

К.В. Мельник

*Національний технічний університет «Харківський політехнічний інститут», Харків*

## МОДЕЛЮВАННЯ ПРОЦЕСУ ІНТЕЛЕКТУАЛЬНОЇ ОБРОБКИ МЕДИЧНИХ ДАНИХ

*В статті розглядається один з варіантів моделювання міркувань лікаря при обробці медичних даних. Для цього використовується теорія інтелекту, а конкретно метод компараторної ідентифікації. Виділена множина ознак порушень серцево-судинних захворювань. Пропонується модель оцінки ризику розвитку захворювань на прикладі серцево-судинних захворювань.*

**Ключові слова:** медична інформаційна система, обробка медичних даних, медична картка пацієнта, теорія інтелекту, план лікувально-профілактичних заходів, компараторна ідентифікація, оцінка ризику захворювань.

### Вступ

На сьогоднішній день інформаційні технології (ІТ) використовує кожна медична установа у всіх своїх бізнес-процесах: прийом і лікування пацієнтів, ведення медичної, адміністративної та фінансової документації, проведення різних лабораторних і приборно-комп'ютерних досліджень та інше. З огляду існуючих медичних інформаційних систем (МІС) [1; 2] можна виділити наступні МІС: системи електронної історії хвороби пацієнта; автоматизовані робочі місця лікарів; консультативно-діагностичні системи; інформаційно-довідкові системи; аптечні інформаційні системи; системи архівування та передачі діагностичних даних; телемедичні системи; скринінгові системи. В результаті роботи кожної МІС накопичується величезна кількість медичних даних: інформація щодо пацієнтів, статистичні дослідження, результати перевірки працездатності медичних приладів, онтології захворювань та інше. Інформація щодо конкретного пацієнта, яка зберігається в його медичній картці, має різну природу даних: кількісні дані, якісні, дати, зображення, історія розвитку різних захворювань, тому обсяг медкарти може бути дуже великим.

Як видно з проведеного аналізу, існуючі МІС вирішують задачі діагностики, збору та зберігання медичної інформації, обробки спеціалізованих медико-біологічних сигналів. Але на сьогоднішній день в Україні практично немає таких МІС, які б обробляли інформацію з медкарти для виявлення ризиків різних захворювань та проведення профілактики. Згідно чинного наказу МОЗ України «Про диспансеризацію населення» [3] кожен заклад охорони здоров'я повинен забезпечувати організацію проведення диспансеризації населення. Це включає в себе проведення комплексу лікувально-оздоровчих заходів та впровадження заходів з пропаганди здорового способу життя. План лікувально-профілактичних заходів повинен відповідати чин-

ному законодавству. Корекцію плану для конкретного пацієнта може провести лікар в залежності від того, яка інформація представлена в медичній картці цього пацієнта, але на це витрачається багато часу. Тому стає актуальною задача розробки інструменту обробки медичних даних з картки пацієнта для розробки плану лікувально-профілактичних заходів на основі міркувань лікаря.

**Мета даної роботи** – вирішення задачі інтелектуальної обробки медичних даних на основі моделювання міркувань лікаря.

### Інтелектуальна обробка медичних даних

Розглянемо процес інтелектуальної обробки медичних даних на прикладі ідентифікації знань з точки зору оцінки розвитку і профілактики серцево-судинних захворювань (ССЗ). Аналітичний огляд існуючих методів обробки медичної інформації показав, що для цього можна використовувати апарат нейронних мереж [4; 5], нечітку логіку [4–9], Байєсові мережі [5; 11; 12], теорію інтелекту [5; 13; 14] та інше. Але будь-які висновки, зроблені за допомогою математичних методів обробки інформації, на основі яких буде розроблятися план лікувально-профілактичних заходів, повинні відповідати чинним медико-технологічним документам та засадам доказової медицини. Такими документами є уніфікований клінічний протокол надання медичної допомоги при порушеннях ССЗ та для профілактики ССЗ [15], а також адаптована клінічна настанова, заснована на доказах, при проведенні профілактики ССЗ [16]. В залежності від захворювання план лікувально-профілактичних заходів може відрізнятись, що теж прописано в законодавстві у локальних протоколах профілактики і лікування певних ССЗ [17].

Посилаючись на протоколи та клінічну настанову [15–17] комплекс медико-організаційних заходів щодо профілактики ССЗ можна представити у вигляді схеми на рис. 1.

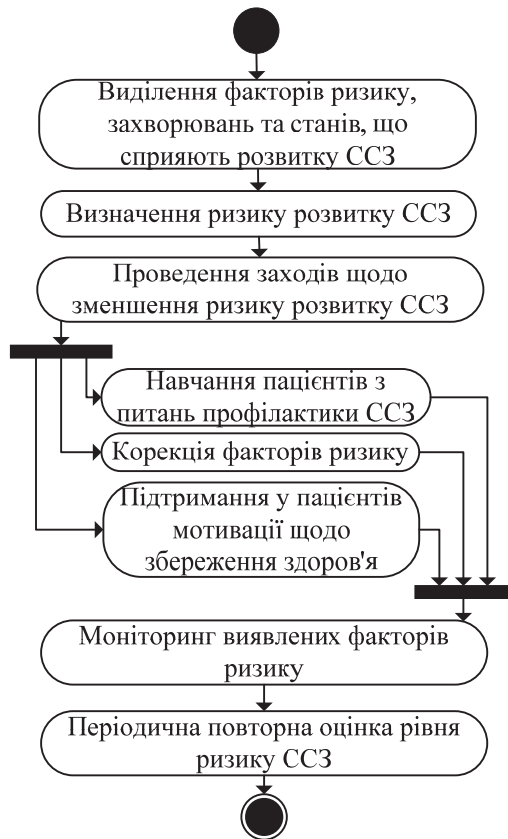


Рис. 1. План лікувально-профілактичних заходів згідно клінічному протоколу [15]

Згідно приведеного плану, можна виділити наступні задачі, які потребують вирішення при складанні комплексу лікувально-профілактичних заходів:

1) оцінка групи ризику щодо певного захворювання на основі даних конкретної медкартки, що в подальшому дозволить розробити набір рекомендацій пацієнту для елімінації ризику захворювання;

2) групування медичних карток в залежності від цілей: проведення диспансеризації, проведення скри-

нінгових процедур, проведення вакцинації тощо; що дозволить більш ефективно планувати роботу ЛПУ;

3) виявлення переліку відсутніх аналізів або процедур, обробка результатів яких дозволить лікарю приймати більш ефективні медичні рішення щодо порушень ССЗ.

