

УДК 004.9

## **ВИРІШЕННЯ БАГАТОКРИТЕРІАЛЬНОЇ ЗАДАЧІ ОПТИМІЗАЦІЇ З ВИКОРИСТАННЯМ ГЕНЕТИЧНОГО АЛГОРИТМУ ТА МЕТОДУ АНАЛІЗУ ІЄРАРХІЙ**

В.О. Бабенко, О.К. Носовець

*Національний технічний університет України  
«Київський політехнічний інститут імені Ігоря Сікорського»  
vbabenko2191@gmail.com, o.nosovets@gmail.com*

Розглянуто розробку алгоритму для вирішення багатокритеріальної задачі оптимізації з використанням генетичного алгоритму та методу аналізу ієрархій. В роботі наведено етапи виконання алгоритму та результати його використання на прикладі підбору варіанта лікування пацієнтів зі складними вродженими вадами серця. Моделювання виконувалось методом групового урахування аргументів (МГУА) з використанням програмного забезпечення *GMDH Shell DS*.

*Ключові слова: багатокритеріальна оптимізація, генетичні алгоритми, метод аналізу ієрархій, метод групового урахування аргументів.*

The development of algorithm for solving multi-objective optimization problem using genetic algorithm with analytic hierarchy process and the results of its use in selecting a treatment option for patients with complex congenital heart defects are considered. Modeling was performed using group method of data handling (GMDH) in *GMDH Shell DS* software.

*Keywords: multi-objective optimization, genetic algorithms, analytic hierarchy process, group method of data handling.*

Рассмотрено разработку алгоритма для решения многокритериальной задачи оптимизации с использованием генетического алгоритма и метода анализа иерархий, а также результаты его использования на примере подбора варианта лечения пациентов с тяжелыми врожденными пороками сердца. Моделирование выполнялось методом группового учёта аргументов (МГУА) с использованием программного обеспечения *GMDH Shell DS*.

*Ключевые слова: многокритериальная оптимизация, генетические алгоритмы, метод анализа иерархий, метод группового учета аргументов.*

**Вступ.** Задача одночасної оптимізації двох або більше цільових функцій в заданій області визначення є відомою, та зустрічається в багатьох областях науки, техніки та економіки. Даний вид задач широко застосовується і в сфері біології та медицини. Методи вирішення подібних задач можна використати, в тому числі, і для знаходження вірогідного лікування пацієнтів з вродженими вадами серця, де в ролі цільових функцій виступають післяопераційні ускладнення.

Дана стаття розглядає вирішення подібної задачі шляхом знаходження екстремуму однієї цільової функції, яка була отримана за допомогою згортки  $n$ -ої кількості функцій методом аналізу ієрархій Сааті.

### **1. Постановка задачі**

Нехай існує база клінічних даних пацієнтів з вродженими вадами серця, в якій містяться:

а)  $X$  – матриця змінних початкового стану пацієнта  $x_1, \dots, x_n$ , де  $n$  – кількість змінних в матриці  $X$ ;

б)  $Y$  – матриця змінних лікування пацієнта  $y_1, \dots, y_m$ , де  $m$  – кількість змінних в матриці  $Y$ ;

в)  $Z$  – матриця змінних стану пацієнта після лікування  $z_1, \dots, z_o$ , де  $o$  – кількість змінних в матриці  $Z$ ;

Матрицю  $Z$  можна представити як:  $Z = XY$ . Таким чином, виникає задача оптимізації, де необхідно знайти  $Y$  при відомому  $X$  [1].

При вирішенні подібної задачі виникають наступні проблеми:

1) необхідність одночасної оптимізації змінних  $z_1, \dots, z_o$ .

2) знаходження змінних  $y_1, \dots, y_m$ , які дають глобальний екстремум (задача класу складності  $NP$ );

Перший пункт можна вирішити за допомогою методу аналізу ієрархій, який дає змогу «згорнути» усі змінні в одну. Таким чином залишається знайти екстремум лише однієї функції.

Другий пункт можна вирішити за допомогою використання генетичного алгоритму, який є евристичним алгоритмом пошуку екстремуму.

Таким чином, поєднання методу аналізу ієрархій та генетичного алгоритму дозволяє створити алгоритм, який вирішує задачу одночасної оптимізації більше, ніж однієї змінної.

