ISSN 2079-0031 (Print) ISSN 2411-0558 (Online)

УДК 004.8 **DOI: 10.20998/2411-0558.2023.01.10**

Т. О. БІЛОБОРОДОВА, канд. техн. наук, доц., ІПМЕ, м. Київ,
І. С. СКАРГА-БАНДУРОВА, докт. техн. наук, проф., ТНТУ, м. Тернопіль

МЕТОДОЛОГІЯ ПРИЙНЯТТЯ РІШЕНЬ НА ОСНОВІ СУКУПНОЇ ДОСТОВІРНОСТІ ТА ІНТЕПРЕТОВАНОСТІ РЕКОМЕНДАЦІЙ ШТУЧНОГО ІНТЕЛЕКТУ

Стаття розглядає проблему переходу в медичній діагностиці від клінічно-залежних методологій до підходів, базованих на доказах, з використанням штучного інтелекту (ШІ). Основна мета дослідження полягає у розробці методології прийняття рішення на основі об'єднаного рішення людини та ШІ, та інтерпретованості результату ШІ для людини. Запропонована методологія включає формування рішень на основі людського інтелекту (ЛІ) та ШІ, оцінку корисності рекомендацій, і формування спільного рішення на основі сукупної ймовірності. Практичне застосування методології було демонстровано на прикладі експерименту з класифікації немедичних зображень. Експеримент показав, що спільне рішення, засноване на об'єднаному підході, досягає більшої достовірності (0,77) порівняно з окремими підходами ЛІ (0,73) та ШІ (0,82). Результати дослідження підкреслюють важливість прозорості, інтерпретованості та довіри до результатів ШІ для успішного використання ШІ в медицині. Іл.: 1. Бібліогр.: 16 назв.

Ключові слова: прийняття рішень, штучний інтелект, достовірність, інтепретованість

Постановка проблеми. У сфері медичної діагностики спостерігається перехід від традиційних клінічно-залежних методологій обстеження до підходів, що переважно базуються на доказах, коли лікарі значною мірою покладаються на об’єктивні дані та стандартизовані критерії діагностики, щоб підвищити точність і послідовність діагностики. У цьому контексті залучення штучного інтелекту (ШІ) є цінним інструментом для аналізу медичних доказів, надаючи суттєву підтримку медичним працівникам. Рішення на основі ШІ сприяють швидкому наданню медичних послуг за рахунок прискорення часу діагностики та оптимізації розподілу ресурсів у системах охорони здоров’я. Теледерматологія є одним із прикладів демонстрації потенціалу впровадження ШІ. Дослідження [1, 2] показали, що модель згорткової нейронної мережі перевершила діагностичну точність більшості з 58 дерматологів у ідентифікації меланоми. Цей успіх підкреслює можливості подібного прогресу в інших областях автоматизованої діагностики, таких як скринінг раку молочної залози та раку шийки матки [3]. Однак прийнятність рішень на основі ШІ викликає певні занепокоєння.

ISSN 2079-0031 (Print) ISSN 2411-0558 (Online)

Покладаючись виключно на можливості моделей штучного інтелекту, передбачається, що людські судження є достатньо надійними, щоб компенсувати будь-які неточності в результатах, створених ШІ. Тим не менш, цей підхід не бере до уваги людські помилки та можливу неузгодженість, особливо при роботі з великими обсягами даних. Більшість методів спільного прийняття рішень людиною та штучним інтелектом в першу чергу враховують достовірність рішень штучного інтелекту, недооцінюючи значення достовірності рішення людини в прийнятті рішень у остаточному процесі прийняття рішень. Таким чином, для забезпечення ефективного та точного прийняття рішень стає обов'язковим враховувати впевненість рішень як ШІ, так і людського інтелекту (ЛІ).

