
Магістерська кваліфікаційна робота на тему:

ІНФОРМАЦІЙНА ТЕХНОЛОГІЯ МЕДИЧНОГО ДІАГНОСТУВАННЯ ЗА АНАЛІЗОМ КРОВІ НА ОСНОВІ НЕЙРОННОЇ МЕРЕЖІ КОХОНЕНА

Виконала: ст.гр. 1КН-14мн Журавська Ю.О.

Науковий керівник: к.т.н., доц.каф.КН Колесницький О.К.

Рецензент: к.т.н., доц.каф ОТ Крупельницький Л.В.

АКТУАЛЬНІСТЬ ДОСЛІДЖЕННЯ

Недоліки сучасних систем діагностування:

- недостатні швидкість та достовірність визначення діагнозу;
- невеликий спектр діагностованих хвороб;
- обмежений функціонал;
- вимогливість до програмного та апаратного забезпечення;
- неможливість одночасного діагностування декількох пацієнтів.

МЕТА ТА ЗАВДАННЯ ДОСЛІДЖЕННЯ

Мета – підвищення достовірності та швидкості медичного діагностування з використанням інформаційної технології у порівнянні з програмами-аналогами за рахунок застосування модифікованої нейронної мережі Кохонена.

Завдання:

- провести аналіз проблеми розв'язання задачі медичного діагностування за аналізом крові;
- розглянути існуючі методи вирішення задачі медичного діагностування та обрати й обґрунтувати вибір методу, який задовольняє мету даної магістерської кваліфікаційної роботи;
- розробити математичну модель інформаційної системи медичного діагностування та удосконалити її згідно з метою роботи;
- сформулювати стадії інформаційної технології та їх основі розробити структуру та алгоритм роботи програмного засобу;
- виконати програмну реалізацію запропонованої інформаційної технології діагностування;
- провести тестування програмного продукту та виконати аналіз отриманих результатів.

ОБ'ЄКТ, ПРЕДМЕТ ТА МЕТОДИ ДОСЛІДЖЕННЯ

Об'єкт дослідження – це процес медичного діагностування пацієнтів з використанням інформаційних технологій.

Предмет дослідження – це методи та програмні засоби визначення діагнозу пацієнта за аналізом крові.

Методи дослідження:

- системного аналізу для аналізу структури інформаційної системи;
- теорія нейронних мереж для реалізації інформаційної технології до медичного діагностування;
- методи математичної статистики для розробки процесу кластеризації та обрахунків результатів експериментів над програмним засобом;
- об'єктно-орієнтованого програмування для автоматизації розрахунків.

НАУКОВА НОВИЗНА ОДЕРЖАНИХ РЕЗУЛЬТАТІВ

- вперше запропоновано інформаційну технологію медичного діагностування, яка використовує показники загального аналізу крові та нейронну мережу Кохонена для визначення відхилення від показників здорової людини, що дозволяє діагностувати у пацієнта більш великий спектр хвороб, ніж відомі інформаційні технології діагностування;
- **удосконалено модель навчання** нейронної мережі Кохонена за рахунок введення процедури підбору для центрів кластерів наперед визначених значень, що являють собою нормальні показники аналізу крові та відхилені від норми показники, які відповідають визначеним хворобам, що дозволило суттєво скоротити швидкість навчання нейронної мережі;
- **удосконалено метод функціонування** нейронної мережі Кохонена за рахунок використання як функції сусідства функцію «Мексиканський капелюх» та використання зваженої Евклідової метрики як міри довжини, якою розраховується відстань між вхідними показниками аналізу крові та центрами кластерів, що дозволяє досягнути вищої достовірності навчання мережі та, відповідно, вищої достовірності діагностування;
- **знайшла подальшого розвитку інформаційна технологія** медичного діагностування за аналізом крові на основі нейронної мережі Кохонена з використанням у якості моделі конкурентного навчання моделі м'якої конкуренції та введенням процесу нормалізації векторів вхідних даних.

ПРАКТИЧНЕ ЗНАЧЕННЯ ОДЕРЖАНИХ РЕЗУЛЬТАТІВ

- розроблено новий спосіб медичного діагностування, який використовує значення показників загального аналізу крові.
- розроблено алгоритм навчання нейронної мережі Кохонена для підвищення швидкості даного процесу.
- розроблено алгоритм функціонування нейронної мережі Кохонена для підвищення достовірності та швидкодії кластеризації числових об'єктів.
- розроблено програмний засіб для медичного діагностування на основі нейронної мережі Кохонена.

Розроблені алгоритми можуть бути впроваджені в початковий процес у якості лекції на тему «Нейромережевий метод медичного діагностування з аналізом крові з використанням нейронної мережі Кохонена» дисципліни «Нейромережеві методи обчислювального інтелекту», дослідження методу діагностування за аналізом крові та роботи нейронної мережі Кохонена.

ПОСТАНОВКА ПРОБЛЕМИ

У даній ситуації доцільно застосовувати метод медичного діагностування, заснований на використанні показників загального аналізу крові.

Переваги: всебічний огляд пацієнта, висока достовірність результатів.

Вхідні дані: стать пацієнта, 13 показників аналізу крові (гемоглобін, еритроцити, колірний показник, ретикулоцити, тромбоцити, ШОЕ, лейкоцити, паличкоядерні, сегментоядерні, еозинофіли, базофіли, лімфоцити, моноцити).

Вихідні дані: діагноз, рекомендований лікар (виведення на екран та/або в файл).

ВИБІР ТЕОРЕТИЧНОЇ ОСНОВИ

1. Системи логічного висновку.
2. Системи, які використовують нечітку логіку .
3. Експертні системи.
4. Імітаційне моделювання.
5. Генетичні алгоритми .
6. **Штучні нейронні мережі:**

Переваги: висока достовірність результатів та швидкість обробки інформації, не потребують надвеликих обсягів даних, можливість оновлення інформації у системі за рахунок процедури навчання.

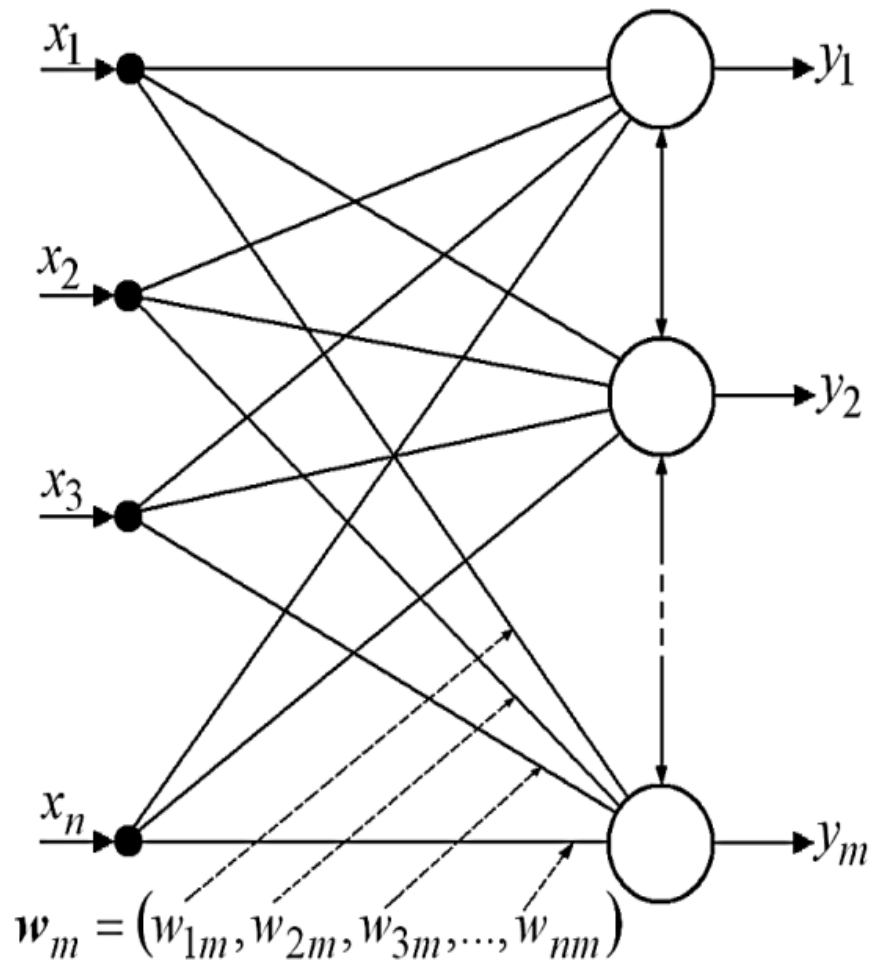
Метод обробки інформації – кластеризація даних, де вхідні дані – вектори показників аналізу крові, кластери – можливі діагнози.

