

Міністерство освіти і науки України  
Харківський національний університет радіоелектроніки

Факультет Інформаційно-аналітичних технологій та менеджменту  
(повна назва)

Кафедра Інформатики  
(повна назва)

## КВАЛІФІКАЦІЙНА РОБОТА Пояснювальна записка

рівень вищої освіти перший (бакалаврський)

### МОДЕЛЮВАННЯ МЕТОДУ НЕЧІТКИХ С-СЕРЕДНІХ З ВИКОРИСТАННЯМ ЕВОЛЮЦІЙНОЇ ОПТИМІЗАЦІЇ

(тема)

Виконав:

здобувач 4 року навчання,

групи ІТІНФ-21-1

Редекоп І. Є.

(прізвище, ініціали)

Спеціальність 122 Комп'ютерні науки

(код і повна назва спеціальності)

Тип програми освітньо-професійна

Освітня програма Інформатика

(повна назва освітньої програми)

Керівник доц. Шафроненко А.Ю.

(посада, прізвище, ініціали)

Допускається до захисту

Завідувач кафедри інформатики \_\_\_\_\_  
(підпис)

Кобилін О. А.  
(прізвище, ініціали)

2025 р.

## Харківський національний університет радіоелектроніки

Факультет Інформаційно-аналітичних технологій та менеджментуКафедра ІнформатикиРівень вищої освіти перший (бакалаврський)Спеціальність 122 Комп'ютерні науки  
(код і повна назва)Тип програми освітньо-професійнаОсвітня програма Інформатика  
(повна назва)

ЗАТВЕРДЖУЮ:

Зав. кафедри \_\_\_\_\_  
(підпис)

« \_\_\_\_\_ » \_\_\_\_\_ 2025 р.

**ЗАВДАННЯ**  
НА КВАЛІФІКАЦІЙНУ РОБОТУздобувачеві Редекопу Івану Євгеновичу  
(прізвище, ім'я, по батькові)1. Тема роботи Моделювання методу нечітких с-середніх з використанням еволюційної оптимізації

затверджена наказом університету від 19 травня 2025 року № 381Ст

2. Термін подання здобувачем роботи до екзаменаційної комісії 09 червня 2025 р.

3. Вихідні дані до роботи науково-методична та науково-технічна література, матеріали конференцій, дані інтернет-мережі, публікації з відкритим доступом щодо алгоритмів кластеризації, синтетичні та реальні набори даних, інструменти математичного моделювання, бібліотека для машинного навчання Scikit-learn.

4. Перелік питань, що потрібно опрацювати в роботі \_\_\_\_\_

1. Обґрунтування актуальності задачі кластеризації.

2. Огляд методів нечіткої кластеризації.

3. Формування методу гібридної нечіткої кластеризації та математичні моделі.

4. Експериментальне дослідження та порівняння ефективності з іншими методами.

5. Перелік графічного матеріалу із зазначенням креслеників, схем, плакатів, комп'ютерних ілюстрацій (п.5 включається до завдання за рішенням випускової кафедри) Кластеризація наборів даних, мета та завдання роботи, теоретичні основи нечіткої кластеризації, математичне моделювання методів нечіткої кластеризації, програмна реалізація та експерименти, порівняльний аналіз методів, висновки.

---



---



---



---



---



---

6. Консультанти розділів роботи (п.6 включається до завдання за наявності консультантів згідно з наказом, зазначеним у п.1)

Найменування розділу	Консультант (посада, прізвище, ім'я, по батькові)	Позначка консультанта про виконання розділу	
		підпис	дата

### КАЛЕНДАРНИЙ ПЛАН

№ з/п	Назва етапів роботи	Строк / терміни виконання етапів роботи	Примітка
1	Отримання завдання на кваліфікаційну роботу	07.04.2025	
2	Аналіз завдання, підбір літератури	08.04.25-10.04.25	
3	Аналіз літератури з досліджуваної проблеми	11.04.25-14.04.25	
4	Аналіз технічних засобів	15.04.25-20.04.25	
5	Розробка методу	21.04.25-27.04.25	
6	Програмна реалізація	28.04.25-11.05.25	
7	Оформлення пояснювальної записки	12.05.25-20.05.25	
8	Перевірка на нормоконтроль	21.05.25-01.06.25	
9	Перевірка на плагіат	21.05.25-01.06.25	
10	Рецензування	21.05.25-01.06.25	
11	Підготовка презентації та доповіді	21.05.25-18.06.25	
12	Занесення роботи в електронний архів	02.06.25-18.06.25	
13	Попередній захист кваліфікаційної роботи	02.06.25-18.06.25	

Дата видачі завдання 7 квітня 2025 р.

Здобувач \_\_\_\_\_  
(підпис)

Керівник роботи \_\_\_\_\_ доц. Шафроненко А.Ю.  
(підпис) (посада, прізвище, ініціали)

## РЕФЕРАТ/ABSTRACT

Пояснювальна записка до кваліфікаційної роботи: 68 с., 66 табл., 3 рис., 1 дод., 30 джерел.

**КЛАСТЕРИЗАЦІЯ, МЕТОД НЕЧІТКИХ С-СЕРЕДНІХ, МЕТОД НЕЧІТКОЇ ОПТИМІЗАЦІЇ РОЇВ, АЛГОРИТМ КЛАСТЕРИЗАЦІЇ.**

Об'єктом роботи є розробка комбінованого алгоритму, що поєднує метод нечітких с-середніх та метод нечіткої оптимізації роїв часток.

Метою даної роботи є створення та імітаційне дослідження продуктивного комбінованого підходу до нечіткої кластеризації, що базується на поєднанні методу нечітких с-середніх та методу нечіткої оптимізації роїв часток. Особливий акцент у роботі зроблено на обробці великих обсягів даних, де традиційні методи часто виявляються недостатньо ефективними.

У результаті роботи здійснена програмна реалізація комбінованого алгоритму методу нечітких с-середніх та методу нечіткої оптимізації роїв часток та проведена серія експериментів з метою зіставлення ефективності запропонованого рішення.

**CLUSTERIZATION, FUZZY C-MEANS METHOD, FUZZY SWARM OPTIMIZATION METHOD, CLUSTERIZATION ALGORITHM.**

The object of the work is the development of a combined algorithm that combines the fuzzy c-means method and the fuzzy particle swarm optimization method.

The aim of the work is to develop and simulate a productive combined approach to fuzzy clustering, based on a combination of the fuzzy c-means method and the fuzzy particle swarm optimization method. Special emphasis is placed on processing large amounts of data, where traditional methods often prove to be insufficiently effective.

As a result of the work, a software implementation of the combined algorithm of the fuzzy c-means method and the fuzzy particle swarm optimization method was carried out and a series of experiments were conducted to compare the effectiveness of the proposed solution.

## ЗМІСТ

Перелік умовних позначень, символів, одиниць, скорочень і термінів .....	6
Вступ.....	7
1 Кластеризація, види кластеризації .....	9
1.1 Огляд поняття «кластеризації».....	9
1.2 Цілі кластеризації.....	11
1.3 Методи кластеризації .....	13
1.4 Постановка задачі .....	18
2 Гібридні нечіткі с-середні і нечітка оптимізація роїв часток .....	19
2.1 Види алгоритмів кластеризації.....	19
2.1.1 Ієрархічні та плоскі алгоритми .....	19
2.1.2 Чіткі алгоритми кластеризації .....	21
2.1.3 Нечіткі алгоритми кластеризації .....	22
2.2 Алгоритм нечітких с-середніх .....	25
2.3 Оптимізація роїв часток .....	28
2.4 Нечітка оптимізація роїв часток для нечіткої кластеризації .....	30
2.5 Алгоритм нечіткої кластеризації з використанням гібридних нечітких с-середніх і нечіткої оптимізації роїв часток .....	32
3 Експериментальні результати .....	36
Висновки .....	63
Перелік джерел посилання .....	64
Додаток А Результати роботи алгоритмів .....	66

**ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СИМВОЛІВ, ОДИНИЦЬ,  
СКОРОЧЕНЬ І ТЕРМІНІВ**

FCM – Fuzzy C-Means (метод нечіткої кластеризації *c*-середніх)

PSO – Particle Swarm Optimization (метод рою часток)

FPSO – Fuzzy Particle Swarm Optimization (метод оптимізації рою часток)

## ВСТУП

В епоху великих даних, інтелектуальний аналіз даних став одним із головних тем як для академічних досліджень, так промислових застосувань. Зростаюча доступність великомасштабних, багатовимірних і неоднорідних наборів даних створює зростаючий попит на гнучкі, надійні та масштабовані методи аналізу даних та розпізнавання образів. Серед багатьох методів навчання кластеризація грає фундаментальну роль у розпізнаванні структур даних, зменшення вимірності та покращення інтерпретації. Кластеризація включає в себе групування схожих точок даних у кластери таким чином, щоб об'єкти в самих кластерах були більше схожі один на одного ніж на об'єкти у інших кластерах. Традиційні методи кластеризації, такі як  $k$ -середні, ієрархічна кластеризація та моделі суміші Гауса, широко використовувалися завдяки їх простоті та ефективності обчислень. Однак, ці методи часто припускають жорстке призначення точок кластерам, що обмежує їх ефективність коли дані демонструють накладення або неоднозначність.

Щоб подолати цю проблему, були розроблені методи моделювання нечіткої кластеризації, зокрема алгоритм нечітких  $c$ -середніх (FCM) представлений Бездеком. На відміну від жорсткої кластеризації, алгоритм нечітких  $c$ -середніх дозволяє кожній точці даних відноситись до кількох кластерів з різним ступенем приналежності, забезпечуючи більш виважений та реалістичний погляд на групування даних. Незважаючи на свою теоретичну елегантність, FCM алгоритм має певні обмеження. Зокрема, він дуже чутливий до початкових умов і може легко сходитися до кальних мінімумів, особливо коли застосовується до великих, шумних або невиконаних наборів даних. Крім того, FCM не гарантує глобальної оптимальності, оскільки він використовує оптимізацію на основі градієнта, якому не вистачає можливостей дослідження.

Вирішити цю проблему непросто, але останні дослідження вивчали інтеграцію еволюційних обчислювальних методів у процес нечіткої кластеризації. Один із найбільш багатообіцяючих підходів є гібридизація FCM

з оптимізацією роєм частинок (PSO) – популяційною стохастичною оптимізацією, натхненою соціальною поведінкою птахів і риб. PSO був успішно застосований до широкого кола задач комбінаторної та безперервної оптимізацію включаючи кластеризацію. У PSO рій частинок рухається простором рішення, динамічно регулюючи свої позиції на основі власного досвіду та досвіду своїх сусідів. Кожна частинка представляє потенційне рішення, і рій спільно шукає глобальний оптимум, збалансовуючи дослідження та експлуатацію. При застосуванні до кластеризації PSO може оптимізувати центри кластерів, матриці членства або обидва, залежно від схеми представлення. Комбінування FCM та PSO призводить до гібридної моделі нечіткої кластеризації, яка використовує швидкість локальної конвергенції FCM і можливості глобального пошуку PSO. Було запропоновано декілька гібридних моделей, включаючи FCM-PSO, FPSO-FCM і адаптивну нечітку PSO, кожна з яких стосується різних аспектів підвищення продуктивності. Ці моделі зазвичай ініціалізують рій частинок центрами кластерів, отриманими з FCM, і ітеративно вдосконалюють структуру кластеризації за допомогою PSO.

Актуальність роботи полягає у тому, що в умовах постійного зростання обсягів і складності даних, які необхідно обробляти в режимі реального часу, класичні методи кластерного аналізу втрачають ефективність. Поєднання алгоритму нечітких  $c$ -середніх та еволюційного алгоритму у межах гібридного методу кластеризації дозволяє усунути недоліки кожного з них окремо, що відкриває нові можливості для створення адаптивних та стійких систем аналізу даних у таких критично важливих сферах, як медицина, кібербезпека, фінансова аналітика, екологічний моніторинг тощо.

# 1 КЛАСТЕРИЗАЦІЯ, ВИДИ КЛАСТЕРИЗАЦІЇ

## 1.1 Огляд поняття «кластеризації»

Кластеризація, також відома як кластерний аналіз, є одним із основних методів неконтрольованого машинного навчання та інтелектуального аналізу даних. Це стосується процесу поділу набору даних на підмножини, які називаються кластерами, так що точки даних у кожному кластері є більш схожими одна на одну, ніж на точки в інших кластерах. Кластеризація розкриває приховану структуру даних, дозволяючи аналітикам ідентифікувати властиві групування, тенденції або шаблони без необхідності попередньо визначених міток або категорій [1], [2].

Загальною метою кластеризації є досягнення високої внутрішньокластерної подібності та низької міжкластерної подібності на основі попередньо визначеної метрики подібності чи відстані. Ці показники можуть бути геометричними, такими як евклідова відстань, або більш складними, такими як косинус подібності або відстань Махаланобіса, залежно від природи та структури набору даних.

Кластеризація виконує кілька завдань:

- узагальнення даних: шляхом зменшення великих наборів даних у репрезентативні групи стає легше інтерпретувати й аналізувати дані на вищому рівні абстракції;
- виявлення шаблонів: кластеризація допомагає виявити раніше невідомі зв'язки та структури в даних [3];
- попередня обробка для контрольованого навчання: кластери можна використовувати для ініціалізації контрольованих моделей або як функції для завдань класифікації;
- виявлення аномалій: об'єкти, які явно не належать до жодного кластера, можуть вважатися викидами або новинками.