Аналіз останніх досліджень і публікацій. Достовірність рішень ЛІ разом із існуючими упередженнями навколо них є темою, яка широко вивчається в різних областях досліджень, включаючи прийняття рішень [4]. Автори [5] досліджують процес формування рішення людиною та пропонують двокомпонентний простір моделі поведінки людини, включаючи компонент корисності та компонент вибору, для визначення яким чином впливає рекомендація ШІ на прийняття рішення людиною в залежності від довіри людини до результатів ШІ. Результати показують, що моделі, скориговані людиною, перевершують вихідні дані моделі штучного інтелекту, що свідчить про те, що люди схильні використовувати власне судження під час процесу прийняття рішень, щоб оцінити, чи слід приймати рекомендації штучного інтелекту. Порівняння параметрів моделі показує, що коли ставки рішень стають більшими, люди, як правило, знижують свою віру в правильність рекомендацій ШІ та більше покладаються на власні судження в ухваленні рішень за допомогою ШІ.

Дослідження [6] було зосереджено на співпраці людини та машини для прийняття різноманітних рішень. Автори запропонували рішення, спрямоване на використання взаємодоповнюваності між людиною та машиною для максимізації винагороди за рішення, яке перевершує як алгоритм, так і людину, коли вони приймають рішення самостійно.

Специфічний тип взаємодії людини та штучного інтелекту, коли людина-наглядач вирішує або прийняти рекомендацію штучного інтелекту, або самостійно вирішити завдання, вивчається в дослідженні [7]. Автори максимізують очікувану корисність команди людини та штучного інтелекту, виражену в термінах якості остаточного рішення, вартості перевірки та індивідуальної точності людини та машин. Дослідження показує переваги моделювання командної роботи під час навчання через покращення очікуваної командної користі для наборів даних, враховуючи такі параметри, як людські навички та вартість помилок.

Достовірність, визначена як суб'єктивна ймовірність того, що рішення буде правильним [8, 9], забезпечує засіб для оцінки впевненості людського рішення. Одним із підходів до оцінки ступеня достовірності в прийнятих рішенням людини є калібрування, яке перевіряє узгодженість між рівнями достовірностями окремих людей і фактичною точністю рішень [10, 11]. Крім того, клінічна діагностика дотримується принципу не завдавати шкоди, що вимагає зусиль для забезпечення інтерпретації та прозорості клінічних рішень на основі ШІ. Успішне впровадження рішень на основі ШІ залежить від встановлення довіри між медичними експертами та системами ШІ. Якщо пояснення, надані ШІ, не відповідають очікуванням лікарів, вони навряд чи будуть прийняті, що підкреслює важливість ефективних механізмів комунікації та пояснення в результатах ШІ. Ці факти обумовлюють актуальність поставленої мети та вибір методів її реалізації.

Мета статті. Метою дослідження є розробка методології прийняття рішення на основі об'єднаного рішення ЛІ та ШІ та інтепретованості результату ШІ для людини.

Формалізація параметрів прийняття рішення на основі рекомендацій ШІ. Формально, параметри прийняття рішень за допомогою ШІ можуть бути визначені наступним чином. Визначимо множину випробувань з прийняття рішень у вигляді n-вимірного вектору ознак x, таке, що $x \in X$. Визначимо бінарне рішення, яке потрібно прийняти в цьому випробуванні як $y \in \{+1, -1\}$. Результати моделі ШІ $m(x)$ під час прийняття рішень є розподілом ймовірностей набору можливих рішень, тобто $m(x) = \{+1\colon P(y = +1|x, -1\colon P(y = -1|x)\}$. Враховуючи $m(x)$, модель ШІ надає людині рекомендацію щодо прийняття рішення, яка складається з двох частин – рекомендованого рішення $\hat{y}^m = \arg\max m(x)$, і достовірності рекомендованого рішення $c^m = \max m(x) = m(x)[y = \widehat{=y}^m]$. Припустимо, що достовірність моделі ШІ відкалібрована, тобто $c^m = P(y = \hat{y}^m)$. Також, припустимо, що людина, яка приймає рішення, також формує свою власну пропозицію щодо рішення – $h(x)$, що визначає результат ЛІ в процесі прийняття рішення, та, також, складається з рішення людини $\hat{y}^h = \arg\max h(x)$, і достовірності рішення людини $c^h = \max h(x) = h(x)[y = \hat{y}^h]$.