ВИБІР НЕЙРОННОЇ МЕРЕЖІ

Результати дослідження достовірності та швидкості кластеризації нейронними мережами на прикладі кластеризації ірису

Нейронна мережа	Кількість правильно кластеризованих елементів (з 150)	Кількість неправильно кластеризованих елементів (з 150)	Достовірність кластеризації	Швидкість кластеризації (с)	Швидкість навчання (с)
LVQ	135	15	90%	15,8	41,3
Нейронна мережа Кохонена	144	6	96%	7	40,2
SOM	144	6	96%	10,2	71
Нейронний газ	135	15	90%	35,4	194,1
ART-1	123	27	82%	78,1	180
ART-2	123	27	82%	61,4	191,1

СТРУКТУРА НЕЙРОННОЇ МЕРЕЖІ КОХОНЕНА ДЛЯ ЗАДАЧІ МЕДИЧНОГО ДІАГНОСТУВАННЯ



- $x(i) = (x_1(i), x_2(i), \dots, x_n(i))$, де $i = 1, 2, \dots, P$ – вектор вхідних сигналів; $P = 13$ за кількістю показників аналізу крові.

- $y = (y_1, y_2, \dots, y_m)$ – вихідний вектор, де m – кількість кластерів; $m = 43$ за кількістю діагнозів.

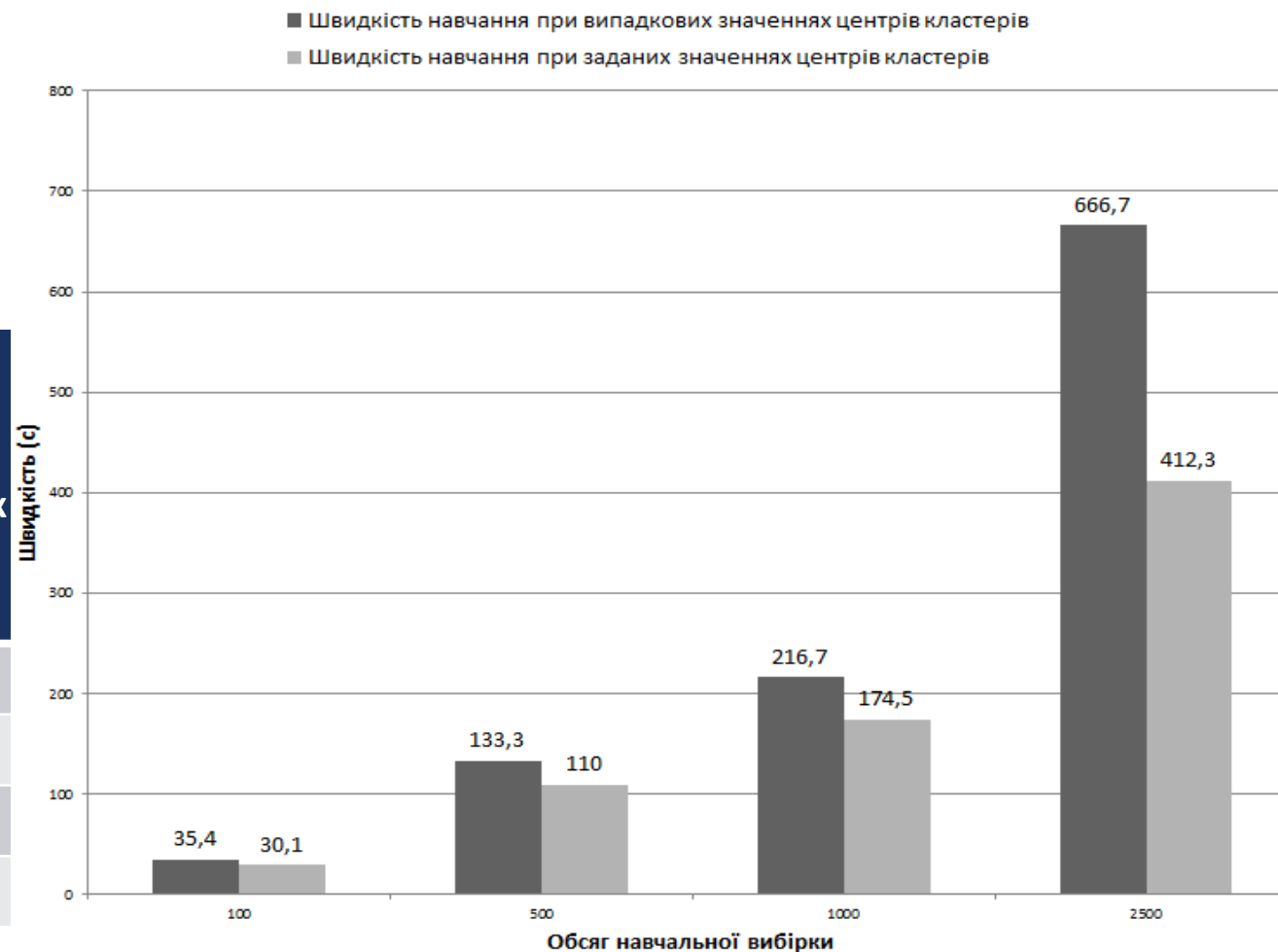
- $W = \begin{pmatrix} W^{11} & W^{12} & \dots & W^{1m} \\ W^{21} & W^{22} & \dots & W^{2m} \\ \dots & \dots & \dots & \dots \\ W^{k1} & W^{k2} & \dots & W^{km} \end{pmatrix}$ - матриця множини вагових коефіцієнтів;

- $W^{ij} = (w_1^{ij}, w_2^{ij}, \dots, w_n^{ij})$ - елементи матриці (вектори вагових коефіцієнтів).

ВИБІР МЕТОДУ ВИЗНАЧЕННЯ ЦЕНТРІВ КЛАСТЕРІВ

Задані значення центрів кластерів відповідають показникам аналізу крові для можливих діагнозів.

Обсяг навчальної вибірки	Швидкість навчання при випадкових значеннях центрів кластерів (с)	Швидкість навчання при заданих значеннях центрів кластерів (с)
100	35,4	30,1
500	133,3	110
1000	216,7	174,5
2500	666,7	412,3



ВИБІР МОДЕЛІ КОНКУРЕНТНОГО НАВЧАННЯ ТА НОРМАЛІЗАЦІЯ ВХІДНИХ ДАНИХ

1. Модель жорсткої конкуренції WTA («Winner Takes All») – виявлення одного нейрону-переможця.
2. Модель справедливої конкуренції CWTA («Conscience WTA») – введення механізму «стомлення» переможців для уникнення явища порожнього кластера.
- 3. Модель м'якої конкуренції WTM («Winner Takes Most») – навчання області навколо нейрона-переможця для пришвидшення збіжності та, відповідно, часу навчання.**

$$w_m := w_m + \alpha(x_i - w_m)K(p(x_i, w_m)), m = 1, \dots, M. \quad K(p) = \exp(-\beta p^2) \quad \beta > 0$$

Для оптимальних розрахунків нейронної мережі була введена процедура попередньої обробки вхідних даних, при якій значення ознак, що утворюють вхідний вектор, приводяться до заданого діапазону [0...1].

