

визначатися мінімумом цього ж логарифма, але зі знаком «-», що відповідає умові

$$\sup\{\ln F(\bar{X}/\Theta_k)\} \Leftrightarrow \inf\{\ln F^{-1}(\bar{X}/\Theta_k)\}. \quad (3.25)$$

Максимум $\ln F(\bar{X}/\Theta_k)$ еквівалентний за визначенням мінімуму $\ln F^{-1}(\bar{X}/\Theta_k)$, де перетворений логарифм матиме вигляд

$$\ln F^{-1}(\bar{X}/\Theta_k) = \frac{p}{2} \ln(2\pi) + \sum_{i=1}^p \ln \sigma_i^{(k)} + \frac{1}{2} \sum_{i=1}^p \left(\frac{x_i - \bar{x}_i^{(k)}}{\sigma_i^{(k)}} \right)^2. \quad (3.26)$$

Останній доданок правої частини виразу (3.26) є лінійною функцією геометричної відстані між двома точками в p -мірному просторі інформативних ознак. Перша точка відноситься до ідентифікованого об'єкту, а друга – до нормованого (по дисперсії) центру точок, що належать до класу з номером k . Можна показати [80, 79], що така відстань еквівалентна міжфункціональній відстані $\rho_L(\cdot)$ (підрозділ 2.6 розділу 2) для порівнюваних поліноміальних функцій, значення яких пронормовані по дисперсіям. Таке нормування не змінює властивостей міжфункціональної відстані, оскільки є лінійною операцією, але дозволяє використовувати відстань $\rho_L(\cdot)$ як еквівалент геометричної відстані.

Для всіх $k = \overline{1, m}$ лінійно перетворимо $\xi^{(k)} = \ln \{F^{-1}(\bar{X}/\Theta_k)\}$, прибравши константу $\frac{p}{2} \ln(2\pi)$ і помноживши отриману різницю на 2:

$$\hat{\xi}^{(k)} = \left(\xi^{(k)} - \frac{p}{2} \ln(2\pi) \right) 2. \quad (3.27)$$

Введемо позначення

$$\hat{x}_i^{(k)} = \frac{x_i - \bar{x}_i^{(k)}}{\sigma_i^{(k)}}. \quad (3.28)$$

З урахуванням введених додаткових перетворень і позначень отримаємо остаточно мінімізований вираз для цільової функції мінімізації

$$\hat{\xi}^{(k)} = \sum_{i=1}^p \left[\ln \left(\sigma_i^{(k)} \right)^2 + \hat{x}_i^{(k)} \right], \quad (3.29)$$

для якої зберігається умова

$$\inf\{\xi^{(k)}\} \Leftrightarrow \inf\{\hat{\xi}^{(k)}\}. \quad (3.30)$$

Слід зазначити, що статистика $\hat{\xi}^{(k)}$ виразу (3.29) є функцією числових характеристик для інформативних ознак, середнє і дисперсія яких, в свою чергу, визначається номером k біологічного стану, що діагностується. Екстремум даної статистики, як показує вираз (3.30), дозволяє визначити конкретне значення номера k , тобто фактично ідентифікувати модель динаміки глікемії.

3.7 Висновки по розділу

1. Розроблено метод порівняльного оцінювання інформативності параметрів ідентифікації моделей динаміки глікемії на основі одностороннього тесту дисперсійного аналізу. Доведено інформаційну значимість як традиційних параметрів, так і параметрів спектральних.

2. Вперше доведена можливість (без зниження вірогідності ідентифікації) використання тільки двох глюкозотолерантних тестів із залученням додаткової спектральної інформації за спрощеними поліноміальними моделями динаміки глікемії.

3. Доведено ефективність розробленої процедури ідентифікації моделей динаміки глікемії для задач багатосторонньої класифікації (безлічі ідентифікованих біологічних станів), у порівнянні з існуючими методами глюкозотолерантного тестування.

