



Харківський національний університет радіоелектроніки

ГІБРИДНІ НЕЙРО-ФАЗЗИ СИСТЕМИ ДЛЯ ЗАВДАНЬ ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ ПОТОКІВ МЕДИЧНИХ ДАНИХ

ПЕРОВА ІРИНА ГЕННАДІЇВНА

д.т.н., с.н.с., доцент, професор кафедри біомедичної
інженерії ХНУРЕ

Проблема, що розв'язується в роботі: розроблення методів для аналізу потоків медичної інформації в послідовному режимі з використанням підходів Medical Data Mining для підтримки впровадження системи eHealth в Україні і підвищення ефективності медичного діагностування в цілому.

Об'єктом дослідження є процес аналізу потоків медичних даних в онлайн-режимі.

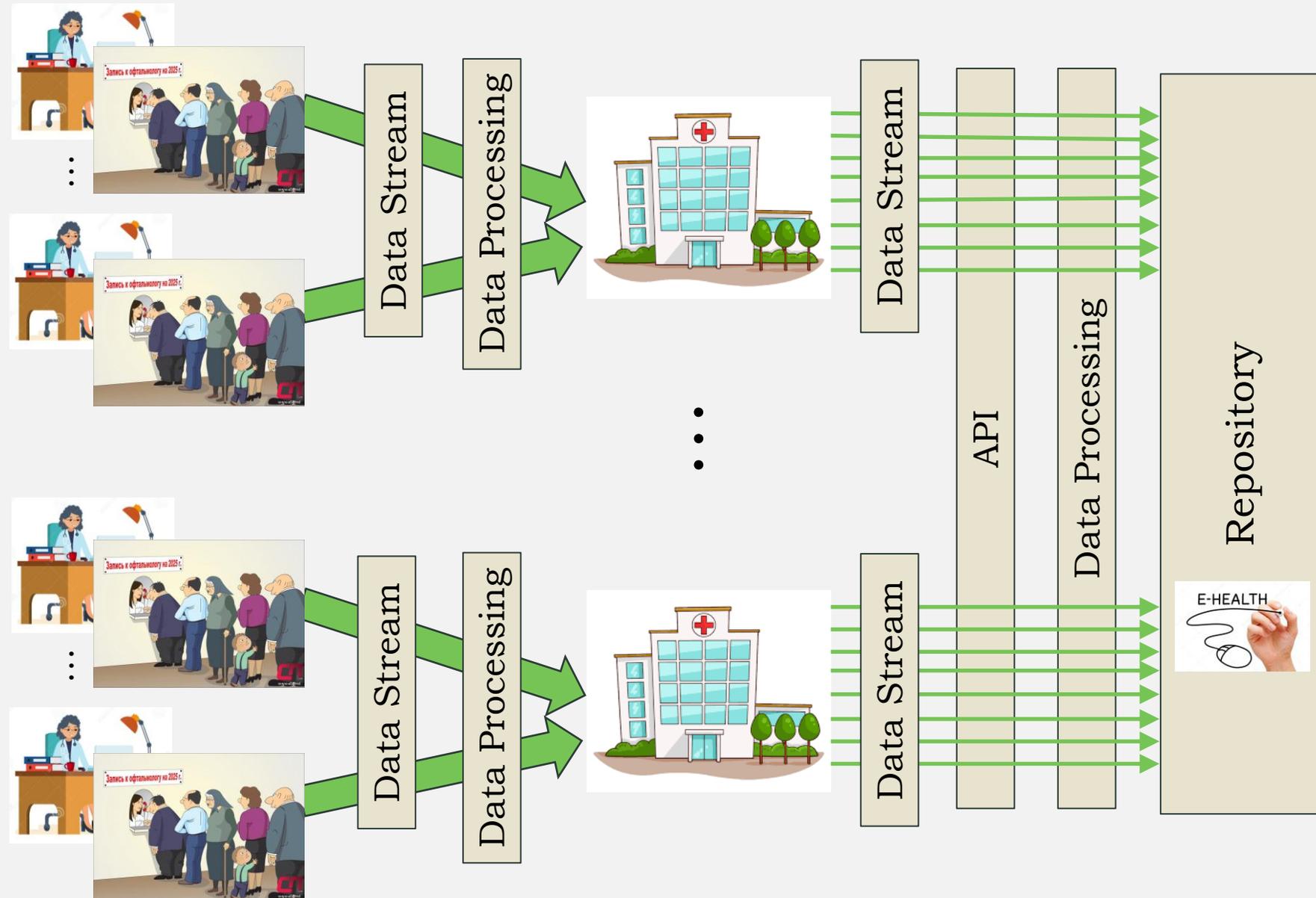
Предметом дослідження є моделі й методи послідовного аналізу масивів медичних даних в умовах обмеженої інформації в онлайн-режимі.

Метою роботи є підвищення ефективності медичного діагностування в онлайн-режимі на основі інтелектуального аналізу масивів медичних даних з використанням гібридних нейро-фаззі систем.

Завдання роботи:

1. Розробити адаптивний метод попереднього оброблення медичних даних в послідовному режимі з урахуванням типів даних, включаючи заповнення пропущених значень, гібридний метод оцінювання інформативності медичних показників в онлайн-режимі.
2. Розробити методи аналізу медичних даних під час медичного онлайн-діагностування в режимі контрольованого навчання на основі гібридних адаптивних нейро-фаззі систем для ситуацій репрезентативної навчальної вибірки.
3. Розробити методи аналізу медичних даних під час медичного онлайн-діагностування в режимі активного навчання та асоціації на основі гібридних нейро-фаззі систем для ситуацій нерепрезентативної навчальної вибірки.
4. Розробити методи аналізу медичних даних під час медичного онлайн-діагностування в режимі самонавчання, що базуються на кластеризації та асоціативній кластеризації за допомогою нейро-фаззі мереж.
5. Провести апробацію усіх розроблених методів на клінічних даних.

Потік медичних даних



Критерій ефективності

Для порівняння ефективності діагностування різних підходів між собою було введено єдиний критерій ефективності $PerCr$, що складається з помилки при навчанні/тестуванні діагностувальної системи ($Fault$) та часу ($Time$), що необхідний для опрацювання одного пацієнта.

$$PerCr = \lambda_1 Fault + \lambda_2 Time,$$

λ_1, λ_2 – вагові коефіцієнти, що обираються за умов $\sum \lambda = 1$.

Помилка $Fault$ та час $Time$ нормуються в інтервал від 0 до 1.

$\min(Time) = 0$ мс, $\max(Time) = 1$ мс.

$\min(Fault) = 0$, $\max(Fault) = 1$.

$$0 < PerCr < 1$$

Найкраща ситуація: $PerCr < 0,25$.

Заповнення пропусків

$$x(k) = (x_1(k), x_2(k), \dots, x_i(k), \dots, x_n(k))^T$$

При цьому деякі вектори спостережень можуть мати пропуски.

1. Розрахунок часткових відстаней між поточним пацієнтом $x(k)$ з пропусками і всіма іншими пацієнтами $x(p)$, $p \neq k$

$$d_p(k, p) = \frac{1}{n - n_k - n_p + n_{kp}} \sum_{i=1}^n |x_i(k) - x_i(p)|$$

n_k – число пропусків у k -го вектору спостережень (пацієнта),

n_p – число пропусків у p -го вектору спостережень (пацієнта),

n_{kp} – число загальних пропусків в одній і тій же позиції у обох пацієнтів.