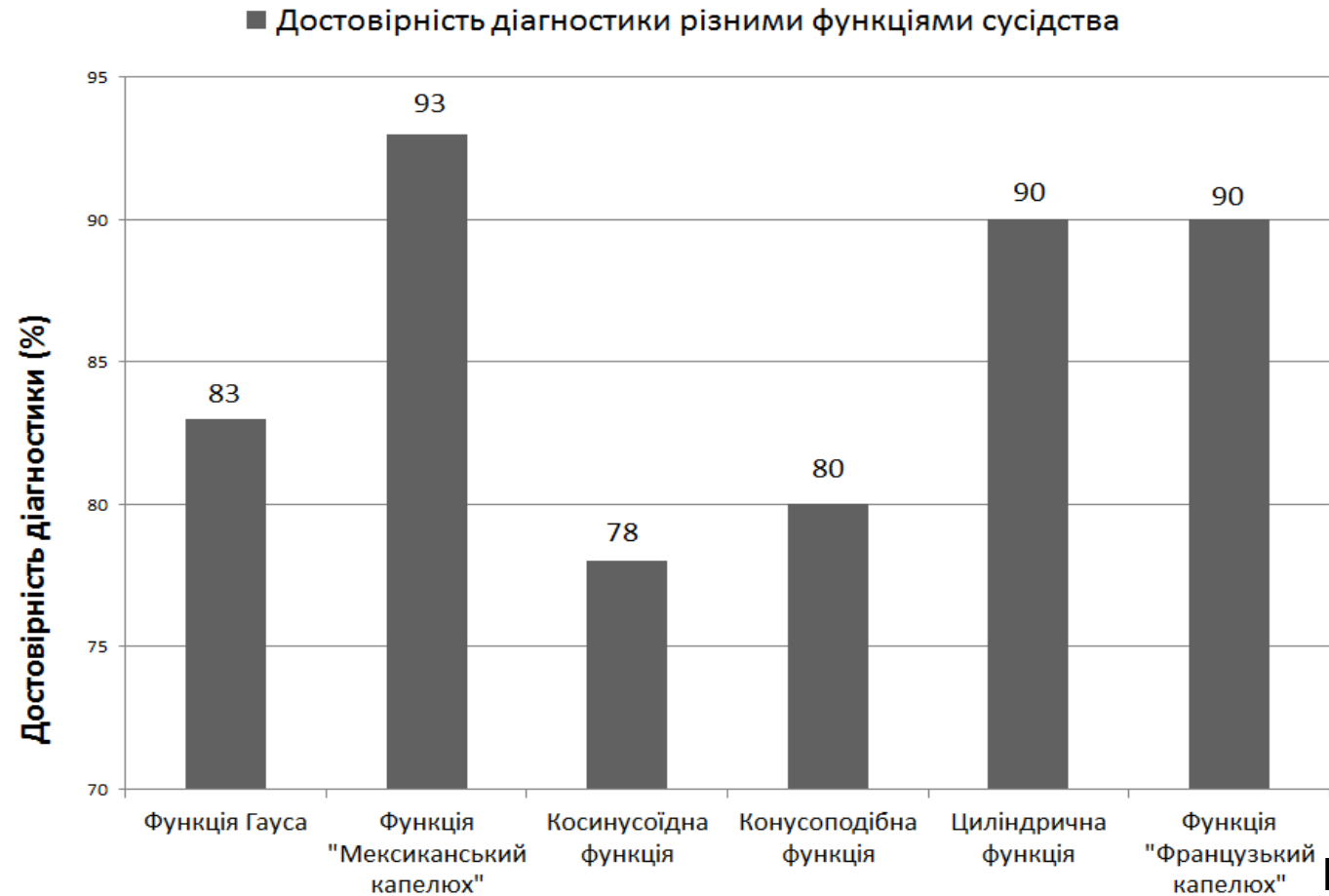
$$x_0' = \frac{(x_0 - x_{\min})}{x_{\max} - x_{\min}}$$

де x_0' - значення після проведеної нормалізації;
 x_0 – значення, що підлягає нормалізації;
 $[x_{\min}, x_{\max}]$ – інтервал вхідних значень x .

ОБҐРУНТУВАННЯ ВИБОРУ ФУНКЦІЇ СУСІДСТВА

Функції сусідства та досягнена достовірність діагностики

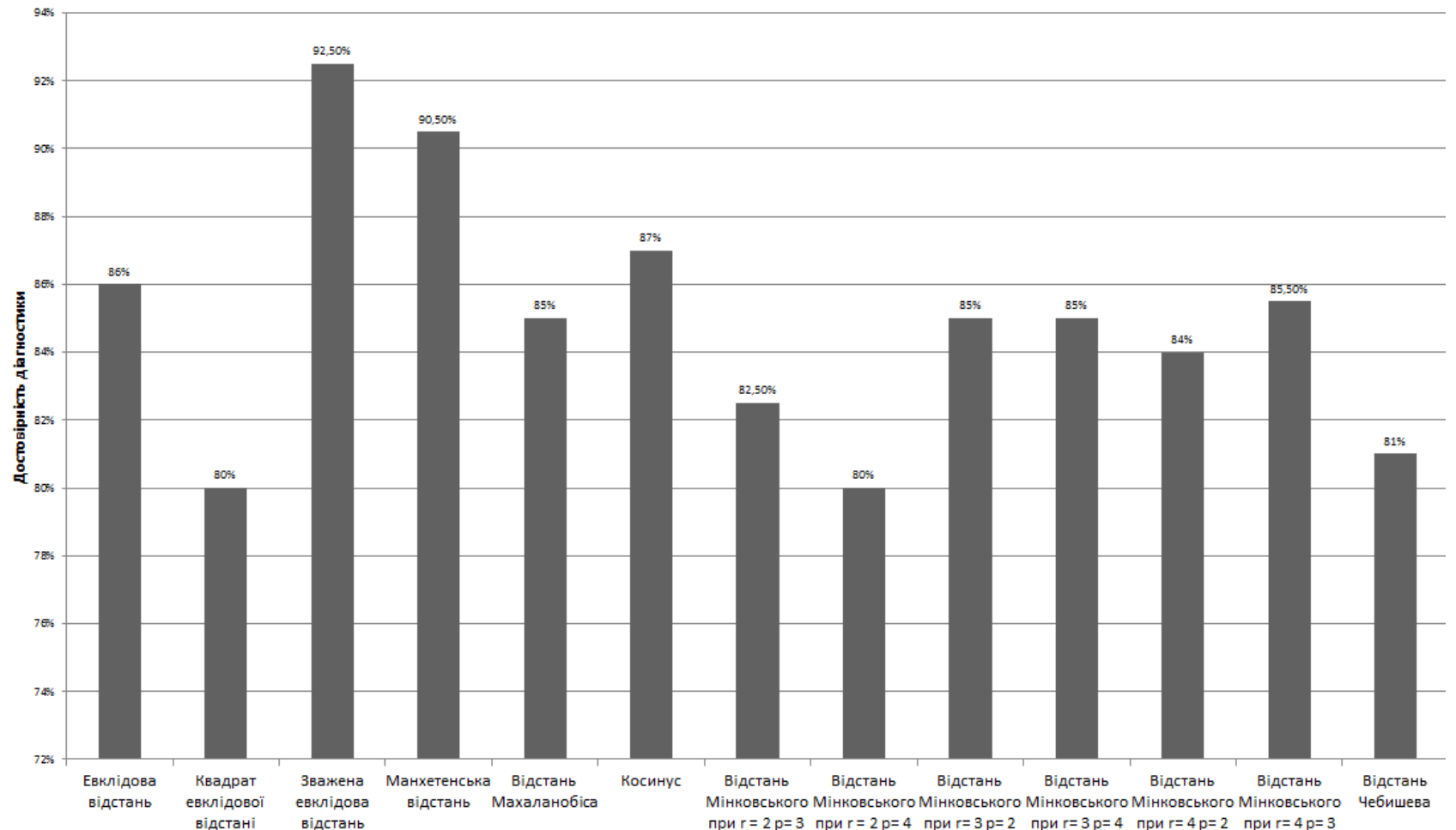
Функція сусідства	Кількість правильно розпізнаних векторів з 200	Кількість неправильно розпізнаних векторів з 200	Достовірність діагностики
Гауса	166	34	83%
«Мексиканський капелюх»	186	14	93%
Косинусоїдна	156	44	78%
Конусоподібна	160	40	80%
Циліндрична	180	20	90%
«Французький капелюх»	180	20	90%

