

О. В. Чмихова<sup>1</sup>, П. Ф. Шапов<sup>1</sup>, В. В. Куліченко<sup>1</sup>, О. В. Горбуліч<sup>2</sup>

<sup>1</sup> Національний технічний університет «ХПІ», Харків, Україна

<sup>2</sup> Харківська медична академія післядипломної освіти, Харків, Україна

## ПАРАМЕТРИЧНА ІДЕНТИФІКАЦІЯ ДІАБЕТИЧНИХ СТАНІВ ПРИ ВИКОРИСТАННІ ПОЛІНОМІАЛЬНИХ МОДЕЛЕЙ ДИНАМІКИ ГЛІКЕМІЇ

**Предметом дослідження** є елементи аналітичного і структурного синтезу процедури ідентифікації моделі динаміки глікемії за критерієм максимальної правдоподібності. **Метою роботи** є зменшення ризиків ідентифікації діабетичних станів при використанні поліноміальних моделей динаміки глікемії. **Завдання.** На основі результатів глюкозотолерантного тестування дослідити можливості параметричного розпізнавання діабетичних станів, коли динаміка рівня глюкози апроксимується поліноміальною моделлю будь-якого порядку. Використати аналітичний вираз для мінімуму середнього ризику для структурного синтезу процедури ідентифікації виду діабету. Підтвердити ефективність використання отриманої структурної схеми процедури ідентифікації моделі динаміки глікемії при глюкозотолерантному тестуванні. **Висновки.** Розроблено блок-схему з адаптивною (по відношенню до кількості ідентифікованих глікемічних моделей) структурою і відповідна їй комп'ютеризована система багатосторонньої ідентифікації моделей динаміки глікемії за двома глюкозотолерантними тестами. Доведено можливість підвищення достовірності діагностики при використанні запропонованого методу, у порівнянні зі стандартним від значення 0,972 до 0,974 (обсяг навчальної вибірки  $N = 60$ ).

**Ключові слова:** діабетичний стан, динаміка глікемії, параметрична ідентифікація, поліноміальна модель, середній ризик, алгоритм ідентифікації.

### Вступ

В даний час розроблений досить широкий клас моделей динаміки глікемії стандартна модель [1], вдосконалена модель [2], поліноміальна [3], які базуються на використанні фіксованої кількості інформативних параметрів (ознак). Всі перераховані моделі оперують точними [2] або статистично наближеними [3] залежностями вимірюваної рівня глюкози в функції часу тестування. Перевага точних моделей – адекватність біохімічним змінам кількісного складу крові. Недолік – невизначеність математичної моделі за функціональним змістом. Перевага наближених моделей – опис біохімічного об'єкта в просторі використовуваних параметрів, що дозволяє замінити задачу кількісної ідентифікації (визначається точністю математичного опису динаміки глікемії) завданням параметричного розпізнавання образів, зберігаючи, практично, незмінною достовірність ідентифікації. Недолік – залежність моделі розпізнавання від якості навчання (точності оцінювання коефіцієнтів), пов'язаного з обсягом навчальної вибірки. Однак, такий недолік можна асимптотично звести до статистично обгрунтованого мінімуму, збільшуючи обсяги навчальних вибірок.

**Метою роботи** є мінімізація ризиків ідентифікації діабетичних станів, для глюкозотолерантних тестів, при використанні поліноміальних моделей динаміки глікемії. Для досягнення поставленої мети необхідно вирішити такі завдання:

1) розробити аналітичний вираз для мінімуму середнього ризику при параметричному розпізнаванні діабетичних станів методом максимальної правдоподібності;

2) синтезувати структурну схему алгоритму ідентифікації моделі динаміки глікемії на основі розпізнавання по мінімуму середнього ризику;

3) підтвердити ефективність розробленого алгоритму ідентифікації на експериментальному матеріалі клінічних глюкозотолерантних тестів.

