

УДК 004.032.26

О.К. КОЛЕСНИЦЬКИЙ, Ю.О. ЖУРАВСЬКА

## ЕКСПЕРИМЕНТАЛЬНІ ДОСЛІДЖЕННЯ СИСТЕМИ МЕДИЧНОГО ДІАГНОСТУВАННЯ ПАЦІЄНТІВ ЗА АНАЛІЗОМ КРОВІ НА ОСНОВІ НЕЙРОННОЇ МЕРЕЖІ КОХОНЕНА

*Вінницький національний технічний університет,  
вул. Хмельницьке шосе, 95, м. Вінниця, Україна  
Телефон: (0432) 56-08-48, факс: (0432) 46-57-72,  
E-mail: vntu@vntu.vinnica.ua*

**Анотація:** В даній статті описано результати експериментальних досліджень програмної системи медичного діагностування пацієнтів за аналізом крові. У запропонованій системі використовується нейронна мережа Кохонена, в якій центри кластерів визначаються заздалегідь згідно з нормальними та відхиленими від норми значеннями показників загального аналізу крові.

Ключові слова: нейронна мережа Кохонена, кластеризація, медична діагностика.

**Анотация:** В данной статье описаны результаты экспериментальных исследований программной системы медицинского диагностирования пациентов по анализу крови. В предлагаемой системе используется нейронная сеть Кохонена, в которой центры кластеров определяются заранее согласно с нормальными и отклоненными от нормы значениями показателей общего анализа крови. Ключевые слова: нейронная сеть Кохонена, кластеризация, медицинская диагностика.

**Abstract:** This paper describes the results of experimental researches of program system of the medical diagnostics of patients based on blood test. The proposed system uses a Kohonen neural network, in which the cluster centers are determined in advance in accordance with normal and deviations from the normal values of the parameters of a general blood test.

**Keywords:** Kohonen neural network, clustering, medical diagnostics.

### ВСТУП

На сьогоднішній день існуючий процес діагностування пацієнтів державних лікарень є досить довгим та багатоетапним, оскільки для підтвердження або спростування діагнозу потрібно здати низку аналізів, що може призвести через довготривалість даного процесу до загострення хвороби або поставлення хибного діагнозу. Існуючі автоматизовані методи діагностики є недостатньо ефективними – більшість систем орієнтована на визначення конкретного захворювання, що через специфіку медичної галузі, має високу ймовірність виявитись хибним, дані системи мають невиправдану складність алгоритмів та їх програмної реалізації та загалом можуть використовуватись лише медичними фахівцями.

Таким чином, існує проблема недосконалості сучасних методів діагностування, що не забезпечують високу точність виявлення захворювання та зручність використання із залученням персональних мобільних засобів, не можуть бути використані пацієнтами дистанційно без відвідування лікарні [1].

Для усунення даної проблеми пропонується підхід, заснований на визначенні показників загального аналізу крові, що проводиться при більшості захворювань і профілактичних обстеженнях, і включає в себе 13 показників, кожен з яких при відхиленні від норми передбачає схильність до певних хвороб [2]. Діагностування за допомогою аналізу отриманих показників аналізу крові допоможе виявити тип захворювання пацієнта та надати рекомендації щодо лікування.

Для реалізації даного підходу використана кластеризація даних, що забезпечує високу точність їх обробки, та при подальшій програмній реалізації має нескладну програмну структуру. Для кластеризації результатів показників загального аналізу крові потрібні лише числові значення та велика точність в їх визначенні, тому для реалізації даної задачі була обрана мережа Кохонена [3,4].

Отже, метою дослідження є доведення правильності функціонування та визначення достовірності роботи запропонованої системи медичного діагностування пацієнтів за аналізом крові шляхом проведення експериментів над нею. Експерименти полягають у порівнянні кількості правильно поставлених діагнозів

запропонованою системою, робота якої основана на використанні модифікованої мережі Кохонена, з системою, що використовує стандартну мережу Кохонена та з програмами-аналогами. Також буде проведене дослідження, що полягає у визначенні достовірності роботи системи з різними обсягами навчальної вибірки, а також визначення точності діагностування на навчальній вибірці.