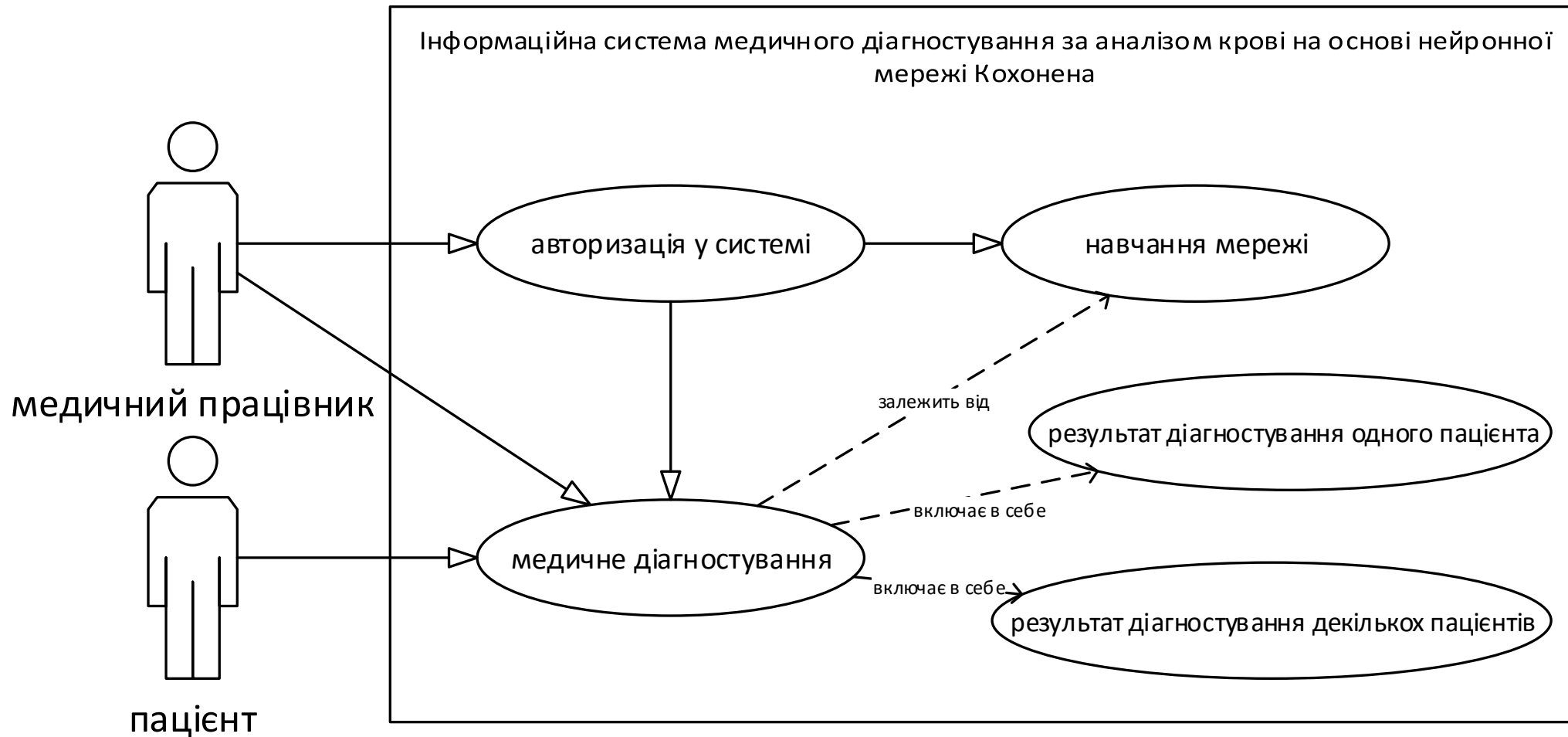


ОБҐРУНТУВАННЯ ВИБОРУ МЕТРИКИ

■ Міри довжини (метрики)

Результати експериментів по визначенню достовірності діагностування в залежності від використовуваної міри відстані

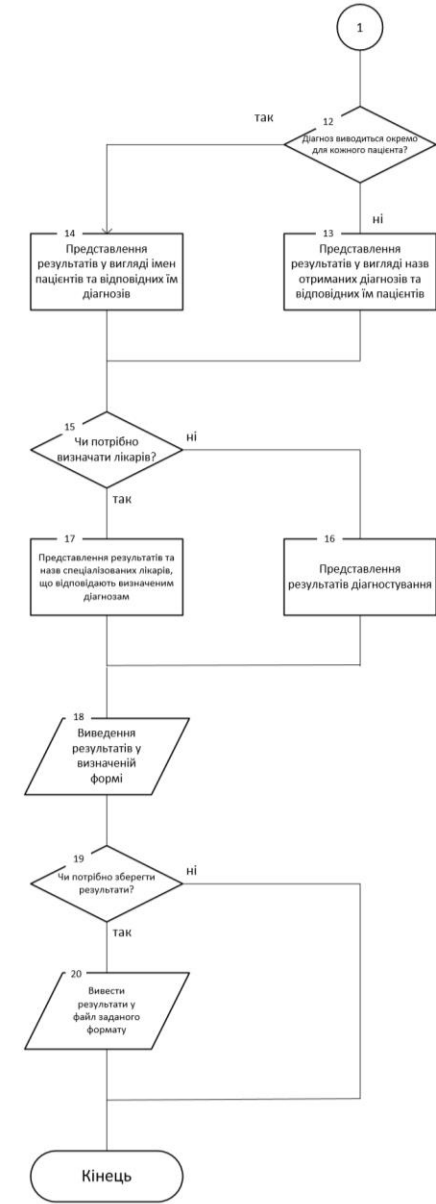
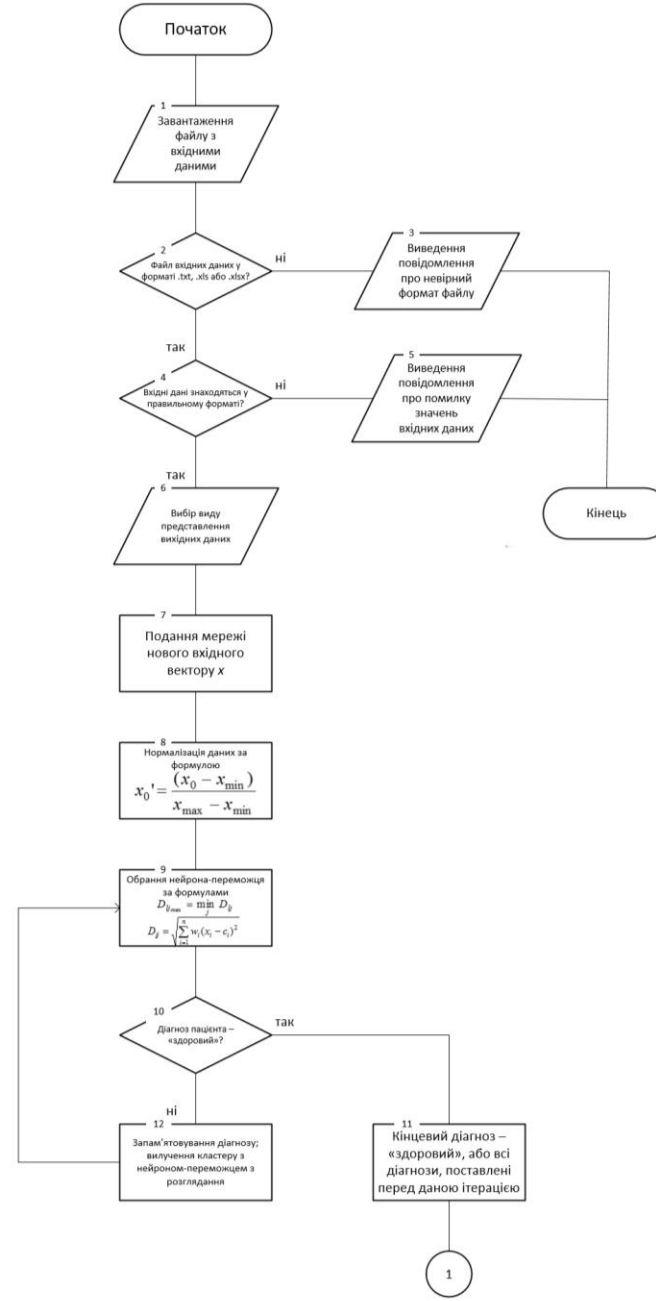
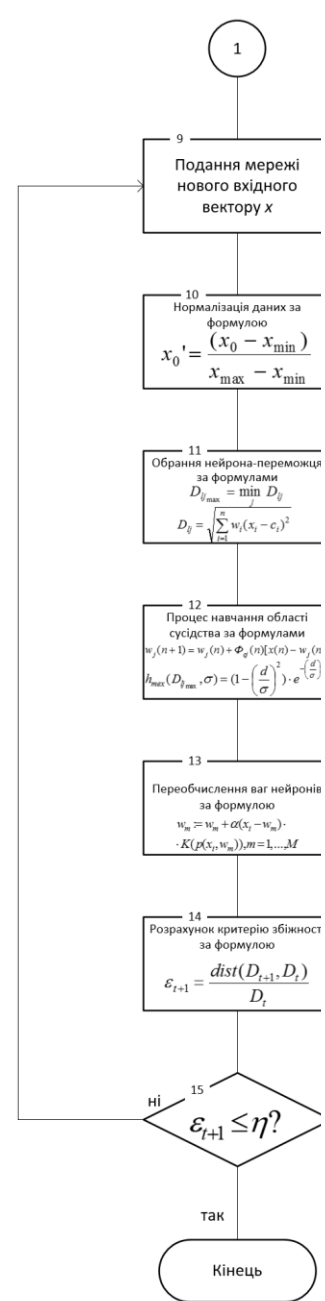
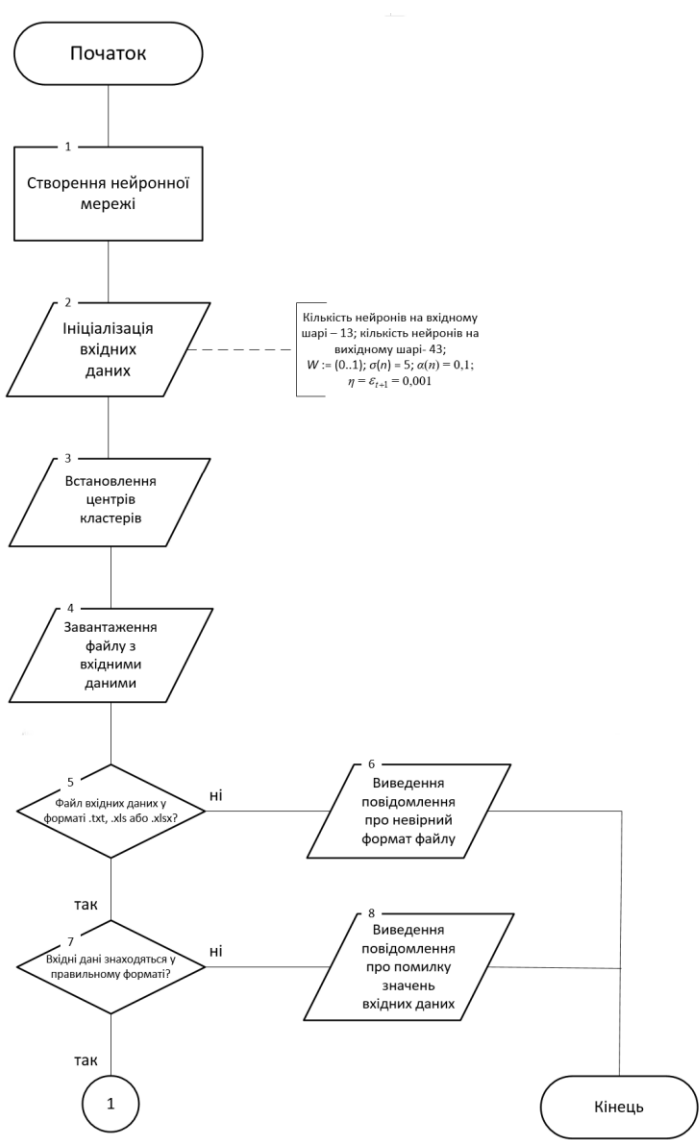