### Результати досліджень

**1. Формалізація моделей ідентифікації.** Будемо розглядати процедуру ідентифікації таких моделей як параметричну, за умови, що числові характеристики інформативних ознак для кожної з моделей – відомі. Якщо в якості числових характеристик використовується математичне очікування і дисперсія ознак при апріорних припущеннях нормальності (або хоча б симетричності) законів розподілу ознак, то процедура ідентифікації буде еквівалентна процедурі параметричного розпізнавання виду моделі динаміки глікемії. Введемо поняття вектору параметрів  $\bar{X}$ , складовими якого є  $p$  інформативних ознак:

$$\bar{X} = (x_1, \dots, x_p),$$

де  $p$  – розмірність простору інформативних ознак.

Нехай  $(x_1, \dots, x_p)$  – сукупність результатів вимірювань значень  $p$  контрольованих величин, що утворюють вектор  $\bar{x} = (x_1, \dots, x_p)$ , і залежать від того, який з станів має місце. Нехай  $f(\bar{x} | \Theta_k)$  – умовний розподіл цього вектору, що відповідає стану  $\Theta_k$ . Є також: набір рішень  $\gamma_1, \dots, \gamma_m$  ( $m$  – кількість ідентифікованих моделей динаміки глікемії). Є також правило вибору рішень  $\gamma(\bar{x})$ , що приписує кожному можливому результату одне з рішень, матриця втрат  $\Pi_{jk}$ , що враховують наслідок вибору рішення  $\gamma_j$  при дійсному стані  $\Theta_k$ . Заданий також критерій якості  $R_0$  правила вибору рішення [4]. В рамках теорії перевірки статистичних гіпотез [4, 5] набір рішень являє собою ряд логічних тверджень про вид стану  $\Theta$ , тобто про те, яка з гіпотез,  $h_1 : \Theta \in \Theta_1, \dots, h_m : \Theta \in \Theta_m$ , істинна. Одна з гіпотез множини  $\{\gamma_1, \dots, \gamma_m\}$  може бути нульовою, а

інші тоді, утворюють складну альтернативну гіпотезу. Ідентифікація одного з станів  $\{\Theta_1, \dots, \Theta_m\}$  зводиться, таким чином, до параметричного тестування вектора вимірних значень за допомогою функції, еквівалентній середньому ризику

$$R_0 = F \left[ \left\{ f(\bar{x} | \Theta_k) \right\}, \Pi_{jk}, \{P_1, \dots, P_m\} | \bar{x} \right] \quad (1)$$

при заданому  $\gamma(\bar{x})$ . Це означає, що простір всіх можливих вибірок значень випадкового вектору контрольованих величин  $\bar{X}$  має бути розділений на  $m$  непересічних областей, з яких одна – це область прийняття рішення, а що залишилися – це критична область  $\bar{\omega}$ . При такому тестуванні можливі помилки першого і другого роду, ймовірності яких

$$\alpha = P \left[ \gamma_j | \bar{x} \in \omega; \Theta \in \Theta_k \right], k \neq j; ;$$

$$\beta = P \left[ \gamma_k | \bar{x} \in \bar{\omega}; \Theta \notin \Theta_k \right].$$

Слід зазначити, що правило вибору рішення  $\gamma(\bar{x})$  задається до початку тестування.

Теоретичною основою розглянутої загальної моделі прийняття рішень є байєсівська теорія рішень [6], яка складає основу дискримінантного аналізу [7]. В рамках такого аналізу добре розроблені моделі параметричної дискримінації для об'єктів контролю з двома, ( $m = 2$ ), станами  $\Theta_1, \Theta_2$ .

У класичній теорії статистичних рішень [8] такі моделі реалізують параметричні тести двосторонньої класифікації. Якщо число діагностованих станів більше двох ( $m \geq 3$ ), то дискримінація базується на параметричних тестах багатосторонньої класифікації. Збільшення числа діагностованих (дискримінованих) станів різко ускладнює модель прийняття рішень, оскільки для кожного з станів необхідно апіорі знати числові характеристики інформативних ознак (хоча б на рівні середніх значень і дисперсій). Така апіорна інформація може бути отримана тільки на етапі навчання системи дискримінації, що еквівалентно створенню імовірнісних моделей для стандартних зразків інформативних ознак по апіорі встановленим видам діагностованих станів. Оскільки етап навчання вимагає досить великих обсягів виміральної багатовимірної інформації, що досить проблематично для більшості прикладних задач медичної діагностики, то в більшості випадків моделі такої діагностики – це двосторонні тести.