2. Розрахунок рівнів належності $\mu_{p_l}(k)$ k -го пропуску до p_l -го спостереження:

$$\mu_{p_l}(k) = \frac{d_p^{-1}(k, p_l)}{\sum_{p_l=1}^{N-1} d^{-1}(k, p_l)}$$

3. Заповнення пропуску в одній з позицій:

$$\hat{x}_{\tilde{i}}(k) = \sum_{p_l} \mu_{p_l}(k) x_{\tilde{i}}(p_l) \mathcal{N} \bar{p}_l$$

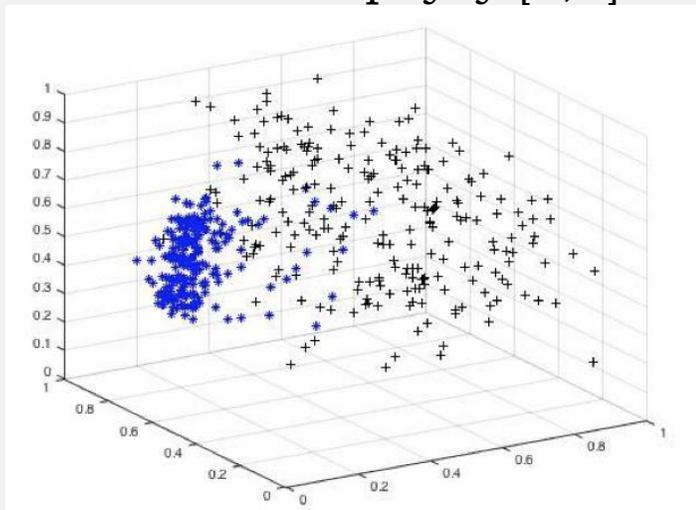
Процедура повторюється доти, доки всі пропуски не будуть заповнені.

Нормування даних

Кількісна метрика

Нормування на гіперкуб:

Всі дані належать гіперкубу $[0;1]^n$ або $[-1;1]^n$



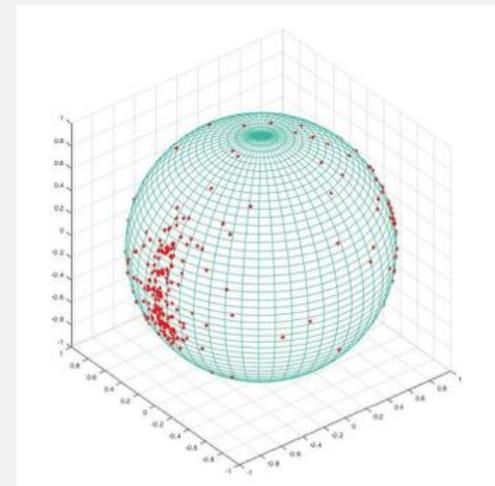
Відстань вимірюється за допомогою метрики Мінковського

$$d_{il}^{mum} = \left(\frac{1}{N} \sum_{j=1}^N |x_{ij} - x_{lj}|^p \right)^{\frac{1}{p}}$$

при $p=2$ – евклідова метрика,
при $p=1$ – манхеттенська метрика, для усіх $i=1,2,\dots,n$.

Нормування на гіперкулю:

1. центрування $\sum_{k=1}^k x(k) = 0$
2. нормування $\|x(k)\| = 1$



Відстанню є міра близькості об'єктів

$$\text{sim}(c_j(k), x(k)) = \cos \theta_j(k)$$

Нормування даних

Рангова метрика

кожен ознака x_i має R_i рангів, тобто об'єктом аналізу є відповідний ранг, $x_{ij}^{r_i} j = 1, 2, \dots, N$.

Далі вводиться в розподіл частот $f_i^{r_i} = \frac{N_i^{r_i}}{N}$

де $N_i^{r_i}$ - число появи r_i -го рангу в i -му стовпчику.

Далі вводяться частотні числові характеристики, які і кодуються на гіперкуб

$$x_{ij}^1 = 0,5 f_i^1, x_{ij}^{r_i} = x_{ij}^{r_i-1} + 0,5 (f_i^{r_i-1} + f_i^{r_i})$$

Відстань між двома пацієнтами в просторі буде визначатися як

$$0 \leq d_{il}^{rank} = \frac{1}{N} \sum_{j=1}^N |x_{ijr_i} - x_{ljr_i}| \leq 1$$

Категоріальна метрика

В якості категоріальної метрики використовується вираз:

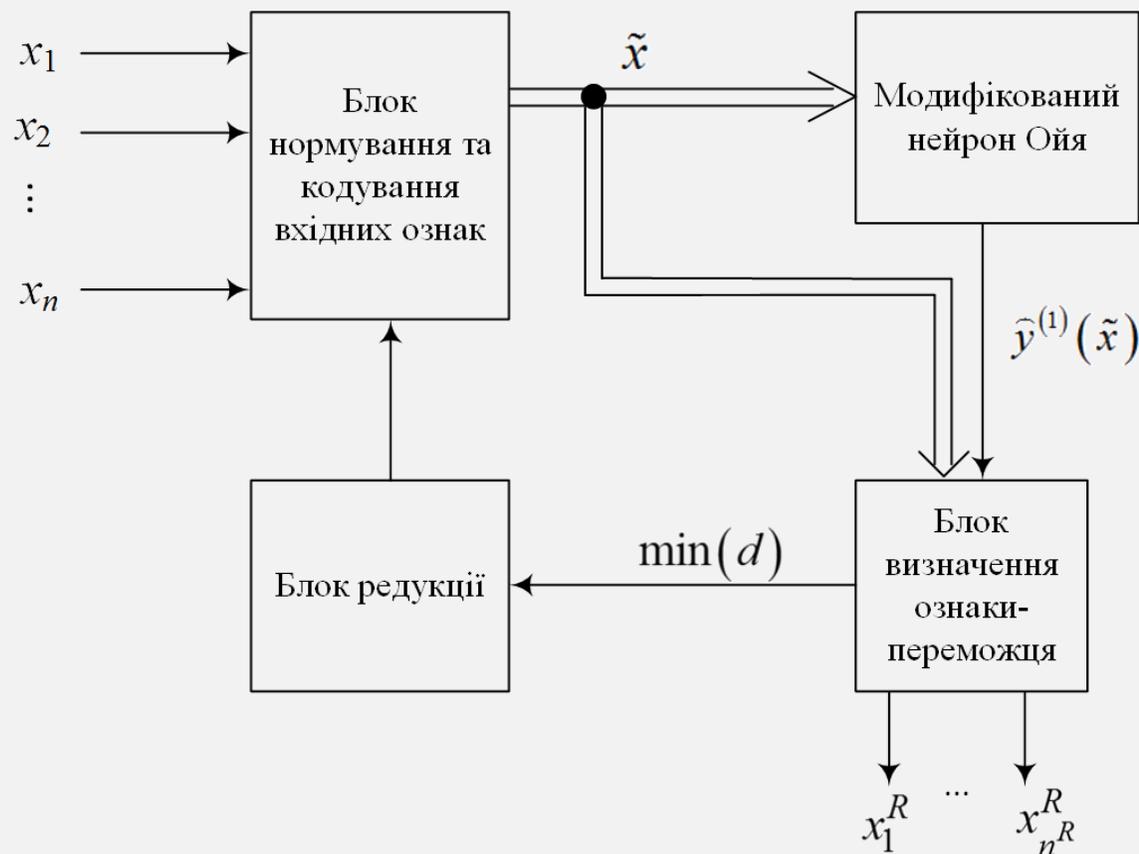
$$0 \leq d_{il}^{cat} = \frac{1}{N} \sum_{j=1}^N \delta(x_{ij}, x_{lj}) \leq 1; \delta(x_{ij}, x_{lj}) = \begin{cases} 1, & \text{если } x_{ij} \neq x_{lj} \\ 0, & \text{если } x_{ij} = x_{lj} \end{cases}$$

Загальна відстань між об'єктами аналізу:

$$0 \leq d_{il} = \left(\frac{1}{n_{num}} d_{il}^{num} + \frac{1}{n_{rank}} d_{il}^{rank} + \frac{1}{n_{cat}} d_{il}^{cat} \right) \leq 1$$

