

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
БУДІВНИЦТВА І АРХІТЕКТУРИ

Кваліфікаційна наукова
праця на правах рукопису

БИКОВ СЕРГІЙ ВОЛОДИМИРОВИЧ

УДК 004.82:69:002;72.025;721

ДИСЕРТАЦІЯ

ІНФОРМАЦІЙНА ТЕХНОЛОГІЯ ОЦІНКИ ТЕХНІЧНОГО СТАНУ
ОБ'ЄКТІВ БУДІВНИЦТВА В УМОВАХ НЕВИЗНАЧЕНОСТІ ІЗ
ЗАСТОСУВАННЯМ МЕТОДИКИ КЛАСТЕРИЗАЦІЇ

126 – інформаційні системи та технології

(шифр і назва спеціальності)

12 «Інформаційні технології»

(галузь знань)

Подається на здобуття наукового ступеня
кандидата технічних наук
(доктора філософії)

Дисертація містить результати власних досліджень.
Використання ідей, результатів і текстів інших авторів мають посилання
на відповідне джерело
/БИКОВ С.В./

Науковий керівник: 1. Терент'єв Олександр Олександрович,
доктор технічних наук, професор, професор кафедри інформаційних
технологій проєктування та прикладної математики Київського
національного університету будівництва та архітектури
2. Горбатюк Євгеній Володимирович,
кандидат технічних наук, доцент кафедри будівельних машин Київського
національного університету будівництва та архітектури

Київ– 2026

АНОТАЦІЯ

Биков С.В. Інформаційна технологія оцінки технічного стану об'єктів будівництва в умовах невизначеності із застосуванням методики кластеризації. - *Кваліфікаційна наукова праця на правах рукопису.*

Дисертація на здобуття наукового ступеня доктора філософії за спеціальністю 126 - «Інформаційні системи та технології». Київський національний університет будівництва і архітектури. Київ, 2026.

Роботу присвячено розробленню інформаційної технології оцінювання технічного стану об'єктів будівництва в умовах невизначеності на основі застосування методів кластеризації. У межах дослідження передбачено формування відповідного математичного апарату та розроблення алгоритмічного забезпечення кластерного аналізу з використанням нечітких множин першого і другого типів та генетичних алгоритмів оптимізації. Запропонована технологія орієнтована на практичну реалізацію з можливістю інтеграції до сучасних інформаційних систем моніторингу технічного стану будівель і споруд.

Наукова новизна одержаних результатів. Наукова новизна дисертаційної роботи полягає у науковому обґрунтуванні інформаційної технології Clustering4Build, що інтегрує ІНМТ2, нечіткі множини та еволюційні алгоритми для класифікації технічного стану об'єктів в умовах невизначеності. Розроблено GA-Optimizer і TypeReducer для автоматизованої оптимізації параметрів та підвищення стійкості кластеризації, а також QualityEvaluator для обґрунтованого вибору оптимальних моделей. Інтеграція з DSS і мікросервісною архітектурою забезпечує масштабованість, прозорість рішень і зростання точності технічного моніторингу будівель.

В дисертаційній роботі **удосконалено:**

- інформаційну технологію Clustering4Build для оцінки технічного стану об'єктів будівництва в умовах невизначеності, яка інтегрує методи кластеризації на основі нечітких множин першого та другого типів з

генетичною оптимізацією параметрів. Забезпечено адаптивне групування об'єктів із урахуванням перехідних станів і різномірності даних.

- розширення множини об'єктів кластеризації до IHMT2 для PCM-алгоритму, що дозволяє враховувати невизначеність фазифікатора та «ширини зони» для кожного кластера. Реалізація в Clustering4Build через мікросервіси FuzzificationEngine і TypeReducer забезпечує обчислення інтервальних функцій типовості та центрів кластерів із збереженням даних у JSONB для DSS. Інноваційність полягає в інтеграції автоматизованого підбору параметрів PCM-алгоритму за допомогою GA-Optimizer та оцінки якості кластеризації через QualityEvaluator для точного технічного моніторингу.

- застосування комбінованих алгоритмів FCM+GA та PCM+GA до задачі кластеризації технічного стану об'єктів будівництва, у яких на основі відомих підходів до еволюційного кодування параметрів нечіткої кластеризації реалізовано автоматизоване визначення центрів кластерів, параметрів фазифікації та ступенів приналежності, адаптоване до специфіки шумових і неповних даних технічного моніторингу, що сприяє підвищенню точності, збіжності та обчислювальної ефективності оцінювання технічного стану;

- науково-прикладну реалізацію мікросервісної архітектури з REST-взаємодією для централізованого збереження результатів та інтеграції з DSS, BIM і IoT-потокami, що забезпечує масштабованість, модульність, відтворюваність обчислювальних експериментів та практичну підтримку прийняття рішень інженерами.

Набули подальшого розвитку:

- науково-методичні засади побудови інформаційної технології оцінки технічного стану об'єктів будівництва в умовах невизначеності, що, на відміну від існуючих BIM-, SHM- та DSS-рішень, базуються на інтегрованому використанні методів нечіткої кластеризації, апарату нечітких множин першого і другого типів та інтервальних моделей подання знань. Запропонований підхід забезпечує формалізацію перехідних станів конструкцій, поєднання інструментальних вимірювань і експертних оцінок, а

також адаптивну обробку різномірних і неповних даних моніторингу. Це створює теоретичне підґрунтя для розроблення універсального програмного інструментарію кластерного аналізу технічного стану в межах спеціальності 126 «Інформаційні системи та технології»;

- комбіновані методи інтелектуальної кластеризації FCM+GA та PCM+GA в складі інтегрованої інформаційної технології Clustering4Build, які передбачають генетичну оптимізацію центрів кластерів, параметрів фазифікації та ступенів приналежності з кодуванням хромосом відповідними ваговими коефіцієнтами. На відміну від існуючих підходів, запропонована архітектура реалізує замкнений цикл «оптимізація – аналіз – рекомендація – візуалізація», інтегрується з PostgreSQL, BIM-моделями та IoT-потокami даних і забезпечує підвищення швидкодії, стійкості збіжності та точності оцінювання технічного стану будівельних об'єктів.

Основний зміст дисертаційної роботи. Дисертаційну роботу присвячено вирішенню науково-прикладного завдання підвищення достовірності та автоматизації оцінки технічного стану об'єктів будівництва в умовах невизначеності шляхом розроблення інтегрованої інформаційної технології кластерного аналізу. У межах дослідження обґрунтовано перехід від фрагментарного аналізу показників до функціонування комплексних систем технічного моніторингу з використанням нечітких та інтервальних моделей подання знань. Розроблено методи кластеризації на основі нечітких множин першого і другого типів у поєднанні з генетичними алгоритмами оптимізації, що забезпечують адаптивне визначення параметрів моделей в умовах шумових, неповних і неоднорідних даних. Створено мікросервісну архітектуру Clustering4Build із реалізацією замкненого циклу «оптимізація – аналіз – рекомендація – візуалізація» та інтеграцією з BIM-моделями, IoT-потокami і базою даних PostgreSQL. Запропонована технологія забезпечує підвищення точності, швидкодії та відтворюваності обчислювальних експериментів, формування обґрунтованих рекомендацій у межах DSS та

створює надійну основу для автоматизованого моніторингу і прогнозування технічного стану будівель у реальному часі.

За результатами дослідження розроблено:

1. Розроблено інтегровану інформаційну технологію Clustering4Build для оцінки технічного стану об'єктів будівництва в умовах невизначеності, що поєднує інтервальні нечіткі множини другого типу, алгоритми FCM і PCM та генетичну оптимізацію параметрів у єдиному адаптивному середовищі. Створено програмні модулі GA-Optimizer, TypeReducer і QualityEvaluator, які забезпечують автоматизоване налаштування фазифікаторів, центрів кластерів і параметрів «ширини зони», а також обґрунтований вибір моделей за показниками якості кластеризації. Апробація підтвердила підвищення точності, стійкості збіжності та відтворюваності результатів у задачах технічного моніторингу за наявності шумових і неповних даних.

2. Обґрунтовано мікросервісну архітектуру інформаційного середовища технічного моніторингу з REST-взаємодією, що забезпечує модульність, масштабованість та інтеграцію з DSS, BIM-моделями, IoT-потокami і базою даних PostgreSQL (JSONB). Реалізовано замкнений цикл «оптимізація – аналіз – рекомендація – візуалізація», який підтримує автоматизовану кластеризацію та формування інженерних рішень у реальному часі. Запропонована структура гарантує прозорість обчислювальних експериментів, аудит результатів і практичну підтримку прийняття рішень щодо безпечної експлуатації будівельних об'єктів.

У **першому розділі** сформовано концептуально-теоретичну базу оцінки технічного стану об'єктів будівництва в умовах невизначеності. Обґрунтовано доцільність використання нечітких та інтервальних моделей для формалізації знань у системах технічного моніторингу. Показано, що нечітка кластеризація дозволяє враховувати перехідні стани конструкцій, різномірність і неповноту вхідних даних. Поєднання інструментальних вимірювань з експертними оцінками забезпечує комплексний аналіз стану об'єктів. Проаналізовано сучасні BIM-, SHM- та DSS-рішення та виявлено обмеженість їх аналітичних

можливостей щодо адаптивної кластеризації. Встановлено відсутність універсального програмного інструментарію для інтеграції різнорідних даних моніторингу. Визначено необхідність створення інтегрованої інформаційної технології з поєднанням нечітких множин, кластерного аналізу та засобів візуалізації результатів. Розділ закладає теоретичне підґрунтя для розробки адаптивної та масштабованої системи оцінки технічного стану будівель.

Другий розділ присвячено формуванню науково-методичного підґрунтя та розробці комбінованих методів кластеризації FCM+GA та PCM+GA для оцінки технічного стану об'єктів будівництва. Розроблено алгоритми адаптивного визначення центрів кластерів, ступенів приналежності та параметрів фазифікації в умовах шумових, неповних та неоднорідних даних. Створено модулі фазифікації, аналітичного ядра, оптимізації та DSS, інтегровані з BIM-моделями, IoT-потокami та базою PostgreSQL. Інноваційність полягає у поєднанні нечіткої логіки, генетичної оптимізації та кодування хромосом ступенями приналежності. Розроблена архітектура Clustering4Build реалізує замкнений цикл «оптимізація → аналіз → рекомендація → візуалізація». Це забезпечує підвищення стабільності, швидкодії та точності кластеризації. Запропоновані методи інтегрують математичну ефективність, інформаційні технології та практичну адаптивність. Розділ встановлює основу для автоматизованого та надійного прогнозування технічного стану будівель.

У **третьому розділі** обґрунтовано застосування FCM та PCM-алгоритмів з IHMT2 для оцінки технічного стану будівель у умовах невизначеності. Розроблено модулі FuzzificationEngine та FCM-Core для автоматичного формування інтервальних функцій приналежності. Визначено показники якості кластеризації, такі як гіпероб'єм та індекс сферичності, для вибору оптимальної моделі. Застосовано генетичні алгоритми для оптимізації фазифікаторів і параметрів «ширини зони», що підвищує точність і відтворюваність кластеризації. Розширено множину об'єктів PCM-алгоритму для врахування різної щільності та об'єму даних. Запропоновано методику

вибору між НМТ1 та ІНМТ2 на основі аналізу структурної варіативності даних. Інтеграція алгоритмів з мікросервісною архітектурою забезпечує автоматизацію та надійність технічного моніторингу. Наведено приклади класифікації об'єктів із різною щільністю та об'ємом для демонстрації ефективності підходу.

Четвертий розділ присвячено розробці та дослідженню модулів інформаційної технології Clustering4Build для автоматизованої оцінки технічного стану будівель. Запропоновано мікросервісну архітектуру з REST-взаємодією, що забезпечує модульність, масштабованість і технологічну незалежність компонентів. Реалізовано підтримку кластеризації на основі НМТ1 та ІНМТ2 із процедурою зниження типу. Запроваджено централізоване збереження результатів у PostgreSQL (JSONB), що гарантує відтворюваність і аудит обчислювальних експериментів. Інтеграція генетичної оптимізації та DSS дозволяє адаптивно налаштовувати моделі та формувати практичні рекомендації. Створено комплекс пакетів прикладних програм для реального часу, що забезпечує надійний та точний моніторинг будівель. Запропоновані рішення інтегруються з BIM, IoT та DSS, підвищуючи інтелектуальність системи. Розділ демонструє практичну реалізацію комплексного, автоматизованого та адаптивного технічного моніторингу об'єктів будівництва.

Ключові слова: оцінка технічного стану будівель засобами інформаційної технології Clustering4Build, автоматизований моніторинг, нечіткі множини, інтервальні моделі, кластерний аналіз, генетична оптимізація, адаптивні моделі, комбіновані алгоритми FCM+GA та PCM+GA, мікросервісна архітектура, цифрові дані, інтеграція з BIM та IoT.

ABSTRACT

Bykov S.V. Information technology for assessing the technical condition of construction objects under conditions of uncertainty using the clustering method. - *Qualification scientific work in the form of a manuscript.*

Dissertation for the degree of Doctor of Philosophy in specialty 126 – “Information Systems and Technologies”. Kyiv National University of Construction and Architecture. Kyiv, 2026.

The work is devoted to the development of information technology for assessing the technical condition of construction objects under conditions of uncertainty based on the use of clustering methods. The research provides for the formation of an appropriate mathematical apparatus and the development of algorithmic support for cluster analysis using fuzzy sets of the first and second types and genetic optimization algorithms. The proposed technology is focused on practical implementation with the possibility of integration into modern information systems for monitoring the technical condition of buildings and structures.

Scientific novelty of the results obtained. The scientific novelty of the dissertation work lies in the scientific substantiation of the Clustering4Build information technology, which integrates INMT2, fuzzy sets and evolutionary algorithms for classifying the technical condition of objects under conditions of uncertainty. GA-Optimizer and TypeReducer have been developed for automated optimization of parameters and increasing the stability of clustering, as well as QualityEvaluator for a reasoned selection of optimal models. Integration with DSS and microservice architecture provides scalability, transparency of solutions and increased accuracy of technical monitoring of buildings.

The dissertation work has improved:

- Clustering4Build information technology for assessing the technical condition of construction objects under conditions of uncertainty, which integrates clustering methods based on fuzzy sets of the first and second types with genetic optimization of parameters. Adaptive grouping of objects is provided, taking into account transient states and data heterogeneity.

- expansion of the set of clustering objects to INMT2 for the PCM algorithm, which allows taking into account the uncertainty of the fuzzifier and the "zone width" for each cluster. Implementation in Clustering4Build through the FuzzificationEngine and TypeReducer microservices provides calculation of

interval typicality functions and cluster centers with data storage in JSONB for DSS. Innovation lies in the integration of automated selection of PCM algorithm parameters using GA-Optimizer and clustering quality assessment using QualityEvaluator for accurate technical monitoring

- combined algorithms FCM+GA and PCM+GA with automated determination of cluster centers, degrees of membership and fuzzification parameters, which increases the accuracy, stability and speed of technical condition assessment in conditions of noisy, incomplete and heterogeneous data.

- scientific and applied implementation of microservice architecture with REST interaction for centralized storage of results and integration with DSS, BIM and IoT flows, which ensures scalability, modularity, reproducibility of computational experiments and practical support for decision-making by engineers.

Further development has been made:

- scientific and methodological principles for building information technology for assessing the technical condition of construction objects under conditions of uncertainty, which, unlike existing BIM, SHM and DSS solutions, are based on the integrated use of fuzzy clustering methods, the apparatus of fuzzy sets of the first and second types and interval models of knowledge representation. The proposed approach provides formalization of transient states of structures, combination of instrumental measurements and expert assessments, as well as adaptive processing of heterogeneous and incomplete monitoring data. This creates a theoretical basis for the development of a universal software toolkit for cluster analysis of technical condition within the scope of specialty 126 "Information Systems and Technologies";

- combined methods of intelligent clustering FCM+GA and PCM+GA as part of the integrated information technology Clustering4Build, which provide for genetic optimization of cluster centers, fuzzification parameters and degrees of membership with chromosome encoding by appropriate weight coefficients. Unlike existing approaches, the proposed architecture implements a closed loop "optimization - analysis - recommendation - visualization", integrates with

PostgreSQL, BIM models and IoT data streams and provides increased speed, stability of convergence and accuracy of assessing the technical condition of construction objects.

The main content of the dissertation work. The dissertation work is devoted to solving the scientific and applied problem of increasing the reliability and automation of assessing the technical condition of construction objects under conditions of uncertainty by developing an integrated information technology of cluster analysis. The study substantiates the transition from fragmentary analysis of indicators to the functioning of complex technical monitoring systems using fuzzy and interval knowledge representation models. Clustering methods based on fuzzy sets of the first and second types in combination with genetic optimization algorithms have been developed, which provide adaptive determination of model parameters in conditions of noisy, incomplete and heterogeneous data. The Clustering4Build microservice architecture has been created with the implementation of a closed loop “optimization - analysis - recommendation - visualization” and integration with BIM models, IoT streams and the PostgreSQL database. The proposed technology provides increased accuracy, speed and reproducibility of computational experiments, the formation of substantiated recommendations within the DSS and creates a reliable basis for automated monitoring and forecasting of the technical condition of buildings in real time.

According to the results of the study, the following was developed:

1. An integrated information technology Clustering4Build was developed to assess the technical condition of construction objects under uncertainty, which combines interval fuzzy sets of the second type, FCM and PCM algorithms and genetic optimization of parameters in a single adaptive environment. The GA-Optimizer, TypeReducer and QualityEvaluator software modules were created, which provide automated tuning of fuzzifiers, cluster centers and "zone width" parameters, as well as a reasoned selection of models based on clustering quality indicators. The testing confirmed the increase in accuracy, stability of convergence

and reproducibility of results in technical monitoring tasks in the presence of noisy and incomplete data.

2. The microservice architecture of the technical monitoring information environment with REST interaction is substantiated, which provides modularity, scalability and integration with DSS, BIM models, IoT streams and the PostgreSQL database (JSONB). A closed loop “optimization - analysis - recommendation - visualization” is implemented, which supports automated clustering and the formation of engineering solutions in real time. The proposed structure guarantees the transparency of computational experiments, audit of results and practical support for decision-making regarding the safe operation of construction objects.

In the first section, a conceptual and theoretical basis for assessing the technical condition of construction objects under uncertainty is formed. The feasibility of using fuzzy and interval models for formalizing knowledge in technical monitoring systems is substantiated. It is shown that fuzzy clustering allows taking into account transient states of structures, heterogeneity and incompleteness of input data. The combination of instrumental measurements with expert assessments provides a comprehensive analysis of the condition of objects. Modern BIM, SHM and DSS solutions are analyzed and their analytical capabilities regarding adaptive clustering are limited. The absence of a universal software toolkit for integrating heterogeneous monitoring data is established. The need to create an integrated information technology combining fuzzy sets, cluster analysis and means of visualizing results is determined. The section lays the theoretical foundation for the development of an adaptive and scalable system for assessing the technical condition of buildings.

The second section is devoted to the formation of a scientific and methodological basis and the development of combined clustering methods FCM+GA and PCM+GA for assessing the technical condition of construction objects. Algorithms for adaptive determination of cluster centers, degrees of membership and fuzzification parameters in conditions of noisy, incomplete and heterogeneous data have been developed. Modules for fuzzification, analytical core,

optimization and DSS have been created, integrated with BIM models, IoT streams and the PostgreSQL database. The innovation lies in the combination of fuzzy logic, genetic optimization and chromosome coding by degrees of membership. The developed Clustering4Build architecture implements a closed loop “optimization → analysis → recommendation → visualization”. This ensures increased stability, speed and accuracy of clustering. The proposed methods integrate mathematical efficiency, information technology and practical adaptability. The section establishes the basis for automated and reliable prediction of the technical condition of buildings.

The third section justifies the use of FCM and PCM algorithms with INMT2 for assessing the technical condition of buildings under uncertainty. The FuzzificationEngine and FCM-Core modules are developed for automatic generation of interval membership functions. Clustering quality indicators, such as hypervolume and sphericity index, are determined for selecting the optimal model. Genetic algorithms are used to optimize fuzzifiers and “zone width” parameters, which increases the accuracy and reproducibility of clustering. The set of objects of the PCM algorithm is expanded to take into account different data densities and volumes. A method for choosing between NMT1 and INMT2 based on the analysis of structural variability of data is proposed. The integration of algorithms with a microservice architecture ensures automation and reliability of technical monitoring. Examples of classification of objects with different densities and volumes are given to demonstrate the effectiveness of the approach.

The fourth section is devoted to the development and research of information technology modules Clustering4Build for automated assessment of the technical condition of buildings. A microservice architecture with REST interaction is proposed, which ensures modularity, scalability and technological independence of components. Support for clustering based on NMT1 and INMT2 with a type reduction procedure is implemented. Centralized storage of results in PostgreSQL (JSONB) is introduced, which guarantees reproducibility and audit of computational experiments. Integration of genetic optimization and DSS allows adaptively tuning

models and generating practical recommendations. A set of application software packages for real-time has been created, which provides reliable and accurate monitoring of buildings. The proposed solutions are integrated with BIM, IoT and DSS, increasing the intelligence of the system. The section demonstrates the practical implementation of complex, automated and adaptive technical monitoring of construction sites.

Keywords: assessment of the technical condition of a building based on IT Clustering4Build, automated monitoring, fuzzy sets, interval models, cluster analysis, genetic optimization, adaptive models, combined FCM+GA and PCM+GA algorithms, microservice architecture, digital data, integration with BIM and IoT.

СПИСОК ОПУБЛІКОВАНИХ ПРАЦЬ ЗА ТЕМОЮ ДИСЕРТАЦІЇ

Статті у наукових фахових виданнях України категорії «Б»:

1. Gorbatyuk Ie., Terentyev O., Mishchuk D., **Bykov S.** Integrated models of forecasting reliability of decision making of the system of diagnostics of the technical condition of buildings. *Гірничі, будівельні, дорожні та меліоративні машини*. 2025. № 105. С. 73–78. DOI: <https://doi.org/10.32347/gbdmm.2025.105.0501>.

2. Tyslenko O., **Bykov S.**, Katin O. System of organization and functioning of mechanisms for ensuring the operational safety of buildings. *Шляхи підвищення ефективності будівництва*. 2025. Вип. 2(56). С. 91–104. DOI: [https://doi.org/10.32347/2707-501x.2025.56\(2\).91-104](https://doi.org/10.32347/2707-501x.2025.56(2).91-104).

3. Тисленко О. Б., **Биков С. В.** Аналітичні та діагностичні блоки контролю в цифровому середовищі оцінки структурної надійності. *Управління розвитком складних систем*. 2025. № 64. С. 259–266. DOI: <https://doi.org/10.32347/2412-9933.2025.64.259-266>.

Статті в закордонних періодичних виданнях, прирівняних до фахових видань категорії «Б» (за умови індексації/цитування):

4. Terentyev O. O., Gorbatyuk Ie. V., Tyslenko O. B., **Bykov S. V.** Methods of damage of the information system of building technical condition diagnostics. *Modern engineering and innovative technologies*. Karlsruhe, 2024. Iss. 35, part 2. P. 35–44. DOI: <https://doi.org/10.30890/2567-5273.2024-35-00-049>.

5. Terentyev O. O., Gorbatyuk Ie. V., Tyslenko O. B., **Bykov S. V.** Program and technical complex of implementation of the information system of diagnostics of the technical condition of buildings. *SWorldJournal*. 2025. Iss. 30, part 1. P. 82–96. DOI: <https://doi.org/10.30888/2663-5712.2025-30-01-004>.

6. Terentyev O. O., Gorbatyuk Ie. V., Tyslenko O. B., **Bykov S. V.** Creation of information methods in the diagnosis of industrial buildings. *Modern engineering and innovative technologies*. 2025. Iss. 38, part 1. P. 146–151. DOI: <https://doi.org/10.30890/2567-5273.2025-38-01-030>.

Наукові праці апробаційного характеру

(тези доповідей на науково-практичних конференціях)

7. Terentyev O. O., Gorbatyuk Ie. V., Tyslenko O. B., **Bykov S. V.** Methods of diagnosing deformations of building structures. *International scientific conference "Technique and technology of the future '2024"*, Karlsruhe, Germany, 20 October 2024. P. 36–40. DOI: <https://doi.org/10.30890/2709-1783.2024-35-00-027>.
8. Терентьев О. О., Горбатюк Є. В., Тисленко О. Б., **Биков С. В.** Інтегрована модель оцінки надійності рішень для діагностики будівель. *Proceedings of the 6th International scientific and practical conference "Scientific achievements of contemporary society"*, London, United Kingdom, 10–12 January 2025. P. 280–286.
9. Терентьев О. О., Горбатюк Є. В., Тисленко О. Б., **Биков С. В.** Інформаційна технологія управління життєвим циклом будівельних конструкцій і споруд. *Modernization of today's science: experience and trends: coll. of sci. papers "SCIENTIA"*, Singapore, 24 January 2025. P. 192–195. DOI: <https://doi.org/10.36074/scientia-24.01.2025>.
10. Терентьев О. О., Горбатюк Є. В., Тисленко О. Б., **Биков С. В.** Моделювання загроз безпеки та вразливості у віртуальній мережевій системі. *Proceedings of the 5th International scientific and practical conference "Science in the modern world: innovations and challenges"*, Toronto, Canada, 23–25 January 2025. P. 169–173.
11. Terentyev O. O., Gorbatyuk Ie. V., Tyslenko O. B., **Bykov S. V.** Creating a model of mobile location. *International scientific conference "Organization of scientific research in modern conditions '2024"*, Seattle, USA, 21 March 2025. P. 25–28. DOI: <https://doi.org/10.30888/2709-2267.2025-30-00-003>.