Спроби вирішення завдання багатосторонньої класифікації здійснено лише при обмеженнях на ймовірність помилки першого роду  $\alpha$  і являють собою моделі тестування на значимість, з апіорі невідомою ймовірністю помилки другого роду  $\beta$ . Практично, всі існуючі моделі дискримінації розроблені для випадків, коли багатовимірною щільністю  $f(\bar{x} | \Theta_k)$ ,  $k = \overline{1, m}$ , є невідродженим  $p$ -мірним нормальним розподілом з однаковими або різними за класами  $\Theta_1, \dots, \Theta_m$  коваріаційними матрицями.

**2. Синтез цільової функції мінімізації середнього ризику.** Практика використання багатовимірного дискримінантного аналізу при обмеженнях на обсяги навчальних вибірок і на кількість апіорних припущень про їх імовірнісні властивості [9] показує, що в основу процедури ідентифікації моделей динаміки глікемії зручно поставити метод максимальної правдоподібності. Особливу цінність метод максимальної правдоподібності, в нашому випадку, набуває через його головну властивість – забезпечувати мінімум середнього ризику при глобальному максимумі функції правдоподібності. Критерій максимальної правдоподібності може бути реалізований в векторній формі як цільова функція вигляду [10]:

$$W(\bar{x} | \Theta_k) = (2\pi)^{\frac{p}{2}} |\Sigma|^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} (\bar{x} - \mu^{(k)})' \Sigma^{-1} (\bar{x} - \mu^{(k)}) \right]$$

для стану  $\Theta_k$  – максимальна ( $\bar{x}$  – вектор стовпець вимірних значень;  $\mu^{(k)}$  – вектор умовних середніх;  $\Sigma$  – матриця коваріацій). Слід врахувати, що визначник коваріаційної матриці повинен відповідати умові невідродженості розподілу  $W(\bar{x} | \Theta_k)$ , тобто  $|\Sigma| > 0$ .

В якості інформативних ознак використовуються статистично незалежні величини, випадковість яких обумовлена непереборною невизначеністю моделей динаміки глікемії. Така невизначеність посилюється обмеженістю виміральної багатовимірної інформації, використовуваної для навчання системи ідентифікації.

Кожну зі складових вектору ознак  $\bar{X}$  будемо вважати випадковою величиною  $x_i^{(k)}$  з умовною нормальною щільністю розподілу  $f(x_i^{(k)})$ , де  $k$  – номер ідентифікованого стану:

$$f(x_i^{(k)}) = \text{NORM} \left( \bar{X}_i^{(k)}, (\sigma_i^{(k)})^2 \right); i = \overline{1, p}; k = \overline{1, m}.$$

Функція правдоподібності в скалярній формі для всіх  $p$  інформативних ознак є такою:

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

Для зручності подальшого аналізу прологарифмуємо цю функцію правдоподібності, враховуючи, що така операція не змінює координати логарифмічного перетворення по відношенню до значень інформативних параметрів [10]:

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

Для зручності роботи з позитивними значеннями логарифма функції правдоподібності врахуємо, що максимум вихідного логарифма буде визначатися мінімумом цього ж логарифма, але зі знаком «-», що відповідає умові

$$\sup \left\{ \ln F(\bar{X} / \Theta_k) \right\} = \inf \left\{ \ln F^{-1}(\bar{X} / \Theta_k) \right\}. \quad (2)$$