Кількісна метрика

Гібридна система оцінювання інформативності медичних показників Feature Selection-Extraction



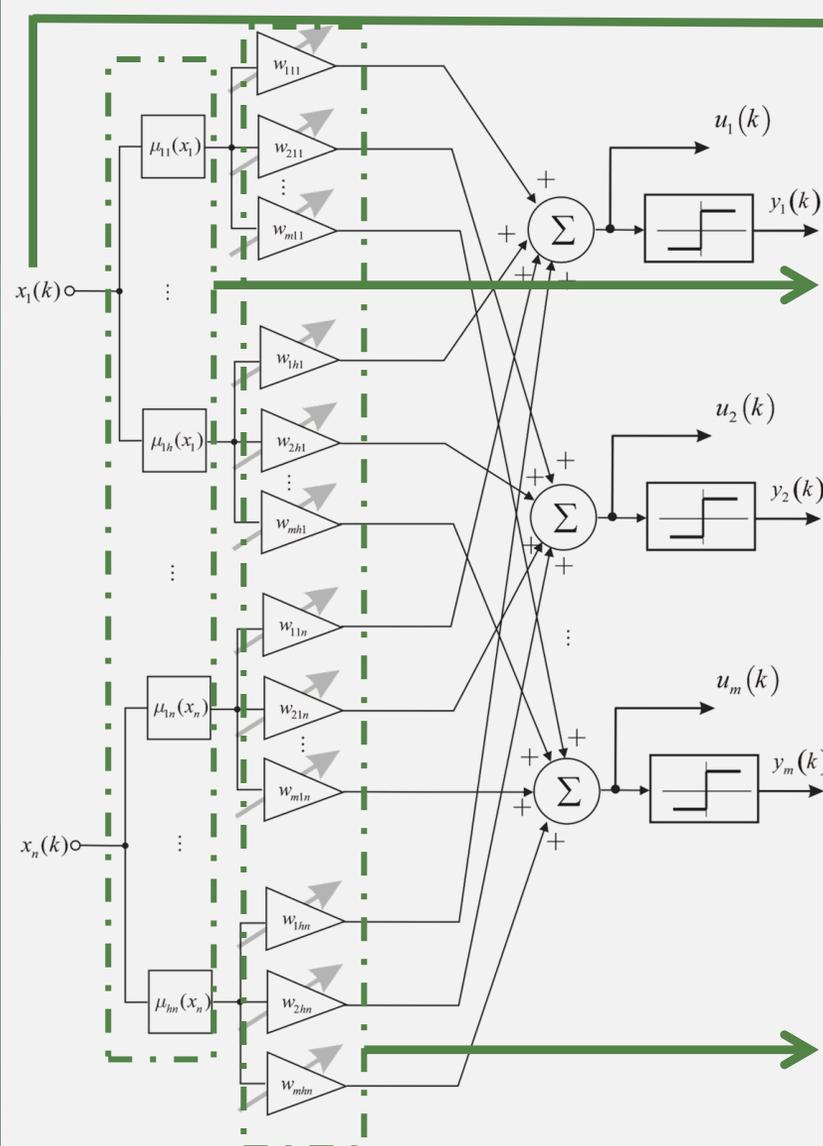
У блоці визначення ознаки-«переможця» визначається та ознака, відстань до якої у сенсі манхеттенської метрики до першої головної компоненти є найменшою згідно із:

$$d(\tilde{x}(z), \hat{y}^{(1)}) = \sum_{i=1}^N |\tilde{x}_i(z) - \hat{y}_i^{(1)}|.$$

Етап 1

Етап 2: діагностування (вибірка репрезентативна)

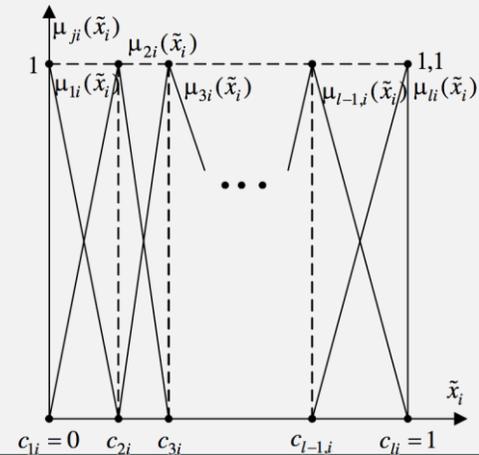
Модифікація багатовимірного нео-фаззі-нейрону



$$x(k) = (x_1(k), x_2(k), \dots, x_n(k))^T \in R^n$$

$$\mu_{li}(x_i) = \begin{cases} \frac{x_i - c_{l-1,i}}{c_{li} - c_{l-1,i}}, & x_i \in [c_{l-1,i}, c_{li}] \\ \frac{c_{l+1,i} - x_i}{c_{l+1,i} - c_{li}}, & x_i \in [c_{li}, c_{l+1,i}] \\ 0, & \text{otherwise,} \end{cases}$$

$$\sum_{l=1}^h \mu_{li}(x_i) = 1$$



Вектори вихідних сигналів $u(k) = W(k)\mu(k)$, $y(k) = \text{sign } u(k)$

Для навчання всіх виходів одночасно вводиться $(m \times nh)$ -матриця синаптичних ваг:

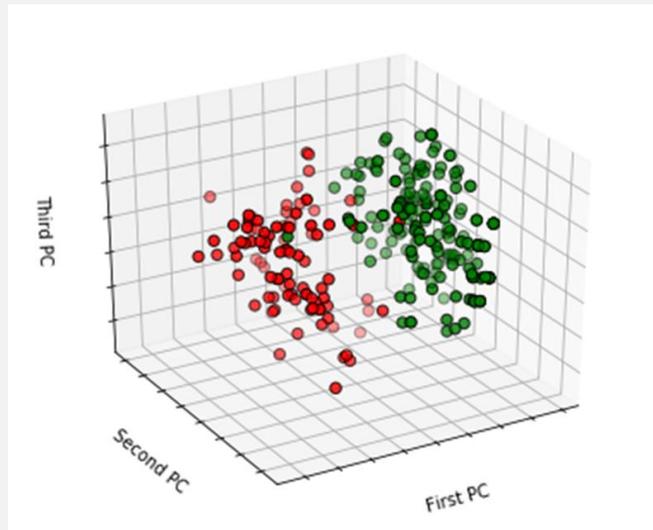
$$W(k) = \begin{pmatrix} w_{111}(k), \dots, w_{1h1}(k), \dots, w_{1li}(k), \dots, w_{1hn}(k) \\ w_{211}(k), \dots, w_{2h1}(k), \dots, w_{2li}(k), \dots, w_{2hn}(k) \\ \vdots \in \\ w_{m11}(k), \dots, w_{mh1}(k), \dots, w_{mli}(k), \dots, w_{mhn}(k) \end{pmatrix}$$

Багатовимірний аналог алгоритму навчання $\begin{cases} W(k+1) = W(k) + r^{-1}(k)(d(k) - \text{sign } W(k)\mu(k))\mu^T(k), \\ r(k) = \alpha r^{-1}(k-1) + \|\mu(k)\|^2, \quad 0 \leq \alpha \leq 1, \end{cases}$

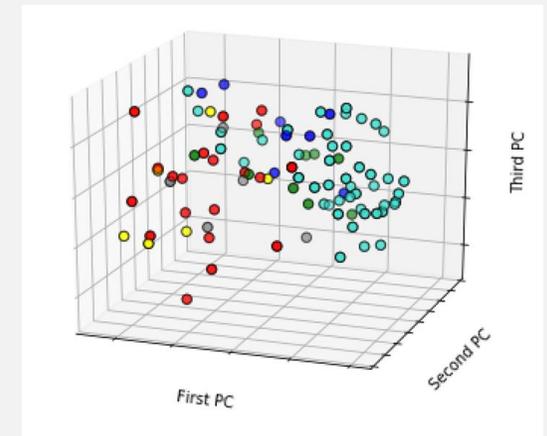
Апробація роботи на клінічних даних

Відсоток невірно класифікованих пацієнтів для всіх 106 ознак		Значення критерію ефективності роботи системи <i>PerCr</i>	Відсоток невірно класифікованих пацієнтів для найбільш інформативних 12 ознак		Значення критерію ефективності роботи системи <i>PerCr</i>
Навчальна вибірка	Тестова вибірка		Навчальна вибірка	Тестова вибірка	
7.19%	11.11%	0,085	5.3%	6.4%	0,0545

Дані містять 188 пацієнтів, кожен з яких описаний 106 ознаками та має один з 6 діагнозів (сечокам'яна хвороба (камені нирки та сечоводу), сечосоловий діатез, нефраптоз, гіпоплазія та пієлонефрит).