РОЗДІЛ 4

ЕКСПЕРИМЕНТАЛЬНЕ ДОСЛІДЖЕННЯ РОЗРОБЛЕНИХ ПРОЦЕДУР
ІДЕНТИФІКАЦІЇ ВИДІВ ДІАБЕТИЧНИХ ПОРУШЕНЬ

Цей розділ присвячено перевірці ефективності розроблених процедур ідентифікації видів діабетичних порушень за результатами множинного глюкозотолерантного тестування для верифікованих варіантів діабетичних ускладнень. Такі дослідження дають можливість отримання кількісних оцінок (в формі діагностичних ризиків і вірогідностей), що характеризують ефективність розроблених математичних моделей і процедур багатоальтернативної ідентифікації. Можна виділити два основних етапи таких досліджень:

- розробка структурної схеми мікропроцесорної системи для процедури параметричної ідентифікації діабетичних порушень;
- оцінка вірогідності і ризиків 1-го і 2-го роду при параметричній ідентифікації діабетичних порушень по відношенню до стандартного тесту та до групи глюкозотолерантних тестів.

4.1 Синтез структурної схеми процедури ідентифікації діабетичних порушень

Використовуючи математичну модель (3.26), можна побудувати структурну схему процедури ідентифікації номера k параметричної моделі динаміки глікемії за результатами вимірювання p інформативних ознак x_1, \dots, x_p , рис. 4.1.

