

О.А. ЖУКОВСЬКА

Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», Київ, Україна, e-mail: zhukovskaya71@gmail.com.

Л.С. ФАЙНЗІЛЬБЕРГ

Міжнародний науково-навчальний центр інформаційних технологій і систем НАН та МОН України; Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», Київ, Україна, e-mail: fainzilberg@gmail.com.

ОЦІНЮВАННЯ КОРИСНОСТІ БІНАРНОГО КЛАСИФІКАТОРА НА ОСНОВІ УДОСКОНАЛЕНОГО МЕТОДУ ROC-АНАЛІЗУ

Анотація. Сформульовано означення корисності бінарного класифікатора в сенсі зменшення апіорного ризику помилкової класифікації. Запропоновано достатні умови, що гарантують корисність діагностичного тесту згідно з цим означенням. Отримані умови надали змогу вдосконалити традиційний ROC-аналіз завдяки обмеженню відповідного фрагмента ROC-кривої. Показано, що пряма, яка обмежує фрагмент ROC-кривої гарантовано корисного тесту, збігається з відомою ізолінією ефективності, що відповідає рівню апіорного ризику. Визначено допустимі межі співвідношення втрат від помилки пропуску цілі та хибної тривоги, за яких тест із відповідними операційними характеристиками залишається корисним для скринінгу захворювання з відомим преваленсом. На основі отриманих результатів обґрунтовано ефективність нового методу аналізу та інтерпретації електрокардіограм, який ґрунтується на визначенні оригінальної діагностичної ознаки у фазовому просторі та надає змогу виявляти осіб з високим ризиком ішемічної хвороби серця на ранніх стадіях захворювання.

Ключові слова: бінарний класифікатор, ROC-крива, діагностична ознака, аналіз та інтерпретація ЕКГ.

ВСТУП

Бінарні класифікатори широко застосовують у різних галузях. Так, наприклад, для профілактики захворювань потрібно проводити масові обстеження населення (скринінг) для виявлення пацієнтів з високим ризиком потенційно небезпечних хвороб [1]. У банківській сфері активно використовують системи скорингу [2], що забезпечують виявлення неблагонадійних позичальників. Перелік таких прикладів може бути продовжений.

Відомі різні підходи до оцінювання ефективності бінарних класифікаторів [3]. Зручним засобом є метод, який базується на аналізі так званої операційної характеристичної кривої (Receive Operating Characteristic curve, ROC) [4, 5]. Традиційний ROC-аналіз ґрунтується на оцінюванні двох характеристик діагностичного тесту: чутливості та специфічності [6–15], які, по суті, оцінюють імовірності помилок пропуску цілі та хибної тривоги, прийнятих у теорії статистичних рішень. Для інтегральної оцінки ефективності бінарного класифікатора, що реалізує діагностичне вирішувальне правило, найчастіше визначають площу під ROC-кривою [16].

У роботах [17–21] запропоновано розвинення ROC-аналізу для випадку, коли множина класів, що розпізнаються, містить більше двох діагнозів. Удосконалення методу ROC-аналізу для порівняння бінарних класифікаторів щодо очікуваних середніх втрат від неправильних рішень запропоновано в [22–24].

Мета цієї статті — подальше вдосконалення методу ROC-аналізу для забезпечення вибору діагностичного тесту, який гарантовано зменшує апіорний ризик помилкової діагностики.

ТРАДИЦІЙНИЙ ROC-АНАЛІЗ

Розглянемо потрібні у подальших дослідженнях основні деталі традиційного ROC-аналізу [4–15]. Для цього розглянемо задачу скринінгу, яка зводиться до виявлення хворих людей у великій групі обстежуваних, де є і хворі (клас V_1), і умовно здорові (клас V_2) особи, у яких немає аналізованого захворювання.

Нехай для виявлення хворих осіб передбачається використовувати бінарний класифікатор, який на підставі доступної інформації приймає одне з двох рішень: пацієнт «Хворий» (позитивний результат тестування) або «Здоровий» (негативний результат тестування). Для оцінювання ефективності такого тесту проводиться його тестування на репрезентативній групі пацієнтів із заздалегідь відомими діагнозами. Результати тестування можна навести у вигляді матриці (табл. 1), у клітинках якої зазначено кількість відповідних спостережень.

За даними табл. 1 легко оцінити прийняті у медичній діагностиці операційні характеристики тесту, а саме чутливість

$$S_E = \frac{TP}{TP + FN}, \quad (1)$$

яка визначає частку правильних позитивних результатів TP , отриманих для всіх хворих осіб $TP + FN$, та специфічність

$$S_P = \frac{TN}{TN + FP}, \quad (2)$$

яка визначає частку правильних негативних результатів TN , отриманих для всіх здорових осіб $TN + FP$.

Бінарний класифікатор з фіксованими операційними характеристиками (1) та (2) відображають точкою в ROC-просторі з координатами S_E та $1 - S_P$ [4]. Зауважимо, що величина

$$1 - S_P = 1 - \frac{TN}{TN + FP} = \frac{TN + FP - TN}{TN + FP} = \frac{FP}{TN + FP}$$

визначає частку помилкових позитивних результатів, отриманих для усіх здорових осіб $TN + FP$.

Розглядуваний ROC-простір надає змогу графічно продемонструвати діагностичну цінність тесту та порівнювати ефективність різних тестів [22].

Ідеальний тест A (рис. 1) розташований у точці з координатами $(0, 1)$. За результатом цього тесту всіх хворих осіб відносять до класу V_1 і не виникає помилок хибної тривоги. Звідси випливає, що чим ближчим до точки A є тест з операційними характеристиками S_E і S_P , тим він ефективніший.