Максимум  $\ln F(\bar{X} / \Theta_k)$  еквівалентний за визначенням мінімуму

$$\ln F^{-1}(\bar{X} / \Theta_k) = \rho \cdot \ln(2\pi) / 2 + \sum_{i=1}^p \ln \sigma_i^{(k)} + \frac{1}{2} \sum_{i=1}^p \left( \frac{x_i - \bar{x}_i^{(k)}}{\sigma_i^{(k)}} \right)^2 \quad (3)$$

Останній доданок правої частини виразу (3) є лінійною функцією геометричної відстані між двома точками в  $p$ -вимірному 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 [10], що така відстань еквівалентна міжфункціональній відстані  $\rho_L(\cdot)$  для порівнюваних поліноміальних функцій, значення яких про- нормовані по дисперсіях.

**3. Синтез структурної моделі процедури ідентифікації діабетичних порушень** Використовуючи математичну модель (6), можна побудувати структурну модель процедури ідентифікації номера  $k$  параметричної моделі динаміки глікемії за результатами вимірювання  $p$  інформативних ознак  $x_1, \dots, x_p$ , рис. 1, який 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. Експериментальна перевірка можливостей розробленої структурної схеми для ідентифікації діабетичних станів, що використовуються в лікарській практиці і клінічних дослідженнях, показала, що показники ефективності існуючих глюкозотолерантних тестів істотно розрізняються. Оцінки ризиків і достовірності результатів ідентифікації для 3-х порівнюваних тестів представлені в табл. 1.