Множина випробувань включає дані з прийняття рішення ЛІ на основі послідовності T випробувань прийняття рішень за допомогою моделі ШІ. У кожному випробуванні $t(1 \le t \le T)$ людині надається вектор ознак x^t разом із результатом та рекомендацією моделі ШІ $\hat{y}^{m,t}$ і достовірністю $c^{m,t}$. Людина також формує власне судження $h(x^t)$ щодо вхідного зразка. Маючи всю цю інформацію, людина повинна прийняти рішення $\hat{y}^t$ виконавши дію $d^t \in \{accept, reject\}$, щоб прийняти рекомендацію моделі ШІ або відхилити її.

Таким чином, мета прийняття остаточного рішення ШІ полягає у формалізації процесу прийняття рішення людиною щодо прийняття чи відхилення рекомендації ШІ, тобто, визначення d^t для кожного випробування t.

Методологія прийняття рішення на основі об‘єднаного рішення ЛІ та ШІ та інтепретованості результату ШІ для людини. Запропонована методологія включає наступні етапи: (1) формування рішень ЛІ та ШІ, (2) оцінка сукупної корисності прийняття/відхилення рекомендації, (3) формування спільного рішення на основі сукупної ймовірності прийняття/відхилення рекомендації.

На першому етапі, при формуванні сукупної достовірності відносно кожного рішення, враховується результат передбачення ЛІ та моделі ШІ, а також достовірність цього передбачення. Далі розраховується сукупна корисність кожної дії яка використовується на останньому етапі за уваги довіри людини до рекомендації ШІ для формування остаточного рішення щодо дії на основі рекомендації ШІ: прийняти або відхилити рекомендацію ШІ, та, також, розраховується ймовірність цієї дії.

Оцінка сукупної достовірності рішень ЛІ та ШІ. Формування рішень відбувається в залежності від поставленого завдання. Результатом формування рішень є передбачення та достовірність передбачення, тобто впевненість у правильності цього передбачення, для кожного зразку випробування від кожного представника ЛІ та ШІ незалежно.

Достовірність ЛІ розраховується з використанням моделі на основі реальних даних від групи експертів. Для розрахунку достовірності ЛІ для низки випробувань, декілька експертів повинні прийняти рішення чи сформулювати заключення на основі низки випробувань. В цьому випадку частка точних рішень визначається як достовірність людини c^h наступним чином (1).

$$c^h = \frac{1}{|\Omega_H|}\sum_{i \epsilon \Omega_H} 1(h_i = y), \qquad (1)$$

де Ω_H група експертів, h_i – результат, пропонований експертом-людиною, y – дійсний результат, i – ідентифікатор експерта.

ISSN 2079-0031 (Print) ISSN 2411-0558 (Online)

Враховуючи достовірності ЛІ та ШІ, об‘єднане рішення формується на основі сукупної достовірності $c^{m+h,t}$ для подальшого прийняття або відхилення рекомендації моделі ШІ.

Сукупна достовірність ЛІ та ШІ розраховується на основі критерію середньозважених логарифмічних шансів, що представляє собою комбінацію усереднення та Байєсова правила [12], відповідно до якого логарифмічні шанси агрегованої ймовірності представлені як середнє значення логарифмічних шансів окремої оцінки наступним чином (2):

$$c^{m+h,t} = \frac{\exp(\alpha)}{1+\exp(\alpha)}, \tag{2}$$

де $\alpha = \frac{1}{2}\left(\ln\frac{c^{m,t}}{1-c^{m,t}} + \ln\frac{c^{h,t}}{1-c^{h,t}}\right)$. Оскільки логарифмічне значення шансів ймовірності підкреслює відмінності екстремальних ймовірностей (тобто ймовірностей, близьких до 0 або 1), чистий ефект цього правила, таким чином, полягає в тому, щоб призвести екстремальні ймовірності до середнього значення.