Кластеризація широко використовується в різних сферах:

- у біології для класифікації видів на основі генетичних даних;
- у маркетингу сегментувати клієнтів на основі поведінки та демографічних показників;
- у медицині для виявлення закономірностей у симптомах пацієнтів або результатах діагностики [4];
- під час обробки зображень [5], для сегментації на основі регіону;
- у обробці природної мови, для тематичного моделювання та групування документів.

У кластеризації однією з ключових вимог є однорідність даних. Це означає, що всі елементи, які підлягають кластеризації, повинні бути схожими між собою за набором характеристик, що їх описують. Наприклад, при кластеризації клієнтів за їхніми покупками, важливо, щоб усі об'єкти (клієнти) мали однакові або схожі ознаки, які можна порівняти. Якщо дані мають різноманітні властивості, кластеризація може дати непередбачувані або неочевидні результати, адже алгоритм може помилково об'єднати різні групи об'єктів у один кластер.

Факторний аналіз є потужним інструментом, що допомагає досягти однорідності даних перед кластеризацією. Факторний аналіз дозволяє зменшити вимірність даних і виокремити основні фактори, що визначають їхню структуру. Ця процедура автоматично вирішує проблему однорідності, оскільки фактично перетворює вихідні ознаки на групи, які найбільше впливають на структуру вибірки. Завдяки такій трансформації зберігається важлива інформація, а додаткові шуми і несуттєві характеристики усуваються, що дозволяє покращити результат кластеризації.

Однак, якщо факторний аналіз не застосовується, то вибірку потрібно коригувати вручну, щоб забезпечити необхідну однорідність. В цьому випадку можуть використовуватися методи стандартної нормалізації або z-стандартизації, яка перетворює різні шкали вимірів на єдиний стандарт. Однак, зважаючи на можливі негативні наслідки, як, наприклад, зменшення чіткості поділу груп, такі стандартизації мають бути виконані обережно, щоб

не спотворити результат кластеризації. З цієї причини коригування вибірки є важливою складовою попередньої підготовки даних для кластеризації.

Концепція нечіткої кластеризації особливо актуальна в задачах, пов'язаних із невизначеністю, невизначеністю або перекриттям меж класів. Найвідомішим алгоритмом у цій області є алгоритм нечітких с-середніх (FCM), який мінімізує цільову функцію шляхом повторного оновлення центрів кластерів і значень членства [2, 6].

Однак традиційні методи кластеризації, включаючи FCM, стикаються з проблемами масштабованості, конвергенції до локальних оптимумів і чутливості до ініціалізації. Ці проблеми призвели до інтеграції методів метаевристичної оптимізації, таких як оптимізація роєм частинок (PSO), які допомагають краще досліджувати простір рішень і знаходити більш оптимальні конфігурації кластерів.

## 1.2 Цілі кластеризації

Кластеризація є основним інструментом дослідницького аналізу даних, призначеним для виявлення прихованої структури набору даних без попереднього знання міток класів. Основна мета полягає в тому, щоб визначити природні групи в даних, які відображають базові зв'язки або закономірності. Однак конкретні цілі кластеризації можуть відрізнятися залежно від області застосування та передбачуваного використання результатів. Нижче наведено найпоширеніші та важливі цілі, пов'язані з кластеризацією.

Розуміння даних і виявлення структури.

Однією з головних цілей кластеризації є отримання уявлення про базовий розподіл даних шляхом виявлення кластерів, які представляють значущі групи. Ці структури можуть відповідати окремим категоріям, функціональним ролям, поведінці користувачів або природним явищам.

Виявляючи ці кластери, дослідники та практики можуть краще інтерпретувати складні набори даних і формувати гіпотези для подальшого аналізу.

Узагальнення та стиснення даних.

Кластеризацію можна використовувати для зменшення розміру та складності великих наборів даних шляхом групування подібних точок даних і представлення їх через прототипи кластерів або центроїди. Це стиснення є корисним для зберігання, візуалізації та ефективності обчислень. У багатьох випадках кожен кластер може бути представлений одним репрезентативним екземпляром, що дозволяє спростити процеси моделювання та прийняття рішень.

Попередня обробка та розробка функцій.

Кластерні призначення можуть служити додатковими функціями в контрольованих навчальних завданнях або можуть керувати стратегіями зменшення розмірності та вибірки. Кластеризація допомагає визначити структуру даних і може бути застосована для виявлення надмірностей, ідентифікації прихованих змінних і сегментації даних для цілеспрямованого навчання моделі.

Виявлення викидів і новизни.

Іншою метою кластеризації є виявлення аномальних або нетипових точок даних, які чітко не належать жодному кластеру. Ці точки можуть вказувати на шум, помилки введення даних, шахрайські дії або справді рідкісні явища. Виявлення викидів широко використовується в системах кібербезпеки [7], виявлення шахрайства та моніторингу працездатності.

Підтримка прийняття рішень і формування політики.

У бізнесі, медицині та державній політиці кластеризація підтримує прийняття рішень на основі даних шляхом групування суб'єктів (наприклад, клієнтів, пацієнтів, регіонів) зі схожими характеристиками [8, 9]. Це допомагає формувати цільові втручання, персоналізовані рекомендації та оптимізовані стратегії розподілу ресурсів.

Побудова таксономії.

У науковому та академічному контекстах кластеризація часто використовується для побудови таксономій і систем класифікації. Наприклад, у біології кластеризація використовується для створення філогенетичних дерев, а в лінгвістиці для побудови мовних сімей. Ці таксономії забезпечують ієрархічну структуру, яка допомагає в концептуальній організації та відкритті знань.

Кожна з цих цілей висуває різні вимоги до алгоритму кластеризації. Наприклад, виявлення викидів вимагає чутливості до граничних випадків, тоді як побудова таксономії може вимагати підходів ієрархічної кластеризації. Подібним чином підсумовування даних надає перевагу компактності та репрезентативності над детальними межами.

У нечіткій кластеризації ці цілі розширюються, дозволяючи об'єктам частково належати до кількох кластерів. Це особливо актуально, коли дані мають структуру, що збігається, або властиву невизначеність. Ступені членства забезпечують більш детальне розуміння зв'язків між точками даних і кластерами, що часто є більш інформативним, ніж жорстке призначення [2].

### 1.3 Методи кластеризації

Методи кластеризації можна класифікувати на основі різних критеріїв, таких як їх базова математична модель, представлення пилу, рівень нагляду або обробка членства точки даних. Різноманітність алгоритмів відображає широкий спектр проблем, які постають перед даними реального світу, починаючи від масштабу, форми та рівнів шуму до розмірності та типу даних.

Далі наведено огляд основних класів методів кластеризації.

Методи поділу.

Методи розділення кластеризації ділять набір даних на заздалегідь визначену кількість кластерів, які зазвичай позначаються як  $k$ . Ці методи намагаються оптимізувати попередньо визначений критерій, зазвичай

пов'язаний із внутрішньокластерною компактністю та міжкластерним поділом.

- $k$ -means є найпопулярнішим алгоритмом розбиття. Він використовує евклідову відстань для мінімізації дисперсії всередині кластерів [1];
- $k$ -medoids покращує  $k$ -means, використовуючи фактичні точки даних замість центроїдів, що робить його більш стійким до викидів;
- нечіткі  $c$ -середні (FCM) розширюють  $k$ -середні, дозволяючи м'яке призначення точок даних декільком кластерам [2].

Методи поділу є обчислювально ефективними та масштабованими, але вони часто вимагають кількості кластерів як вхідних даних і можуть сходитися до локальних оптимумів.

Ієрархічні методи.

Ієрархічна кластеризація застосовується, коли необхідно поділити дані на різні рівні або категорії, починаючи від великих кластерів і поступово дроблячи їх на дрібніші. Цей підхід дозволяє будувати структуру, що нагадує дерево, де кожен вузол представляє собою кластер, а його підвузли – більш дрібні підкласи, що належать до цього кластеру. Такий тип кластеризації є особливо корисним для задач, де важливо зберегти інформацію про зв'язки між різними рівнями класифікації, як, наприклад, у завданнях таксономії, що зустрічаються в біології, соціології або інших науках.

Результатом такої ієрархічної кластеризації є деревоподібна структура, в якій об'єкти класифікуються на різних рівнях. На найвищому рівні розміщуються найбільш загальні кластери, що включають всі елементи, а на найнижчому рівні – конкретніші підкласи, що поділяються на ще менші групи. Наприклад, в таксономії рослин або тварин класифікація може починатися від великих категорій, таких як роди або види, і потім дробитися на підкатегорії, наприклад, підвиди чи популяції. Така структура дозволяє легко орієнтуватися в даних і зрозуміти, як елементи класифікуються за різними ознаками [10].

Крім того, кожен об'єкт в ієрархічній класифікації може належати до кількох кластерів, що характеризують його на різних рівнях. Цей підхід

дозволяє отримати більш детальну інформацію про об'єкти, визначаючи, до яких кластерів вони належать і з якими іншими елементами мають схожість. Таким чином, ієрархічна кластеризація не лише забезпечує класифікацію, але й створює чітку структуру зв'язків між об'єктами, що дозволяє використовувати її в ряді різних дисциплін для аналізу складних систем.

Загальні критерії підключення включають:

- одиночна тяга (мінімальна відстань);
- повне з'єднання (максимальна відстань);
- середнє зчеплення (середня відстань);
- метод Уорда (мінімізує внутрішньокластерну дисперсію).

Методи на основі щільності.

Кластеризація на основі щільності ідентифікує кластери як області високої щільності, розділені областями низької щільності. Ці алгоритми особливо гарні для ідентифікації кластерів довільної форми та виявлення викидів.

– DISCAN (просторова кластеризація додатків із шумом на основі щільності) – це популярний алгоритм, який групує тісно розташовані точки та позначає розріджені області як шум [11].

– OPTICS розширює DBSCAN, обробляючи кластери з різною щільністю.

Ці методи не вимагають попередньої кількості кластерів, але чутливі до параметрів щільності, таких як радіус околиці ( $\epsilon$ ) і мінімальні точки.

Методи на основі моделі.

Ці методи припускають, що дані генеруються сумішшю базових розподілів ймовірностей, як правило, Гауса. Моделі суміші Гауса (GMMs) [12] використовують алгоритм очікування-максимізації (EM) для оцінки параметрів кожного розподілу.

Кластеризація на основі моделі підходить, коли дані відповідають імовірнісному шаблону та забезпечують статистичну інтерпретацію, але вона вимагає припущень щодо розподілу даних.

Грид-базовані моделі.

Методи на основі сітки ділять простір даних на кінцеву кількість комірок для узагальнення даних і виконання кластеризації на основі цих сіток [13].

STING (Сітка статистичної інформації) і CLIQUE є добре відомими алгоритмами на основі сітки, які використовуються в просторовому та високовимірному аналізі даних.

Ці методи ефективні з точки зору обчислювального часу, особливо для великих баз даних.

Спектральна кластеризація.

Спектральні методи використовують власні вектори матриць подібності для виконання зменшення розмірності перед застосуванням традиційного алгоритму кластеризації, такого як  $k$ -середні [14].

Вони особливо корисні, коли структура кластера є невивуклою або нероздільною у вихідному просторі ознак.

Методи нечіткої кластеризації.

Нечітка кластеризація використовується коли об'єкти можуть частково належати до кількох кластерів. Цей підхід дозволяє кожному об'єкту бути представленим не чітким належанням до кластеру, а певним ступенем приналежності, що варіюється від 0 до 1. Це дозволяє більш гнучко класифікувати дані, що особливо важливо для задач, де об'єкти знаходяться на межі між кількома категоріями, і не можуть бути чітко віднесені до одного з кластерів. Наприклад, при класифікації даних, де об'єкти мають спільні характеристики з кількома групами, нечітка кластеризація може допомогти визначити їх належність до кожної з груп з певним ступенем ймовірності.

Один із найбільш широко використовуваних алгоритмів нечіткої кластеризації – це нечіткі  $c$ -середні (FCM). Цей метод мінімізує зважену цільову функцію, яка збалансовує ступінь приналежності об'єктів до кластерів і відстань між об'єктами та їх центроїдами. FCM дозволяє кожному елементу належати до кількох кластерів одночасно, що дає змогу більш точно класифікувати об'єкти, які можуть бути схожими на кілька груп, але не

відносяться до них на 100%. Цей підхід широко використовується в обробці зображень, біологічних даних, а також у фінансовому та маркетинговому аналізі.

Існують і інші варіанти та розширення цього алгоритму, такі як можливі *c*-середні (PCM) [15] та нечіткий алгоритм Густафсона-Кесселя [16]. Алгоритм PCM є розширенням класичних *c*-середніх, який вводить нову функцію належності, що дає змогу краще відображати можливі належності до кластерів, враховуючи невизначеність даних. Водночас алгоритм Густафсона-Кесселя розширює базову модель, використовуючи адаптивні мірки відстані для більш точного класифікування об'єктів, що мають різні типи та масштаби вхідних даних. Такі модифікації допомагають зробити алгоритм більш універсальним і застосовним для складних або високорозмірних даних, де стандартні методи кластеризації можуть не забезпечити достатньої точності [17].