ЗМІСТ

ВСТУП.....	21
РОЗДІЛ 1. ОГЛЯД І АНАЛІЗ ПІДХОДІВ ДО ВИРІШЕННЯ ЗАДАЧІ ОЦІНКИ І КЛАСИФІКАЦІЇ ТЕХНІЧНОГО СТАНУ ОБ'ЄКТІВ БУДІВНИЦТВА В ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЯХ.....	31
1.1. Особливості задач оцінювання технічного стану об'єктів будівництва як інформаційної проблеми.....	31
1.2. Джерела та види невизначеності в інформаційних системах технічного моніторингу.....	37
1.3. Аналіз інформаційних технологій та методів інтелектуальної обробки даних для задач технічного моніторингу.....	40
1.4. Методи кластеризації в інформаційних системах: можливості та обмеження.....	45
1.4.1. Ієрархічна кластеризація об'єктів.....	48
1.4.2 Алгоритм чітких c-середніх.....	49
1.4.3 Можливісна кластеризація	50
1.4.4 Показники якості.....	51
1.5. Нечіткі та інтервальні моделі представлення знань в інформаційних технологіях.....	52
1.6. Аналіз сучасних програмних та інформаційних систем оцінки технічного стану.....	58
1.7. Постановка науково-прикладної задачі та вимоги до інформаційної технології оцінки технічного стану об'єктів будівництва.....	61
Висновки до розділу 1.....	66
РОЗДІЛ 2. РОЗРОБКА МЕТОДІВ КЛАСТЕРИЗАЦІЇ НА ОСНОВІ НЕЧІТКИХ МНОЖИН ПЕРШОГО ТИПУ І ГЕНЕТИЧНИХ АЛГОРИТМІВ.....	69
2.1. Математичний базис інформаційної технології кластеризації технічного стану об'єктів будівництва.....	69
2.1.1. Формалізація технічних показників у вигляді нечітких множин.....	69

2.1.2. Математичні основи FCM та обробка даних у середовищі невизначеності.....	75
2.2. Генетичні алгоритми оптимізації кластеризації з використанням РСМ-алгоритму на основі нечітких множин першого типу.....	80
2.2.1. Кодування хромосоми нечіткими ступенями приналежності об'єктів центрам кластерів.....	82
2.2.2. Генетичний алгоритм з хромосомою постійної довжини, що закодована координатами центрів кластерів.....	83
2.2.3. Генетичний алгоритм з хромосомою постійної довжини, що закодована нечіткими ступенями приналежності об'єктів центрам кластерів.....	87
2.2.4. Генетичний алгоритм з хромосомою змінної довжини, що закодована координатами центрів кластерів.....	92
2.2.5. Аналіз проблеми реалізації генетичного алгоритму з хромосомою змінної довжини, що закодована нечіткими ступенями приналежності об'єктів центрам кластерів.....	96
2.3. Комбінування алгоритму FCM на основі нечітких множин першого типу з генетичним алгоритмом	97
2.3.1. Комбінування FCM-алгоритму на основі нечітких множин першого типу і генетичного алгоритму з хромосомою постійної довжини при кодуванні хромосом координатами центрів кластерів.....	99
2.3.2. Комбінування FCM-алгоритму на основі нечітких множин першого типу і генетичного алгоритму з хромосомою змінної довжини при кодуванні хромосом координатами центрів кластерів.....	103
2.4. Комбінування РСМ-алгоритму на основі нечітких множин першого типу і генетичного алгоритму.....	115
2.4.1. Алгоритмічна схема поєднання РСМ-методу та генетичного алгоритму з фіксованою хромосомою.....	117

2.4.2 Комбінування РСМ-алгоритму на основі нечітких множин першого типу і генетичного алгоритму з хромосомою змінної довжини при кодуванні хромосом координатами центрів кластерів.....	120
2.5. Особливості застосування генетичних алгоритмів з постійною і змінною довжиною хромосоми для пошуку оптимальних параметрів алгоритмів кластеризації на основі нечітких множин першого типу.....	121
2.6. Проблема вибору комбінованого методу кластеризації на основі нечітких множин першого типу.....	122
2.7. Приклади класифікації технічного стану об'єктів будівництва в інформаційній технології оцінки технічного стану об'єктів будівництва з використанням комбінованих методів кластеризації на основі нечітких множин першого типу.....	123
2.8. Архітектура інформаційної технології кластеризації технічного стану Clustering4Build.....	133
Висновок до розділу 2.....	134
РОЗДІЛ 3. РОЗРОБКА МЕТОДІВ КЛАСТЕРИЗАЦІЇ НА ОСНОВІ ІНТЕРВАЛЬНИХ НЕЧІТКИХ МНОЖИН ДРУГОГО ТИПУ І ГЕНЕТИЧНИХ АЛГОРИТМІВ.....	137
3.1 Кластеризація з використанням FCM-алгоритму на основі інтервальних нечітких множин другого типу.....	137
3.1.1 Проблема невизначеності фазифікатора в FCM-алгоритмі.....	138
3.1.2 Розширення множини об'єктів кластеризації на інтервальні нечіткі множини другого типу для FCM-алгоритму.....	142
3.1.3 Ітераційний алгоритм Карника–Менделя для обчислення інтервальних центрів кластерів.....	146
3.2 Проблема вибору показника якості кластеризації на основі інтервальних нечітких множин другого типу.....	149
3.3 Генетичний алгоритм пошуку оптимальної комбінації значень фазифікаторів для FCM-алгоритму на основі інтервальних нечітких множин другого типу.....	152

3.4 Кластеризація з використанням FCM-алгоритму на основі інтервальних нечітких множин другого типу.....	155
3.5. Розширення множини об'єктів кластеризації на інтервальні нечіткі множини другого типу для РСМ-алгоритму.....	157
3.6 Генетичний алгоритм пошуку оптимальної комбінації значень фаззифікаторів, що реалізують управління невизначеністю, і значень «ширини зони» для РСМ-алгоритму на основі інтервальних нечітких множин другого типу.....	161
3.7. Генетичний алгоритм пошуку оптимальної комбінації значення фаззифікатора і значень «ширини зони», що реалізують управління невизначеністю, для РСМ-алгоритму на основі інтервальних нечітких множин другого типу.....	163
3.8 Проблема вибору методу кластеризації на основі нечітких множин першого типу або інтервальних нечітких множин другого типу в інформаційних технологіях оцінки технічного стану об'єктів будівництва.....	167
3.9 Приклади класифікації технічного стану об'єктів будівництва в інформаційних технологіях оцінки технічного стану об'єктів будівництва з використанням методів кластеризації на основі інтервальних нечітких множин другого типу.....	168
3.9.1 Кластеризація множини об'єктів на два кластери суттєво різного об'єму і суттєво різної щільності з використанням методу можливої кластеризації на основі інтервальних нечітких множин другого типу для фіксованої комбінації значень фаззифікаторів.....	182
3.9.2 Кластеризація множини об'єктів на два кластери суттєво різного об'єму і суттєво різної щільності з використанням методу можливої кластеризації на основі інтервальних нечітких множин другого типу для довільної комбінації значень фаззифікаторів.....	184
Висновки до розділу 3	189

РОЗДІЛ 4. ПРОГРАМНА РЕАЛІЗАЦІЯ МЕТОДІВ КЛАСТЕРИЗАЦІЇ НА ОСНОВІ НЕЧІТКИХ МНОЖИН.....	191
4.1 Загальні характеристики пакету прикладних програм «Класифікація технічного стану об'єктів будівництва в умовах невизначеності».....	191
4.2 Особливості розробки пакету прикладних програм інтегрованого в сучасну інформаційну технологію Clustering4Build.....	193
4.3. Пакет прикладних програм «Класифікація технічного стану об'єктів будівництва в умовах невизначеності».....	198
4.4. Загальна структура пакету прикладних програм «Класифікація технічного стану об'єктів будівництва в умовах невизначеності».....	207
4.5 Програмна реалізація комплексу ПП як складових інформаційної технології Clustering4Build.....	225
Висновки до розділу 4.....	249
ЗАГАЛЬНІ ВИСНОВКИ	250
Список літератури.....	254
Додатки.....	267

ВСТУП

Актуальність теми дисертації обумовлена зростанням вимог до надійності, безпеки та експлуатаційної стійкості будівельних об'єктів в умовах інтенсивної експлуатації, воєнних ушкоджень і старіння інфраструктури. За даними Київської школи економіки, станом на листопад 2024 року в Україні пошкоджено або зруйновано близько 236 тисяч об'єктів житлового фонду (8,6% загальної площі житлового фонду країни), а прямі збитки житловому сектору оцінюються у понад 60 млрд доларів [2]. Це зумовлює безпрецедентне навантаження на систему технічних обстежень і потребу в автоматизованих засобах масової та оперативної оцінки технічного стану об'єктів. Сучасні системи моніторингу генерують великі обсяги різномірних, неповних і шумових даних, що ускладнює застосування традиційних методів оцінки технічного стану. Невизначеність вимірювань, суб'єктивність експертних оцінок і складність конструкцій створюють потребу у розробленні методів обробки та інтеграції таких даних. Наукове завдання полягає в забезпеченні адаптивного аналізу технічного стану будівельних об'єктів з урахуванням перехідних станів, різномірності та неповноти інформації. Виникає потреба у застосуванні інтелектуальних моделей, що поєднують кластеризацію, нечіткі множини та еволюційні алгоритми для формалізації знань у системах технічного моніторингу. Суттєвим є створення універсального підходу, здатного інтегрувати різномірні джерела даних, включно з інструментальними вимірюваннями та експертними оцінками. Необхідно забезпечити науково-обґрунтовану, дієву та достовірну автоматизовану оцінку та прогнозування стану об'єктів у реальному часі. Це відкриває перспективи для розвитку комплексних інформаційних технологій, що поєднують аналітичні моделі, обчислювальні алгоритми та моделі прийняття рішень для ефективного наукового дослідження будівельних систем.

Зв'язок роботи з науковими програмами, планами, темами. Дисертаційне дослідження узгоджується з пріоритетами державної науково-технологічної політики та цифрової трансформації, включно з положеннями

Концепції розвитку штучного інтелекту в Україні та стратегічними ініціативами цифрової модернізації держави. Робота відповідає завданням науково-технічних програм і планів повоєнного відновлення України на 2023–2032 роки, а також стратегій цифрового розвитку та трансформації. Тематика дослідження корелює із законодавчими положеннями у сфері наукової, інформаційної та технічної діяльності. Дисертація відповідає профілю спеціальності 126 – «Інформаційні системи та технології» та спрямована на вдосконалення методів аналізу та кластеризації даних технічного моніторингу будівель. Робота виконана на кафедрі інформаційних технологій проєктування та прикладної математики КНУБА. Отримані результати використовуються для формування змісту навчальних дисциплін та підготовки магістрів за спеціальністю 126. Дослідження інтегрує методи обробки невизначених даних, кластеризації та математичної оптимізації в інформаційних системах будівництва. Робота формує науково-методичну основу для розвитку інтелектуальних систем моніторингу та прогнозування технічного стану будівель. Вона сприяє підвищенню точності аналізу даних і підтримці прийняття рішень у наукових і освітніх програмах університету.

Мета і завдання дослідження. Мета дисертаційної роботи полягає у створенні інтегрованої інформаційної технології для оцінки технічного стану об'єктів будівництва в умовах невизначеності шляхом застосування методів кластеризації, що передбачає формування математичного апарату та розробку алгоритмів кластерного аналізу з використанням нечітких множин першого і другого типів та генетичних алгоритмів оптимізації; дослідження сфокусоване на забезпеченні адаптивного, точного та стабільного визначення параметрів моделей технічного стану, технологія має інтегруватися із сучасними інформаційними системами моніторингу будівель і споруд, забезпечуючи автоматизацію оцінки та підтримку прийняття рішень, реалізація підходу спрямована на підвищення надійності, безпеки та ефективності експлуатації об'єктів будівництва. Результатом дослідження є формування засад

універсального програмного інструментарію для автоматизованого моніторингу та аналізу технічного стану будівель.

Досягнення мети роботи забезпечено постановкою та вирішенням наступних завдань дослідження:

1. Провести аналіз існуючих методів та інформаційних технологій оцінки технічного стану об'єктів будівництва в умовах невизначеності та сформулювати вимоги до інформаційної технології кластеризації.
2. Розробити математичний апарат і алгоритми кластерного аналізу з використанням нечітких множин першого і другого типів для адаптивного визначення стану конструкцій.
3. Створити комбіновані алгоритмічні підходи на основі FCM, РСМ та генетичної оптимізації для обробки неповних, шумових і різнорідних даних моніторингу.
4. Розробити програмні модулі для фазифікації, аналітичного ядра, оптимізації та інтеграції з DSS, BIM-моделями і потоками IoT для підтримки прийняття рішень.
5. Впровадити методику оцінки технічного стану будівель із використанням інтервальних нечітких множин другого типу та автоматизованого підбору параметрів через генетичну оптимізацію.
6. Забезпечити реалізацію інформаційної технології як інтегрованого програмного продукту для адаптивного, точного і відтворюваного моніторингу технічного стану об'єктів будівництва у реальному часі.

Об'єктом дослідження є процес оцінки технічного стану будівельних об'єктів у системах технічного моніторингу, що включає збирання, обробку та інтерпретацію багатовимірних даних про конструкції в умовах невизначеності та неповноти інформації. **Предметом дослідження** визначена інформаційна технологія кластеризації технічного стану будівельних об'єктів, що охоплює методи та алгоритми на основі нечітких множин і генетичної оптимізації для підвищення точності та адаптивності аналізу.

Обґрунтованість та достовірність отриманих результатів. Надійність і наукова обґрунтованість результатів дисертаційного дослідження забезпечуються коректністю постановки задачі кластеризації, використанням апробованого математичного апарату теорії нечітких множин першого і другого типів та узгодженістю отриманих рішень із класичними результатами теорії нечіткої кластеризації (FCM, PCM). Достовірність результатів підтверджується порівнянням розроблених комбінованих методів FCM+GA і PCM+GA з базовими методами кластеризації (k-means, ієрархічна кластеризація) на тестових і практичних наборах даних, а також аналізом показників якості кластеризації (гіпероб'єм, індекс сферичності) для різних комбінацій параметрів. Коректність запропонованих підходів до формалізації технічного стану об'єктів будівництва узгоджується з вимогами чинних нормативних документів, зокрема ДСТУ-Н Б В.1.2-18:2016. Обчислювальні експерименти виконано на реалізованому програмному комплексі Clustering4Build із фіксацією проміжних результатів (центрів кластерів, матриць приналежності, показників якості) для забезпечення відтворюваності обчислень.

Методи дослідження. Методологічне підґрунтя дисертаційного дослідження сформовано як цілісну систему взаємопов'язаних підходів, що інтегрують загальнонаукові, спеціалізовані та прикладні методи для оцінювання технічного стану будівельних об'єктів за умов невизначеності. До загальнонаукових методів належать системний і структурно-функціональний аналіз, абстрагування та узагальнення, індукція і дедукція, формалізація, моделювання, порівняльний аналіз та метод наукової гіпотези, які забезпечують цілісне осмислення проблеми та наукове обґрунтування розроблених підходів.

Спеціальні методи та моделі дослідження включають методи теорії нечітких множин першого типу та інтервальних нечітких множин другого типу, методи нечіткого (FCM) та можливісного (PCM) кластерного аналізу, генетичні алгоритми оптимізації параметрів кластеризації, а також

ітераційний алгоритм Карника–Менделя для зниження типу інтервальних нечітких моделей. Особливу увагу приділено методам кластеризації технічного стану будівельних об'єктів на основі нечітких множин першого типу та генетичних алгоритмів. Розроблені комбіновані підходи FCM + GA та PCM + GA дозволяють адаптивно визначати центри кластерів, ступені приналежності та параметри фазифікації, ефективно працюючи з шумовими, неповними й неоднорідними даними.

Важливим елементом дослідження стало створення програмних модулів для фазифікації, аналітичного ядра, оптимізації та DSS, інтегрованих із базою даних PostgreSQL, BIM-моделями та IoT-потокami. Інноваційність підходу полягає у поєднанні нечіткої логіки та генетичної оптимізації з кодуванням хромосом ступенями приналежності, що забезпечує прискорену збіжність алгоритмів і підвищену точність результатів. Архітектура Clustering4Build реалізує замкнений цикл «оптимізація → аналіз → рекомендація → візуалізація», підвищуючи стабільність та швидкодію процесу кластеризації.

Особлива увага приділялася застосуванню генетичних алгоритмів із хромосомою постійної та змінної довжини: хромосома змінної довжини дозволяє автоматично визначати оптимальну кількість кластерів, що підвищує ефективність пошуку та скорочує час обчислень. Для вибору оптимального методу кластеризації (FCM, PCM або комбінованого) реалізовано порівняння показників якості кластеризації (гіпероб'єм, індекс сферичності) з формуванням рекомендацій на основі отриманих метрик, що забезпечує обґрунтований вибір моделі на кожному етапі аналізу.

Застосування розробленої інформаційної технології передбачає класифікацію технічного стану об'єктів будівництва за допомогою комбінованих методів нечіткої кластеризації з розподілом об'єктів на цільові класи технічного стану. Експериментальні дослідження показали, що комбінований метод FCM + GA із кодуванням хромосом ступенями приналежності забезпечує оптимальний баланс між точністю кластеризації, стабільністю алгоритму та швидкістю, а раціональний розмір популяції (30–

50 хромосом) і рівень мутації (5–10%) дозволяють досягати глобального мінімуму цільової функції без надмірних витрат часу.

Таким чином, методичне забезпечення дослідження поєднує сучасні математичні методи, інформаційні технології та практичну адаптивність, що забезпечує достовірну оцінку та прогнозування технічного стану будівельних об'єктів, інтегруючи результати кластеризації з натурними обстеженнями та підвищуючи практичну цінність розробленої системи для інженерного моніторингу.

Наукова новизна одержаних результатів полягає в науково-методичному обґрунтуванні та практичній реалізації інформаційної технології Clustering4Build, що поєднує інтервальні нечіткі множини другого типу, апарат нечітких множин та еволюційні алгоритми для класифікації технічного стану будівельних об'єктів за умов невизначеності. У межах дослідження створено модулі GA-Optimizer і TypeReducer, призначені для автоматизованого налаштування параметрів моделей та підвищення стійкості процедур кластеризації, а також QualityEvaluator для формалізованого вибору найкращих конфігурацій за показниками якості. Інтеграція технології з DSS та мікросервісною архітектурою забезпечує масштабованість, прозорість формування рішень і підвищення достовірності технічного моніторингу будівель.

У дисертаційній роботі **удосконалено:**

- інформаційну технологію Clustering4Build для оцінювання технічного стану об'єктів будівництва в умовах неповноти та нечіткості даних, у якій методи кластеризації на основі нечітких множин першого і другого типів інтегровано з генетичною оптимізацією параметрів, що забезпечує адаптивне групування об'єктів з урахуванням перехідних станів та різноманітності інформації;

- підхід до розширення множини об'єктів кластеризації на основі інтервальних нечітких множин другого типу для РСМ-алгоритму, що дозволяє враховувати невизначеність фазифікатора та параметра «ширини зони»

окремо для кожного кластера. Реалізація через мікросервіси FuzzificationEngine і TypeReducer забезпечує обчислення інтервальних функцій типовості й центрів кластерів із централізованим збереженням даних у форматі JSONB для подальшого використання в DSS. Новизна полягає у поєднанні автоматизованої параметричної оптимізації PCM за допомогою GA-Optimizer та комплексної оцінки якості кластеризації через QualityEvaluator;

- комбіновані алгоритми FCM+GA та PCM+GA, у яких автоматизовано визначення центрів кластерів, параметрів фазифікації та ступенів приналежності, що сприяє підвищенню точності, збіжності та обчислювальної ефективності оцінювання технічного стану за наявності шумових і неповних даних;

- науково-прикладну реалізацію мікросервісної архітектури з REST-взаємодією, яка забезпечує централізоване збереження результатів, інтеграцію з DSS, BIM-моделями та IoT-потокami, а також підтримує модульність, масштабованість і відтворюваність обчислювальних експериментів у процесі прийняття інженерних рішень.

Набули подальшого розвитку:

- науково-методичні засади побудови інформаційної технології оцінювання технічного стану будівельних об'єктів в умовах невизначеності, які, на відміну від наявних BIM-, SHM- та DSS-підходів, ґрунтуються на комплексному застосуванні нечіткої кластеризації, моделей нечітких множин першого і другого типів та інтервальних форм представлення знань. Запропонований підхід забезпечує формалізацію перехідних станів конструкцій, узгодження інструментальних вимірювань з експертними оцінками та адаптивне опрацювання різномірних моніторингових даних, формуючи теоретичну основу для створення універсального програмного інструментарію кластерного аналізу в межах спеціальності 126 «Інформаційні системи та технології»;

- комбіновані методи інтелектуальної кластеризації FCM+GA та PCM+GA у складі інтегрованої технології Clustering4Build, що передбачають

генетичну оптимізацію центрів кластерів, параметрів фазифікації та ступенів належності на основі схем кодування хромосом, відомих із задач еволюційної нечіткої кластеризації. На відміну від цих підходів, розроблених для типової нечіткої (FCM) кластеризації, у роботі запропоновано поширення генетичної оптимізації на можливісну (PCM) кластеризацію на основі інтервальних нечітких множин другого типу, а також інтеграцію в архітектуру, що реалізує замкнений цикл «оптимізація — аналіз — рекомендація — візуалізація», інтегрується з PostgreSQL, BIM-моделями та потоками IoT-даних і забезпечує підвищення швидкодії, стійкості збіжності та точності оцінювання технічного стану будівель.

Практичне значення одержаних результатів дисертації полягає у створенні інформаційної технології Clustering4Build, що забезпечує інтелектуальну кластеризацію технічного стану будівельних об'єктів за неоднорідними, неповними та шумовими даними в реальному часі. Розроблені модулі GA-Optimizer, TypeReducer, QualityEvaluator та DSS дозволяють автоматизовано підбирати оптимальні параметри нечітких кластеризаторів, виконувати оцінку якості кластеризації та візуалізацію результатів. Це підвищує точність, стійкість та відтворюваність оцінки технічного стану, забезпечує підтримку прийняття рішень у реальних умовах експлуатації і інтеграцію з BIM-, SHM- та DSS-системами. Використання генетичних алгоритмів і інтервальних нечітких множин другого типу дозволяє адаптувати систему до різних типів будівельних об'єктів, уникати локальних мінімумів і забезпечувати формально обґрунтовану обробку невизначеної інформації. Технологія дозволяє здійснювати багатокритеріальну класифікацію об'єктів, прогнозувати їх технічний стан та формувати рекомендації щодо експлуатації та обслуговування, що робить її ефективним інструментом для комплексного моніторингу та управління безпекою будівельних конструкцій.

Запропонований підхід сприяє вдосконаленню процедур проходження технічних обстежень шляхом автоматизації оброблення результатів інструментальних вимірювань та скорочення часу їх експертного аналізу. Це

забезпечує зменшення тривалості обстежень без втрати якості оцінювання, підвищує достовірність висновків щодо фактичного технічного стану будівель і споруд. Крім того, впровадження технології покращує інтерпретацію результатів технічного діагностування завдяки формалізації критеріїв оцінювання та узгодженню багатопараметричних показників у єдиному аналітичному середовищі. Практичне впровадження результатів дисертаційного дослідження підтверджено відповідними довідками про впровадження на ТОВ «Марстон-Груп», ТОВ «Альфа-Сервіс», Інститут місцевого розвитку.

Публікації. Основні положення, наукові результати та елементи новизни дисертаційної роботи відображено у 11 наукових публікаціях, серед яких 6 статей опубліковано у фахових виданнях України категорії «Б», а також 5 матеріалів доповідей у збірниках наукових і науково-практичних конференцій.

Апробація результатів дослідження. Ключові результати та висновки дисертації були представлені на 5 всеукраїнських і міжнародних науково-практичних конференціях, де отримали позитивну оцінку наукової спільноти.

Особистий внесок здобувача у здобутті наукових і практичних результатів дисертаційної роботи являє собою самостійну наукову діяльність, яка відображає його оригінальні ідеї, концепції та розробки, що сприяли розв'язанню поставлених наукових завдань. Значна частина ключових результатів отримана безпосередньо автором, що підтверджується дев'ятьма публікаціями, при цьому всі використані джерела належним чином цитовані. До складу особистого внеску здобувача у науковому опрацюванні обраної тематики входять такі результати:

1. Комплексна інформаційна технологія Clustering4Build, призначена для оцінки технічного стану будівельних об'єктів у умовах невизначеності. Вона поєднує інтервальні нечіткі множини другого типу, алгоритми FCM і РСМ, а також генетичну оптимізацію параметрів у єдиному адаптивному середовищі. Створено програмні модулі GA-Optimizer, TypeReducer та

QualityEvaluator, які забезпечують автоматичне налаштування фазифікаторів, центрів кластерів і параметрів «ширини зони», а також обґрунтований вибір моделей за показниками якості кластеризації. Результати апробації підтвердили підвищення точності класифікації, стабільності збіжності алгоритмів та відтворюваності результатів навіть при наявності шумових і неповних даних.

2. Мікросервісна архітектури для інформаційного середовища технічного моніторингу з REST-взаємодією, що забезпечує модульність, масштабованість та інтеграцію з DSS, BIM-моделями, IoT-потокami і базою даних PostgreSQL (JSONB). Реалізовано замкнений цикл «оптимізація – аналіз – рекомендація – візуалізація», який дозволяє автоматизувати кластеризацію та формування інженерних рішень у режимі реального часу. Така структура гарантує прозорість обчислювальних експериментів, аудит результатів та практичну підтримку процесів прийняття рішень щодо безпечної експлуатації будівельних об'єктів.

Структура та обсяг дисертації. Побудова дисертаційної роботи визначена послідовністю розв'язання поставлених наукових завдань, етапами проведеного дослідження та логікою викладення одержаних результатів. Дисертація включає анотації, перелік публікацій автора за темою дослідження, зміст, вступ, чотири розділи основної частини, загальні висновки, список використаних джерел і додатки. Повний обсяг роботи становить 285 сторінку, з них 237 сторінок основного тексту, 21 таблиця, 103 рисунка та 165 використаних джерел.

РОЗДІЛ І. ОГЛЯД І АНАЛІЗ ПІДХОДІВ ДО ВИРІШЕННЯ ЗАДАЧІ ОЦІНКИ І КЛАСИФІКАЦІЇ ТЕХНІЧНОГО СТАНУ ОБ'ЄКТІВ БУДІВНИЦТВА В ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЯХ

1.1. Особливості задач оцінювання технічного стану об'єктів будівництва як інформаційної проблеми

У сучасному містобудуванні якість будівництва та експлуатації безпосередньо визначає довговічність і безпеку будівель і споруд, а також рівень фінансових витрат на їх утримання, ремонт і реконструкцію. Як показують численні дослідження, дефекти, допущені на етапах проектування, будівництва або експлуатації, призводять до суттєвого зростання експлуатаційних витрат і підвищення ризику аварійних ситуацій [1, 20]. У зв'язку з цим оцінювання технічного стану об'єктів будівництва набуває системного характеру та виходить за межі суто інженерної задачі, трансформуючись у складну інформаційну проблему. Таким чином, оцінювання технічного стану об'єктів будівництва доцільно розглядати не як окрему інженерну процедуру, а як задачу побудови та функціонування інформаційної системи, що забезпечує збір, інтеграцію, формалізацію та інтелектуальну обробку багатовимірних даних з метою підтримки прийняття управлінських рішень.

Методика та обсяг обстеження об'єктів будівництва залежать від передбачуваного сценарію їх подальшої експлуатації, зокрема від наявності або відсутності збільшення навантажень, реконструкції, надбудови чи зміни функціонального призначення. Напрямок і глибина обстеження визначаються також зовнішніми проявами експлуатаційних факторів, такими як деформації несучих конструкцій, підвищена вологість, наявність води в підвальних приміщеннях, деградаційні процеси матеріалів. При цьому обов'язковим є аналіз інженерно-геологічних умов, стану ґрунтів основи та оцінка впливу навколишнього середовища з прогнозуванням можливих наслідків [3, 4, 21]. Основними документами для обліку технічного стану об'єкта будівництва є паспорт та технічний журнал експлуатації. Паспорт містить дані про проект

будівлі з урахуванням змін під час зведення, а технічний журнал слугує для контролю за експлуатацією і ремонтом. Під час обстеження будівель складаються зведені відомості про дефекти та пошкодження конструкцій, а також обсяги робіт для їх усунення. Важливо також складати пояснювальну записку з рекомендаціями щодо ремонту або посилення конструкцій. Оцінка технічного стану включає не лише визначення зносу, але й аналіз дефектів, їх причин та необхідних витрат для приведення об'єкта в експлуатаційний стан.