червоний колір – пацієнти із ефективністю медикаментозного лікування; зелений колір – пацієнти із фармакорезистентністю.



Дані містять 309 пацієнтів з епілепсією, кожен з яких описаний 25 ознаками.

Дані були поділені на навчальну та тестову вибірки у співвідношенні 206 пацієнтів на 103 пацієнти.

Навчання:

`numTrainErrors` = 2 пацієнти (0,97%)

PerCr = 0,0683.

Тестування:

`numTestErrors` = 6 пацієнтів (5.82%),

PerCr = 0,0501.

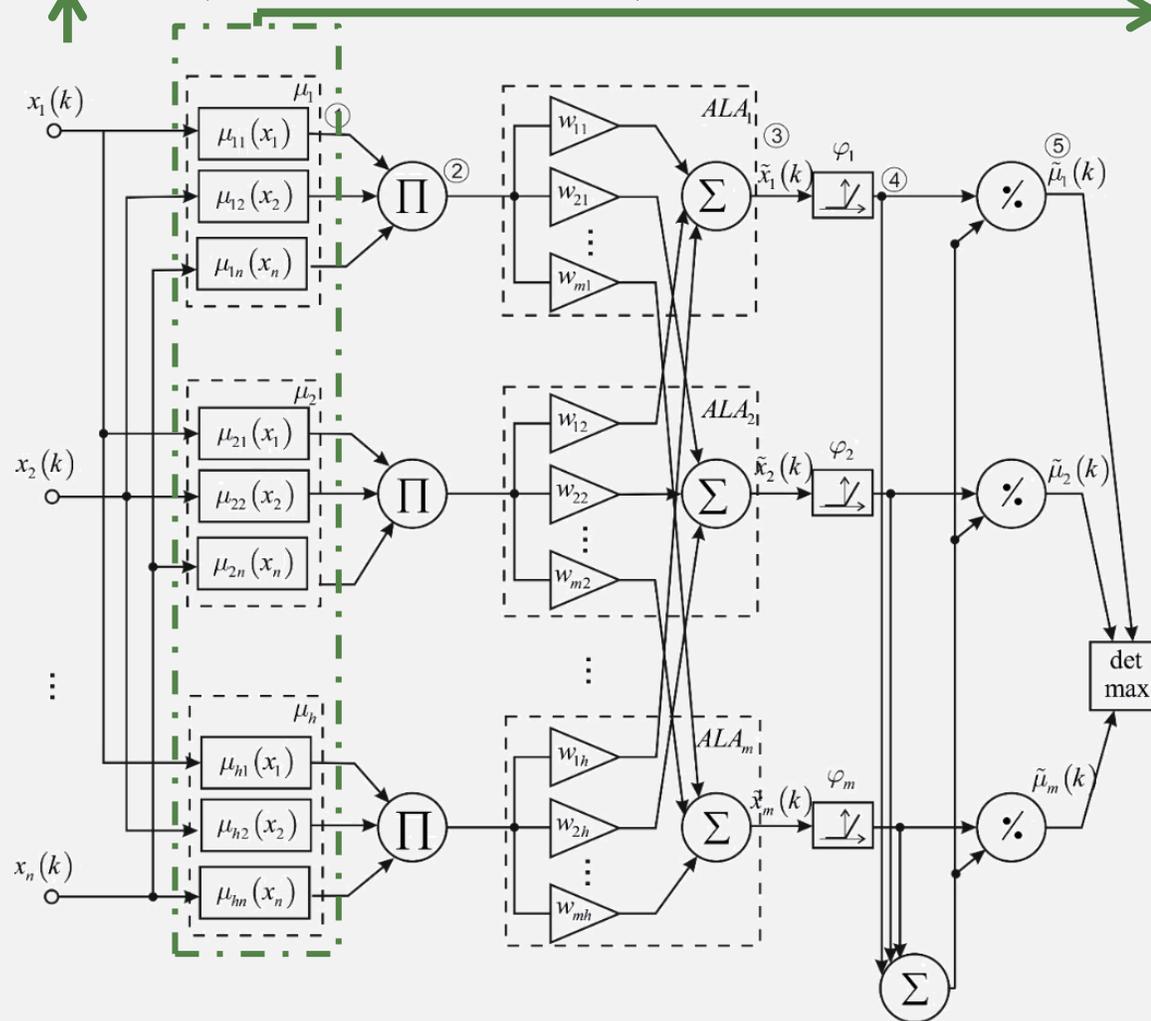
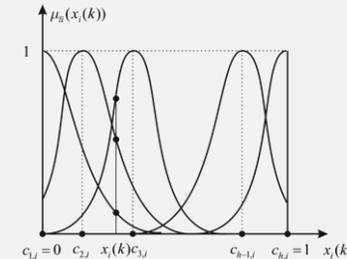
Етап 1

Етап 2: діагностування (вибірка репрезентативна)

Гібридна нейро-фаззі мережа для медичного діагностування DNFSm

$$x(k) = (x_1(k), x_2(k), \dots, x_n(k))^T \in R^n$$

$$\mu_{li}(x_i(k)) = \exp\left(-\frac{(x_i(k) - c_{li})^2}{2\sigma_i^2}\right)$$



2-й прихований шар

$$\hat{x}_l = \prod_{i=1}^n \mu_{li}(x_i(k))$$

3-й прихований шар

$$x_j(k) = \sum_{l=1}^h w_{jl} \prod_{i=1}^n \mu_{li}(x_i(k)) = \sum_{l=1}^h w_{jl} \hat{x}_l(k)$$

4-й прихований шар

$$u_j(k) = \varphi(x_j(k)) = \begin{cases} x_j(k) & \text{if } x_j(k) > 0, \\ 0 & \text{if } x_j(k) \leq 0, \end{cases}$$

5-й прихований шар

$$\mu_j(k) = \frac{u_j(k)}{\sum_{j=1}^m u_j(k)} = \frac{\varphi\left(\sum_{l=1}^h w_{jl} \hat{x}_l(k)\right)}{\sum_{j=1}^m \varphi\left(\sum_{l=1}^h w_{jl} \hat{x}_l(k)\right)}$$

6-й прихований шар

$$\mu^*(k) = \max\{\mu_1(k), \mu_2(k), \dots, \mu_m(k)\}$$

Етап 1

Етап 2: діагностування (вибірка репрезентативна)

Гібридна нейро-фаззі мережа для медичного діагностування DNFS_mn

Для навчання цієї мережі використовується квадратичний критерій:

$$E_j(k) = \frac{1}{2} \|d(k) - \tilde{x}(k)\|^2$$

Алгоритм налаштування синаптичних ваг має вигляд:

$$W(k+1) = W(k) + (d(k) - W(k)\hat{x}(k))\hat{x}^+(k),$$

де $d(k) = (d_1(k), d_2(k), \dots, d_m(k))^T$ навчальний сигнал (0 або 1)