Еволюційний та ройовий методи.

Нечітка кластеризація є особливо корисною для даних з неоднозначними межами, де традиційні методи можуть зазнати труднощів при чіткому поділі на групи [18, 19]. Вона дозволяє отримати більш інформативні призначення, забезпечуючи не лише класифікацію, а й ступінь приналежності кожного об'єкта до кожної категорії.

До них належать алгоритми кластеризації, які використовують еволюційні обчислення, такі як генетичні алгоритми (GA) [20], оптимізація роєм частинок (PSO) [21] і оптимізація колонії мурах (ACO) [22]. Ці методи оптимізують цілі кластеризації за допомогою стратегій пошуку на основі населення.

Кластеризація на основі PSO була успішно застосована для оптимізації центроїдів кластерів і матриць приналежності в нечітких моделях.

Такі методи особливо корисні для уникнення локальних мінімумів і пошуку глобально оптимальних рішень у складних просторах пошуку.

#### 1.4 Постановка задачі

Об'єктом роботи є розробка комбінованого алгоритму, що поєднує метод нечітких  $c$ -середніх та метод нечіткої оптимізації роїв часток.

Метою даної роботи є створення та імітаційне дослідження продуктивного комбінованого підходу до нечіткої кластеризації, що базується на поєднанні методу нечітких  $c$ -середніх та методу нечіткої оптимізації роїв часток. Особливий акцент у роботі зроблено на обробці великих обсягів даних, де традиційні методи часто виявляються недостатньо ефективними.

Для реалізації цієї мети передбачено виконання наступних завдань:

- здійснити огляд та аналіз існуючих підходів до нечіткої кластеризації, зокрема визначити їх сильні сторони та недоліки;
- спроектувати та реалізувати комбінований алгоритм, що інтегрує механізми FCM та PSO для покращення точності кластеризації;
- провести серію експериментів з метою зіставлення ефективності запропонованого рішення з відомими алгоритмами та підтвердити переваги гібридного підходу.

## 2 ГІБРИДНІ НЕЧІТКІ С-СЕРЕДНІ І НЕЧІТКА ОПТИМІЗАЦІЯ РОЇВ ЧАСТОК

### 2.1 Види алгоритмів кластеризації

#### 2.1.1 Ієрархічні та плоскі алгоритми

Методи кластеризації прийнято поділяти на дві основні групи: ієрархічні та плоскі алгоритми. Ієрархічні алгоритми, які також позначаються як таксономічні, не обмежуються формуванням одного-єдиного розбиття вибірки на взаємонеперетинні кластери. Замість цього вони створюють ієрархічну структуру вкладених кластерів. У результаті застосування таких алгоритмів формується дерево кластерів, де кореневий вузол відповідає всій вибірці, а листкові вузли – найдрібнішим кластерам, що містять окремі об'єкти.

На відміну від них, плоскі алгоритми кластеризації передбачають побудову лише одного рівня розбиття множини об'єктів на кластери, без врахування ієрархічної структури або вкладеності.

У межах ієрархічної кластеризації розрізняють два підходи:

– низхідний (дивізивний) – кластеризація здійснюється за принципом «згори донизу». На початковому етапі всі об'єкти об'єднуються в один загальний кластер, який поступово поділяється на все дрібніші підкласи. Процес триває доти, доки не буде досягнуто необхідного ступеня деталізації;

– висхідний (агломеративний) – кластеризація виконується за принципом «знизу догори». Початково кожен об'єкт розглядається як окремий кластер. Далі кластери поетапно об'єднуються між собою згідно з певним критерієм подібності, допоки вся вибірка не буде згрупована в один загальний кластер.

Найбільш поширеним у практиці є саме висхідний підхід, оскільки він дозволяє зберігати послідовність злиття об'єктів і краще відображає природну структуру даних.

Результати ієрархічної кластеризації зазвичай візуалізуються у вигляді дендрограми – дерева, що демонструє процес злиття (або розділення) кластерів на кожному кроці алгоритму [23]. Класичним прикладом побудови подібної структури є біологічна класифікація живих організмів (тварин і рослин).

Одним із ключових аспектів реалізації ієрархічних алгоритмів є вибір способу обчислення відстані між кластерами [24]. Від нього залежить порядок об'єднання об'єктів і якість отриманої кластерної структури. Найчастіше застосовуються такі метрики:

- **одиначний зв'язок** – відстань між двома кластерами визначається як мінімальна відстань між будь-якою парою об'єктів, що належать до різних кластерів. Цей метод має тенденцію до утворення витягнутих кластерів (ланцюжків), оскільки об'єднання здійснюється за найменшою відстанню;

- **повний зв'язок** – у цьому випадку враховується максимальна відстань між будь-якими двома об'єктами з різних кластерів. Метод забезпечує формування компактних кластерів та стійкий до викидів, але є менш ефективним при наявності витягнутих або нерівномірних груп;

- **невиважене попарне середнє** – середня відстань між усіма можливими парами об'єктів із різних кластерів. Цей підхід є збалансованим і добре працює як у випадках компактних, так і протяжних кластерів;

- **виважене попарне середнє** – аналог попереднього методу, однак під час обчислення середнього значення враховується кількість об'єктів у кожному кластері. Таким чином, великі кластери мають більший вплив на розрахунок міжкластерної відстані;

- **невиважений центроїдний метод** – відстань між центрами тяжіння (центроїдами) двох кластерів. Метод є менш чутливим до крайніх значень, але може призводити до ефекту злиття кластерів, розташованих на значній відстані;

- **виважений центроїдний метод** – модифікація попереднього методу, яка враховує вагові коефіцієнти, що залежать від розміру кластерів. Метод

доцільно застосовувати, коли об'єкти розподілені нерівномірно, і присутні суттєві відмінності у кількості елементів у кожному кластері.

Серед основних недоліків ієрархічних алгоритмів слід відзначити їх незмінність після виконання: кожен крок злиття або розбиття є остаточним. Крім того, отримана система повних розбиттів може виявитися надмірною з огляду на конкретну задачу кластеризації, що потребує вибору лише релевантного рівня деталізації.

### 2.1.2 Чіткі алгоритми кластеризації

У чітких методах кластеризації початкову множину об'єктів  $X$  поділяють на кілька підмножин таким чином, щоб кожен елемент належав лише одному кластеру. Інакше кажучи, ці підмножини не перетинаються між собою, і кожен об'єкт асоціюється виключно з однією групою.

Одним з найвідоміших представників цього класу методів є алгоритм  $k$ -середніх. Він здійснює розподіл об'єктів на  $k$  кластерів, де значення  $k$  задається користувачем наперед, виходячи з конкретних потреб задачі. Основний принцип роботи алгоритму полягає в ітеративному оновленні центрів кластерів до тих пір, поки не буде досягнута певна стабільність у розподілі об'єктів.

Попри популярність, метод  $k$ -середніх має низку обмежень. Зокрема, він погано справляється із задачами, де кластери не мають чітких меж – тобто, у випадках, коли об'єкти не можна однозначно віднести до однієї групи. Крім того, алгоритм гарантує лише знаходження одного з локальних мінімумів функції цільового відхилення (зазвичай це сума квадратів відстаней об'єктів до відповідних центрів кластерів), але не обов'язково глобального мінімуму. Значною мірою якість кластеризації залежить від початкового вибору центрів, а спосіб їх оптимального задання наперед невідомий, що також знижує надійність результату.

### 2.1.3 Нечіткі алгоритми кластеризації

У 1965 році Лофті Заде вперше запропонував теорію нечітких множин, яку він використовував для моделювання та аналізу систем, що включають людину. Основна ідея цього підходу полягає в тому, що елементи людського мислення не є точними числами, а елементами нечітких множин. Важливою особливістю цих множин є те, що перехід від «належності» до «неналежності» не є різким, а відбувається безперервно, що дозволяє створювати моделі, близькі до реальних процесів мислення. Теорія нечітких множин, запропонована Заде, сьогодні знайшла широке застосування в багатьох прикладних сферах, зокрема в управлінні підприємствами, контролі якості продукції та оптимізації технологічних процесів.

Цю теорію було адаптовано до кластеризації, в результаті чого з'явилися нечіткі методи кластеризації. На відміну від класичних методів, які чітко відносять об'єкти до одного кластера, нечітка кластеризація дозволяє одному об'єкту належати до кількох кластерів одночасно, причому ступінь належності до кожного з кластерів може бути різним. Алгоритми нечіткої кластеризації дозволяють розподіляти об'єкти на кілька кластерів з визначенням ступеня приналежності кожного об'єкта до кожного з кластерів. Ступінь приналежності визначається на основі відстані між об'єктом і центром кластеру.

Нечітка кластеризація є синтезом класичних ідей кластерного аналізу і теорії нечітких множин, що описані в роботах Заде. Замість традиційного чіткого розподілу об'єктів на кластери застосовується нечітке розбиття, де кожному об'єкту ставляться функції приналежності, значення яких змінюються в межах від 0 до 1. Це дозволяє отримати більш точне представлення про належність об'єктів до кластерів, оскільки кожен об'єкт може належати до кількох кластерів одночасно з певним ступенем ймовірності.

Нехай маємо  $n$  об'єктів  $X_1, X_2, \dots, X_n$  та  $k$  кластерів, на які необхідно розподілити ці об'єкти. Кожному об'єкту  $X_j$  присвоюється ймовірність належності до кожного з кластерів  $\mu_{ji}$ . У класичній задачі кожному об'єкту ставиться у відповідність лише один кластер, але в контексті нечіткої кластеризації для кожного об'єкта ми отримуємо набори функцій приналежності  $\mu_{j1}, \mu_{j2}, \dots, \mu_{jk}$ , де кожне значення  $\mu_{ji}$  є ймовірністю того, що об'єкт  $X_j$  належить до  $i$ -го кластеру.

Нечіткі кластери можна описати за допомогою матриці нечіткого розбиття  $F$ , де  $\mu_{ki}$  є ступенем належності об'єкта  $X_k$  до кластеру  $i$ , і ці значення лежать в інтервалі  $[0, 1]$ . Формула для такої матриці виглядає так:

$$F = [\mu_{ki}], \mu_{ki} \in [0, 1], k = 1, M, i = 1, c, \quad (2.1)$$

де  $M$  – кількість об'єктів;

$c$  – кількість кластерів.

Існують певні умови для матриці нечіткого розбиття, які гарантують коректність результатів кластеризації. Перша умова (2.2) вимагає, щоб для кожного об'єкта  $k$  сума ступенів належності цього об'єкта до всіх кластерів дорівнювала одиниці:

$$\sum_{i=1}^c u_{ki} = 1, k = 1, M. \quad (2.2)$$

Друга умова (2.3) вказує, що сума ступенів належності всіх об'єктів до кожного з кластерів повинна бути строго більше нуля, але менше певного значення  $N$ :

$$0 < \sum_{k=1}^M u_{ki} < N, i = 1, c. \quad (2.3)$$

Нечітке розбиття дозволяє коректно працювати з об'єктами, які знаходяться на межі між двома або більше кластерами, присвоюючи їм ступінь

належності, рівний 0.5. Однак проблема може виникнути при обробці об'єктів, що знаходяться далеко від центрів кластерів. Для таких об'єктів інтуїтивно слід призначити низькі ступені належності, але при цьому сума ступенів приналежності всіх об'єктів до кластерів повинна залишатися рівною одиниці. Для усунення цієї проблеми пропонується можливе розбиття, яке гарантує, що кожен об'єкт буде належати хоча б до одного кластеру. Це можна забезпечити ослабленням умови (2.2):

$$\sum_{i=1}^c u_{ki} \leq 1. \quad (2.4)$$

Для оцінки якості нечіткого розбиття використовується критерій розкиду:

$$J(U, V) = \sum_{i=1}^c \sum_{k=1}^M u_{ki}^m \|V_i - X_k\|^2, \quad (2.5)$$

де  $V_i$  – центри нечітких кластерів;

$\|V_i - X_k\|^2$  – відстань між об'єктами  $X_k$  та центром кластеру  $V_i$ ;

$m$  – експонентна вага, що визначає нечіткість і розмитість кластерів.

Цей критерій дозволяє визначити, як добре кластеризація описує структуру даних. Щоб знайти оптимальні центри кластерів, застосовуються такі формули для розрахунку центрів нечітких кластерів:

$$V_i = \frac{\sum_{k=1}^M (u_{ki})^m X_k}{\sum_{k=1}^M (u_{ki})^m}. \quad (2.6)$$

Ці формули дозволяють точно обчислити центри нечітких кластерів на основі ступеня приналежності кожного об'єкта до кожного кластеру.

## 2.2 Алгоритм нечітких $c$ -середніх

Алгоритм нечітких  $c$ -середніх (FCM) є популярним методом кластеризації, що використовує методи невизначених множників Лагранжа для визначення оптимальних кластерів[25]. Цей алгоритм, розроблений в 1974 році, надає змогу здійснювати класифікацію об'єктів, які належать до кількох кластерів одночасно, що дає більш гнучке представлення даних.