Для ефективної оцінки технічного стану об'єктів будівництва, зокрема для визначення потреби у ремонті або купівлі-продажу, необхідна детальна інформація про стан будівлі та її конструкцій. Аналіз проектної документації може лише загально оцінити характеристики об'єкта, оскільки вона часто не фіксує відхилень від проекту, які сталися в процесі будівництва чи експлуатації. Найбільш достовірним джерелом є натурні обстеження, виконані фахівцями за чіткими методичними принципами.

Однією з умов проведення таких досліджень є точне визначення функціонального призначення об'єкта та врахування змін навантажень на його конструкції, а також оновлення нормативних вимог з моменту проектування. Важливим є також врахування природно-кліматичних умов району та їх зміни через техногенну діяльність.

Обстеження включає оцінку не лише самого об'єкта, а й його інженерних систем, транспорту, впливу сусідніх об'єктів і екологічної ситуації. В результаті обстеження порівнюються фактичні дані з нормативними вимогами, а також визначається доцільність збереження або реконструкції об'єкта з урахуванням необхідних робіт.

Основні роботи натурального обстеження включають оцінку стану фундаментів, несучих і огорожувальних конструкцій, експлуатаційних середовищ, а також ремонтпридатності об'єкта та економічної доцільності ремонту і реконструкції.

Натурні дослідження несучих і огорожувальних конструкцій включають візуальний огляд, інструментальні вимірювання та лабораторні

дослідження зразків. Визначають ступінь корозії і механічного пошкодження матеріалу, а також агресивний вплив на конструкції. Проводяться заміри фактичних розмірів елементів і визначення навантажень, що діють на конструкції. Особлива увага приділяється монтажним і стиковим з'єднанням, а також стану гідроізоляції.

Оцінка доцільності ремонту або реконструкції включає визначення терміну експлуатації, обсягу ремонтних робіт та можливості реконструкції. Конструкції поділяються на непридатні для ремонту (аварійні) і ремонтпридатні, для яких оцінюється складність ремонту.

Обстеження технічного стану об'єктів є важливим для визначення страхового ризику, зокрема в умовах експлуатації. Особливо це стосується будівель на карстових ґрунтах, де необхідно проводити заходи для запобігання карстових деформацій, зокрема шляхом глибокого закладання опор або цементації карстових порід.

Особливу увагу слід звернути на руйнування стін у приміщеннях з підвищеною вологістю та на фасадах цегляних будівель. Це проявляється в осипанні матеріалу та зміні форми цегли.

Оцінка технічного стану залізобетонних конструкцій включає виявлення тріщин, деформацій, корозії арматури та дефектів зварних сполучень, що знижують міцність і стійкість конструкцій до зовнішніх впливів. Особлива увага приділяється вузловим з'єднанням, оскільки дефекти часто виникають саме в цих зонах. При обстеженні необхідно врахувати зміни навантажень через реконструкцію, погану експлуатацію чи зміни в матеріалах, що можуть призвести до аварійних ситуацій. Дефекти в монолітних і збірних конструкціях, такі як зменшення товщини захисного шару або корозія арматури, впливають на довговічність і безпеку об'єкта. Оцінка дефектів, зокрема тріщин, має включати динамічну перевірку їх стабільності та впливу води, що часто є основною причиною руйнувань.

Металеві конструкції використовуються в висотних та одноповерхових будівлях, де пошкодження металу спричиняється вологістю, агресивним

середовищем, неправильним монтажем та експлуатацією. Основні дефекти включають корозію, викривлення, вигини та деформації елементів, що знижують міцність конструкцій. Пошкоджені конструкції потребують періодичного очищення та фарбування для зупинки корозії, а також обстеження вузлових з'єднань та сполучень, щоб уникнути аварійних ситуацій.

Дерев'яні конструкції схильні до біологічного руйнування через порушення тепло-вологісного режиму та гідрогеологічних умов. Проникнення води сприяє розвитку грибкових уражень деревини, що веде до руйнувань. При обстеженні необхідно ретельно перевіряти всі елементи конструкцій, особливо в місцях можливого пошкодження, і вжити заходів для попередження подальшого руйнування.

Фасади будівель часто пошкоджуються через протікання води, погану герметизацію швів та порушення зчеплення штукатурки з бетоном, що може призвести до відшарування покриття. Оцінка стану оздоблення включає виявлення дефектів на фасадах і внутрішніх стінах, перевірку зчеплення штукатурки і виявлення проблем з вологістю та теплоізоляцією.

Підлоги, особливо з керамічної плитки або асфальтобетону, потребують оцінки на наявність тріщин та вибоїн, які можуть бути причиною пошкоджень покриття. Важливо забезпечити гідроізоляцію зруйнованих ділянок та дотримання технології укладання для відновлення покриття.

Стан покрівель перевіряється на наявність протікань, пошкоджених місць, таких як тріщини на шифері або металевих листах. Важливо також перевірити якість покрівельних матеріалів, герметичність швів та відповідність профілю покрівлі для забезпечення належного стоку води.

З позицій інформаційних технологій оцінювання технічного стану характеризується необхідністю обробки багатовимірних, неоднорідних і часто неповних даних, що надходять з різних джерел. До таких джерел належать результати інструментальних вимірювань, дані лабораторних випробувань, експертні оцінки, а також експлуатаційна та середовищна інформація. У

роботах з моніторингу технічного стану підкреслюється, що такі системи фактично працюють з великими масивами даних, для яких характерні шум, похибки вимірювань і часові варіації параметрів [5, 6]. У сучасних роботах підкреслюється, що відсутність уніфікованих підходів до подання та узгодження таких даних істотно ускладнює їх аналіз і порівняння, особливо при довготривалому моніторингу великої кількості об'єктів [20, 22, 26]. У межах інформаційних технологій оцінки технічного стану доцільно розглядати повний інформаційний конвеєр: збір первинних даних, їх валідацію, нормалізацію, фазифікацію, інтелектуальну обробку та інтерпретацію результатів. Такий підхід дозволяє перейти від фрагментарного аналізу окремих показників до системної обробки даних у межах єдиної інформаційної технології.

Особливою рисою задач технічного моніторингу є наявність різних видів невизначеності: вимірювальної, модельної, інформаційної та експертної. Як зазначається у [7, 8, 24], класичні детерміновані або ймовірнісні моделі не завжди дозволяють адекватно відобразити розмиті межі між станами об'єктів, особливо у випадках, коли оцінювання базується на лінгвістичних або бальних шкалах. Це зумовлює необхідність застосування моделей, здатних формально враховувати нечіткість і невизначеність на всіх етапах обробки даних.

При аналізі множин об'єктів будівництва (наприклад, у межах міської забудови або портфеля експлуатації) важливою задачею є класифікація та групування об'єктів з метою пріоритизації управлінських рішень. У наукових роботах показано, що методи кластеризації є ефективним інструментом для автоматизованого виявлення структур у даних технічного моніторингу та формування класів об'єктів зі схожим технічним станом [9, 10, 27]. При цьому використання жорстких методів кластеризації може призводити до втрати інформації про проміжні або перехідні стани об'єктів.

У зв'язку з цим значного поширення набули методи нечіткої кластеризації, зокрема алгоритм Fuzzy C-Means, які дозволяють описувати ступінь належності об'єкта до кількох кластерів одночасно. Разом із тим у

випадках, коли невизначеність стосується не лише належності об'єкта до кластера, але й параметрів самих функцій приналежності, доцільним є застосування інтервальних нечітких множин другого типу. У роботах [11–13, 25] обґрунтовано, що такі моделі забезпечують підвищену стійкість до варіацій вхідних даних та дозволяють формалізувати «невизначеність щодо нечіткості» за рахунок використання інтервальних функцій приналежності.

Останніми роками спостерігається активний розвиток інформаційних технологій моніторингу технічного стану будівель і споруд, орієнтованих на використання методів інтелектуального аналізу даних та машинного навчання. У сучасних дослідженнях підкреслюється, що класичні підходи до оцінювання технічного стану, засновані на детермінованих моделях і жорсткій класифікації, не забезпечують достатньої робастності в умовах неповноти та мінливості вхідної інформації [14, 15, 18, 19, 28]. У зв'язку з цим дедалі більшої уваги набувають нечіткі та гібридні методи аналізу, здатні враховувати невизначеність різної природи.

У роботах останніх років показано ефективність застосування нечіткої та можливої кластеризації для задач оцінювання технічного стану будівельних конструкцій і систем структурного моніторингу. Зокрема, у [16, 17] доведено, що використання інтервальних нечітких множин другого типу дозволяє підвищити стійкість результатів кластеризації до шумів і похибок вимірювань, а також забезпечує коректніше формування груп об'єктів зі складною та неоднорідною структурою даних. Це є особливо важливим для задач міського моніторингу, де об'єкти істотно відрізняються за віком, конструктивними схемами та умовами експлуатації.

Крім того, у сучасних публікаціях акцентується увага на необхідності інтеграції алгоритмів кластеризації в сервісно-орієнтовані та мікросервісні архітектури інформаційних систем, що забезпечує масштабованість, повторюваність експериментів і можливість обробки великих обсягів даних у реальному часі [18, 19]. Такий підхід розглядається як перспективний напрям

розвитку інформаційних технологій технічного моніторингу та підтримки прийняття рішень у будівельній галузі.

Таким чином, задача оцінювання технічного стану об'єктів будівництва в умовах невизначеності є багатокомпонентною інформаційною проблемою, що потребує інтеграції методів збору, зберігання та інтелектуальної обробки даних. Це обґрунтовує необхідність розроблення спеціалізованої інформаційної технології, орієнтованої на автоматизовану кластеризацію, формалізований облік невизначеності та підтримку прийняття рішень щодо експлуатації, ремонту і реконструкції об'єктів будівництва.

1.2. Джерела та види невизначеності в інформаційних системах технічного моніторингу

Інформаційні системи технічного моніторингу об'єктів будівництва функціонують в умовах суттєвої невизначеності, яка обумовлена як фізичною природою процесів деградації конструкцій, так і особливостями формування та обробки даних. У сучасних наукових дослідженнях підкреслюється, що невизначеність є невід'ємною властивістю систем моніторингу складних інженерних об'єктів і має враховуватися на всіх етапах інформаційного конвеєра — від збору даних до підтримки прийняття рішень [7, 8, 24].

Одним із основних джерел невизначеності є вимірювальна невизначеність, пов'язана з похибками сенсорів, нестабільністю умов вимірювань, обмеженою роздільною здатністю приладів та впливом шумів. Для систем інструментального контролю будівельних конструкцій характерними є як випадкові, так і систематичні похибки, що призводить до варіацій результатів вимірювань навіть за незмінного технічного стану об'єкта [5, 6]. Унаслідок цього первинні дані моніторингу потребують попередньої обробки, фільтрації та оцінки достовірності.

Важливу роль відіграє інформаційна невизначеність, яка виникає через неповноту або фрагментарність даних. На практиці не всі елементи об'єктів будівництва можуть бути обстежені з однаковою періодичністю та деталізацією, а результати моніторингу різних періодів часто мають різний склад ознак. Як зазначається у працях [20, 22, 26], відсутність уніфікованих

форматів подання даних і стандартів їх інтеграції ускладнює порівняльний аналіз результатів та знижує ефективність довготривалого моніторингу.

Ще одним суттєвим видом є модельна невизначеність, зумовлена спрощенням математичних моделей, що використовуються для інтерпретації результатів обстежень. Реальні процеси деградації будівельних конструкцій є нелінійними, багатофакторними та залежать від великої кількості випадкових впливів. Тому будь-яка модель технічного стану є лише наближенням до реального процесу, а її параметри можуть змінюватися в часі [3, 4, 24].

Окрему групу становить експертна невизначеність, яка виникає при використанні бальних або лінгвістичних оцінок технічного стану. Навіть за наявності нормативних документів та методичних рекомендацій результати експертного оцінювання залежать від досвіду, кваліфікації та суб'єктивного сприйняття фахівця. У роботах [21, 25] показано, що різні експерти можуть по-різному інтерпретувати один і той самий стан конструкції, що призводить до розмитості меж між категоріями технічного стану.

У контексті інформаційних технологій важливою є також алгоритмічна невизначеність, яка пов'язана з вибором методів обробки даних, параметрів алгоритмів та критеріїв прийняття рішень. Зокрема, при застосуванні методів кластеризації результати можуть істотно залежати від вибору кількості кластерів, метрики відстані, параметрів фазифікації та критеріїв зупинки алгоритму [9, 10, 27]. Це потребує використання адаптивних і робастних підходів до налаштування алгоритмів аналізу.

Важливим чинником формування невизначеності в інформаційних системах технічного моніторингу є відсутність єдиного підходу до стандартизації процедур оцінювання технічного стану об'єктів будівництва. Проблеми, пов'язані з використанням різних стандартів, методик та шкал оцінювання, істотно ускладнюють порівняння результатів обстежень, отриманих у різний час або різними організаціями, що підтверджується як вітчизняними, так і зарубіжними дослідженнями [20, 22, 29, 30]. У таких умовах виникає потреба в застосуванні комплексних інформаційних технологій, здатних забезпечити уніфікацію подання даних і відтворюваність результатів моніторингу.

Процеси обстеження і моніторингу технічного стану об'єктів будівництва супроводжуються формуванням значних обсягів різнорідних даних, що надходять у систему в умовах невизначеності. Джерелами такої невизначеності є неповнота та неточність вимірювань, використання різних шкал оцінювання, суб'єктивність експертних суджень, а також обмеження чинної нормативно-методичної бази [21, 24, 26, 31]. Як зазначається у сучасних роботах, ці фактори істотно впливають на якість інформаційного забезпечення систем технічного моніторингу та ускладнюють автоматизовану інтерпретацію результатів [32].

Невизначеність виникає також унаслідок різних підходів до оцінювання окремих елементів будівельних конструкцій, відсутності єдиних критеріїв визначення значущості елементів моніторингу та складності прогнозування розвитку дефектів у часі. Додатковим ускладнювальним чинником є різнотипність об'єктів будівництва, які відрізняються конструктивними схемами, матеріалами, віком та умовами експлуатації [23, 27, 33]. У таких умовах навіть коректно виконані інструментальні обстеження не завжди дозволяють однозначно інтерпретувати технічний стан об'єкта без залучення додаткових аналітичних моделей.

У ряді досліджень підкреслюється, що класичні детерміновані методи обробки інформації виявляються недостатньо ефективними для задач технічного моніторингу, оскільки вони не враховують розмитий та інтервальний характер вхідних даних [7, 8, 34]. Це зумовлює необхідність застосування моделей і методів, здатних формалізувати нечіткість, інтервальність та ймовірнісний характер інформації в межах сучасних інформаційних систем.

Саме тому в останні роки активно досліджуються нечіткі, інтервальні та гібридні підходи до аналізу даних технічного моніторингу, які дозволяють інтегрувати вимірювальні та експертні дані, зберігати інформацію про рівень невизначеності та підвищувати робастність прийнятих рішень [11–13, 25, 35]. Використання таких підходів створює методологічне підґрунтя для побудови інформаційних технологій оцінки технічного стану об'єктів будівництва,

орієнтованих на підтримку прийняття рішень в умовах неповноти та суперечливості даних.

Сучасні дослідження в галузі інформаційних систем технічного моніторингу вказують на доцільність застосування нечітких та інтервальних моделей для інтегрального врахування різних видів невизначеності. Зокрема, використання інтервальних нечітких множин другого типу дозволяє формалізувати невизначеність не лише на рівні належності об'єкта до класу, але й на рівні самих функцій приналежності, що є принципово важливим для задач класифікації технічного стану в умовах суперечливих і неповних даних [11–13, 25].

Таким чином, інформаційні системи технічного моніторингу об'єктів будівництва функціонують в умовах комплексної невизначеності, що має багатокomпонентну природу. Ефективне розв'язання задач оцінювання технічного стану потребує застосування спеціалізованих інформаційних технологій, які поєднують механізми контролю якості даних, адаптивні алгоритми обробки та робастні моделі представлення знань. Це створює методологічне підґрунтя для використання нечіткої та інтервальної кластеризації як основи сучасних систем технічного моніторингу.

1.3. Аналіз інформаційних технологій та методів інтелектуальної обробки даних для задач технічного моніторингу

Оцінка технічного стану будівельних конструкцій полягає у визначенні ступеня ушкодження, категорії стану і можливості подальшої експлуатації конструкцій за прямим або зміненим функціональним призначенням. Процес включає порівняння гранично допустимих і фактичних значень міцності, стійкості, деформації та експлуатаційних характеристик конструкцій, визначених за результатами візуальних, інструментальних обстежень і перевірочних розрахунків. Критерії оцінки поділяються на два типи: ті, що характеризують несучу здатність і здатність до деформації, та ті, що оцінюють

експлуатаційну придатність. Оцінка враховує пошкодження, дефекти і впливи, що визначають необхідність ремонту або реконструкції.

З позицій розробники інформаційної системи чи технології задача оцінювання технічного стану об'єктів будівництва розглядається як задача обробки, інтеграції та інтелектуального аналізу багатовимірних, різнорідних і невизначених даних. При цьому ключовим є не лише отримання окремих оцінок, а побудова формалізованої інформаційної моделі, яка забезпечує автоматизовану класифікацію, кластеризацію та підтримку прийняття рішень.

У задачах оцінки технічного стану об'єктів будівництва часто застосовуються методи класифікації та кластеризації, що дозволяють групувати об'єкти за їх властивостями, вираженими через оцінки за критеріями або елементами моніторингу. Вибір групування залежить від властивостей об'єктів, що визначаються їх характеристиками (оцінками). У формально-алгоритмічних методах використовуються технології математичної теорії розпізнавання, класифікації та кластерного аналізу, де кожен об'єкт описується набором властивостей. Такі методи дозволяють вирішувати задачі класифікації, діагностики, прогнозування, структурної ідентифікації та побудови аналітичних описів множин об'єктів. Об'єкти представляються у вигляді точок у метричному просторі, де відстань між ними визначається для оцінки їх близькості. Вибір простору для представлення об'єктів здійснюється за допомогою евклідової метрики.

Класифікація об'єктів у інформаційних технологіях оцінки технічного стану будівельних конструкцій є одним із базових інструментів інтелектуальної обробки даних, що забезпечує формалізоване групування об'єктів за сукупністю їх технічних характеристик, поданих у вигляді критеріїв або елементів моніторингу [55, 56]. Застосування класифікації дозволяє впорядкувати множину об'єктів, зменшити складність аналізу багатовимірних даних та створити передумови для автоматизованої інтерпретації технічного стану в системах моніторингу [57].

У межах інформаційних технологій технічного моніторингу розрізняють пряму та непряму класифікацію. Пряма класифікація передбачає віднесення

об'єктів до наперед визначених класів технічного стану на основі заданого набору ознак і є ефективною за відносно невеликої кількості об'єктів та стабільної структури критеріїв оцінювання [58]. Водночас такий підхід вимагає ретельного вибору інформативних ознак, здатних забезпечити однозначне розрізнення класів технічного стану [59].

Непряма класифікація ґрунтується на формуванні класів шляхом аналізу комбінацій ознак та їх значень і дозволяє враховувати складні залежності між параметрами технічного стану об'єктів [60, 61]. Однак зі збільшенням кількості критеріїв і можливих значень різко зростає число потенційних класів, що ускладнює інтерпретацію результатів і потребує застосування додаткових методів відбору та узагальнення ознак [62].

З формальної точки зору процедура класифікації в інформаційних технологіях описується у вигляді системи вирішальних правил логічного типу «ЯКЩО <умова>, ТО <рішення>» [63]. При прямій класифікації умови задаються переліком ознак або характеристик, типових для певного класу. У випадку непрямої класифікації умови формуються як відношення між ознаками або їх значеннями, що дозволяє враховувати багатовимірний і часто нечіткий характер інформації, отриманої в процесі технічного моніторингу [64, 65]. Результатом застосування таких правил є віднесення кожного об'єкта до відповідного класу технічного стану, що створює основу для подальшого використання методів кластеризації, прогнозування та підтримки прийняття управлінських рішень [66].

Класифікація та кластеризація об'єктів технічного моніторингу є одним із ключових інструментів автоматизації процесів оцінювання технічного стану будівель і споруд, а також виявлення об'єктів, що потребують поточного або капітального ремонту чи реконструкції. У сучасних умовах зростання обсягів моніторингових даних та ускладнення структури об'єктів будівництва традиційні експертні підходи виявляються недостатньо ефективними, що зумовлює необхідність застосування інтелектуальних інформаційних технологій [35–37].

Сучасні інформаційні технології оцінки технічного стану ґрунтуються на методах інтелектуального аналізу даних, які забезпечують автоматизовану

обробку великих масивів різномірної інформації, що формується в результаті інструментальних вимірювань, лабораторних досліджень та експертного оцінювання. Особливе місце серед таких методів займають алгоритми класифікації та кластеризації, що дозволяють групувати об'єкти за подібністю їх технічних характеристик без попереднього задання жорстких правил прийняття рішень [38–40].

Застосування кластерного аналізу в системах технічного моніторингу дає змогу виявляти приховані закономірності у даних, формувати однорідні групи об'єктів за рівнем технічного стану та визначати пріоритети щодо виконання ремонтно-відновлювальних заходів. Це є особливо важливим в умовах обмежених фінансових ресурсів та необхідності оперативного прийняття управлінських рішень на основі неповної або суперечливої інформації [41–43].

У роботах останніх років активно досліджуються підходи, що поєднують методи машинного навчання, статистичного аналізу та нечіткої логіки для розв'язання задач технічного моніторингу. Зокрема, використовуються нейронні мережі, методи ансамблевого навчання, самоорганізовані карти, а також гібридні моделі, які дозволяють враховувати складний нелінійний характер деградаційних процесів у будівельних конструкціях [44–46].

Водночас у ряді досліджень зазначається, що суто детерміновані або ймовірнісні моделі не завжди забезпечують адекватну інтерпретацію результатів моніторингу через високий рівень невизначеності вхідних даних. У зв'язку з цим дедалі більшого поширення набувають методи, засновані на нечітких та інтервальних моделях, які дозволяють формалізувати розмитість меж між класами технічного стану та інтегрувати експертні знання у процес автоматизованого аналізу [47–49].

Окрему увагу в сучасних дослідженнях приділено кластеризації на основі нечітких множин другого типу, яка забезпечує підвищену стійкість до варіацій вхідних даних та дозволяє явно враховувати невизначеність самих функцій приналежності. Застосування таких підходів у задачах технічного

моніторингу будівель підтверджує їх ефективність при аналізі множин об'єктів з істотно різною щільністю та об'ємом кластерів [50–52].

Важливим напрямом розвитку інформаційних технологій технічного моніторингу є інтеграція методів інтелектуальної обробки даних у розподілені та мікросервісні архітектури, що забезпечують масштабованість, гнучкість та можливість роботи в режимі реального часу. Такі архітектурні рішення дозволяють поєднувати алгоритмічні модулі кластеризації, оптимізації параметрів і підтримки прийняття рішень у єдину інформаційну технологію [53, 54].

Таким чином, аналіз сучасних інформаційних технологій та методів інтелектуальної обробки даних свідчить про доцільність використання класифікації та кластеризації як базових інструментів автоматизованої оцінки технічного стану об'єктів будівництва. Поєднання кластерного аналізу з нечіткими та інтервальними моделями створює методологічне підґрунтя для розробки ефективних інформаційних технологій технічного моніторингу в умовах невизначеності.

У задачах технічного моніторингу один і той самий об'єкт іноді представлений кількома реалізаціями з різними значеннями ознак (результати повторних вимірювань або оцінки різних експертів), що ускладнює формування єдиного вектора ознак і потребує додаткової обробки (агрегації, усереднення або відбору представницького значення) перед виконанням кластеризації [70, 74, 75]. Надалі об'єкти технічного стану розглядаються як точки q -вимірного простору ознак — вектори кількісних, якісних та інтервальних показників, що є стандартним підходом для методів нечіткої (FCM) та можливісної (PCM) кластеризації, використаних у роботі.

Узагальнюючи наведене, можна зробити висновок, що для задач оцінки технічного стану будівельних конструкцій доцільним є застосування методів кластеризації, здатних працювати з багатовимірними, неповними та нечіткими даними, характерними для результатів технічного моніторингу. У зв'язку з цим подальший розгляд зосереджується на аналізі методів кластеризації, що застосовуються в сучасних інформаційних системах, а також на оцінюванні їх

можливостей і обмежень у контексті задач технічного стану об'єктів будівництва.

1.4. Методи кластеризації в інформаційних системах: можливості та обмеження

Інтервальні нечіткі множини другого типу забезпечують підвищену стійкість до варіацій вхідних даних і дозволяють розв'язувати проблему класифікації та моніторингу в умовах невизначеності.

Класичні методи кластеризації, зокрема ієрархічні алгоритми та алгоритми чітких с-середніх, широко застосовуються для аналізу технічного стану об'єктів. Водночас їх використання пов'язане з низкою обмежень, зокрема необхідністю попереднього задання кількості кластерів і чутливістю до шумів у даних.

У даній роботі кластеризація розглядається не як самоціль, а як складова інформаційної технології підтримки прийняття рішень, що забезпечує узагальнення технічного стану об'єктів будівництва на основі інтелектуальної обробки даних. Таким чином, акцент робиться на адаптивності, інтерпретованості та інтегрованості алгоритмів кластеризації в інформаційні системи.

У реальних умовах моніторингу технічного стану об'єктів будівництва дані часто мають розмитий характер, а межі між класами є нечіткими. Це обмежує застосування чітких методів кластеризації та зумовлює потребу в більш гнучких підходах.

Кластерний аналіз (автоматична класифікація, розпізнавання образів без вчителя, чисельна таксономія) займає одне з центральних місць серед методів аналізу даних і являє собою сукупність підходів, методів, алгоритмів, призначених для формування однорідних класів в довільній проблемній області - знаходження деякого розбиття досліджуваної сукупності об'єктів на підмножини відносно подібних, схожих між собою об'єктів.

Вихідним припущенням для виділення таких підмножин - кластерів - служить неформальне припущення про те, що об'єкти, що відносяться до

одного кластеру, повинні мати більшу схожість між собою, ніж з об'єктами з інших кластерів [33].

Необхідність аналізу великих обсягів об'єктивної і суб'єктивної інформації, пов'язаних з задачами різної фізичної природи, що неможливо формалізувати або складно формалізувати, вимагає розвитку нових наукових напрямів, в тому числі прикладної статистики та методів аналізу даних.

Застосування методів прикладної статистики засновано на припущенні про ймовірнісну інтерпретацію інформації, що аналізується, і отриманні за допомогою цих методів закономірностей, що мають стохастичний характер [33].

Кластерний аналіз не використовує апріорних припущень про ймовірнісну природу вихідної інформації і застосовує тільки евристичні міркування про характер і особливості досліджуваної сукупності об'єктів.

Задача кластерного аналізу в інформаційних технологіях оцінки технічного стану об'єктів будівництва полягає в знаходженні деякого теоретико-множинного розбиття (покриття) вихідної множини об'єктів на підмножини - кластери - таким чином, щоб елементи, що відносяться до однієї підмножини, відрізнялися між собою в значно меншій мірі, ніж елементи з різних підмножин [33] .

Знаходження або виявлення кластерів в вихідній множині об'єктів повинно відповідати таким вимогам:

- кожен кластер повинен являти собою концептуально однорідну категорію і містити схожі об'єкти з близькими за значеннями характеристик об'єктами (значеннями ознак, оцінок за критеріями або елементів моніторингу);

- сукупність всіх кластерів повинна бути вичерпною, тобто охоплювати всі об'єкти вихідної множини;

- кластери повинні бути такими, що є взаємовиключними, тобто жоден з об'єктів вихідної множини не повинен одночасно належати двом різним кластерам.