### РОБОТА МЕРЕЖІ КОХОНЕНА

Оскільки стандартна реалізація мережі Кохонена не передбачає визначеної кількості кластерів, та їх початкові центри приймають випадкові малі значення, була проведена модифікація даної мережі, що полягає у попередньому визначенні кількості кластерів та їх центрів, які відповідають нормальним та відхиленням від норми значенням показників загального аналізу крові.

Задача, що розглядається, запропонована до вирішення модифікованою мережею Кохонена. Дана мережа використовує неконтрольоване навчання та навчальна множина складається лише із значень вхідних змінних.

Мережа розпізнає кластери в навчальних даних і розподіляє дані до відповідних кластерів. Якщо в наступному мережа зустрічається з набором даних, несхожим на жодний із відомих зразків, вона відносить його до нового кластеру. Якщо в даних містяться мітки класів, то мережа спроможна вирішувати задачі класифікації [5].

Шар Кохонена складається з деякої кількості  $n$  паралельно діючих лінійних елементів. Всі вони мають однакову кількість входів  $m$  та отримують на свої входи один і той самий вектор вхідних сигналів  $x=(x_1 \dots x_m)$ . На виході  $j$ -го лінійного елемента отримуємо сигнал, що розраховується за формулою 1.

$$y_j = w_{j0} \sum_{i=1}^m w_{ji} x_i, \quad (1)$$

де  $w_{ji}$  — ваговий коефіцієнт  $i$ -го входу  $j$ -го нейрону,  $w_{j0}$  — пороговий коефіцієнт.

Після проходження шару лінійних елементів сигнали посилаються на обробку за правилом «переможець забирає все»: серед вихідних сигналів  $y_i$  шукається максимальний; його номер  $j_{\max} = \arg \max_j \{y_j\}$ . Остаточо, на виході сигнал з номером  $j_{\max}$  дорівнює одиниці, всі інші — нулю. Якщо максимум одночасно досягається для декількох  $j_{\max}$ , то приймають всі відповідні сигнали рівними одиниці.

Навчання розроблюваної мережі відбувається наступним чином. На початку роботи визначається кількість кластерів та їх центри. Після цього деякий вхідний вектор з набору навчальних вибирається і встановлюється на вході нейронної мережі. На цьому етапі відмінності між вхідним вектором та всіма векторами обчислюються за формулою 2.

$$D_{ij} = |X^l - W_{ij}| = \sqrt{(x_1 - w_{ij1})^2 + \dots + (x_n - w_{ijn})^2}, \quad (2)$$

де  $i$  та  $j$  — показники нейронів у вихідному шарі. Після цього нейронна мережа обирає нейрон-переможець з переліку визначених центрів кластерів, тобто такий, щоб його ваговий вектор був схожий на вхідний за формулою 3.

$$D(k_1, k_2) = \min_{i,j} D_{i,j}, \quad (3)$$

де  $k_1$  та  $k_2$  — показники нейрона-переможця. Після цього проводиться корекція вагових векторів переможця та сусідніх нейронів. Близькі нейрони до переможця визначаються топологічною функцією сусідства, яка розраховується за формулою 4.

$$h(p, t) = \exp\left(-\frac{p^2}{\sigma^2(t)}\right) \left(1 - \frac{2}{\sigma^2(t)} p^2\right), \quad (4)$$

де  $p$  — відстань до нейрона переможця, яка знаходиться за формулою 5.

$$p = \sqrt{(k_1 - i)^2 + (k_2 - j)^2}, \quad (5)$$

де  $\sigma$  – функція, що визначає радіус сусідства. На початку функціонування програмного модулю вона включає весь простір сенсорного поля (сітки), але з часом, значення її зменшується.

Після обчислення топологічної функції ваги всіх нейронів переобчислюються за формулою 6.

$$W_{ij}(t+1) = W_{ij}(t) + \alpha(t)h(p,t)(X^l(t) - W_{ij}(t)), \quad (6)$$

де  $\alpha(t)$  – функція швидкості навчання, яка також змінюється з часом. Якщо нейрон є переможцем, або сусіднім до нього, його вектор ваг оновлюється або залишається незмінним в іншому випадку. На кожному кроці нейронна мережа визначає нейрон, чий ваговий вектор найбільш схожий до вхідного, та коригує його ваги та ваги сусідів, щоб наблизити їх до вхідного вектора.