Завдання алгоритму полягає в тому, щоб поділити множину об'єктів, представлених векторами ознак, на кілька кластерів. У цьому підході кожен об'єкт може одночасно належати до кількох кластерів, причому з різним ступенем приналежності. Відмінною рисою є те, що кордони між кластерами також нечіткі, що дозволяє більш точно моделювати реальні ситуації, де об'єкти не можна однозначно віднести до одного кластеру.

Алгоритм працює таким чином, що кожен об'єкт має певний ступінь належності до всіх кластерів, що обчислюється за допомогою відстані між об'єктом і центром кожного кластера. Ітераційно обчислюються нові центри кластерів і оновлюються ступені приналежності, поки зміни в результатах не стануть мінімальними.

Алгоритм нечітких середніх передбачає, що набір  $n$  об'єктів  $O = \{o_1, o_2, \dots, o_n\}$ , що знаходяться в  $d$ -мірному просторі  $R^d$ , сегментується на  $c$  нечітких кластерів, де  $1 < c < n$ . Кожен кластер має свій центр або центроїд  $Z = \{z_1, z_2, \dots, z_c\}$ , і класифікація об'єктів описується матрицею приналежності  $\mu$ , яка має розміри  $n \times c$ . Елемент  $\mu_{ij}$  матриці вказує на ступінь приналежності  $i$ -го об'єкта до  $j$ -го кластеру.

Для матриці  $\mu$  використовуються такі умови:

$$\mu_{ij} \in [0,1], \quad \forall i = 1,2, \dots, n; \quad \forall j = 1,2, \dots, c; \quad (2.7)$$

$$\sum_{j=1}^c \mu_{ij} = 1, \quad \forall i = 1,2, \dots, n; \quad (2.8)$$

$$0 < \sum_{i=1}^n \mu_{ij} < n, \forall j = 1, 2, \dots, c. \quad (2.9)$$

Цільова функція алгоритму нечітких середніх полягає в мінімізації наступного рівняння:

$$J_m = \sum_{j=1}^c \sum_{i=1}^n u_{ki}^m d_{ij}^2; \quad (2.10)$$

$$d_{ij} = \|o_i - z_j\|, \quad (2.11)$$

де  $m$  – скаляр, який визначає ваговий показник і контролює нечіткість кластерів.

Відстань  $d_{ij}$  обчислюється як евклідова відстань між об'єктами  $o_i$  і  $z_j$ , де  $z_j$  є центроїдом  $j$ -го кластеру. Центроїд  $z_j$  визначається за формулою:

$$z_j = \frac{\sum_{i=1}^n \mu_{ij}^m o_i}{\sum_{i=1}^n \mu_{ij}^m}. \quad (2.12)$$

Цей вираз визначає центр  $j$ -го кластеру як середнє значення всіх об'єктів, зважених за ступенем їх належності до цього кластеру.

Процес кластеризації за допомогою нечітких  $c$ -середніх може бути представленим наступними етапами:

Крок 1. Вибір параметра  $m$  (де  $m > 1$ ), визначення функцій приналежності  $\mu_{ij}$  для кожного об'єкта та кластеру (де  $i=1, 2, \dots, n, j=1, 2, \dots, c$ ).

Крок 2. Обчислення центрів кластерів  $z_j$  за допомогою рівняння (2.12).

Крок 3. Визначення евклідової відстані  $d_{ij}$  для кожного об'єкта та кластера.

Крок 4. Оновлення функцій приналежності  $\mu_{ij}$  згідно з рівнянням (2.13).

$$\mu_{ij} = \frac{1}{\sum_{k=1}^c \left(\frac{d_{ij}}{d_{ik}}\right)^{\frac{2}{m-1}}}. \quad (2.13)$$

Крок 5. Перевірка на зближення: якщо зміни незначні, алгоритм завершується, інакше переходить до Кроку 2.

Для зупинки алгоритму можна застосовувати кілька правил. Одним із них є припинення при досягненні невеликої зміни значень центроїдів або неможливості подальшої мінімізації цільової функції. Важливим аспектом є вибір кількості кластерів  $c$ , оскільки правильно визначити кількість кластерів для реальних даних без додаткової інформації складно.

Існують два основних підходи для вибору оптимальної кількості кластерів:

– підхід на основі критеріїв компактності та відокремленості: за умови правильного вибору кількості кластерів, об'єкти повинні бути розподілені по компактним і добре відокремленим групам. Якщо кластеризація вибрана неправильно, то кластери будуть менш компактними і погано відокремленими. Одним із критеріїв для оцінки компактності є індекс Хіє-Бені:

$$\chi = \frac{\sum_{i=1,c} \sum_{k=1,M} (\mu_{ik})^m \cdot \|X_k - V_i\|^2}{M \cdot \min_{i \neq j} (\|X_k - V_i\|^2)}. \quad (2.14)$$

– підхід на основі послідовного об'єднання кластерів: починаючи з великої кількості кластерів, вони поступово об'єднуються на основі схожості, визначеної за допомогою формальних критеріїв.

Іншим важливим параметром алгоритму є експоненціальна вага  $m$ . Збільшення значення  $m$  призводить до «розмивання» кластерів, а при  $m \rightarrow \infty$  матриця приналежності  $F$  набуває вигляду  $F = [1/c]$ , що є непридатним рішенням, оскільки всі об'єкти будуть однаково належати до кожного кластеру. Використання більшого значення  $m$  дозволяє посилити вплив об'єктів з великим ступенем належності і зменшити вплив об'єктів з малими значеннями ступеня належності. Зазвичай  $m$  встановлюють рівним 2, однак оптимальне значення експоненціальної ваги ще не визначено теоретично.

### 2.3 Оптимізація роїв часток

Оптимізація роїв часток (Particle Swarm Optimization, PSO) є потужним методом чисельної оптимізації, який належить до класу еволюційних алгоритмів. Цей підхід заснований на моделюванні поведінки соціальних організмів, таких як зграї птахів, косяки риб [26] або колонії комах. Основна ідея полягає у тому, що група агентів (часток), які взаємодіють між собою, здатна знаходити оптимальні рішення задачі завдяки обміну інформацією про вже досліджені області простору пошуку.

Метод було вперше запропоновано Джеймсом Кеннеді та Расселом Еберхартом у 1995 році [27]. Спочатку PSO розроблявся як інструмент моделювання соціальної поведінки, проте з часом перетворився на універсальний механізм оптимізації, який активно застосовується у різних галузях – від машинного навчання та комп'ютерного зору до біоінформатики та інженерних задач.

У класичному варіанті PSO алгоритм працює з популяцією часток, кожна з яких представляє потенційне рішення задачі. Частки мають дві основні характеристики – положення в просторі рішень та швидкість переміщення. На початку роботи алгоритму ці параметри ініціалізуються випадковим чином. У кожній наступній ітерації частки оновлюють своє положення на основі попереднього досвіду, а також інформації, отриманої від інших часток популяції.

Кожна частка запам'ятовує:

- власне найкраще знайдене положення  $pbest$ ;
- глобальне найкраще положення  $gbest$ , досягнуте всією популяцією.

Швидкість та положення частки на наступному кроці розраховуються за наступними формулами:

$$V(t + 1) = wV(t) + c_1r_1(pbest(t) - X(t)) + c_2r_2(gbest(t) - X(t)); \quad (2.15)$$

$$K = 1, 2, \dots, P; \quad (2.16)$$

$$X_0(t + 1) = X(t) + V(t + 1), \quad (2.17)$$

де  $X(t)$  – поточне положення частки в просторі пошуку;

$V(t)$  – швидкість;

$w$  – маса інерції;

$c_1, c_2$  – коефіцієнти прискорення, що визначають інтенсивність впливу особистого та колективного досвіду;

$r_1, r_2$  – випадкові значення в інтервалі  $[0, 1]$ , які додають стохастичність у процес пошуку;

$P$  – загальна кількість часток у популяції.

Ітеративне оновлення положення часток забезпечує поступову конвергенцію до глобального або локального оптимуму задачі. Алгоритм PSO має високу швидкість збіжності та здатен ефективно працювати навіть у складних, багатовимірних просторах рішень. Крім того, на відміну від деяких інших еволюційних методів (наприклад, генетичних алгоритмів), PSO не потребує операцій схрещування чи мутації, що спрощує його реалізацію.

У 2004 році Пенг та його колеги [28] розширили класичну модель PSO, запропонувавши так звану нечітку оптимізацію роїв часток (Fuzzy Particle Swarm Optimization, FPSO). Це розширення передбачає використання концепцій нечіткої логіки для кращого моделювання невизначеності в процесі пошуку та адаптивного налаштування параметрів алгоритму. Зокрема, у FPSO можуть використовуватися нечіткі правила для управління значенням коефіцієнтів інерції або прискорення, що дозволяє зробити процес пошуку більш гнучким і ефективним, особливо у задачах, де присутня нечіткість в описі параметрів або обмежень.

## 2.4 Нечітка оптимізація роїв часток для нечіткої кластеризації

Метод нечіткої оптимізації роїв часток (FPSO) був запропонований Пенном як модифікація класичної оптимізації роїв часток для вирішення задачі комівояжера. Цей підхід включає інтеграцію нечітких елементів у модель, що дозволяє краще працювати з нечіткими даними. У цьому розділі ми адаптуємо метод FPSO для розв'язання задач кластеризації, де необхідно визначити оптимальні центри нечітких кластерів[29].

У алгоритмі нечіткої оптимізації роїв часток кожен елемент рою має позицію  $X$ , яка відображає нечіткий зв'язок між об'єктами даних  $o = \{o_1, o_2, \dots, o_n\}$  та відповідними центрами кластерів  $Z = \{z_1, z_2, \dots, z_c\}$ . Позиція частки в рою виражається через функції приналежності  $\mu_{ij}$ , де  $\mu_{ij}$  відображає ступінь належності об'єкта  $o_i$  до кластеру  $z_j$ , з обмеженнями, що описані в рівняннях (2.7) і (2.8). Таким чином, кожен елемент рою представляє нечітку матрицю функцій приналежності, яка дозволяє точно відобразити ступінь належності об'єктів до кожного кластеру.

Матриця положень часток аналогічна до матриці приналежності в алгоритмі нечітких  $c$ -середніх. Швидкість часток у рою задається матрицею розміру  $n \times c$ , де елементи перебувають в межах від  $[-1, 1]$ . Оновлення позицій і швидкостей часток відбувається за допомогою наступних рівнянь:

$$V(t + 1) = w \odot V(t) \oplus (c_1 r_1) \odot (pbest(t) \oplus X(t)) \oplus (c_2 r_2) \odot (gbest(t) \oplus X(t)); \quad (2.18)$$

$$X(t + 1) = X(t) \oplus V(t + 1), \quad (2.19)$$

де  $V(t)$  – швидкість частки на кроці  $t$ ;

$pbest(t)$  – найкраще локальне положення частки на кроці  $t$ ;

$gbest(t)$  – глобально найкраще положення в популяції.

Після оновлення матриці положень може виникнути ситуація, коли деякі елементи порушують умови. Для виправлення цього необхідно нормалізувати матрицю положень: всі негативні елементи переводяться в нуль. Якщо в рядку всі елементи матриці дорівнюють нулю, вони мають бути повторно оцінені за допомогою випадкових чисел в інтервалі  $[0, 1]$  після чого застосовується перетворення для коригування цих значень.

$$X_{normal} = \begin{bmatrix} \frac{\mu_{11}}{\sum_{j=1}^c \mu_{1j}} & \dots & \frac{\mu_{1c}}{\sum_{j=1}^c \mu_{1j}} \\ \vdots & \ddots & \vdots \\ \frac{\mu_{n1}}{\sum_{j=1}^c \mu_{nj}} & \dots & \frac{\mu_{nc}}{\sum_{j=1}^c \mu_{nj}} \end{bmatrix}. \quad (2.20)$$

Оцінка придатності рішень здійснюється через функцію оцінки, що в даному контексті визначається як:

$$f(X) = \frac{K}{J_m}, \quad (2.21)$$

де  $K$  – константа;

$J_m$  – функція, що визначає якість кластеризації, яку можна мінімізувати (2.10).

Чим менше значення функції  $J_m$ , тим краще результат кластеризації, що призводить до покращення індивідуальної придатності частки  $f(X)$ .

Алгоритм нечіткої оптимізації роїв часток для кластеризації:

Крок 1. Ініціалізувати параметри алгоритму, включаючи розмір популяції  $P$ , коефіцієнти  $c_1, c_2$ , ваговий коефіцієнт  $w$  та максимальну кількість ітерацій.

Крок 2. Створити початкову популяцію з  $P$  часток. Кожна частка має своє положення  $X$ , локальне найкраще положення  $pbest$ , глобальне найкраще положення  $gbest$  та швидкість  $V$ , де кожен з цих параметрів є матрицею розміру  $n \times c$ .

Крок 3. Ініціалізувати положення  $X$ , швидкість  $V$ , локальне найкраще положення  $pbest$  для кожної частки і глобальне найкраще положення  $gbest$  для всього рою.

Крок 4. Обчислити центри кластерів для кожної частки за допомогою рівняння (2.12), використовуючи її поточне положення.

Крок 5. Для кожної частки розрахувати значення придатності за допомогою функції оцінки, визначеної в рівнянні (2.21).

Крок 6. Для кожної частки оновити її локальне найкраще положення  $pbest$ , вибираючи те, яке має найкраще значення придатності.

