

**ДОНЕЦЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ**

На правах рукопису

**ВАСЯЄВА ТЕТЯНА ОЛЕКСАНДРІВНА** 

УДК 004.891:004.85:004.82:004.622

**НЕЙРОМЕРЕЖЕВІ ТА ЕВОЛЮЦІЙНІ МЕТОДИ АНАЛІЗУ  
ДАНИХ У ЕКСПЕРТНИХ СИСТЕМАХ МЕДИЧНОЇ  
ДІАГНОСТИКИ**

Спеціальність 05.13.06 –  
«Інформаційні технології»

Автореферат дисертації на здобуття наукового ступеня  
кандидата технічних наук

Донецьк – 2010

Дисертацією є рукопис.

Роботу виконано в Донецькому національному технічному університеті Міністерства освіти і науки України.

**Науковий керівник:**

доктор технічних наук, професор  
Скобцов Юрій Олександрович,  
Донецький національний технічний  
університет МОН України, м. Донецьк,  
завідувач кафедри автоматизованих систем  
управління.

**Офіційні опоненти**

доктор технічних наук, професор  
Філатов Валентин Олександрович,  
Харківський національний університет  
радіоелектроніки МОН України, м. Харків,  
професор кафедри штучного інтелекту,  
науковий керівник центру інформаційних  
систем і технологій.

доктор технічних наук, професор  
Зубов Дмитро Анатолійович,  
Київський міжнародний університет МОН  
України, м. Київ, професор кафедри  
міжнародної інформації та інформатики.

Захист відбудеться «19» лютого 2010 р. о 13:00 год. на засіданні спеціалізованої вченої ради К 11.051.08 Донецького національного університету за адресою: 83001, м. Донецьк, пр. Театральний, 13, корп. 4, ауд. 416.

З дисертацією можна ознайомитись у бібліотеці Донецького національного університету за адресою: 83001, м. Донецьк, вул. Університетська, 24.

Автореферат розіслано «18» січня 2010 р.

Вчений секретар спеціалізованої  
вченої ради К 11.051.08  
кандидат технічних наук, доцент



Д.В. Шевцов

## ЗАГАЛЬНА ХАРАКТЕРИСТИКА РОБОТИ

**Актуальність теми.** Ефективність роботи будь-якої галузі значною мірою визначається рівнем інформаційних технологій і широтою їх застосування. У медицині особливого значення набуває напрям, пов'язаний з розробкою інтелектуальних систем (ІС), електронних медичних консультацій, розвитком телемедицини, оскільки все це удосконалює і визначає нові діагностичні й лікувальні можливості.

Учені всього світу не можуть пояснити драматичну і давно існуючу проблему медицини – синдром раптової смерті немовляти (СРСН), коли зовні здорова дитина спокійно засинає у своєму ліжку, а через декілька годин її знаходять мертвою. Ця проблема є актуальною, кількість випадків СРСН складає: в Австрії – 5,4, в Німеччині – 7,8, в Італії – 10, в Голландії – 1,4, в Росії – 4,3, в Словаччині – 1,4, в Україні смертність від СРСН складає 7 на 100 дітей, народжених живими. У багатьох країнах проводяться програми профілактики СРСН, а в США існує Національний інститут СРСН.

У зв'язку з цим актуальною науковою задачею є розробка методів інтелектуального аналізу даних (ІАД) і створення на базі розроблених методів експертної системи (ЕС) для визначення ступеня ризику СРСН.

**Зв'язок роботи з науковими програмами, планами, темами.** Дисертаційна робота виконана в рамках держбюджетних тем Донецького національного технічного університету Д-11-04 «Розробка і дослідження нейромережкових та еволюційних методів побудови систем прийняття рішення» (№ державної реєстрації 0104U002394), Н-3-07 «Розробка наукових основ побудови комп'ютерних систем технічної і медичної діагностики» (98-15 від 04.04.07) і Д-11-07 «Розробка і дослідження еволюційних і нейромережкових моделей, методів і алгоритмів в системах діагностики та прогнозування» (№ державної реєстрації 0107U001482).

**Мета роботи і задачі дослідження.** Мета дисертаційної роботи – підвищення достовірності визначення ступеня ризику СРСН за рахунок розробки методів інтелектуального аналізу даних, створення на базі розроблених методів ЕС та її впровадження. Для розробки ЕС визначення ступеня ризику СРСН необхідно вирішити такі задачі:

- розробити структуру ЕС визначення ступеня ризику СРСН;
- реалізувати методи кодування для підготовки даних;
- розробити і реалізувати методи відбору інформативних даних;
- розробити і реалізувати методи добування знань на основі методів ІАД.

*Об'єкт дослідження:* процес проектування і розробки методів добування знань для медичних експертних систем.

*Предмет дослідження:* методи розробки добування знань для експертної системи визначення ступеня ризику СРСН.

*Методи дослідження,* застосовані у роботі, базуються на методах теорії штучних нейронних мереж, еволюційних обчислень, дискретної математики, основних положеннях теорії ймовірності та математичної статистики.

### **Наукова новизна отриманих результатів.**

1. Вперше розроблено експертну систему визначення ступеня ризику СРСН, що дозволило підвищити його достовірність і визначати ступінь ризику на ранньому терміні вагітності і відразу після народження дитини.

2. Набув подальшого розвитку еволюційний метод відбору інформативних факторів ризику за рахунок розробленої фітнес-функції, що дозволило підвищити ефективність відбору інформативних факторів ризику СРСН шляхом регулювання співвідношення кількості факторів і помилки класифікації.

3. Вдосконалено деревоподібний спосіб кодування особини для булевої функції, що дало можливість будувати класифікаційні дерева для визначення ступеня ризику СРСН.

4. Вдосконалено спосіб кодування особини у вигляді дерева, що реалізовує булеву функцію в диз'юнктивній нормальній формі, що дозволило отримувати класифікаційні правила для визначення ступеня ризику СРСН.

5. Набула подальшого розвитку троїчна логіка в булевих обчисленнях класифікаційного дерева, що дало можливість виконувати класифікацію при невизначених значеннях деяких факторів ризику.

### **Практичне значення роботи.**

1. Розроблено і впроваджено експертну систему визначення ступеня ризику СРСН у Донецькому регіональному центрі охорони материнства і дитинства і в науково-дослідному інституті медичних проблем сім'ї Донецького національного медичного університету ім. Горького, що дозволило:

- підвищити достовірність визначення ступеня ризику СРСН на 10% порівняно з використовуваним раніше методом;
- отримати можливість визначати ступінь ризику СРСН на різних термінах вагітності і відразу після народження дитини на відміну від використовуваного раніше методу, який дозволяв визначати ступінь ризику СРСН лише після народження дитини;
- в разі високого ступеня ризику СРСН розпочинати профілактичні заходи, направлені на його зниження ще до народження дитини.

2. У держбюджетній темі Д-11-04 «Розробка і дослідження нейромережових та еволюційних методів побудови систем прийняття рішення» застосовані алгоритми і програмне забезпечення методу класифікації на базі нейронних мереж (НМ). У держбюджетній темі Н-3-07 «Розробка наукових основ побудови комп'ютерних систем технічної і медичної діагностики» застосовані алгоритми і програмне забезпечення: еволюційного методу відбору інформативних факторів ризику СРСН; методу визначення високого ступеня ризику СРСН на основі генетичного програмування (ГП). У держбюджетній темі Д-11-07 «Розробка і дослідження еволюційних і нейромережових моделей, методів і алгоритмів в системах діагностики і прогнозування» застосовані алгоритми і програмне забезпечення методу побудови системи класифікаційних правил для визначення ступеня ризику СРСН на основі генетичного програмування. Передбачено можливість виконувати класифікацію при

невизначених значеннях деяких факторів ризику за рахунок використання троїчної логіки в булевих обчисленнях.

3. Розроблені алгоритми і програмне забезпечення застосовуються в навчальному процесі під час виконання курсових робіт, дипломних проектів і магістерських робіт студентами спеціальностей «Інформаційні управляючі системи і технології» і «Спеціалізовані комп'ютерні системи» Донецького національного технічного університету.

**Особистий внесок дисертанта.** Всі основні результати роботи отримані автором самостійно. Особистий внесок здобувача полягає в такому: [1,2] – розробка методів нейромережевої класифікації; [6-9] – розробка методів підготовки даних; [3-5,10-17] – розробка методів визначення ступеня ризику СРСН.