Кожен вхідний вектор з навчальної вибірки представляється нейронній мережі, і навчання триває або деяке фіксоване число циклів, або доки різниця між вхідним і ваговим векторами досягне заданого значення  $\epsilon$ . Різниця між сусідніми нейронів зменшується з плином часу, і, отже, вони організуються в групи (кластери), які відповідають одному з класів з навчальної множини [6, 7].

Для розпізнавання декількох діагнозів після визначення одного проводиться повторне діагностування з виключенням визначеної хвороби до того моменту, доки пацієнт згідно з проведеним діагностуванням не виявиться здоровим.

Для оптимальних розрахунків мережі була також введена процедура попередньої обробки вхідних даних, при якій значення ознак, що утворюють вхідний вектор, приводяться до деякого заданому діапазону. Нормалізація необхідна, оскільки вихідні значення ознак змінюються в достатньо великому діапазоні, тому існує ймовірність некоректної роботи нейронної мережі з такими даними. Так, в одному вхідному векторі можуть міститися значення, що відрізняються один від одного на кілька порядків: стандартний рівень гемоглобіну – 120-180г/л, колірного показнику – 0,85-1,15%. Такий дисбаланс між значеннями показників може викликати нестійкість роботи моделі, погіршити результати навчання і уповільнити його процес. Після нормалізації всі значення вхідних ознак будуть приведені до деякого вузького діапазону, а саме - [0 ... 1] що дозволяє мережі працювати з даними більш коректно [8].

Для реалізації запропонованого підходу до діагностування була використана нормалізація вхідних даних за формулою 7.

$$y = \frac{(x - x_{\min})(d_2 - d_1)}{x_{\max} - x_{\min}} + d_1, \quad (7)$$

де:  $x$  - значення, що підлягає нормалізації;  $[d_1, d_2]$  - інтервал значень  $x$ ;  $[x_{\min}, x_{\max}]$  - інтервал, до якого буде зведене значення  $x$ .

### РЕАЛІЗАЦІЯ ЗАПРОПОНОВАНОГО ПІДХОДУ

Описаний підхід до діагностування був реалізований за допомогою мови програмування C# (рис.1), що дозволяє визначити діагноз впродовж хвилини.