Вирішити наведені задачі дозволить інтелектуальна обробка інформації з медичних карток. Змодельовано процес міркування лікаря при вирішенні поставлених задач. Для цього пропонується застосувати теорію інтелекту. Як показано в роботах [13; 14] для моделювання міркувань лікаря можна використовувати метод компараторної ідентифікації (КІ) [18], який базується на використанні алгебри скінченних предикатів. Основою методу КІ є здатність людини однозначно віднести об'єкт, що пропонується, до одного з двох класів. За цей процес відповідає предикат впізнання, який дорівнює 1 при впізнанні об'єкту та 0 – в іншому випадку.

Головна умова застосування методу КІ – це дискретність, кінцевість і детермінованість об'єктів предметного простору. В даному випадку предметний простір являє собою декартовий добуток множини об'єктів з предметної області: медичних даних, карток пацієнтів, апаратури тощо.

Для розробки плану лікувально-профілактичних заходів згідно клінічному протоколу [15] спочатку лікарю треба виділити фактори ризику (ФР) та супутні захворювання ССЗ. Для цього з усієї інформації в картці пацієнта необхідно визначити набір інформативних ознак, які мають відношення до ССЗ, або впливають на них.

Результатом огляду медичних карток стала сформована множина інформативних ознак серцевих захворювань, приклад яких представлено в табл. 1.

Таблиця 1

Комплекс ознак СС

№ п/п	Назва ознаки	Позначення	Значення ознаки
1	Стать	$X_1 = \{x_1^1, x_1^2\}$	$x_1^1$ – жінка; $x_1^2$ – чоловік
2	Вік	$X_2 = \{x_2^1, x_2^2, x_2^3\}$	$x_2^1$ – менш, ніж 40 років; $x_2^2$ – 40-50 років; $x_2^3$ – за 50 років
3	Цукровий діабет	$X_3 = \{x_3^1, x_3^2, x_3^3, x_3^4\}$	$x_3^1$ – є; $x_3^2$ – нема, актуальний діагноз; $x_3^3$ – нема, не актуальний діагноз; $x_3^4$ – невідомо
4	Артеріальна гіпертензія	$X_4 = \{x_4^1, x_4^2, x_4^3, x_4^4\}$	$x_4^1$ – є; $x_4^2$ – нема, актуальний діагноз; $x_4^3$ – нема, не актуальний діагноз; $x_4^4$ – невідомо
5	Порушення роботи нирок	$X_5 = \{x_5^1, x_5^2, x_5^3\}$	$x_5^1$ – присутні; $x_5^2$ – нема; $x_5^3$ – невідомо
6	Тахікардія	$X_6 = \{x_6^1, x_6^2, x_6^3, x_6^4, x_6^5\}$	$x_6^1$ – є, актуальний діагноз; $x_6^2$ – є, не актуальний діагноз; $x_6^3$ – нема, актуальний діагноз; $x_6^4$ – нема, не актуальний діагноз; $x_6^5$ – невідомо

7	Спадковість щодо ССЗ	$X_7 = \{x_7^1, x_7^2, x_7^3\}$	$x_7^1 - \epsilon$ ; $x_7^2 - \text{нема}$ ; $x_7^3 - \text{невідомо}$
8	Куріння	$X_8 = \{x_8^1, x_8^2, x_8^3\}$	$x_8^1 - \text{курить}$ ; $x_8^2 - \text{не курить}$ ; $x_8^3 - \text{невідомо}$
9	Зловживання алкоголем	$X_9 = \{x_9^1, x_9^2, x_9^3\}$	$x_9^1 - \text{присутнє}$ ; $x_9^2 - \text{нема}$ ; $x_9^3 - \text{невідомо}$
10	Гіподинамія	$X_{10} = \{x_{10}^1, x_{10}^2, x_{10}^3\}$	$x_{10}^1 - \text{присутня}$ ; $x_{10}^2 - \text{нема}$ ; $x_{10}^3 - \text{невідомо}$

Область зміни заданих 10-ти ознак можна формально задати наступним чином, використовуючи закони істинності, хибності та заперечення:

$$\begin{cases} x_j^1 \wedge x_j^2 = 0; \\ x_j^1 \vee x_j^2 = 1, \end{cases} \text{ де } j \in \{1, 7\};$$

$$\begin{cases} x_j^1 \wedge x_j^2 \wedge x_j^3 = 0; \\ x_j^1 x_j^2 \vee x_j^1 x_j^3 \vee x_j^2 x_j^3 = 0; \end{cases} \text{ де } j \in \{2, 5, 7, 8, 9, 10\};$$

$$\begin{cases} x_j^1 \vee x_j^2 \vee x_j^3 = 1, \\ x_j^1 \wedge x_j^2 \wedge x_j^3 \wedge x_j^4 = 0; \\ x_j^1 x_j^2 x_j^3 \vee x_j^1 x_j^2 x_j^4 \vee x_j^1 x_j^3 x_j^4 \vee x_j^2 x_j^3 x_j^4 = 0; \\ x_j^1 x_j^2 \vee x_j^1 x_j^3 \vee x_j^1 x_j^4 \vee x_j^2 x_j^3 \vee x_j^2 x_j^4 \vee x_j^3 x_j^4 = 0; \\ x_j^1 \vee x_j^2 \vee x_j^3 \vee x_j^4 = 1, \end{cases}$$

де  $j \in \{3, 4\}$ ;

$$\begin{cases} x_6^1 \wedge x_6^2 \wedge x_6^3 \wedge x_6^4 \wedge x_6^5 = 0; \\ x_6^1 x_6^2 x_6^3 x_6^4 \vee x_6^1 x_6^2 x_6^3 x_6^5 \vee x_6^1 x_6^2 x_6^4 x_6^5 \vee x_6^1 x_6^3 x_6^4 x_6^5 \vee \\ \vee x_6^2 x_6^3 x_6^4 x_6^5 = 0; \\ x_6^1 x_6^2 x_6^3 \vee x_6^1 x_6^2 x_6^4 \vee x_6^1 x_6^2 x_6^5 \vee x_6^1 x_6^3 x_6^4 \vee x_6^1 x_6^3 x_6^5 \vee \\ \vee x_6^1 x_6^4 x_6^5 \vee x_6^2 x_6^3 x_6^4 \vee x_6^2 x_6^3 x_6^5 \vee x_6^2 x_6^4 x_6^5 \vee x_6^3 x_6^4 x_6^5 = 0; \\ x_6^1 x_6^2 \vee x_6^1 x_6^3 \vee x_6^1 x_6^4 \vee x_6^1 x_6^5 \vee x_6^2 x_6^3 \vee x_6^2 x_6^4 \vee x_6^2 x_6^5 \vee \\ \vee x_6^3 x_6^4 \vee x_6^3 x_6^5 \vee x_6^4 x_6^5 = 0; \\ x_6^1 \vee x_6^2 \vee x_6^3 \vee x_6^4 \vee x_6^5 = 1. \end{cases}$$

Введемо множину значень класу ризику згідно протоколу [15]:  $R = \{r_1, r_2, r_3, r_4\}$ , де  $r_1$  – низький рівень ризику ССЗ,  $r_2$  – помірний,  $r_3$  – високий,  $r_4$  – дуже високий рівень. Значення виділених ознак з медичної карти конкретного пацієнта можна відобразити графічно (рис. 2), де номер замальованого кола відповідає номеру значення ознаки  $X_i (i = \overline{1, 10})$ , а останній стовпчик R показує номер класу ризику виникнення ССЗ, до якого віднесли медкарту пацієнта. Кожна комбінація значень виділених ознак відноситься до одного з чотирьох класів, що наведені вище. Відображення набору ознак у конкретний клас ризику  $X \rightarrow R$  визначається однозначно на основі опитування фахівців у медицині та згідно чинним медико-технологічним документам.