$\hat{x}(k) = (\hat{x}_1(k)I_m^T, \hat{x}_2(k)I_m^T, \dots, \hat{x}_h(k)I_m^T)^T$ вихід другого прихованого шару,

$I_m = (m \times 1)$ вектор, утворений одиницями

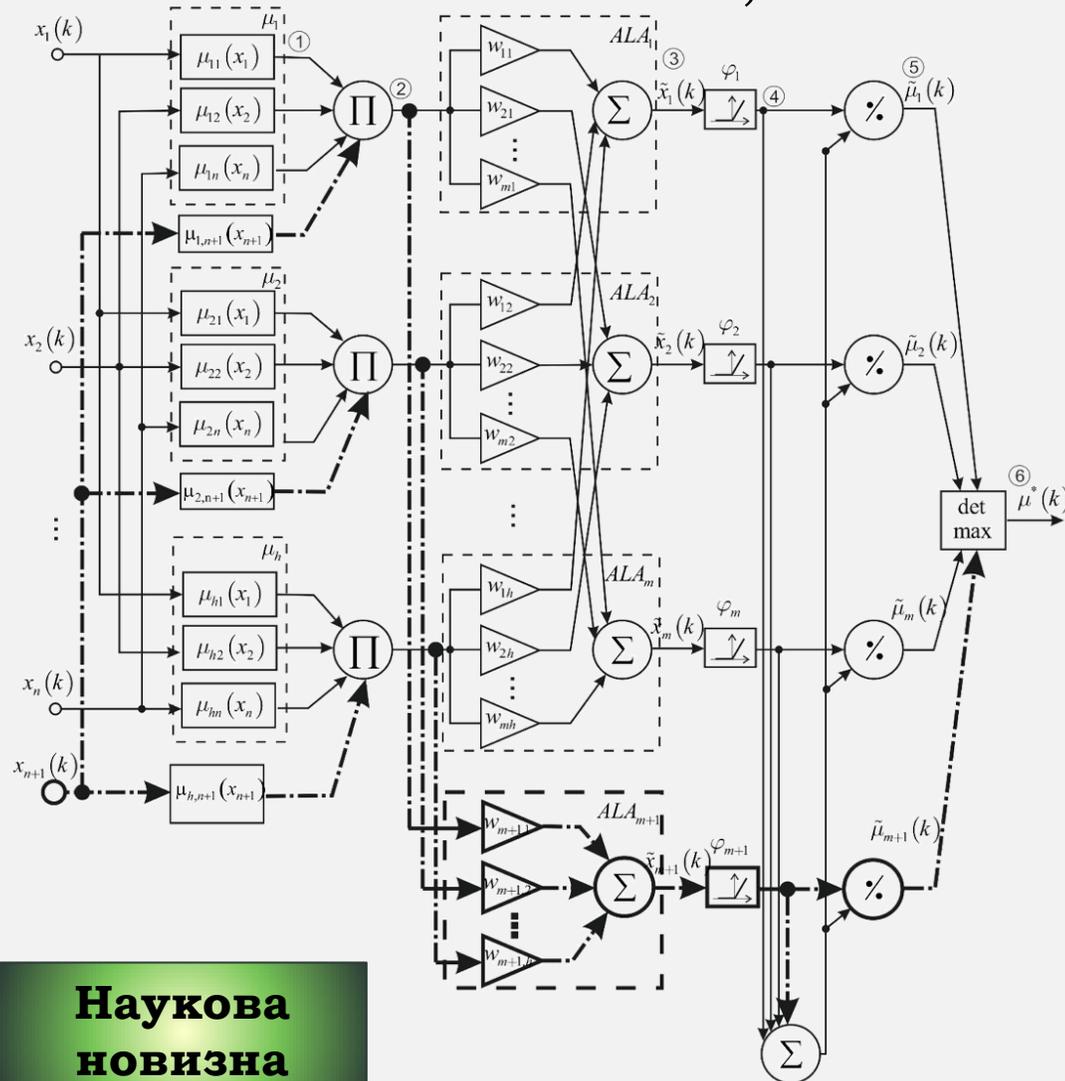
Також можна додати процедуру налаштування центрів функцій належності $\mu_{li}(x_i(k))$ і тим самим зробити можливу адаптацію системи DNFS до поточного набору даних

$$c(k+1) = c(k) - \eta(k) \frac{\partial E}{\partial c} \quad \text{де} \quad \frac{\partial E}{\partial c} = \frac{\partial E}{\partial e} \cdot \frac{\partial e}{\partial x} \cdot \frac{\partial x}{\partial \hat{x}} \cdot \frac{\partial \hat{x}}{\partial \mu} \cdot \frac{\partial \mu}{\partial c}$$

Етап 1

Етап 2: діагностування (вибірка репрезентативна)

Еволюційна глибинна гібридна нейро-фаззі мережа DNFS_{m+1,n+1}



Еволюція такої системи реалізується на рівні першого прихованого шару

$$\mu_{1,n+1}(x_{n+1}), \mu_{2,n+1}(x_{n+1}), \dots, \mu_{h,n+1}(x_{n+1})$$

Алгоритм переналаштування синаптичних ваг:

$$W^*(k+1) = \begin{pmatrix} W(k) \\ \dots \\ w_{m+1}^T(k) \end{pmatrix} + \left(\begin{pmatrix} d(k) \\ \dots \\ d_{m+1}(k) \end{pmatrix} - \begin{pmatrix} \tilde{x}(k) \\ \dots \\ \tilde{x}_{m+1}(k) \end{pmatrix} \right) \hat{x}^+(k)$$

Наукова
новизна

Апробація роботи DNFS_{m+1,n+1} (дані репозиторія)

Еволюціонування входів мережі

	Відсоток невірно класифікованих пацієнтів при n ознаках / час опрацювання 1 пацієнта		Відсоток невірно класифікованих пацієнтів при $n+1$ ознаках / час опрацювання 1 пацієнта	
	training	testing	training	testing
Iris	1,4% / 130 мкс	2,68% / 70 мкс	3,57% / 128 мкс	8,23% / 70 мкс
Dermatology	0,87% / 171 мкс	6,5% / 91 мкс	4,75% / 158 мкс	16,55% / 90 мкс
Pima Indians Diabetes	0,45% / 136 мкс	7,33% / 79 мкс	0,25% / 135 мкс	7,94% / 77 мкс
Parkinson	1,33% / 135 мкс	4,66% / 81 мкс	0,56% / 132 мкс	7,92% / 76 мкс

Еволюціонування виходів мережі

	Відсоток невірно класифікованих пацієнтів для m діагнозів / час опрацювання 1 пацієнта		Відсоток невірно класифікованих пацієнтів для $m+1$ діагнозів / час опрацювання 1 пацієнта	
	training	testing	training	testing
Iris	1% / 820 мкс	2% / 640 мкс	9% / 850 мкс	16,6% / 630 мкс
Dermatology	0,5% / 980 мкс	4% / 690 мкс	13,16% / 990 мкс	20% / 720 мкс

PerCr < 0,076

PerCr < 0,127

Метод активного навчання нейро-фаззі мережі для online медичного діагностування

Режим самонавчання для нерозмічених даних

1. Конкуренція. Розраховується відстань між центрами кластерів і пацієнтом в прийнятій метриці або у вигляді міри подібності:

$$d^2(c_j(k), x(k)) = \|x(k) - c_j(k)\|^2 \qquad \text{sim}(c_j(k), x(k)) = \cos \theta_j(k)$$