Враховуючи вхідний зразок $x \in X$, функцію передбачення моделі $m(x)$ і функцію передбачення ЛІ $h(x)$, метою схеми прийняття рішень є оптимізація правила прийняття рішення на цьому етапі $g: X \to \{0,1\}$, яка визначає рішення на цьому етапі на основі сукупної достовірності $c^{m+h,t}$ наступним чином (3):

$$(m,h,g)(x) \triangleq \begin{cases} h(x) \text{ якщо } g(x) = 1, \\ m(x) \text{ якщо } g(x) = 0. \end{cases} \tag{3}$$

Таким чином, враховуючи оцінку сукупної достовірності $c^{m+h,t}$ формулюємо $g(c^t)$ наступним чином (4):

$$g\big(m(x), h(x)\big) = \begin{cases} 1, \text{якщо } c^t \geq c^{m+h,t}, \\ 0, \text{в іншому випадку}. \end{cases} \tag{4}$$

Шляхом оптимізації порогового значення ε пропонується спільне рішення між ЛІ та ШІ на основі достовірності.

Формування рішення на основі довіри людини до рекомендації ШІ. Незаподіяння шкоди є основним правилом охорони здоров‘я. Етап формування рішення на основі довіри застосовується лише у випадку, коли на попередньому етапі результатом є рішення на основі ШІ $m(x)$.

Формування рішення на основі довіри людини до результату ШІ формується на основі результату ШІ та довіри експерта до цього результату на основі заключень норма або патологія з використанням наступних сформульованих сценаріїв [13].

ISSN 2079-0031 (Print) ISSN 2411-0558 (Online)

Сценарій 1: лікар довіряє правильним рекомендаціям, що надані ШІ.

Сценарій 2: лікар не довіряє неправильній рекомендації ШІ і відкидає її

Сценарій 3: лікар довіряє невірній рекомендації ШІ.

Сценарій 4: лікар відхиляє правильну рекомендацію ШІ.

Запропоновані сценарії можуть бути розширені в залежності від кількості досліджуваних ступенів захворювання.

Сценарій 1 та 2 є безпечними практиками, але сценарій 2 є результатом, коли ШІ не має довіри експерту. Сценарій 3 та 4 є ризиком для пацієнта.

Результати ШІ та ЛІ порівнюються з еталонним значенням. На основі тестування рішення ШІ отримуються результати хибно позитивних (ХП), хибно негативних (ХН), істино позитивних (ІП), істино негативних (ІН) спрацьовувань та розраховується частота хибно позитивних результатів FP як ХП/(ХП+ІН). Це ймовірність того, що буде дано позитивний результат, коли справжнє значення є негативним. Хибно негативний показник FN – це ймовірність того, що модель ШІ пропустить справді позитивний результат. Він розраховується як ХН/(ХН+ІП).

Кількість помилок ШІ з точки зору експерту має бути зведена до мінімуму, оскільки кількість хибно позитивних (твердження, що здоровий пацієнт має певний діагноз) і хибно негативних (твердження, що хворий пацієнт здоровий) може призвести до надмірного виявлення пацієнтів, які здорові або не лікують пацієнтів, які насправді є нездоровими чи хворими. Важливо підтримувати баланс між надмірним виявленням і зниженням глобальних витрат на медичне обслуговування та оплату праці, зберігаючи при цьому високий рівень справжнього позитивного виявлення, отже, гарантуючи швидке й адекватне лікування людей із позитивними випадками.

Кількість помилкових спрацьовувань має бути збалансована з кількістю помилково негативних результатів. В ідеалі це означає, що як хибно позитивні, так і хибно негативні результати ШІ мають бути нижчими, ніж результати ЛІ.