Тест B (див. рис. 1), розташований у лівому нижньому кутку, називають «консервативним», оскільки за малого відсотка помилок хибної тривоги $1 - S_P$ він має

Таблиця 1. Результати тестування на репрезентативній групі пацієнтів

Поточний діагноз	Рішення, прийняте за результатом тестування	
	Пацієнт «Хворий»	Пацієнт «Здоровий»
«Хворий» (клас V_1)	Кількість правильних позитивних результатів TP	Кількість помилкових негативних результатів FN
«Здоровий» (клас V_2)	Кількість помилкових позитивних результатів FP	Кількість правильних негативних результатів TN

низьку чутливість S_E . Оскільки тест B має високе значення специфічності S_P , прийняття рішення: пацієнт «Хворий», слід сприймати з високим ступенем довіри. Водночас рішення: пацієнт «Здоровий», може бути неправильним через низьку чутливість S_E .

Тест C (див. рис. 1), розташований у правому верхньому кутку, називають «ліберальним», оскільки за великих значень чутливості S_E він має великий відсоток помилок хибної тривоги $1 - S_P$. Через високу чутливість S_E рішення: пацієнт «Здоровий», імовірно, правильне, тоді як рішення: пацієнт «Хворий», може бути помилковим через низьку специфічність S_P .

Тест D (див. рис. 1), розташований на діагоналі $S_E = 1 - S_P$ ROC-простору, є некорисним. Такі тести еквівалентні класифікаторам, які використовують стратегію «випадкового вгадування» класів V_1 та V_2 .

У медичній діагностиці часто послуговуються пороговими вирішувальними правилами вигляду

$$\begin{aligned} &\text{«Хворий»}, && \text{якщо } x > x_0, \\ &\text{«Здоровий»}, && \text{якщо } x \leq x_0, \end{aligned} \quad (3)$$

де x — діагностична ознака, а x_0 — порогове значення ознаки x . У такому випадку з'являється можливість визначити пари S_E та S_P для різних значень x_0 , які відповідають послідовності точок бінарного класифікатора у ROC-просторі.

З правила (3) випливає простий алгоритм побудови експериментальної ROC-кривої:

- відсортувати значення спостережуваної ознаки x у порядку зменшення;
- послідовно зменшувати (з деяким кроком) поріг x_0 та за екзамінаційною вибіркою спостережень обчислювати відповідні пари значень S_E та S_P ;
- відобразити отриману послідовність пар S_E та S_P у ROC-просторі (рис. 2).

Позначимо $p(x|V_1)$ і $p(x|V_2)$ умовні розподіли ймовірностей ознаки x у класах V_1 і V_2 . Множину $X_1 \triangleq \{x : p(x|V_1) \neq 0\}$ будемо називати носієм розподілу x для класу V_1 , а множину $X_2 \triangleq \{x : p(x|V_2) \neq 0\}$ — носієм розподілу x для класу V_2 .

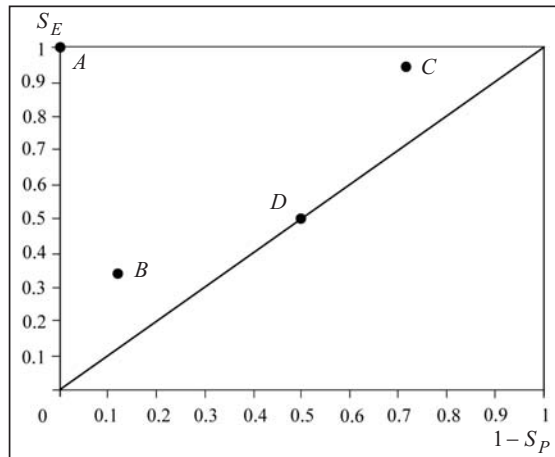


Рис. 1. Бінарні класифікатори в ROC-просторі

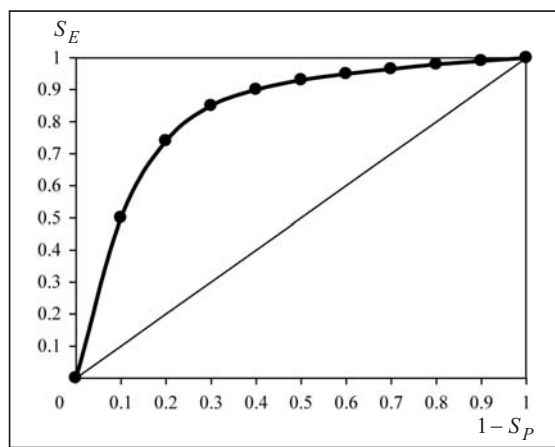


Рис. 2. ROC-крива тесту, що базується на вирішувальному правилі (3)