Обрані ознаки ССЗ тісно пов'язані між собою, наприклад, погіршення значення однієї ознаки може призвести до погіршення значення другої ознаки, або можлива ситуація, коли сукупність певних значень різних ознак є причиною для зміни значення якоїсь залежної ознаки. Для предметної області, що розглядається у даній роботі, можна привести такі приклади: наявність діабету може слугувати причиною поганої роботи нирок, які, в свою чергу, можуть призвести до появи артеріальної гіпертензії. На основі сукупності обраних ознак та взаємозв'язків між ними можна побудувати логічну мережу, яка представляє собою графічне представлення декартового добутку усіх ознак, що розглядаються.

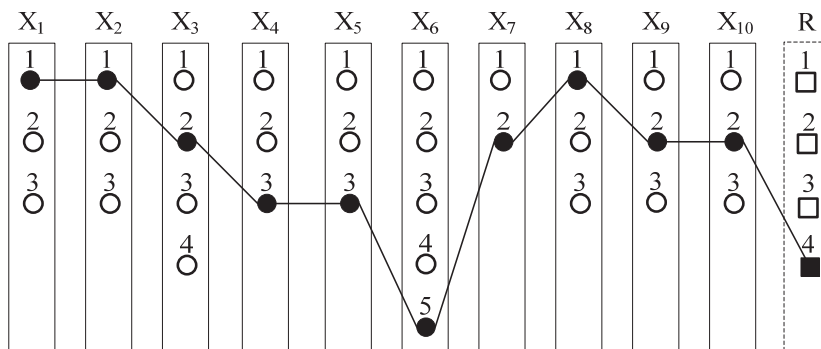


Рис. 2. Відображення ідентифікованих даних конкретного пацієнта

Згідно протоколу профілактики ССЗ та надання допомоги при лікуванні ССЗ [15] спочатку треба виділити фактори ризику та загальні ознаки, а потім

розглядаються наявні симптоми. Тому побудову логічної мережі доцільно розподілити на два етапи: спочатку групуються пов'язані за типом ознаки,

створюються агреговані показники по кожному типу ознак, а потім елементи декартового добутку агрегованих показників розподіляються по класам ризику виникнення захворювань. Загальний вигляд дерева, яке показує процес створення агрегованих показників, представлений на рис. 3.

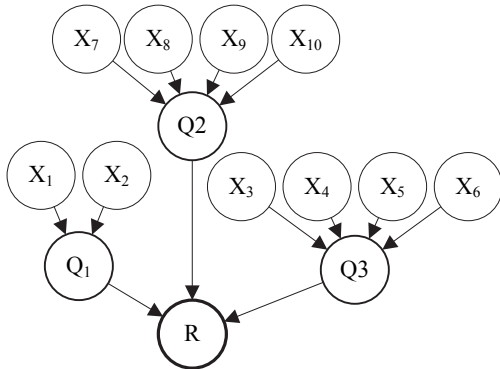


Рис. 3. Процес створення агрегованих показників

Перша група ознак – це загальна інформація щодо пацієнта:  $X_1$  – стать та  $X_2$  – вік, вони формують агрегований показник  $Q_1$ . Друга група ознак складається з факторів ризику виникнення ССЗ:  $X_7$  – спадковість,  $X_8$  – куріння,  $X_9$  – зловживання алкоголем та  $X_{10}$  – гіподинамія або відсутність фізичного навантаження в житті пацієнта. За цю групу відповідає агрегований показник  $Q_2$ . Перелік сигнальних симптомів щодо наявності порушень в серцево-судинній системі складається у показник  $Q_3$ :  $X_3$  – наявність цукрового діабету,  $X_4$  – артеріальна гіпертензія,  $X_5$  – тенденції до порушень роботи нирок,  $X_6$  – тахікардія.

Значення кожного агрегованого показника також доцільно розподілити на чотири класи ризику виникнення ССЗ, що відповідає протоколу профілактики ССЗ та надання допомоги при лікуванні ССЗ [15]. Логічна мережа агрегованого показника  $Q_1$ , яка відображає відносини приналежності тих чи інших медичних карток за загальними ознаками  $X_1$  – стать та  $X_2$  – вік до конкретних класів ризику, зображена на рис. 4.

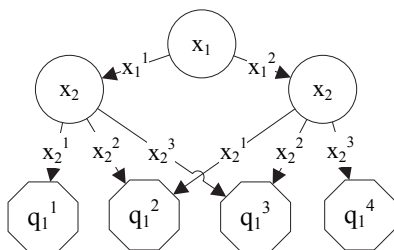


Рис. 4. Логічна мережа для загальних ознак

Показник  $Q_1 = \{q_1^1, q_1^2, q_1^3, q_1^4\}$  має наступні значення:  $q_1^1$  – низький ризик ССЗ,  $q_1^2$  – помірний ризик ССЗ,  $q_1^3$  – високий ризик ССЗ,  $q_1^4$  – дуже високий ризик ССЗ.

Розподіл шести можливих ситуацій для показника  $Q_1$  між чотирма класами ризику можна представити у наступній системі предикатних рівнянь:

$$\begin{cases} q_1^1 = x_1^1 x_2^1; \\ q_1^2 = x_1^1 x_2^2 \vee x_1^2 x_2^1; \\ q_1^3 = x_1^1 x_2^3 \vee x_1^2 x_2^2; \\ q_1^4 = x_1^2 x_2^3. \end{cases}$$

Другий агрегований показник  $Q_2$  формується на групі ознак, які є ФР для розвитку серцевих захворювань. ФР безпосередньо не є симптомами різних захворювань, але можуть бути каталізаторами до розвитку різних порушень в організмі. Показник  $Q_2$  згідно протоколу [15] також складається з наступної сукупності значень  $Q_2 = \{q_2^1, q_2^2, q_2^3, q_2^4\}$ , де  $q_2^1$  – низький ризик ССЗ,  $q_2^2$  – помірний ризик ССЗ,  $q_2^3$  – високий ризик ССЗ,  $q_2^4$  – дуже високий ризик ССЗ. Декартовий добуток факторів ризику  $X_7, X_8, X_9$  та  $X_{10}$ , на основі яких будується показник  $Q_2$ , складається з 81 можливої ситуації. Логічна мережа, яка представлена на рис. 5, відображає розподіл значень ФР на 4 класи ризику виникнення ССЗ.