Проблема кластерного аналізу набуває самостійного значення в контексті інтелектуального аналізу даних (Data Mining) [59]. Задача кластеризації подібна до задачі класифікації, є її логічним продовженням, але її відмінність полягає в тому, що класи технічного стану будівельних об'єктів досліджуваної множини заздалегідь не визначені. Кластеризація є описовою процедурою, вона не робить ніяких статистичних висновків, але дає можливість провести розвідувальний аналіз і вивчити «структуру даних» [117, 118]. Існують деякі негласні відмінності в термінах «кластеризація» і «пошук структур даних».

Поняття «кластеризація» означає просторове розділення таксономічних об'єктів, тобто враховує їх близькість в координатних площинах. Поняття «пошук структур даних» несе в собі велике смислове навантаження, позначаючи поділ статистичних або облікових об'єктів на основі деякого кількісного або якісного критерію, семантичної траєкторії або прихованого об'єднуючого початку, дозволяючи об'єднати в кластери об'єкти на перший погляд різноманітні і просторово далекі.

Розрізняють задачі кластерного аналізу (і відповідно алгоритми) з заданою кількістю кластерів і задачі кластерного аналізу з незаданою кількістю кластерів в інформаційних технологіях технічного стану об'єктів будівництва. В останньому випадку оптимальна кластеризація і число кластерів визначається в результаті вирішення єдиного завдання.

В результаті застосування різних алгоритмів кластеризації можуть бути отримані неоднакові результати, що є нормальним явищем і пояснюється результатами роботи алгоритму кластеризації. Знаючи принципи, згідно з якими об'єкти об'єднуються в один кластер, можна в певних межах інтерпретувати результати кожного конкретного алгоритму кластеризації. Вирішення задачі кластеризації принципово неоднозначне, так як не існує однозначно найкращого критерію якості кластеризації. Кластеризація може виконуватися для об'єктів з кількісними (числовими), якісними або змішаними ознаками.

Класифікація і розпізнавання, будучи найбільш простими завданнями Data Mining, відносяться до стратегії «навчання з учителем» [12, 58].

Для вирішення такого роду задач навчальна вибірка в інформаційній технології повинна містити значення як вхідних, так і вихідних (цільових) змінних. Кластеризація є задачею Data Mining, що відноситься до стратегії «навчання без учителя», тобто не вимагає наявності вихідних (цільових) змінних в навчальній вибірці [88].

Найбільш відомими класичними методами кластеризації є ієрархічна кластеризація і кластеризація з використанням базового алгоритму чітких середніх [69, 71, 72, 85, 86, 102, 108].

1.4.1 Ієрархічна кластеризація об'єктів

Ієрархічна кластеризація широко застосовується у задачах оцінки технічного стану будівель, оскільки дозволяє формувати систему вкладених кластерів, що відображає різні рівні подібності об'єктів за сукупністю показників технічного моніторингу [91, 92]. Результати ієрархічного групування зазвичай подаються у вигляді дендрограми, яка наочно демонструє структуру технічного стану будівель і може бути використана для виділення груп з близьким рівнем зносу або пошкодженості. В ієрархічній кластеризації виділяють два основні типи алгоритмів: низхідні (дивізимні), у яких множина об'єктів послідовно розбивається на дрібніші кластери, та висхідні (агломеративні), у яких об'єкти поетапно об'єднуються у все більші кластери. Найбільш поширеними є агломеративні алгоритми, де на початковому етапі кожен об'єкт розглядається як окремий кластер, а відстань між об'єктами визначається за обраною метрикою, зокрема евклідовою $d(x_i, x_j)$. На кожній ітерації два найближчі кластери U і V об'єднуються у новий кластер W , після чого відстань $d(W, S)$ до інших кластерів S обчислюється за правилами ближнього сусіда, далекого сусіда або середнього зв'язку, що зумовлює різні результати кластеризації [31, 91].

Для задач оцінювання технічного стану будівель, дані яких часто формуються на основі багаторазових інструментальних вимірювань або

експертних оцінок, перспективним є поєднання ієрархічної кластеризації з повторюваним або гранульованим поданням об'єктів. Це дозволяє враховувати повторюваність і варіативність показників технічного моніторингу та підвищує інформативність ієрархічної структури кластерів [95].

1.4.2 Алгоритм чітких с-середніх

Алгоритм чітких с-середніх (Hard C-Means, HCM) застосовується для кластеризації об'єктів будівництва, технічний стан яких описується набором кількісних показників, отриманих у процесі інструментального або експертного моніторингу, зокрема показників деформацій, міцності матеріалів, рівня фізичного зносу та результатів неруйнівного контролю [96, 97]. Алгоритм передбачає попереднє задання кількості кластерів c , які інтерпретуються як рівні або категорії технічного стану будівель і споруд [98].

Нехай $X = \{x_1, x_2, \dots, x_n\}$ — множина об'єктів будівництва, що підлягають оцінюванню, а $P = \{p_1, p_2, \dots, p_q\}$ — множина кількісних критеріїв (елементів моніторингу), які характеризують їх технічний стан. Кожному об'єкту x_i ставиться у відповідність вектор оцінок

$$x_i = (x_{i1}, x_{i2}, \dots, x_{iq}), \quad x_{ik} \in \mathbb{R}, \quad (1.1)$$

де x_{ik} — значення k -го показника технічного стану для i -го об'єкта [99].

Подібність між об'єктами будівництва визначається у метричному просторі через відстань між векторами їх технічних характеристик або між об'єктами і центрами кластерів, що інтерпретуються як типові технічні стани конструкцій [100].

При застосуванні алгоритму чітких с-середніх множина об'єктів X розбивається на c підмножин X_j з такими властивостями:

$$X = \bigcup_{j=1}^c X_j, \quad X_j \cap X_k = \emptyset \ (j \neq k), \quad X_j \neq \emptyset \ (j = 1, \dots, c). \quad (1.3)$$

Ці умови відповідають чіткому віднесенню кожного об'єкта будівництва до єдиного класу технічного стану [101].

Для формального опису розбиття вводиться характеристична функція

$$u_{ij} = \begin{cases} 1, & x_i \in X_j, \\ 0, & x_i \notin X_j, \end{cases}$$

на основі якої формується матриця належності $U = \{u_{ij}\}$, що описує c -розбиття множини об'єктів X . Матриця U повинна задовольняти умови:

$$u_{ij} \in \{0,1\}; \quad \sum_{j=1}^c u_{ij} = 1, \quad i = 1, \dots, n; \quad 0 < \sum_{i=1}^n u_{ij} < n, \quad j = 1, \dots, c. \quad (1.4)$$

Ці обмеження забезпечують повне та несуперечливе розбиття об'єктів будівництва за технічним станом [102].

Якість кластеризації оцінюється за допомогою цільової функції розкидання. У випадку використання евклідової метрики вона має вигляд:

$$J = \sum_{j=1}^c \sum_{i=1}^n u_{ij} \|x_i - v_j\|^2, \quad (1.5)$$

де v_j — центр j -го кластера, який інтерпретується як усереднений технічний стан групи будівельних об'єктів [103].

Координати центрів кластерів визначаються як:

$$v_j = \frac{1}{|X_j|} \sum_{x_i \in X_j} x_i, \quad (1.6)$$

де $|X_j|$ — кількість об'єктів, що належать до j -го кластеру [104].

Для додаткової оцінки якості розбиття можуть застосовуватися узагальнені показники, зокрема загальний гіпероб'єм кластерів:

$$FHV = \sum_{j=1}^c \sqrt{\det(F_j)}, \quad (1.7)$$

$$F_j = \frac{\sum_{i=1}^n u_{ij} (x_i - v_j)(x_i - v_j)^T}{\sum_{i=1}^n u_{ij}}, \quad (1.8)$$

що дозволяє враховувати внутрішню варіативність показників технічного стану в межах кожного кластера [105].

Дискретний характер чіткого розбиття призводить до негладкості цільової функції та ускладнює пошук глобального оптимуму, що обмежує застосовність алгоритму чітких c -середніх у задачах технічного моніторингу будівель з перехідними або нечіткими станами конструкцій [106].

1.4.3 Можливісна кластеризація (Possibilistic C-Means)

Поряд з алгоритмом нечітких c -середніх (FCM), що зберігає умову нормування ступенів приналежності (сума приналежностей об'єкта до всіх

кластерів дорівнює одиниці), у задачах аналізу зашумлених і неповних даних набув поширення можливісний підхід до кластеризації — алгоритм Possibilistic C-Means (PCM) [87]. На відміну від FCM, у PCM ступінь приналежності об'єкта до кластера інтерпретується не як відносна частка належності серед інших кластерів, а як міра типовості (сумісності) об'єкта з даним кластером незалежно від інших кластерів, що усуває вплив умови нормування і підвищує стійкість результатів кластеризації до викидів та аномальних вимірювань [87, 116].

Водночас відомим обмеженням класичного PCM-алгоритму є схильність до формування «збіжних» (coincident) кластерів, коли за невеликого початкового наближення центри різних кластерів можуть збігатися в одній області ознакового простору, що знижує інформативність розбиття [116, 128]. Крім того, якість роботи PCM істотно залежить від коректного підбору параметра фаззифікатора (зони нечіткості) для кожного кластера окремо, що ускладнює його застосування на практиці без додаткових механізмів автоматизованого налаштування [123, 130].

У задачах оцінки технічного стану об'єктів будівництва, де вхідні дані характеризуються шумом вимірювань, неповнотою та наявністю потенційних викидів (аварійні або нетипові об'єкти), стійкість PCM до аномальних значень є суттєвою перевагою порівняно з FCM. Разом з тим необхідність ручного підбору параметрів фаззифікації та відсутність механізмів урахування невизначеності самого фаззифікатора зумовлюють доцільність поєднання PCM з методами автоматизованої оптимізації параметрів (розділ 2) та з апаратом інтервальних нечітких множин другого типу для формалізації невизначеності «ширини зони» кожного кластера (розділ 3), що і розглядається в подальших розділах роботи.

1.4.4 Показники якості (валідності) кластеризації

Оскільки задача кластеризації є принципово багатоваріантною (результат залежить від обраного алгоритму, метрики відстані та параметрів), важливою складовою методології є застосування формалізованих показників

якості (валідності) кластеризації, які дозволяють об'єктивно порівнювати альтернативні розбиття та обирати серед них найбільш адекватне [115, 122, 127].

У літературі описано низку індексів валідності кластеризації, серед яких найбільш відомі: індекс Данна (Dunn index), що оцінює співвідношення міжкластерної та внутрішньокластерної відстані; індекс Xie–Beni, який ураховує компактність кластерів відносно мінімальної відстані між їх центрами; силует-аналіз (silhouette), що дозволяє оцінити якість належності кожного окремого об'єкта до свого кластера порівняно із сусідніми [131]. Для нечіткої та можливісної кластеризації додатково застосовуються показники, побудовані на основі геометричних характеристик кластерів у просторі ознак, зокрема показники гіпероб'єму та компактності (сферичності) кластерів, які дозволяють оцінювати не лише розділення між кластерами, а й форму й однорідність кожного окремого кластера [123, 130].

Аналіз наведених підходів показує, що більшість класичних індексів валідності орієнтовані на чіткі або нечіткі (НМТ1) моделі кластеризації і не враховують специфіку інтервального подання функцій приналежності, характерного для інтервальних нечітких множин другого типу. Це зумовлює необхідність адаптації та обґрунтованого вибору показників якості (зокрема гіпероб'єму та індексу сферичності) для задач кластеризації технічного стану об'єктів будівництва в умовах ІНМТ2, що детально розглядається в розділі 3 даної роботи.

1.5. Нечіткі та інтервальні моделі представлення знань в інформаційних технологіях

Існуючі системи моніторингу потребують інтеграції сучасних методів аналізу даних, зокрема кластеризації на основі нечітких множин, для підвищення точності та ефективності оцінки технічного стану.

Застосування апарату нечітких множин дозволяє формалізувати суб'єктивні експертні оцінки та інтервальні значення параметрів технічного

стану. Особливу ефективність у задачах моніторингу демонструють інтервальні нечіткі множини другого типу, які забезпечують підвищену стійкість до варіацій вхідних даних.

Такі моделі дають змогу адекватно враховувати невизначеність і забезпечують більш надійну кластеризацію об'єктів у складних умовах експлуатації.

Вимога однозначного віднесення об'єктів до кластерів є надто жорсткою для задач оцінки технічного стану будівель, які характеризуються слабкою структурованістю, неповнотою даних і наявністю перехідних станів конструкцій. Методи нечіткої кластеризації послаблюють цю вимогу за рахунок введення нечітких кластерів і функцій приналежності, що приймають значення в інтервалі $[0,1]$, дозволяючи об'єктам частково належати кільком кластерам одночасно [107, 108]. Такий підхід є адекватним для опису реальних технічних станів будівель, де межі між категоріями «справний», «обмежено працездатний» та «аварійний» мають розмитий характер.

Концептуальний зв'язок між кластерним аналізом і теорією нечітких множин ґрунтується на тому, що класи технічного стану будівель є нечіткими за своєю природою, а перехід між ними відбувається поступово, а не дискретно [109, 110]. У межах нечіткого кластерного аналізу невизначеність інтерпретується у можливісному сенсі, що дозволяє враховувати кластерну відносність і типовість об'єктів будівництва на основі значень функцій приналежності [111].

Застосування апарату теорії нечітких множин, зокрема поняття лінгвістичної змінної, є доцільним при оцінюванні технічного стану будівельних конструкцій, коли частина показників формується у вигляді експертних або вербальних оцінок (наприклад, «незначні пошкодження», «помірний знос», «критичний стан») [112, 113]. Це дозволяє формалізувати суб'єктивні судження експертів і інтегрувати їх з результатами інструментальних вимірювань у єдиній кластерній моделі [114].

Використання нечіткої кластеризації у поєднанні з повторюваним поданням об'єктів технічного моніторингу дає змогу враховувати повторюваність вимірювань та різноманітність даних, однак потребує значних

обсягів вхідної інформації, що ускладнює етап підготовки даних з точки зору часових і фінансових витрат [115]. Додаткові труднощі виникають при кластеризації об'єктів, що формують кластери з істотно різною щільністю або об'ємом, що є типовим для вибірок будівель з різними умовами експлуатації та строками служби [116].

Суттєвим недоліком відомих алгоритмів нечіткої кластеризації є необхідність багаторазового підбору їх параметрів для отримання адекватних результатів. У зв'язку з цим актуальним є розроблення методів кластеризації технічного стану будівель, які дозволяють автоматизувати пошук оптимальних параметрів і забезпечити стабільні результати з мінімальними часовими витратами. Для розв'язання цієї задачі доцільним є застосування генетичних алгоритмів, які ефективно використовуються для оптимізації параметрів нечітких кластерних моделей [117–120].

Генетичні алгоритми (ГА) — клас еволюційних методів оптимізації, що ґрунтуються на принципах природного добору, схрещування та мутації [117, 118], — набули поширення як інструмент автоматизованого налаштування параметрів нечітких кластерних моделей завдяки здатності ефективно досліджувати багатовимірний простір параметрів без потреби в обчисленні похідних цільової функції [119]. У задачах гібридизації нечіткої кластеризації з ГА параметри моделі (центри кластерів, ступені приналежності, параметри фазифікації) кодуються у вигляді хромосоми, придатність (fitness) якої оцінюється через цільову функцію кластеризації, після чого шляхом ітеративного застосування операторів селекції, схрещування та мутації популяція хромосом еволюціонує в напрямку покращення якості розбиття [117, 119].

Окремої уваги потребує вибір схеми кодування хромосоми. У класичному підході застосовується хромосома **сталोї довжини**, яка кодує параметри для наперед заданої кількості кластерів; такий підхід є обчислювально ефективним, проте вимагає апріорного визначення кількості кластерів [117, 118]. Альтернативою є хромосома **змінної довжини**, яка

дозволяє алгоритму одночасно з оптимізацією параметрів кластерів визначати й саму кількість кластерів у процесі еволюційного пошуку, що є особливо актуальним для задач технічного моніторингу, де кількість категорій технічного стану об'єктів заздалегідь не завжди відома [119].

Незважаючи на доведену ефективність поєднання генетичних алгоритмів з алгоритмом FCM у низці прикладних задач [119], питання систематичного застосування генетичної оптимізації до **можливісної** кластеризації (PCM), зокрема з урахуванням специфіки її параметрів фазифікації та «ширини зони» кожного кластера, а також поєднання генетичної оптимізації з інтервальними нечіткими множинами другого типу, залишається недостатньо дослідженим, що й визначає напрям подальших розробок, представлених у розділах 2 і 3 даної роботи.

Аналіз сучасних інформаційних технологій показує, що наявні підходи до кластеризації технічного стану об'єктів будівництва мають низку суттєвих обмежень, пов'язаних із жорсткою фіксацією кількості кластерів та обмеженим набором елементів моніторингу [120, 121]. Водночас розвиток будівельних матеріалів і методів інструментального контролю зумовлює постійне зростання кількості потенційних показників технічного стану, що ускладнює застосування класичних схем кластеризації [94, 122].

Додаткові труднощі пов'язані з невизначеністю вибору інформативних елементів моніторингу та оптимальної кількості кластерів, що є критичним при прийнятті рішень щодо розподілу обмежених фінансових ресурсів на ремонт і реконструкцію будівель [123]. Складність задачі посилюється різноманітністю вхідних даних, оскільки оцінювання технічного стану конструкцій може здійснюватися як у вигляді кількісних показників (наприклад, відсоток фізичного зносу за результатами інструментальних вимірювань), так і у формі бальних або лінгвістичних експертних оцінок [124, 125]. Це зумовлює необхідність застосування гнучких методів кластеризації, здатних інтегрувати різноманітну інформацію та адекватно відображати реальний технічний стан об'єктів будівництва, при цьому $n_i (i = \overline{1,4})$ - кількість будівельних об'єктів даного типу. При наявності у будівельного

об'єкту «уявного» елементу моніторингу відсоток зношеності по ньому дорівнює 0.

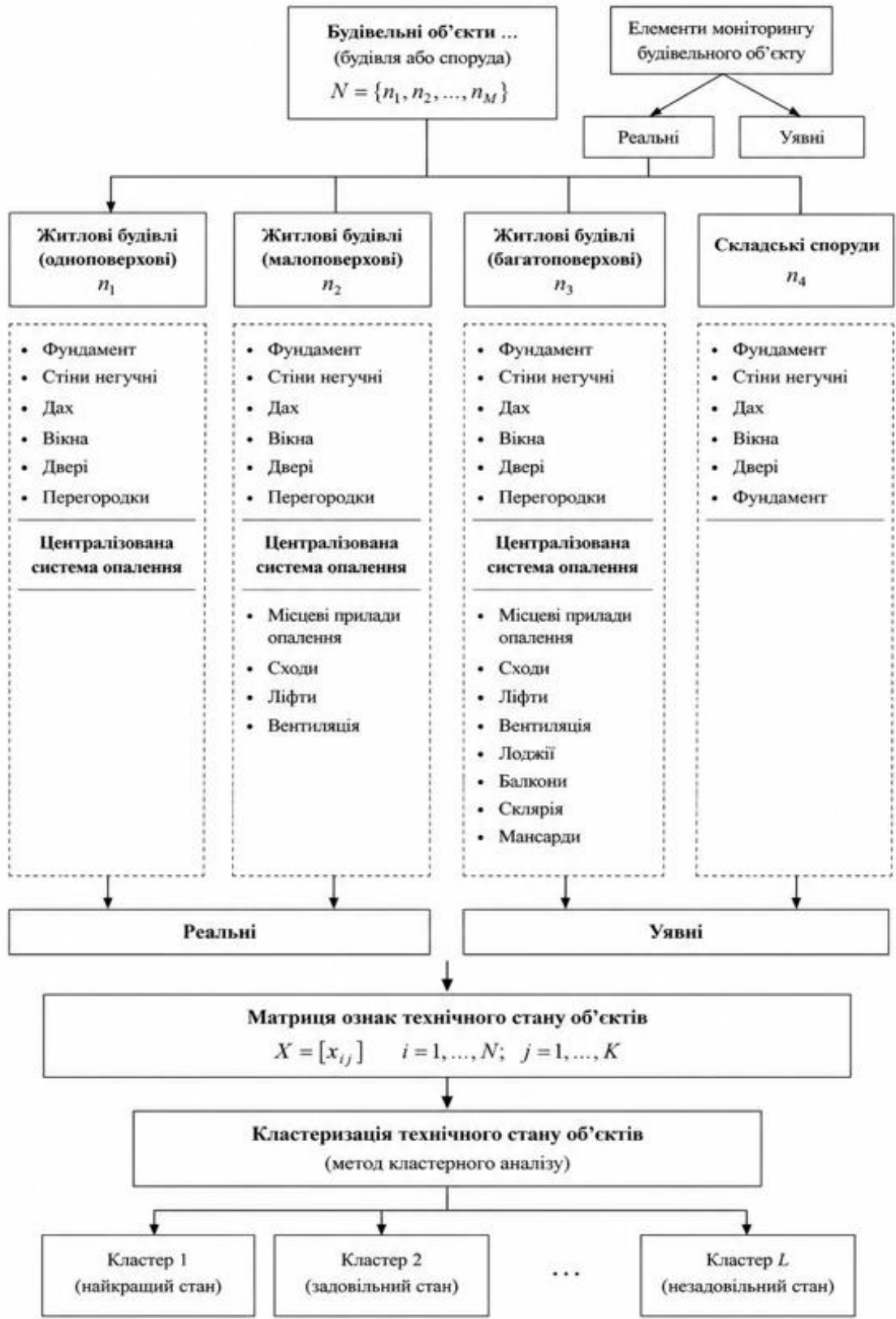


Рис.1.1. Задача кластеризації технічного стану будівельних об'єктів з довільною кількістю елементів моніторингу (Розроблено автором)

Суттєвою проблемою кластеризації технічного стану будівель є різноманітність типів об'єктів, які відрізняються як кількістю елементів моніторингу ці (для малоповерхових і багатоповерхових будівель, складських приміщень тощо), так і чисельністю об'єктів кожного типу пі, [126, 127]. Така неоднорідність ускладнює застосування класичних кластерних алгоритмів, орієнтованих на фіксовану розмірність ознакового простору [73].

Для забезпечення комплексної оцінки та коректної кластеризації всієї сукупності об'єктів будівництва пропонується методика, що ґрунтується на формуванні єдиного повного переліку елементів моніторингу для всієї множини об'єктів [128]. Для кожного типу будівельних об'єктів елементи моніторингу поділяються на реальні (наявні) та умовні (уявні), відсутні в даному типі об'єкта. Оцінювання реальних елементів виконується за результатами інструментальних вимірювань або експертного бального оцінювання, тоді як для умовних елементів значення показників приймаються рівними нулю [129].

Такий підхід дозволяє уніфікувати подання об'єктів незалежно від їх типу та забезпечити кластеризацію всієї множини будівель на основі порівнянних векторів технічного стану. У результаті можливе групування об'єктів з близькими або контрастними оцінками технічного стану в один кластер залежно від обраного алгоритму та критерію якості кластеризації [130].

Застосування загальних принципів кластерного аналізу дає змогу виконувати розбиття як наперед задану кількість кластерів, так і на оптимальну кількість кластерів, визначену за обраним показником якості [81, 131]. Додаткове використання нечітких множин першого та інтервальних нечітких множин другого типу забезпечує адекватну кластеризацію вибірок, що містять кластери різної щільності та об'єму, що є типовим для задач оцінки технічного стану будівель [81, 132, 133].

Запропонована методика створює основу для розроблення сучасних інформаційних систем підтримки прийняття рішень у сфері експлуатації,

ремонту та реконструкції будівель, забезпечуючи підвищення об'єктивності та адекватності рішень за мінімальних часових витрат [134].

1.6. Аналіз сучасних програмних та інформаційних систем оцінки технічного стану

Сучасні програмні та інформаційні системи оцінки технічного стану об'єктів будівництва можна умовно поділити на кілька класів залежно від функціонального призначення та рівня інтелектуальної обробки даних. Найбільш поширеними є системи, орієнтовані на документування результатів обстежень і розрахунок фізичного зносу на основі нормативних методик [135].

До цієї групи належать програмні комплекси типу PlanRadar, Archibus та IBM Maximo, які використовуються для фіксації дефектів, управління технічним обслуговуванням і накопичення експлуатаційних даних протягом життєвого циклу будівель [136–138]. Такі системи забезпечують зручну роботу з базами даних, формування звітів і контроль виконання ремонтних робіт, однак їх аналітичні можливості щодо автоматизованої оцінки технічного стану та кластеризації об'єктів є обмеженими.

Окрему категорію становлять BIM-орієнтовані інформаційні системи, зокрема Autodesk Revit, Navisworks, Bentley iTwin та Tekla Structures, які дозволяють інтегрувати геометричну модель будівлі з атрибутивними даними про конструкції та інженерні системи [139, 140, 141]. Незважаючи на високий рівень візуалізації та можливість накопичення експлуатаційної інформації, більшість BIM-платформ не містить вбудованих механізмів інтелектуальної кластеризації та нечіткого аналізу технічного стану, що обмежує їх застосування для підтримки прийняття рішень в умовах невизначеності.

У сфері структурного моніторингу (Structural Health Monitoring, SHM) застосовуються спеціалізовані програмні системи, такі як SCADA-based SHM systems, StrainSmart, GeoMoS та OptiMon, які забезпечують автоматизований збір даних з датчиків деформацій, переміщень і вібрацій [142–144]. Ці системи ефективні для оперативного контролю окремих параметрів технічного стану,

однак зазвичай орієнтовані на аналіз часових рядів і порогових значень, а не на комплексну кластеризацію об'єктів за багатовимірними показниками.

Окремим напрямом розвитку інформаційних систем технічного моніторингу є перехід від монолітної архітектури програмних комплексів до сервісно-орієнтованих і мікросервісних архітектур, що набули поширення в сучасній розробці розподілених інформаційних систем завдяки модульності, незалежному розгортанню окремих компонентів та можливості горизонтального масштабування [148, 149]. Використання REST-інтерфейсів взаємодії між компонентами системи (модулями збору даних, аналітичної обробки, оптимізації та візуалізації) дозволяє забезпечити технологічну незалежність окремих підсистем та спростити їх інтеграцію із зовнішніми джерелами даних, зокрема BIM-моделями та потоками даних IoT-пристроїв [148, 149].

Аналіз розглянутих вище класів систем (BIM, SHM, DSS) показує, що переважна більшість існуючих рішень реалізована як монолітні або слабко модульні програмні комплекси, що ускладнює їх адаптацію під специфічні задачі кластерного аналізу та інтеграцію нових аналітичних модулів без істотної переробки системи в цілому. Це зумовлює доцільність застосування мікросервісного підходу до побудови інформаційної технології оцінки технічного стану об'єктів будівництва, що забезпечує можливість незалежного розвитку та масштабування окремих аналітичних компонентів (фазифікації, генетичної оптимізації, оцінки якості кластеризації, підтримки прийняття рішень) у межах єдиної інтегрованої архітектури, розробка якої представлена в розділі 4 даної роботи.