Діаграма варіантів використання

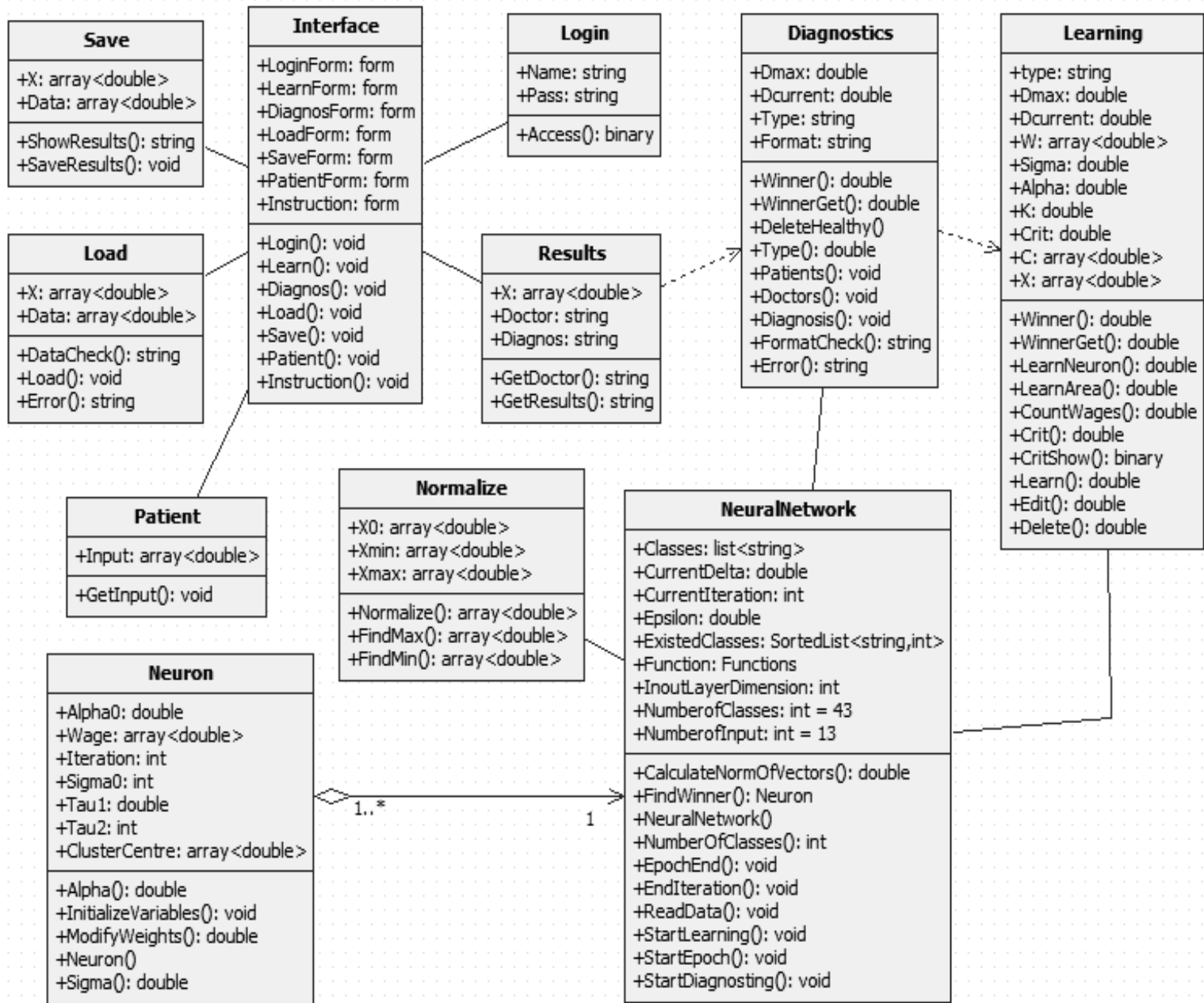


Структура інформаційної системи

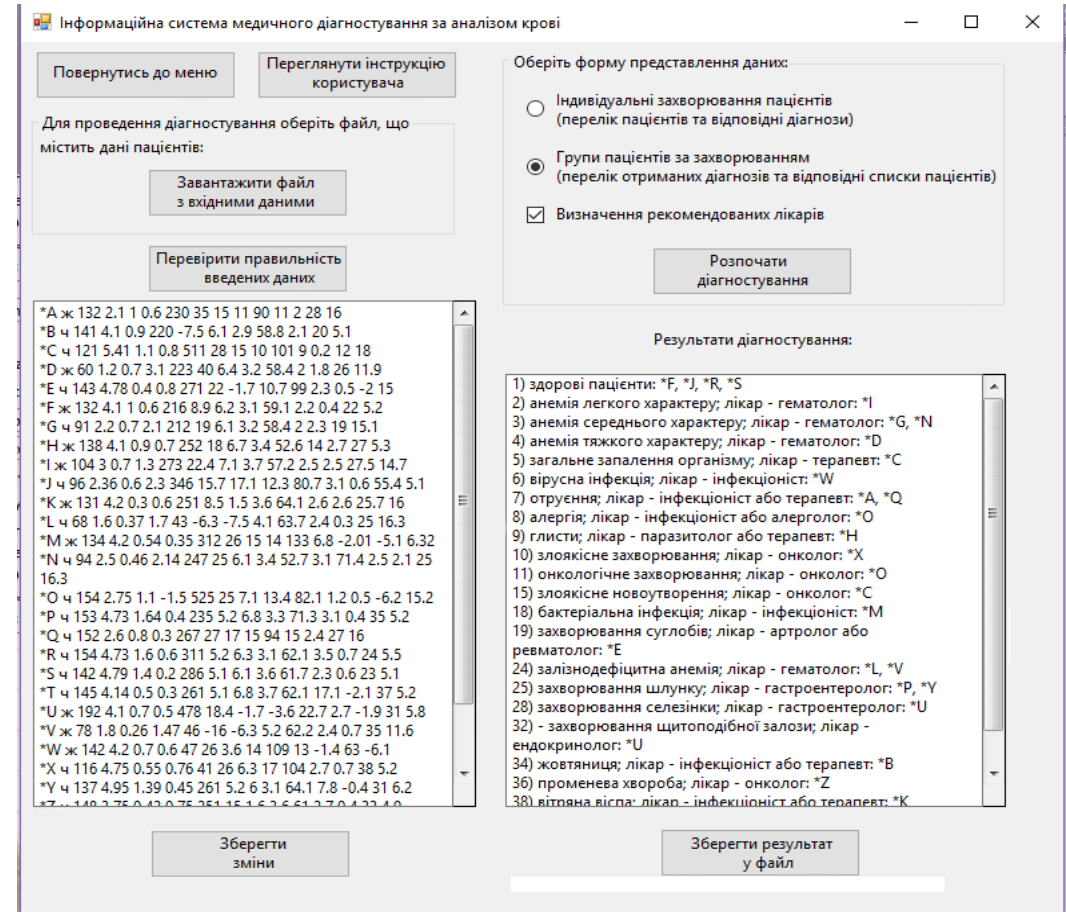