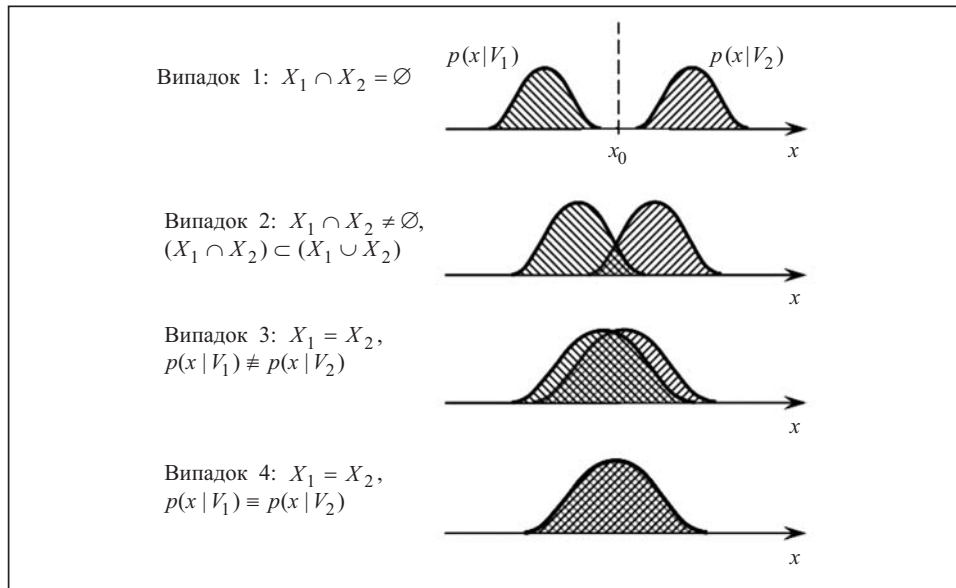


Рис. 3. Топологія носіїв X_1 і X_2 класів V_1 і V_2

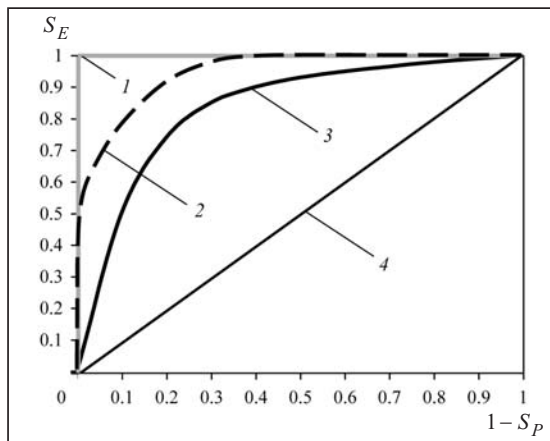


Рис. 4. Взаємозв'язок ROC-кривих з топологією носіїв класів V_1 і V_2

Можливі чотири випадки, які характеризують топологію носіїв X_1 , X_2 (рис. 3) і відповідний цим випадкам вигляд ROC-кривих (рис. 4). У випадку 1 множини X_1 і X_2 не перетинаються, і може бути встановлений поріг x_0 , який повністю розділяє значення ознаки x для класів V_1 та V_2 . Тоді правило (3) визначить ідеальний тест, що забезпечує безпомилкову класифікацію V_1 та V_2 . Графік відповідної ROC-кривої (див. рис. 4, крива 1) проходить від точки з координатами (0, 0)

у точку з координатами (0, 1) і далі в точку з координатами (1, 1).

У випадку 2 множини X_1 і X_2 частково перетинаються, і відповідною ROC-кривою є крива 2 (див. рис. 4).

У випадку 3, коли множини X_1 і X_2 збігаються, але умовні розподіли $p(x|V_1)$ і $p(x|V_2)$ різні, відповідною ROC-кривою є крива 3 (див. рис. 4). І насамкінець, у випадку 4, коли збігаються не тільки множини X_1 і X_2 , але й умовні розподіли $p(x|V_1)$ і $p(x|V_2)$, діагностична ознака x стає некорисною і відповідна ROC-крива лежить на діагоналі $S_E = 1 - S_P$ ROC-простору (див. рис. 4, крива 4).

Зауважимо, що дві окремо некорисні ознаки: x_1 і x_2 , які мають однакові одновимірні умовні розподіли в класах, тобто $p(x_i|V_1) \equiv p(x_i|V_2)$, $i=1,2$, разом можуть бути не тільки корисними, а й забезпечити безпомилкову класифікацію, коли $X_1^{(2)} \cap X_2^{(2)} = \emptyset$, де $X_m^{(2)} \triangleq \{(x_1, x_2) : p(x_1, x_2|V_m) \neq 0\}$ — носії двовимірних умовних розподілів $p(x_1, x_2|V_m)$, $m=1,2$ (рис. 5).

У [25] доведено, що такий цікавий випадок принципово можливий лише тоді, коли існує умовна статистична залежність між ознаками x_1 і x_2 одразу в обох класах, тобто

$$p(x_1, x_2 | V_m) \neq p(x_1 | V_m) p(x_2 | V_m), \quad m = 1, 2.$$

Для порівняння ефективності діагностичних тестів зручно використовувати інтегральну характеристику, якою зазвичай вважають площу AUC (Area Under Curve), обчислену під ROC-кривою [16]. Величина AUC визначає середню чутливість $\bar{S}_E$ тесту за можливих значень

специфічності $0 \leq S_P \leq 1$ або середню специфічність $\bar{S}_P$ тесту за можливих значень чутливості $0 \leq S_E \leq 1$, причому $AUC = 1$ для ідеального тесту та $AUC = 0.5$ — для некорисного. Тому вважається, що чим більше величина AUC наближається до одиниці, тим ефективніший тест.

Зрозуміло, що для операційних характеристик потенційно корисних тестів, які лежать вище діагоналі ROC-простору, виконується умова

$$S_E > 1 - S_P. \quad (4)$$

ROC-АНАЛІЗ З УРАХУВАННЯМ ТЕОРІЇ СТАТИСТИЧНИХ РІШЕНЬ

Розглянемо тепер деякі удосконалення традиційного ROC-аналізу, що дають змогу оцінювати ефективність діагностичного тесту щодо очікуваних втрат (середнього ризику).