## 2. Мета дослідження

Створення алгоритму для вирішення задачі багатокритеріальної оптимізації на прикладі реальної медичної проблеми.

## 3. Аналіз літературних даних

Для вирішення задачі одночасної оптимізації критеріїв було обрано методи багатокритеріального прийняття рішень. Найбільш відомими з них є: метод аналізу ієрархій [2] та метод аналізу мереж, які були розроблені американським математиком Сааті, метод *ELECTRE* (ELimination and Choice Expressing REality – виключення і вибір, які виражають реальність), модель Брауна-Гібсона, метод *Best Worst* (від найкращого до найгіршого) та метод *VIKOR*. Проаналізувавши дані методи, було обрано метод аналізу ієрархій, оскільки він є більш простим в реалізації і дозволяє оптимізувати достатньо велику кількість критеріїв. Підхід Сааті більш детально описано в роботах [3-4].

Для вирішення задачі знаходження лікувальної стратегії було обрано принцип генетичних алгоритмів [5-8], оскільки вони дають змогу вирішити  $NP$ -повну задачу, де присутня велика кількість рішень, за короткий час, з використанням принципів еволюції.

#### **4. Характеристика клінічного матеріалу**

Для тестування розробленого алгоритму, обрано задачу вибору оптимальної лікувальної стратегії для пацієнтів зі складними вродженими вадами серця, застосування якої дозволить знизити ризик виникнення ранніх післяопераційних ускладнень.

В роботі було використано клінічні дані 128 пацієнтів з вродженими вадами серця. Вік пацієнтів склав від 3 до 28 років. З них 70 (54,7%) чоловіків та 58 (45,3%) жінок. Дослідні періоди включали в себе: доопераційний, операційний та ранній післяопераційний періоди. Проаналізовані клініко-морфологічні характеристики хворих, дані лабораторних та інструментальних досліджень, характеристики операційного та післяопераційного етапів. Усього база даних містила 313 показників, з яких було відібрано значимі для поставленої мети дослідження:

*1) змінні початкового стану пацієнта (12 змінних):*

Вік в місяцях ( $x_{101}$ ); діаметр лівої легеневої артерії ( $x_{102}$ ); індекс Наката – площа поперечного перерізу гілок легеневої артерії до площі поверхності тіла ( $x_{103}$ ); кінцево-діастолічний тиск в лівому шлуночку ( $x_{104}$ ); тиск в легеневій артерії ( $x_{105}$ ); кінцево-діастолічний індекс ( $x_{106}$ ); опір легеневих судин ( $x_{107}$ ); бали по шкалі факторів ризику для операції тотального кавопульмонального сполучення ( $x_{108}$ ); z-індекс – оцінка, яка виражається як відхилення від середнього значення в одиницях стандартного відхилення ( $x_{109}$ ); наявність гіпертрофії ( $x_{110}$ ); наявність морфології правого шлуночку ( $x_{111}$ ); пацієнту проводили процедуру Рашкінда ( $x_{112}$ ).

*2) змінні операційного лікування (9 змінних):*

Загальний об'єм альбуміну – анаболічного лікарського засобу, який підтримує онкотичний тиск (oncotic pressure), збільшує об'єм циркулюючої крові і підвищує артеріальний тиск, та використовується при операційному шоці ( $x_{201}$ ); рівень добутаміну ( $x_{202}$ ); рівень адреналіну ( $x_{203}$ ); діаметр вшитого кондуїту ( $x_{204}$ ); усунення коарктації аорти ( $x_{205}$ ); процедура Damus-Kaye-Stansel ( $x_{206}$ ); пластика атріовентрикулярних клапанів під час операції тотального кавопульмонального сполучення ( $x_{207}$ ); пластика гілок легеневої артерії під час операції тотального кавопульмонального сполучення ( $x_{208}$ ); процедура Кавашима з відведенням печінкової вени ( $x_{209}$ ).

