели замість двопараметрового значення – трипараметрове, то кількість ознак різко збільшилася, а отже і зросла можливість появи помилки першого і другого роду.

Таблиця 1

Атрибути закритого перелому

Незалежні змінні та їх атрибути				Залежна змінна (цільова функція)
Сила болю	Набряк, пухлина	Зміна форми кінцівки	Крепітація	Чи є закритий перелом?
1	1	1	1	Так
2	0	0	1	Так
2	0	1	1	Так
2	1	1	1	Так
0	1	1	1	Так
1	0	1	1	Так
2	0	1	0	Так
1	0	1	1	Так
1	1	0	1	Так
1	0	0	1	Ні
2	1	1	0	Ні
0	1	0	0	Ні
0	1	0	1	Ні
1	1	0	0	Ні
0	0	1	1	Ні
0	0	1	0	Ні
1	0	0	0	Ні
0	0	0	0	Ні
0	0	0	0	Ні
0	0	0	1	Ні
2	0	0	0	Ні
1	1	1	0	Ні

Пояснення до таблиці 1:

- а) сила болю 0 – біль не відчувається, 1 – біль помірний, 2 – біль нестерпний;
- б) набряк, пухлина 1 – набряк(пухлина) наявний, 0 – набряк (пухлина) відсутній;
- в) зміна форми кінцівки 1– зміна форми відбувається, 0 – форма кінцівки не змінюється;
- г) крепітація 1 – присутня, 0 – крепітація не діагностована.

Обчислимо ентропію цільової функції по властивості $S = \text{"так"}$, для цього порахуємо кількість відповідей так в 6-ому стовпчику і співвіднесемо їх до загальної кількості відповідей, за рахунок формули ентропії:

> restart;

> H:=unapply(-m × log[2](m/n)/n-(n-m) × log[2]((n-m)/n)/n,n,m);

$$H := (n, m) \rightarrow -\frac{m \ln\left(\frac{m}{n}\right)}{\ln(2)n} - \frac{(n - m) \ln\left(\frac{n - m}{n}\right)}{\ln(2)n}$$

> H_perelom:=evalf(H(21,9));

Як видно з нових розрахунків, значення ентропії змінилося з 1 до 0,985 внаслідок зміни виборки по травмі.

Обчислимо за формулою зміну ентропії. Сила болю характеризуються трьома атрибутами:

- біль не відчувається (0): $n_1 = 7$ (перша колонка таблиці), $m_1 = 1$;
- біль не відчувається (0): $n_1 = 7$ (перша колонка таблиці), $m_1 = 1$;
- біль помірний (1): $n_2 = 8$, $m_2 = 4$;
- біль нестерпний (2): $n_2 = 6$, $m_2 = 4$.

> gain_pain:=evalf(H_perelom-(7*H(7,1)+8*H(8,4)+6*H(6,4))/21); # приріст інформації за умови показнику болю

$$\text{gain_pain} := 0.1446812571$$

Після розрахунку зміни ентропії по показнику сила болю, розрахуємо показник зміни ентропії для інших параметрів:

> gain_nabriak:=evalf(H_perelom-(9*H(9,4)+11*H(12,5))/21);

$$\text{gain_nabriak} := 0.0472166663$$

> gain_form:=evalf(H_perelom-(11*H(11,7)+10*H(10,2))/21);

```
gain_form := 0.1461069788
> gain_krepitacia:=evalf(H_perelom-(12*H(12,8)+9*H(6,3))/21);
gain_krepitacia := 0.0319162306
```

Таблиця 3. Приріст інформації по закритому перелому

Ознака класифікації	Приріст інформації
Біль	0.145
Набряк	0.047
Зміна форми	0.146
Крепітація	0.032

Аналіз інформативності ознак показує, що найбільш інформативною незалежною змінною (ознакою образу) є зміна форми, одразу за нею – показник міри болю, і лише потім набряк і крепітація. Іншими словами вектор ознак варто створювати у такій послідовності $x = (f, p, n, k)$, і в такій самій послідовності проводити класифікацію. Таким чином ентропійно-інформаційний аналіз дозволяє ранжувати ознаки образу за їх інформативністю.