**Апробація результатів дисертації.** Результати дисертаційної роботи викладено в наукових доповідях на міжнародній науковій конференції «Нейросетевые технологии и их применение» (м. Краматорськ, 2002); 1-й і 3-й міжнародних конференціях «Advanced Computer Systems and Networks: Design and Application» (м. Львів, 2003, 2007); II науковій конференції «ДОНБАС – 2020: НАУКА І ТЕХНІКА – ВИРОБНИЦТВУ» (м. Донецьк, 2004); V міжнародній науково-практичній конференції «Комп'ютерні системи в автоматизації виробничих процесів» (м. Хмельницький, 2007); IX міжнародній науково-практичній конференції «Інформаційні технології в освіті та управлінні» (м. Нова Каховка, 2007); міжнародній науковій конференції «Компьютерные науки и информационные технологии» (м. Саратов, 2007); міжнародних науково-практичних конференціях «ИНФОТЕХ – 2007», «ИНФОТЕХ – 2009» (м. Севастополь, 2007, 2009); IX міжнародній науково-технічній конференції «Искусственный интеллект. Интеллектуальные системы – ИИ – 2008» (м. Кацивели, 2008); міжнародній науковій конференції «Інтелектуальні системи прийняття рішень і проблеми обчислювального інтелекту (ISDMCI'2008)» (м. Євпаторія, 2008); II Всеукраїнській науково-практичній конференції «Сучасні тенденції розвитку інформаційних технологій в науці, освіті та економіці» (м. Луганськ, 2008); на розширеному засіданні кафедри автоматизованих систем управління Донецького національного технічного університету (м. Донецьк, 2009); на розширеному науковому семінарі кафедри комп'ютерних технологій Донецького національного університету (м. Донецьк, 2009).

**Публікації.** За темою дисертації опубліковано 17 наукових праць, з них 7 – у провідних спеціальних виданнях, що входять до переліку ВАК України, 2 патенти на корисну модель, інші – у матеріалах конференцій.

**Структура й обсяг дисертації.** Дисертація складається з переліку умовних скорочень, вступу, 5 розділів, висновків, списку літератури з 93 джерел і 10 додатків. Вона містить 21 ілюстрацію, 53 таблиці. Загальний обсяг дисертації складає 230 сторінок, у тому числі 145 сторінок основного тексту.

## ОСНОВНИЙ ЗМІСТ РОБОТИ

У вступі обґрунтовано актуальність теми дисертаційної роботи, сформульовано мету і задачі дослідження. Викладено основні результати, отримані в роботі, показано їх новизну, а також відзначено їх наукову й практичну значимість.

Перший розділ роботи присвячено огляду наукової літератури, визначенню напрямів досліджень. Розкрито сутність завдань діагностики і прийняття рішень в медицині як комплексного процесу і розглянуто його етапи.

Виконано аналіз способів репрезентації навчальних даних, подано класифікацію вхідної інформації та способи її кодування залежно від типу даних. Досліджено методи оцінки інформативності й редукції вхідних даних. Розглянуто статистичні методи й методи штучного інтелекту для відбору інформативних факторів ризику.

Проведено аналіз сучасних методів добування знань. Показано, що методи машинного навчання на основі використання навчальних даних мають переваги порівняно з іншими математичними методами. Розглянуто методи штучного інтелекту на базі машинного навчання. Відзначено, що генетичне програмування має потужні можливості пошуку закономірностей, розглянуто параметри, що впливають на його ефективність, значення яких залежать від конкретного завдання.

На підставі проведеного аналізу сформульовано мету, поставлено задачі та вибрано напрями досліджень для розробки медичної ЕС визначення ступеня ризику СРСН.

У другому розділі розроблено структуру ЕС визначення ступеня ризику СРСН, яку представлено на рис. 1 ( $\bar{X}$  і  $\bar{X}_{inf}$  – повний і інформативний набір факторів ризику СРСН,  $\bar{Y}$  – результат розвитку СРСН).

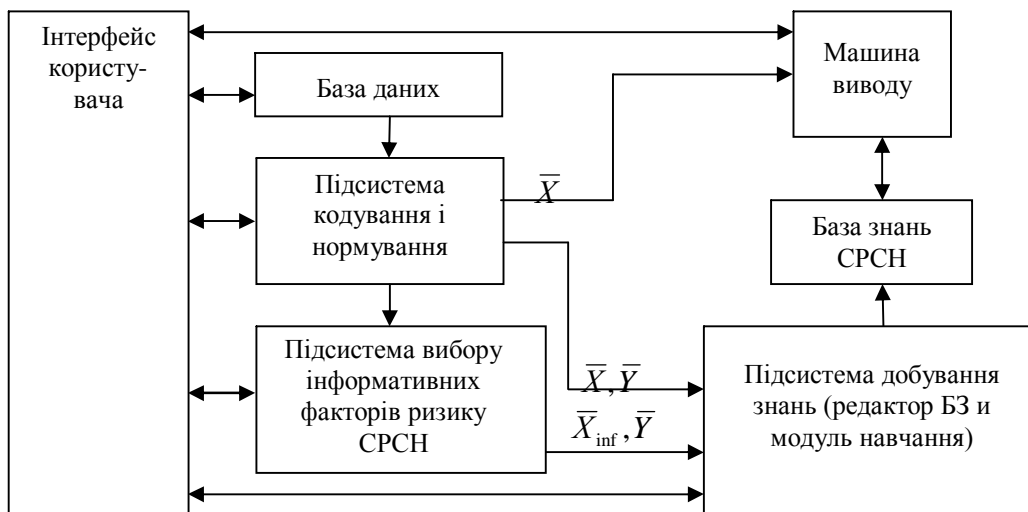


Рис. 1. Укрупнена структура ЕС визначення ступеня ризику СРСН

Здійснено постановку задачі до формування знань для ЕС. Показано, що під час вирішення задачі прогнозування необхідно визначити функцію (1):

$$Y = f(x_1, x_2, \dots, x_n), \quad (1)$$

де  $Y$  – досліджувана величина, залежна від факторів  $x_1, x_2, \dots, x_n$ .

Прогнозування, залежно від того який вигляд має цільова змінна, може бути виконане на основі класифікації або регресії. Регресія передбачає, що вихідна змінна є безперервною величиною, тоді як в даному випадку цільова змінна набуває дискретних значень, саме тому в контексті цієї задачі використано класифікацію.

Стосовно задачі визначення ступеня ризику СРСН  $Y$  – ступінь ризику, що класифікується,  $x_1, x_2, \dots, x_n$  – фактори ризику СРСН, функція (1) визначає ступінь ризику СРСН при деяких певних значеннях факторів  $x_1, x_2, \dots, x_n$ . Розглянуто декілька підходів до рішення задачі класифікації, які передбачають етап навчання з учителем. Сформовано навчальну вибірку, яка в нашому випадку становить собою реальні медичні дані, отримані під час обстеження вагітних жінок.

1. Насамперед необхідно виконати класифікацію за ступенями ризику: дуже низький ступінь ризику, низький ступінь ризику, високий ступінь ризику, дуже високий ступінь ризику. В загальному вигляді це задача класифікації на  $l$  класів, де кожен клас відповідає певному ступеню ризику СРСН, і функція для класифікації має вигляд, представлений формулами (2) і (3):

$$F = \sum_{i=1}^n x_i \cdot w_i, \quad (2)$$

$$Y = \begin{cases} K_1, F \in [0; N_1), \\ K_2, F \in [N_1; N_2), \\ \dots \\ K_l, F \in [N_{l-1}; N_l]; \end{cases} \quad (3)$$

де  $K=(K_1, K_2, \dots, K_l)$  – можливі результати класифікації;  $X=(x_1, x_2, \dots, x_n)$  – фактори ризику;  $W=(w_1, w_2, \dots, w_n)$  – вагові коефіцієнти факторів ризику;  $N=(N_1, N_2, \dots, N_l)$  – порогові значення. Розробка методу класифікації полягає в знаходженні коефіцієнтів  $W$  і порогів  $N$  (етап навчання), а безпосередньо класифікація виконується підсумовуванням коефіцієнтів наявних у пацієнта факторів ризику, і залежно від результату визначенням номера класу за формулою (3).