*3) змінні післяопераційного стану пацієнта (ранні післяопераційні ускладнення, 9 змінних):*

Ексудація більше 14 днів ( $x_{301}$ ); наявність усіх післяопераційних порушень ритму ( $x_{302}$ ); наявність плевриту ( $x_{303}$ ); плікація діафрагми ( $x_{304}$ ); порушення мозкового кровообігу ( $x_{305}$ ); наявність тромбозу ( $x_{306}$ ); наявність хілотораксу ( $x_{307}$ ); наявність повної атріовентрикулярної блокади ( $x_{308}$ ); дисфункція синусового вузла ( $x_{309}$ ).

Змінні післяопераційного стану пацієнта були відібрані за допомогою лікаря-експерта. Вони виступали в ролі вихідних змінних, в той час, як змінні початкового стану і операційного лікування – в ролі вхідних. Для уникнення проблеми мультиколінеарності було застосовано метод рекурсивного відбору ознак (*Recursive Feature Elimination*) [9].

Алгоритм відбору найбільш значимих вхідних змінних виглядав наступним чином: для вихідних змінних розраховувалися моделі класифікації на основі вхідних за допомогою алгоритму випадкового лісу (*Random Forest*) [10], після чого змінна з найменшим значенням «важливості» виключалася з аналізу; дана операція повторювалася, поки не було знайдено моделей з найбільшими значеннями точності, чутливості та специфічності.

### 5. Опис алгоритму

Загальний алгоритм вирішення багатокритеріальної задачі оптимізації представлено на рис. 1.

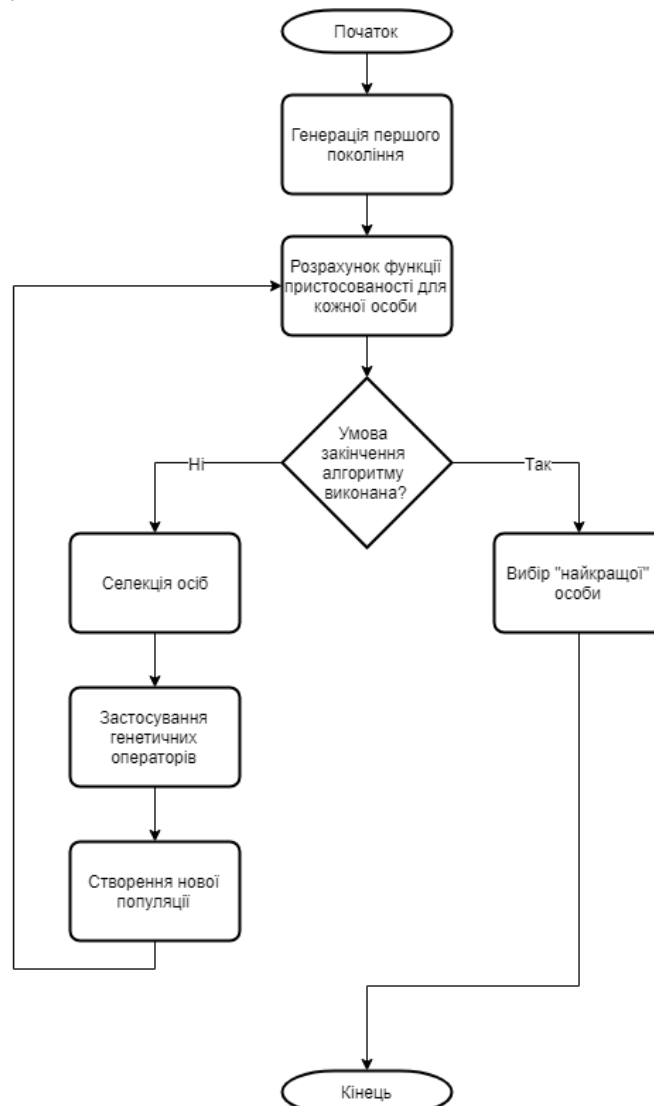


Рис. 1. Комплексний алгоритм знаходження вірогідного операційного лікування

Задача дослідження полягала в знаходженні комбінації змінних операційного лікування, яка забезпечить «оптимальний» стан пацієнта в ранньому післяопераційному періоді. Під «оптимальним» мається на увазі стан пацієнта з мінімальною кількістю післяопераційних ускладнень. Через велику кількість можливих комбінацій, пошук необхідного рішення займає об'ємну кількість часу, та вимагає значних ресурсів. При цьому час вирішення збільшується пропорційно збільшенню змінних операційного лікування. Запропонований алгоритм вирішує дану проблему наступним шляхом:

1. Випадковим чином формується вибірка із  $N$  «осіб» (хромосом) [5-6]. Кожна особа є масивом із 9 значень змінних операційного лікування, які лежать в наступних проміжках:  $x_{201} \in [1; 1000]$ ;  $x_{202} \in [1; 250]$ ;  $x_{203} \in [5; 60]$ ;  $x_{204} \in [1; 5]$ ;  $x_{205} \in [1; 2]$ ;  $x_{206} \in [1; 2]$ ;  $x_{207} \in [1; 2]$ ;  $x_{208} \in [1; 2]$ ;  $x_{209} \in [1; 2]$ .

2. Для кожної особи розраховується функція згортки, яка в даному алгоритмі виступає в ролі функції пристосованості, і виглядає наступним чином:

$$\Phi_{\text{згортки}} = -0,353 \cdot x_{301} - 0,177 \cdot x_{302} - 0,118 \cdot x_{303} - 0,088 \cdot x_{304} - 0,071 \cdot x_{305} - 0,059 \cdot x_{306} - 0,050 \cdot x_{307} - 0,044 \cdot x_{308} - 0,039 \cdot x_{309},$$

де  $x_{301}, \dots, x_{309}$  – ранні післяопераційні ускладнення, які виступають в ролі критеріїв.

Дану функцію було отримано за допомогою методу аналізу ієрархій, шляхом проведення парного порівняння критеріїв, з використанням реляційної шкали чисел Сааті та врахуванням аксіом пов'язаності, гомогенності та синтезу [2-4]. Результати попарного порівняння представлені в таблиці 1.

Таблиця 1. Попарне порівняння критеріїв

	$x_{301}$	$x_{302}$	$x_{303}$	$x_{304}$	$x_{305}$	$x_{306}$	$x_{307}$	$x_{308}$	$x_{309}$
$x_{301}$	1	2	3	4	5	6	7	8	9
$x_{302}$	0.5	1	1.5	2	2.5	3	3.5	4	4.5
$x_{303}$	0.33	0.67	1	1.33	1.67	2	2.33	2.67	3
$x_{304}$	0.25	0.5	0.75	1	1.25	1.5	1.75	2	2.25
$x_{305}$	0.2	0.4	0.6	0.8	1	1.2	1.4	1.6	1.8
$x_{306}$	0.167	0.33	0.5	0.67	0.83	1	1.167	1.33	1.5
$x_{307}$	0.143	0.286	0.429	0.571	0.714	0.857	1	1.143	1.286
$x_{308}$	0.125	0.25	0.375	0.5	0.625	0.75	0.875	1	1.125
$x_{309}$	0.11	0.22	0.33	0.44	0.56	0.67	0.78	0.89	1

Після порівняння, для кожного критерію розраховується його власний вектор (ваговий коефіцієнт) [2-4]. Результати розрахунків представлені в таблиці 2.

Таблиця 2. Вагові коефіцієнти критеріїв

Критерій	Середнє геометричне	Ваговий коефіцієнт
$x_{301}$	4.147	0.353
$x_{302}$	2.074	0.177
$x_{303}$	1.382	0.118
$x_{304}$	1.037	0.088
$x_{305}$	0.829	0.071
$x_{306}$	0.691	0.059
$x_{307}$	0.592	0.05
$x_{308}$	0.518	0.044
$x_{309}$	0.461	0.039

Критерії є бінарними змінними, тому для них були побудовані моделі класифікації за допомогою покрокового змішаного алгоритму МГУА [11], реалізованого в програмному забезпеченні *GMDH Shell DS*. Приклад однієї з побудованих моделей класифікації представлено нижче:

$$\begin{aligned}
 x_{301} = & -0.382 + \frac{0.001 \cdot x_{201}}{x_{111}} - \frac{0.007 \cdot x_{106}}{x_{202}} - 0.013x_{105} \cdot x_{109} + \frac{0.011x_{104}}{x_{208}} + \\
 & + \frac{0.452 \cdot x_{112}}{x_{111}} + \frac{1.506 \cdot x_{109}}{x_{203}} - \frac{0.027 \cdot x_{101}}{x_{102}} + \frac{17.16}{x_{106} \cdot x_{204}} + \frac{0.896}{x_{112} \cdot x_{208}} - \\
 & - \frac{0.023 \cdot x_{103}}{x_{106}} .
 \end{aligned}$$