Нехай апіорна ймовірність $P(V_1)$ появи хворої людини в групі обстежуваних осіб відома за даними медичної статистики і характеризує поширеність (преваленс) тієї чи іншої хвороби. Відповідно величина $P(V_2) = 1 - P(V_1)$ характеризує апіорну ймовірність наявності у групі обстежуваних осіб людини, яка не має досліджуваного захворювання.

Будемо характеризувати можливі помилки діагностичного тесту величинами L_{12} , L_{21} , які визначають втрати від помилки пропуску цілі (не виявлено «Хворого» пацієнта з класу V_1) та хибної тривоги («Здорового» пацієнта відносять до класу хворих V_1).

Тоді очікувані втрати від помилкової діагностики визначає співвідношення

$$R = P(V_1)(1 - S_E)L_{12} + [1 - P(V_1)](1 - S_P)L_{21}. \quad (5)$$

Умова (5) надає змогу за відомими $P(V_1)$, L_{12} і L_{21} обчислити середній ризик помилкової діагностики (апостеріорний ризик) для тесту, який має операційні характеристики S_E і S_P . Виникає природне запитання: чи завжди такий тест буде корисним?

Щоб відповісти на це запитання, передусім потрібно сформулювати саме означення корисності діагностичного тесту. На погляд авторів цієї статті, таким означенням може бути умова, що тест забезпечує менші середні втрати, ніж ті,

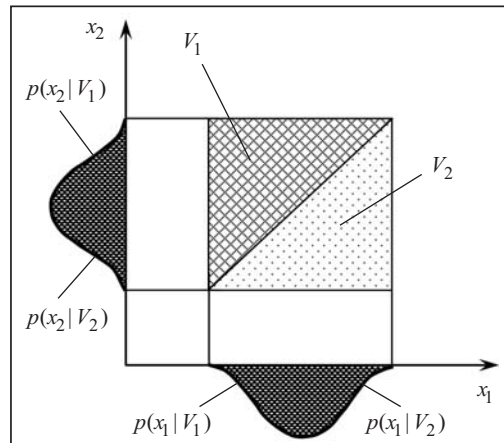


Рис. 5. Безпомилкова класифікація V_1 і V_2 за двома окремо некорисними ознаками

які спостерігалися би без використання діагностичного тесту [26]. Інакше кажучи, поняття корисності логічно сформулювати у такий спосіб.

Означення 1. Діагностичний тест є корисним, якщо виконується строга нерівність

$$R < R_0, \quad (6)$$

тобто апостеріорний ризик R менший за апріорний ризик R_0 , що реалізує стратегія прийняття рішень лише на підставі відомих значень $P(V_1)$, L_{12} і L_{21} .

Наведемо деякі пояснення. Зрозуміло, якщо бінарний класифікатор реалізує баєсівську стратегію прийняття рішень, яка мінімізує апостеріорний ризик, то виконується або умова (6), або умова $R = R_0$. Інакше кажучи, формально побудований алгоритм діагностики не може бути «шкідливим». Однак на практиці часто відсутня інформація, потрібна для реалізації стратегії Баєса, і діагностичний тест будується не формальним шляхом.

У таких важливих для практичного застосування випадках за фіксованих $P(V_1)$, L_{12} , L_{21} і відомих S_E , S_P замість умови (6) може виконуватися на перший погляд «парадоксальна» умова $R > R_0$.

Проілюструємо таку можливість на модельному прикладі. Нехай для виявлення «Хворих» пацієнтів у групі з 10000 осіб запропоновано використовувати діагностичний тест із достатньо високими операційними характеристиками $S_E = 0.9$ і $S_P = 0.9$. Нехай преваленс захворювання $P(V_1) = 0.02$, а втрати від помилки пропуску цілі та хибної тривоги вважаються однаковими: $L_{12} = L_{21} = 1$.

На рис. 6 наведено очікувані результати тестування. Якщо не використовувати діагностичного тесту, то оптимальна стратегія має бути такою: всіх обстежуваних визнавати «Здоровими» пацієнтами (інакше втрати R_0 будуть більшими!!!). Тоді середні апріорні втрати на одного обстежуваного, зумовлені тим, що 200 «Хворих» пацієнтів не будуть виявлені, складуть

$$R_0 = (L_{12} \cdot 200) / 10000 = (1 \cdot 200) / 10000 = 0.02. \quad (7)$$

У разі використання діагностичного тесту з чутливістю $S_E = 0.9$ та специфічністю $S_P = 0.9$ середні втрати, зумовлені помилково позитивними результатами діагностики 980 «Здорових» пацієнтів та невиявленням 20 «Хворих» пацієнтів, становитимуть

$$R = (L_{21} \cdot 980 + L_{12} \cdot 20) / 10000 = (1 \cdot 980 + 1 \cdot 20) / 10000 = 0.1. \quad (8)$$

Зі співвідношень (7) і (8) випливає, що $R > R_0$ і такий тест не можна визнати корисним, оскільки за заданих втрат від помилок пропуску цілі та хибної тривоги він лише збільшує очікувані втрати!