2. Кооперація і синаптична адаптація

$$\left\{ \begin{array}{l} c_p(k+1) = \frac{c_p(k) + \eta(k) u_p^2(k) (x(k) - c_p(k))}{\|c_p(k) + \eta(k) u_p^2(k) (x(k) - c_p(k))\|} \quad \forall p = 1, 2, \dots, m, \\ 0 \leq u_p(k) = \|x(k) - c_p(k)\|^{-2} \left(\sum_{l=1}^m \|x(k) - c_l(k)\|^{-2} \right)^{-1} \leq 1, \\ \eta(k) = r^{-1}(k), \quad r(k) = \alpha r(k-1) + 1, \quad 0 \leq \alpha \leq 1. \end{array} \right. \quad \left\{ \begin{array}{l} c_p(k+1) = \frac{c_p(k) + \eta(k) [\cos \theta_p(k)]_+ (x(k) - c_p(k))}{\|c_p(k) + \eta(k) [\cos \theta_p(k)]_+ (x(k) - c_p(k))\|} \quad \forall p = 1, 2, \dots, m, \\ 0 \leq u_p(k) = [\cos \theta_p(k)]_+ \left(\sum_{l=1}^m [\cos \theta_l(k)]_+ \right)^{-1} \leq 1, \\ \eta(k) = r^{-1}(k), \quad r(k) = \alpha r(k-1) + 1, \quad 0 \leq \alpha \leq 1. \end{array} \right.$$

Метод активного навчання нейро-фаззі мережі для online медичного діагностування

Режим контрольованого навчання для розмічених даних

Якщо пацієнт з відомим діагнозом потрапляє в комірку Вороного, яка відповідає відомому діагнозу цього пацієнта, то центр кластера підтягуються до цього пацієнта.

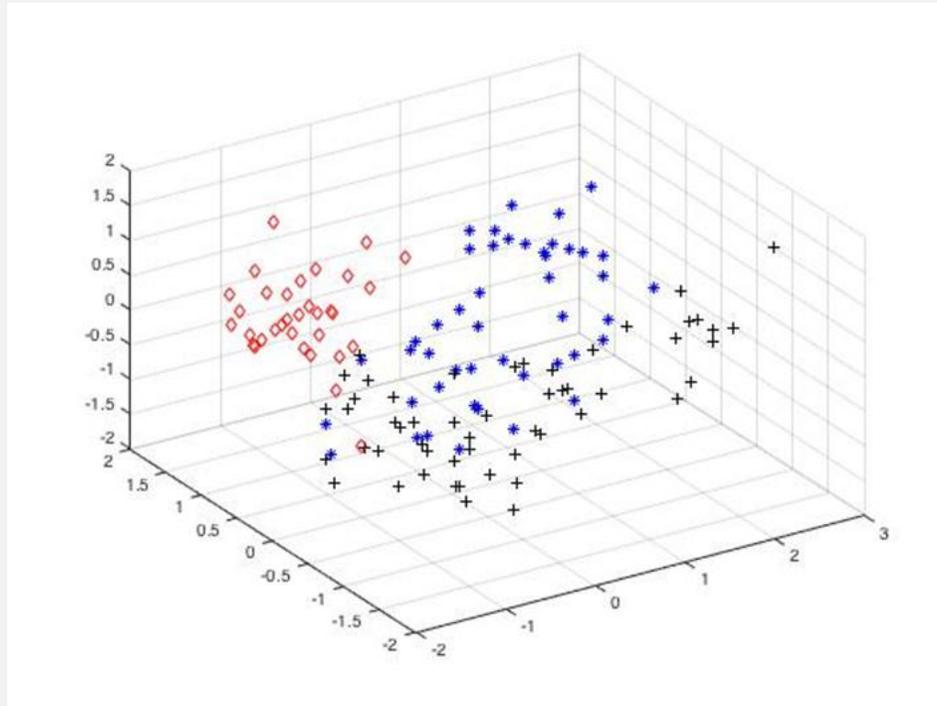
$$\begin{cases} c_p(k+1) = \frac{c_p(k) + \eta(k) u_p^*(k) (x(k) - c_p(k))}{\|c_p(k) + \eta(k) u_p^*(k) (x(k) - c_p(k))\|}, \quad \forall p = 1, 2, \dots, q, \\ \eta(k) = r^{-1}(k), \quad r(k) = \alpha r(k-1) + 1, \quad 0 \leq \alpha \leq 1. \end{cases}$$

Якщо пацієнт з відомим діагнозом потрапляє в комірку Вороного, яка не відповідає відомому діагнозу цього пацієнта, то центр цього кластера відштовхується від цього пацієнта.

$$\begin{cases} c_j(k+1) = \frac{c_j^*(k) - \eta(k) \frac{(1 - \cos \theta_j(k))^{-1}}{\sum_{l=1}^m (1 - \cos \theta_l(k))^{-1}} (x(k) - c_j^*(k))}{\left\| c_j^*(k) - \eta(k) \frac{(1 - \cos \theta_j(k))^{-1}}{\sum_{l=1}^m (1 - \cos \theta_l(k))^{-1}} (x(k) - c_j^*(k)) \right\|}, \\ \eta(k) = r^{-1}(k), \quad r(k) = \alpha r(k-1) + 1, \quad 0 \leq \alpha \leq 1. \end{cases}$$

де $\frac{(1 - \cos \theta_j(k))^{-1}}{\sum_{l=1}^m (1 - \cos \theta_l(k))^{-1}}$ – нечітка імовірнісна функція «відштовхування» центроїда.

Апробація роботи нейро-фаззі системи з активним навчанням на клінічних даних



◇ - пацієнти із захворюванням на пневмонію; * - пацієнти з хронічним обструктивним захворюванням легень; + - пацієнти з бронхіальною астмою

Загальна кількість пацієнтів складає 132 людини з пульмонологічними захворюваннями.

Кожен пацієнт характеризується 104 ознаками.

46 пацієнтів мають ХОЗЛ;

53 пацієнти – бронхіальну астму;

33 пацієнти – пневмонію.

У 80% пацієнтів був штучно видалений діагноз (106 пацієнтів в режимі самонавчання, 26 пацієнтів в режимі контрольованого навчання).