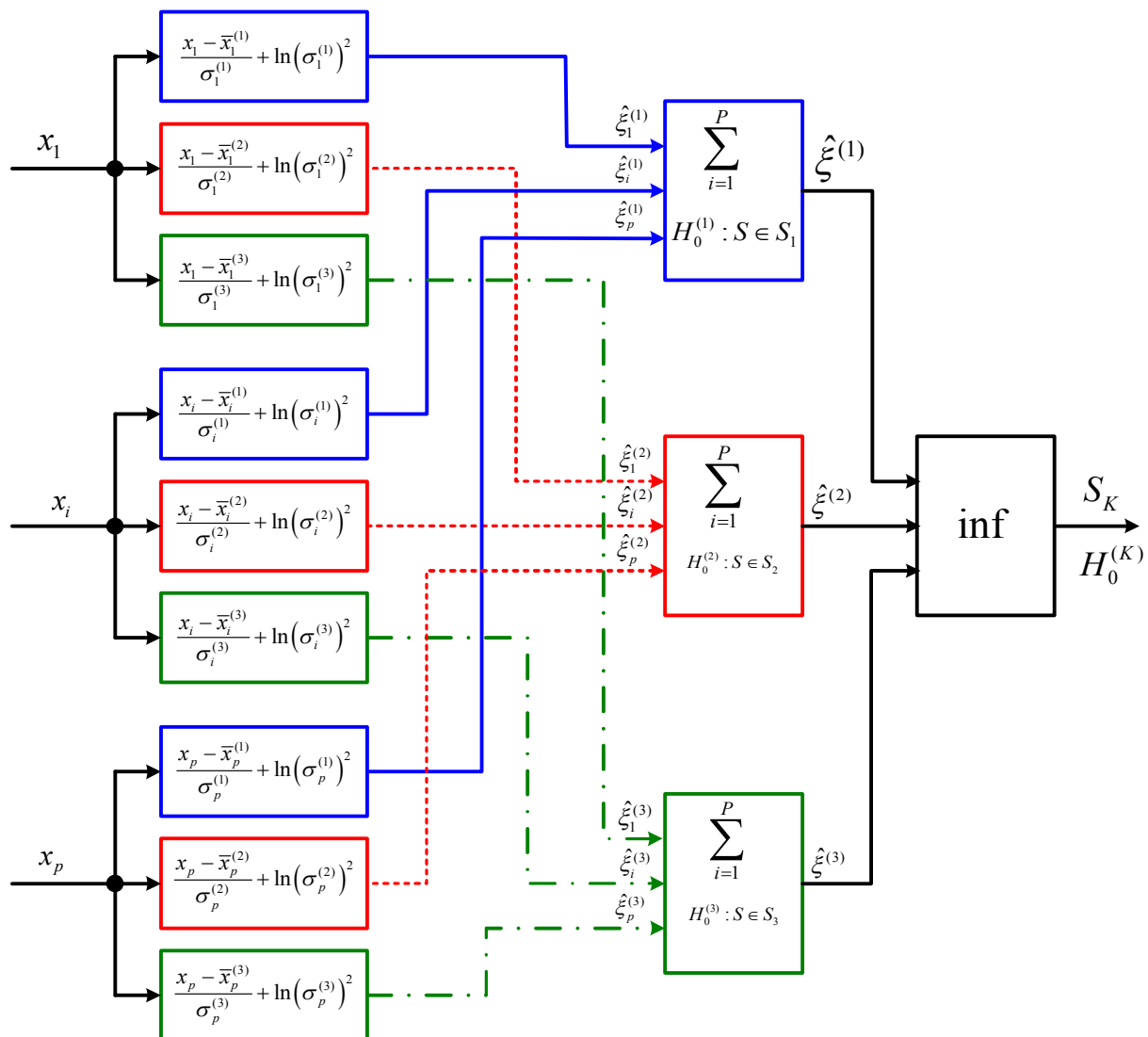


Рис. 1. Структурна схема процедури ідентифікації моделі динаміки глікемії

**Висновки**

1. Розроблено блок-схему з адаптивною (по відношенню до кількості ідентифікованих глікемічних моделей) структурою і відповідна їй комп'юте-

ризована система багатосторонньої ідентифікації моделей динаміки глікемії за двома глюкозотолерантними тестами. Збережена можливість ідентифікації глікемічних моделей для розширеної (більше 2-х) кількості тестів при необхідності підвищення досто-

вірності ідентифікації за умови збереження мінімально допустимого значення ризику 2-го роду.

Таблиця 1 – Показники ефективності глюкозотолерантних тестів

Ризики і достовірності	Метод тестування		
	Стандарт	Вдосконалений	Розроблений
Ризик 1-го роду	0	0	0,026
Ризик 2-го роду	0,091	0,409	0,045
Середній ризик	0,033	0,15	0,033
Достовірність	0,967	0,85	0,967

Розрахунок достовірності в табл. 1 проводиться для апіорних ймовірностей

$$q_1 = 0,633 \quad \text{і} \quad q_2 = 0,367.$$

2. Доведено можливість підвищення достовірності діагностики при використанні запропонованого методу, у порівнянні зі стандартним від значення 0,972 до 0,974 (обсяг навчальної вибірки  $N = 60$ ), причому останню достовірність теоретично можна підвищити до значення 0,979, якщо обсяг навчальної вибірки буде збільшуватися (що обумовлено потенційним зменшенням ризику другого роду  $\beta$ ).

#### СПИСОК ЛІТЕРАТУРИ

- Сахарный диабет: Доклад исследовательской группы. сер. техн. докл. ВОЗ. М. Мед. 1987. 125 с.
- Лапта С.С. Методы повышения диагностической эффективности глюкозотолерантных тестов (на основе математического моделирования динамики гликемии): Дис. канд. техн. наук: 05.11.17 / ХНУРЕ. – Х., 2004. – 220 с.
- Сокол Є.І. Підвищення ефективності глюкозотолерантних тестів при використанні поліноміальних моделей динаміки глікемії / Є.І. Сокол, П.Ф. Щапов, О.В. Чмихова // Сучасні інформаційні системи. Т. 3, № 3. Харків, 2019. С. 138-141.
- Левин Б.Р. Теоретические основы статистической радиотехники. Кн. 2 / Б.Р. Левин – М.: Сов. радио, 1975. – 392 с.
- Айвазян С.А. Прикладная статистика: основы моделирования и первичная обработка данных / С.А. Айвазян, И.С. Енюков, Л.Д. Мешалкин. – М.: Финансы и статистика, 1983. – 471 с.
- Дуда Р. Распознавание образов и анализ сцен / Р. Дуда, П.Харт. – М.: Мир, 1976. – 512 с.
- Зыбов В.Н. Моделирование функции преобразования первичного преобразователя в задачах многофакторных изменений / В.Н. Зыбов // Измерительная техника. – 2006. - № 4. – С. 26-31.
- Джонсон Н. Статистика и планирование эксперимента в технике и науке: Методы планирования эксперимента / Н. Джонсон, Ф. Лион; пер. с англ. – М.: Мир, 1981. – 520 с.
- Раудис Ш. Ограниченность выборки в задачах классификации / Ш. Раудис // Статистические проблемы управления. – Вильнюс. – 1976. – Вып. 18. – С. 1-185.
- Щапов П.Ф. Повышение достоверности контроля и диагностики объектов в условиях неопределённости: монография / П.Ф. Щапов, О.Г. Аврунин. – Х.: ХНАДУ, 2011. – 191 с.