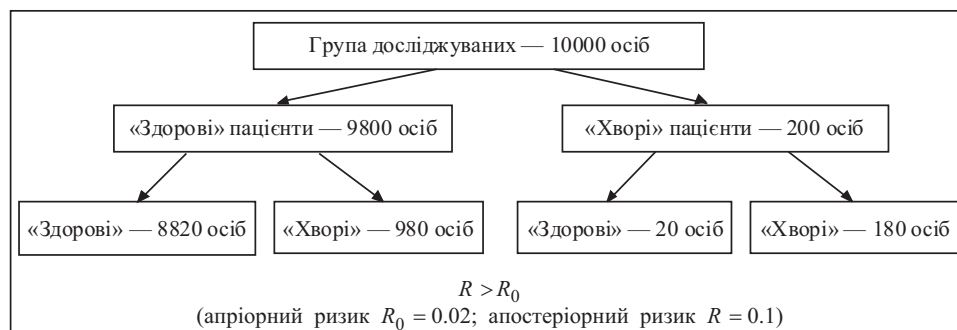


Рис. 6. Очікувані результати тестування

Незважаючи на «парадоксальність» того, що $R > R_0$, отриманий результат цілком обґрунтований. Неважко показати, що за розглянутими значеннями $P(V_1)$, L_{12} , L_{21} , S_E і S_P рішення: пацієнт «Хворий», не є обґрунтованим для мінімізації середнього ризику. Інакше кажучи, для кожного обстежуваного завжди маємо приймати рішення: пацієнт «Здоровий», яке збігається з апіорними рішеннями.

Покажемо, що (4) є лише необхідною, але не достатньою умовою, яка гарантує виконання строгої нерівності (6).

Твердження 1. Діагностичний тест гарантовано корисний у сенсі означення 1, якщо існує залежність між його чутливістю S_E і специфічністю S_P , виражена співвідношеннями

$$S_E > m(1 - S_P), \text{ якщо } m \geq 1, \quad (9)$$

або

$$S_E > 1 - m + m(1 - S_P), \text{ якщо } m < 1, \quad (10)$$

де

$$m = \frac{[1 - P(V_1)]}{\omega P(V_1)} \quad (11)$$

є величиною, яка залежить від апіорної ймовірності $P(V_1)$ і співвідношення втрат від помилок пропуску цілі та хибної тривоги

$$\omega = \frac{L_{12}}{L_{21}}. \quad (12)$$

Доведення. Якщо не використовувати діагностичного тесту і приймати рішення лише на підставі значень $P(V_1)$, L_{12} та L_{21} , то апіорна стратегія зводиться до одного з двох можливих варіантів:

— всіх обстежуваних відносити до класу «Хворих» V_1 , і тоді апіорний ризик має вигляд

$$R_0^+ = L_{21}[1 - P(V_1)]; \quad (13)$$

— приймати рішення про те, що кожен обстежуваний «Здоровий», і тоді апіорний ризик має вигляд

$$R_0^- = L_{12}P(V_1). \quad (14)$$

Із (13), (14) з урахуванням (11) випливає, що мінімальний апіорний ризик, який фігурує у правій частині нерівності (6), можна записати так:

$$R_0 = \begin{cases} L_{21}[1 - P(V_1)], & \text{якщо } m < 1, \\ L_{12}P(V_1), & \text{якщо } m \geq 1. \end{cases} \quad (15)$$

Унаслідок підстановки виразів (5) та (15) у нерівність (6) після елементарних перетворень отримуємо співвідношення (9) та (10). Твердження доведено.

Як видно з виразу (11), для перевірки умов (9) і (10) достатньо мати лише інформацію про співвідношення втрат (12), і така інформація більш доступна під час розв'язування прикладних задач, ніж інформація про абсолютні значення втрат від помилки пропуску цілі L_{12} та хибної тривоги L_{21} .

Умови (9) та (10) надають змогу обґрунтовано оцінювати корисність нових діагностичних тестів. Для цього доцільно удосконалити традиційний ROC-аналіз, обмеживши відповідний фрагмент ROC-кривої (рис. 7).

Пряма 1 (рис. 7, а) відповідає випадку, коли $P(V_1) = 0.15$ і $\omega = L_{12}L_{21}^{-1} = 1$. Отже, тест некорисний, оскільки пряма 1 не перетинає ROC-кривої. Але якщо припустити, що втрати від помилки пропуску цілі вчетверо вищі втрат від хибної тривоги, тобто $L_{12}L_{21}^{-1} = 4$, то відповідна пряма 2 вже перетинає ROC-криву. Та-

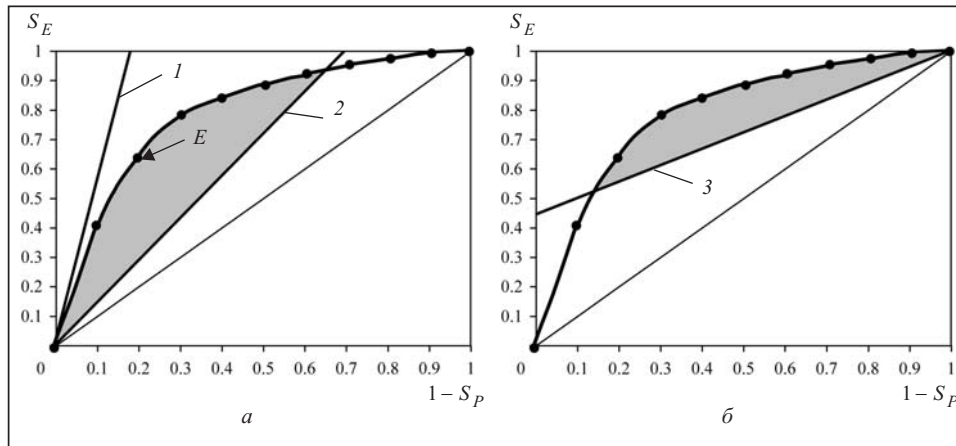


Рис. 7. Ілюстрація ідеї удосконаленого ROC-аналізу для випадків $m \geq 1$ (а) та $m < 1$ (б)

ким чином, тест, що має чутливість $S_E = 62.5\%$ і специфічність $S_P = 80\%$, виявиться гарантовано корисним згідно з (6), оскільки за таких умов точка E , яка відповідає наведеним операційним характеристикам, розташована на допустимій частині ROC-кривої. На рис 7, б, коли $m < 1$, обмежувальна пряма 3 починається зі значень $S_P = 0$, $S_E = 1$. І якщо вона перетинає ROC-криву, то на відповідному її фрагменті існують значення S_P і S_E , за яких тест буде корисним.