Крок 7. Оновити глобальне найкраще положення  $gbest$  для рою, вибираючи найбільш придатну частку серед усіх.

Крок 8. Оновити швидкість кожної частки за допомогою рівняння (2.18), використовуючи поточні значення її положення, найкращі локальні та глобальні положення.

Крок 9. Оновити положення кожної частки за допомогою рівняння (2.19), коригуючи її положення згідно з новими значеннями швидкості.

Крок 10. Якщо умова зупинки не виконана, повернутися до Кроку 4.  
Умови зупинки: Алгоритм завершується, коли досягнута максимальна кількість ітерацій або коли не відбувається покращення значення  $gbest$  протягом кількох ітерацій.

## 2.5 Алгоритм нечіткої кластеризації з використанням гібридних нечітких $s$ -середніх і нечіткої оптимізації роїв часток

Метод нечіткої кластеризації з використанням  $s$ -середніх працює швидше, ніж метод нечіткої оптимізації роїв часток завдяки меншій кількості обчислень, необхідних для функцій, але він схильний до потрапляння в локальні оптимуми. У даному підході поєднуються алгоритм нечітких  $s$ -середніх та метод нечіткої оптимізації роїв часток, створюючи гібридну

стратегію під назвою FCM-FPSO, яка об'єднує сильні сторони обох підходів: швидкість кластеризації  $c$ -середніх і здатність роїв часток уникати локальних мінімумів [30].

Цей комбінований метод застосовує алгоритм  $c$ -середніх до часток рою на кожній ітерації, що дозволяє поступово поліпшувати функцію придатності кожної частки. Початкові оптимальні рішення отримуються за допомогою алгоритму  $c$ -середніх, що служить основою для подальшої роботи алгоритму FPSO. Спочатку результати від алгоритму  $c$ -середніх (позначені як  $X_{FCM}$ ) використовуються для ініціалізації частини рою для оптимізації рою часток. Інші частки рою,  $X_{rndj}$  (де  $j=1,2,\dots,N_{PSO}-N_{FCM}$ ), заповнюються випадковим чином, створюючи початкову популяцію для подальшої оптимізації.

У методі нечіткої кластеризації кожна частка представляє собою вектор, що відповідає центрам кластерів. Тобто кожна частка  $part_l$  є набором таких векторів:

$$part_l = (V_1, V_2, \dots, V_i, \dots, V_c), \quad (2.22)$$

де  $l$  – номер частки;

$V_i$  – центр  $i$ -го кластеру.

Таким чином, рій складається з кількох кандидатів на класифікацію для поточного набору даних. Кожна точка або вектор даних має певну ступінь належності до кожного кластеру згідно з його функцією приналежності. Це означає, що кожен об'єкт має нечітке належність до кластерів. Кожен кластер має свій центр, і на кожній ітерації частки представляють рішення у вигляді набору центрів кластерів.

Алгоритм нечітких  $c$ -середніх зазвичай сходиться швидше, ніж метод нечіткої оптимізації роїв часток, однак може мати менш точні результати кластеризації. Для покращення продуктивності кластеризації в цьому розділі пропонується комбінувати обидва підходи, заповнюючи початковий рій

результатами алгоритму  $c$ -середніх, що дозволяє використовувати швидкість і точність обох методів.

Алгоритм нечіткої кластеризації, який комбінує методи нечітких  $c$ -середніх і нечіткої оптимізації роїв часток, може бути сформульовано наступним чином:

Крок 1: Ініціалізувати необхідні параметри для алгоритмів нечітких  $c$ -середніх і нечіткої оптимізації роїв часток, включаючи кількість часток в популяції  $P$ , коефіцієнти  $c_1$ ,  $c_2$ , вагу  $w$  і експоненціальний показник  $m$ .

Крок 2. Створити рій, що складається з  $P$  часток. Для кожної частки визначити її положення  $X$ , локально найкраще положення  $pbest$ , глобальне найкраще положення  $gbest$  і швидкість  $V$ , де кожен з цих параметрів представляє собою матрицю розміру  $n \times c$ .

Крок 3. Для кожної частки ініціалізувати положення  $X$ , швидкість  $V$ , локальне найкраще положення  $pbest$  і глобальне найкраще положення  $gbest$  для всього рою.

Крок 4. Алгоритм нечіткої оптимізації роїв часток:

Крок 4.1. Розрахувати центри кластерів для кожної частки, використовуючи рівняння (2.12).

Крок 4.2. Обчислити значення придатності кожної частки, використовуючи рівняння (2.21).

Крок 4.3. Визначити нове  $pbest$  для кожної частки.

Крок 4.4. Обчислити нове  $gbest$  для всього рою.

Крок 4.5. Оновити матрицю швидкості для кожної частки, використовуючи рівняння (2.18).

Крок 4.6. Оновити матрицю положення для кожної частки відповідно до рівняння (2.19).

Крок 4.7. Якщо умова завершення для алгоритму нечіткої оптимізації роїв часток не виконана, поверніться до Кроку 4.

Крок 5. Алгоритм кластеризації нечітких  $c$ -середніх:

Крок 5.1. Розрахувати центри кластерів для кожної частки, використовуючи рівняння (2.12).

Крок 5.2. Обчислити відстань  $d_{ij}$  для кожної частки, використовуючи рівняння (2.11).

Крок 5.3. Оновити функцію приналежності  $\mu_{ij}$  для кожної частки за допомогою рівняння (2.13).

Крок 5.4. Оновити  $pbest$  для кожної частки.

Крок 5.5. Оновити  $gbest$  для рою.

Крок 5.6. Умова завершення: Якщо умова завершення для алгоритму кластеризації нечітких  $c$ -середніх не виконана, поверніться до Кроку 5.

Крок 6. Якщо умова завершення для гібридного алгоритму не виконана, поверніться до Кроку 4.

### 3 ЕКСПЕРИМЕНТАЛЬНІ РЕЗУЛЬТАТИ

Для оцінки результатів кластеризації застосовують функцію придатності. Окрім цього, вводиться додатковий параметр для оцінки, а саме ентропія, яка служить числовим показником рівня впорядкованості системи. Ентропія розбиття (3.1) досягає найменшого значення, коли система максимально впорядкована, що дозволяє використовувати цей показник для визначення якості кластеризації.

$$PE = - \frac{\sum_{q=1}^Q \sum_{k=1}^K u_{qk} \ln(u_{qk})}{Q}, PE \in [0, \ln K]. \quad (3.1)$$

Для оцінювання методів кластеризації — нечітких  $c$ -середніх, нечіткої оптимізації рою частинок, а також гібридного підходу, що поєднує ці два методи, було розглянуто сім реальних наборів даних.

– набір даних для рослини – ірису Фішера, складається з трьох видів квітки ірису. Для кожного виду підібрано 50 зразків з чотирма ознаками:

1. Довжина чашолистка.
2. Ширина чашолистка.
3. Довжина пелюстки.
4. Ширина пелюстки.
5. Клас:

- ірис *Setosa*;
- ірис *Versicolor*;
- ірис *Virginica*.

– набір даних центру переливанні крові, що складається з 748 записів, 4 ознак та 2 класів:

1. Давність – місяці з моменту останньої пожертви.
2. Частота – загальна кількість пожертв.
3. Гроші – загальна кількість зданої крові в кубічних центрах.

4. Час – місяці з моменту першої пожертви.

5. Клас:

– 1 = здавав кров у марті 2007;

– 2 = не здавав кров у марті 2007.

– набір даних вина, що складається з 3 типів вина, 178 предметів та 13 ознак:

1. Спирт.

2. Яблучна кислота.

3. Зола.

4. Лужність золи.

5. Магній.

6. Загальна кількість фенолів.

7. Флаваноїди.

8. Нефлаваноїдні феноли.

9. Проантоціани.

10. Інтенсивність кольору.

11. Відтінок.

12. OD280/OD315 розведених вин.

13. Пролін.

– набір даних діабету, складається з 768 предметів, 9 ознак:

1. Результат.

2. Вагітності.

3. Глюкоза.

4. Артеріальний тиск.

5. Товщина шкіри.

6. Інсулін.

7. ІМТ.

8. Цікавий діабет.

9. Вік.

– набір даних раку молочної залози Коїмбра, складається з 116 предметів, що мають 10 ознак:

1. Вік.
2. ІМТ.
3. Глюкоза.
4. Інсулін.
5. НОМА.
6. Лептин.
7. Адипонектин.
8. Резистин.
9. МСР-1.

– набір даних окулярів, які мають 3 типи та складаються з 214 предметів, 9 ознак:

1. Ідентифікаційний номер – від 1 до 214.
2. *RI* – показник заломлення.
3. *Na* – натрій.
4. *Mg* – магній.
5. *Al* – алюміній.
6. *Si* – кремній.
7. *K* – калій.
8. *Ca* – кальцій.
9. *Ba* – барій.

– набір даних місця локалізації білків, що має 336 предметів та 8 ознак:

1. Назва послідовності – номер доступу для бази даних SWISS-PROT.
2. *msg* – метод розпізнавання сигнальної послідовності МакГеоха.
3. *gvh* – метод розпізнавання сигнальної послідовності фон Хейне.
4. *lip* – консенсусна оцінка послідовності сигнальної пептидази II фон Хейне. Бінарний атрибут.

5. *chg* – наявність заряду на N-кінці передбачуваних ліпопротеїнів.  
Бінарний атрибут.

6. *aac* – оцінка дискримінантного аналізу амінокислотного вмісту зовнішньої мембрани та периплазматичних білків.

7. *alm1* – оцінка програми прогнозування мембранної області ALOM.

8. *alm2* – оцінка програми ALOM після виключення передбачуваних розщеплюваних сигнальних областей з послідовності.

Ці набори даних охоплюють приклади даних низької, середньої і високої розмірності. Ці алгоритми реалізовані з використанням Python та бібліотеки для машинного навчання Scikit-learn. Результати експериментів, отримані з понад 100 незалежних запусків методу нечітких *c*-середніх, а також з 10 незалежних запусків методу нечіткої оптимізації рою часток і гібридного методу, що поєднує нечіткі *c*-середні та нечітку оптимізацію рою часток, наведено в таблицях. Значення в таблицях відповідають значенням цільової функції (2.10).

У таблицях 3.1–3.3 наведено результати застосування методів кластеризації на основі нечітких *c*-середніх, нечіткої оптимізації рою часток та нечіткої кластеризації з використанням гібридних нечітких *c*-середніх і нечіткої оптимізації роїв часток до набору даних «Іриси».

Таблиця 3.1 - Результат використання методу кластеризації з використанням нечітких *c*-середніх на наборі даних «Іриси».

$(n, c, d)$	FCM		
	<i>найгірше</i>	<i>середнє</i>	<i>найкраще</i>
(150, 3, 4)	196,5	146,3	114,8
Ентропія	1,03	0,73	0,61

Таблиця 3.2 - Результат використання методу нечіткої оптимізації роїв часток на наборі даних «Іриси».

$(n, c, d)$	FPSO		
	<i>найгірше</i>	<i>середнє</i>	<i>найкраще</i>
(150, 3, 4)	182,9	141,4	118,6
Ентропія	0,85	0,67	0,56

Таблиця 3.3 - Результат використання методу нечіткої кластеризації з використанням гібридних нечітких  $c$ -середніх і нечіткої оптимізації роїв часток на наборі даних «Іриси».

$(n, c, d)$	FCM-FPSO		
	<i>найгірше</i>	<i>середнє</i>	<i>найкраще</i>
(150, 3, 4)	179,8	131,1	111,6
Ентропія	0,83	0,64	0,55

У таблицях 3.4–3.6 наведено результати застосування методів кластеризації на основі нечітких  $c$ -середніх, нечіткої оптимізації рою часток та нечіткої кластеризації з використанням гібридних нечітких  $c$ -середніх і нечіткої оптимізації роїв часток до набору даних «Кров».

Таблиця 3.4 - Результат використання методу кластеризації з використанням нечітких  $c$ -середніх на наборі даних «Кров».

$(n, c, d)$	FCM		
	<i>найгірше</i>	<i>середнє</i>	<i>найкраще</i>
(748, 3, 4)	993,4	976,4	951,3
Ентропія	1,07	1,02	0,94

Таблиця 3.5 - Результат використання методу нечіткої оптимізації роїв часток на наборі даних «Кров».

$(n, c, d)$	<b>FPSO</b>		
	<i>найгірше</i>	<i>середнє</i>	<i>найкраще</i>
(748, 3, 4)	1005,6	937,7	879,2
Ентропія	0,92	0,82	0,70

Таблиця 3.6 - Результат використання методу нечіткої кластеризації з використанням гібридних нечітких  $c$ -середніх і нечіткої оптимізації роїв часток на наборі даних «Кров».

$(n, c, d)$	<b>FCM-FPSO</b>		
	<i>найгірше</i>	<i>середнє</i>	<i>найкраще</i>
(748, 3, 4)	967,1	897,7	839,9
Ентропія	0,9	0,77	0,67

У таблицях 3.7–3.9 наведено результати застосування методів кластеризації на основі нечітких  $c$ -середніх, нечіткої оптимізації рою часток та нечіткої кластеризації з використанням гібридних нечітких  $c$ -середніх і нечіткої оптимізації роїв часток до набору даних «Вино».