Для оцінки спрацьовувань для порівняння стандартних методів ЛІ із методами ШІ використовуються еталонні результати діагностики. Міра підтверджує рівень похибки, який, у свою чергу, повинен перевищувати рівень поточної найкращої практики. У тесті чотири випадки з результатами, де лише два дають однозначну відповідь:

(1) $FN^M < FN^H$ та $FP^M < FP^H$,
(2) $FN^M > FN^H$ та $FP^M < FP^H$.
(3) $FN^M < FN^H$ та $FP^M > FP^H$,

ISSN 2079-0031 (Print) ISSN 2411-0558 (Online)

(4) $\mathrm{FN}^{\mathrm{M}} > \mathrm{FN}^{\mathrm{H}}$ та $\mathrm{FP}^{\mathrm{M}} > \mathrm{FP}^{\mathrm{H}}$.

Розглядаючи наслідки груп результатів:

– у випадку (1) ШІ штучний інтелект вважатиметься таким, що відповідає вимозі неспричинення шкоди пацієнту;

– у випадку результату (4) ШІ не проходить перевірку, оскільки обидві змінні мають менш сприятливий результат, ніж результати ЛІ;

– чи потрапляє результат (2) і (3) у прийнятний етичний діапазон, залежатиме від тяжкості шкоди, до якої призведе хибний результат.

Таким чином, керуючись вимогою "Не причини шкоди", враховуючи хибнопозитивні *FP* та хибнонегативні помилки *FN* ШІ та ЛІ, правило оцінки довіри до ШІ $q: X \to \{0,1\}$ формулюємо $q(\mathrm{FN}^{\mathrm{AI}}, \mathrm{FP}^{\mathrm{AI}}, \mathrm{FN}^{\mathrm{H}}, \mathrm{FP}^{\mathrm{H}})$ наступним чином (5)

$$q(\mathrm{FN}^{\mathrm{M}}, \mathrm{FP}^{\mathrm{M}}, \mathrm{FN}^{\mathrm{H}}, \mathrm{FP}^{\mathrm{H}}) = \begin{cases} 1, \text{якщо } \mathrm{FN}^{\mathrm{M}} < \mathrm{FN}^{\mathrm{H}} \text{ та } \mathrm{FP}^{\mathrm{M}} < \mathrm{FP}^{\mathrm{H}}, \\ 0, \text{в іншому випадку.} \end{cases} \quad (5)$$

За правилом прийнятності рекомендації ШІ єдиним можливим варіантом рішення є сценарій за яким хибнопозитивні *FP* та хибнонегативні помилки *FN* результату ШІ з точки зору еталонного результату є нижчими за помилки ЛІ.

Формування остаточного рішення. Враховуючи результати правил прийняття рішення на основі сукупної достовірності та на основі довіри до результат ШІ, правило остаточного прийняття рішення щодо результату отриманого клінічним рішенням на основі ШІ $D: X \to \{0,1\}$ формулюємо $D(g, q)$ наступним чином (6).

$$D(g, q) = \begin{cases} 1, \text{якщо } m(x) \text{ та } q = 1, \\ 0, \text{в іншому випадку.} \end{cases} \quad (6)$$

Враховуючи вхідний зразок $x \in X$, результат правил прийняття рішення на основі достовірності та на основі довіри, правило остаточного прийняття рішення, що визначає кінцевий результат між передбаченням моделі ШІ та прогнозом людини наступним чином (7)

$$(m, h, g, D)(x) \triangleq \begin{cases} d^t = accept\ m(x) \text{ якщо } D(x) = 0, \\ d^t = reject\ m(x) \text{ якщо } D(x) = 1. \end{cases} \quad (7)$$