2. Для практики найважливішим є визначення високого ступеня ризику СРСН з метою призначення відповідного лікування, направленою на його зниження. У цьому випадку, по суті, необхідно виконати класифікацію на два класи:

а) як приклад реалізації такого підходу розглянуто НМ. При цьому необхідно підібрати архітектуру НМ і навчити її на відповідній вибірці. На етапі класифікації на входи НМ подаються закодовані фактори ризику, а на виході НМ формується результат: високий або низький ступінь ризику СРСН;

б) оскільки вирішується задача класифікації на два класи, то для її вирішення використано математичний апарат булевих функцій, реалізованих у вигляді деревоподібних структур. При цьому листям дерева є фактори ризику, які приведені до бінарних значень, функціональними вузлами – логічні функції

(І, АБО, НЕ), а корінь дерева визначає значення булевої функції, де одиниця відповідає високому ступеню ризику СРСН, а нуль – низькому. Після етапу навчання отримане дерево реалізовує певну булеву функцію, яка виконуватиме класифікацію на високий і низький ступінь ризику СРСН.

3. Для поліпшення інтерпретації вирішується задача побудови класифікаційних правил, що дає можливість не лише обчислювати значення булевої функції (визначати ступінь ризику), але і давати інтерпретацію, як саме було отримане це значення.

Припустимо, що для вектора факторів ризику  $X=(x_1, x_2, \dots, x_n)$ , для кожного  $i$ -го фактора визначено деякий вектор допустимих значень  $P_i=(p_{i,1}, p_{i,2}, \dots, p_{i,s})$ , де  $s$  – кількість допустимих значень, різна для кожного фактора  $x_i$ .

Під час рішення задачі класифікації на два класи класифікаційне правило може бути представлено, наприклад, таким чином:

ЯКЩО  $x_1=p_{1,2} \text{ І } x_2 \neq p_{2,7} \text{ І } x_3=p_{3,4} \text{ І } \dots \text{ І } x_i \neq p_{i,2}$  ТО  $Y=K_1$ , ІНАКШЕ  $Y=K_2$ .  
Часто є декілька таких правил, які приводять до правильної класифікації, наприклад:

ЯКЩО  $x_1=p_{1,2} \text{ І } x_2 \neq p_{2,7} \text{ І } x_3=p_{3,4} \text{ І } \dots \text{ І } x_i \neq p_{i,2} \text{ І } \dots \text{ І } x_n=p_{n,1}$

АБО

$x_1=p_{1,5} \text{ І } x_2=p_{2,4} \text{ І } x_3 \neq p_{3,4} \text{ І } \dots \text{ І } x_i \neq p_{i,1} \text{ І } \dots \text{ І } x_n=p_{n,1}$

АБО

...

АБО

$x_1 \neq p_{1,1} \text{ І } x_2=p_{2,3} \text{ І } x_3 \neq p_{3,5} \text{ І } \dots \text{ І } x_i \neq p_{i,8} \text{ І } \dots \text{ І } x_n=p_{n,4}$

ТО  $Y=K_1$ , ІНАКШЕ  $Y=K_2$ .

Представлені правила є класифікаційними, кожне з яких є самостійним правилом класифікаційної системи, виконання якого віднесе результат до класу  $K_1$ .

Тепер систему правил можна описати так:

ЯКЩО умова 1=так АБО умова 2=так АБО ... АБО умова  $m$ =так

ТО  $Y=K_1$ , ІНАКШЕ  $Y=K_2$ .

Тут умова  $i$ =да може бути, наприклад, складною умовою:

$x_1=p_{1,2} \text{ І } x_2 \neq p_{2,7} \text{ І } x_3=p_{3,4} \text{ І } \dots \text{ І } x_i \neq p_{i,2} \text{ І } \dots \text{ І } x_n=p_{n,1}$ .

Представлені правила можна формально описати таким чином:

ЯКЩО  $f_1(x_1, x_2, \dots, x_n)=1$  АБО  $f_2(x_1, x_2, \dots, x_n)=1$  АБО ... АБО  $f_m(x_1, x_2, \dots, x_n)=1$

ТО  $Y=K_1$ , ІНАКШЕ  $Y=K_2$ .

При такій інтерпретації для побудови класифікаційних правил необхідно знайти булеві функції виду (4):

$$f_j(x_1, x_2, \dots, x_n)=1. \quad (4)$$

У такому разі систему класифікаційних правил представлено так:

$$F(x_1, x_2, \dots, x_n) = \bigvee_{j=1}^m f_j(x_1, x_2, \dots, x_n), \quad (5)$$

а вся конструкція ЯКЩО має вигляд:

$$\text{ЯКЩО } F(x_1, x_2, \dots, x_n) = 1 \text{ ТО } Y=K_1, \text{ ІНАКШЕ } Y=K_2. \quad (6)$$



Тут  $F(x_1, x_2, \dots, x_n)$  визначена в (5), а  $f_j(x_1, x_2, \dots, x_n) = x_1 \cdot x_2 \cdot \dots \cdot x_n$ , при цьому наявність кожного фактора не обов'язкова, і кожен наявний фактор може бути в прямому або інверсному стані (операція НЕ).

Далі реалізовано методи обробки вхідних даних, включаючи кодування нечислової інформації і нормування для числової. Виконано кодування даних різними способами, що дозволило отримати можливість працювати з ЕС з різними формами репрезентації бази знань (БЗ). Проведено експерименти з вибору методів попередньої обробки для роботи з НМ. За допомогою кореляційного аналізу оцінено значимість кожного фактора ризику СРСН окремо. Виконано реалізацію гібридної системи, що реалізовує відбір інформативних факторів ризику за допомогою генетичних алгоритмів (ГА) і НМ, де на першому етапі підбирається архітектура НМ, яка дозволяє з достатньою точністю визначати високий ступінь ризику СРСН. Експериментально вибрано архітектуру НМ, яка містить один прихований шар. Вибрані активаційні функції – гіперболічний тангенс на прихованому шарі і лінійна на вихідному. Кількість нейронів у мережі: на прихованому шарі 6 нейронів, на вихідному – один, входів – 99. Кількість входів обумовлена максимальною кількістю факторів ризику (після кодування отримали 99). Далі за допомогою ГА виконується відбір інформативних факторів ризику. Схема роботи ГА така: кожен можливий варіант набору вхідних змінних представлено у вигляді бітової маски. Нуль у відповідній позиції означає, що ця вхідна змінна не включена у вхідний набір, одиниця – що включена. Таким чином, хромосома є рядком бітів – по одному на кожену можливу вхідну змінну. ГА контролює набір таких хромосом, оцінюючи кожену з них за поточною помилкою навчання НМ. За значеннями помилки відбираються варіанти хромосом, які комбінуються за допомогою штучних генетичних операторів: схрещування і мутації.

Розроблено фітнес-функцію, що представлена формулою (7):

$$F = \left( \frac{X_i}{X_n} \right) \cdot Q_1 + \left( \frac{E_i - E_n}{E_n} \right) \cdot Q_2, \quad (7)$$

де  $X_i$  – кількість одиниць для  $i$ -ої хромосоми,  $X_n$  – максимальна кількість одиниць,  $E_i$  – помилка навчання НМ для  $i$ -ої хромосоми,  $E_n$  – помилка навчання НМ при використанні максимальної кількості факторів,  $Q_1$  і  $Q_2$  – коефіцієнти, за допомогою яких виконується регулювання співвідношення точності класифікації та кількості факторів ризику. Така фітнес-функція дозволяє вибрати співвідношення точності класифікації та кількості факторів ризику.

Проведено експерименти з вибору факторів ризику СРСН. Для подальшої роботи вибрано інформативний набір даних, в якому є 46 факторів ризику, що в два рази менше початкового.

**У третьому розділі** розроблено методи добування знань для ЕС відповідно до постановки задачі в розділі 2.

1. Функція класифікації описується формулами (2) і (3), з чого виходить:

- необхідно задати кількість та інтерпретацію варіантів класифікації –  $K=(K_1, K_2, \dots, K_l)$  (можливі результати класифікації ступеня ризику СРСН, де  $l$  – кількість варіантів класифікації);
- визначити вектор вагових коефіцієнтів  $W=(w_1, w_2, \dots, w_n)$  для факторів ризику  $X=(x_1, x_2, \dots, x_n)$ , де  $n$  – кількість факторів ризику;
- визначити порогові значення  $N=(N_1, N_2, \dots, N_l)$ , де  $l$  – кількість варіантів класифікації.