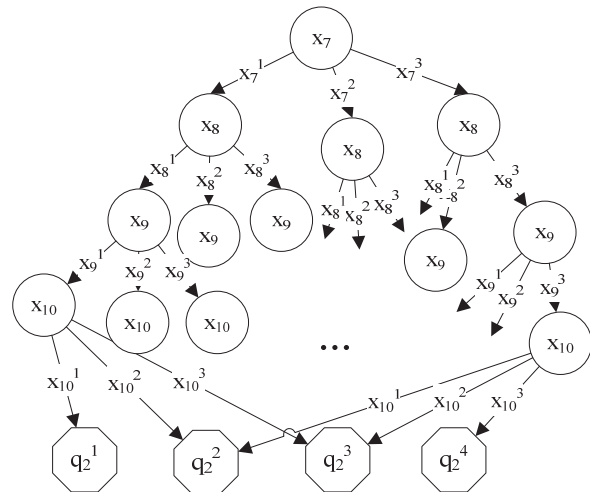


Рис. 5. Логічна мережа для показника  $Q_2$

Класу з низьким ризиком  $q_2^1$  відповідає лише 9 ситуацій, які описують відсутність усіх ФР, або наявність лише одного, окрім спадковості щодо ССЗ. Спадковість чи схильність до виникнення ССЗ має більше вагове значення, ніж інші ФР, бо цей фактор неможливо усунути з життя пацієнта. Вплив цього фактору можна зменшити за допомогою здорового способу життя.

$$q_2^1 = x_7^2 x_8^2 (x_9^2 \vee x_9^3 (x_{10}^2 \vee x_{10}^3)) \vee x_7^2 x_8^3 x_9^2 x_{10}^2 \vee \\ \vee x_7^3 x_8^2 x_{10}^2 (x_9^2 \vee x_9^3).$$

До класу з помірним ризиком  $q_2^2$  відноситься 26 ситуацій: наявність одного чи двох ФР, часткова інформація щодо ФР.

$$q_2^2 = x_7^2 (x_8^1 (x_9^1 x_{10}^2 \vee x_9^2) \vee x_9^3 (x_8^1 x_{10}^2 \vee x_8^2 x_{10}^1)) \vee \\ \vee (x_7^2 (x_8^2 x_9^1 \vee x_8^3 x_9^2) \vee x_7^3 x_8^2 (x_9^2 \vee x_9^3)) (x_{10}^1 \vee x_{10}^3) \vee \\ \vee (x_7^2 x_8^3 \vee x_7^3 x_8^2) x_9^1 (x_{10}^2 \vee x_{10}^3) \vee (x_7^2 x_9^3 \vee x_7^3 x_9^1) x_8^3 x_{10}^2 \vee \\ \vee x_7^3 x_8^1 (x_9^2 x_{10}^2 \vee x_9^3) \vee x_7^3 x_8^3 x_9^2 (x_{10}^1 \vee x_{10}^2).$$

Найчисленніший клас – це клас пацієнтів з високим ризиком  $q_2^3$ , якому відповідає 33 ситуації. Причиною такої великої кількості ситуацій слугує те, що наявність двох і більше факторів ризику зразу відносить медкарту пацієнта до класу з високим ризиком серцевих захворювань. Також до цього класу відносяться пацієнти, щодо яких частина інформації про фактори ризику відсутня.

$$q_2^3 = x_7^1 x_{10}^2 (x_8^1 x_9^2 \vee x_8^2 (x_9^1 \vee x_9^2)) \vee (x_7^1 x_9^3 (x_8^1 \vee x_8^2) \vee \\ \vee (x_7^1 x_8^3 \vee x_7^3 x_8^1) x_9^1) (x_{10}^2 \vee x_{10}^3) \vee x_7^1 x_8^3 (x_9^2 \vee x_9^3) \vee \\ \vee (x_7^2 (x_8^1 (x_9^1 \vee x_9^2) \vee x_8^3 x_9^3) \vee (x_7^2 x_8^3 \vee x_7^3 x_8^2) x_9^1 x_{10}^1 \vee \\ \vee x_7^3 (x_8^1 x_9^2 \vee x_8^3 x_9^3)) (x_{10}^1 \vee x_{10}^3) \vee x_7^3 x_8^3 (x_9^2 x_{10}^3 \vee x_9^3).$$

Клас  $q_2^4$  відповідає ситуаціям, коли присутні три або більше ФР у пацієнта:

$$q_2^4 = x_7^1 x_9^3 x_{10}^1 (x_8^1 \vee x_8^2) \vee (x_7^1 x_8^3 \vee x_7^3 x_8^1) x_9^1 x_{10}^1 \vee \\ \vee (x_7^1 x_8^2 x_9^1 \vee x_7^1 x_9^2 (x_8^1 \vee x_8^2)) (x_{10}^1 \vee x_{10}^3) \vee x_7^1 x_8^1 x_9^1.$$

Агрегований показник  $Q_3$  будується за допомогою сигнальних симптомів  $X_3, X_4, X_5$  та  $X_6$ , які являють собою маркери роботи серцево-судинної системи, або безпосередньо впливають на неї. Переліку виділених сигнальних симптомів відповідає показник  $Q_3 = \{q_3^1, q_3^2, q_3^3, q_3^4\}$ , у якого значення класів – низький ризик ССЗ, помірний, високий та дуже високий ризик відповідно. При побудові декартового добутку значень  $X_3, X_4, X_5$  та  $X_6$  формується множина з 240 можливих ситуацій.

До класу з низьким ризиком ССЗ щодо сигнальних симптомів відносяться пацієнти, у яких або немає супутніх захворювань, або спостерігається одне захворювання з тих, що розглядаються, яке є фактором ризику розвитку ССЗ, але при належному лікуванні не буде таким. Предикатне рівняння, що описує клас з низьким ризиком ССЗ щодо сигнальних симптомів, має наступний вигляд:

$$q_3^1 = ((x_3^2 \vee x_3^4) x_4^1 x_5^3 \vee (x_3^2 x_5^1 \vee x_3^1 x_5^2) x_4^2 \vee (x_4^2 \vee x_4^4) x_5^3 \vee \\ \vee x_3^1 x_4^3 (x_5^2 \vee x_5^3)) (x_6^3 \vee x_6^4) \vee (x_3^3 \vee x_3^4) x_4^3 x_5^3 \vee \\ \vee (x_6^3 \vee x_6^4 \vee x_6^5) ((x_4^2 \vee x_4^4) \vee (x_3^3 \vee x_3^4) (x_5^1 \vee x_5^2)) \vee \\ \vee x_3^2 x_5^2 \vee x_3^2 x_4^3 x_5^3 \vee ((x_3^3 \vee x_3^4 \vee x_3^5) (x_4^2 \vee x_4^3) x_5^2 \vee \\ \vee x_3^2 x_4^3 x_5^3) x_6^2.$$