Таблиця 1.1. Порівняння сучасних програмних та інформаційних систем оцінки технічного стану *(Розроблено автором на основі [136–145])*

Клас систем	Основне призначення	Типи даних	Аналітичні можливості	Робота з невизначеністю	Кластеризація	Основні обмеження
BIM-системи (Revit, iTwin, Tekla)	Моделювання та візуалізація об'єктів	Геометричні, атрибутивні, експлуатаційні	Переважно описові, сценарні	Відсутня або обмежена	Відсутня	Немає інтелектуального аналізу

						технічного стану
SHM-системи (GeoMoS, SCADA-SHM)	Моніторинг окремих параметрів	Часові ряди, сенсорні дані	Пороговий, статистичний аналіз	Обмежена	Відсутня або локальна	Орієнтація на окремі показники, а не комплексну оцінку
DSS-системи (Maximo, Archibus)	Управління експлуатацією і ремонтами	Табличні, нормативні	Правилова логіка, звіти	Частково (експертні правила)	Відсутня	Жорсткі сценарії, відсутність адаптивності
AI/ML-рішення (DL, ML-моделі)	Діагностика окремих дефектів	Візуальні, числові, великі масиви	Високі (у вузьких задачах)	Опосередковано	Частково	Вузька спеціалізація, слабка інтерпретованість
Запропонована методика	Комплексна оцінка технічного стану	Кількісні, експертні, різномірні	Кластерний, нечіткий, інтегральний аналіз	Повна (HMT1, IHMT2)	Адаптивна	Орієнтація на методичну, а не лише прикладну реалізацію

Останніми роками активно розвиваються інтелектуальні системи оцінки технічного стану, що базуються на методах машинного навчання та аналізу великих даних. До них належать експериментальні та комерційні рішення, які використовують нейронні мережі, нечітку логіку та кластерний аналіз для діагностики пошкоджень і прогнозування деградації конструкцій [145–147]. Проте більшість таких систем орієнтована на вузькі класи задач (наприклад, тріщиноутворення або корозію) і не забезпечує універсальної інтеграції різномірних елементів моніторингу.

Таким чином, аналіз сучасних програмних та інформаційних систем показує, що існуючі рішення або зосереджені на зберіганні та візуалізації даних, або реалізують окремі аналітичні функції без комплексної підтримки кластеризації та роботи з невизначеністю. Це підтверджує доцільність розробки інтегрованої інформаційної технології, яка поєднує методи кластерного аналізу, апарат нечітких множин першого та другого типів і засоби візуалізації результатів для обґрунтованої оцінки технічного стану об'єктів будівництва.

Проведений аналіз сучасних програмних та інформаційних систем показав відсутність універсального підходу, здатного забезпечити комплексну кластеризацію технічного стану об'єктів будівництва в умовах невизначеності та різноманітності даних. У зв'язку з цим виникає необхідність формалізованої постановки науково-прикладної задачі та визначення вимог до інформаційної технології, яка забезпечує адекватну оцінку технічного стану об'єктів будівництва, що і є предметом розгляду в підрозділі 1.7.

1.7. Постановка науково-прикладної задачі та вимоги до інформаційної технології оцінки технічного стану об'єктів будівництва

На основі проведеного аналізу сучасних методів, моделей і програмних засобів оцінки технічного стану будівель і споруд формулюється науково-прикладна задача розробки інтегрованої інформаційної технології оцінки технічного стану об'єктів будівництва в умовах невизначеності, неповноти та різноманітності вихідних даних. Дана задача розглядається як задача розробки методів і моделей інтелектуальної обробки даних та побудови спеціалізованої інформаційної технології для підтримки прийняття рішень.

Сутність задачі полягає у створенні такої інформаційної технології, яка забезпечує формалізоване подання результатів інструментального та експертного моніторингу, їх інтелектуальну обробку і адаптивну кластеризацію об'єктів будівництва з метою класифікації технічного стану та підтримки прийняття управлінських рішень. При цьому кластеризація розглядається як основний інструмент узагальнення багатовимірної інформації про стан об'єктів і виявлення груп об'єктів з подібними або критично різними характеристиками.

Розроблювана інформаційна технологія повинна забезпечувати:

- інтеграцію результатів інструментальних вимірювань і експертних оцінок технічного стану;

- обробку великих обсягів даних різної природи (кількісних, якісних, інтервальних, лінгвістичних);
- адаптивну кластеризацію об'єктів будівництва з використанням нечітких та інтервальних нечітких моделей;
- формування класифікації технічного стану об'єктів (наприклад, справний, задовільний, передаварійний, аварійний);
- підтримку прийняття управлінських рішень щодо експлуатації, ремонту, реконструкції або виведення об'єктів з експлуатації;
- автоматизоване налаштування параметрів моделей кластеризації (центрів кластерів, ступенів приналежності, параметрів фазифікації) за допомогою генетичних алгоритмів, що зменшує залежність результатів від ручного підбору параметрів та скорочує часові витрати на підготовку моделі;
- формалізовану оцінку якості (валідності) отриманих розбиттів на основі кількісних показників (зокрема гіпероб'єму та індексу сферичності кластерів) для обґрунтованого вибору моделі серед альтернативних варіантів кластеризації;
- реалізацію у вигляді модульної мікросервісної архітектури з REST-взаємодією між компонентами (фазифікації, кластеризації, оптимізації, оцінки якості та підтримки прийняття рішень), що забезпечує масштабованість, незалежне розгортання та інтеграцію із зовнішніми системами (BIM, IoT).

У даній дисертаційній роботі об'єкти будівництва розглядаються як складні технічні системи. Об'єктами аналізу є будівлі та споруди (або їх сукупності), технічний стан яких оцінюється за результатами інструментального та/або експертного моніторингу.

У межах розроблюваної інформаційної технології Clustering4Build формалізація об'єкта аналізу здійснюється таким чином:

- об'єкт кластеризації — окрема будівля, споруда або конструктивний елемент будівлі;
- опис об'єкта — вектор кількісних і якісних показників технічного стану (елементів моніторингу), що включає показники фізичного зносу,

результати інструментальних вимірювань (деформації, тріщини, осідання), експертні оцінки стану конструктивних елементів, а також інтегральні техніко-експлуатаційні характеристики;

- множина об'єктів — сукупність будівельних об'єктів, які можуть істотно відрізнятися за щільністю та об'ємом у просторі ознак і характеризуються невизначеністю та неповнотою вихідної інформації.

У контексті математичної моделі кожен об'єкт представляється точкою у багатовимірному просторі ознак, а кластеризація використовується для класифікації технічного стану об'єктів. Результати кластеризації слугують інформаційною основою для прийняття рішень щодо подальшого обстеження, ремонту, реконструкції або обмеження експлуатації будівель і споруд.

Таким чином, об'єкти будівництва в межах дослідження розглядаються як інформаційні об'єкти, кожен з яких описується вектором ознак у багатовимірному просторі даних. Результати кластеризації цих об'єктів інтерпретуються як інформаційні моделі технічного стану, що використовуються у системах підтримки прийняття рішень.

Таблиця 1.2. Вхідні параметри та характеристики об'єктів будівництва
(Розроблено автором)

Група параметрів	Позначення	Назва характеристики	Тип даних	Джерело отримання	Призначення в кластеризації
Ідентифікаційні	ID	Ідентифікатор об'єкта	Ціле	Реєстр об'єктів	Однозначна ідентифікація
	ObjType	Тип об'єкта (будівля / споруда)	Категоріальний	Паспорт об'єкта	Попереднє групування
Геометричні	A	Площа, м ²	Дійсне	Проектна документація	Нормалізація впливу
	H	Висота, м	Дійсне	Проектна документація	Порівняльний аналіз
Експлуатаційні	Age	Термін експлуатації, років	Ціле	Технічний паспорт	Оцінка фізичного зносу
	Load	Коефіцієнт навантаження	Дійсне	Розрахункові дані	Аналіз ризиків
Інструментальні	D	Осідання, мм	Дійсне	Геодезичні вимірювання	Формування кластерів
	C	Ширина тріщин, мм	Дійсне	Обстеження	Критичні ознаки стану

	Def	Деформації, мм	Дійсне	Інструментальний контроль	Аналіз аварійності
Експертні	$E_1 \dots E_k$	Експертні оцінки стану	Дійсне (0–1)	Експертне оцінювання	Фаззифікація
Інтегральні	Z	Ступінь фізичного зносу	Дійсне	Розрахунок / експерт	Класифікація стану
	R	Інтегральний показник ризику	Дійсне	DSS	Формування рекомендацій
Службові	μ	Функція приналежності	Дійсне / інтервальне	FCM / ІНМТ2	Формування кластерів

Наведені параметри формують вектор ознак об'єкта будівництва, який використовується як вхідна інформація для кластеризації технічного стану. Характеристики охоплюють геометричні, експлуатаційні, інструментальні та експертні показники, що дозволяє комплексно описати стан об'єктів в умовах невизначеності. Перед виконанням кластеризації дані проходять етапи валідації, нормалізації та фаззифікації в межах інформаційної технології Clustering4Build.

У процесі експлуатації будівель і споруд виникає необхідність регулярної оцінки їх технічного стану з метою обґрунтування рішень щодо подальшої експлуатації, ремонту, реконструкції або виведення з експлуатації. Така оцінка ґрунтується на аналізі різномірної інформації, отриманої в результаті інструментального та експертного моніторингу, і характеризується неповнотою, невизначеністю та суперечливістю даних.

Науково-прикладна задача полягає у розробці інтегрованої інформаційної технології оцінки технічного стану об'єктів будівництва, яка забезпечує формалізоване подання багатовимірних даних технічного моніторингу та їх адаптивну кластеризацію з використанням нечітких і інтервальних моделей. Результатом розв'язання задачі має бути класифікація об'єктів за технічним станом і формування інформаційної основи для підтримки управлінських рішень.

У межах цієї роботи кластеризація розглядається як ключовий інструмент узагальнення інформації, що дозволяє виявляти групи об'єктів з подібними характеристиками технічного стану, а також ідентифікувати граничні та передаварійні стани.

При розробленні інформаційної технології оцінки технічного стану об'єктів будівництва необхідно враховувати такі обмеження:

Різномірність вхідних даних - поєднання кількісних, якісних, експертних, інтервальних та лінгвістичних показників.

Неповнота інформації - відсутність частини елементів моніторингу для окремих типів будівельних об'єктів.

Невизначеність вимірювань - похибки інструментальних вимірювань та суб'єктивність експертних оцінок.

Різна щільність і об'єм кластерів - наявність груп об'єктів, що істотно відрізняються за кількістю та ступенем подібності.

Обмежені часові та фінансові ресурси - необхідність мінімізації кількості повторних обчислень і налаштувань алгоритмів.

Вимога інтерпретованості результатів - результати кластеризації повинні бути зрозумілими фахівцям з експлуатації будівель.

Зазначені обмеження зумовлюють доцільність застосування методів нечіткої кластеризації та інтервальних нечітких множин другого типу.

Ефективність розроблюваної інформаційної технології оцінюється за такими критеріями: адекватність кластеризації - здатність формувати кластери, що відповідають реальним категоріям технічного стану; стійкість результатів - низька чутливість до варіацій вхідних даних; інформативність - можливість виявлення прихованих закономірностей у даних технічного моніторингу; інтерпретованість - зрозумілість кластерів та значень функцій приналежності для експертів; обчислювальна ефективність - прийнятні часові витрати при обробці великих вибірок; практична корисність - придатність результатів для підтримки управлінських рішень.

Нехай задано множину об'єктів будівництва

$$X = \{x_1, x_2, \dots, x_n\},$$

де кожен об'єкт x_i описується вектором ознак

$$x_i = (x_{i1}, x_{i2}, \dots, x_{iq}),$$

які відповідають елементам технічного моніторингу (інструментальним, експертним, інтегральним).

З урахуванням невизначеності частина ознак може бути подана у вигляді нечітких або інтервальних значень. Об'єкти представлені у багатовимірному просторі ознак

$$X \subset R^q. \quad (1.9)$$

Потрібно знайти розбиття множини X на c кластерів

$$X = \bigcup_{j=1}^c X_j, \quad X_j \subset X, \quad j = 1, \dots, c, \quad (1.10)$$

де кожному об'єкту x_i ставиться у відповідність функція приналежності

$$\mu_j(x_i) \in [0,1], \quad i = 1, \dots, n; \quad j = 1, \dots, c,$$

або інтервальна функція приналежності

$$\tilde{\mu}_j(x_i) = [\mu_j^-(x_i), \mu_j^+(x_i)] \subset [0,1] \quad (1.11)$$

у випадку використання ІНМТ2.

Задача кластеризації формулюється як задача оптимізації:

$$J(U, V) = \sum_{i=1}^n \sum_{j=1}^c (\mu_{ij})^m \|x_i - v_j\|^2 \rightarrow \min_{U, V}, \quad (1.12)$$

де J — критерій якості кластеризації (розсіювання, гіпероб'єм, комбінований критерій), v_j — центри кластерів у просторі ознак.

Результати кластеризації використовуються для віднесення об'єктів до категорій технічного стану та формування рекомендацій щодо подальших управлінських дій.

Спираючись на результати аналізу, наведені в розділі 1, подальший виклад логічно зосереджується на формалізації та розвитку запропонованих підходів. Розглянуті підходи та методи орієнтовані на подальшу програмну реалізацію у вигляді модульної інформаційної системи, що включає підсистеми збору даних, аналітичної обробки, кластеризації та візуалізації результатів.

ВИСНОВКИ ДО РОЗДІЛУ 1

У першому розділі здійснено аналіз сучасного стану проблеми оцінки технічного стану об'єктів будівництва в умовах невизначеності та обґрунтовано доцільність застосування методів кластеризації на основі

нечітких і інтервальних нечітких множин як основного інструменту вирішення цієї задачі в межах інформаційних технологій.

Встановлено, що оцінювання технічного стану будівель і споруд доцільно розглядати не як окрему інженерну процедуру, а як задачу побудови інформаційної технології, що забезпечує збір, інтеграцію та інтелектуальну обробку багатовимірних даних в умовах вимірювальної, інформаційної, модельної, експертної та алгоритмічної невизначеності. Показано, що традиційні детерміновані та жорсткі методи класифікації не відображають розмитого й перехідного характеру реальних технічних станів конструкцій, що зумовлює потребу в нечітких і можливісних моделях подання знань.

У межах розділу було:

- проаналізовано джерела та види невизначеності в системах технічного моніторингу і показано необхідність застосування моделей, здатних формально враховувати нечіткість на всіх етапах обробки даних;

- розглянуто класичні методи кластеризації (ієрархічна кластеризація, алгоритм чітких с-середніх) і встановлено обмеження їх застосування в задачах з перехідними станами об'єктів;

- виявлено, що можливісна кластеризація (PCM) забезпечує підвищену стійкість до шумів і викидів порівняно з FCM, однак потребує автоматизованого підбору параметрів фазифікації, що обґрунтовує доцільність її поєднання з генетичними алгоритмами оптимізації;

- проаналізовано існуючі показники якості (валідності) кластеризації та встановлено, що класичні індекси не враховують специфіку інтервального подання функцій приналежності, характерного для інтервальних нечітких множин другого типу;

- обґрунтовано вибір генетичних алгоритмів (зокрема хромосом змінної довжини) як інструменту автоматизованого налаштування параметрів нечітких кластерних моделей;

- проведено порівняльний аналіз сучасних програмних та інформаційних систем оцінки технічного стану (BIM-, SHM- та DSS-рішень) і

встановлено відсутність у них вбудованих механізмів адаптивної нечіткої кластеризації, а також доцільність переходу до мікросервісної архітектури для забезпечення масштабованості й інтеграції з BIM- і IoT-джерелами даних;

-- сформульовано науково-прикладну задачу дослідження та вимоги до інформаційної технології оцінки технічного стану об'єктів будівництва, що включають адаптивну кластеризацію на основі нечітких моделей, автоматизоване налаштування параметрів засобами генетичних алгоритмів, формалізовану оцінку якості розбиттів та реалізацію у вигляді модульної мікросервісної архітектури.

Проведений аналіз створює методологічне підґрунтя для розроблення у наступних розділах роботи комбінованих методів кластеризації на основі нечітких множин першого типу в поєднанні з генетичними алгоритмами (розділ 2), а також методів на основі інтервальних нечітких множин другого типу (розділ 3).

РОЗДІЛ 2. РОЗРОБКА МЕТОДІВ КЛАСТЕРИЗАЦІЇ НА ОСНОВІ НЕЧІТКИХ МНОЖИН ПЕРШОГО ТИПУ І ГЕНЕТИЧНИХ АЛГОРИТМІВ

Розділ присвячено розробці інформаційної технології кластеризації технічного стану об'єктів будівництва, що реалізує обробку, аналіз і оптимізацію даних у середовищі невизначеності. Запропоновані методи нечіткої кластеризації та генетичної оптимізації є складовими програмного модуля інтелектуальної інформаційної системи підтримки прийняття рішень (DSS) у галузі цифрового моніторингу будівельних конструкцій. Розробка включає архітектурне, алгоритмічне та програмне забезпечення для автоматизованого формування кластерів технічного стану об'єктів із використанням сучасних ІТ-підходів до аналізу даних (Data Mining, Soft Computing, Fuzzy Logic).

2.1. Передумови формування інформаційної технології кластеризації технічного стану об'єктів будівництва

2.1.1. Формалізація технічних показників у вигляді нечітких множин

Оцінювання технічного стану будівельних об'єктів ускладнюється впливом множинних факторів невизначеності, пов'язаних із неповнотою, нечіткістю та суб'єктивністю вхідних даних. У зв'язку з цим виникає необхідність формалізації технічних показників у вигляді нечітких множин, що дозволяє адекватно відображати розпливчастість меж між станами конструкцій та забезпечити інтеграцію як сенсорних даних, так і експертних оцінок. Розроблена процедура формалізації реалізована у вигляді програмного модуля нечіткої фазифікації, який входить до складу розробленої інформаційної технології оцінки технічного стану. Цей модуль автоматизує перетворення технічних параметрів у нечіткі множини та передає сформовані вектори ознак у базу даних для подальшої кластеризації.

Розроблений модуль є частиною архітектури інформаційної системи кластеризації технічного стану, побудованої за принципом модульної ІТ-

структури. Він реалізує інтерфейс обміну даними між підсистемами збору сенсорної інформації, базою даних PostgreSQL та аналітичним ядром кластеризації. Модуль відноситься до підсистеми обробки даних та попередньої фазифікації, що є ключовим елементом інформаційної технології цифрового моніторингу об'єктів будівництва. Завдяки цьому забезпечено автоматичний зв'язок між даними моніторингу, експертними оцінками та математичними моделями кластеризації.

У межах цієї роботи було обґрунтовано вибір основних параметрів, які виступають індикаторами технічного стану об'єктів будівництва. До таких параметрів віднесено:

- фізичні характеристики (осідання, ширина тріщин, деформації);
- механічні характеристики (модуль пружності, залишкова несуча здатність);
- функціональні характеристики (герметичність, теплоізоляція, вологість);
- експертні показники (ступінь пошкодження за класифікаційними шкалами).

Для кожного параметра було запропоновано лінгвістичну інтерпретацію на основі трьох або більше рівнів стану: добрий, задовільний, незадовільний тощо. Ці рівні описуються нечіткими множинами з відповідними функціями належності, що моделюють поступовий перехід між класами.

Наприклад, для параметра «ширина тріщини» x (мм) було сформовано три нечіткі множини: Low (мала тріщина) — максимальна належність при $x \leq 0,2$ мм; Medium (середня тріщина) — максимальна належність у діапазоні $x \in [0,3; 0,6]$ мм; High (велика тріщина) — при $x > 1,0$ мм, функція належності прямує до 1. Формально, кожна нечітка множина $A_i \subset X$, що відповідає конкретному параметру x_i , визначається парою $\{x, \mu_{A_i}(x)\}$, де $\mu_{A_i}(x): X \rightarrow [0,1]$ — функція належності, яка задається у вигляді:

- трикутної функції належності, якщо параметр має чітко виражений центр;
- трапецієподібної, якщо існує стійкий інтервал оптимальних значень;
- гауссової або сигмоїдальної, якщо необхідне згладжування переходів.

При цьому враховується наявність інтервальних значень, експертної нечіткості, а також шумів у сенсорних даних, які можуть бути представлені у вигляді розмитих множин з різним ступенем довіри до кожного виміру.

На основі нечіткої формалізації всіх вхідних параметрів будується вектор нечітких ознак для кожного об'єкта діагностики:

$$\tilde{x} = (\mu_1(x_1), \mu_2(x_2), \dots, \mu_n(x_n)), \quad (2.1)$$

де $\mu_i(x_i)$ — значення функції належності для i -го параметра.

Цей вектор є вхідним у процес нечіткої кластеризації (розділ 2.1.2) та дозволяє проводити гнучке групування об'єктів з урахуванням невизначеності в даних.

У рамках практичної реалізації розроблено бібліотеку функцій для автоматичного формування функцій належності на основі статистичних даних (для сенсорних параметрів) та з використанням знань експертів (для суб'єктивних оцінок).

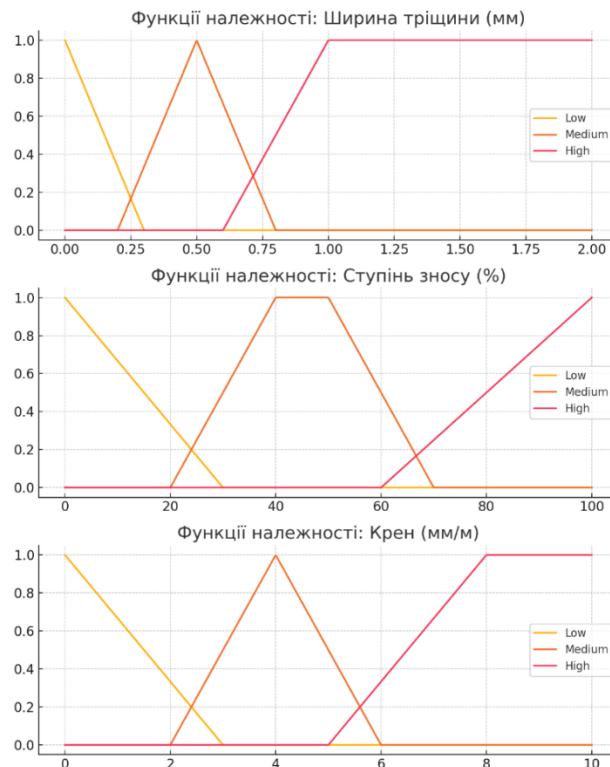


Рис.2.1. Функції належності: Крен (мм/м) (Розроблено автором на основі [7, 107])

На графіках вище зображено функції належності для трьох технічних показників: ширина тріщини, ступінь зносу та крен конструкції, що

використовуються для формалізації в нечітку модель технічного стану. Приклади лінгвістичних термів для 5 основних параметрів будівельного стану подано в таблиці 2.1.

Таблиця 2.1. Лінгвістичні терми для технічних показників *(Розроблено автором)*

Показник	Low (Добрий)	Medium (Задовільний)	High (Незадовільний)
Ширина тріщини, мм	≤ 0.2	0.3–0.6	> 1.0
Ступінь зносу, %	≤ 30	30–70	> 70
Крен, мм/м	≤ 2	2–5	> 5
Вологість, %	≤ 8	8–16	> 16
Модуль пружності, МПа	≥ 25000	15000–25000	< 15000
Корозія арматури, %	≤ 5	5–15	> 15
Прогин балки, мм	≤ 3	3–8	> 8
Температура бетону, °C	≤ 35	35–60	> 60
Відшарування штукатурки, %	≤ 5	5–20	> 20
Осідання фундаменту, мм	≤ 10	10–25	> 25
Час реакції датчика, мс	≤ 100	100–300	> 300
Глибина деформації, мм	≤ 5	5–15	> 15
Рівень шуму, дБ	≤ 55	55–75	> 75
Товщина ізоляції, мм	≥ 60	40–60	< 40
Коефіцієнт теплопровідності, Вт/м·К	≤ 0.035	0.035–0.06	> 0.06

Розроблена формалізація технічних показників у вигляді нечітких множин слугує методологічною основою для побудови інформаційної технології кластеризації технічного стану об'єктів будівництва в умовах невизначеності. Отримані в результаті формалізації вектори нечітких ознак не лише описують стан об'єкта у вигляді багатовимірного простору ступенів

належності, а й забезпечують адаптацію всієї системи до умов неповноти, суб'єктивності та шумової природи вхідних даних.

У розробленій інформаційній технології оцінювання технічного стану, зазначені нечіткі множини формуються автоматично у модулі попередньої обробки даних, який є частиною програмного комплексу Clustering4Build.

Цей модуль взаємодіє з базою даних технічних параметрів і забезпечує підготовку вхідних векторів для подальшої кластеризації у центральному аналітичному модулі системи. Таким чином, нечіткі множини слугують не лише математичною моделлю, а й структурним елементом інформаційної технології.

Програмна реалізація модуля нечіткої формалізації виконана на мові Python із використанням бібліотек NumPy, Pandas і Fuzzy Logic Toolbox. Архітектурна структура модуля представлена на рис. 2.2.

Вона містить три основні компоненти:

- модуль імпорту даних;
- блок фазифікації;
- API-інтерфейс передачі нечітких векторів у базу даних.

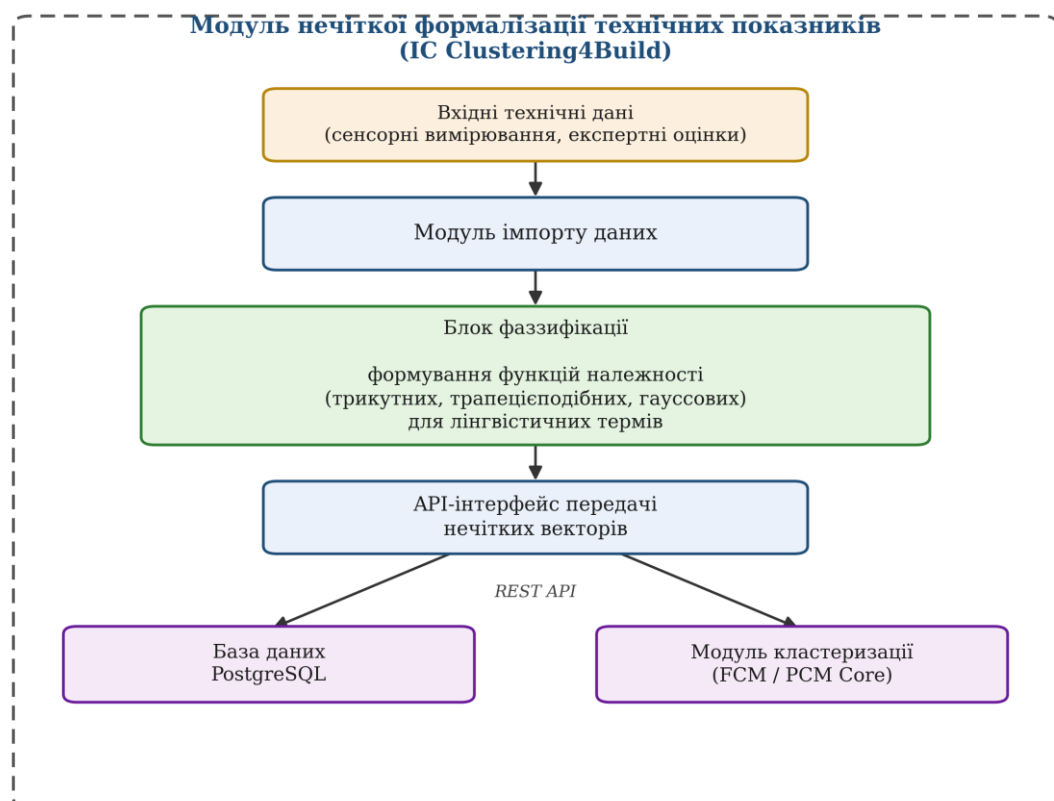


Рис.2.2. Архітектура програмного модуля нечіткої формалізації технічних показників у складі IC Clustering4Build. *(Розроблено автором)*

Така структура забезпечує модульність, масштабованість і можливість повторного використання у складі розподіленої інформаційної системи. Дані зберігаються у реляційній базі PostgreSQL, а взаємодія з іншими модулями інформаційної технології здійснюється через REST API. Такий підхід дозволяє інтегрувати модуль у складні системи моніторингу та використовувати його як складову інтелектуальної інформаційної системи оцінки технічного стану об'єктів будівництва.