**Рецензент:** д-р техн. наук, доц. Г. Е. Філатова,  
Національний технічний університет “ХПІ”, Харків  
Received (Надійшла) 10.10.2019  
Accepted for publication (Прийнята до друку) 26.10.2019

#### Параметрическая идентификация диабетических состояний при использовании полиномиальных моделей динамики гликемии

О. В. Чмыхова, П. Ф. Щапов, В. В. Куличенко, А. В. Горбулич

**Предмет исследования.** Элементы аналитического и структурного синтеза процедуры идентификации модели динамики гликемии по критерию максимального правдоподобия. **Цель работы.** Уменьшение рисков идентификации диабетических состояний при использовании полиномиальных моделей динамики гликемии. **Задание.** На основе результатов глюкозотолерантного тестирования исследовать возможности параметрического распознавания диабетических состояний, когда динамика уровня глюкозы аппроксимируется полиномиальной моделью любого порядка. Использовать аналитическое выражение для минимума среднего риска для структурного синтеза процедуры идентификации вида диабета. Подтвердить эффективность использования полученной структурной схемы процедуры идентификации модели динамики гликемии при глюкозотолерантном тестировании. **Выводы.** Разработана блок-схема с адаптивной (по отношению к количеству идентифицируемых гликемических моделей) структурой и соответствующая ей компьютеризованная система многосторонней идентификации моделей динамики гликемии по двум глюкозотолерантным тестам. Доказана возможность повышения достоверности диагностики при использовании предлагаемого метода, по сравнению со стандартным от значения 0,972 до 0,974 (объем обучающей выборки  $N=60$ ).

**Ключевые слова:** диабетическое состояние, динамика гликемии, параметрическая идентификация, полиномиальная модель, средний риск, алгоритм идентификации.

#### Parametric identification of diabetic conditions using polynomial glycemic dynamics models

O. Chmykhova, P. Shchapov, V. Kulichenko, A. Gorbulitch

**The subject.** Elements of analytical and structural synthesis of the procedure for identifying the glycemic dynamics model by the criterion of maximum likelihood. **Objective.** Reducing the risks of identifying diabetic conditions using polynomial glycemic dynamics models. **The task.** Based on the results of glucose tolerance testing, investigate the possibilities of parametric recognition of diabetic conditions when the dynamics of glucose levels are approximated by a polynomial model of any order. Use an analytical expression to minimize the average risk for structural synthesis of the type of diabetes identification procedure. Confirm the effectiveness of using the obtained structural diagram of the procedure for identifying the glycemic dynamics model during glucose tolerance testing. **Conclusions.** A block diagram has been developed with an adaptive (with respect to the number of identifiable glycemic models) structure and a corresponding computerized system for multilateral identification of glycemic dynamics models using two glucose tolerance tests. The possibility of increasing the reliability of diagnosis using the proposed method is proved, compared with the standard one from a value of 0,972 to 0,974 (training sample size  $N = 60$ ).

**Keywords:** diabetic state, glycemic dynamics, parametric identification, polynomial model, medium risk, identification algorithm.