Зауважимо, що для $m = 1$ достатні умови (9) і (10) збігаються з необхідною умовою (5). У цьому випадку пряма, що обмежує фрагмент ROC-кривої гарантовано корисного тесту, лежить на діагоналі ROC-простору.

Цікаво порівняти отримані умови з результатами, наведеними в [22], де розглядаються так звані ізолінії ефективності (iso-performance lines). З виразу (5) випливає, що рівняння ізолінії з очікуваними втратами $R = R^*$ має вигляд лінійної залежності в координатах S_E , $1 - S_P$:

$$S_E = m(1 - S_P) + 1 - \frac{R^*}{L_{12}P(V_1)}. \quad (16)$$

Рівняння (16) збігається з рівняннями прямих (9) і (10) за виконання умови

$$R^* = \begin{cases} L_{21}P(V_1), & \text{якщо } m \geq 1, \\ L_{12}[1 - P(V_1)], & \text{якщо } m < 1, \end{cases}$$

права частина якої збігається з правою частиною (15). Тобто ділянку корисних тестів обмежує ізолінія ефективності

$$S_E = m(1 - S_P) + 1 - \frac{R_0}{L_{12}P(V_1)}, \quad (17)$$

яка відповідає рівню апіорного ризику R_0 та обов'язково проходить через точку з координатами $(0, 0)$, якщо $m \geq 1$, або точку з координатами $(1, 1)$, якщо $m < 1$.

Зрозуміло, що можливі такі випадки:

- ізолінія (17) перетинає ROC-криву і тоді $R < R_0$;
- ізолінія (17) дотична до ROC-кривої і тоді $R = R_0$;
- ізолінія (17) не перетинає ROC-криву і тоді $R > R_0$.

Умови (9) і (10) надають змогу визначити допустимий інтервал співвідношення втрат $\omega = L_{12}L_{21}^{-1}$, за яким тест із чутливістю S_E та специфічністю S_P залишається корисним для скринінгу захворювання з преваленсом $P(V_1)$. Такий інтервал визначається за формулою

$$\frac{1 - P(V_1)}{P(V_1)} \frac{1 - S_P}{S_E} \leq \omega \leq \frac{1 - P(V_1)}{P(V_1)} \frac{S_P}{1 - S_E}. \quad (18)$$

ПРАКТИЧНІ РЕЗУЛЬТАТИ

Отримані результати дали змогу обґрунтувати ефективність інноваційного в кардіології методу аналізу та інтерпретації електрокардіограм — методу фазаграфії [27]. Метод ґрунтується на переході від скалярного сигналу $x(t)$, який містить інформацію про електричну активність серця, до траєкторії на фазовій площині з координатами $x(t)$, $\dot{x}(t)$, де $\dot{x}(t)$ — швидкість зміни сигналу [28].

Завдяки такому переходу вдалося визначити нову діагностичну ознаку β_T — симетрію ділянки реполяризації електрокардіограми (ЕКГ). Для перевірки діагностичної цінності β_T проводилося статистичне оброблення клінічних даних: 441 запис ЕКГ попередньо верифікованих хворих на ішемічну хворобу серця (ІХС) та 387 записів ЕКГ здорових осіб із контрольної групи.

Результат оброблення даних показав, що середні значення β_T мають статистично значущі відмінності у групах ($p < 0.01$) і становлять 0.956 ± 0.43 і 0.665 ± 0.12 у групі ІХС і контрольній групі відповідно.

Для оцінювання корисності діагностичного правила

$$\begin{aligned} &\text{ІХС, якщо } \beta_T > \beta_0, \\ &\text{Норма, якщо } \beta_T \leq \beta_0, \end{aligned} \quad (19)$$

що базується на порівнянні β_T з пороговим значенням β_0 , використовувався розроблений програмний застосунок, в якому реалізовано удосконалений ROC-аналіз. На рис. 8 наведено експериментальну ROC-криву. Сірим кольором виокремлено область значень S_E і S_P , які за умов $P(V_1) = 0.1$ та $\omega = 5$ задовольняють співвідношення (9), (10).

Встановлено, що прийняття рішень за правилом (19) забезпечує чутливість $S_E = 81\%$ і специфічність $S_P = 78\%$, якщо $\beta_0 \approx 0.72$ — оптимальне значення порога, якому відповідає точка, найбільш віддалена від діагоналі.

Відповідно до (18), якщо $S_E = 81\%$ і $S_P = 78\%$, діагностичне правило (19) гарантовано корисне для скринінгу ІХС з преваленсом $P(V_1) = 0.1$ у достатньо широкому діапазоні значень співвідношення втрат: $2.44 \leq \omega \leq 37$.