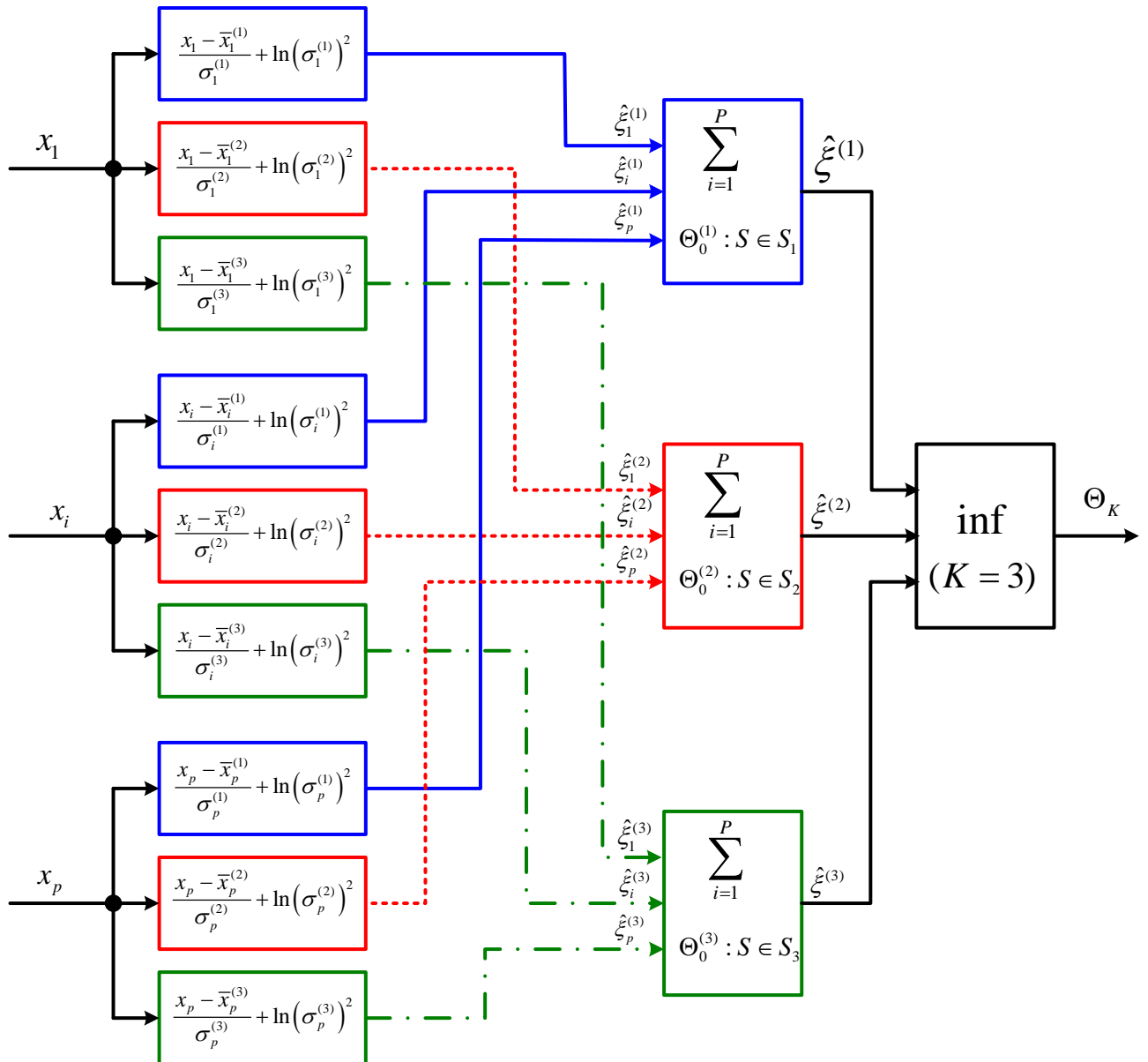


Рисунок 4.1 – Структурна схема процедури ідентифікації моделі динаміки глікемії

Рис. 4.1 показує, що структура процедури може бути адаптована для будь-якого числа m ідентифікованих біологічних станів ($m \geq 2$). Допускається і адаптація до числа p використовуваних інформативних ознак, які можна розділити на дві групи:

- p_1 ознак, які визначаються результатами вимірювання рівня глюкози в ході глюкозотолерантного тестування (величини середніх \bar{x}_i^k , де $i = \overline{1, p_1}$, задаються значеннями поліноміальних регресій відновлених в ході навчання системи ідентифікації);

- p_2 ознак, які визначаються коефіцієнтами автокогерентності при спектральному розкладанні відповідної поліноміальної регресії.

При цьому розмірність простору ознак $p = p_1 + p_2$.

4.2 Експериментальна перевірка ефективності розробленої процедури ідентифікації по відношенню до стандартного тесту

Така перевірка має на меті підтвердження адекватності розробленої процедури ідентифікації по відношенню до стандартного тесту [50] для максимально можливого числа ідентифікованих біологічних станів.

Дослідження підрозділів 3.2-3.4 дозволили виділити набір інформативних параметрів у вигляді двох груп:

- двох параметрів $g_b(X_1)$ і $g_{120}(X_4)$, що відображають зміни математичного очікування рівня глікемії в часі, як поліноміальної регресії 2-го порядку ($p_1 = 2$);

- чотирьох параметрів $Z_1 - Z_4$, які відображають зміни дисперсії рівнів глікемії, розподілених по частотах (за масштабами) вейвлет-перетворення ($p_1 = 4$).

Введемо позначення для множини параметрів $\{X_1, X_4; Z_1, \dots, Z_4\}$ як нової множини $\{y_1, \dots, y_6\}$, що відображає динамічні властивості (по математичному очікуванню і спектру) функції відгуку $Y(t)$ рис. 2.1 розділу 2.

Виходячи з обраної в підрозділі 2.6 системи чотирьох класів $\Theta_1, \dots, \Theta_4$ ($m = 4$) діабетичних станів (табл. 2.3) була проведена ідентифікація всіх чотирьох станів для 120 пацієнтів [1]. При цьому, для зменшення зсувів в оцінках функції правдоподібності повна вибірка розбивалася на дві групи по 60 пацієнтів (метод іспиту), за аналогією з методикою підрозділу 2.6. З іншого боку для класу Θ_1 (норма – відсутність діабетичної патології) ідентифікація проводилась методом ковзаючого іспиту U [91], в якому результати глюкозотолерантного тестування всіх 40 пацієнтів контрольної вибірки для 1-го класу були розбиті на дві вибірки:

- навчальна вибірка обсягом $N_1 = 39$;
- контрольна вибірка, яка використовується для ідентифікації, обсягом $N_2 = 1$.

Навчання системи ідентифікації для пацієнтів зі станом Θ_1 здійснювалося багаторазово з послідовною зміною навчальної вибірки, яка передбачає і зміни номера контрольованого пацієнта. Така процедура дозволяє зменшити негативний зсув оцінки ризику 1-го роду, об'єктивно збільшуючи дійсне значення цього ризику [91, 75].

У табл. 4.1 представлені значення логарифма функції максимальної правдоподібності (3.23), отримані методом ковзаючого іспиту. Правильна ідентифікація пацієнтів першого класу відповідає незатонованим максимумам логарифма функції правдоподібності $\ln F^{-1}(\bar{x}/\Theta_1)$ по стану Θ_1 . Помилкові класифікації даної таблиці – затоновані.