Класифікаційну таблицю (запропоновані порогові значення і кількість градацій ступеня ризику залежно від терміну вагітності) представлено в табл. 1. Вагові коефіцієнти отримано в розділі 2. Класифікація запропонованим методом зводиться до підрахунку балів для всіх наявних факторів ризику (2) і визначення за табл. 1 ступеня ризику СРСН.

Таблиця 1

### Визначення ступеня ризику СРСН

Термін вагітності	Ступінь ризику			
	Дуже низький ризик	Низький ризик	Високий ризик	Дуже високий ризик
До 24 тижнів	До 14	14—26	27—55	Більше 55
Після 24 тижнів	До 15	15—30	31—61	Більше 61
Після пологів	До 20	20—39	40—80	Більше 80

2. Для реалізації ЕС, коли немає потреби визначати функцію (1) явно, використано НМ. Добрі результати класифікації показують такі архітектури НМ:

- 99 входів (відповідає кількості факторів ризику після кодування для НМ), 8 нейронів на першому прихованому шарі (функції активації – гіперболічний тангенс), 6 нейронів на другому прихованому шарі (функції активації – сигмоїд), 3 нейрони на третьому прихованому шарі (функції активації – гіперболічний тангенс) і один нейрон на виході з лінійною функцією активації;
- 99 входів, 8 нейронів на першому прихованому шарі (функції активації – гіперболічний тангенс), 4 нейрони на другому прихованому шарі (функції активації – сигмоїд) і один нейрон на виході з лінійною функцією активації;
- 99 входів, 6 нейронів на першому прихованому шарі (функції активації – гіперболічний тангенс) і один нейрон на виході з лінійною функцією активації.

У випадку двох класів також використано апарат ГП, який дозволяє побудувати за допомогою еволюційного алгоритму на основі навчальної вибірки програму, яка виконує класифікацію. Узагальнений алгоритм роботи ГП представлено на рис. 2.

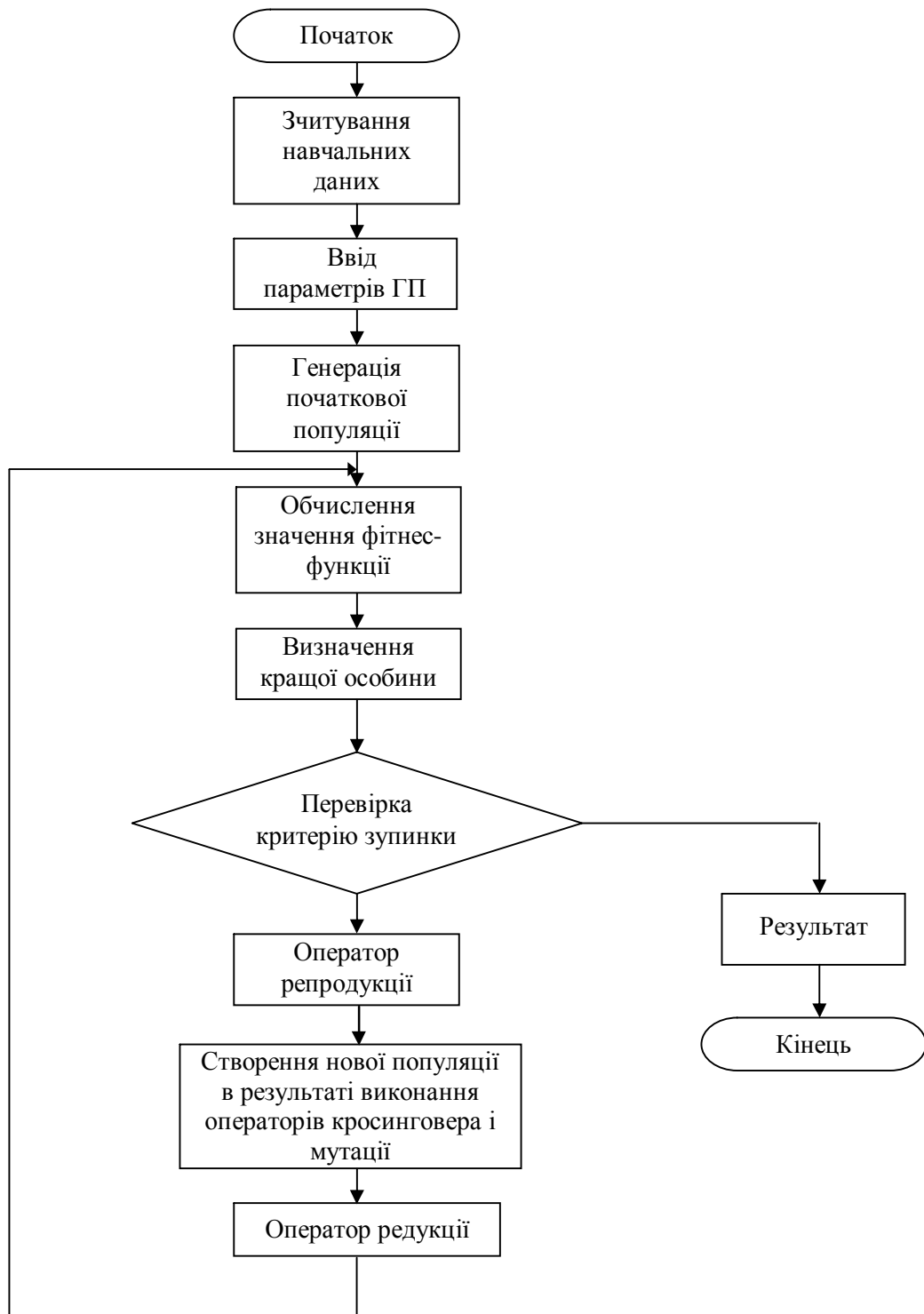


Рис. 2. Узагальнений алгоритм побудови дерева за допомогою ГП

Використано деревоподібну структуру особини. Приклад деревоподібної структури показано на рис. 3, де особина є деревом, якому відповідає логічна функція (8):

$$F(x_1, x_2, \dots, x_{176}) = \overline{(x_1 \vee x_5)} \cdot \overline{(x_{12} \cdot x_7)} \cdot \overline{(x_{21} \vee (x_{25} \vee x_8))} \cdot (x_4 \cdot x_6). \quad (8)$$

Тут значення функції  $F(x_1, x_2, \dots, x_{176}) = 1$  відповідає високому ступеню ризику СРСН, 0 – низькому.