Зауважимо, що в групу ІХС було включено лише пацієнтів, у яких традиційний ЕКГ аналіз у 12 відведеннях не виявив жодних відхилень від норми. Водночас запропоноване діагностичне правило (19) підтвердило достатньо високі показники на такому «складному» клінічному матеріалі, і тому вважатимемо його цілком прийнятним для скринінгу ІХС.

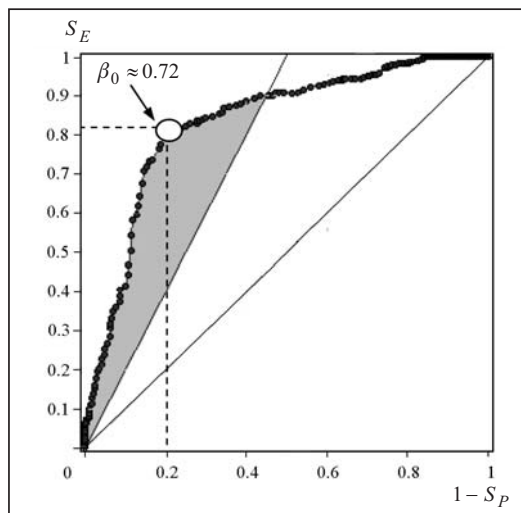


Рис. 8. Результат оцінювання корисності діагностичного правила (19)

ВИСНОВКИ

Доведено, що умови (9), (10) гарантують корисність діагностичного тесту у сенсі зменшення апіорного ризику. Отримані умови надали змогу вдосконалити традиційний ROC-аналіз завдяки обмеженню відповідного фрагмента ROC-кривої. Показано, що пряма, яка обмежує фрагмент ROC-кривої гарантовано корисного тесту, збігається з ізолінією ефективності, що відповідає рівню апіорного ризику. Визначено допустимі межі співвідношення втрат від помилки пропуску цілі та хибної тривоги, за якими тест із чутливістю S_E та специфічністю S_P залишається корисним для скринінгу захворювання з відомим преваленсом $P(V_1)$.

Результати дослідження мають практичне застосування для підтвердження ефективності інноваційного у кардіології методу фазаграфії, який надає змогу під час проведення масових профілактичних обстежень виявляти осіб з прихованими початковими ознаками ішемічної хвороби серця.

СПИСОК ЛІТЕРАТУРИ

1. Maxim D., Niebo R., Utel M.J. Screening tests: A review with examples. *Inhalation Toxicology*. 2014. Vol. 26. Iss. 13. P. 811–828. <http://doi.org/10.3109/08958378.2014.955932>.
2. Zhukovska O. Decision-making model on potential borrower lending for in-dependent experts group. *Proc. IEEE 3rd Intern. Conf. on System Analysis & Intelligent Computing (SAIC)* (4–7 Oct. 2022, Kyiv, Ukraine). P. 118–121. <http://doi.org/10.1109/SAIC57818.2022.9923015>.
3. Dendek C., Mandziuk J. Improving performance of a binary classifier by training set selection. *Proc. of the 18th International Conference on Artificial Neural Networks* (September 3–6, 2008, Prague Czech Republic). Part I. 2008. P. 1–8. http://doi.org/10.1007/978-3-540-87536-9_14.
4. Metz C.E. Fundamental ROC analysis. In: *Handbook of Medical Imaging*. R.L. Van Metter; Beutel J., Kundel H.L. (Eds). Vol. 1. Physics and Psychophysics. Bellingham: SPIE Press, 2000. P. 754–769. <https://doi.org/10.1117/3.832716.ch15>.
5. Fawcett T. Using rule sets to maximize ROC performance. *Proc. of the IEEE International Conference on Data Mining (ICDM-2001)*. Los Alamitos, CA: IEEE Computer Society, 2001. P. 131–138. <https://doi.org/10.1109/ICDM.2001.989510>.
6. Flach P., Wu S. Repairing concavities in ROC curves. *Proc. of the 2003 UK Workshop on Computational Intelligence*. University of Bristol, 2003. P. 38–44.
7. Sonogo P., Kocsor A., Pongor S. ROC analysis: Applications to the classification of biological sequences and 3D structures. *Briefings in Bioinformatics*. 2008. Vol. 9, Iss. 3. P. 198–209. <https://doi.org/10.1093/bib/bbm064>.
8. Feng K., Hong H., Nang K., Wang J. Decision making with machine learning and ROC curves. *arXiv:1905.02810v1 [stat.ME]* 5 May 2019. <https://doi.org/10.48550/arXiv.1905.02810>.
9. Davis J., Goadrich M. The relationship between Precision-Recall and ROC curves. *Proc. of the 23rd International Conference on Machine Learning (ICML'06)*. 2006. P. 233–240. <https://doi.org/10.1145/1143844.1143874>.
10. Van den Hout W. The area under an ROC curve with limited information. *Medical Decision Making*. 2003. Vol. 23, Iss. 2. P. 160–166. <https://doi.org/10.1177/0272989X03251246>.
11. Pepe M.S., Longton G., Janes H. Estimation and comparison of receiver operating characteristic curves. *The Stata Journal*. 2009. Vol. 9, N 1. P. 1–16. PMID: 20161343.
12. Alonzo T.A., Pepe M.S. Distribution-free ROC analysis using binary regression techniques. *Biostatistics*. 2002. N 3. P. 421–432. <https://doi.org/10.1093/biostatistics/3.3.421>.
13. Provost F., Fawcett T. Robust classification for imprecise environments. *Machine Learning*. 2001. Vol. 42, N 3. P. 203–231. <https://doi.org/10.1023/A:1007601015854>.
14. Majnik M., Bosnić Z. ROC analysis of classifiers in machine learning: A survey. *Intelligent Data Analysis*. 2013. Vol. 17, N 3. P. 531–558. <https://doi.org/10.3233/IDA-130592>.
15. Hu N. Using receiver operating characteristic (ROC) analysis to evaluate information-based decision-making. In: *Advanced Methodologies and Technologies in Business Operations and Management*. 2019. P. 764–776. <https://doi.org/10.4018/978-1-5225-7362-3.ch057>.