Практичне застосування запропонованої методології. Експеримент було реалізовано для збору даних про відношення, довіру та поведінку людини в умовах прийняття рішення з використанням рекомендацій ШІ. В експерименті прийняло участь 66 суб‘єктів. Експеримент проводився з використанням Cats and Dogs Breeds Classification Oxford Dataset [14]. В експерименті суб‘єктам поставлено завдання на класифікацію зображень. Всього в експерименті використано 100 зображень, що належать до 10 класів. Перед початком експерименту кожен суб‘єкт отримав відповідні інструкції щодо завдання. Після закінчення навчання суб'єкт почав працювати над набором зі 100 завдань з прийняття рішень, які демонструвалися йому випадковим чином. На першому етапі суб'єкт бачив зображення тварини і потім мав визначити породу тварини. На другому етапі суб'єкт бачив результат моделі ШІ щодо породи тварини та достовірність цього результату. З урахуванням цих даних суб'єкт повинен був ухвалити остаточне рішення щодо породи тварини, прийнявши або відхиливши рекомендацію моделі ШІ. Дані, зібрані в ході дослідження, дозволяють вивчити функцію прийняття рішень людиною $h(x)$, щоб зробити висновок про прогноз і достовірність людини, яка приймає рішення, в ході випробування з прийняття рішення.

Дизайн експерименту. Реалізовано наступні режими експерименту:

– виключно ЛІ: режим має на увазі використання виключно ЛІ для прийняття рішень щодо набору даних зображень;

– виключно ШІ: режим має на увазі використання виключно ШІ для прийняття рішень щодо набору даних зображень;

– спільне рішення: режим має на увазі використання запропонованої методології для формування остаточного рішення щодо класифікації зображень.

На етапі випробувань з прийняття рішень виключно людиною без підтримки ШІ кожен суб‘єкт повинен виконати 100 завдань з класифікації породи кішки чи собаки на зображенні, що демонструвалися випадковим чином. Перед початком експерименту кожен суб‘єкт отримав відповідні інструкції щодо завдання. В кожному завданні людину просили дати відповідь щодо породи кішки чи собаки. Крім того, для кожної відповіді суб’єкт повинен вказати очікувану достовірність свого рішення щодо класифікації зображення за шкалою від 0 до 10, де 0 – найнижча, а 10 найвища достовірність. Таким чином, отримано результат класифікації зображень людським інтелектом та очікувана достовірність цього результату, яка використовується далі для оптимізація правила прийняття рішення на основі сукупної достовірності на цьому етапі. Розраховано істинну достовірність ЛІ c^h на основі (1). Середня істинна достовірність результатів прийняття рішення у 100 випробуваннях 6 експертами склала 0,73.

ISSN 2079-0031 (Print) ISSN 2411-0558 (Online)

На етапі випробувань з прийняття рішень виключно ШІ використано попередньо навчені моделі ШІ GoogLeNet Inceptionv3 [15] та Local Interpretable Model-agnostic Explanations (LIME) [16], які використані для класифікації зображення та генерації пояснення на основі теплокарти (англ. heatmap), що представляє візуалізацію на основі кольорової матриці ймовірностей пікселів, віднесених до певного класу зображення для пояснення результатів класифікації. Таким чином, отримано результати передбачення породи тварини на зображенні ШІ, достовірність цього результату та інтепретація результату класифікації для людини. Достовірність ШІ c^h результатів прийняття рішення у 100 випробуваннях склала 0,82.

Далі розраховано сукупну достовірність ЛІ та ШІ $c^{m+h,t}$ на основі (2), що склала 0,77. Сукупна достовірність використовується для оптимізації правила прийняття рішення на цьому етапі.

На етапі випробувань з прийняття спільного рішення суб’єктам продемонстровано результати класифікації моделі, тобто породу тварин, разом із поясненнями LIME, які включали теплокарту із зазначенням області найвищої ймовірності для класифікованого класу. Суб'єкти повинні прийняти або відхилити результат класифікації ШІ та зробити висновок про те, була інтерпретація результату моделі зрозумілою чи ні. Приклад інтерфейсу завдання для суб‘єктів наведено на рис. 1.