До класу з помірним ризиком виникнення ССЗ належать пацієнти, що мають два захворювання, або одне захворювання та невідомість щодо інших значень сигнальних ознак. Рівняння, яке описує помірний ризик виникнення ССЗ, визначається наступним чином:

$$q_3^2 = ((x_3^2 x_1^1 \vee x_3^1 x_4^4) x_5^2 \vee x_3^1 x_5^1 (x_4^2 \vee x_4^3)) (x_6^3 \vee x_6^4) \vee \\ \vee (x_4^1 ((x_3^3 \vee x_3^4) x_5^2 \vee x_3^3 x_5^3) \vee x_3^2 x_4^4 x_5^1) (x_6^3 \vee x_6^4 \vee x_6^5) \vee \\ \vee ((x_3^2 \vee x_3^4) x_4^1 \vee x_4^2 \vee x_3^1 x_4^3 \vee x_4^4) x_5^3 (x_6^1 \vee x_6^2 \vee x_6^5) \vee \\ \vee ((x_3^2 \vee x_3^3 \vee x_3^4) (x_4^2 x_5^1 \vee x_4^4 x_5^2) \vee \\ \vee (x_3^3 \vee x_3^4) x_4^4 x_5^1) (x_6^1 \vee x_6^2) \vee \\ \vee x_3^1 x_5^2 (x_4^2 \vee x_4^3) (x_6^2 \vee x_6^5) \vee x_3^2 (x_4^2 x_5^1 x_6^5 \vee x_4^3 x_5^3 x_6^1) \vee \\ \vee (x_4^2 \vee x_4^3) x_5^2 x_6^1 \vee x_3^2 x_4^3 x_5^1 (x_6^2 \vee x_6^3 \vee x_6^4 \vee x_6^5).$$

Високий рівень ризику ССЗ відноситься до тих пацієнтів, інформація щодо яких частково невідома, та перелік наявних сигнальних симптомів може варіювати від двох до трьох.

$$q_3^3 = (x_3^1 x_4^1 (x_5^2 \vee x_5^3) \vee (x_3^2 x_4^1 \vee x_3^1 x_4^4) x_5^1) (x_6^3 \vee x_6^4) \vee \\ \vee x_4^1 x_6^5 (x_3^1 (x_5^2 \vee x_5^3) \vee x_3^2 x_5^2) \vee \\ \vee (x_4^1 ((x_3^2 \vee x_3^3 \vee x_3^4) x_5^2 \vee x_3^3 x_5^3) \vee \\ \vee x_4^1 x_5^1 (x_6^2 \vee x_6^3 \vee x_6^4 \vee x_6^5) (x_3^3 \vee x_3^4) \vee \\ x_5^1 ((x_3^3 \vee x_3^4) x_4^3 \vee x_3^2 x_4^4)) (x_6^1 \vee x_6^2) \vee \\ \vee x_3^1 ((x_4^2 \vee x_4^3) x_5^1 \vee x_4^4 x_5^2) (x_6^1 \vee x_6^2 \vee x_6^5).$$

До класу з дуже високим ризиком ССЗ відносяться пацієнти, у яких спостерігаються усі супутні сигнальні захворювання:

$$q_3^4 = x_4^1 x_5^1 (x_3^1 \vee (x_3^3 \vee x_3^4) x_6^1) \vee \\ \vee x_3^1 x_4^1 (x_5^2 \vee x_5^3) (x_6^1 \vee x_6^2) \vee \\ \vee (x_3^2 x_4^1 \vee x_3^1 x_4^4) x_5^1 (x_6^1 \vee x_6^2 \vee x_6^5).$$

Сукупність значень показників  $Q_1, Q_2$  та  $Q_3$  використовуються для побудови агрегованого показника  $R = \{r_1, r_2, r_3, r_4\}$  другого рівня, що остаточно відносить медкарту пацієнту до конкретного класу ризику ССЗ, в залежності від загальних ознак, ФР та сигнальних симптомів. Предикатні рівняння, що описують класи ризику, мають наступний вигляд:

$$r_1 = q_1^1 q_2^1 (q_3^1 \vee q_3^2) \vee (q_1^1 q_2^2 \vee (q_1^2 \vee q_1^3) q_2^1) q_3^1;$$

$$r_2 = q_1^1(q_2^1q_3^3 \vee q_2^2q_3^2) \vee (q_1^1(q_2^3 \vee q_2^4) \vee q_1^2(q_2^2 \vee q_2^3) \vee q_1^3q_2^2 \vee q_1^4(q_2^1 \vee q_2^2))(q_3^1 \vee q_3^2) \vee (q_1^2 \vee q_1^3)q_2^1(q_3^2 \vee q_3^3) \vee (q_1^2q_2^4 \vee (q_1^3 \vee q_1^4)q_2^3)q_3^1;$$

$$r_3 = q_2^1q_3^4 \vee (q_1^1 \vee q_1^2 \vee q_1^3)(q_2^2 \vee q_2^3)(q_3^3 \vee q_3^4) \vee q_1^3q_2^3(q_3^2 \vee q_3^4) \vee (q_1^3 \vee q_1^4)q_2^4q_3^1 \vee (q_1^1q_2^4 \vee q_1^4(q_2^1 \vee q_2^2))q_3^3 \vee (q_1^2q_2^4 \vee q_1^4q_2^3)(q_3^2 \vee q_3^3);$$

$$r_4 = (q_1^1 \vee q_1^2)q_2^4q_3^4 \vee q_1^3q_2^4(q_3^3 \vee q_3^4) \vee q_1^4q_3^4(q_2^2 \vee q_2^3) \vee q_1^4q_2^4(q_3^2 \vee q_3^3).$$

В залежності від того, до якого класу буде віднесена медична картка, лікарем буде розроблятися комплекс лікувально-профілактичних процедур згідно чинного законодавства та набір рекомендацій для підтримання стану здоров'я пацієнта в належній стані.

### Приклад використання компараторної ідентифікації при обробці медичних даних

Розглянемо приклад використання розробленої технології визначення групи ризику розвитку серцевих захворювань: треба визначити, в яку групу ризику віднести чоловіка 1960 р.н., у якого немає цукрового діабету, а також спостерігається тахікардія, хоча спадкової схильності до ССЗ в сім'ї немає.

З представленої інформації можна зробити наступні висновки:

Предикати впізнання для ознаки  $X_1$  такі:

$$x_1^1 = \begin{cases} 1, \text{ якщо стат} = \text{"ж"}; \\ 0, \text{ в інш. випадку,} \end{cases}$$

$$i \quad x_1^2 = \begin{cases} 1, \text{ якщо стат} = \text{"ч"}; \\ 0, \text{ в інш. випадку,} \end{cases}$$

тому  $x_1^2 = 1 \wedge x_1^1 = 0$ .