Таблиця 4.1

Значення логарифмів перетвореної функції правдоподібності $\ln F^{-1}(\bar{x}/\Theta_1)$
для варіанту ковзаючої ідентифікації стану Θ_1

№ пацієнта	δ_1	δ_2	δ_3	δ_4
	2	3	4	5
1	1,708196579	19,09185251	27,95209102	61,4072615
2	8,718943252	25,94245031	47,07763051	87,21866328
4	4,772511163	12,56167749	18,11384454	33,97079801
5	16,5055035	38,71997199	61,09048925	97,54785586
6	0,41700987	14,57630667	17,96636517	45,99950846
7	9,02354868	8,934817267	32,70815202	49,37213397
8	9,099036603	34,80726648	25,12656826	51,08843204
9	2,145973858	18,87182515	29,75460439	59,89898604
10	3,399505909	24,84859494	19,89795226	46,03270176
11	12,93065314	21,37627606	55,50779457	84,89691063
12	18,65145054	20,73293644	13,53983697	15,57569593
13	12,66128901	6,645692191	15,34214652	19,27529284
14	15,44136581	27,25860027	15,23429136	24,38814591
15	2,111585557	18,95600297	27,15529947	60,6465943
16	7,292585533	7,243185915	38,02314606	60,03670075
20	2,2166526	20,29901157	8,567170181	32,00534202
21	3,881338192	26,34546363	34,3843595	72,60438735

Закінчення табл. 4.1

1	2	3	4	5
22	4,968544108	17,82500321	34,5144917	69,30499322
24	30,93843541	45,41547371	41,33663274	51,02309664
25	4,406364753	8,811589093	14,60186265	35,92177984
28	8,748144024	11,40184432	26,66112213	40,10880144
29	1,256376849	17,03187338	11,09869567	36,8430014
32	13,27448193	46,20471847	24,41008822	53,63345408
33	0,783964466	9,410916209	15,28306429	38,09458507
34	0,354540826	10,70035222	16,66828016	40,71987525
35	12,68963025	35,69199698	27,17648909	49,64996227
36	18,01698497	37,29510369	32,94104257	71,54708972
37	3,769004198	25,01231509	9,784853487	33,13058686
38	6,044734067	9,158357578	26,113994	51,66203032
40	3,198198972	11,56087016	20,7408437	47,72396737
41	1,638408186	13,49541468	8,861391477	30,49739495
43	4,072647342	18,02064788	21,67946212	52,9301536
46	19,43174317	34,3849658	25,16263443	36,47106839
47	21,61916835	46,79800215	37,98227875	81,35210307
50	6,251722945	29,52558231	34,69167761	75,62374384
52	1,390138763	22,10038335	21,93647329	55,41694507
55	48,25126945	69,30864172	60,39957003	105,9748444
58	7,966627296	26,44377264	10,47539406	38,49132013
59	8,159431307	21,07209515	13,32134299	39,98257657
60	16,09074419	37,32735543	40,53814736	83,36934196

Слід зазначити, що в табл. 4.1 через використання методу ковзного іспиту, отриманий додатковий помилковий результат для пацієнта № 7, що збільшило кількість помилкових рішень з 4 до 5 (значення цільової функції для 1-го і 2-го стовпчиків, при навчанні системи ідентифікації на всіх 40 пацієнтів, приймали такі відповідні значення: 8,034617367 і 9,02270149). Слід також враховувати, що операція ковзного іспиту відбувалася за умови, що коефіцієнти полінома 2-го порядку (2.18) при кожній ітерації розраховувалися заново. Поліноми (2.19) – (2.21) залишалися при цьому без змін.

У табл. 4.2 представлені значення логарифма функції максимальної правдоподібності (3.23), отримані для патологічних станів Θ_2 , Θ_3 і Θ_4 . Правильна ідентифікація пацієнтів класів, що відображають дані стану відповідає незатонованим максимумам логарифма функції правдоподібності

$\ln F^{-1}(\bar{x}/\Theta_1)$ по стану Θ_1 . Помилкові класифікації даної таблиці – затоновані.