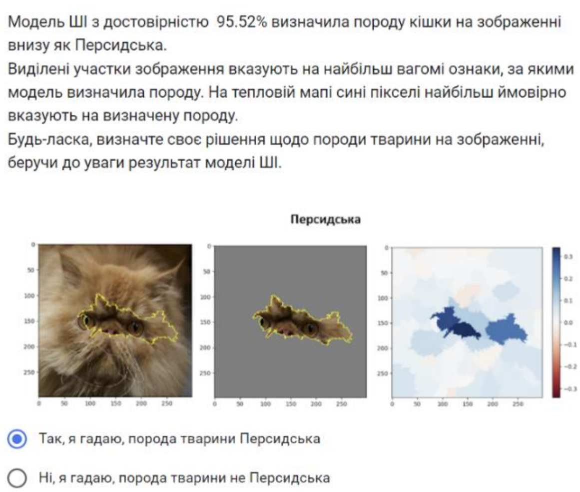

Рис. 1. Приклад інтерфейсу завдання для суб‘єктів щодо довіри до результатів ШІ

На основі довіри до інтепретації результату ШІ формується остаточне рішення щодо прийняття чи відхилення рекомендації ШІ.

ISSN 2079-0031 (Print) ISSN 2411-0558 (Online)

Висновки. Результати цього дослідження роблять внесок у загальну прийнятність використання рішень на основі ШІ. Цей підхід підкреслює важливість підтримки прозорості, інтерпретації та довіри до результатів штучного інтелекту для досягнення визнання серед медичних працівників. Вирішуючи проблеми, пов'язані з достовірністю штучного інтелекту, методологія забезпечує структуру, яка підвищує впевненість медичних працівників у використанні систем ШІ та також враховує фактор прийняття невірних рішень людиною.

References:

1. Esteva, A., Kuprel, B., Novoa, R.A., Ko, J., Swetter, S.M., Blau, H.M. and Thrun, S., (2017), Dermatologist-level classification of skin cancer with deep neural networks. nature, 542(7639), pp.115-118.

2. Goldstein, B.A. and Pencina, M.J., (2014), Developing Implementable Risk Prediction Models with Electronic Health Records Data. *Wiley StatsRef: Statistics Reference Online*, pp.1-8.

3. Hendy, J., Chrysanthaki, T., Barlow, J., Knapp, M., Rogers, A., Sanders, C., Bower, P., Bowen, R., Fitzpatrick, R., Bardsley, M. and Newman, S., (2012), An organisational analysis of the implementation of telecare and telehealth: the whole systems demonstrator, *BMC health services research*, 12, pp.1-10.

4. Kremer, M., Moritz, B. and Siemsen, E., (2011), Demand forecasting behavior: System neglect and change detection, *Management Science*, 57(10), pp.1827-1843.

5. Wang, X., Lu, Z., Yin, M., (2022), April. Will you accept the ai recommendation? predicting human behavior in ai-assisted decision making. *In Proceedings of the ACM Web Conference*, pp. 1697-1708.

6. Gao, R., Saar-Tsechansky, M., De-Arteaga, M., Han, L., Lee, M. K., Lease, M., (2021), *Human-AI collaboration with bandit feedback. arXiv preprint arXiv*:2105.10614.

7. Bansal, G., Nushi, B., Kamar, E., Horvitz, E., Weld, D. S., (2021), Is the most accurate ai the best teammate? optimizing ai for teamwork. *In Proceedings of the AAAI Conference on Artificial Intelligence*, 35(13), pp. 11405-11414.

8. Guo, C., Pleiss, G., Sun, Y. and Weinberger, K.Q., (2017), July. On calibration of modern neural networks, *In International conference on machine learning*, pp. 1321-1330.

9. Pouget, A., Drugowitsch, J. and Kepecs, A., (2016), Confidence and certainty: distinct probabilistic quantities for different goals, *Nature neuroscience*, 19 (3), pp. 366-374.

10. Fischhoff, B., Slovic, P. and Lichtenstein, S., 1977. Knowing with certainty: The appropriateness of extreme confidence, *Journal of Experimental Psychology: Human perception and performance*, 3 (4), pp. 552.

11. Pulford, B.D. and Colman, A.M., (1997), Overconfidence: Feedback and item difficulty effects, *Personality and individual differences*, 23 (1), pp. 125-133.