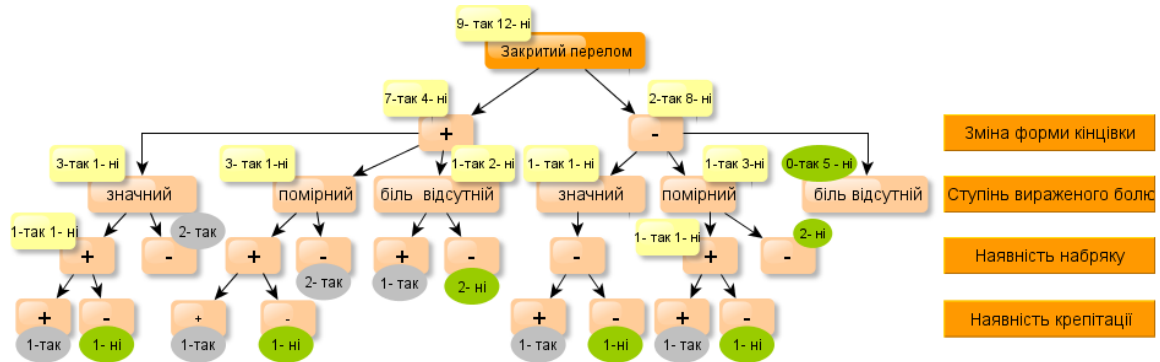


Рис. 2. Дерево прийняття рішень по закритим переломам

Клінічні тести, використовувані в клінічному обстеженні, не ідеальні. Завжди залишається ймовірність того, що результати діагностичного дослідження не відображають об'єктивне наявність або відсутність захворювання. Зазвичай для будь-якого клінічного тесту існує декілька (в різного ступеня точних) альтернатив. Різна доступність, вартість, безпека (а також фактори часу і здорового глузду) обмежують широке застосування тих чи інших методів обстеження. Для кожного захворювання існує «золотий стандарт діагностики» – найбільш точний діагностичний метод, за допомогою якого можна встановити наявність або відсутність даного захворювання. Як правило, застосування еталонного методу діагностики обмежується його незручностями - від високого ризику ускладнень до вартості. Порівнюючи новий більш точний метод зі старим стандартним будуть виявлятися додаткові позитивні і негативні результати. Ці результати будуть лише здаватися помилково-позитивними і помилково-негативними.

Терміни позитивний результат тесту і негативний результат тесту використовуються в їх звичайному значенні для позначення наявності чи відсутності захворювання відповідно.

Побудуємо чотирихпольову таблицю, для ілюстрації співвідношень між результатами даного клінічного тесту і «золотим стандартом діагностики».

Всього обстежувалось $a + c$ хворих пацієнтів (за результатами тесту) та $b + d$ здорових пацієнтів. Нашим гіпотетичним тестом правильно виявлено a хворих (з $a + c$ всього хворих) і d здорових (з $b + d$ всього здорових).

Ці співвідношення правильно діагностованих пацієнтів носять назви чутливості і специфічності клінічного дослідження.

Ймовірності Se та Sp у діагностичній практиці називають відповідно чутливістю та специфічністю діагнозу. Чутливість $Se = a / (a + c)$. Специфічність $Sp = d / (b + d)$.

Таблиця 2

		Захворювання		
		Присутнє	Відсутнє	
Тест	Позитивний	A	B	A+B
	Негативний	C	D	C+D
		A+C	B+D	

```

> restart:
xvor:=proc (f,p,pu,kr)
if f=1 and p=1 and pu=1 then print(`patologia`)
elif f=1 and p=2 and pu=1 then print(`patologia`)
elif f=1 and p=1 and kr=1 then print(`patologia`)
elif f=1 and pu=1 and kr=1 then print(`patologia`)
elif p=2 and f=1 and kr=1 then print(`patologia`)
elif p=1 and f=0 and kr=0 then print(`patologia`)
else print(`norma`) end if end proc:

```

Далі прогнозування (діагностика) можуть відбуватися автоматично. Закодований вектор ознак подаємо на вхід процедури, яка надає прогноз:

```

> xvor(1,1,1,1);# Вектори ( зміна форми кистки, середня сила болю, наявна пухлина, наявна крепітація)
xvor(1,1,0,0);
xvor(1,0,1,1);
xvor(0,0,0,1);
xvor(0,0,1,1);