Предикати впізнання для ознаки  $X_2$  обчислюються наступним чином: спочатку треба знайти значення допоміжної величини  $A_2$ , яка потім дозволить розрахувати значення ознаки  $X_2$ .

$$A_2 = \begin{cases} 1, \text{ якщо (поточний\_рік - рік\_народж.)} \leq 40; \\ 2, \text{ якщо } 41 < (\text{поточний\_рік} - \text{рік\_нар.}) \leq 55; \\ 3, \text{ якщо (поточний\_рік - рік\_народж.)} > 55, \end{cases}$$

$$x_2^1 = \begin{cases} 1, \text{ якщо } (A_2 = 1); \\ 0, \text{ в інш. випадку,} \end{cases} \quad x_2^2 = \begin{cases} 1, \text{ якщо } (A_2 = 2); \\ 0, \text{ в інш. випадку,} \end{cases}$$

$$x_2^3 = \begin{cases} 1, \text{ якщо } (A_2 = 3); \\ 0, \text{ в інш. випадку,} \end{cases}$$

рік\_народж = 1960  $\Rightarrow A_2 = 3 \Rightarrow x_2^1 = 0; x_2^2 = 0; x_2^3 = 1$ .

Для ознаки  $X_3$ , яка характеризує наявність цукрового діабету, спочатку треба обчислити допоміжну величину  $A_3$ :

$$A_3 = \begin{cases} 1, \text{ якщо "цукр.діаб"} \leq \text{"e"}; \\ 2, \text{ якщо } \begin{cases} \text{"глюкоза"} = \text{"норма"}; \\ (\text{поточн\_рік} - \text{дата\_аналізу}) < 0,5 \text{ рік}; \end{cases} \\ 3, \text{ якщо } \begin{cases} \text{"глюкоза"} = \text{"норма"}; \\ (\text{поточн\_рік} - \text{дата\_аналізу}) \geq 0,5 \text{ рік}; \end{cases} \\ 4, \text{ в інш. випадку.} \end{cases}$$

Предикати для розрахунку значень  $X_3$ :

$$x_3^1 = \begin{cases} 1, \text{ якщо } (A_3 = 1); \\ 0, \text{ в інш. випадку,} \end{cases} \quad x_3^2 = \begin{cases} 1, \text{ якщо } (A_3 = 2); \\ 0, \text{ в інш. випадку,} \end{cases}$$

$$x_3^3 = \begin{cases} 1, \text{ якщо } (A_3 = 3); \\ 0, \text{ в інш. випадку,} \end{cases} \quad x_3^4 = \begin{cases} 1, \text{ якщо } (A_3 = 4); \\ 0, \text{ в інш. випадку.} \end{cases}$$

Виходячи з умов задачі, отримуємо:  $x_3^1 = 0$ ,  $x_3^2 = 1$ ,  $x_3^3 = 0$ ,  $x_3^4 = 0$ .

У медкарті записано, що не спостерігається схильності до ССЗ, тому  $x_7^1 = 0$ ;  $x_7^2 = 1$ ;  $x_7^3 = 0$ .

Також відомо, що в медкарті є запис про тахікардію – це свідчить про наступні значення ознаки  $X_6$ :  $x_6^1 = 1$ ,  $x_6^2 = x_6^3 = x_6^4 = x_6^5 = 0$ .

Відсутність інших індивідуальних медичних даних дає такі значення для ознак  $X_4$ ,  $X_5$ ,  $X_8$ ,  $X_9$ ,  $X_{10}$ : значення  $x_4^4$ ,  $x_5^3$ ,  $x_8^3$ ,  $x_9^3$ ,  $x_{10}^3$  дорівнюють одиниці, а усі інші дорівнюють нулю.

При підстановці отриманих даних можна обчислити значення рівнянь:

$$q_1^1 = x_1^1x_2^1 = 0 \wedge 1 = 0;$$

$$q_1^2 = x_1^1x_2^2 \vee x_1^2x_2^1 = (0 \wedge 0) \vee (1 \wedge 0) = 0;$$

$$q_1^3 = x_1^1x_2^3 \vee x_1^2x_2^2 = (0 \wedge 1) \vee (1 \wedge 0) = 0;$$

$$q_1^4 = x_1^2x_2^3 = 1 \wedge 1 = 1.$$

Аналогічно були обчислені усі інші значення агрегованого показника  $Q_2$ :

$$q_2^1 = 1 \wedge 0 \wedge (0 \vee 1 \wedge (0 \vee 1)) \vee 1 \wedge 1 \wedge 0 \wedge 0 \vee 0 \wedge 0 \wedge 0 \wedge (0 \vee 1) = 0;$$

$$q_2^2 = 1 \wedge (0 \wedge (0 \wedge 0 \vee 0) \vee 1 \wedge (0 \wedge 0 \vee 0 \wedge 0)) \vee (1 \wedge (0 \wedge 0 \vee 1 \wedge 0) \vee 0 \wedge 0 \wedge (0 \vee 1)) \wedge (0 \vee 1) \vee (1 \wedge 1 \vee 0 \wedge 0) \wedge 0 \wedge (0 \vee 1) \vee (1 \wedge 1 \vee 0 \wedge 0) \wedge 1 \wedge 0 \vee 0 \wedge 0 \wedge (0 \wedge 0 \vee 1) \vee 0 \wedge 1 \wedge 0 \wedge (0 \vee 0) = 0;$$

$$q_2^3 = 0 \wedge 0 \wedge (0 \wedge 0 \vee 0 \wedge (0 \vee 0)) \vee (0 \wedge 1 \wedge (0 \vee 0) \vee (0 \wedge 1 \vee 0 \wedge 0) \wedge 0) \wedge (0 \vee 1) \vee 0 \wedge 1 \wedge (0 \vee 1) \vee (1 \wedge (0 \wedge (0 \vee 1) \vee 1 \wedge 1) \vee (1 \wedge 1 \vee 0 \wedge 0) \wedge 0 \wedge 0 \vee 0 \wedge (0 \wedge 0 \vee 1 \wedge 0)) \wedge (0 \vee 1) \vee 0 \wedge 1 \wedge (0 \wedge 1 \vee 1) = 1;$$

$$q_2^4 = 0 \wedge 1 \wedge 0 \wedge (0 \vee 0) \vee (0 \wedge 1 \vee 0 \wedge 0) \wedge 0 \wedge 0 \vee \\ \vee (0 \wedge 0 \wedge 0 \vee 0 \wedge 0 \wedge (0 \vee 0)) \wedge (0 \vee 1) \vee 0 \wedge 0 \wedge 0 = 0.$$