У подальших розділах ці вектори використовуються як вхідні дані для алгоритмів нечіткої кластеризації — зокрема, Fuzzy C-Means (FCM) та його узагальнених модифікацій на базі інтервальних нечітких множин другого типу (IT2-FCM). Завдяки попередньо сформованим функціям належності до лінгвістичних термів («добрий», «задовільний», «незадовільний») забезпечується можливість кластеризації не за жорсткими значеннями, а з урахуванням розмитих границь між класами, що є характерною рисою практичної інженерної діагностики.

Також формалізовані показники використовуються при оптимізації параметрів кластеризації за допомогою генетичних алгоритмів, що реалізовано у розділі 2.2. Зокрема, функції належності стають елементами хромосомної репрезентації рішень у багатокритеріальних алгоритмах оптимізації (наприклад, NSGA-III). Це дозволяє ефективно налаштовувати ключові параметри кластеризації (кількість кластерів, функції подібності, тип нормалізації тощо) з урахуванням нечітких ознак об'єктів.

У рамках програмної реалізації, описаної у розділі 4, нечіткі множини застосовуються для автоматичної обробки вхідних даних із сенсорних систем та BIM-моделей, а також для реалізації модулів кластеризації у прикладному програмному комплексі «Clustering4Build». У графічному інтерфейсі користувача (GUI) результати нечіткої класифікації виводяться у вигляді інтерактивних мап, кольорових шкал і рівнів небезпеки, що підвищує зручність для прийняття рішень.

Таким чином, застосування нечітких множин для формалізації технічних показників забезпечує підвищену адаптивність кластеризаційних алгоритмів до умов реальної експлуатації будівельних об'єктів, дозволяючи відобразити неоднозначність інженерних оцінок у цифровій формі.

У розробленій інформаційній технології кластеризації технічного стану FCM виступає ключовим аналітичним компонентом, який забезпечує обробку нечітких і неповних даних.

Алгоритм інтегрований з модулями оптимізації (генетичні алгоритми) та оцінки якості кластерів, що дозволяє створити єдину архітектуру для автоматизованого аналізу технічного стану будівельних об'єктів.

Реалізація FCM-алгоритму у складі розробленої інформаційної технології кластеризації дозволяє виконувати комплексну обробку даних, автоматичне оновлення кластерів і інтеграцію результатів у систему підтримки прийняття рішень.

Таким чином, розроблена процедура нечіткої формалізації реалізує повний цикл обробки даних - від збору до формування нечітких векторів ознак і передачі їх у кластеризаційне ядро. Вона поєднує методи нечіткої логіки з сучасними інформаційними технологіями, базами даних і веб-сервісами, що забезпечує її відповідність вимогам спеціальності до створення інтелектуальних інформаційних систем аналізу технічного стану будівельних об'єктів.

2.1.2. Математичні основи FCM та обробка даних у середовищі невизначеності

Кластеризація даних у задачах технічної діагностики об'єктів будівництва має справу з неоднорідними, неповними та часто нечіткими вхідними даними, що суттєво знижує ефективність класичних методів аналізу. Враховуючи це, у рамках даного дослідження як базовий інструмент для групування технічних станів об'єктів було обрано алгоритм нечіткої кластеризації типу Fuzzy C-Means (FCM), який дозволяє формувати м'яке

групування об'єктів, тобто забезпечує часткову належність кожного елементу до кількох кластерів одночасно.

Основна ідея алгоритму FCM полягає в такому: нехай маємо множину вхідних векторів технічного стану $X = \{x_1, x_2, \dots, x_N\}$, де $x_i \in \mathbb{R}^d$ — вектор нечітких ознак для i -го об'єкта, отриманий згідно з процедурою, описаною у п. 2.1.1. Метою FCM є мінімізація цільової функції:

$$J_m(U, V) = \sum_{i=1}^N \sum_{j=1}^C u_{ij}^m \|x_i - v_j\|^2, \quad (2.2)$$

де: C — кількість кластерів; $u_{ij} \in [0,1]$ — ступінь належності об'єкта x_i до кластера j ; $m > 1$ — параметр нечіткості (типово $m \in [1,5; 3,0]$); $v_j \in \mathbb{R}^d$ — центр кластера j ; $\|x_i - v_j\|$ — евклідова відстань або інша метрика.

Центри кластерів і матриця приналежностей $U = [u_{ij}]$ оновлюються ітеративно за такими правилами:

$$v_j = \frac{\sum_{i=1}^N u_{ij}^m x_i}{\sum_{i=1}^N u_{ij}^m}, \quad u_{ij} = \left(\sum_{k=1}^C \left(\frac{\|x_i - v_j\|}{\|x_i - v_k\|} \right)^{\frac{2}{m-1}} \right)^{-1} \quad (2.3)$$

Ці рівняння повторюються доти, доки не буде досягнута збіжність — зміна матриці U стає меншою за заданий поріг ε . Реалізація алгоритму FCM у складі інформаційної технології забезпечує інтеграцію з модулем збору даних і базою знань.

Реалізація алгоритму FCM виконана у вигляді окремого обчислювального модуля FCM Core, що входить до складу інформаційної технології кластеризації технічного стану. Цей модуль реалізує серверну частину аналітичного ядра системи, взаємодіє з базою даних PostgreSQL через REST API та забезпечує обробку нечітких векторів, отриманих із модуля фазифікації.

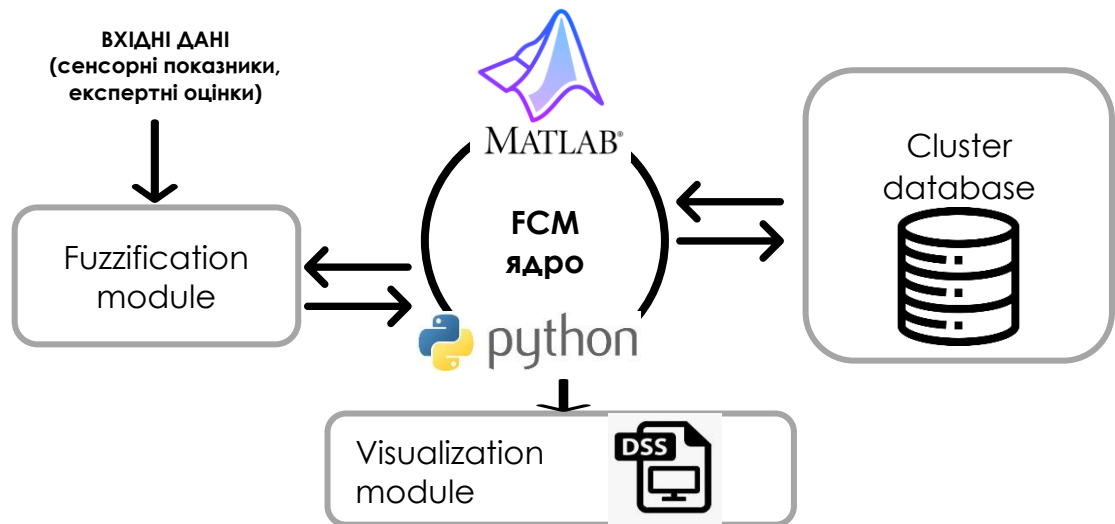


Рис. 2.3. Інформаційна схема реалізації FCM-алгоритму у складі технології кластеризації технічного стану об'єктів будівництва Clustering4Build (Розроблено автором)

У процесі обчислень модуль формує журнали ітерацій, зберігає проміжні матриці $U(k)$ і центри кластерів $V(k)$ у базі даних, що дозволяє виконувати контроль збіжності та адаптивне оновлення результатів кластеризації в режимі реального часу.

Кожен об'єкт зберігається у базі даних у вигляді вектора нечітких ознак, що дозволяє здійснювати оновлення результатів кластеризації в реальному часі.

Такий підхід забезпечує зв'язок математичної моделі з реальними інформаційними потоками системи моніторингу. У межах інформаційної архітектури Clustering4Build, модуль FCM Core реалізовано як вебсервіс аналітичного рівня. Він приймає вхідні дані від модулів збору сенсорних показників, виконує кластеризацію, зберігає результати у сховище кластерів та передає їх у модуль візуалізації DSS (Decision Support System).

Однією з ключових переваг FCM є його придатність до роботи з невизначеними і розмитими даними, які характерні для систем технічного моніторингу:

1. Нечіткі вхідні ознаки (результати формалізації технічного стану) природно вбудовуються у простір R^d , де кожна координата — значення функції належності до відповідного лінгвістичного терму.

2. М'яке групування дозволяє уникнути жорсткої класифікації, що є принципово важливим у прикладних задачах, де об'єкти можуть перебувати на межі між двома станами (наприклад, «задовільний» / «незадовільний»).

3. Розширення FCM для роботи з недостовірними або нечіткими відстанями передбачає модифікацію метрики або включення коефіцієнтів достовірності $\alpha_i \in [0,1]$ для кожного об'єкта:

$$d'(x_i, v_j) = \alpha_i \|x_i - v_j\|^2 \quad (2.4)$$

Це дозволяє враховувати ступінь довіри до окремих вимірювань (наприклад, сенсор із втратою калібрування).

У рамках поставленої задачі кластеризації технічного стану об'єктів будівництва алгоритм FCM було адаптовано з урахуванням особливостей технічного моніторингу, а саме до:

- роботи з високовимірними нечіткими векторами, які містять інформацію про тріщини, знос, крен, деформації, температуру, вологість тощо;
- обробки поточкових або періодично оновлюваних даних з сенсорних систем;

використання альтернативних метрик (манхеттенської, косинусної) для задач із нерівнорозмірними параметрами. Алгоритм FCM надає ефективний інструментарій для аналізу нечітких даних у задачах оцінки технічного стану. Його використання у поєднанні з формалізацією технічних показників у вигляді нечітких множин дозволяє формувати адаптивні моделі кластеризації, стійкі до неточностей, суб'єктивності та неоднорідності вхідної інформації.

У рамках інформаційної технології кластеризації цей алгоритм реалізовано як аналітичне ядро програмного комплексу Clustering4Build, що забезпечує автоматичне групування об'єктів за ступенем технічного стану. Програмна реалізація здійснена на Python із використанням бібліотек NumPy, SciPy та scikit-fuzzy. Вона підтримує інтеграцію з BIM-моделями через API, а також роботу з поточковими сенсорними даними через MQTT або HTTP-з'єднання.

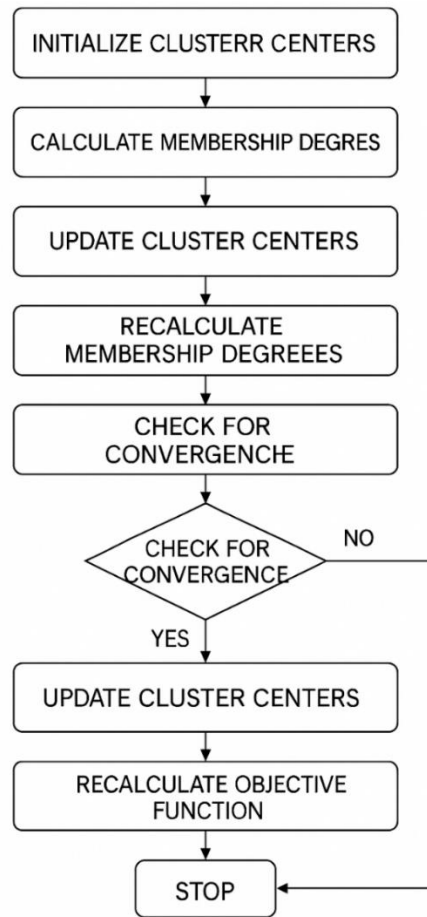


Рис.2.4. Алгоритм FCM як інструментарій для аналізу нечітких даних у задачах оцінки технічного стану *(Розроблено автором на основі [9, 102])*

Реалізація передбачає серверний модуль обчислень, інтерфейс API для взаємодії з клієнтським додатком та можливість інтеграції з ВІМ-системами. Отримані матриці належностей зберігаються у базі даних і використовуються для візуалізації технічного стану в інформаційній системі моніторингу.

Результатом застосування FCM є матриця ступенів належності U , яка дозволяє: побудувати відкриту інтерпретацію технічного стану на основі найбільш домінуючого кластера; реалізувати гнучке групування з урахуванням багатofакторного стану об'єкта; застосувати алгоритм у рамках гібридних та оптимізованих моделей, описаних у наступних підрозділах (2.3, 3.1).

Таким чином, підрозділ реалізує не лише математичну, а й інформаційно-технологічну складову нечіткої кластеризації. Алгоритм FCM виступає ядром аналітичної підсистеми, що забезпечує автоматичну обробку нечітких даних, інтеграцію з базою знань, формування візуалізаційних звітів

та підтримку прийняття рішень у цифрових системах моніторингу будівельних конструкцій. У результаті застосування алгоритму FCM формується матриця ступенів належності U , елементи якої $u_{ij} \in [0;1]$ відображають ступінь належності i -го об'єкта до j -го кластера. На відміну від жорсткої класифікації, де об'єкт повністю відноситься лише до одного класу, FCM дозволяє відображати розмиту природу технічного стану, що особливо актуально в умовах міжкласових та прикордонних ситуацій.

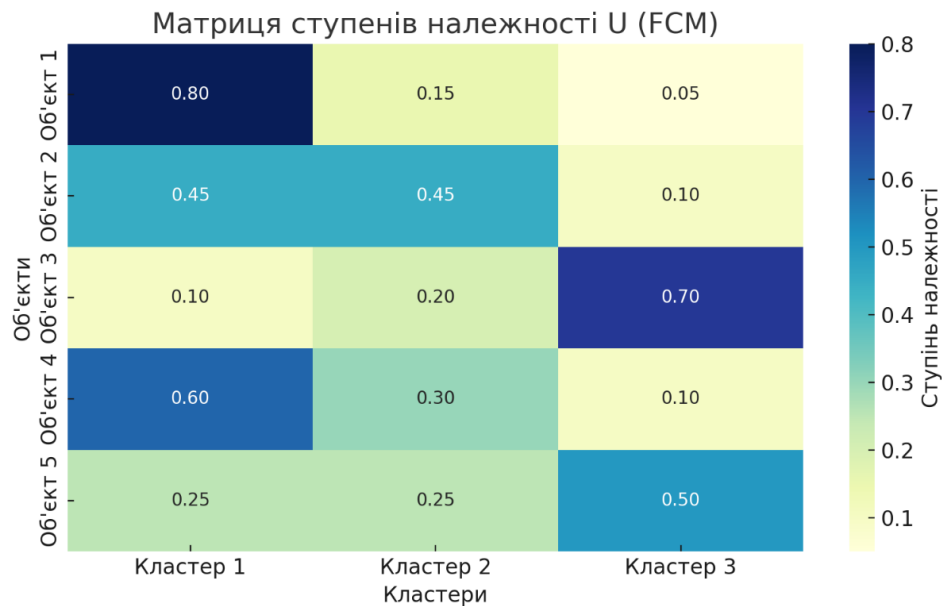


Рис.2.5. Матриця ступенів належності U (FCM). (Розроблено автором на основі [9])

Наприклад, у демонстраційному прикладі для 5 об'єктів і 3 кластерів, об'єкт 1 має високу ступінь належності до першого кластера (0.80), тоді як об'єкт 2 майже порівну належить до двох кластерів (0.45 до першого і другого), що вказує на нечіткість діагнозу. Така структура матриці дозволяє не лише точно виявляти кластери, до яких належать об'єкти, а й визначати ступінь довіри до рішення, використовувати ці значення для побудови графічних інтерфейсів користувача, автоматизованої візуалізації результатів, а також інтегрувати їх у системи підтримки прийняття рішень при оцінюванні технічного стану будівельних конструкцій.

2.2. Генетичні алгоритми оптимізації кластеризації з використанням FCM-алгоритму на основі нечітких множин першого типу

Використання генетичного алгоритму (ГА) дозволяє істотно скоротити час пошуку оптимального нечіткого розбиття, для якого значення цільової функції буде екстремальним [48, 102]. У межах даного дослідження реалізація генетичного алгоритму виконується у складі інформаційної технології кластеризації технічного стану об'єктів будівництва. ГА реалізовано як окремий обчислювальний модуль GeneticOptimizer, що функціонує на рівні аналітичного ядра системи Clustering4Build.

Модуль взаємодіє з базою даних PostgreSQL та підсистемою FCM/PCM Core через REST API, отримуючи поточні центри кластерів та передаючи оптимізовані параметри назад у систему для оновлення результатів кластеризації.

Структурна інтеграція ГА у систему Clustering4Build передбачає автоматизований обмін даними між підсистемами через API-рівень. Зокрема, оптимізаційний модуль GeneticOptimizer взаємодіє з FCM/PCM Core, базою даних PostgreSQL та візуалізаційним інтерфейсом DSS, що дозволяє використовувати результати оптимізації безпосередньо у процесах діагностики будівельних об'єктів.

Така інтеграція забезпечує автоматизовану оптимізацію параметрів кластеризації, відповідає принципам побудови інтелектуальних інформаційних систем, а також підвищує продуктивність аналізу технічного стану в режимі реального часу. У подальшому в якості опції відповідності буде використовуватися індекс Се - Бені відповідно до формули (2.2), який повинен бути мінімізований.

При реалізації ГА необхідно розглянути два випадки:

- шукана кількість кластерів c задана заздалегідь (фіксована);
- шукана кількість кластерів c заздалегідь невідома (не задана).

В першому випадку слід використовувати ГА з хромосомою постійної довжини, а в другому — ГА з хромосомою змінної довжини.

Для деякої кількості кластерів c хромосома може бути закодована координатами центрів кластерів або ступенями нечіткої приналежності об'єктів центрам кластерів - числами з інтервалу $[0,1]$.

Кодування хромосоми у системі Clustering4Build реалізовано у вигляді JSON-структур, що зберігаються у базі даних для подальшої візуалізації результатів оптимізації. Модуль GeneticOptimizer є складовою аналітичного шару IC Clustering4Build, що взаємодіє з FCM/PCM Core через REST API. Така побудова дозволяє масштабувати обчислення, підтримувати розподілену обробку даних і інтегрувати систему з зовнішніми вебсервісами. Кожна хромосома супроводжується метаданими — номером покоління, значенням функції пристосованості, типом схрещування, що забезпечує відстежуваність та контроль якості обчислень у середовищі невизначеності.

Такий підхід дозволяє інтегрувати математичну модель ГА в архітектуру інформаційної системи та забезпечити її повторне використання при обробці різних типів технічних даних. GeneticOptimizer виконує роль аналітичного сервісу в архітектурі розподіленої інформаційної системи. Його структура побудована за принципами сервісно-орієнтованої архітектури (SOA). Використання REST API забезпечує незалежність компонентів, спрощує інтеграцію з підсистемами збору даних і DSS-модулями, а також дозволяє реалізувати масштабованість і модульність у хмарному середовищі.

2.2.1. Кодування хромосоми нечіткими ступенями приналежності об'єктів центрам кластерів

При кодуванні хромосоми координатами центрів кластерів довжина хромосоми дорівнює sq , де s - кількість кластерів, q - кількість елементів моніторингу. При цьому перші q елементів хромосоми відповідають координатам центру першого кластеру, другі q елементів - координатам центру другого кластеру і т.п. [151]. Діапазон зміни значення кожного гену (координати центру кластера) визначається інтервалом $[D_{\min}, D_{\max}]$, де $D_{\min}$ і $D_{\max}$ - мінімальний і максимальний бали обраної порядкової шкали, так як оцінювання об'єктів за кожним елементом моніторингу проводиться з використанням деякої порядкової шкали шляхом виставлення балів.

При кодуванні хромосоми нечіткими ступенями приналежності об'єктів центрам нечітких кластерів довжина хромосоми буде дорівнювати sp , де s - кількість кластерів, p - кількість об'єктів. При цьому перші s елементів хромосоми відповідають ступеням приналежності першого об'єкта центрам кластерів, другі s елементів хромосоми - ступеням приналежності другого об'єкта центрам кластерів і т.п. Діапазон зміни значення кожного гену визначається інтервалом $[0, 1]$, границі якого відповідають мінімальній і максимальній нечітким ступеням приналежності об'єкта кластеру [153]. Кодування хромосоми нечіткими ступенями приналежності об'єктів центрам нечітких кластерів дозволяє уникнути втрати інформації під час операції схрещування, так як батьківські хромосоми будуть обмінюватися тільки нечіткими ступенями приналежності об'єктів центрам кластерів, а не координатами центрів кластерів. Програмна реалізація описаної схеми кодування хромосом виконана у вигляді окремого модуля оптимізації з REST-взаємодією та централізованим збереженням популяцій у базі даних (детально розглянуто в розділі 4).

2.2.2. Генетичний алгоритм з хромосомою постійної довжини, що закодована координатами центрів кластерів

При формуванні початкової популяції хромосом при заздалегідь заданій кількості кластерів $s=s^*$ хромосома, для якої нечітка кластеризація фактично виконується на кількість кластерів s менше, ніж $s^*(s < s^*)$ повинна бути визнана «нежиттєздатною», при цьому значення її функції відповідності слід вважати рівною деякому максимальному числу $F_{\max}$ (значення якого свідомо більше максимального значення цільової функції), щоб виключити цю хромосому з подальшого відбору, інакше ефективність застосування ГА буде значно знижена [152]. Також слід вчинити при обчисленні значень функції відповідності хромосом-нащадків при виконанні операцій схрещування або мутації.

Вибір батьків здійснюється на основі ймовірнісного відбору, відповідно до якого для кожної хромосоми на основі значення її функції відповідності обчислюється ймовірність:

$$p_l = \frac{F_l}{\sum_{k=1}^P F_k} \quad (2.5)$$

$$q_l = \sum_{k=1}^l p_k \quad (2.6)$$

де F_l — значення цільової функції хромосоми l ; P — розмір популяції; q_l — накопичена (кумулятивна) ймовірність для l -ї хромосоми.

Реалізація відбору та генерації потомства у модулі GeneticOptimizer виконується у вигляді паралельних процесів із використанням бібліотеки multiprocessing у Python. Використання багатопотокових технологій у реалізації ГА дозволяє створити високопродуктивну інформаційну платформу, яка обробляє великі масиви технічних даних у реальному часі. Це дозволяє розподіляти обчислення між ядрами процесора й суттєво зменшувати час оптимізації.

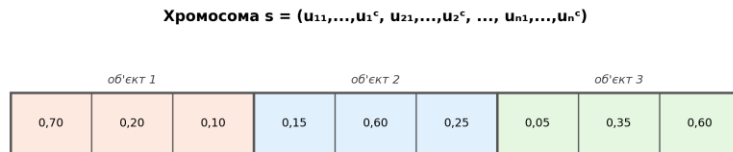


Рис. 2.6.а. Вид хромосоми при кластеризації множини об'єктів на c кластерів по q елементам моніторингу. (Розроблено автором на основі [151, 152, 153])

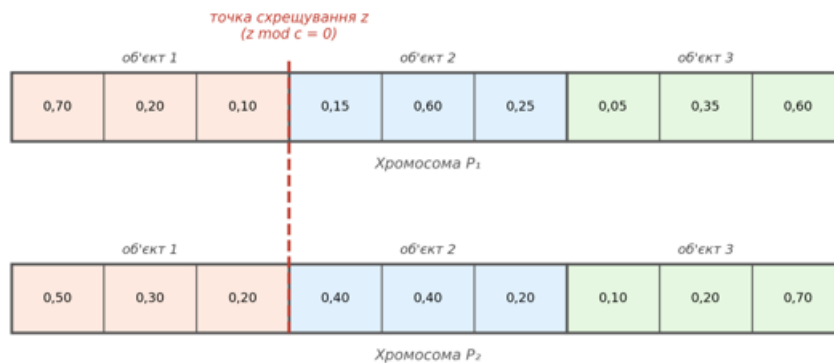


Рис. 2.6.б. Дві хромосоми до виконання операції схрещування. (Розроблено автором на основі [151, 152, 153])

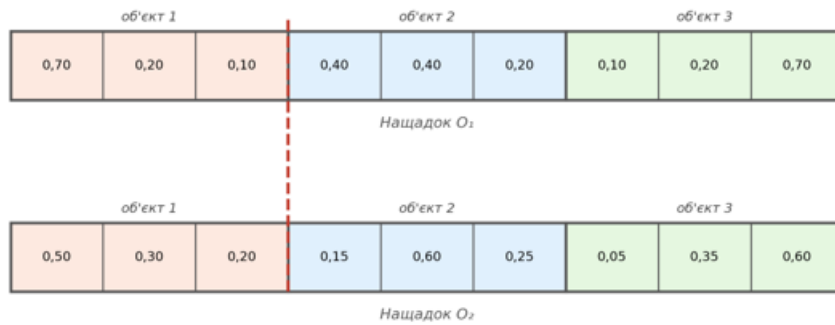


Рис. 2.6.в. Дві хромосоми після виконання операції схрещування.
(Розроблено автором на основі [151, 152, 153])



Рис. 2.6.г. Хромосоми до виконання операції мутації. (Розроблено автором на основі [151, 152, 153])



Рис. 2.6.д. Хромосома після виконання операції мутації. (Розроблено автором на основі [151, 152, 153])

Батьківські хромосоми обираються таким чином.

4. На горизонтальній осі відкладається ряд чисел $q_1, q_2, \dots, q_P$.
5. Генерується випадкове число $z \in [0,1]$.
6. В якості батька обирається хромосома l , якщо випадкове число z потрапило в l -й інтервал: $q_{l-1} < z \leq q_l$, де $q_0 = 0$.
7. Кроки 2–3 повторюються для визначення другої хромосоми батька.

Дві визначені таким чином хромосоми будуть використовуватися в якості батьків при виконанні операції схрещування.

Нехай при реалізації ГА використовується одноточкове схрещування і одноточкова мутація. При виконанні операції схрещування вибирається

ймовірність схрещування R_c і генерується випадкове число N_c . Якщо $R_c > N_c$, то випадковим чином вибирається точка схрещування z і виконується схрещування.

При виконанні операції мутації вибирається ймовірність мутації R_m і генерується випадкове число N_m . Якщо $R_m > N_m$, то випадковим чином вибирається точка мутації z і виконується мутація. Генетичний алгоритм для комбінування з FCM-алгоритмом реалізовано за аналогією до підходів, представлених у [151, 152], з такими особливостями:

1. Випадковим чином створюється популяція розміром P відповідно до формули (2.3) для функції пристосованості (fitness function) об'єктів кластерам і формулою (2.3) для центрів кластерів з перевіркою умови «життєздатності» хромосом, тобто вимоги про розбиття на задану кількість кластерів: $c = c^*$ (рисунок 2.1).