Кожна модель була оцінена за показниками чутливості, специфічності та точності (загальному відсотку правильно передбачених подій), розрахованими на навчальній, екзаменаційній та перевірочній вибірках. Навчальна вибірка складала 70% від всіх спостережень, екзаменаційна – 20%, перевірочна – 10%. Оцінки якості кожної моделі наведені в табл. 3.

Таблиця 3. Оцінки якості моделей

Модель	Вибірка								
	Навчальна			Екзаменаційна			Перевірочна		
	Чутл.	Спеціф.	%	Чутл.	Спеціф.	%	Чутл.	Спец.	%
$x_{301}$	0.905	1	91.8	0.932	1	94.2	0.667	0.3	38.5
$x_{302}$	1	0.884	89.8	1	0.872	88.5	0.875	0.8	84.6
$x_{303}$	1	0.96	96.2	1	0.979	98.1	0.167	0.857	53.8
$x_{304}$	1	0.971	97.3	1	0.918	92.3	0	1	92.3

Продовження таблиці 3.

Модель	Вибірка								
	Навчальна						Навчальна		
	Чутл.		Чутл.		Чутл.		Чутл.		Чутл.
$x_{305}$	1	0.954	95.6	1	0.96	96.2	0	0.727	61.5
$x_{306}$	1	0.896	90.1	1	0.92	92.3	–	0.769	76.9
$x_{307}$	1	0.993	99.3	–	0.981	98.1	0	0.818	69.2
$x_{308}$	1	0.993	99.3	1	0.98	98.1	0.5	1	92.3
$x_{309}$	1	0.993	99.3	1	0.98	98.1	0	1	76.9

3. Після отримання функцій пристосованості перевіряється умова закінчення алгоритму. В даному випадку умовою є знаходження *максимуму* значення функції пристосованості. В разі виконання умови обирається «найкраща» особа, тобто необхідне рішення.

В протилежному випадку виконуються «класичні» методи генетичних алгоритмів [5-6]:

- Селекція осіб рулеточним відбором. Селекція полягає в тому, що формується  $N$ -а кількість пар осіб, які стають «батьками» осіб наступного покоління. Їх відбирають «запусками» рулетки, де для кожної особи виділено по одному сектору колеса. Розмір сектору пропорційний ймовірності попадання в нову популяцію, таким чином особи з більш високою функцією пристосованості будуть вибиратись частіше.

- Після того, як були обрані пари «батьків», відбувається їхнє схрещення (кросовер), таким чином формуючи особи нового покоління. Принцип схрещення полягає в тому, що всередині батьків визначається точка розриву, яка ділить їх на дві частини. Нова особа буде містити в собі першу частину від «батька», і другу частину від «матері». Схематично це показано на рис. 2.