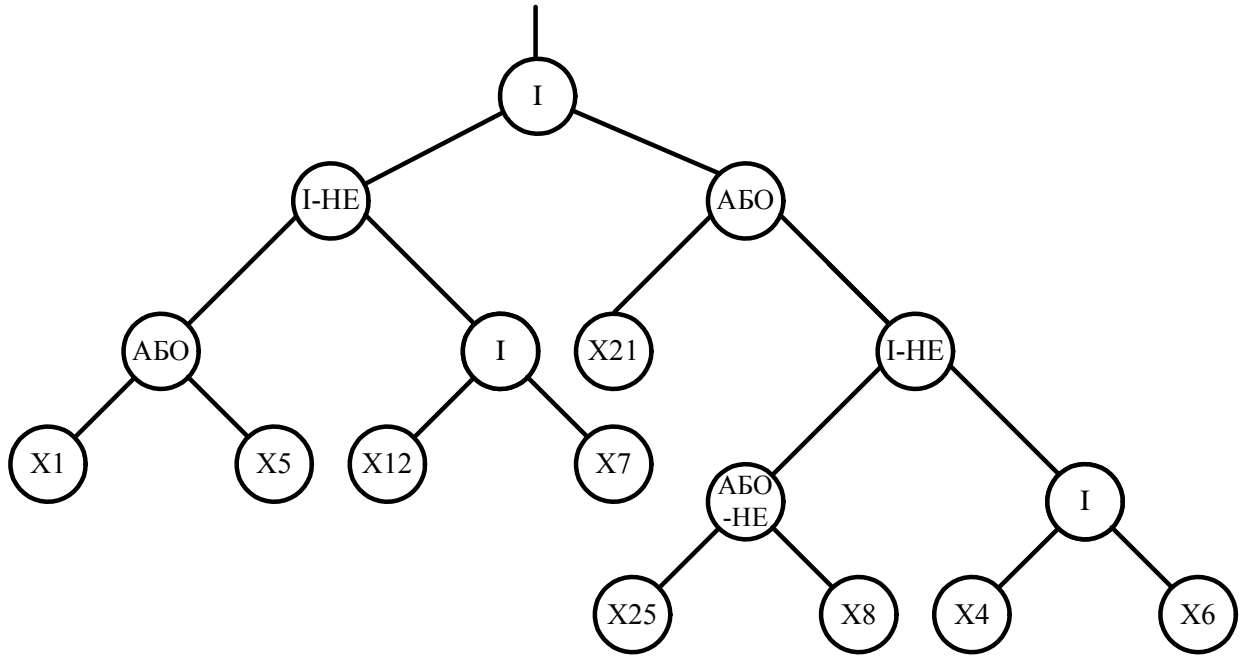


Рис. 3. Приклад структури дерева для класифікації ступеня ризику СРСН

Термінальна множина в цьому випадку складається з виділених раніше інформативних факторів ризику, які після кодування є булевозначним вектором. Функціональна множина складається з логічних операцій: I, АБО, НЕ. Оскільки перші дві операції можуть мати два і більше входи і один вихід, а остання операція завжди має один вхід і один вихід, то для зручної програмної реалізації замінимо операцію НЕ на I-НЕ та АБО-НЕ. Таку заміну виконано з метою уніфікувати кількість входів (2) для всіх операцій. Таким чином, функціональна множина складається з 4 логічних операцій I, АБО, I-НЕ та АБО-НЕ. Такий логічний базис дозволяє позбутися інтронів, в цьому випадку подвійної інверсії (НЕ(НЕ(X))). Фітнес-функція становить собою частку пацієнтів з правильною класифікацією ступеня ризику СРСН (9):

$$E = \frac{1}{M} \cdot (M - \sum_{i=1}^M (|F_i - Y_i|)), \quad (9)$$

де  $M$  – кількість навчальних прикладів,  $F$  – отриманий результат класифікації,  $Y$  – бажаний результат класифікації.

3. Розглянутий підхід, як і у випадку НМ, не дозволяє логічно пояснювати отриманий результат. Тому запропоновано новий метод кодування особини для ГП, який дозволяє інтерпретувати булеву функцію у вигляді класифікаційних правил. Як і раніше, особина є деревом, але тепер це дерево є булевою функцією у диз'юнктивній нормальній формі (ДНФ). На рис. 4 наведено приклад дерева, що відповідає функції у ДНФ.

Дерево представлено чотирма правилами. Така репрезентація особини значно спрощує інтерпретацію результату. У вказаному прикладі розшифровка буде такою:

ЯКЩО виконується правило 1 АБО виконується правило 2 АБО виконується правило 3 АБО виконується правило 4, ТО результат 1, ІНАКШЕ результат 2.

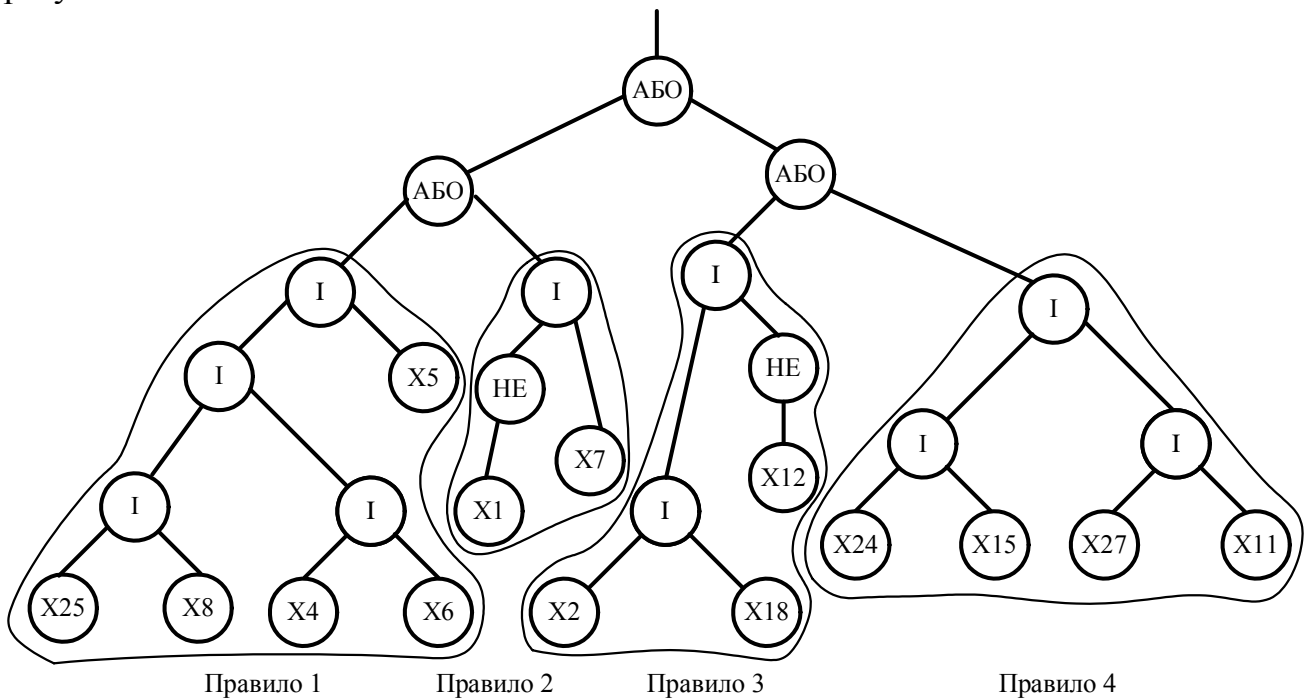


Рис. 4. Приклад дерева, що є булевою функцією у ДНФ

Тут правила 1-4 відповідають булевым функціям (10)-(13), а виконання правила означає, що значення відповідної функції дорівнює 1:

$$f_1(x_1, x_2, \dots, x_{176}) = x_{25}x_8 \cdot x_4x_6 \cdot x_5, \quad (10)$$

$$f_2(x_1, x_2, \dots, x_{176}) = \bar{x}_1x_7, \quad (11)$$

$$f_3(x_1, x_2, \dots, x_{176}) = x_2x_{18}\bar{x}_{12}, \quad (12)$$

$$f_4(x_1, x_2, \dots, x_{176}) = x_{24}x_{15} \cdot x_{27}x_{11}. \quad (13)$$

Відповідно все дерево можна представити виразом (14):

$$F(x_1, x_2, \dots, x_{176}) = x_{25}x_8 \cdot x_4x_6 \cdot x_5 \vee \bar{x}_1x_7 \vee x_2x_{18}\bar{x}_{12} \vee x_{24}x_{15} \cdot x_{27}x_{11}. \quad (14)$$

Якщо значення виразу (14) дорівнює 1, то це означає виконання конструкції ЯКЩО, тобто результат 1, не виконання – ІНАКШЕ, тобто результат 2.

У загальному вигляді дереву відповідає формула (15):

$$F(x_1, x_2, \dots, x_{176}) = \bigvee_{j=1}^m f_j(x_1, x_2, \dots, x_{176}), \quad (15)$$

де  $m$  – кількість класифікаційних правил. Якщо значення формули (15) дорівнює одиниці – то ступень ризику СРСН високий, а якщо нулю – низький.

У процесі роботи з медичними даними досить часто виникає ситуація, коли значення деяких параметрів невідомі. З метою мінімізувати втрату даних під час навчання й отримати можливість роботи при невідомих значеннях деяких факторів ризику запропоновано використовувати трічну логіку в булевих обчисленнях.

У табл. 2-4 наведено таблиці істинності для таких логічних функцій: І, АБО і НЕ. Символ «\*» означає стан, що може бути 0 або 1, але невідомо, що саме.