2. При $g < G$ (G і g — максимальна і поточна кількість поколінь ГА відповідно) обчислюється значення цільової функції (2.2) для кожної хромосоми і створюється $R_c \cdot P/2$ пар хромосом-батьків.

3. Виконуються операції схрещування (рисунки 2.2 і 2.3) і мутації (рисунки 2.4 і 2.5) для поточної популяції з перевіркою умови «життєздатності» хромосом, тобто вимоги про розбиття на задану кількість кластерів: $c = c^*$.

4. Створюється нова популяція розміром P , доповнена хромосомами-нащадками в кількості $R_c \cdot P$, потім $R_c \cdot P$ хромосом з гіршими значеннями функції відповідності (2.2) відкидаються. Якщо $g < G$, здійснюється перехід до кроку 2. Якщо $g \geq G$, то робота ГА завершується і здійснюється перехід до кроку 5.

5. Вибирається краща хромосома, яка мінімізує функцію відповідності (2.2). Для кожного об'єкта кластеризації визначаються нечіткі ступені приналежності до нечітких кластерів з новими центрами кластерів відповідно до формули (2.3).

Нечітка кластеризація на основі ГА з використанням хромосом постійної довжини, закодованих координатами центрів кластерів, дозволяє отримати більш адекватні результати кластеризації з досягненням субмінімуму індексу Се–Бені за формулою (2.2), однак вимагає великої кількості змін поколінь G .

Для реалізації цього процесу у складі інформаційної технології створено підсистему адаптивної оптимізації кластеризації, що координується контролером `OptimizationController`. Контролер здійснює запуск еволюційного циклу, моніторинг збіжності популяцій, збереження історії поколінь у базі даних та передачу оновлених центрів кластерів у модуль `PCM Core`.

Архітектурно підсистема реалізована у вигляді мікросервісу `OptimizationController`, який має REST-інтерфейс для обміну даними з `FCM/PCM Core`. Мікросервісний підхід у побудові підсистеми `OptimizationController` відображає ключові принципи сучасних інформаційних технологій, орієнтованих на розподілені обчислення. Кожен сервіс працює автономно, має власний API та виконує визначену функцію в межах загальної архітектури `Clustering4Build`.

Кожна ітерація еволюційного циклу записується у системний журнал та візуалізується у панелі моніторингу `DSS`, що забезпечує прозорість процесу оптимізації для користувача.

У межах інформаційної технології кластеризації технічного стану, генетичний алгоритм реалізує функції адаптивної оптимізації параметрів `FCM` — кількості кластерів, функцій подібності, коефіцієнтів нечіткості. Програмна реалізація виконана у вигляді окремого оптимізаційного модуля, який взаємодіє з аналітичним ядром через API. Це забезпечує масштабованість системи, повторне використання алгоритмів та інтеграцію з іншими компонентами інформаційної системи.

2.2.3. Генетичний алгоритм з хромосомою постійної довжини, що закодована нечіткими ступенями приналежності об'єктів центрам кластерів

При кодуванні хромосоми нечіткими ступенями приналежності об'єктів центрам кластерів сума перших s координат хромосоми, сума других s координат хромосоми і т.п. дорівнює одиниці, так як перші s координат відповідають нечітким ступеням приналежності першого об'єкту нечітким кластерам, другі s координат - нечітким ступеням приналежності другого об'єкту кластерам і т.п. Приклад хромосоми для випадку розбиття множини об'єктів на 4 кластери наведено на рисунку 2.7.

Генетичний алгоритм реалізується так само, як і ГА, наведений в п. 2.3.3, але є ряд відмінностей при виконанні операцій схрещування і мутації. Початкова популяція, як і в п. 2.3.3, формується таким чином, щоб вона складалася лише з «життєздатних» хромосом, інакше ефективність застосування ГА буде значно знижена.

Вибір батьків здійснюється з використанням імовірнісного відбору на основі формули (2.5).

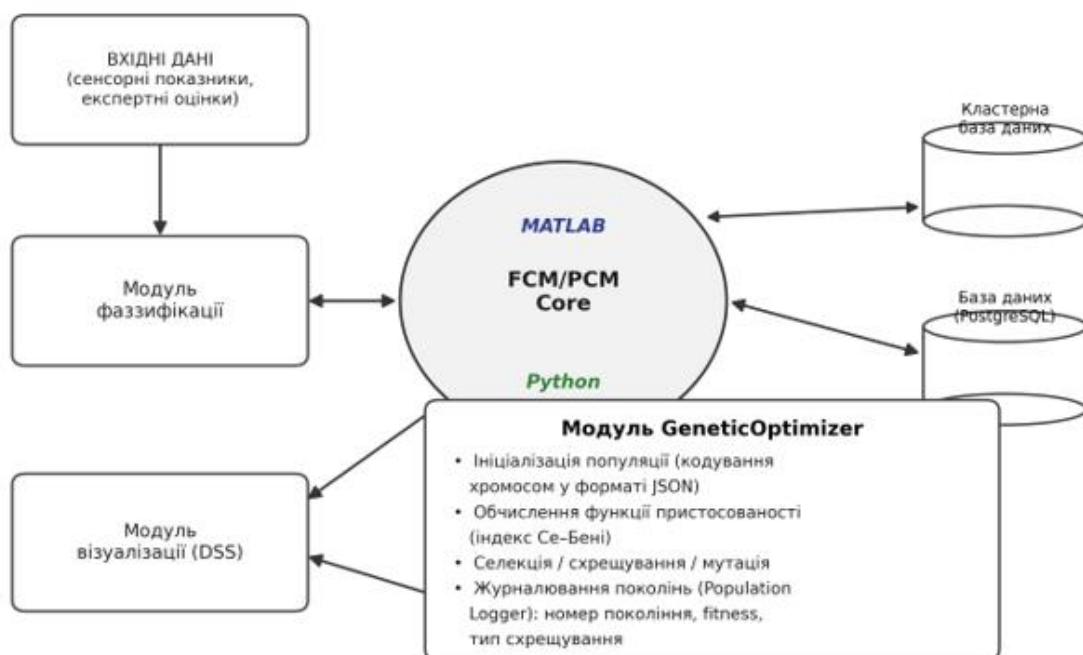


Рис. 2.7. Інформаційна архітектура модуля GeneticOptimizer у складі IT Clustering4Build (Розроблено автором)

Розроблена архітектура відповідає вимогам побудови модульних інформаційних систем, оскільки забезпечує:

- відокремлення рівнів даних, логіки та представлення;
- можливість горизонтального масштабування;
- сумісність із середовищами Python та веб-інтерфейсами DSS;
- підтримку інтеграції з IoT-системами збору даних.

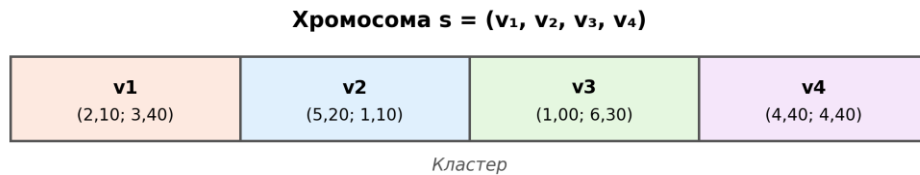


Рис. 2.8.а. Вид хромосоми при кластеризації множини об'єктів 4 кластерами.
(Розроблено автором на основі [117, 118, 151, 152])

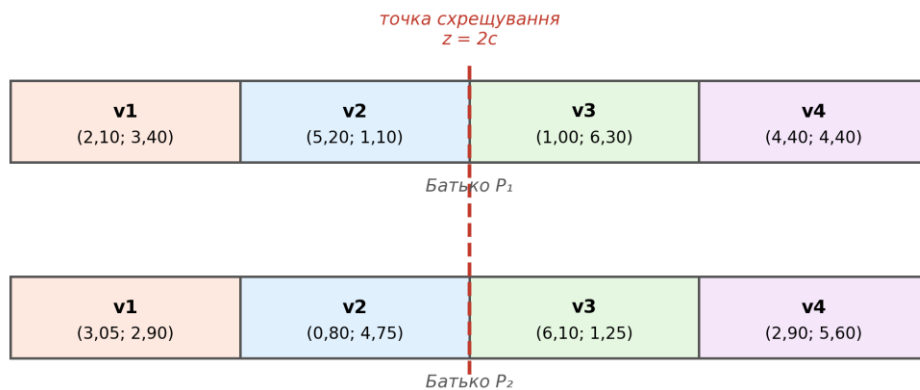


Рис. 2.8.б. Дві хромосоми до виконання операції схрещування. (Розроблено автором на основі [117, 118, 151, 152])

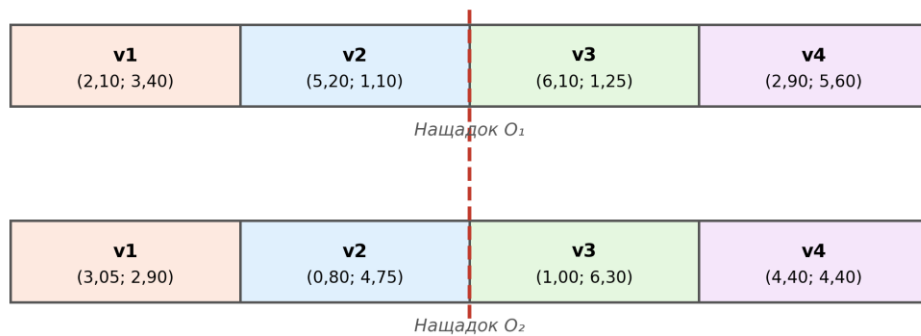


Рис. 2.8.в. Дві хромосоми після операції схрещування. (Розроблено автором на основі [117, 118, 151, 152])

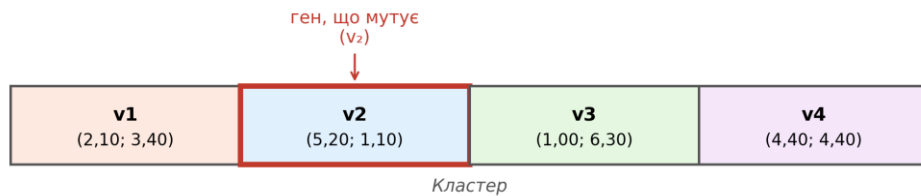


Рис. 2.8.г. Хромосоми до виконання операції мутації. (Розроблено автором на основі [117, 118, 151, 152])

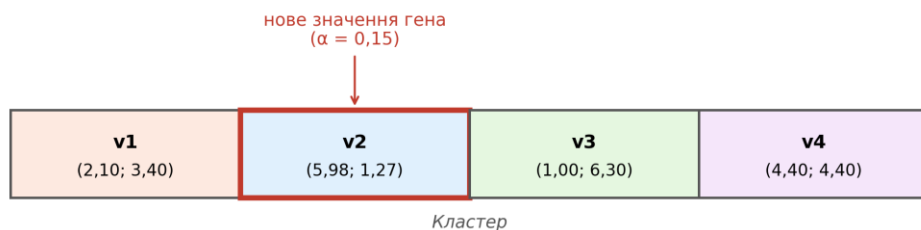


Рис. 2.8.д. Хромосоми після виконання операції мутації при $\alpha=0,15$. (Розроблено автором на основі [117, 118, 151, 152])

При виконанні схрещування вибирається ймовірність схрещування R_c і генерується випадкове число N_c . Якщо $R_c > N_c$, то випадковим чином вибирається точка схрещування z ($z \bmod c = 0$) і виконується схрещування

(рисунки 2.14 і 2.15). Точка схрещування z визначає номер об'єкта кластеризації, починаючи з нечітких ступенів приналежності якого відбувається обмін частинами хромосом-батьків. Умова $z \bmod c = 0$ забезпечує обмін усіма нечіткими ступенями приналежності деякого об'єкта, тобто виконання умови:

$$z = k \cdot c, \quad k = 1, 2, \dots, n \quad (2.7)$$

де c — кількість кластерів, u_{ij} — ФП об'єкта x_i кластеру j .

При виконанні мутації вибирається ймовірність мутації R_m і генерується випадкове число N_m . Якщо $R_m > N_m$, то випадковим чином вибирається точка мутації z ($z \bmod c = 0$) і виконується мутація. Точка мутації z визначає номер об'єкта, для нечітких ступенів приналежності якого буде виконуватися мутація. Потім випадковим чином вибирається точка w , що визначає номер гена, який мутує (нечіткого ступеня приналежності об'єкта кластеру w) для об'єкта з номером z (рис. 2.10 і 2.11). Також генерується випадкове число $s(w) \in [0, 1]$, де $s(w)$ — значення нечіткого ступеня приналежності об'єкта z кластеру w .

Тоді нове значення нечіткого ступеня приналежності в точці мутації, що визначає приналежність об'єкта z кластеру w , знаходиться як:

$$u_{zw}^{\text{new}} = s(w) \quad (2.8)$$

Нечіткі ступені приналежності даного об'єкта z іншим кластерам з номерами, що не дорівнюють w , змінюються таким чином:

$$u_{zk}^{\text{new}} = u_{zk} \cdot \frac{1-s(w)}{1-u_{zw}}, \quad k \neq w \quad (2.9)$$

В цьому випадку генетичний алгоритм має такий вигляд.

Випадковим чином створюється популяція розміром P відповідно до формули (2.3) для нечітких ступенів належності об'єктів кластерам і формулою (2.3) для центрів кластерів з перевіркою умови «життєздатності» хромосом, тобто вимоги про розбиття на задане число кластерів: $c = c^*$ (рисунок 2.7). При $g < G$ (g і G — поточна і максимальна кількість поколінь

ГА відповідно) обчислюється значення цільової функції (2.2) для кожної хромосоми і створюється $R_c \cdot P/2$ пар хромосом-батьків.

Виконуються операції схрещування (рис. 2.8 і 2.9) і мутації (рис. 2.10 і 2.11) для поточної популяції з перевіркою умови «життєздатності» хромосом, тобто вимоги про розбиття на задану кількість кластерів: $c = c^*$.

Створюється нова популяція розміром P , доповнена хромосомами-нащадками в кількості $R_c \cdot P$, потім $R_c \cdot P$ хромосоми з гіршими значеннями функції відповідності (2.2) відкидаються. Якщо $g < G$, виконується перехід до кроку 2. Якщо $g \geq G$, то робота ГА завершується і відбувається перехід до кроку 5.

Вибирається найкраща хромосома, яка мінімізує функцію відповідності (2.2), і для неї обчислюються нові координати центрів нечітких кластерів відповідно до (2.6).

Нечітка кластеризація на основі ГА з використанням хромосом постійної довжини, закодованих нечіткими ступенями приналежності об'єктів центрам кластерів, дозволяє отримати більш адекватні результати кластеризації з знаходженням субмінімуму індексу C_e - Бені, але як і в разі використання ГА використання хромосом постійної довжини, закодованих координатами центрів кластерів, вимагає великої кількості змін поколінь G для отримання шуканого нечіткого розбиття.

Програмна реалізація ГА виконана на Python із використанням бібліотеки DEAP. Окрім DEAP, для аналітичної візуалізації процесу еволюції використовується модуль Plotly/Dash, який інтегровано в інтерфейс DSS.

Це дозволяє в реальному часі спостерігати зміну функції пристосованості, кількість поколінь і збіжність параметрів. Реалізація модуля GeneticOptimizer передбачає обробку великих обсягів даних із використанням бібліотек NumPy, Pandas, SciPy та DEAP. Для інтеграції з іншими модулями використовується внутрішній REST API, що дозволяє взаємодіяти з компонентами FCM/PCM Core, базою даних PostgreSQL та візуалізаційним інтерфейсом DSS. У системі передбачено логування кожного покоління, що забезпечує

можливість візуального аналізу еволюції популяції у вигляді графіків і таблиць у GUI. Під час виконання алгоритму створюється журнал параметрів популяцій, який зберігається у базі даних для подальшого аналізу. Робота модуля координується контролером інформаційної технології, який керує виконанням обчислень і взаємодією між модулями.

2.2.4. Генетичний алгоритм з хромосомою змінної довжини, що закодована координатами центрів кластерів

Якщо оптимальна кількість кластерів заздалегідь невідома, то в ГА довжина хромосоми може змінюватися від $c_{\min}$ до $c_{\max}$, де $c_{\min}$ і $c_{\max}$ — мінімальна і максимальна можлива кількість кластерів. Головною відмінністю ГА з хромосомою змінної довжини від ГА з хромосомою постійної довжини є реалізація операції схрещування [152]. У рамках інформаційної системи Clustering4Build алгоритм з хромосомою змінної довжини реалізований як модуль Adaptive Genetic Optimizer (AGO), який функціонує у зв'язці з підсистемою PCM Core. Обмін даними між цими модулями здійснюється через REST API у форматі JSON. Модуль AGO приймає поточні координати центрів кластерів, виконує еволюційний цикл оптимізації та повертає оновлені параметри через endpoint `/api/pcm/update`.

Результати кожної ітерації зберігаються у базі даних PostgreSQL у таблицях `population_log` і `evolution_trace`, що забезпечує можливість відтворення та аналітичного моніторингу процесу збіжності.

Вибір батька буде полягати в виборі найкращої хромосоми, яка мінімізує індекс Се-Бені, з двох випадково обраних.

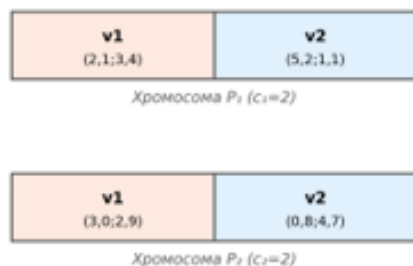


Рис. 2.9.а. Дві хромосоми однакової довжини ($c_1=c_2=2$). (Розроблено автором на основі [152])

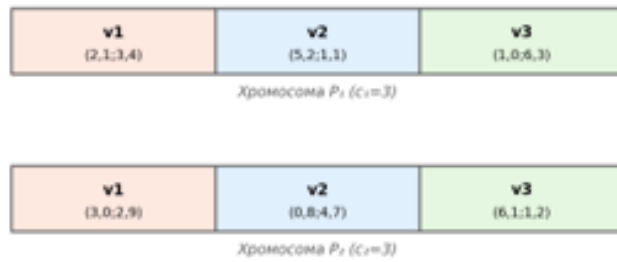


Рис. 2.9.б. Дві хромосоми однакової довжини ($c_1=c_2=3$). (Розроблено автором на основі [152])



Рис. 2.9.в. Приклади можливих варіантів пар хромосом перед реалізацією операції схрещування. Дві хромосоми різної довжини: c_1, c_2 . (Розроблено автором на основі [152])

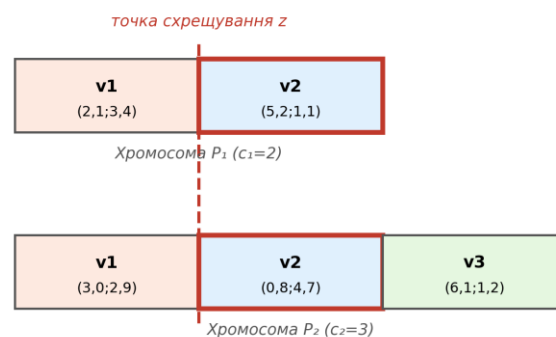


Рис. 2.9.г. Дві хромосоми різної довжини до виконання операції схрещування. (Розроблено автором на основі [152])

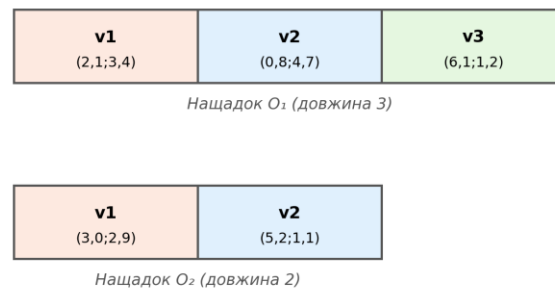


Рис. 2.9.д. Дві хромосоми різної довжини після виконання операції схрещування. (Розроблено автором на основі [152])

Потім дві обрані батьківські хромосоми будуть використовуватися для виконання схрещування. При виконанні схрещування вибирається ймовірність схрещування R_c і генерується випадкове число N_c . Якщо $R_c > N_c$, то випадковим чином вибирається точка схрещування z , $z \in \{1, \dots, \tilde{c}q\}$, $c = \min(c_1, c_2)$ (c_1 і c_2 — кількість кластерів для першої і другої хромосоми, що беруть участь у схрещуванні), і виконується схрещування.

При виконанні мутації вибирається ймовірність мутації R_m і генерується випадкове число N_m . Якщо $R_m > N_m$, то випадковим чином вибирається точка мутації z і виконується мутація.

На рисунку 2.9 наведено приклади пар хромосом-батьків, обраних на основі ймовірнісного відбору для виконання операції схрещування, для випадку, коли кількість кластерів c дорівнює 2 або 3. Тоді можливі три варіанти групування хромосом:

- дві хромосоми однакової довжини при $c_1 = c_2 = 2$, де c_1 і c_2 — кількість кластерів для першої і другої хромосом відповідно;
- дві хромосоми однакової довжини при $c_1 = c_2 = 3$;
- дві хромосоми різної довжини при $c_1 = 2$ і $c_2 = 3$.

У разі, коли батьківські хромосоми мають однакову довжину, операція схрещування виконується так само, як в ГА з хромосомою постійної довжини (п. 2.3.3, рисунки 2.9.а і 2.9.б). На рисунках 2.9.г і 2.9.д наведено приклади

хромосом різної довжини (c_1 і c_2) до і після виконання операції схрещування. Операція мутації в ГА з хромосомою змінної довжини аналогічна операції мутації в ГА з хромосомою постійної довжини. Загалом реалізація ГА з хромосомою змінної довжини аналогічна реалізації ГА з хромосомою постійної довжини, за винятком особливостей виконання операції схрещування [152].

Програмна реалізація алгоритму виконана з використанням бібліотек DEAP, NumPy та FastAPI. Компоненти реалізовані у вигляді мікросервісів:

- PCM Core — обчислення центрів кластерів;
- Adaptive GA Service — управління популяцією;
- Database Service — збереження результатів;
- DSS Visualization — побудова графіків збіжності.

Така модульна побудова відповідає архітектурі інтелектуальних ІС, що передбачає розподілену обробку даних і масштабованість.

З інформаційно-технологічної точки зору, модуль Adaptive GA Service виступає ядром еволюційної оптимізації в системі Clustering4Build. Він здійснює автоматичне логування етапів еволюції, передає ключові параметри у базу даних та формує REST-повідомлення для DSS-підсистеми, яка відображає динаміку процесу в режимі реального часу.

2.2.5. Аналіз проблеми реалізації генетичного алгоритму з хромосомою змінної довжини, що закодована нечіткими ступенями приналежності об'єктів центрам кластерів

Зважаючи на особливості способу кодування хромосоми нечіткими ступенями приналежності об'єктів центрам кластерів і особливості реалізації операції схрещування для хромосом, що закодовані нечіткими ступенями приналежності об'єктів центрам кластерів реалізація ГА з хромосомою змінної довжини, закодованої нечіткими ступенями приналежності об'єктів центрам кластерів, не представляється можливим.

. Це пояснюється так. Хромосома, закодована нечіткими ступенями приналежності об'єктів центрам кластерів, складається з n «блоків» (за кількістю об'єктів n) довжиною c (за кількістю кластерів c). Тоді виконання операції схрещування може призвести до абсурдної ситуації, коли частина об'єктів кластеризації розбивається на одне число кластерів (наприклад, $c_1 = 2$), а інша частина об'єктів кластеризації — на інше число кластерів (наприклад, $c_2 = 3$), і все це — в одній хромосомі. Отже, ГА з хромосомою змінної довжини, закодованою нечіткими ступенями приналежності об'єктів центрам кластерів, не може бути реалізований.

Крім математичних аспектів, у системі реалізовано інформаційно-управлінський механізм взаємодії модулів, який забезпечує обмін метаданими між GeneticOptimizer, FCM/PCM Core та базою знань. Це дозволяє зберігати параметри оптимізації, значення функцій та структуру кластерів у єдиному форматі JSON через внутрішній протокол обміну.

Розроблені алгоритми інтегровані в архітектуру інформаційної технології Clustering4Build, що складається з модулів збору даних, аналітичного ядра, оптимізаційного ядра, бази даних та системи візуалізації. Це забезпечує повний цикл обробки нечітких даних — від сенсорних вимірів до оптимізованих результатів кластеризації, що подаються користувачу у вигляді інтерактивних графіків у системі DSS.

2.3 Комбінування алгоритму FCM на основі нечітких множин першого типу з генетичним алгоритмом

Розроблена в межах підрозділу 2.2 архітектура забезпечує повний цикл роботи інформаційної системи — від збору сенсорних даних до оптимізації параметрів кластеризації та візуалізації результатів у DSS. Поєднання ГА, PCM і REST-інфраструктури створює основу інтелектуальної ІТ-платформи для цифрового моніторингу будівельних конструкцій.

Розроблена архітектура Clustering4Build може бути класифікована як інтелектуальна ІТ-платформа, що реалізує принципи розподілених обчислень, модульної інтеграції та адаптивного управління даними. У системі реалізовано механізми автоматичного прийняття рішень на основі нечітких правил та генетичної оптимізації, що відповідає напрямкам розробки інтелектуальних інформаційних систем аналізу технічного стану. Передбачено інформаційні інтерфейси REST API, базу даних PostgreSQL з JSONB-структурами, модуль DSS для підтримки прийняття рішень, а також засоби захисту даних на рівні транзакцій і контролю доступу.

Поєднання алгоритму FCM, заснованого на нечітких множинах першого типу (НМП1), з генетичним алгоритмом (ГА) із хромосомами постійної або змінної довжини дозволяє скоротити кількість поколінь G, необхідних для досягнення адекватних результатів нечіткої кластеризації. [151, 152, 153]. Комбінований метод нечіткої кластеризації передбачає почергове виконання FCM-алгоритму на основі НМТ1 і ГА з хромосомою постійної (змінної) довжини [152]. Схема комбінованого алгоритму кластеризації FCM + GA подана в додатку 3..

Схема ілюструє поетапне поєднання FCM-алгоритму на основі нечітких множин першого типу з генетичним алгоритмом у процесі формування та оптимізації кластерів. Чергування кроків FCM і ГА дозволяє забезпечити більш гнучке налаштування параметрів кластеризації та уникати локальних мінімумів цільової функції. З інформаційної точки зору, комбінований метод реалізує взаємодію двох аналітичних сервісів у межах архітектури Clustering4Build: модулів FCM Core та GeneticOptimizer. Їхня інтеграція здійснюється через REST API, що дозволяє автоматизувати обмін параметрами кластерів, функціями належності та показниками якості у форматі JSON. Такий підхід забезпечує узгодженість між аналітичними модулями, базою даних та візуалізаційним інтерфейсом DSS.