Вхідні дані										Результати									
Завантажити результати аналізів крові з файлу.																			
Діагнози пацієнтів																			
A1	x	120	3.7	0.93	0.5	242	8	6	2	61	4	0.4	29	2	пацієнт A1 - здорова				
A2	x	121	4.1	0.97	0.9	216	8	7	3	57	7	0.6	37	8	пацієнт A2 - можлива астма				
A3	x	157	4.1	0.90	0.9	299	5	4	1	55	6	0.5	34	6	пацієнт A3 - можлива астма				
A4	x	153	4.9	0.91	1.2	346	15	11	7	72	5	0.4	24	8	пацієнт A4 - можливе запалення в організмі				
A5	x	140	4.2	0.99	1.2	301	7	4	3	58	2	0.6	19	4	пацієнт A5 - здорова				
A6	x	144	3.8	1.1	0.7	232	5	3	1	52	2	0.5	19	8	пацієнт A6 - можливі проблеми з селезінкою				
A7	x	158	4.0	0.91	0.2	206	4	4	2	66	5	0.7	22	3	пацієнт A7 - здоровий				
A8	x	145	4.2	1.05	1.3	192	4	4	1	53	3	0.3	39	9	пацієнт A8 - можливі проблеми з нирками				
A9	x	146	3.7	1.04	0.4	380	16	10	7	84	5	0.4	36	7	пацієнт A9 - можливе запалення в організмі				
A10	x	159	5.7	1.14	0.2	225	6	5	6	71	2	0.2	27	7	пацієнт A10 - здоровий				
B1	x	151	3.9	0.99	1.0	186	5	3	6	72	3	0.5	35	3	пацієнт B1 - можливі проблеми з селезінкою				
B2	x	159	3.7	1.02	0.5	215	5	6	5	56	2	0.3	43	11	пацієнт B2 - ймовірність захворювання туберкульозом				
B3	x	157	4.2	0.92	0.9	271	4	4	3	58	2	0.2	19	9	пацієнт B3 - здоровий				
B4	x	133	3.9	1.12	0.6	256	10	4	2	48	6	0.8	21	2	пацієнт B4 - можлива астма				
B5	x	188	5.7	1.17	0.2	300	6	6	4	62	4	0.6	21	3	пацієнт B5 - можливе зневоднення організму				
B6	x	151	4.0	1.12	0.2	239	6	5	2	48	3	0.5	26	5	пацієнт B6 - здорова				
B7	x	141	3.9	0.92	0.2	205	5	4	1	57	7	0.8	38	9	пацієнт B7 - здорова				
B8	x	157	4.3	1.03	0.2	252	6	6	4	67	4	0.4	31	3	пацієнт B8 - можливі проблеми з щитоподібною залозою				
B9	x	142	4.8	0.94	0.3	218	5	5	2	51	2	0.3	29	7	пацієнт B9 - здоровий				
B10	x	134	5.1	0.74	1.3	172	12	4	1	62	3	0.4	30	3	пацієнт B10 - можлива анемія				
C1	x	140	4.7	1.01	0.6	214	7	4	1	55	2	0.2	37	7	пацієнт C1 - можливі проблеми з цитоподібною залозою				
C2	x	157	4.9	0.98	0.9	209	6	5	2	66	4	0.4	23	5	пацієнт C2 - здоровий				
C3	x	157	4.6	1.28	0.5	312	12	4	1	55	2	0.3	36	4	пацієнт C3 - можливе зневоднення організму				
C4	x	130	4.6	0.85	0.6	244	5	4	1	55	2	0.3	25	4	пацієнт C4 - здоровий				
C5	x	145	4.1	0.92	0.1	301	8	6	3	65	4	0.6	30	6	пацієнт C5 - можливі проблеми з нирками				
C6	x	131	4.1	0.94	1.0	247	9	7	3	70	5	0.9	34	6	пацієнт C6 - здоровий				
C7	x	133	3.7	1.05	1.1	315	13	8	2	68	-1	-0.3	21	8	пацієнт C7 - можливе запалення в організмі, можлива алергія				
C8	x	119	3.0	0.91	1.0	217	6	5	3	57	2	0.2	29	5	пацієнт C8 - можлива анемія				
C9	x	142	4.7	0.97	0.1	287	8	6	3	63	2	0.2	39	5	пацієнт C9 - можливі проблеми з нирками				
C10	x	138	4.7	0.79	1.1	320	5	3	7	76	5	0.5	32	8	пацієнт C10 - здоровий				
D1	x	150	4.1	0.87	0.9	107	8	6	7	73	1	1	19	10	пацієнт D1 - можлива інфекція в організмі				
D2	x	139	4.1	1.00	1.1	245	6	6	3	63	4	0.5	34	5	пацієнт D2 - здоровий				

Рис.1. Приклад роботи програми

Вхідні дані представлені у вигляді стовпців даних, де перший стовпець відповідає імені пацієнту, другий – його статі, стовпці з третього по одинадцятий відповідають показникам загального аналізу крові. Значення, що виходяться за межі заданого діапазону вважаються відхиленими від норми та можуть означати наявність хвороби (або декількох хвороб) у пацієнта (табл. 1).

Таблиця 1 .

**Нормальні значення показників загального аналізу крові**

Показник загального аналізу крові	Нормальне значення	
	чоловіки	жінки
гемоглобін	130 - 160	120 - 140
еритроцити	4 - 5,1	3,7 - 4,7
колірний показник	0,85 – 1,15	
ретикулоцити	0,2 – 1,2	
тромбоцити	180 - 320	
ШОЕ	1 - 10	2 - 15
лейкоцити	4 – 9	
паличкоядерні	1 – 6	
сегментоядерні	47 – 72	
еозинофіли	0 – 5	
базофіли	0 – 1	
лімфоцити	18 – 40	
моноцити	2 - 9	

**ПРОВЕДЕННЯ ЕКСПЕРИМЕНТІВ НАД СИСТЕМОЮ**

Перевіримо точність діагностування за допомогою стандартної нейронної мережі Кохонена [9] та запропонованою модифікованою мережею. Результати, що є усередненим значенням з 25 проведених експериментів, представлені у таблиці 2.