Таблиця 3.7 - Результат використання методу кластеризації з використанням нечітких  $c$ -середніх на наборі даних «Вино».

$(n, c, d)$	<b>FCM</b>		
	<i>найгірше</i>	<i>середнє</i>	<i>найкраще</i>
(178, 3, 14)	1106,1	1068,2	1027,1
Ентропія	1,09	1,09	1,09

Таблиця 3.8 - Результат використання методу нечіткої оптимізації роїв часток на наборі даних «Вино».

$(n, c, d)$	<b>FPSO</b>		
	<i>найгірше</i>	<i>середнє</i>	<i>найкраще</i>
(178, 3, 14)	1041,4	949,9	872,2
Ентропія	1,07	1,02	0,97

Таблиця 3.9 - Результат використання методу нечіткої кластеризації з використанням гібридних нечітких  $c$ -середніх і нечіткої оптимізації роїв часток на наборі даних «Вино».

$(n, c, d)$	<b>FCM-FPSO</b>		
	<i>найгірше</i>	<i>середнє</i>	<i>найкраще</i>
(178, 3, 14)	949,2	895,6	856,9
Ентропія	1,07	1,01	0,88

У таблицях 3.10–3.12 наведено результати застосування методів кластеризації на основі нечітких  $c$ -середніх, нечіткої оптимізації рою часток та нечіткої кластеризації з використанням гібридних нечітких  $c$ -середніх і нечіткої оптимізації роїв часток до набору даних «Діабет».

Таблиця 3.10 - Результат використання методу кластеризації з використанням нечітких  $c$ -середніх на наборі даних «Діабет».

$(n, c, d)$	<b>FCM</b>		
	<i>найгірше</i>	<i>середнє</i>	<i>найкраще</i>
(768, 3, 8)	2700	2640,3	2595,7
Ентропія	1,1	1,09	1,09

Таблиця 3.11 - Результат використання методу нечіткої оптимізації роїв часток на наборі даних «Діабет».

$(n, c, d)$	FPSO		
	<i>найгірше</i>	<i>середнє</i>	<i>найкраще</i>
(768, 3, 8)	2478,9	2321,9	2159,9
Ентропія	1,07	1,04	1,01

Таблиця 3.12 - Результат використання методу нечіткої кластеризації з використанням гібридних нечітких  $c$ -середніх і нечіткої оптимізації роїв часток на наборі даних «Діабет».

$(n, c, d)$	FCM-FPSO		
	<i>найгірше</i>	<i>середнє</i>	<i>найкраще</i>
(768, 3, 8)	2396,5	2270,4	2118,2
Ентропія	1,07	1,05	0,99

У таблицях 3.13–3.15 наведено результати застосування методів кластеризації на основі нечітких  $c$ -середніх, нечіткої оптимізації рою часток та нечіткої кластеризації з використанням гібридних нечітких  $c$ -середніх і нечіткої оптимізації роїв часток до набору даних «Рак».

Таблиця 3.13 - Результат використання методу кластеризації з використанням нечітких  $c$ -середніх на наборі даних «Рак».

$(n, c, d)$	FCM		
	<i>найгірше</i>	<i>середнє</i>	<i>найкраще</i>
(116, 3, 9)	488	448,7	420,2
Ентропія	1,1	1,09	1,08

Таблиця 3.14 - Результат використання методу нечіткої оптимізації роїв часток на наборі даних «Рак».

$(n, c, d)$	<b>FPSO</b>		
	<i>найгірше</i>	<i>середнє</i>	<i>найкраще</i>
(116, 3, 9)	458,8	393,8	370,7
Ентропія	1,07	1,03	0,99

Таблиця 3.15 - Результат використання методу нечіткої кластеризації з використанням гібридних нечітких  $c$ -середніх і нечіткої оптимізації роїв часток на наборі даних «Рак».

$(n, c, d)$	<b>FCM-FPSO</b>		
	<i>найгірше</i>	<i>середнє</i>	<i>найкраще</i>
(116, 3, 9)	396,2	378,6	363,6
Ентропія	1,08	1,04	0,98

У таблицях 3.16–3.18 наведено результати застосування методів кластеризації на основі нечітких  $c$ -середніх, нечіткої оптимізації рою часток та нечіткої кластеризації з використанням гібридних нечітких  $c$ -середніх і нечіткої оптимізації роїв часток до набору даних «Окуляри».

Таблиця 3.16 - Результат використання методу кластеризації з використанням нечітких  $c$ -середніх на наборі даних «Окуляри».

$(n, c, d)$	<b>FCM</b>		
	<i>найгірше</i>	<i>середнє</i>	<i>найкраще</i>
(214, 3, 9)	870,2	821	778
Ентропія	1,09	1,09	1,08

Таблиця 3.17 - Результат використання методу нечіткої оптимізації роїв часток на наборі даних «Окуляри».

$(n, c, d)$	<b>FPSO</b>		
	<i>найгірше</i>	<i>середнє</i>	<i>найкраще</i>
(214, 3, 9)	751	694,7	659,3
Ентропія	1,05	0,98	0,87

Таблиця 3.18 - Результат використання методу нечіткої кластеризації з використанням гібридних нечітких  $c$ -середніх і нечіткої оптимізації роїв часток на наборі даних «Окуляри».

$(n, c, d)$	<b>FCM-FPSO</b>		
	<i>найгірше</i>	<i>середнє</i>	<i>найкраще</i>
(214, 3, 9)	717,3	674,4	647,4
Ентропія	1,05	0,96	0,85

У таблицях 3.19–3.21 наведено результати застосування методів кластеризації на основі нечітких  $c$ -середніх, нечіткої оптимізації рою часток та нечіткої кластеризації з використанням гібридних нечітких  $c$ -середніх і нечіткої оптимізації роїв часток до набору даних «Білки».

Таблиця 3.19 - Результат використання методу кластеризації з використанням нечітких  $c$ -середніх на наборі даних «Білки».

$(n, c, d)$	<b>FCM</b>		
	<i>найгірше</i>	<i>середнє</i>	<i>найкраще</i>
(336, 3, 7)	1075,9	1006,9	960
Ентропія	1,1	1,09	1,07

Таблиця 3.20 - Результат використання методу нечіткої оптимізації роїв часток на наборі даних «Білки».

$(n, c, d)$	<b>FPSO</b>		
	<i>найгірше</i>	<i>середнє</i>	<i>найкраще</i>
(336, 3, 7)	851,8	804,5	751,8
Ентропія	0,98	0,91	0,82

Таблиця 3.21 - Результат використання методу нечіткої кластеризації з використанням гібридних нечітких  $c$ -середніх і нечіткої оптимізації роїв часток на наборі даних «Білки».

$(n, c, d)$	<b>FCM-FPSO</b>		
	<i>найгірше</i>	<i>середнє</i>	<i>найкраще</i>
(336, 3, 7)	822,2	782,8	750,2
Ентропія	1,02	0,91	0,78

Для підвищення ефективності роботи алгоритмів методів кластеризації на основі нечітких  $c$ -середніх, нечіткої оптимізації рою часток та нечіткої кластеризації з використання гібридних нечітких  $c$ -середніх і нечіткої оптимізації роїв часток було виконано налаштування параметрів з метою вибору оптимальних значень. За результатами експериментів встановлено, що найкращі результати досягаються при таких налаштуваннях:  $P = 10$ ,  $c_1 = 2$ ,  $c_2 = 2$ ,  $w = 0.9$ .

Умови завершення для кожного алгоритму такі:

- у алгоритмі кластеризації з використанням нечітких  $c$ -середніх зупинка відбувається, коли подальше покращення рішення неможливе.
- у алгоритмі нечіткої оптимізації роїв часток зупинка відбувається, коли найкраще глобальне рішення  $gbest$  не покращується протягом 1000 послідовних ітерацій.

– у алгоритмі нечіткої кластеризації з використання гібридних нечітких  $c$ -середніх і нечіткої оптимізації роїв часток: FPSO зупиняється при досягненні 1000 ітерацій або якщо  $gbest$  не змінюється протягом 200 ітерацій, FCM зупиняється після 5 ітерацій; додатково, уся гібридна процедура припиняється, якщо  $gbest$  не покращується протягом 2 послідовних ітерацій.

У всіх алгоритмах ступінь нечіткості  $m$  було встановлено на рівні 2.

Як видно з таблиць результатів, гібридний метод нечітких  $c$ -середніх і нечіткої оптимізації роїв часток демонструє кращу якість кластеризації на всіх наборах даних, зокрема завдяки здатності уникати локальних мінімумів. Крім того, експерименти показали, що при невеликих обсягах даних метод нечіткої оптимізації роїв часток перевершує метод кластеризації нечітких  $c$ -середніх.

Додатково для тестування різної кількості кластерів візьмемо ще чотири набори даних:

– набір даних болтів, що складається з 40 предметів та 7 ознак:

1. Швидкість обертання 1.
2. Загальна кількість болтів.
3. Швидкість обертання 2.
4. Чутливість.
5. Кількість болтів на другій швидкості.
6. Час.
7. Порядок запису.

– набір даних людей з хворобою вовчаковий нефрит, складається з 87 предметів та 3 ознак:

1. Час.
2. Тривалість.
3. Логарифм 1+Тривалість.

– набір даних тканин молочної залози, що складається з 106 предметів та 9 ознак;

– набір даних виноградника, складається з 52 предметів та 2 ознак.

У таблицях 3.22–3.24 наведено результати застосування методів кластеризації на основі нечітких  $c$ -середніх, нечіткої оптимізації рою часток та нечіткої кластеризації з використанням гібридних нечітких  $c$ -середніх і нечіткої оптимізації роїв часток до набору даних «Болти» при розбитті на 2 кластери.

Таблиця 3.22 - Результат використання методу кластеризації з використанням нечітких  $c$ -середніх на наборі даних «Болти» при розбитті на 2 кластери.

$(n, c, d)$	FCM		
	<i>найгірше</i>	<i>середнє</i>	<i>найкраще</i>
(40, 2, 7)	176,8	167,5	156,2
Ентропія	0,69	0,68	0,67

Таблиця 3.23 - Результат використання методу нечіткої оптимізації роїв часток на наборі даних «Болти» при розбитті на 2 кластери.

$(n, c, d)$	FPSO		
	<i>найгірше</i>	<i>середнє</i>	<i>найкраще</i>
(40, 2, 7)	180,9	162,1	149,3
Ентропія	0,67	0,62	0,58

Таблиця 3.24 - Результат використання методу нечіткої кластеризації з використанням гібридних нечітких  $c$ -середніх і нечіткої оптимізації роїв часток на наборі даних «Болти» при розбитті на 2 кластери.

$(n, c, d)$	FCM-FPSO		
	<i>найгірше</i>	<i>середнє</i>	<i>найкраще</i>
(40, 2, 7)	169,3	156,5	148,8
Ентропія	0,67	0,63	0,55

У таблицях 3.25–3.27 наведено результати застосування методів кластеризації на основі нечітких  $c$ -середніх, нечіткої оптимізації рою часток та нечіткої кластеризації з використанням гібридних нечітких  $c$ -середніх і нечіткої оптимізації роїв часток до набору даних «Болти» при розбитті на 3 кластери.

Таблиця 3.25 - Результат використання методу кластеризації з використанням нечітких  $c$ -середніх на наборі даних «Болти» при розбитті на 3 кластери.

$(n, c, d)$	<b>FCM</b>		
	<i>найгірше</i>	<i>середнє</i>	<i>найкраще</i>
(40, 3, 7)	130,2	117,6	109
Ентропія	1,09	1,08	1,05

Таблиця 3.26 - Результат використання методу нечіткої оптимізації роїв часток на наборі даних «Болти» при розбитті на 3 кластери.

$(n, c, d)$	<b>FPSO</b>		
	<i>найгірше</i>	<i>середнє</i>	<i>найкраще</i>
(40, 3, 7)	115	106,2	101,5
Ентропія	1,06	1,01	0,94

Таблиця 3.27 - Результат використання методу нечіткої кластеризації з використанням гібридних нечітких  $c$ -середніх і нечіткої оптимізації роїв часток на наборі даних «Болти» при розбитті на 3 кластери.

$(n, c, d)$	<b>FCM-FPSO</b>		
	<i>найгірше</i>	<i>середнє</i>	<i>найкраще</i>
(40, 3, 7)	109,6	102,8	98,52
Ентропія	1,08	1,01	0,88

У таблицях 3.28–3.30 наведено результати застосування методів кластеризації на основі нечітких  $c$ -середніх, нечіткої оптимізації рою часток та нечіткої кластеризації з використанням гібридних нечітких  $c$ -середніх і нечіткої оптимізації роїв часток до набору даних «Болти» при розбитті на 4 кластери.

Таблиця 3.28 - Результат використання методу кластеризації з використанням нечітких  $c$ -середніх на наборі даних «Болти» при розбитті на 4 кластери.

$(n, c, d)$	FCM		
	<i>найгірше</i>	<i>середнє</i>	<i>найкраще</i>
(40, 4, 7)	94,3	89,7	84,5
Ентропія	1,38	1,37	1,36

Таблиця 3.29 - Результат використання методу нечіткої оптимізації роїв часток на наборі даних «Болти» при розбитті на 4 кластери.