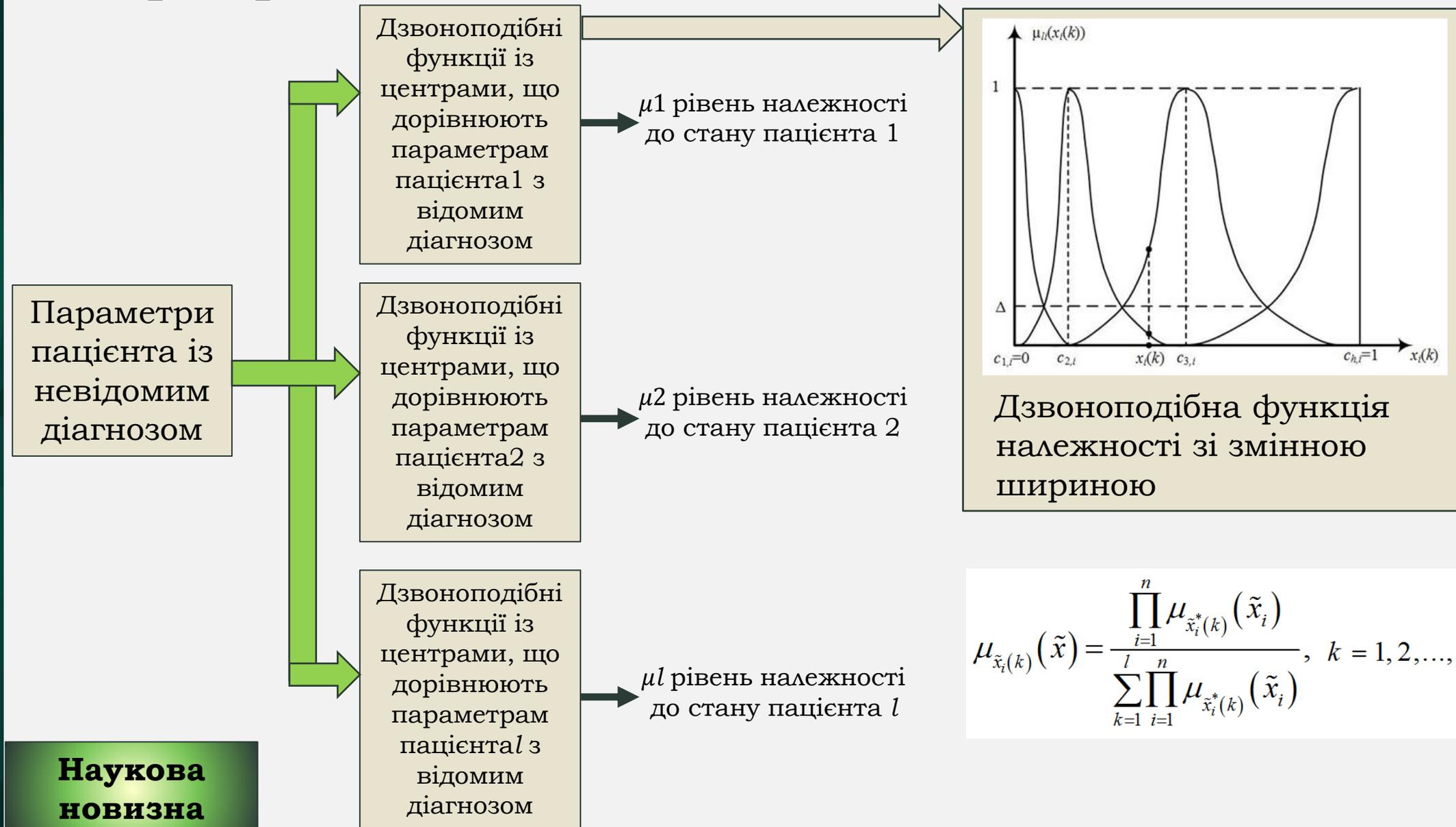
Нейро-фаззі система із активним навчанням продемонструвала відсоток вірно класифікованих пацієнтів на рівні 84,38%.

$PerCr = 0,096$

Етап 1

Етап 2: діагностування (вибірка нерепрезентативна)

Нейро-фаззі автоасоціативна пам'ять



Апробація роботи на клінічних даних

Варіанти найтипівіших представників	Значення параметру	Відсоток неправильно асоційованих пацієнтів	<i>PerCr</i>
ID25	0.2	18,15	0,469
	0.5	24,7	0,502
ID64	0.2	21,44	0,486
	0.5	35,8	0,558
ID18	0.2	26,7	0,512
	0.5	30,66	0,532
ID32	0.2	19,75	0,477
	0.5	25,78	0,507
ID65	0.2	17,89	0,468
	0.5	19,9	0,478
ID70	0.2	21,33	0,485
	0.5	30,02	0,529

Кількість пацієнтів склала 95 осіб, кожен з яких має 44 ознаки.

Кількість діагнозів – 2: ішемічна хвороба серця, гіпертонічна хвороба.

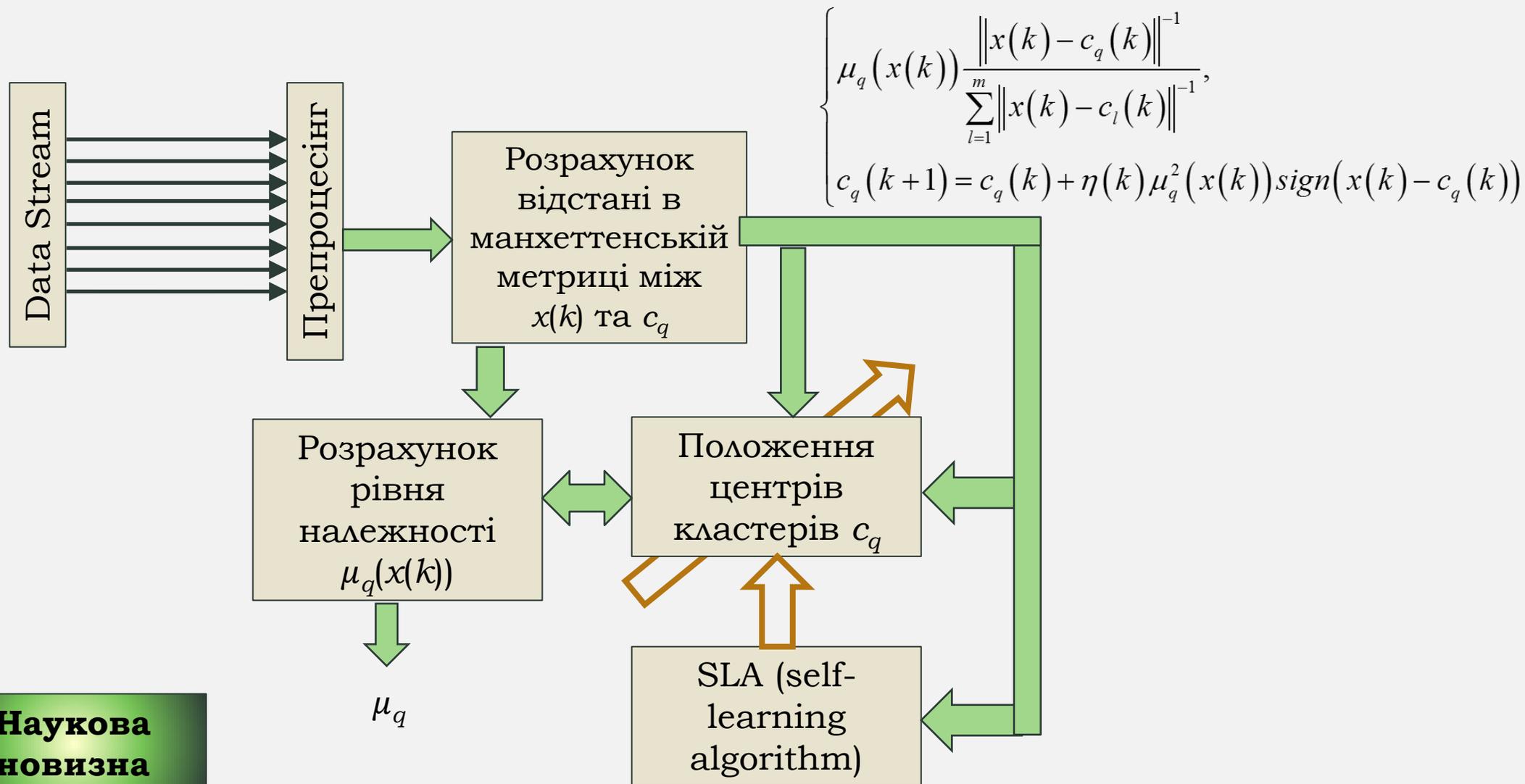
Пацієнти з ID18, ID25, ID64 є найтипівішими з діагнозом гіпертонічна хвороба.

Пацієнти з ID32, ID65, ID70 є найтипівішими з діагнозом ішемічна хвороба серця.