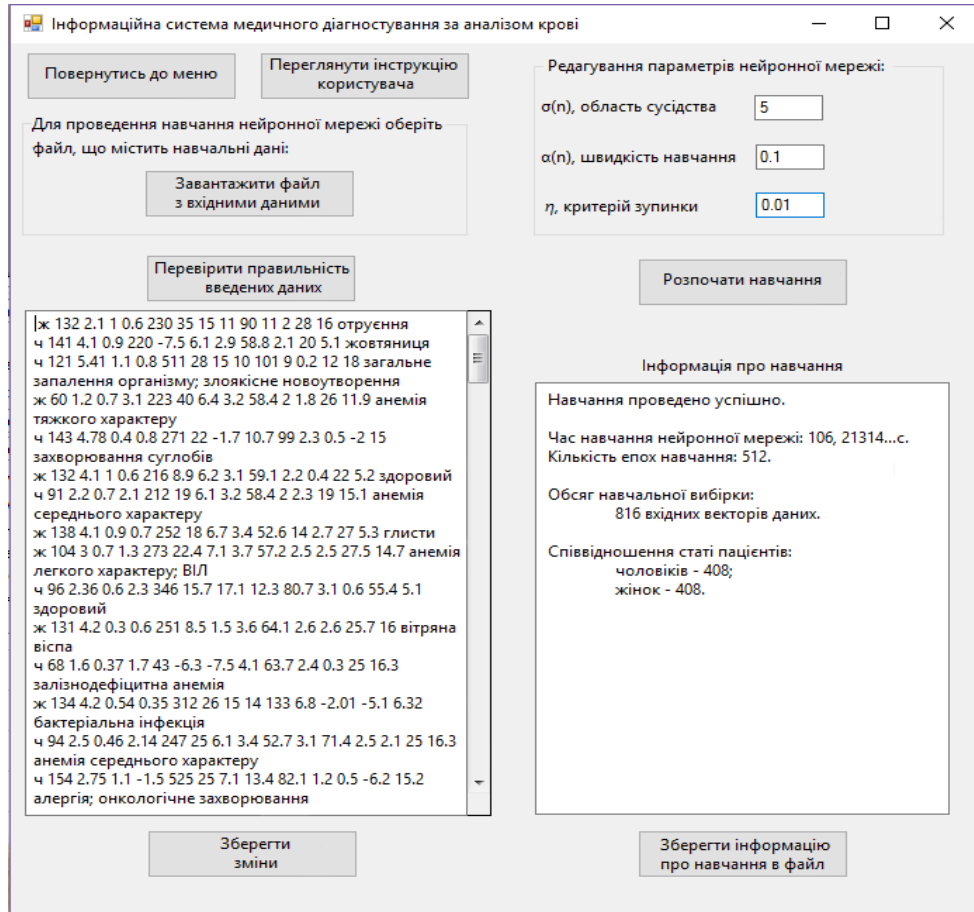


Алгоритм режиму навчання

Алгоритм режиму діагностування

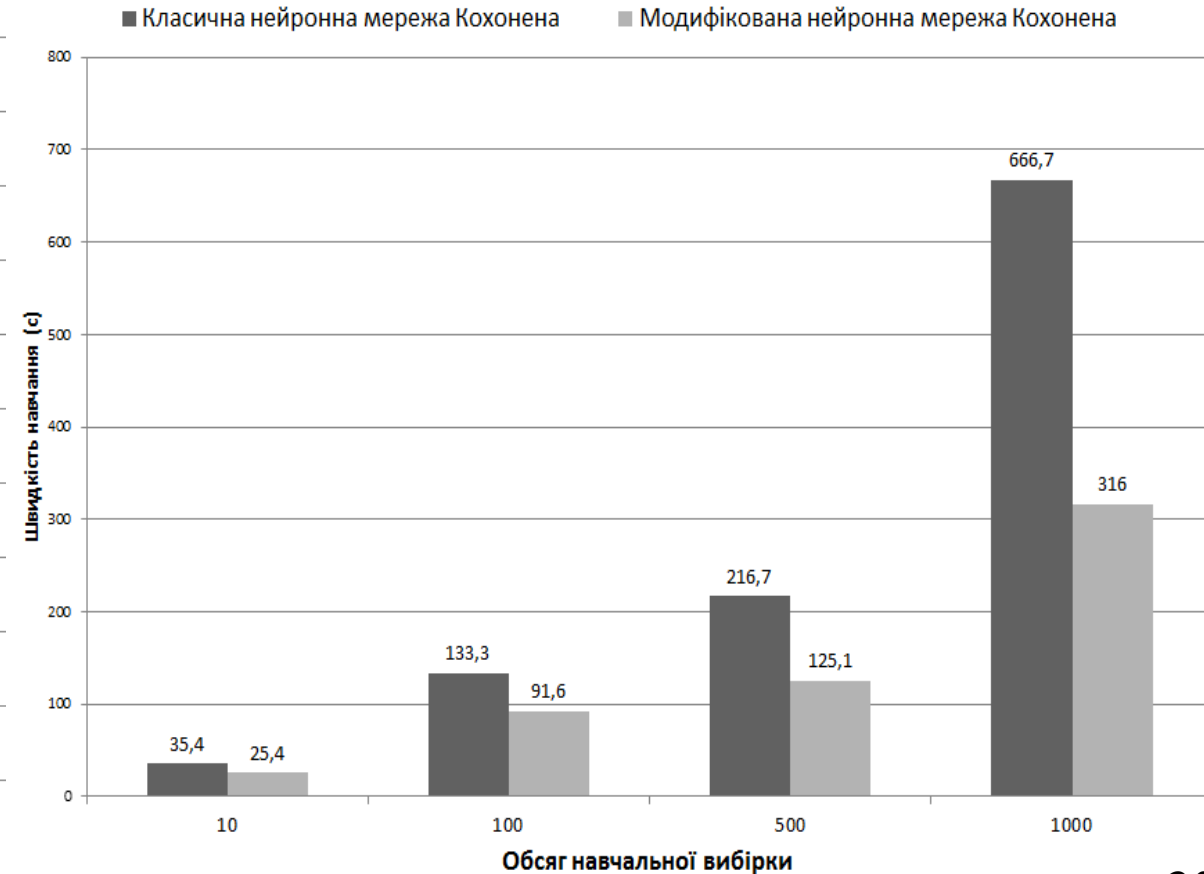
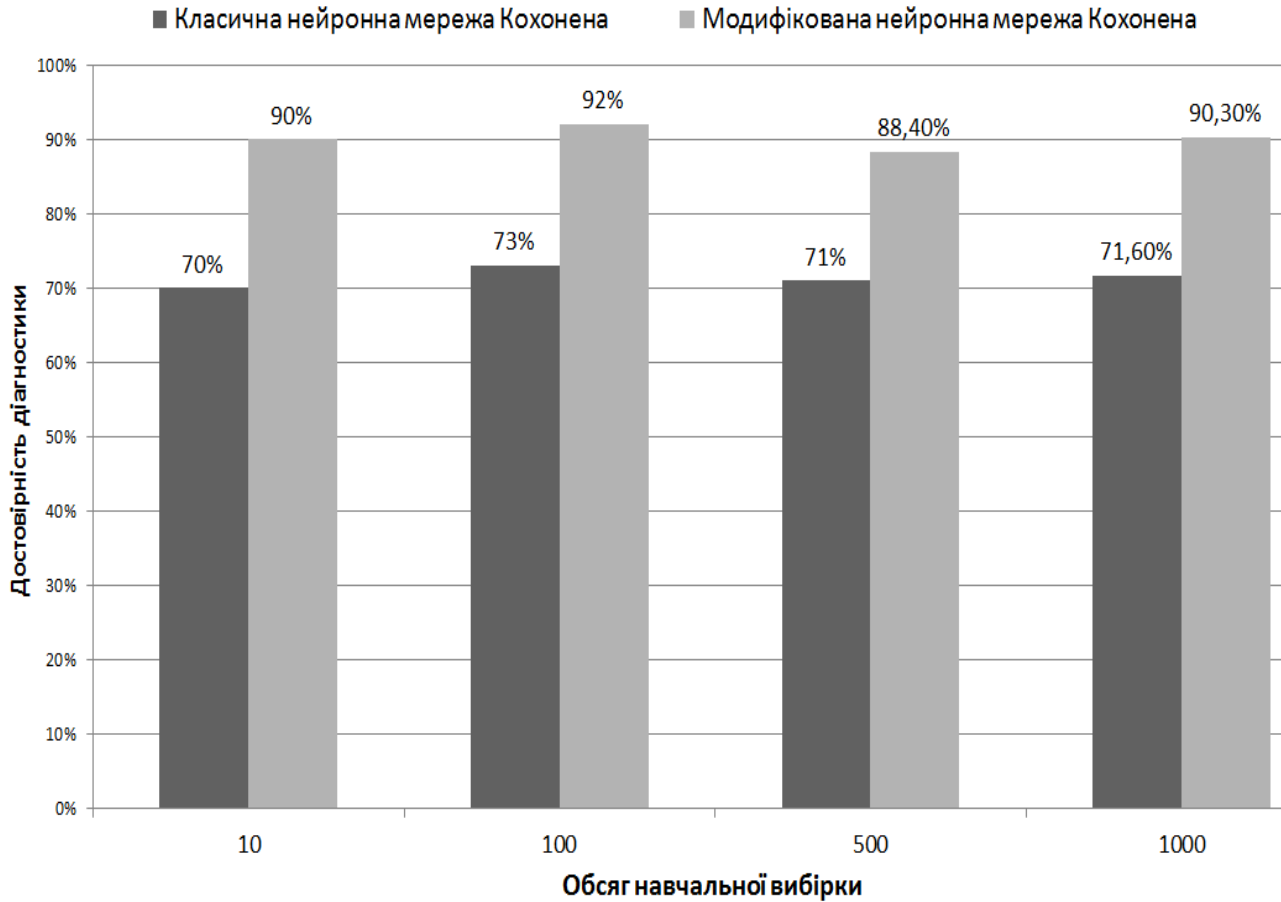