$(n, c, d)$	FPSO		
	<i>найгірше</i>	<i>середнє</i>	<i>найкраще</i>
(40, 4, 7)	86,2	79,4	75,1
Ентропія	1,34	1,28	1,19

Таблиця 3.30 - Результат використання методу нечіткої кластеризації з використанням гібридних нечітких  $c$ -середніх і нечіткої оптимізації роїв часток на наборі даних «Болти» при розбитті на 4 кластери.

$(n, c, d)$	FCM-FPSO		
	<i>найгірше</i>	<i>середнє</i>	<i>найкраще</i>
(40, 4, 7)	86,1	76,9	72,7
Ентропія	1,34	1,27	1,13

У таблицях 3.31–3.33 наведено результати застосування методів кластеризації на основі нечітких  $c$ -середніх, нечіткої оптимізації рою часток та нечіткої кластеризації з використанням гібридних нечітких  $c$ -середніх і нечіткої оптимізації роїв часток до набору даних «Болти» при розбитті на 5 кластерів.

Таблиця 3.31 - Результат використання методу кластеризації з використанням нечітких  $c$ -середніх на наборі даних «Болти» при розбитті на 5 кластерів.

$(n, c, d)$	FCM		
	<i>найгірше</i>	<i>середнє</i>	<i>найкраще</i>
(40, 5, 7)	77,5	71,8	68,4
Ентропія	1,60	1,59	1,58

Таблиця 3.32 - Результат використання методу нечіткої оптимізації роїв часток на наборі даних «Болти» при розбитті на 5 кластерів.

$(n, c, d)$	FPSO		
	<i>найгірше</i>	<i>середнє</i>	<i>найкраще</i>
(40, 5, 7)	67,8	62,2	58,2
Ентропія	1,55	1,47	1,32

Таблиця 3.33 - Результат використання методу нечіткої кластеризації з використанням гібридних нечітких  $c$ -середніх і нечіткої оптимізації роїв часток на наборі даних «Болти» при розбитті на 5 кластерів.

$(n, c, d)$	FCM-FPSO		
	<i>найгірше</i>	<i>середнє</i>	<i>найкраще</i>
(40, 5, 7)	63,4	61,2	58
Ентропія	1,57	1,48	1,29

У таблицях 3.34–3.36 наведено результати застосування методів кластеризації на основі нечітких  $c$ -середніх, нечіткої оптимізації рою часток та нечіткої кластеризації з використанням гібридних нечітких  $c$ -середніх і нечіткої оптимізації роїв часток до набору даних «Нефрит» при розбитті на 2 кластери.

Таблиця 3.34 - Результат використання методу кластеризації з використанням нечітких  $c$ -середніх на наборі даних «Нефрит» при розбитті на 2 кластери.

$(n, c, d)$	FCM		
	<i>найгірше</i>	<i>середнє</i>	<i>найкраще</i>
(87, 2, 3)	129,9	123,1	116,6
Ентропія	0,67	0,53	0,45

Таблиця 3.35 - Результат використання методу нечіткої оптимізації роїв часток на наборі даних «Нефрит» при розбитті на 2 кластери.

$(n, c, d)$	FPSO		
	<i>найгірше</i>	<i>середнє</i>	<i>найкраще</i>
(87, 2, 3)	137,2	127,3	119,3
Ентропія	0,65	0,47	0,37

Таблиця 3.36 - Результат використання методу нечіткої кластеризації з використанням гібридних нечітких  $c$ -середніх і нечіткої оптимізації роїв часток на наборі даних «Нефрит» при розбитті на 2 кластери.

$(n, c, d)$	FCM-FPSO		
	<i>найгірше</i>	<i>середнє</i>	<i>найкраще</i>
(87, 2, 3)	127,7	119,9	115,4
Ентропія	0,55	0,44	0,34

У таблицях 3.37–3.39 наведено результати застосування методів кластеризації на основі нечітких  $c$ -середніх, нечіткої оптимізації рою часток та нечіткої кластеризації з використанням гібридних нечітких  $c$ -середніх і нечіткої оптимізації роїв часток до набору даних «Нефрит» при розбитті на 3 кластери.

Таблиця 3.37 - Результат використання методу кластеризації з використанням нечітких  $c$ -середніх на наборі даних «Нефрит» при розбитті на 3 кластери.

$(n, c, d)$	FCM		
	<i>найгірше</i>	<i>середнє</i>	<i>найкраще</i>
(87, 3, 3)	85,6	80,6	76,4
Ентропія	1,02	0,88	0,79

Таблиця 3.38 - Результат використання методу нечіткої оптимізації роїв часток на наборі даних «Нефрит» при розбитті на 3 кластери.

$(n, c, d)$	FPSO		
	<i>найгірше</i>	<i>середнє</i>	<i>найкраще</i>
(87, 3, 3)	83,4	79	74,7
Ентропія	0,89	0,75	0,63

Таблиця 3.39 - Результат використання методу нечіткої кластеризації з використанням гібридних нечітких  $c$ -середніх і нечіткої оптимізації роїв часток на наборі даних «Нефрит» при розбитті на 3 кластери.

$(n, c, d)$	FCM-FPSO		
	<i>найгірше</i>	<i>середнє</i>	<i>найкраще</i>
(87, 3, 3)	78,8	75,2	71,2
Ентропія	0,78	0,69	0,57

У таблицях 3.40–3.42 наведено результати застосування методів кластеризації на основі нечітких  $c$ -середніх, нечіткої оптимізації рою часток та нечіткої кластеризації з використанням гібридних нечітких  $c$ -середніх і нечіткої оптимізації роїв часток до набору даних «Нефрит» при розбитті на 4 кластери.

Таблиця 3.40 - Результат використання методу кластеризації з використанням нечітких  $c$ -середніх на наборі даних «Нефрит» при розбитті на 4 кластери.

$(n, c, d)$	FCM		
	<i>найгірше</i>	<i>середнє</i>	<i>найкраще</i>
(87, 4, 3)	63,5	59,7	55,9
Ентропія	1,27	1,12	0,96

Таблиця 3.41 - Результат використання методу нечіткої оптимізації роїв часток на наборі даних «Нефрит» при розбитті на 4 кластери.

$(n, c, d)$	FPSO		
	<i>найгірше</i>	<i>середнє</i>	<i>найкраще</i>
(87, 4, 3)	61,7	56,3	51,6
Ентропія	1,07	0,91	0,82

Таблиця 3.42 - Результат використання методу нечіткої кластеризації з використанням гібридних нечітких  $c$ -середніх і нечіткої оптимізації роїв часток на наборі даних «Нефрит» при розбитті на 4 кластери.

$(n, c, d)$	FCM-FPSO		
	<i>найгірше</i>	<i>середнє</i>	<i>найкраще</i>
(87, 4, 3)	57,8	52,7	48,4
Ентропія	1,01	0,89	0,78

У таблицях 3.43–3.45 наведено результати застосування методів кластеризації на основі нечітких  $c$ -середніх, нечіткої оптимізації рою часток та нечіткої кластеризації з використанням гібридних нечітких  $c$ -середніх і нечіткої оптимізації роїв часток до набору даних «Нефрит» при розбитті на 5 кластери.

Таблиця 3.43 - Результат використання методу кластеризації з використанням нечітких  $c$ -середніх на наборі даних «Нефрит» при розбитті на 5 кластерів.

$(n, c, d)$	FCM		
	<i>найгірше</i>	<i>середнє</i>	<i>найкраще</i>
(87, 5, 3)	50,5	47,7	44,7
Ентропія	1,46	1,32	1,17

Таблиця 3.44 - Результат використання методу нечіткої оптимізації роїв часток на наборі даних «Нефрит» при розбитті на 5 кластерів.

$(n, c, d)$	FPSO		
	<i>найгірше</i>	<i>середнє</i>	<i>найкраще</i>
(87, 5, 3)	47,2	42,1	38,3
Ентропія	1,31	1,07	0,94

Таблиця 3.45 - Результат використання методу нечіткої кластеризації з використанням гібридних нечітких  $c$ -середніх і нечіткої оптимізації роїв часток на наборі даних «Нефрит» при розбитті на 5 кластерів.

$(n, c, d)$	FCM-FPSO		
	<i>найгірше</i>	<i>середнє</i>	<i>найкраще</i>
(87, 5, 3)	48	40,5	35,8
Ентропія	1,27	1,05	0,84

У таблицях 3.46–3.48 наведено результати застосування методів кластеризації на основі нечітких  $c$ -середніх, нечіткої оптимізації рою часток та нечіткої кластеризації з використанням гібридних нечітких  $c$ -середніх і нечіткої оптимізації роїв часток до набору даних «Залози» при розбитті на 2 кластери.

Таблиця 3.46 - Результат використання методу кластеризації з використанням нечітких  $c$ -середніх на наборі даних «Залози» при розбитті на 2 кластери.

$(n, c, d)$	FCM		
	<i>найгірше</i>	<i>середнє</i>	<i>найкраще</i>
(106, 2, 9)	617,6	579,6	556,9
Ентропія	0,69	0,69	0,68

Таблиця 3.47 - Результат використання методу нечіткої оптимізації роїв часток на наборі даних «Залози» при розбитті на 2 кластери.

$(n, c, d)$	FPSO		
	<i>найгірше</i>	<i>середнє</i>	<i>найкраще</i>
(106, 2, 9)	606,3	571,9	530,5
Ентропія	0,68	0,66	0,64

Таблиця 3.48 - Результат використання методу нечіткої кластеризації з використанням гібридних нечітких  $c$ -середніх і нечіткої оптимізації роїв часток на наборі даних «Залози» при розбитті на 2 кластери.

$(n, c, d)$	FCM-FPSO		
	<i>найгірше</i>	<i>середнє</i>	<i>найкраще</i>
(106, 2, 9)	603,4	544,5	513,4
Ентропія	0,68	0,66	0,61

У таблицях 3.49–3.51 наведено результати застосування методів кластеризації на основі нечітких  $c$ -середніх, нечіткої оптимізації рою часток та нечіткої кластеризації з використанням гібридних нечітких  $c$ -середніх і нечіткої оптимізації роїв часток до набору даних «Залози» при розбитті на 3 кластери.

Таблиця 3.49 - Результат використання методу кластеризації з використанням нечітких  $c$ -середніх на наборі даних «Залози» при розбитті на 3 кластери.

$(n, c, d)$	FCM		
	<i>найгірше</i>	<i>середнє</i>	<i>найкраще</i>
(106, 3, 9)	416,5	404,6	384,1
Ентропія	1,09	1,09	1,08

Таблиця 3.50 - Результат використання методу нечіткої оптимізації роїв часток на наборі даних «Залози» при розбитті на 3 кластери.

$(n, c, d)$	FPSO		
	<i>найгірше</i>	<i>середнє</i>	<i>найкраще</i>
(106, 3, 9)	422,6	386,5	350,6
Ентропія	1,07	1,05	1,04

Таблиця 3.51 - Результат використання методу нечіткої кластеризації з використанням гібридних нечітких  $c$ -середніх і нечіткої оптимізації роїв часток на наборі даних «Залози» при розбитті на 3 кластери.

$(n, c, d)$	FCM-FPSO		
	<i>найгірше</i>	<i>середнє</i>	<i>найкраще</i>
(106, 3, 9)	385	366,4	345,2
Ентропія	1,07	1,05	1,03

У таблицях 3.52–3.54 наведено результати застосування методів кластеризації на основі нечітких  $c$ -середніх, нечіткої оптимізації рою часток та нечіткої кластеризації з використанням гібридних нечітких  $c$ -середніх і нечіткої оптимізації роїв часток до набору даних «Залози» при розбитті на 3 кластери.

Таблиця 3.52 - Результат використання методу кластеризації з використанням нечітких  $c$ -середніх на наборі даних «Залози» при розбитті на 4 кластери.

$(n, c, d)$	FCM		
	<i>найгірше</i>	<i>середнє</i>	<i>найкраще</i>
(106, 4, 9)	317,9	307,7	294,8
Ентропія	1,39	1,38	1,38

Таблиця 3.53 - Результат використання методу нечіткої оптимізації роїв часток на наборі даних «Залози» при розбитті на 4 кластери.

$(n, c, d)$	FPSO		
	<i>найгірше</i>	<i>середнє</i>	<i>найкраще</i>
(106, 4, 9)	304,5	283,8	267,8
Ентропія	1,35	1,33	1,30

Таблиця 3.54 - Результат використання методу нечіткої кластеризації з використанням гібридних нечітких  $c$ -середніх і нечіткої оптимізації роїв часток на наборі даних «Залози» при розбитті на 4 кластери.

$(n, c, d)$	FCM-FPSO		
	<i>найгірше</i>	<i>середнє</i>	<i>найкраще</i>
(106, 4, 9)	298,9	273,3	259
Ентропія	1,36	1,33	1,29

У таблицях 3.55–3.57 наведено результати застосування методів кластеризації на основі нечітких  $c$ -середніх, нечіткої оптимізації рою часток та нечіткої кластеризації з використанням гібридних нечітких  $c$ -середніх і нечіткої оптимізації роїв часток до набору даних «Виноградник» при розбитті на 2 кластери.