Значення агрегованого показника  $Q_3$  :

$$q_3^1 = ((1 \vee 0) \wedge 0 \wedge 1 \vee (1 \wedge 0 \vee 0 \wedge 0) \wedge 0 \vee (0 \vee 1) \wedge 1 \vee \\ \vee 0 \wedge 0 \wedge (0 \vee 1)) \wedge (0 \vee 0) \vee (0 \vee 0) \wedge 0 \wedge 1 \vee \\ \vee (0 \vee 0 \vee 0) \wedge ((0 \vee 0 \vee 1) \wedge ((0 \vee 0) \wedge (0 \vee 0) \vee 1 \wedge 0) \vee \\ \vee 1 \wedge 0 \wedge 1) \vee ((1 \vee 0 \vee 0) \wedge (0 \vee 0) \wedge 0 \vee 1 \wedge 0 \wedge 1) \wedge 0 = 0;$$

$$q_3^2 = ((1 \wedge 0 \vee 0 \wedge 1) \wedge 0 \vee 0 \wedge 0 \wedge (0 \vee 0)) \wedge (0 \vee 0) \vee \\ \vee (0 \wedge ((0 \vee 0) \wedge 0 \vee 0 \wedge 1) \vee 1 \wedge 1 \wedge 0) \wedge (0 \vee 0 \vee 0) \vee \\ \vee ((1 \vee 0) \wedge 0 \vee 0 \vee 0 \wedge 0 \vee 1) \wedge 1 \wedge (1 \vee 0 \vee 0) \vee \\ \vee ((1 \vee 0 \vee 0) \wedge (0 \wedge 0 \vee 1 \wedge 0) \vee (0 \vee 0) \wedge 1 \wedge 0) \wedge (1 \vee 0) \vee \\ \vee 0 \wedge 0 \wedge (0 \vee 0) \wedge (0 \vee 0) \vee 1 \wedge (0 \wedge 0 \wedge 0 \vee 0 \wedge 1 \wedge 1) \vee \\ \vee (0 \vee 0) \wedge 0 \wedge 1 \vee 1 \wedge 0 \wedge 0 \wedge (0 \vee 0 \vee 0 \vee 0) = 1;$$

$$q_3^3 = (0 \wedge 0 \wedge (0 \vee 1) \vee (1 \wedge 0 \vee 0 \wedge 1) \wedge 1) \wedge (0 \vee 0) \vee \\ \vee 0 \wedge 0 \wedge (0 \wedge (0 \vee 1) \vee 1 \wedge 0) \vee (0 \wedge (1 \vee 0 \vee 0) \wedge 0 \vee 0 \wedge 1) \vee \\ \vee 0 \wedge 0 \wedge (0 \vee 0 \vee 0 \vee 0) \wedge (0 \vee 0) \vee \\ \vee 0 \wedge ((0 \vee 0) \wedge 0 \vee 1 \wedge 1) \wedge (1 \vee 0) \vee \\ \vee 0 \wedge ((0 \vee 0) \wedge 0 \vee 1 \wedge 0) \wedge (1 \vee 0 \vee 0) = 0;$$

$$q_3^4 = 0 \wedge 0 \wedge (0 \vee (0 \vee 0) \wedge 1) \vee 0 \wedge 0 \wedge (0 \vee 1) \wedge (1 \vee 0) \vee \\ \vee (1 \wedge 0 \vee 0 \wedge 1) \wedge 0 \wedge (1 \vee 0 \vee 0) = 0.$$

Обчислення рівнянь агрегованих показників призвело до наступних результатів:

$$\begin{matrix} q_1^1 = 0 & q_2^1 = 0 & q_3^1 = 0 \\ q_1^2 = 0 & q_2^2 = 0 & q_3^2 = 1 \\ q_1^3 = 0 & q_2^3 = 1 & q_3^3 = 0 \\ q_1^4 = 1 & q_2^4 = 0 & q_3^4 = 0 \end{matrix}$$

Виявити групу ризику ССЗ медкарти, що розглядається, дозволить обчислення рівнянь, які описують побудову показника R :

$$\begin{aligned} r_1 &= 0 \wedge 0 \wedge (0 \vee 1) \vee (0 \wedge 0 \vee (0 \vee 0) \wedge 0) \wedge 0 = 0; \\ r_2 &= 0 \wedge (0 \wedge 0 \vee 0 \wedge 1) \vee (0 \wedge (1 \vee 0) \vee 0 \wedge (0 \vee 1) \vee \\ &\vee 0 \wedge 0 \vee 1 \wedge (0 \vee 0)) \wedge (0 \vee 1) \vee (0 \vee 0) \wedge 0 \wedge (1 \vee 0) \vee \\ &\vee (0 \wedge 0 \vee (0 \vee 1) \wedge 1) \wedge 0 = 0; \\ r_3 &= 0 \wedge 0 \vee (0 \vee 0 \vee 0) \wedge (0 \vee 1) \wedge (0 \vee 0) \vee \\ &\vee 0 \wedge 1 \wedge (1 \vee 0) \vee (0 \vee 1) \wedge 0 \wedge 0 \vee \\ &\vee (0 \wedge 0 \vee 1 \wedge (0 \vee 0)) \wedge 0 \vee (0 \wedge 0 \vee 1 \wedge 1) \wedge (1 \vee 0) = 1; \\ r_4 &= (0 \vee 0) \wedge 0 \wedge 0 \vee 0 \wedge 0 \wedge (0 \vee 0) \vee 1 \wedge 0 \wedge (0 \vee 1) \vee \\ &\vee 1 \wedge 0 \wedge (1 \vee 0 \vee 0) = 0. \end{aligned}$$

Остаточні значення:  $r_1 = 0$ ,  $r_2 = 0$ ,  $r_3 = 1$ ,  $r_4 = 0$ .

Інтелектуальна обробка даних з медкарти з великою кількістю відсутніх даних призвела до того, що пацієнт попав в групу з великим ризиком розвитку серцевих захворювань.

Отримана інформація враховується при подальшому визначенні комплексу профілактичних заходів для цього пацієнта, а також для розробки ре-

комендацій, що дозволять зменшити ризик ССЗ, і порад щодо здорового способу життя.

## Висновки

1. Розроблена модель інтелектуальної обробки медичних даних на основі моделювання процесу міркування лікаря за допомогою методу компаративної ідентифікації. Ця модель дозволяє зробити попередню обробку даних з медкарт пацієнтів щодо певного захворювання без втручання лікаря. Використання розробленої моделі на практиці має свої переваги:

- лікар не витрачає час на попередню обробку медичних даних;
- автоматизована обробка даних дозволяє знизити до мінімуму такий чинник, як людський фактор;
- визначена група ризику щодо певного захворювання дозволяє розробити набір рекомендацій і порад щодо здорового способу життя та приведення здоров'я пацієнта до норми;
- розроблена модель дозволяє підготувати інформацію щодо пацієнта, яка необхідна для розробки чи корекції набору лікувально-профілактичних процедур для профілактики чи лікування певних захворювань.