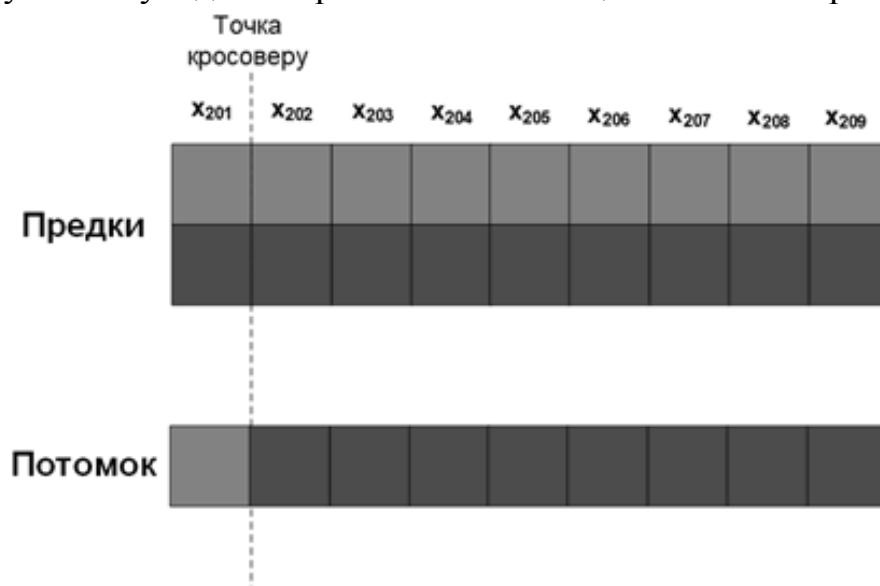


Рис. 2. Принцип схрещення

Таким чином формується нове покоління осіб. У випадку, коли всі особи в популяції є однаковими, схрещення втрачає сенс. Тоді застосовується мутація, яка полягає в тому, що  $N-1$  особи замінюються на абсолютно нові, випадково згенеровані особи.

4. Кроки 2-3 повторюються до тих пір, поки не буде виконана умова закінчення алгоритму.

## 6. Результати перевірки алгоритму

Для створення програмного забезпечення були використані мови програмування *Python* (для реалізації алгоритму) та *Java* (для створення *UI/UX* додатку). Щоб перевірити працездатність алгоритму була взята вибірка із 10 пацієнтів. Основними критеріями були час знаходження рішення та кількість виконаних ітерацій алгоритму (рис. 2). Результати тестування наведені в таблиці 4.

Таблиця 4. Результати тестування

Пацієнт №	Середній час знаходження рішення, с	Середня кількість ітерацій для знаходження
1	1.69	71
2	0.28	12
3	5.5	227
4	3.16	130
5	9.41	363
6	3.84	141
7	11.39	440
8	2.25	72
9	3.81	140
10	15.5	643

На рис. 3 зображено приклад зміни найкращого значення функції згортки при виконанні алгоритму. На ньому видно, що функція прямує до -1. Це пов'язано перш за все з тим, що критерії можуть давати лише два значення: 1, тобто немає ускладнення, і 2, тобто є. Так як 1 краще ніж 2, відповідно в формулі функції згортки вагові коефіцієнти критеріїв мають від'ємні значення, а максимально допустиме значення, яке може набути функція, дорівнює -1.

На рис. 4 представлено вигляд програми після виконання алгоритму для знаходження лікувальної стратегії пацієнтові №1 (таблиця 4). На даному зображенні показано, що варіант лікування, який знайшов алгоритм, вірогідно дасть на виході лише два ускладнення із дев'яти можливих. При цьому в реальному житті у пацієнта, при його лікуванні, виникло сім ускладнень із дев'яти.



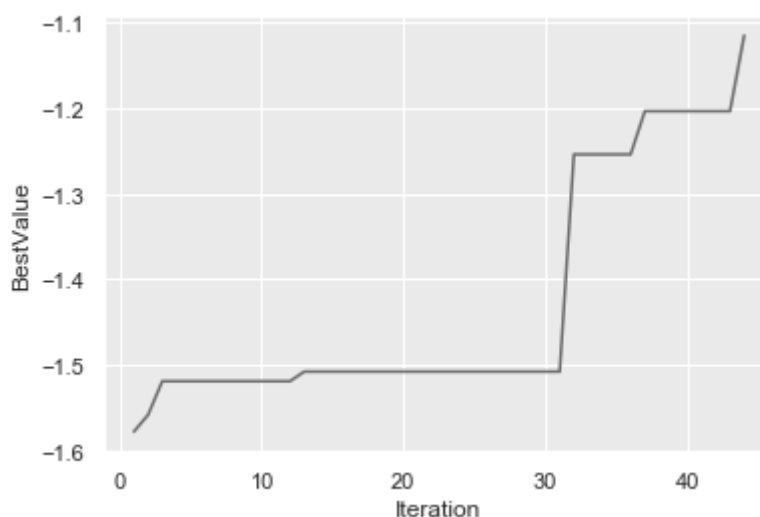


Рис. 3. Знаходження найкращого значення функції згортки

Оскільки існує деяка кількість рішень, які дають однаковий результат на виході, програма знаходить 5 можливих варіантів лікування. Право вибору стратегії залишається за лікарем.