Таблиця 3.55 - Результат використання методу кластеризації з використанням нечітких  $c$ -середніх на наборі даних «Виноградник» при розбитті на 2 кластери.

$(n, c, d)$	FCM		
	<i>найгірше</i>	<i>середнє</i>	<i>найкраще</i>
(52, 2, 2)	51,9	43,1	33,1
Ентропія	0,69	0,45	0,32

Таблиця 3.56 - Результат використання методу нечіткої оптимізації роїв часток на наборі даних «Виноградник» при розбитті на 2 кластери.

$(n, c, d)$	FPSO		
	<i>найгірше</i>	<i>середнє</i>	<i>найкраще</i>
(52, 2, 2)	40,8	35,8	31,8
Ентропія	0,39	0,34	0,31

Таблиця 3.57 - Результат використання методу нечіткої кластеризації з використанням гібридних нечітких  $c$ -середніх і нечіткої оптимізації роїв часток на наборі даних «Виноградник» при розбитті на 2 кластери.

$(n, c, d)$	FCM-FPSO		
	<i>найгірше</i>	<i>середнє</i>	<i>найкраще</i>
(52, 2, 2)	36,9	33,3	31,7
Ентропія	0,37	0,32	0,30

У таблицях 3.58–3.60 наведено результати застосування методів кластеризації на основі нечітких  $c$ -середніх, нечіткої оптимізації рою часток та нечіткої кластеризації з використанням гібридних нечітких  $c$ -середніх і нечіткої оптимізації роїв часток до набору даних «Виноградник» при розбитті на 3 кластери.

Таблиця 3.58 - Результат використання методу кластеризації з використанням нечітких  $c$ -середніх на наборі даних «Виноградник» при розбитті на 3 кластери.

$(n, c, d)$	FCM		
	<i>найгірше</i>	<i>середнє</i>	<i>найкраще</i>
(52, 3, 2)	34,3	27,8	22,4
Ентропія	1,05	0,7	0,60

Таблиця 3.59 - Результат використання методу нечіткої оптимізації роїв часток на наборі даних «Виноградник» при розбитті на 3 кластери.

$(n, c, d)$	FPSO		
	<i>найгірше</i>	<i>середнє</i>	<i>найкраще</i>
(52, 3, 2)	24,6	22,3	19,3
Ентропія	0,74	0,64	0,57

Таблиця 3.60 - Результат використання методу нечіткої кластеризації з використанням гібридних нечітких  $c$ -середніх і нечіткої оптимізації роїв часток на наборі даних «Виноградник» при розбитті на 3 кластери.

$(n, c, d)$	FCM-FPSO		
	<i>найгірше</i>	<i>середнє</i>	<i>найкраще</i>
(52, 3, 2)	22,3	20,8	19,1
Ентропія	0,65	0,60	0,56

У таблицях 3.61–3.63 наведено результати застосування методів кластеризації на основі нечітких  $c$ -середніх, нечіткої оптимізації рою часток та нечіткої кластеризації з використанням гібридних нечітких  $c$ -середніх і нечіткої оптимізації роїв часток до набору даних «Виноградник» при розбитті на 4 кластери.

Таблиця 3.61 - Результат використання методу кластеризації з використанням нечітких  $c$ -середніх на наборі даних «Виноградник» при розбитті на 4 кластери.

$(n, c, d)$	FCM		
	<i>найгірше</i>	<i>середнє</i>	<i>найкраще</i>
(52, 4, 2)	24,9	20,5	17,2
Ентропія	1,23	0,98	0,87

Таблиця 3.62 - Результат використання методу нечіткої оптимізації роїв часток на наборі даних «Виноградник» при розбитті на 4 кластери.

$(n, c, d)$	FPSO		
	<i>найгірше</i>	<i>середнє</i>	<i>найкраще</i>
(52, 4, 2)	17,7	15,7	13,4
Ентропія	0,92	0,82	0,77

Таблиця 3.63 - Результат використання методу нечіткої кластеризації з використанням гібридних нечітких  $c$ -середніх і нечіткої оптимізації роїв часток на наборі даних «Виноградник» при розбитті на 4 кластери.

$(n, c, d)$	FCM-FPSO		
	<i>найгірше</i>	<i>середнє</i>	<i>найкраще</i>
(52, 4, 2)	17,5	15,2	13,6
Ентропія	0,90	0,81	0,73

У таблицях 3.64–3.66 наведено результати застосування методів кластеризації на основі нечітких  $c$ -середніх, нечіткої оптимізації рою часток та нечіткої кластеризації з використанням гібридних нечітких  $c$ -середніх і нечіткої оптимізації роїв часток до набору даних «Виноградник» при розбитті на 5 кластерів.

Таблиця 3.64 - Результат використання методу кластеризації з використанням нечітких  $c$ -середніх на наборі даних «Виноградник» при розбитті на 5 кластерів.

$(n, c, d)$	FCM		
	<i>найгірше</i>	<i>середнє</i>	<i>найкраще</i>
(52, 5, 2)	20	16,4	13,4
Ентропія	1,45	1,18	1,03

Таблиця 3.65 - Результат використання методу нечіткої оптимізації роїв часток на наборі даних «Виноградник» при розбитті на 5 кластерів.

$(n, c, d)$	FPSO		
	<i>найгірше</i>	<i>середнє</i>	<i>найкраще</i>
(52, 5, 2)	13,6	12,1	11
Ентропія	1,02	0,94	0,85

Таблиця 3.66 - Результат використання методу нечіткої кластеризації з використанням гібридних нечітких  $c$ -середніх і нечіткої оптимізації роїв часток на наборі даних «Виноградник» при розбитті на 5 кластерів.

$(n, c, d)$	FCM-FPSO		
	<i>найгірше</i>	<i>середнє</i>	<i>найкраще</i>
(52, 5, 2)	13,6	11,4	10,2
Ентропія	1,02	0,91	0,81

## ВИСНОВКИ

Нечітка кластеризація є важливою проблемою, яка активно досліджується в різних наукових сферах.

У цій роботі було проаналізовано три різні моделі нечіткої кластеризації, а також виявлено їхні сильні та слабкі сторони.

Для створення та моделювання ефективного гібридного методу нечіткої кластеризації, що поєднує метод кластеризації з використанням нечітких *c*-середніх та метод нечіткої оптимізації роїв часток, було досліджено три алгоритми кластеризації.

Одним із найпоширеніших методів є нечіткий алгоритм *c*-середніх (FCM). Однак він залежить від початкових умов і часто застрягає в локальних оптимумах. Для вирішення подібних оптимізаційних задач часто використовується метод оптимізації роїв частинок (PSO) – глобальний стохастичний інструмент, який можна застосовувати для різноманітних оптимізаційних проблем.

У цій роботі для усунення недоліків FCM було запропоновано його інтеграцію з нечіткою PSO (FPSO). Таким чином, створено гібридний метод, який поєднує переваги обох підходів.

Експерименти на одинадцяти відомих наборах даних (іриса, кров, вино, діабет, рак, окуляри, білок, болти, нефрит, залози, виноградник) підтвердили ефективність запропонованого гібридного алгоритму, показавши високу якість кластеризації.

Отже, розроблено новий підхід, який вдосконалює алгоритм нечітких *c*-середніх за рахунок інтеграції з PSO. Численні експерименти та аналіз демонструють, що запропонований метод значно підвищує продуктивність FCM.

**ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ**

1. Jain, A. K., Murty, M. N., & Flynn, P. J. (1999). Data clustering: a review. *ACM computing surveys (CSUR)*, 31(3), 264-323.
2. Bezdek, J. C. (2013). *Pattern recognition with fuzzy objective function algorithms*. Springer Science & Business Media.
3. Tvoroshenko, I., Gorokhovatskyi, V., Kobylin, O., & Tvoroshenko, A. (2023). Application of deep learning methods for recognizing and classifying culinary dishes in images.
4. Касумов, А. І. (2023). Розробка застосунку для кластеризації часових рядів у вигляді медичних даних.
5. Кобилін, О. А., & Творошенко, І. С. (2021). Методи цифрової обробки зображень.
6. Hathaway, R. J., & Bezdek, J. C. (2001). Fuzzy c-means clustering of incomplete data. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 31(5), 735-744.
7. Ейхман, Т. І. (2021). Метод виявлення програм-вимагачів з використанням машинного навчання.
8. Коткова, А. А. (2019). Комплекс задач кластеризації даних місцевих фінансів.
9. Герасимова, С. С., & Шафроненко, А. Ю. (2024, April). РОЗРОБКА ЗАСТОСУНКУ ДЛЯ ОНЛАЙН РОЗПІЗНАВАННЯ ОБЛИЧ У ВІДЕОПОТОЦІ. In *The 5 th International scientific and practical conference "Science and society: modern trends in a changing world" (April 15-17, 2024)* MDPC Publishing, Vienna, Austria. 2024. 492 p (p. 151).
10. Ran, X., Xi, Y., Lu, Y., Wang, X., & Lu, Z. (2023). Comprehensive survey on hierarchical clustering algorithms and the recent developments. *Artificial Intelligence Review*, 56(8), 8219-8264.
11. Maulik, U., & Bandyopadhyay, S. (2000). Genetic algorithm-based clustering technique. *Pattern recognition*, 33(9), 1455-1465.

12. Wan, H., Wang, H., Scotney, B., & Liu, J. (2019, October). A novel Gaussian mixture model for classification. In *2019 IEEE International Conference on Systems, Man and Cybernetics (SMC)* (pp. 3298-3303). IEEE.
13. Tareq, M., Sundararajan, E. A., Harwood, A., & Bakar, A. A. (2021). A systematic review of density grid-based clustering for data streams. *Ieee Access*, *10*, 579-596.
14. Tang, C., Li, Z., Wang, J., Liu, X., Zhang, W., & Zhu, E. (2022). Unified one-step multi-view spectral clustering. *IEEE Transactions on Knowledge and Data Engineering*, *35*(6), 6449-6460.
15. Бодяньський, С. В., Шафроненко, А. Ю., & Климова, І. М. (2019). Онлайн достовірна нечітка кластеризація даних з використанням функції належності спеціального типу. *Біоніка інтелекту*, *2*(93), 3-6.
16. Bodyanskiy, Y., Shafronenko, A., Klymova, I., & Polyvoda, V. (2021, May). Robust recurrent credibilistic modification of the Gustafson-Kessel algorithm. In *International Scientific Conference "Intellectual Systems of Decision Making and Problem of Computational Intelligence"* (pp. 613-623). Cham: Springer International Publishing.
17. Bodyanskiy, Y. V., Shafronenko, A., & Klymova, I. (2021, April). Adaptive Recovery of Distorted Data Based on Credibilistic Fuzzy Clustering Approach. In *COLINS* (pp. 6-15).
18. Shafronenko, A., Bodyanskiy, Y. V., & Pliss, I. (2023). Credibilistic Fuzzy Clustering Method Based on Evolutionary Approach of Crazy Wolves in Online Mode. In *CMIS* (pp. 141-150).
19. Shafronenko, A., Bodyanskiy, Y., Pliss, I., & Patlan, K. (2019, June). Fuzzy Clusterization of Distorted by Missing Observations Data Sets Using Evolutionary Optimization. In *2019 9th International Conference on Advanced Computer Information Technologies (ACIT)* (pp. 217-220). IEEE.
20. Ясько, О. С. (2024). *Генетичні алгоритми для вирішення транспортних задач з використанням технологій штучного інтелекту* (Doctoral dissertation, ХНУРЕ).

21. Шафроненко, А. Ю. (2023). Адаптивні методи нечіткої кластеризації потоків даних з використанням еволюційного самонавчання.
22. Ясько, О. С. (2023). Алгоритм мурашиної колонії для вирішення транспортних задач із виходом з локальних мінімумів.
23. Авлякулов, Т. Е. (2023). Аналіз методів кластеризації на основі щільності розподілу даних.
24. Алевська, А. І. (2024). Дослідження методів кластеризації даних у задачах прийняття рішень.
25. Bodyanskiy, Y., Pliss, I., & Shafronenko, A. (2022). Adaptive neurofuzzy clustering of distorted data based on prototype-centroid strategy using evolutionary procedures. *Artificial Intelligence*, 27, 239-244.
26. Патлань, К. В. (2022). Використання та дослідження методів глибинного навчання при вирішенні задач динамічного аналізу даних.
27. Kennedy, J., & Eberhart, R. (1995, November). Particle swarm optimization. In *Proceedings of ICNN'95-international conference on neural networks* (Vol. 4, pp. 1942-1948). ieee.
28. Pang, W., Wang, K. P., Zhou, C. G., & Dong, L. J. (2004, September). Fuzzy discrete particle swarm optimization for solving traveling salesman problem. In *The Fourth International Conference on Computer and Information Technology, 2004. CIT'04.* (pp. 796-800). IEEE.
29. Praseeda, C. K., & Shivakumar, B. L. (2021). Fuzzy particle swarm optimization (FPSO) based feature selection and hybrid kernel distance based possibilistic fuzzy local information C-means (HKD-PFLICM) clustering for churn prediction in telecom industry. *SN Applied Sciences*, 3, 1-18.
30. Yadav, V., & Gupta, I. K. (2022). Modified adaptive inertia weight particle swarm optimisation for data clustering. *International Journal of Innovative Computing and Applications*, 13(1), 34-40.