Таблиця 2.

**Досягнена достовірність діагностування звичайною мережею Кохонена та запропонованою модифікованою мережею**

Обсяг навчальної вибірки	Обсяг тестової вибірки	Класична нейронна мережа Кохонена			Модифікована нейронна мережа Кохонена		
		Середня кількість правильно розпізнаних вхідних векторів	Середня кількість неправильно розпізнаних вхідних векторів	достовірність діагностики	Середня кількість правильно розпізнаних вхідних векторів	Середня кількість неправильно розпізнаних вхідних векторів	достовірність діагностики
50	10	7,96	2,04	79,6%	8,9	1,1	89%
500	100	76,2	24,8	76,2%	88,16	11,84	88,2%
1000	500	389,4	110,6	77,8%	448,2	51,8	89,6%
1500	1000	720,81	229,19	72%	851,94	148,06	85,2%

У результаті порівняння роботи стандартної та модифікованої мереж було встановлено, що запропонований метод має більш високі показники точності роботи у порівнянні із класичним.

Визначимо точність діагностування системою, використовуючи для тестування навчальну вибірку. Результати, що є середнім значенням з 25 проведених експериментів, представлені у таблиці 3.

Перевіримо точність діагностування системи у порівнянні с програмами-аналогами. Для порівняння оберемо інтерактивну систему діагностування Diagnos.ru [10], що має найбільшу в світі базу захворювань (близько 240) та заснована на методі нечіткої логіки, найпопулярнішу систему онлайн діагностування YourDiagnosis [11], яка базується на медичних алгоритмах, та IGematolog, яка є системою визначення захворювань крові. Для діагностування візьмемо 100 наборів вхідних значень та визначимо чи правильно вони були розпізнані. Для порівняння ефективності роботи запропонованої системи у порівнянні з системою IGematolog [12] візьмемо 100 наборів вхідних значень здорових пацієнтів та тих, що мають захворювання крові, оскільки дана система має обмежену базу даних хвороб, що включають в себе лише захворювання крові. Результати представлені у таблицях 4 та 5.

Таблиця 3 .

**Досягнена достовірність діагностування при використанні навчальної вибірки у якості тестової**

Загальний обсяг навчальної вибірки	Обсяг навчальної вибірки, що використана у тестуванні	Кількість правильно розпізнаних вхідних векторів	Кількість неправильно розпізнаних вхідних векторів	Достовірність діагностики
50	10	9,6	0,4	96%
500	100	91,9	8,1	91,9%
1000	500	468,14	31,86	93,6%
1500	1000	915,1	84,9	91,5%

Таблиця 4.

**Досягнена достовірність діагностування запропонованою системою та системами *Diagnos.ru* та *YourDiagnosis***

Система діагностування	Запропонована система діагностування за аналізом крові	<i>Diagnos.ru</i>	<i>YourDiagnosis</i>
Кількість правильно розпізнаних вхідних векторів (з 100)	85	79	71
Кількість неправильно розпізнаних вхідних векторів (з 100)	15	21	29
Достовірність діагностики	85%	79%	71%

Таблиця 5.

**Досягнена достовірність діагностування запропонованою системою та системою *IGematolog***

Система діагностування	Запропонована система діагностування за аналізом крові	<i>IGematolog</i>
Кількість правильно розпізнаних вхідних векторів (з 100)	93	84
Кількість неправильно розпізнаних вхідних векторів (з 100)	7	16
Достовірність діагностики	93%	84%

Отже, результати експериментів показують, що запропонована програма має вищу точність діагностування у порівнянні з програмами-аналогами.

У якості експериментальних даних були використані аналізи пацієнтів та їх діагнози, представлені у загальнодоступних базах даних [13,14].