Діаграма класів



Приклади роботи програми у режимі навчання та діагностування

ЕКСПЕРИМЕНТАЛЬНІ ДОСЛІДЖЕННЯ: ПОРІВНЯННЯ КЛАСИЧНОЇ ТА МОДИФІКОВАНОЇ НЕЙРОННОЇ МЕРЕЖІ КОХОНЕНА

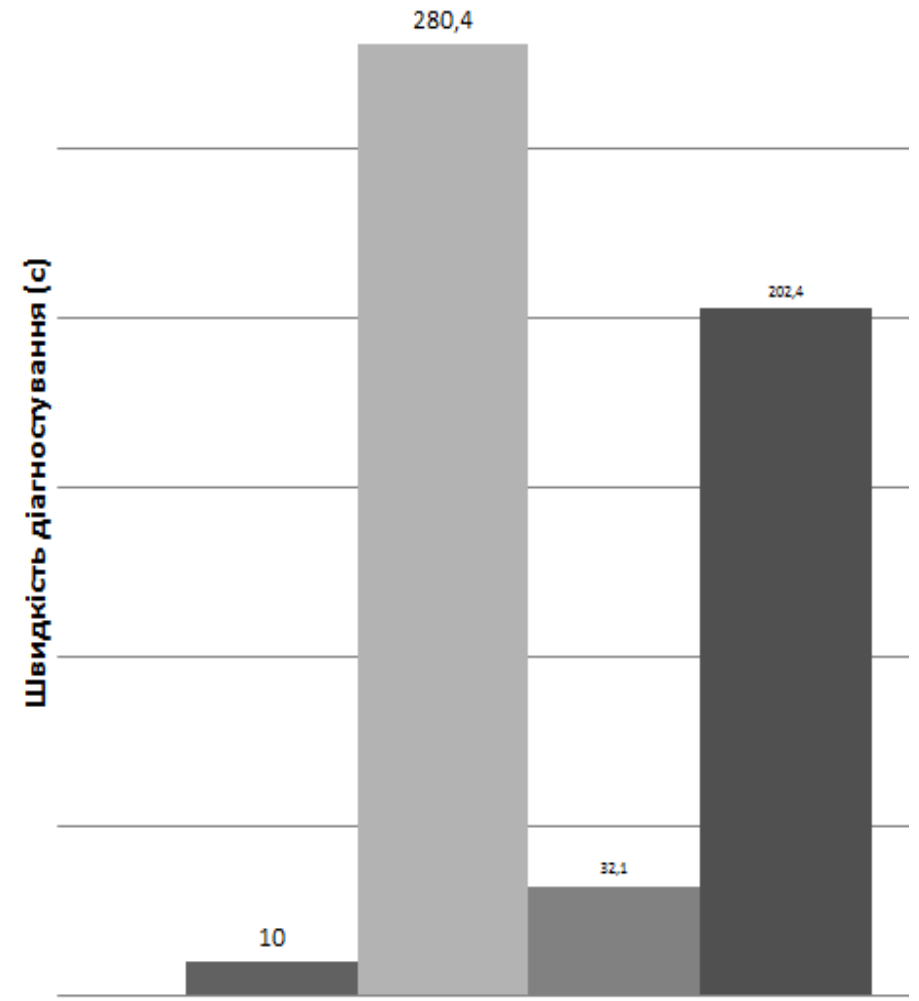
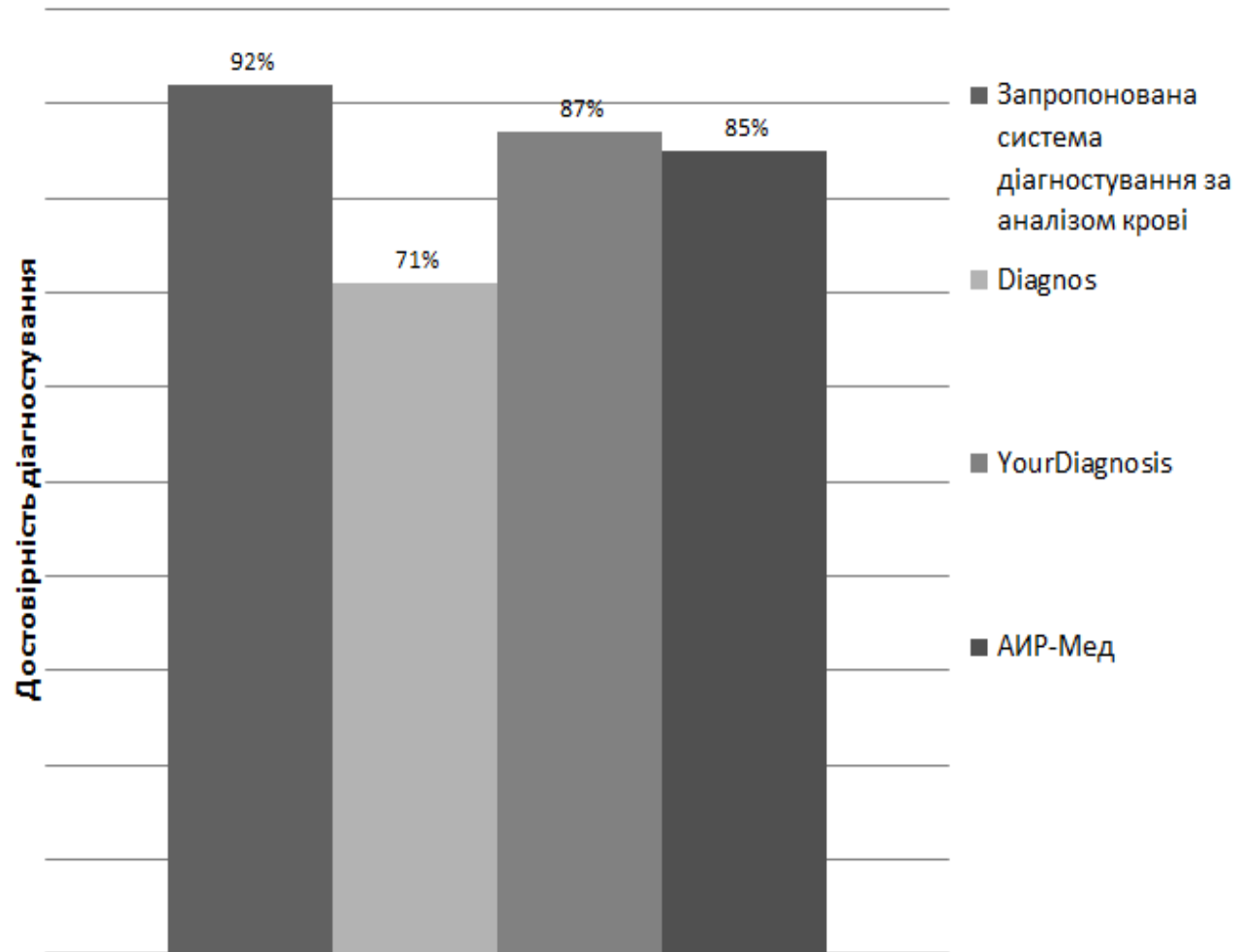