А це, в свою чергу, дозволить приймати лікарю більш ефективні медичні рішення.

2. Проведено тестові розрахунки розробленої моделі, що підтвердило практичну значимість результатів для лікувально-профілактичних установ.

## Список літератури

1. Мельник К.В. Архитектура медицинской скрининговой информационной системы / К.В. Мельник // Материалы 14-й Международной научно-технической конференции «Системный анализ и информационные технологии. SAIT 2012». – Киев: УНК «ИПСА» НТУУ «КПИ», 2012. – С. 291-292.
2. Мельник К.В. Анализ данных для медицинской информационной системы в лечебно-профилактическом учреждении / К.В. Мельник, А.Е. Голоскоков // Вестник НТУУ«ХПИ». – Харьков, 2012. – № 29. – С. 60-67.
3. Наказ Міністерства охорони здоров'я України № 728 від 27.08.2010 «Про диспансеризацію населення» [Електронний ресурс]. – Режим доступу до ресурсу: <http://zakon2.rada.gov.ua/laws/show/z1396-10>.
4. Мельник К.В. Проблема комплексного лечения сердечно-сосудистой системы пациента / К.В. Мельник, А.Е. Голоскоков // Вестник НТУУ«ХПИ». – Харьков, 2006. – № 39. – С. 153-158.
5. Мельник К.В. Проблемы и основные подходы к решению задачи медицинской диагностики / К.В. Мельник, С.И. Ершова // Системы обработки информации. – Х.: ХУПС, 2011. – Вып. 2. – С. 244-248.
6. Мельник К.В. Система принятия решений при управлении лечением сердечных заболеваний / К.В. Мельник, А.Е. Голоскоков // Вестник НТУУ«ХПИ». – Харьков, 2008. – № 26. – С. 13-17.

7. Мельник К.В. Синтез продукционной системы диагностики состояния пациента / К.В. Мельник, А.Е. Голоскоков // *Материалы всеукраинской научно-практической конференции «Системный анализ и управление «Дни науки - 2007»*. Том 3. – Запорожье, 2007. – С. 175-176.

8. Мельник К.В. Разработка системы управления лечением пациента, основанной на нечеткой ситуационной сети / К.В. Мельник, А.Е. Голоскоков // *Материалы международной научно-практической конференции «Информационные технологии и информационная безопасность в науке, технике и образовании «Инфотех - 2007»*. Часть 2. – Севастополь, 2007. – С. 13-17.

9. Мельник К.В. Система управления лечением пациента на основе нечеткой продукционной базы правил / К.В. Мельник, А.Е. Голоскоков // *Материалы международной научно-практической конференции «Современные информационные и электронные технологии»* Том 1. – Одесса, 2008. – С. 59.

10. Мельник К.В. Процедура диагностирования состояния сердечно-сосудистой системы пациента на основе нечеткой логики / К.В. Мельник, А.Е. Голоскоков // *Вестник НТУ «ХПИ»*. – Харьков, 2008. – № 49. – С. 101-104.

11. Мельник К.В. Применение аппарата Байесовых сетей при обработке данных из медицинских карточек / К.В. Мельник, В.Н. Глушко // *Science and Education a New Dimension: Natural and Technical Sciences*. – Венгрия, Будапешт. – 1(2), Issue:15, 2013. – P. 126-129.

12. Мельник К.В. Использование сетей доверия для задачи скрининга / К.В. Мельник, А.Е. Голоскоков // *Тезисы доповідей міжнародної науково-практичної конференції «Інформаційні технології: наука, техніка, технологія, освіта, здоров'я»*. – Харків: НТУ «ХПИ», 2014. – С. 14.

13. Шабанов-Кушнарченко Ю.П. Теория интеллекта. Математические средства / Ю.П. Шабанов-Кушнарченко. – Х.: Вища шк. Изд-во при Харьк. ун-те, 1984. – 144 с.

14. Шабанов-Кушнарченко Ю.П. Теория интеллекта. Технические средства / Ю.П. Шабанов-Кушнарченко. – Х.: Вища шк. Изд-во при Харьк. ун-те, 1986. – 136 с.

15. Уніфікований клінічний протокол первинної, вторинної (спеціалізованої) та третинної (високоспеціалізованої) медичної допомоги. Профілактика серцево-судинних захворювань [Електронний ресурс]. – Режим доступу до ресурсу: [http://www.moz.gov.ua/ua/portal/dn\\_20160613\\_0564.html](http://www.moz.gov.ua/ua/portal/dn_20160613_0564.html).

16. Адаптована клінічна настанова «Профілактика серцево-судинних захворювань. Оновлена та адаптована клінічна настанова, заснована на доказах» [Електронний ресурс]. – Режим доступу до ресурсу: <http://www.dec.gov.ua/mtd/reestr.html>.

17. Серцево-судинні захворювання. Класифікація, стандарти діагностики та лікування [Електронний ресурс]. – Режим доступу до ресурсу: <http://ukrcardio.org/wp-content/uploads/2015/10/Recommendations-UAKSSZ.pdf>

18. Шабанов-Кушнарченко Ю.П. Компараторная идентификация лингвистических объектов: монография / Ю.П. Шабанов-Кушнарченко, Н.В. Шаронова. – К.: ІСДО, 1993. – 116 с.

Надійшла до редколегії 15.05.2017

**Рецензент:** д-р техн. наук проф. М.Д. Годлевський, Національний технічний університет «ХПИ», Харків.

## МОДЕЛИРОВАНИЕ ПРОЦЕССА ИНТЕЛЛЕКТУАЛЬНОЙ ОБРАБОТКИ МЕДИЦИНСКИХ ДАННЫХ

К.В. Мельник

В статье рассматривается один из вариантов моделирования рассуждений врача при обработке медицинских данных. Для этого используется теория интеллекта, а именно метод компараторной идентификации. Выделено множество признаков нарушений сердечно-сосудистых заболеваний. Предлагается модель оценки риска развития заболеваний на примере сердечно-сосудистых заболеваний.

**Ключевые слова:** медицинская информационная система, обработка медицинских данных, медицинская карта пациента, теория интеллекта, план лечебно-профилактических мероприятий, компараторная идентификация, оценка риска заболеваний.

## MODELING OF THE PROCESS OF INTELLIGENT MEDICAL TREATMENT OF MEDICAL DATA

K. Melnyk

The article considers one of the options for modeling the doctor's reasoning in the processing of medical data. For this, the theory of intelligence, namely the method of comparator identification, is used. A number of signs of cardiovascular diseases are identified. A model for assessing the risk of developing diseases is an example of cardiovascular disease.

**Keywords:** medical information system, medical data processing, patient medical records, theory of intelligence, plan of treatment and preventive measures, comparative identification, risk assessment of diseases.