Таблиця 4.2

Значення логарифмів перетвореної функції правдоподібності $\ln F^{-1}(\bar{x}/\Theta_1)$

для варіанту ковзаючої ідентифікації станів Θ_2, Θ_3 і Θ_4

№ пацієнта	δ_1	δ_2	δ_3	δ_4
17	13,8506346	0,231862923	25,43291744	33,94209725
19	15,5124696	0,231862923	32,38596794	40,40533079
3	32,45178999	34,10484872	14,00343665	9,169324061
18	9,523709514	10,60469435	5,619009816	17,54977602
23	13,14594671	23,19886166	1,630942816	16,23949804
26	43,50517998	39,10377232	11,3214588	11,48461773
27	25,62149946	48,26714959	6,99198228	28,55488763
30	20,70216731	29,82513074	5,195970424	10,64446483
31	35,12955773	57,83936575	25,70650318	33,87465621
39	12,87028044	34,89767779	6,30421629	31,54108565
44	51,74021268	84,79974768	27,72792546	36,87984687
45	19,71957583	34,05797963	5,395517598	12,8096574
48	10,00879813	15,00009444	2,533684251	13,40697233
49	33,70073949	49,98705946	7,737027091	23,43938644
51	18,39561222	28,82300455	0,346007154	8,619847602
53	38,13118576	33,31986002	10,68759106	13,08510078
54	33,72236772	47,62952131	7,209948055	21,11987577
56	12,65590978	28,75497731	2,790989753	22,13016124
57	25,84487111	30,70689323	2,199571794	8,405907302
42	43,70321505	36,9418511	10,32666373	0

Табл. 4.1 і табл. 4.2 були використані для отримання оцінки вірогідності ідентифікації як кількісного показника адекватності розробленого комплексного методу ідентифікації за гранично мінімізованим обсягом глюкозотолерантних тестів (кількість тестів дорівнює 2). Для цього були використані наступні позначення:

- можливі рішення $\gamma_j: \Theta \in \Theta_j; j = \overline{1,4}$;
- ризик 1-го роду за станом Θ_1 $\alpha_1 = P(\gamma_{j \neq 1}/\Theta_1)$;
- ризики 2-го роду за станами Θ_2, Θ_3 і Θ_4 $\beta_j = P(\gamma_{r \neq 1}/\Theta_j), j = 2,3,4$;
- апріорні ймовірності станів $q_j = n_j/N$, де $n_1 = 40, n_2 = 2, n_3 = 16$,

$n_4 = 2$ (примітка: вибір $n_1 = 40$ продиктований загальним значенням всіх варіантів ковзного іспиту).

Вірогідність ідентифікації визначалася як лінійна функція середнього ризику \bar{R} :

$$D = 1 - \bar{R}, \quad (4.1)$$

де середній ризик визначається виразом

$$\bar{R} = \alpha_1 q_1 + \sum_{j=2}^4 \beta_j q_j, \quad (4.2)$$

де q_1, \dots, q_4 – апіорні ймовірності станів $\Theta_1, \dots, \Theta_4$.

Результати розрахунку показують, що середній ризик $\bar{R} = 0,09(9)$, а значення вірогідності $D = 0,9$.

Отримане значення вірогідності вказує, що адекватність результатів ідентифікації за допомогою нового розробленого методу не нижче 90%, що, з огляду на використання всього двох глюкозотолерантних тестів, дозволяє стверджувати про досягнення прийнятної ефективності цього методу, особливо при масовому обстеженні пацієнтів.

Для порівняння ефективності розробленого методу з відомим вдосконаленим методом [1] була розрахована (додаток А.2 [1]) вірогідність ідентифікації останнього, як відношення сумарної кількості помилкових рішень за станами $\Theta_1, \dots, \Theta_4$ до загальної кількості пацієнтів. Отримане значення вірогідності $D = 0,767$, що нижче, ніж величина вірогідності ідентифікації нового методу. Слід при цьому підкреслити, що вдосконалений метод [1] є більш розширеним та універсальним, оскільки крім завдань ідентифікації моделей глікемії пояснює і дозволяє ідентифікувати динамічні і ентеральні фактори, що визначають функції всмоктування глюкози в кров в шлунково-кишковому тракті.

4.3 Експериментальна перевірка ефективності процедури ідентифікації по відношенню до групи глюкозотолерантних тестів

Розширення кількості тестів для такої експериментальної перевірки накладає обмеження на кількість біологічних станів, яке необхідно, відповідно, зменшити (в рамках збереження кількості контрольної інформації при фіксованому обсязі $N = 120$ використовуваних експериментальних даних).

Для дослідження була обрана група з 2-х існуючих на даний час тестів: стандартний [50] і вдосконалений тест [1]. Оскільки кількість нормативних тестів було збільшено вдвічі, то число ідентифікованих станів було зменшено з $m = 4$ до $m = 2$:

а) Θ_1 – характеризується набором пацієнтів з діагнозом $N - N$ як по стандартному, так і за вдосконаленим тестом;

б) Θ_2 – характеризується набором пацієнтів з будь-якими іншими діагнозами хоча б по одному з тестів.