Таблиця 2  
Таблиця істинності функції І

N <sub>1</sub>	N <sub>2</sub>	І
0	0	0
0	1	0
1	0	0
1	1	1
*	0	0
*	1	*
0	*	0
1	*	*
*	*	*

Таблиця 3  
Таблиця істинності функції АБО

N <sub>1</sub>	N <sub>2</sub>	АБО
0	0	0
0	1	1
1	0	1
1	1	1
*	0	*
*	1	1
0	*	*
1	*	1
*	*	*

Таблиця 4  
Таблиця істинності функції НЕ

N <sub>1</sub>	НЕ
0	1
1	0
*	*

У четвертому розділі розроблено ЕС, яка є програмним комплексом, призначеним для надійного зберігання інформації, її обробки і виконання необхідних обчислень.

Розроблено структуру програмного забезпечення. Ядром експертної системи є підсистема добування знань, яка реалізована на основі методів ІАД, а також машина виводу. Відповідно до математичного забезпечення, розробленого в розділі 3, ці підсистеми реалізовані у вигляді декількох модулів.

Підсистема добування знань складається з таких програмних модулів:

- модуль формування і навчання НМ, де виконується навчання з учителем заданої користувачем архітектури НМ, на входи мережі подаються закодовані фактори ризику, на вихід – закодований результат; при успішному навчанні мережа запам'ятовується;
- модуль побудови класифікаційного дерева, де на основі навчальної вибірки виконується формування бінарного дерева;
- модуль побудови класифікаційних правил на основі ГП для класифікації ступеня ризику СРСН;
- модуль побудови класифікаційних правил за умови, що в навчальній вибірці значення деяких факторів ризику невідомі.

Машину виводу реалізовано у вигляді декількох модулів:

- модуль визначення ступеня ризику СРСН на основі класифікаційної таблиці, де виконується зчитування з БД інформації про наявність факторів ризику СРСН у пацієнта, та підсумовування вагових коефіцієнтів наявних факторів ризику;
- модуль визначення високого ступеня ризику СРСН з використанням НМ, де виконується моделювання роботи НМ, на входи якої подаються закодовані фактори ризику, на виході формується закодований результат – високий або низький ступінь ризику СРСН;

- модуль визначення ступеня ризику СРСН, що використовує класифікаційне дерево, отримане на основі ГП, та виконує розрахунок значення бінарного дерева, одиничне значення на виході інтерпретується як високий ступінь ризику СРСН, нульове – низький;
- модуль визначення ступеня ризику СРСН, що використовує класифікаційні правила, отримані на основі ГП;
- модуль визначення ступеня ризику СРСН, який використовує класифікаційні правила з можливістю роботи при невизначених значеннях деяких факторів за рахунок використання троїчної логіки в булевих обчисленнях, де класифікаційні правила отримані також з використанням ГП.

На рис. 5 представлено укрупнену блок-схему алгоритму, що реалізовує визначення ступеня ризику СРСН.

**У п'ятому розділі** наведено результати експериментальних досліджень функціонування системи на реальних медичних даних в умовах Донецького регіонального центру охорони материнства і дитинства і в науково-дослідному інституті медичних проблем сім'ї Донецького національного медичного університету ім. Горького. Порівняльний аналіз показав, що:

1. Розроблений метод класифікації ступеня ризику СРСН на чотири класи дозволяє:

- визначати ступінь ризику СРСН на ранніх термінах вагітності і розпочати профілактичне лікування ще до народження дитини, що раніше не було можливим;
- визначати ступінь ризику СРСН з точністю 93,33%, що підвищує достовірність визначення ступеня ризику на 10% порівняно з використовуваним раніше методом.

2. Проаналізовано результати вибору значимих факторів ризику СРСН, показано, що вибір значимих факторів виконано коректно. Сформовано НМ для визначення ступеня ризику СРСН з точністю 97,85%.

3. Експериментально визначено значення параметрів ГП для формування дерева, що виконує класифікацію на високий і низький ступінь ризику СРСН. Отримано класифікаційне дерево, що дозволяє з точністю 98,57% визначати ступінь ризику СРСН.

4. Експериментально визначено значення параметрів ГП для побудови класифікаційних правил. Отримано класифікаційні правила для визначення ступеня ризику СРСН. Проведено експерименти із застосуванням троїчної логіки в булевих обчисленнях. Показано, що до 5% невідомих факторів ризику на входах системи дозволяють класифікувати ступінь ризику СРСН з точністю до 90%. При цьому збільшення невідомих станів не збільшує помилку визначення ступеня ризику СРСН.

5. Експериментально показано, що вживання троїчної логіки дозволяє отримувати класифікаційні правила за умови, що деякі вхідні фактори невідомі. Отримано класифікаційні правила для визначення ступеня ризику СРСН при 1-5% невідомих значень деяких факторів ризику в навчальній вибірці.

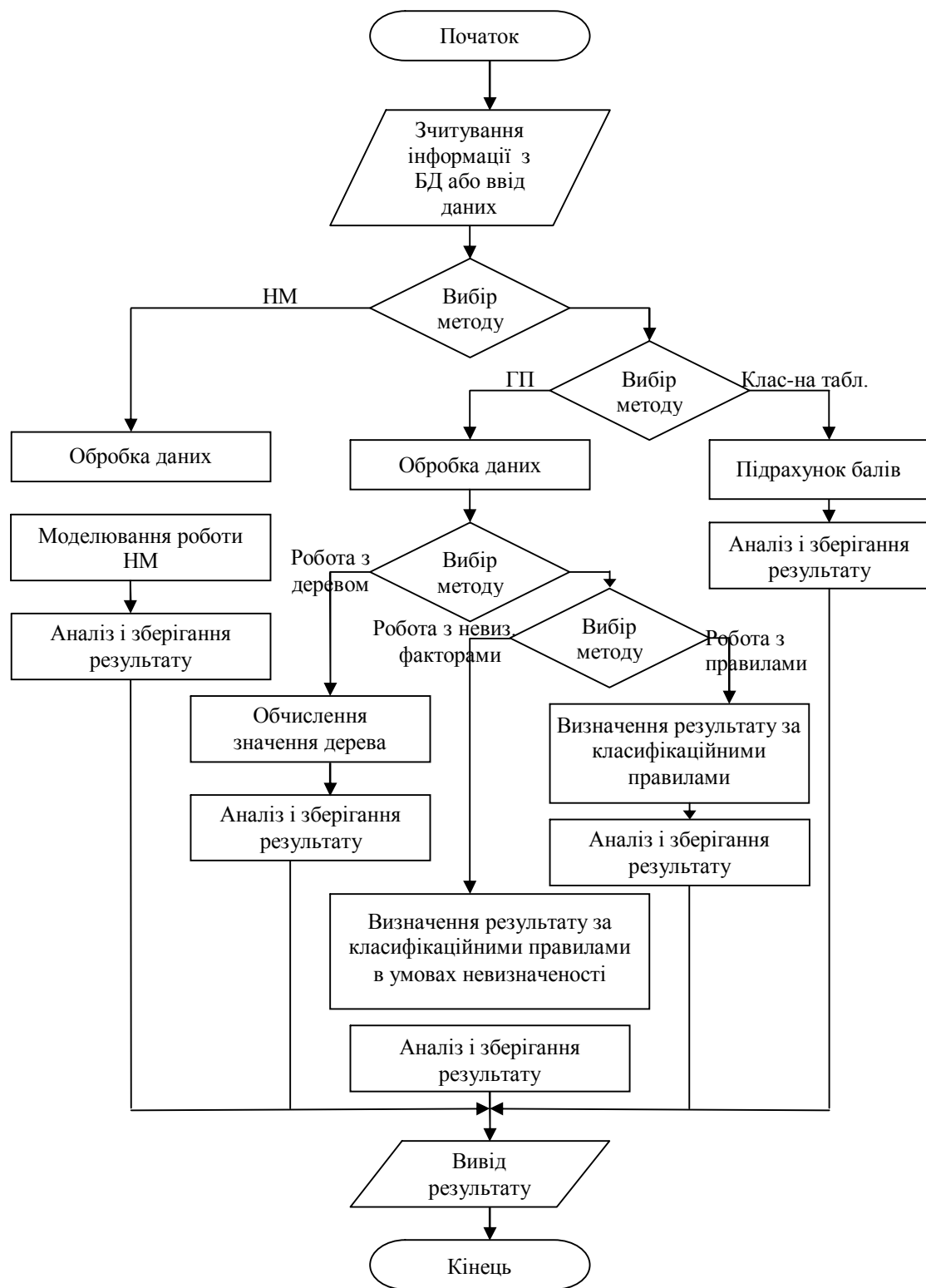


Рис. 5. Укрупнена блок-схема алгоритму, що реалізовує функціонування ЕС визначення ступеня ризику СРСН

**У висновках** викладено найбільш важливі наукові і практичні результати, отримані в дисертаційній роботі.

**У додатках** подано матеріали, що містять результати експериментів, проведених у роботі; екранні форми розробленого проекту; акти впровадження основних результатів дисертаційної роботи.