16. Hand D. J., Till R. J. A simple generalization of the area under the ROC curve to multiple class classification problems. *Machine Learning*. 2001. Vol. 45, N 2. P. 171–186. <https://doi.org/10.1023/A:1010920819831>.
17. Ferri C., Hernández-Orallo J., Salido M.A. Volume under the ROC surface for multi-class problems. In: Machine Learning: ECML 2003. ECML 2003. Lavrac N., Gamberger D., Blockeel H., Todorovski L. (Eds.). *Lecture Notes in Computer Science*. Vol. 2837. Berlin; Heidelberg: Springer, 2003. P. 108–120. https://doi.org/10.1007/978-3-540-39857-8_12.
18. Edwards D.C, Metz C.E, Kupinski M.A. Ideal observers and optimal ROC hypersurfaces in N-class classification. *IEEE Trans. Med. Imaging*. 2004. Vol. 23, N 7. P. 891–895. <https://doi.org/10.1109/TMI.2004.828358>.
19. Edwards D.C, Metz C.E, Nishikawa R.M. The hypervolume under the ROC hypersurface of “Near-Guessing” and “Near-Perfect” observers in N-class classification tasks. *IEEE Trans. Med. Imaging*. 2005. Vol. 24, N 3. P. 293–299. <https://doi.org/10.1109/tmi.2004.841227>.
20. He X, Fry E.C. An optimal three-class linear observer derived from decision theory. *IEEE Trans. Med. Imaging*. 2007. Vol. 26, N 1. P. 77–83. <https://doi.org/10.1109/TMI.2006.885335>.
21. Sahiner B, Chan H-P, Hadjiiski L.M. Performance analysis of 3-class classifiers: Properties of the 3D ROC surface and the normalized volume under the surface. *IEEE Trans. Med. Imaging*. 2008. Vol. 27, Iss. 2. P. 215–227. <https://doi.org/10.1109/TMI.2007.905822>.
22. Fawcett T. ROC graphs with instance-varying costs. *Pattern Recognition Letters*. 2006. Vol. 27, Iss. 8. P. 882–891. <https://doi.org/10.1016/j.patrec.2005.10.012>.
23. Meekins R., Adams S, Beling P.A., Farinholt K., Hipwell N., Chaudhry A., Polter S., Dong Q. Cost-sensitive classifier selection when there is additional cost information. *Proceedings of Machine Learning Research*. 2018. Vol. 88. P. 17–30. PMLR 88:17-30.
24. Holte R.C., Drummond C. Cost-sensitive classifier evaluation using cost curves. In: Advances in Knowledge Discovery and Data Mining. PAKDD 2008. *Lecture Notes in Computer Science*. Berlin; Heidelberg: Springer, 2008. Vol. 5012. P. 26–29. https://doi.org/10.1007/978-3-540-68125-0_4.
25. Fainzilberg L.S. Plausible but groundless premises when constructing diagnostic models. *Journal of Automation and Information Sciences*. 2020. Vol. 52, Iss. 5. P. 38–50. <https://doi.org/10.1615/JAutomatInfScien.v52.i5.40>.
26. Fainzilberg L.S. Conditions of utility of diagnostic tests from the point of view of the statistical theory of decision making. *Journal of Automation and Information Sciences*. 2003. Vol. 35, Iss. 4. P. 63–73. <https://doi.org/10.1615/JAutomatInfScien.v35.i4.30>.
27. Fainzilberg L.S. New opportunities of phasegraphy in medical practice. *Science and Innovation*. 2017. Vol. 1, Iss. 3. P. 37–50. <https://doi.org/10.15407/scine13.03.037>.
28. Fainzilberg L.S. New approaches to the analysis and interpretation of the shape of cyclic signals. *Cybernetics and Systems Analysis*. 2020. Vol. 56, N 4. P. 665–674. <https://doi.org/10.1007/s10559-020-00283-0>.

O.A. Zhukovska, L.S. Fainzilberg

EVALUATION OF THE USEFULNESS OF BINARY CLASSIFIER BASED ON ENHANCED ROC-ANALYSIS

Abstract. The definition of the usefulness of the binary classifier from the point of view of reducing the a priori risk of false classification is formulated. Sufficient conditions are proposed to guarantee the utility of a diagnostic test according to this definition. The obtained conditions improved the traditional ROC analysis by limiting the corresponding region of the ROC curve. The line limiting the region of the guaranteed useful test is shown to coincide with the known iso-performance line corresponding to the a priori risk level. Permissible limits of the ratio of losses from target misses and false alarms were determined, according to which a test with appropriate operational characteristics remains useful for screening a disease with a known prevalence. Based on the obtained results, the authors substantiated the effectiveness of the new method of the analysis and interpretation of electrocardiograms, which is based on determining the original diagnostic feature in the phase space and enables detecting persons with a high risk of coronary heart disease in the early stages of the disease.

Keywords: binary classifier, ROC curve, diagnostic feature, analysis and interpretation of ECG.

Надійшла до редакції 03.01.2023