Використовуючи модель іспиту була проведена розбивка множин пацієнтів зі станами Θ_1 і Θ_2 на дві однакові за обсягом групи ($N = 60$): контрольну та навчальну. Обсяги вибірок за ідентифікованими станами для цих останніх груп склали величини $N_1 = 38$ и $N_2 = 22$.

У табл. 4.3 представлені результати ідентифікації пацієнтів зі станом Θ_1 методом максимальної правдоподібності за мінімумом цільової функції виду $\ln F^{-1}(\bar{x}/\Theta_1)$, рівняння (3.23) (помилкові результати – затоновані).

Таблиця 4.3

Результати ідентифікації пацієнтів за станом Θ_1 за цільовою функцією (3.23)

№ пацієнта	δ_1	δ_2
1	2	3
1	0,922575552	11,91275542
2	4,684079141	20,85217822
4	3,13270898	7,290795777

Закінчення табл. 4.3

1	2	3
5	8,453829643	26,5213595
6	0,294382499	7,448905232
7	4,471032929	12,71702294
8	6,300016566	12,83295961
9	1,061366312	12,14787374
10	2,341948196	9,023576048
11	6,030337477	22,34254126
13	7,019228432	5,715640606
15	1,292046409	11,85623408
16	3,506607526	14,53315242
20	1,043153308	3,694676222
21	1,99191232	15,15358133
22	3,072115459	15,32219142
24	21,76697915	24,22871906
25	2,731181133	6,307508174
28	5,426443425	10,86819006
29	0,61853401	4,735226759
32	8,714575746	13,90098196
33	0,449633223	5,815033258
34	0,210033551	6,415226733
35	8,982686357	14,46701243
36	12,20142543	19,72989508
37	2,108976684	4,659767646
38	4,071427011	11,7120928
40	2,228619885	9,18202352
41	0,753148161	3,402218676
43	2,85867403	10,4188365
46	13,50144202	14,47552949
47	14,26936044	22,94512376
50	3,816606743	16,49685203
52	0,822724226	9,776585772
55	33,1083117	40,26034853
58	4,348601566	6,102576059
59	4,955636747	7,45824601
60	10,92922549	22,42069605

Оскільки вибір пацієнтів зі станом Θ_1 здійснювався при збігу результатів діагнозу $N - N$ стандартного і вдосконаленого тестів, то помилки першого роду для останніх – відсутні.

У табл. 4.4 представлені результати ідентифікації пацієнтів зі станом Θ_2

розробленим методом по мінімуму цільової функції виду $\ln F^{-1}(\bar{x}/\Theta_1)$, рівняння (3.23) (помилкові результати – затоновані).

Таблиця 4.4

Результати ідентифікації пацієнтів зі станом Θ_2 за цільовою функцією (3.23)

№ пацієнта	δ_1	δ_2
12	11,20559839	6,995887558
14	10,07365831	8,609832955
3	16,1602282	16,10921409
17	8,728795795	13,17679404
18	9,695739566	2,18717584
19	4,484134724	1,610206216
23	5,967861022	0,980531807
26	20,65985926	6,32541494
27	11,80682947	5,267263309
30	10,99837421	3,429669375
31	22,9493788	18,17020007
39	6,073525387	4,30577738
42	12,28143345	4,514456197
44	25,17587383	17,09241246
45	9,593974095	4,217164439
48	13,61186333	2,689485445
49	9,864051622	3,13965759
51	16,2461585	2,333852454
53	16,4483482	6,200892287
54	7,311963736	4,039084756
56	11,88028476	1,600851688
57	19,66540519	4,724473526

Використання результатів глюкозотолерантного тестування пацієнтів зі станом Θ_2 для стандартного і вдосконаленого методів (додаток А2 [1]) показало, що для контрольної вибірки з 60 пацієнтів мали місце помилки 2-го роду:

- для стандартного тесту мали місце 2 хибних рішення (неправильно ідентифіковані пацієнти № 12 і № 14);

- для вдосконаленого тесту мали місце 9 помилкових рішень (пацієнти з номерами 18, 23, 31, 39, 44, 45, 51, 56, 57).