Варіант лікування	Загальний об'єм альбуміну	Рівень добутаміну	Рівень адреналіну	Кондуїт вшитий	Усунення коарктації аорти	Процедура DKS	Пластика АВ-клапанів під час ТКПС	Пластика гілок ЛА під час ТКПС	Процедура Кавашима з відведенням ПВ
<input checked="" type="radio"/> Перший варіант лікування	98.19	41.22	13	RPA	Ні	Ні	Ні	Ні	Так
<input type="radio"/> Другий варіант лікування									
<input type="radio"/> Третій варіант лікування									
<input type="radio"/> Четвертий варіант лікування									
<input type="radio"/> П'ятий варіант лікування									

Додати лікування в базу даних

Результати лікування	Так/Ні
Екссудат більше 14 днів	Ні
Всі післяопераційні порушення ритму	Ні
Плеврит в ранньому періоді	Ні
Пликація діафрагми	Ні
Порушення мозкового кровообігу	Так
Тромбоз	Ні
Хілоторакс	Ні
Повна АВ-блокада	Так
Дисфункція синусового вузла	Ні

Рис 4. Результати виконання програми

### Висновки.

Дослідження дали наступні результати:

- Створено алгоритм для знаходження рішень багатокритеріальної задачі оптимізації. Даний алгоритм використовує метод аналізу ієрархій Сааті для одночасної оптимізації множинності критеріїв, а також принципи генетичних алгоритмів для швидкого знаходження рішення. Для моделювання використовувався покроковий змішаний алгоритм МГУА на трьох вибірках. На

екзаменаційній вибірці точність моделей варіюється в межах від 88,5% до 99,1%. Проте на перевіірочній вибірці результати набагато гірші, що скоріш за все пов'язано з малим розміром загальної вибірки даних.

• Для алгоритму було створено програмне забезпечення з використанням мов програмування *Python* та *Java*. Результати тестування показали, що для знаходження рішення в середньому необхідно 5,683 с часу та 224 ітерацій циклу.

Вирішення подібної задачі може дати поштовх для використання генетичних алгоритмів в інших видах задач, наприклад, для моделювання, де також можна застосувати метод аналізу ієрархій для одночасної оптимізації метрик точності моделей.

## Література

1. Nastenko I. Optimal Complexity Models in Individual Control Strategy Task for Objects that Cannot be Re-trialed / I. Nastenko, V. Pavlov, O. Nosovets, K. Zelensky, O. Davidko, O. Pavlov // Lviv, Ukraine : 2019 IEEE 14th International Conference on Computer Sciences and Information Technologies (CSIT). — 2019. — P. 207–210.
2. Saaty T. L. Decision making for leaders: the analytic hierarchy process for decisions in a complex world / T. L. Saaty. — RWS Publications, 1990. — 292 p.
3. Dyer J. S. Remarks on the analytic hierarchy process / J. S. Dyer // Management Science. — 1990. — Vol. 36, No. 3. — P. 247–399.
4. Ларичев О. И. Теория и методы принятия решений, а также хроника событий в волшебных странах / О. И. Ларичев. — Москва : Логос, 2002. — 393 p.
5. Goldberg D. E. Genetic algorithms in search, optimization & machine learning / D. E. Goldberg. — Addison-Wesley Publishing Company, Inc., 1989.
6. Mashwani W. K. Comparison of evolutionary algorithm over multiobjective optimization problems / W. K. Mashwani // P. 13–18.
7. Pfeiffer J. Reference point based multi-objective algorithms for group decisions / J. Pfeiffer, U. Golle, F. Rothlauf // Working Papers in Information Systems and Business Administration Reference Point. — 2008. — P. 9.
8. Whitley D. A genetic algorithm tutorial / D. Whitley. — 1994. — 37 p.
9. Guyon I. Feature extraction: foundations and applications / I. Guyon, S. Gunn, M. Nikravesh, L. A. Zadeh. — Springer Science & Business Media, 2006. — 778 p.
10. Силен Д. Основы data science и big data. python и наука о данных / Д. Силен, А. Мейсман, М. Али. — Питер, 2017. — 336 p.
11. Пидошва Е. А. Основные принципы метода группового учета аргументов и его перспективы / Е. А. Пидошва, А. Б. Иващенко // Vol. 9. — P. 4