ЕКСПЕРИМЕНТАЛЬНІ ДОСЛІДЖЕННЯ: ДОСЛІДЖЕННЯ ДОСТОВІРНОСТІ ДІАГНОСТУВАННЯ В ЗАЛЕЖНОСТІ ВІД СКЛАДНОСТІ ВИБІРКИ

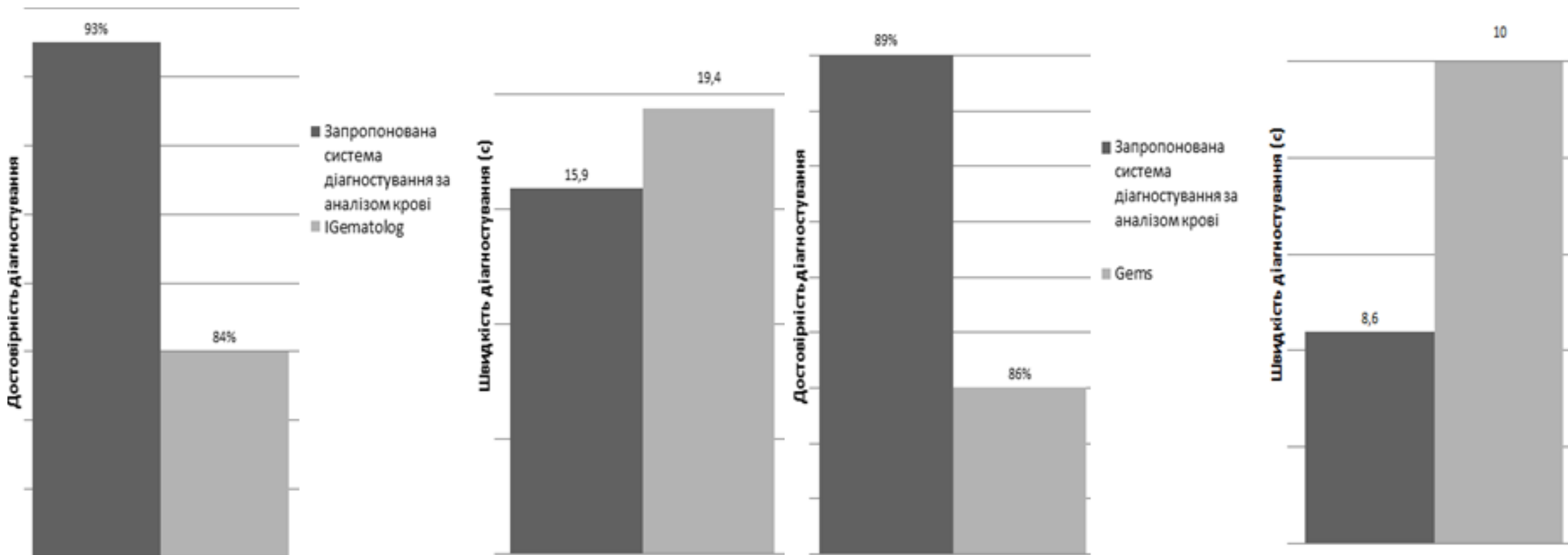
Результати експериментальних досліджень достовірності діагностування інтелектуальної системи в залежності від складності вибірки

Вибірка	Кількість вірно розпізнаних вхідних образів (з 200)	Кількість неправильно розпізнаних вхідних образів (з 200)	Достовірність діагностування
Найскладніша	142	58	71%
Середня	186	14	93%
Найпростіша	200	0	100%

ЕКСПЕРИМЕНТАЛЬНІ ДОСЛІДЖЕННЯ: ПОРІВНЯННЯ З ПРОГРАМАМИ-АНАЛОГАМИ «DIAGNOS», «YOUR DIAGNOSIS» ТА «АІР-МЕД»



ЕКСПЕРИМЕНТАЛЬНІ ДОСЛІДЖЕННЯ: ПОРІВНЯННЯ З ПРОГРАМАМИ-АНАЛОГАМИ «IGEMATOLOG» ТА «GEMS»



ЕКОНОМІЧНА ЧАСТИНА

- Вартість реалізації інформаційної технології складає **192035,7** грн.
- Загальне збільшення прибутку підприємства, тобто **комерційний ефект від впровадження розробки**, за п'ять років складе **1883000** грн.
- Розрахована **абсолютна ефективність вкладених інвестицій** є **позитивним числом**, що свідчить про отримання прибутку інвестором від впровадження програмного продукту у діяльність підприємства.
- **Щорічна ефективність вкладених в наукову розробку інвестицій** складає **23%**, що вище за мінімальну бар'єрну ставку дисконтування, яка складає **20%**. Це означає потенційну зацікавленість інвесторів у фінансуванні розробки.
- **Термін окупності** розробленої інформаційної технології складає **4,2** роки.

АПРОБАЦІЯ

Результати роботи були апробовані на :

- III Міжнародній науково-практичній конференції «Обчислювальний інтелект (результати, проблеми, перспективи)», (м. Черкаси, Україна, 2015 р.);
 - III Всеукраїнській науково-практичній конференції «Інтелектуальні технології в системному програмуванні», (м. Хмельницький, Україна, 2014 р.);
 - IV Всеукраїнській науково-практичній конференції «Інтелектуальні технології в системному програмуванні», (м. Хмельницький, Україна, 2015 р.);
 - XLIII науково-технічній конференції професорсько-викладацького складу, співробітників та студентів Вінницького національного технічного університету (м. Вінниця, Україна, 2014 р.);
- та опубліковані у збірниках даних конференцій.

ПУБЛІКАЦІЇ

За результатами магістерської кваліфікаційної роботи опубліковано:

- 3 статті в фахових виданнях:
 - «Інформаційні технології та комп'ютерна інженерія», №1, 2014 р.;
 - «Оптико-електронні інформаційно-енергетичні технології», №2, 2014 р.;
 - «Інформаційні технології та комп'ютерна інженерія», №3, 2014 р.

- 1 стаття у збірнику праць конференції;

- 3 тези доповіді конференцій.

ОСНОВНІ РЕЗУЛЬТАТИ РОБОТИ

1. Запропоновано інформаційну технологію медичного діагностування, яка використовує показники загального аналізу крові та нейронну мережу Кохонена для визначення відхилення від показників здорової людини, що дозволяє діагностувати у пацієнта більш великий спектр хвороб, ніж відомі інформаційні технології діагностування.
2. Розроблено алгоритми навчання та функціонування нейронної мережі Кохонена, які підвищують швидкість навчання мережі на 18,8% та підвищення достовірності кластеризації числових об'єктів на 38,5%.
3. Удосконалено модель навчання нейронної мережі Кохонена за рахунок введення процедури підбору для центрів кластерів наперед визначених значень, що дозволило скоротити швидкість навчання нейронної мережі на 22,5%.
4. Удосконалено метод функціонування нейронної мережі Кохонена за рахунок використання як функції сусідства функцію «Мексиканський капелюх» та використання зваженої Евклідової метрики як міри довжини, що дозволяє досягнути вищої достовірності навчання та діагностування мережі.
5. Знайшла подальшого розвитку інформаційна технологія медичного діагностування за аналізом крові на основі нейронної мережі Кохонена з використанням у якості моделі конкурентного навчання моделі м'якої конкуренції та введенням процесу нормалізації векторів вхідних даних.
6. Розроблено програмне забезпечення, що реалізує інформаційну технологію медичного діагностування за аналізом крові, яка дозволяє визначати попередній діагноз пацієнта та надавати рекомендації щодо відвідування спеціалізованого лікаря, що забезпечує вищу достовірність та швидкодію, аніж аналогічні системи медичного діагностування.

Поставлені задачі магістерської кваліфікаційної роботи були виконані в повному обсязі.