```

```

patologia
norma
patologia
norma
norma

```

Вектори в даній виборці навчені правильно, а ось вектори подані нижче описані невірно і є похибками.

```

> xvor(0,2,1,1); # Патологія, визначена як норма, оскільки в навчанні не міститься даного вектору;
xvor(1,0,0,0); # Норма, визначена як патологія, оскільки в описі навчання містилася помилка оператора

```

```

norma
norma

```

```

> a:=9; c:=1; d:=12; b:=1; # a – кількість правильно визначених хворих, c – кількість хибно визначених хворих, d – кількість правильно визначених здорових пацієнтів, b – кількість неправильно визначених здорових пацієнтів.

```

```

Se:=a/(a+c) × 100; # Визначення чутливості системи

```

```

a := 9
c := 1
d := 12
b := 1
Se := 90

```

```

> Sp:=d/(b+d) × 100;
evalf(Sp);

```

```

Sp :=  $\frac{1200}{13}$ 
92.30769231

```

Висновки та перспективи подальшого дослідження

Зокрема специфічність означає, що приблизно 92,3 % діагностованих, віднесених нашим класифікатором до категорії "патологія", насправді є хворими, у той час як 7,7 %, віднесені до цієї категорії помилково, і є здоровими. У той же час чутливість, в 90 % гарантує, що 10 % хворих пацієнтів опиняться у групі здорових людей. Якщо від специфічності легко позбавитися шляхом зміни коду в навчальній вибірці, то специфічність виправити складніше, оскільки необхідно більш ретельно розглядати незалежні атрибути.

Література

1. Padiya S. D., Gulhane V. S. Analysis of Bluetooth versions (4.0, 4.2, 5, 5.1, and 5.2) for IoT applications. *In book: Implementing Data Analytics and Architectures for Next Generation Wireless Communications*. Chapter: 10. IGI Global, 2021. P. 153-178. DOI: 10.4018/978-1-7998-6988-7.ch010.
2. Cha S.-C., Yeh K.-H., Chen J.-F. Toward a robust security paradigm for Bluetooth low energy-based smart objects in the Internet-of-Things. *Sensors*. 2017. Vol. 17, no. 2348. P. 1–17. DOI:10.3390/s17102348.
3. Gangwal A., Singh S., Spolaor R., Srivastava A. BLE Whisperer: Exploiting BLE Advertisements for data exfiltration. *In book: Computer Security*. ESORICS, 2022. DOI: 10.1007/978-3-031-17140-6_34.
4. Дерево прийняття рішень для операції. URL: <http://www.d22d.ru/load/29-1-0-523> (дата звернення: 21.12.2023).
5. Визначення чутливості та специфічності діагностичного обстеження. URL: <http://ebm.org.ua/clinical-epidemiology/testing/sensitivity-specificity> (дата звернення: 21.12.2023).

References

1. Padiya S. D., Gulhane V. S. Analysis of Bluetooth versions (4.0, 4.2, 5, 5.1, and 5.2) for IoT applications. *In book: Implementing Data Analytics and Architectures for Next Generation Wireless Communications*. Chapter: 10. IGI Global, 2021. P. 153-178. DOI: 10.4018/978-1-7998-6988-7.ch010.
2. Cha S.-C., Yeh K.-H., Chen J.-F. Toward a robust security paradigm for Bluetooth low energy-based smart objects in the Internet-of-Things. *Sensors*. 2017. Vol. 17, no. 2348. P. 1–17. DOI:10.3390/s17102348.
3. Gangwal A., Singh S., Spolaor R., Srivastava A. BLE Whisperer: Exploiting BLE Advertisements for data exfiltration. *In book: Computer Security*. ESORICS, 2022. DOI: 10.1007/978-3-031-17140-6_34.
4. A decision tree for an operation. URL: <http://www.d22d.ru/load/29-1-0-523> (дата звернення: 21.12.2023).
5. Determination of sensitivity and specificity of diagnostic examination. URL: <http://ebm.org.ua/clinical-epidemiology/testing/sensitivity-specificity> (дата звернення: 21.12.2023).