12. David V Budescu and Hsiu-Ting Yu. (2006), To Bayes or not to Bayes? A comparison of two classes of models of information aggregation, *Decision analysis 3*, pp. 145-162.

13. Choudhury, A., Asan, O., Medow, J. E., (2022), Effect of risk, expectancy, and trust on clinicians' intent to use an artificial intelligence system-Blood Utilization Calculator, *Applied Ergonomics*, 101, pp. 103708.

14. Parkhi, O.M., Vedaldi, A., Zisserman, A. and Jawahar, C.V., (2012), Cats and dogs, *In 2012 IEEE conference on computer vision and pattern recognition*, pp. 3498-3505.
15. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J. and Wojna, Z., (2016), Rethinking the inception architecture for computer vision, *In Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2818-2826.
16. Ribeiro, M.T., Singh, S. and Guestrin, C., (2016), Why should I trust you? Explaining the predictions of any classifier, *In Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, pp. 1135-1144.

Статтю представив д.т.н., проф. НТУ "ХПІ" Поворознюк А.І.

Надійшла (received) 18.07.2023

Biloborodova Tetiana, Cand.Sci.Tech, Assotiate Professor
G.E. Pukhov Institute for Modelling in Energy Engineering
15 General Naumov Street, Kyiv, 03164, Ukraine
e-mail: beloborodova.t@gmail.com
ORCID ID: 0000-0001-7561-7484

Skarga-Bandurova Inna, D.Sci.Tech, Professor
Ternopil Ivan Puluj National Technical University
56 Ruska Street, Ternopil, 46001, Ukraine
e-mail: skarga_bandurova@ukr.net
ORCID ID: 0000-0003-3458-8730

УДК 004.8

Методологія прийняття рішень на основі сукупної достовірності та інтепретованості рекомендації штучного інтелекту / Білобородова Т.О., Скарга-Бандурова І.С. // Вісник НТУ "ХПІ". Серія: Інформатика та моделювання. – Харків: НТУ "ХПІ". – № 1 – 2 (9 – 10). – С. 127 – 139.

У статті розглядається перехід у медичній діагностиці від традиційних клініцистських методологій до доказових підходів із використанням штучного інтелекту (ШІ). Основною метою дослідження є розробка методології прийняття рішень, заснованої на інтеграції людських рішень і рекомендацій на основі ШІ, а також можливості інтерпретації результатів ШІ. Запропонована методологія передбачає формування рішень на основі людського інтелекту (ЛІ) та ШІ, оцінку корисності рекомендацій та генерацію спільного рішення на основі кумулятивної ймовірності. Практичне застосування методики було продемонстровано шляхом експерименту з класифікації немедичних зображень. Результати дослідження підкреслюють важливість прозорості, інтерпретації та довіри до результатів ШІ для успішного використання ШІ в охороні здоров'я. Іл.: 1. Бібліогр.: 16 назв.

Ключові слова: прийняття рішень; штучний інтелект; достовірність; інтепретованість.

UDC 004.8

Decision making methodology based on generalized confidence and interpretability of artificial intelligence recommendation / Biloborodova T.O., Skarga-Bandurova I.S. // Herald of the National Technical University "KhPI". Subject issue: Information Science and Modelling. – Kharkov: NTU "KhPI". – № 1 – 2 (9 – 10). – P. 127 – 139.

The article examines the transition in medical diagnostics from traditional clinician-dependent methodologies to evidence-based approaches using artificial intelligence (AI). The primary objective of the research is to develop a decision-making methodology based on the integration of human decisions and AI-based recommendations, as well as the interpretability of AI results for humans. The proposed methodology involves the formation of decisions based on human intelligence (HI) and AI, the assessment of the utility of recommendations, and generation of a joint decision based on cumulative probability. The practical application of the methodology was demonstrated through an experiment involving the classification of non-medical images. The research findings underscore the importance of transparency, interpretability, and trust in AI results for the successful utilization of AI in healthcare. Figs.: 1. Refs.: 16 titles.

Keywords: decision making; artificial intelligence; confidence; interpretability.