Етап 1

Етап 2: діагностування (вибірка нерепрезентативна)

Метод адаптивної робастної нечіткої кластеризації на основі манхеттенської метрики



**Наукова
новизна**

Наукова новизна отриманих результатів:

1. Вперше запропоновано метод оцінювання інформативності медичних показників, який на основі гібридизації підходів визначення першої головної компоненти за допомогою нейрону Ойя та методів виділення інформативних ознак з множини наявних забезпечує попереднє оброблення медичних даних у послідовному онлайн-режимі та уможливає виділення з потоку медичних даних найінформативніших медичних ознак зі збереженням їхнього змістовного інтерпретування.
2. Удосконалено метод просторової екстраполяції діагностичних ознак пацієнта шляхом використання нечіткого рівня належності, що надає можливість виконання обґрунтованого заповнення пропусків у потоках медичних даних.
3. Удосконалено метод формування багатовимірного нео-фаззі нейрона за рахунок введення в його архітектуру нелінійних активаційних функцій, що надає можливість використовувати його у завданнях нечіткого онлайн-діагностування в умовах наявності класів-діагнозів з перетином у просторі ознак та одночасно підвищити як точність діагностування, так і швидкодію роботи системи.
4. Вперше запропоновано метод формування еволюційної нейро-фаззі мережі в умовах змінної кількості медичних ознак і діагнозів, який відрізняється здатністю адаптувати свою архітектуру в разі зміни розмірності простору вхідних ознак та діагнозів, що дає змогу підвищити якість діагностування за рахунок використання усіх виміряних даних.
5. Вперше запропоновано метод активного навчання нейро-фаззі мережі для медичного діагностування, який здійснює автоматичне перемикання між режимами контрольованого навчання та самонавчання і забезпечує процес налаштування синаптичних ваг нейро-фаззі мережі за ознаками пацієнта як з відомим діагнозом (контрольоване навчання), так і з невідомим (самонавчання), що дає змогу зменшити необхідні обсяги даних у навчальній вибірці та час налаштування мережі.
6. Вперше запропоновано метод адаптивної робастної нечіткої кластеризації багатовимірного простору ознак пацієнта, який використовує манхеттенську метрику як основу нечіткої належності, що надає можливість розв'язувати завдання кластеризації в умовах спотвореності медичних даних аномальними викидами.
7. Вперше розроблено метод асоціативної кластеризації на основі нейро-фаззі автоасоціативної пам'яті, який відрізняється тим, що послідовно уточнює свої параметри під час надходження нових медичних даних та автоматично реалізує асоціативне нечітке висновування на основі аналізу прецедентів, завдяки чому до аналізу додаються медичні випадки з нетиповим поєднанням симптомів.

Практичне значення отриманих результатів:

Розроблене в роботі математичне забезпечення надає можливість заповнювати пропущені значення в медичних показниках пацієнта в послідовному режимі для максимально ефективного використання усіх вимірних ознак, проводити оцінювання інформативної цінності кожної з ознак для обґрунтування необхідності їхнього подальшого використання або вилучення, виконувати діагностування у послідовному онлайн-режимі з розрахунком нечіткої функції належності кожного пацієнта до кожного діагнозу.

Результати роботи апробовано і впроваджено:

- КЗОЗ «Харківська міська клінічна лікарня №13», м. Харків;
- у відділі популяційних досліджень, відділі науково-організаційної роботи та медичної інформації з бібліотекою та відділі атеросклерозу та ішемічної хвороби серця ДУ «Національний інститут терапії ім. Л.Т. Малої національної академії наук України», м. Харків;
- ТзОВ «Клініка сучасних методів діагностики та інноваційних медичних технологій «Водолій», м. Хуст, Закарпатська обл.;
- в діагностично-лікувальний медичний центр «Ужмед», м. Ужгород;
- в інституті електроніки та інформаційних технологій Люблінського технологічного університету, м. Люблін, Польща;
- на кафедрі гігієни та екології №2 Харківського національного медичного університету.

Висновки

У роботі вирішено актуальну науково-технологічну проблему розроблення методів для аналізу потоків медичної інформації в послідовному режимі з використанням підходів Medical Data Mining для підтримки впровадження системи eHealth в Україні і підвищення ефективності медичного діагностування в цілому. Ефективність розроблених методів і засобів підтверджено під час їхнього застосування для розв'язання завдань медичного діагностування за клінічними даними.

Основні наукові й практичні результати роботи:

1. На основі аналізу підходів, методів та систем діагностування і препроцесингу медичних даних виявлено низьку ефективність їхньої роботи з потоками таких даних, встановлено необхідність об'єднання (гібридизації) кількох типів систем аналізу даних в одну та розроблення методів навчання таких систем з урахуванням особливостей первинних медичних даних, що дало змогу визначити необхідні вимоги до програмно-технічних складників систем діагностування, спрямовані на підвищення ефективності медичного діагностування.
2. Удосконалений метод просторової екстраполяції діагностичних ознак пацієнта за рахунок використання нечіткого рівня належності в режимі заповнення пропущених значень надав можливість виконання обґрунтованого заповнення пропусків у початкових медичних даних з метою їхнього подальшого використання для розв'язання завдань діагностування.
3. Розроблення гібридного методу оцінювання інформативності медичних показників за рахунок доповнення нейрону Ойя контуром адаптивного налаштування дає змогу зменшити час опрацювання інформації та підвищити швидкодію процесу самонавчання нейронної мережі для виділення інформативних ознак, що надало можливість виділити з потоку медичних даних найінформативніших медичних ознак зі збереженням їхнього змістовного інтерпретування, що сприяє кращому розумінню процесу подальшого діагностування лікарями.
4. Розроблені методи аналізу медичних даних під час медичного онлайн-діагностування в режимі контрольованого навчання на основі гібридних адаптивних нейро-фаззі систем для ситуацій репрезентативної навчальної вибірки уможливили проведення нечіткого онлайн-діагностування з перетином класів-діагнозів у просторі ознак з підвищенням точності діагностування і швидкодії роботи системи. Під час використання модифікованого багатовимірного нео-фаззі нейрону значення критерію ефективності для пацієнтів із сечокам'яною хворобою з повною кількістю ознак складає $PerCr=0,085$ (точність діагностування 88,89%), з залученням тільки найінформативніших ознак $PerCr=0,054$ (точність діагностування 93,6%). Точність діагностування у модифікованого багатовимірного нео-фаззі нейрона для пацієнтів із фармакорезистентною формою епілепсії склала 94.18%, значення критерію ефективності – $PerCr=0,0501$.
5. Розроблений метод аналізу медичних даних під час медичного онлайн-діагностування в режимі контрольованого навчання з можливістю архітектурної адаптації нейро-фаззі мережі для розширення простору медичних ознак та діагнозів дав змогу підвищити якість діагностування за рахунок використання всіх вимірних даних.
6. Запропонований метод аналізу медичних даних під час медичного онлайн-діагностування в режимі активного навчання на основі гібридної нейро-фаззі мережі здійснює автоматичне перемикавання між режимами контрольованого навчання та самонавчання і забезпечує відповідні зміни процесу налаштування ваг нейро-фаззі мережі за індивідуальними ознаками пацієнта, допомагає зменшити необхідні обсяги даних у навчальній вибірці та час налаштування мережі. Значення критерію ефективності на даних пульмонологічних досліджень склало $PerCr=0,0961$ (точність 84,38%).
7. Розроблений метод аналізу медичних даних під час медичного онлайн-діагностування в режимі самонавчання, який базуються на адаптивній робастній нечіткій кластеризації багатовимірного простору ознак пацієнта з використанням манхеттенської метрики як основи нечіткої належності, надав можливість розв'язувати завдання кластеризації в умовах спотворення медичних даних аномальними викидами.
8. Розроблений метод аналізу медичних даних під час медичного онлайн-діагностування в режимі самонавчання, який базується на асоціативній кластеризації на основі нейро-фаззі автоасоціативної пам'яті, яка послідовно уточнює свої параметри під час надходження нових медичних даних та автоматично реалізує асоціативне нечітке висновування на основі аналізу прецедентів, дав змогу долучати до аналізу медичні випадки з нетиповим поєднанням симптомів. Значення критерію ефективності у діагностуванні гіпертонічної хвороби та ішемічної хвороби серця склало $PerCr=0,468$, значення точності діагностування 82,11%.