**ВИСНОВКИ**

Таким чином, у роботі було проведено дослідження правильності роботи та оцінення достовірності роботи програмної системи медичного діагностування пацієнтів за аналізом крові на основі модифікованої мережі Кохонена. Експериментальні дослідження довели, що підхід з використанням нейронної мережі Кохонена, в якій центри кластерів визначаються заздалегідь згідно з нормальними та відхиленими від норми значеннями показників загального аналізу крові, дає більшу точність діагностування у порівнянні з класичною нейронною мережею Кохонена. Згідно з результатами експериментів, проведених над навчальною вибіркою, можна зробити висновок про правильність роботи системи, оскільки достовірність діагностування навчальної вибірки вище за достовірність діагностування тестової. Було проведено порівняння роботи системи з програмами-аналогами, а саме – *Diagnos.ru*, *YourDiagnosis* та *IGematolog*. У ході порівняння було визначено, що запропонована система має більшу достовірність діагностування.

**СПИСОК ЛІТЕРАТУРИ**

1. Алгоритми діагностування пацієнтів [Електронний ресурс]. Режим доступу: <http://www.vidal.by/vracham/Informatsiya-dlya-spetsialistov/Nevrologiya-psihiatriya/Algoritmy-diaagnostiki-i-vedeniya-patsientov/> - Назва з екрану
2. Загальний аналіз крові [Електронний ресурс]. Режим доступу: <http://ukrhealth.net/zagalnij-analiz-krovi/> - Назва з екрану
3. Обзор методов кластеризации [Електронний ресурс]. Режим доступу: [http://www.dialog21.ru/Archive/2001/volume2/2\\_26.htm](http://www.dialog21.ru/Archive/2001/volume2/2_26.htm) - Назва з екрану
4. Симахин В.А. Кластеризация медицинских данных с помощью нейросетей / В.А. Симахин, А.В. Кашицев [Електронний ресурс]. Режим доступу: <http://neurocomp.ru/klasterizaciya-medicinskih-dannux-s-pomoshhu-nejrosetej/> - Назва з екрану
5. О. С. Амосов. Интеллектуальные информационные системы. Нейронные сети и нечеткие системы: Учеб. Пособие./ О.С. Амосов, Комсомольск-на-Амуре: ГОУВПО «КнАГТУ», 2004. -104 с.
6. Руденко О.Г. Искусственные нейронные сети / О.Г. Руденко, Е.В. Бодянский, Харьков, 2005. – 407с.
7. Дяченко В.А. Адаптивное параллельное обучение модифицированной саморганизующейся карты Кохонена / В.А. Дяченко, О.Ф. Михаль [Електронний ресурс]. Режим доступу: [http://archive.nbuv.gov.ua/portal/natural/bionint/2012\\_1/Mihal1.pdf](http://archive.nbuv.gov.ua/portal/natural/bionint/2012_1/Mihal1.pdf)
8. Парфенович Д.О. Нейронні мережі – від теорії до практики / Д.О. Парфенович [Електронний ресурс]. Режим доступу - <http://www.mql5.com/ru/articles/497>
9. WEBSOM. Проект Тайво Кохонена [Електронний ресурс]. Режим доступу: <http://websom.hut.fi/websom/>
10. Диагноз.ру. Инновационный медицинский центр [Електронний ресурс]. Режим доступу: <http://www.diagnos.ru/> - Назва з екрану.
11. Your Diagnosis. Your Free Online Medical Diagnosis & Symptoms Analysis Tool [Електронний ресурс]. Режим доступу: <http://www.yourdiagnosis.com/> - Назва з екрану.
12. iГематолог – Расшифровка общего анализа крови [Електронний ресурс]. Режим доступу: <http://www.pacientdoma.ru/>
13. Laboratory Procedure Material. Complete Blood Count using HMX – NHANES 2007-200810. Mayo clinic. Diseases and conditions [Електронний ресурс]. Режим доступу: <http://www.mayoclinic.org/diseases-conditions/high-blood-cholesterol/basics/tests-diagnosis/con-20020865> - Назва з екрану.
14. CLL Patient Databases [Електронний ресурс]. Режим доступу: <https://patientdatabases.org/wp/about> - Назва з екрану

Надійшла до редакції 25.05.2014р.

**КОЛЕСНИЦЬКИЙ ОЛЕГ КОСТЯНТИНОВИЧ** – к.т.н., доцент, доцент кафедри комп'ютерних наук, Вінницький національний технічний університет, м. Вінниця, Україна.

**ЖУРАВСЬКА ЮЛІЯ ОЛЕКСАНДРІВНА** – студентка групи 1КН-10 Вінницький національний технічний університет, м. Вінниця, Україна.