## ВИСНОВКИ

У дисертаційній роботі вирішено актуальну науково-практичну задачу розробки методів добування знань для медичної ЕС визначення ступеня ризику СРСН. Розроблена ЕС дозволяє підвищити достовірність визначення ступеня ризику СРСН і раніше розпочинати профілактичні заходи, направлені на його зниження за рахунок використання розроблених методів ІАД. Під час проведення досліджень отримано такі основні результати:

1. Вперше розроблено ЕС, призначену для визначення ступеня ризику СРСН. Створено інформаційну базу ЕС. Запропоновано структурну схему взаємодії програмних модулів. Розроблено алгоритми і реалізовано програмні модулі ЕС.

2. Виконано реалізацію гібридної системи на основі ГА і НМ, що дозволило досягти високої ефективності відбору інформативних факторів ризику СРСН шляхом регулювання співвідношення кількості факторів і помилки класифікації.

3. Проведено експерименти з вибору інформативних факторів ризику СРСН. Обрано інформативний вхідний набір даних, який удвічі менший за початковий. За допомогою кореляційного аналізу визначено значимість кожного окремого фактора ризику СРСН.

4. Розроблено метод класифікації ступеня ризику СРСН на чотири класи, який дозволив визначати ступінь ризику СРСН з точністю 93,33%, що підвищує достовірність на 10% порівняно з використаним раніше методом.

5. Розроблено метод визначення ступеня ризику СРСН на основі НМ. Побудовано НМ для визначення високого ступеня ризику СРСН з точністю 97,85%.

6. Вдосконалено деревоподібний спосіб кодування особини для булевої функції, що дозволило отримувати класифікаційні дерева для визначення ступеня ризику СРСН. Експериментально отримано класифікаційне дерево, що дозволяє з точністю 98,57% визначати ступінь ризику СРСН.

7. Вдосконалено спосіб кодування особини у вигляді дерева, представленого булевою функцією у ДНФ, що дозволило отримувати класифікаційні правила для визначення ступеня ризику СРСН. Експериментально отримано класифікаційні правила для визначення високого ступеня ризику СРСН.

8. Використано троїчну логіку в булевих обчисленнях класифікаційного дерева, що дозволило виконувати класифікацію при невизначених значеннях деяких факторів ризику.

9. Експериментально визначено значення параметрів ГП і отримано класифікаційні правила для визначення ступеня ризику СРСН в умовах невизначеності деяких даних.

10. Проведено випробування ЕС визначення ступеня ризику СРСН в умовах Донецького регіонального центру охорони материнства і дитинства і в науково-дослідному інституті медичних проблем сім'ї Донецького національного медичного університету ім. Горького. Встановлено, що

використання розробленої ЕС дозволяє: визначити необхідність вживання заходів, направлених на зниження ступеня ризику СРСН ще під час вагітності; підвищує достовірність визначення ступеня ризику СРСН на 10% у порівнянні з використовуваним раніше методом.

11. Результати роботи використано у Донецькому регіональному центрі охорони материнства і дитинства і в науково-дослідному інституті медичних проблем сім'ї Донецького національного медичного університету ім. Горького, у держбюджетних темах Д-11-04, Н-3-07, Д-11-07, а також у навчальному процесі кафедри «Автоматизовані системи управління» Донецького національного технічного університету.

## **СПИСОК ПРАЦЬ, ОПУБЛІКОВАНИХ АВТОРОМ ЗА ТЕМОЮ ДИСЕРТАЦІЇ**

1. Vasyaeva T.A. Neural network prediction of atherothrombotic stroke / Skobtsov Y.A., Vasyaeva T.A., Rodin Y.V. // *Advanced Computer Systems and Networks: Design and Application. Proceedings of 1-st International Conference ACSN – 2003, (Lviv, 24-26 September, 2003).* – Lviv, 2003. – P. 134-135.

2. Васяєва Т.О. Нейромережевий підхід до прогнозу результату хвороби Годжкіна / Скобцов Ю.О., Васяєва Т.О. // *Наукові праці Донецького національного технічного університету. Серія: Обчислювальна техніка та автоматизація.* – Донецьк, 2003. – Випуск 64.– С. 233-240.

3. Патент 8642 Україна, МПК<sup>7</sup>: А61В17/42. Спосіб прогнозування розвитку синдрому раптової смерті грудних дітей / Чайка В.К., Яковлева Е.Б., Тутов С.М., Васяєва Т.О.; Заявник і патентовласник Донецький державний медичний університет ім. Горького, Заява № 4200500814 від 31.01.2005. – Опубл. 15.08.2005. – Бюл. № 8.

4. Патент 9146 Україна, МПК<sup>7</sup>: А61В17/42. Спосіб прогнозування розвитку синдрому раптової смерті грудних дітей / Чайка В.К., Яковлева Е.Б., Тутов С.М., Васяєва Т.О.; Заявник і патентовласник Донецький державний медичний університет ім. Горького, Заява № 4200500807 від 31.01.2005. – Опубл. 19.09.2005.– Бюл. № 9.

5. Васяєва Т.О. Визначення групи ризику щодо синдрому раптової смерті немовляти / Васяєва Т.О. // *Наукові праці Донецького національного технічного університету. Серія: Обчислювальна техніка та автоматизація.* – Донецьк, 2005. – Випуск 88.– С. 82-86.

6. Васяєва Т.А. Выделение информативных факторов риска синдрома внезапной смерти грудных детей для формирования обучающей выборки при построении экспертной системы / Васяєва Т.А., Скобцов Ю.А. // *Вісник Хмельницького національного університету.* – Хмельницький, 2007. – №2 Т.1(90). – С. 77-79.

7. Васяєва Т.А. Подготовка данных при разработке медицинских экспертных систем / Скобцов Ю.А., Васяєва Т.А. // *Вісник Херсонського національного технічного університету.* – Херсон, 2007. – №4(27). – С.49-55.

8. Васяева Т.А. Применение искусственного интеллекта для вычисления информативности признаков в медицинских задачах / Скобцов Ю.А., Скобцов В.Ю., Васяева Т.А. // Компьютерные науки и информационные технологии. Тезисы докладов Международной научной конференции, посвященной памяти профессора А.М.Богомолова, (Саратов, 2-4 июля 2007). – Саратов, 2007. – С.110-112.

9. Васяева Т.А. Предобработка входной информации для построения и обучения экспертной системы прогнозирования синдрома внезапной смерти грудных детей / Васяева Т.А. // Наукові праці Донецького національного технічного університету. Серія: Обчислювальна техніка та автоматизація. – Донецьк, 2007. – Випуск 13(121). – С. 118-125.

10. Васяева Т.А. Формирование знаний для медицинских экспертных систем на основе генетического программирования / Скобцов Ю.А., Васяева Т.А., Лобачева М.В. // Информационные технологии и информационная безопасность в науке, технике и образовании «ИНФОТЕХ – 2007» Материалы международной научно-практической конференции, (Севастополь, 10-16 сентября 2007). – Севастополь, 2007. – Ч. 2. – С. 38-42.

11. Vasyaeva T.A. Diagnosis of SIDS using Genetic Programming. / Skobtsov Y.A., Vasyaeva T.A. // Advanced Computer Systems and Networks: Design and Application. Proceedings of 3-st International Conference ACSN – 2007, (Lviv, 20-22 September 2007). – Lviv, 2007. – P. 92-93.

12. Васяева Т.А. Извлечение знаний на основе генетических алгоритмов и генетического программирования / Скобцов Ю.А., Васяева Т.А. // Вісник Східноукраїнського національного університету ім. В. Даля. – Луганськ, 2008. – № 12 (130), Ч. 2. – С. 5-11.

13. Васяева Т.А. Эволюционный подход к формированию знаний для медицинских экспертных систем с учетом неопределенности данных / Скобцов Ю.А., Васяева Т.А. // Штучний інтелект. –Донецьк, 2008. – № 4. – С.631-637.

14. Васяева Т.А. Разработка экспертных систем медицинской диагностики с явным представлением продукционных правил на основе генетического программирования с учетом неопределенности данных / Скобцов Ю.А., Васяева Т.А. // Материалы IX международной научно технической конференции «Искусственный интеллект. Интеллектуальные системы – ИИ – 2008», (Кацивели 22-27 сентября 2008). – Донецк-Таганрог-Минск, 2008. – Т.2. – С.97-101.

15. Васяева Т.А. Разработка экспертных систем медицинской диагностики с явным представлением продукционных правил на основе ГП / Скобцов Ю.А., Васяева Т.А. // Тези міжнародної наукової конференції «Інтелектуальні системи прийняття рішень і проблеми обчислювального інтелекту (ISDMCI'2008)», (Євпаторія 19-23 травня 2008). – Херсон, 2008. – Т3, Ч1. – С. 68-71.

16. Васяева Т.А. Получение правил вывода на основе генетического программирования / Васяева Т.А. // Сучасні тенденції розвитку інформаційних технологій в науці, освіті та економіці: Матеріали II Всеукраїнської науково-

практичної конференції, (Луганськ, 8-10 квітня 2008). – Луганськ, 2008. – С. 19-21.

17. Васяева Т.А. Методы разработки экспертных систем медицинской диагностики / Скобцов Ю.А., Васяева Т.А. // Материалы международной научно-практической конференции «Информационные технологии и информационная безопасность в науке, технике и образовании (ИНФОТЕХ – 2009)» (Севастополь, 7-12 сентября 2009). – Севастополь, 2009. – С. 237-240.

## АНОТАЦІЯ

Васяєва Т.О. Нейромережеві та еволюційні методи аналізу даних у експертних системах медичної діагностики. – Рукопис.

Дисертація на здобуття наукового ступеня кандидата технічних наук за спеціальністю 05.13.06 – «Інформаційні технології». – Донецький національний університет, Донецьк, 2010.

У дисертаційній роботі вирішено актуальну наукову задачу розробки методів добування знань для медичної експертної системи (ЕС). Вперше розроблена ЕС, що дозволяє підвищити достовірність визначення ступеня ризику синдрому раптової смерті немовляти (СРСН) та визначати ступінь ризику на ранніх термінах вагітності за рахунок використання розроблених методів інтелектуального аналізу даних.

Виконано реалізацію гібридної системи на основі нейронних мереж та генетичних алгоритмів, що дозволило досягти високої ефективності відбору інформативних факторів ризику СРСН шляхом регулювання співвідношення кількості факторів і помилки класифікації.

Вдосконалено деревоподібний спосіб кодування особини для булевої функції, що дозволило отримувати класифікаційні дерева для визначення ступеня ризику СРСН. Експериментально отримано класифікаційне дерево, що дозволяє з точністю 98,57% визначати ступінь ризику СРСН.

Вдосконалено спосіб кодування особини у вигляді дерева, представленого булевою функцією у ДНФ, що дозволило отримувати класифікаційні правила для визначення ступеня ризику СРСН. Експериментально отримано класифікаційні правила для визначення високого ступеня ризику СРСН.

Використано троїчну логіку в булевих обчисленнях класифікаційного дерева, що дозволило виконувати класифікацію при невизначених значеннях деяких факторів ризику.

Ключові слова: експертна система, нейронні мережі, генетичні алгоритми, генетичне програмування.

## АННОТАЦИЯ

Васяева Т.А. Нейросетевые и эволюционные методы анализа данных в экспертных системах медицинской диагностики – Рукопись.

Диссертация на соискание ученой степени кандидата технических наук по специальности 05.13.06 – «Информационные технологии». – Донецкий национальный университет, Донецк, 2010.

В диссертационной работе решена актуальная научная задача разработки методов извлечения знаний для медицинской экспертной системы (ЭС) определения степени риска синдрома внезапной смерти грудного ребенка (СВСГР). Впервые разработана ЭС, позволяющая повысить достоверность определения степени риска СВСГР и определять степень риска на раннем сроке беременности и после рождения ребенка за счет использования разработанных методов интеллектуального анализа данных.

Получил дальнейшее развитие эволюционный метод отбора информативных факторов риска, за счет разработанной фитнес-функции, что позволило достигнуть высокой эффективности отбора информативных факторов риска СВСГР путем регулирования соотношения количества факторов и ошибки классификации.

Проведены эксперименты по выбору информативных факторов риска СВСГР, что позволило отобрать информативный входной набор данных, который в два раза меньше исходного. С помощью корреляционного анализа определена значимость каждого отдельного фактора риска СВСГР.

Разработан метод классификации степени риска СВСГР на четыре класса, который позволил определять степень риска СВСГР с точностью 93,33%, что повышает достоверность определения степени риска на 10% по сравнению с использованным ранее методом.

Разработан метод определения степени риска СВСГР с помощью нейронных сетей (НС). Определена и обучена НС для определения высокой степени риска СВСГР с точностью 97,85%.

Усовершенствован древовидный способ кодирования особи для булевой функции, что позволило получать классификационные деревья для определения степени риска СВСГР. Экспериментально получено классификационное дерево, позволяющее с точностью 98,57% определять степень риска СВСГР.

Усовершенствован способ кодирования особи в виде дерева, реализующего булеву функцию в дизъюнктивной нормальной форме, что позволило получать классификационные правила для определения степени риска СВСГР. Экспериментально получены классификационные правила для определения высокой степени риска СВСГР.

Предложено использовать в булевых вычислениях классификационного дерева троичную логику, что позволило выполнять классификацию при неопределенных значениях некоторых факторов риска. Экспериментально получены классификационные правила для определения степени риска СВСГР в условиях неопределенности некоторых данных.

Проведены испытания ЭС определения степени риска СВСГР в Донецком региональном центре охраны материнства и детства и в научно-исследовательском институте медицинских проблем семьи Донецкого национального медицинского университета им. Горького Установлено, что применение разработанной ЭС позволяет еще во время беременности

определять степени риска СВСГР; повышает достоверность определения степени риска на 10% по сравнению с использованным ранее методом.

Результаты работы использованы в Донецком региональном центре охраны материнства и детства и в научно-исследовательском институте медицинских проблем семьи Донецкого национального медицинского университета им. Горького в госбюджетных темах Д-11-04, Н-3-07, Д-11-07, а также в учебном процессе кафедры «Автоматизированные системы управления» Донецкого национального технического университета.

Ключевые слова: экспертная система, нейронные сети, генетические алгоритмы, генетическое программирование.

## ABSTRACT

Vasyaeva T.O. Neural networks and evolutionary methods of data analysis in expert systems for medical diagnostics. – Manuscript.

Candidate's of technical sciences thesis on specialty 05.13.06 – «Information technologies» – Donetsk National University, Donetsk, 2010.

In this thesis an important scientific problem was solved. The problem of development of the methods of data mining for medical expert systems (ES) that is considered. The ES that has been developed allows to detect sudden infant death syndrome (SIDS) risk range with using methods, algorithms and software of information procession with using modern computer hardware. This allows to increase detecting of SIDS risk range reliability and to begin preventive actions earlier.

For the first time an expert system devoted to SIDS risk range detecting has been designed, which allowed to increase reliability of SIDS risk range detecting and to detect the risk range even while the mother is pregnant or right after the child has been born.

The method of informative risk factors selection for a genetic algorithm has received future development, by implementing fitness-function developed. This allowed to achieve a high performance of selection informative SIDS risk factors by regulating factors number and classification errors correlation.

A tree-like encoding approach of chromosome encoding for a boolean function has been improved. This allowed to deduce classification trees for detecting SIDS risk range.

An approach of encoding a chromosome as a tree, which embodies the boolean function in a disjunctive-normal form has been improved. This allowed to deduce classification rules for detecting SIDS risk range.

For the first time a ternary logic in boolean calculations of classification tree was proposed. This allowed to conduct classification even when the risk factors are fuzzy.

Keywords: expert systems, neural networks, genetic algorithms, genetic programming.