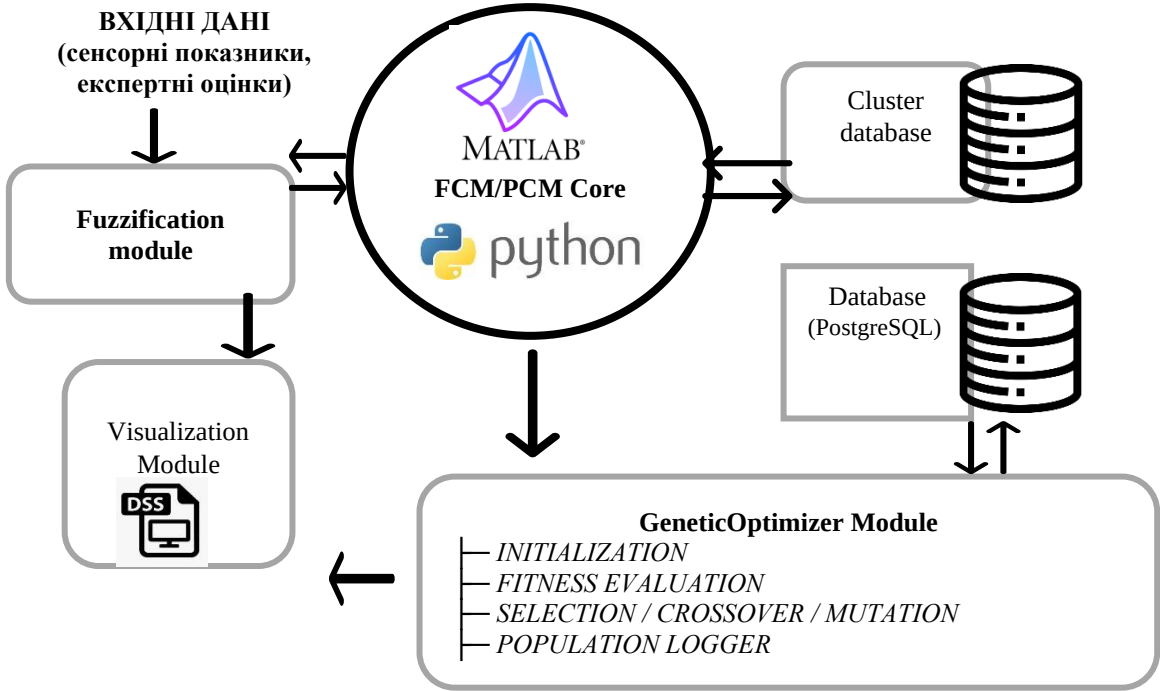


Рис. 2.10. Інформаційна архітектура комбінованого алгоритму FCM+GA у складі Clustering4Build (Розроблено автором)

На рисунку показано взаємодію FCM Core, GeneticOptimizer, бази даних PostgreSQL і модуля DSS через REST API. Обмін даними реалізується у форматі JSON, забезпечуючи автоматизацію процесів кластеризації та оптимізації.

Таблиця 2.2. Інформаційні потоки у комбінованому алгоритмі (Розроблено автором)

Джерело	Приймач	Формат даних	Тип передавання	Призначення
FCM Core	GeneticOptimizer	JSON (centers, U-matrix)	REST POST /api/genetic/init	Ініціалізація хромосом
GeneticOptimizer	PostgreSQL	JSONB	REST POST /api/db/population	Збереження поколінь
PostgreSQL	DSS	SQL / JSON	GET /api/dss/update	Візуалізація результатів
FCM Core / DSS	OptimizationController	REST / WebSocket	Асинхронне оновлення	

Запропонований метод можна застосовувати як при кодуванні хромосом координатами центрів кластерів в ГА з хромосомою постійної (змінної) довжини, так і при кодуванні хромосом нечіткими ступенями приналежності об'єктів центрам кластерів в ГА з хромосомою постійної довжини.

2.3.1. Комбінування FCM-алгоритму на основі нечітких множин першого типу і генетичного алгоритму з хромосомою постійної довжини при кодуванні хромосом координатами центрів кластерів

Комбінований метод полягає в чергуванні кроків FCM-алгоритму та генетичного алгоритму (ГА), що працює з хромосомами постійної або змінної довжини [151, 152]. Здійснюється одна ітерація FCM-алгоритму на основі НМТ1 при формуванні хромосом початкової популяції розміром P .

При $g < G$ (g і G — поточна і максимальна кількість поколінь ГА відповідно) здійснюється одна ітерація ГА з реалізацією операцій схрещування і мутації та обчисленням значень функції відповідності за формулою (2.12) для нової популяції хромосом розміром $R_c \cdot P$.

Для нової популяції розміром $R_c \cdot P$, що представлена хромосомами, закодованими координатами центрів кластерів, здійснюється одна ітерація FCM-алгоритму на основі НМТ1 з обчисленням нових значень ФП (нечітких ступенів належності) об'єктів центрам кластерів відповідно до формули (2.3), нових координат центрів кластерів. Наступним кроком виконується обчислення нових значень ФП об'єктів центрам кластерів відповідно до формули (2.3).

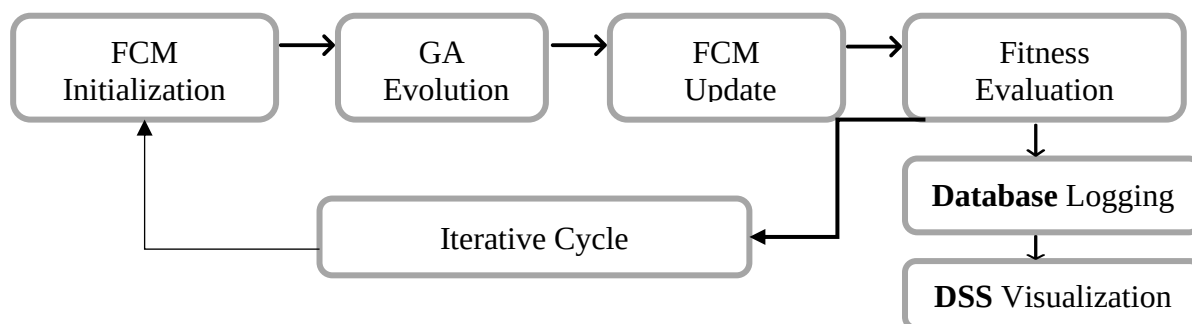


Рис. 2.11. Схема циклу комбінованої кластеризації (Розроблено автором)

У програмній реалізації цей етап виконується у вигляді запиту до аналітичного API FCM Core, який обчислює нові центри кластерів і повертає їх у GeneticOptimizer. Така архітектура реалізує принципи розподіленої обробки даних (distributed computing) та дозволяє масштабувати обчислення у середовищі хмарних сервісів. Кожен компонент Clustering4Build — FCM Core, GeneticOptimizer та DSS - виступає як окремий мікросервіс, який взаємодіє через REST API.

Така побудова дозволяє забезпечити незалежність обчислювальних модулів, горизонтальне масштабування, а також інтеграцію системи з IoT-сенсорними вузлами у реальному часі.

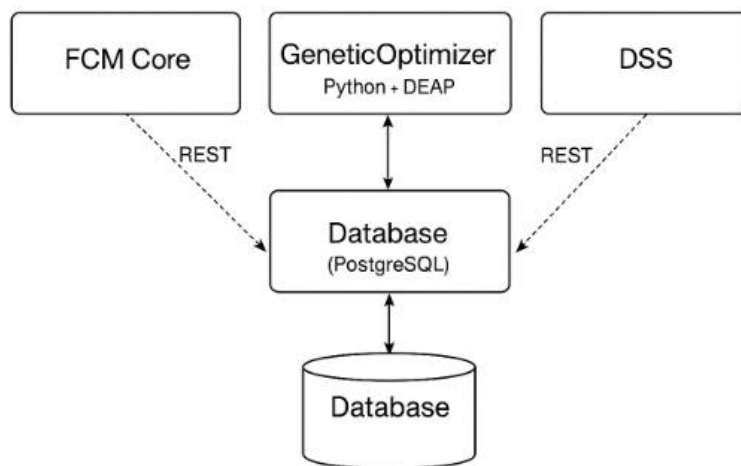


Рис. 2.12. Мікросервісна архітектура комбінованої ІС кластеризації Clustering4Build. (Розроблено автором)

З розширеної популяції розміром $2P + R_c \cdot P$, що отримана шляхом об'єднання популяції розміром P попереднього кроку і популяції розміром $R_c \cdot P$ поточного кроку, видаляються «нежиттєздатні» $P + R_c \cdot P$ хромосоми з максимальними значеннями функції відповідності за формулою (2.2). Якщо $g < G$, виконується перехід до кроку 2. Якщо $g \geq G$, то робота ГА завершується і відбувається перехід до кроку 5.

Обирається найкраща хромосома, яка мінімізує функцію відповідності за формулою (2.2). Шукані координати центрів нечітких кластерів визначаються на основі найкращої хромосоми. В якості результуючих нечітких ступенів належності об'єктів центрам кластерів приймаються

ступені приналежності об'єктів центрам кластерів, що відповідають найкращій хромосомі і вже обчислені в ході реалізації комбінованого методу нечіткої кластеризації.

Таблиця 2.3. Налаштування параметрів генетичного алгоритму для кластеризації IT2-FCM+GA (Розроблено автором)

Параметр	Позначення	Значення	Коментар
Розмір популяції	P	50–100	Залежить від обсягу даних
Максимальна кількість поколінь	G	30–50	Для збіжності оптимального рішення
Ймовірність схрещування	R_c	0.7–0.9	Рекомендується значення > 0.8
Ймовірність мутації	R_m	0.05–0.2	Може бути адаптивною
Тип селекції	—	Tournament	Можливі: roulette, rank, elitist
Тип схрещування	—	Одноточкове	Або блочне чи часткове об'єднання центрів
Тип мутації	—	Гаусова	Або випадкова зміна однієї координати
Критерій зупинки	—	$g \geq G_g$ або без покращень 5 поколінь	Комбінований підхід

Слід зазначити, що на другому етапі комбінованого методу нечіткої кластеризації в якості ГА використовується ГА, описаний в п. 2.3.3.

Так як FCM-алгоритм є ітераційним алгоритмом, то при реалізації ГА в кожному поколінні поряд з «головною» популяцією хромосом, закодованих координатами центрів кластерів, обчислюється «допоміжна» популяція нечітких ступенів належності об'єктів центрам кластерів.

Для підвищення ефективності ГА пропонується в процесі формування розширеної популяції хромосом чергового покоління ГА враховувати популяцію розміром P, отриману в попередньому поколінні, і нову популяцію розміром $(P + R_c \cdot P)$, отриману в поточному поколінні. Для кожної хромосоми нової популяції розміром $(P + R_c \cdot P)$ додатково здійснюється одна ітерація FCM-алгоритму на основі НМТ1, яка дозволяє оновити ступені

належності об'єктів та координати центрів кластерів. У результаті формується розширена популяція розміром $(2P + R_c \cdot P)$, хромосоми якої впорядковуються за зростанням значень функції відповідності (індексу Се-Бені) за формулою (2.2). Фінальна популяція розміром P формується шляхом відсікання хромосом із найгіршими значеннями функції відповідності з розширеної популяції $(2P + R_c \cdot P)$, з гіршими значеннями функції відповідності за формулою (2.2).

Для впорядкування кластерів від «кращого» до «гіршого» можна застосовувати величину відстані d_j від центру кластера $v_j = (v_{j1}, v_{j2}, v_{jq})$ до «ідеального об'єкта» з «ідеальними» оцінками за всіма q елементами моніторингу. Наприклад, якщо еталонний (ідеальний) об'єкт описується вектором оцінок $(1, \dots, 1)$ за всіма q показниками моніторингу, то відстань до центру кластера можна обчислити як:

$$d_j = \sqrt{\sum_{k=1}^q (v_j^k - 1)^2}, \quad j=(1,c) \quad (2.10)$$

Таким чином, чим більша відстань від «ідеального об'єкта» до центра кластера, тим «гіршим» є кластер і об'єкти, що в нього потрапили. Розраховані показники використовуються не лише для аналітичної оцінки якості кластерів, а й для формування інформаційних звітів у підсистемі DSS. Для оцінки ефективності комбінованого підходу проведено порівняння продуктивності різних варіантів алгоритмів кластеризації.

Таблиця 2.4. Порівняння продуктивності різних варіантів алгоритмів кластеризації (Розроблено автором)

Метод кластеризації	Кількість ітерацій	Середній індекс Се-Бені	Час виконання (с)	Відносна продуктивність	Примітка
FCM	80	0,045	18,6	1.0x	Базовий метод
GA (пост. хромосоми)	55	0,041	14,9	1,25ч	Оптимізує центри
FCM + GA	35	0,039	10,4	1,8x	

Комбінований метод FCM+GA демонструє скорочення кількості ітерацій до збіжності на 45–60% та підвищення точності кластеризації, що підтверджує ефективність інформаційної інтеграції алгоритмів у межах архітектури Clustering4Build.

2.3.2 Комбінування FCM-алгоритму на основі нечітких множин першого типу і генетичного алгоритму з хромосомою змінної довжини при кодуванні хромосом координатами центрів кластерів

Алгоритм комбінованої кластеризації на основі FCM і ГА зі змінною довжиною хромосом реалізується аналогічно до методу, описаного в п. 2.3.1, із тією різницею, що замість фіксованої кількості кластерів використовується динамічно змінна. Таким чином, комбінування FCM та ГА у межах інформаційної технології Clustering4Build не лише підвищує ефективність математичного моделювання, але й реалізує сучасні ІТ-принципи побудови аналітичних систем.

При кодуванні хромосом нечіткими ступенями приналежності об'єктів центрам кластерів комбінування FCM-алгоритму і ГА з хромосомою постійної довжини реалізується за схемою, адаптованою з підходу [153], таким чином:

1. Здійснюється одна ітерація FCM-алгоритму на основі НМТ1 при формуванні хромосом початкової популяції розміром P .
2. При $g < G$ (G і g — максимальна і поточна кількість поколінь ГА відповідно) здійснюється одна ітерація ГА з реалізацією операцій схрещування і мутації та обчисленням значень функції відповідності за формулою (2.2) для нової популяції хромосом розміром $(P + R_c \cdot P)$.
3. Для нової популяції розміром $(P + R_c \cdot P)$, що представлена хромосомами, закодованими нечіткими ступенями приналежності об'єктів центрам кластерів, здійснюється одна ітерація FCM-алгоритму на основі НМТ1 з обчисленням нових координат центрів кластерів відповідно до

формули (2.3), нових значень ФП (нечітких ступенів приналежності) об'єктів центрам кластерів відповідно до формули (2.3).

4. З розширеної популяції розміром $2P + R_c \cdot P$, що отримана шляхом об'єднання популяції розміром P попереднього покоління і популяції розміром $P + R_c \cdot P$ поточного покоління, видаляються «нежиттєздатні» $P + R_c \cdot P$ хромосоми з максимальними значеннями функції відповідності за формулою (2.2). Якщо $g < G$, виконується перехід до кроку 2. Якщо $g \geq G$, то робота ГА завершується і відбувається перехід до кроку 5.

5. Обирається найкраща хромосома, яка мінімізує функцію відповідності за формулою (2.2). Шукані ступені приналежності об'єктів центрам кластерів визначаються на основі найкращої хромосоми. В якості результуючих координат центрів нечітких кластерів приймаються координати центрів нечітких кластерів, що відповідають найкращій хромосомі і вже обчислені в ході реалізації комбінованого методу нечіткої кластеризації.

Таблиця 2.5. Порівняння ефективності методів кластеризації

Метод	Цільова функція	Число Генерацій	% Точність класифікації
FCM	0,34	–	87,5%
FCM-GA	0,30	25	92,8%
ГАл	0,31	15	88,7%

(Розроблено автором)

Наведено порівняння класичного методу FCM, комбінованого підходу FCM-GA та генетичного алгоритму ГАл за значенням цільової функції, кількістю поколінь до збіжності та точністю класифікації. Комбінований метод FCM-GA демонструє найкращу ефективність за всіма критеріями.

Таким чином, комбінування FCM-алгоритму з генетичними підходами дає змогу адаптивно шукати оптимальну структуру кластерів, уникати локальних мінімумів функції відповідності та підвищити якість класифікації технічного стану об'єктів будівництва в умовах невизначеності.

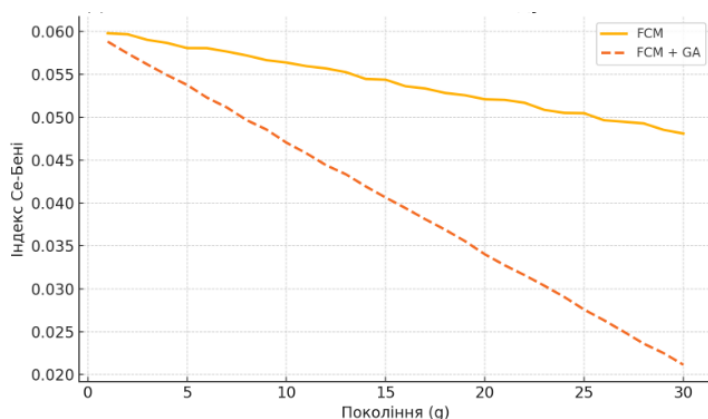


Рис. 2.13. Динаміка збіжності комбінованого методу FCM+GA за поколіннями *(Розроблено автором)*

На рисунку представлено структуру хромосом, блок-схему процесу комбінованої нечіткої кластеризації, а також порівняння реалізації алгоритму FCM + GA в середовищах MATLAB та Python. Ліва частина ілюструє структуру хромосом із нечіткими ступенями належності та координатами центрів кластерів, а також логіку ітеративної взаємодії FCM-алгоритму з генетичними операціями.

Права частина інфографіки відображає переваги та особливості реалізації в MATLAB і Python за критеріями продуктивності, гнучкості, синтаксису та наявності пакетів.

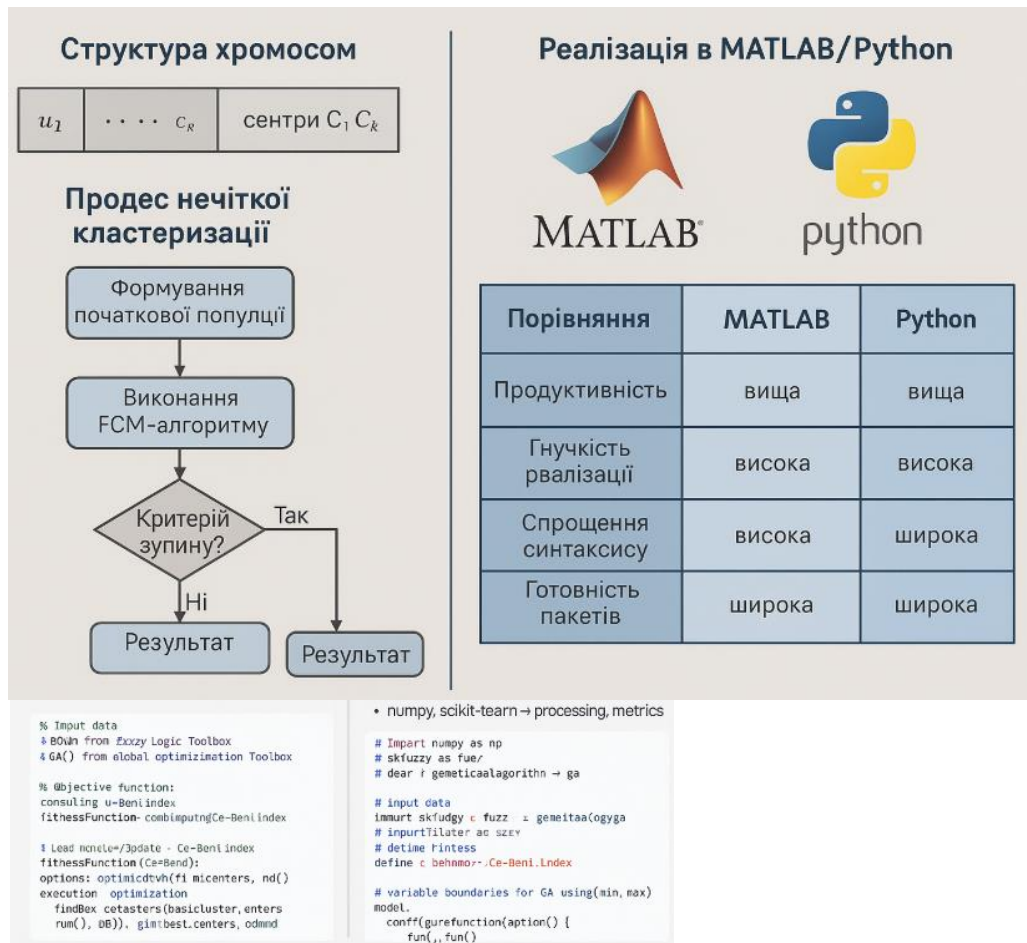


Рис. 2.14. Динаміка збіжності комбінованого методу FCM+GA за поколіннями (Розроблено автором)

У межах комбінованого підходу на основі інтервальних нечітких множин другого типу (IT2-FCM) та генетичного алгоритму (ГА) була реалізована інтегрована обчислювальна схема, зображена на рисунку 2.15.

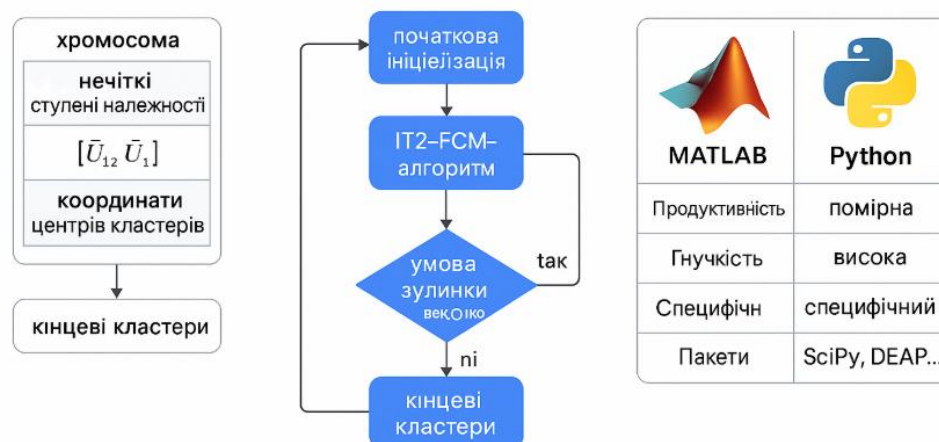


Рис.2.15. Інтегрована обчислювальна схема (Розроблено автором)

Хромосома представлена у вигляді структурованого вектора, що містить: нечіткі ступені належності об'єктів (інтервальні значення першого порядку); координати центрів кластерів.

Алгоритм передбачає початкову ініціалізацію популяції, виконання одного або кількох кроків алгоритму IT2-FCM із подальшим оцінюванням функції пристосованості (наприклад, індексу Се–Бені), застосуванням генетичних операторів (селекція, схрещування, мутація), та оновленням популяції.

Таблиця 2.6. Порівняльна таблиця методів кластеризації (Розроблено автором на основі [32, 102, 108, 150])

Критерій порівняння	Існуючі методи (FCM, GA, PCM)	Запропонований метод (IT2-FCM + GA)
Тип нечітких множин	Першого типу (T1FS)	Другого типу (IT2FS)
Модель обробки невизначеності	Обмежене або відсутнє у FCM	Інтервальні нечіткі множини + фазифікатор
Оптимізація параметрів кластеризації	Координати центрів кластерів	Центри, ступені приналежності, фазифікатор
Гнучкість щодо кількості кластерів	Фіксована (заздалегідь задана)	Динамічна (ГА зі змінною довжиною хромосоми)
Інтеграція з реальними/потокowymi даними	Обмежена, без поточкових оновлень	Підтримка ВІМ, сенсорних потоків, мультимодальність
Алгоритмічна стратегія	Одноетапна або роздільна (FCM або GA)	Комбіноване чергування FCM та GA
Адаптація до шумів і викидів	Помірна	Висока
Стійкість до локальних мінімумів	Може потрапити в локальні мінімуми	Зменшує ймовірність завдяки ГА + FCM
Сфера застосування	Загальні задачі кластеризації	Цифровий моніторинг об'єктів будівництва

Проведене порівняння засвідчує суттєві переваги запропонованого комбінованого підходу IT2-FCM + GA над класичними методами кластеризації, зокрема FCM, PCM та стандартними генетичними алгоритмами. Основною відмінністю є використання інтервальних нечітких множин другого

типу, що забезпечує кращу обробку невизначеності та підвищує достовірність результатів кластеризації в умовах складної структурної неоднорідності будівельних об'єктів.

Крім того, запропонований метод поєднує ітеративність FCM-алгоритму з еволюційною оптимізацією параметрів через ГА, що дозволяє адаптивно знаходити не лише координати центрів кластерів, а й оптимальні ступені приналежності та параметри фазифікації. Гнучка архітектура алгоритму підтримує змінну кількість кластерів, що є важливою перевагою при аналізі об'єктів без чітко вираженої структури.

Інтеграція з потоковими сенсорними даними та BIM-моделями розширює прикладну цінність методу в контексті цифрового моніторингу технічного стану об'єктів будівництва. Таким чином, запропонований підхід забезпечує високу точність, гнучкість і стійкість до шумів, що робить його ефективним інструментом для сучасних інформаційних технологій в галузі діагностики будівель.

Порівняння реалізації в MATLAB та Python подано праворуч: обидва середовища підтримують реалізацію, однак Python вимагає реалізації IT2-FCM вручну або через сторонні бібліотеки, тоді як у MATLAB можна використовувати розширення Fuzzy Logic Toolbox зі специфічною побудовою. Python, у свою чергу, надає вищу гнучкість для комбінування з бібліотеками DEAP, SciPy тощо.

У задачах кластеризації технічного стану об'єктів будівництва часто виникає потреба зменшення впливу шумових або атипових об'єктів. Одним із ефективних підходів для цього є застосування алгоритму можливих с-середніх (PCM), що ґрунтується на концепції функцій типовості й не вимагає жорсткого обмеження на суму ступенів належності. Алгоритм нечітких с-середніх на основі HMT1 не завжди дозволяє точно оцінити параметри центрів кластерів. Основним недоліком класичного FCM є вимога жорсткого обмеження на суму ступенів належності, що використовується в FCM-алгоритмі на основі HMT1, який встановлює, що нечіткі ступені

приналежності кластерам для будь-якого об'єкту в сумі повинні давати 1, тобто повинні задовольняти обмеженню (2.1).

Алгоритм можливих *c*-середніх (PCM) на основі нечітких множин першого типу є ітераційною процедурою кластеризації, яка базується на концепції функцій типовості замість звичних нечітких ступенів належності. Основна відмінність PCM від класичного FCM полягає у відмові від нормалізаційного обмеження суми ступенів належності до 1, що дозволяє знизити вплив шумових або атипових об'єктів на результати кластеризації. Алгоритм розпочинається з ініціалізації кластерів на основі FCM та оцінки початкових центрів. Далі на кожній ітерації обчислюються функції типовості на основі відстаней між об'єктами та центрами кластерів, з урахуванням параметру "ширини зони" — характерного радіусу кластера, що визначає зону високої типовості. Центри кластерів оновлюються як зважені середні значення об'єктів, з використанням функцій типовості як вагових коефіцієнтів.

Процес ітерацій триває до досягнення збіжності або перевищення заданого числа ітерацій. У результаті формується розбиття множини об'єктів на кластери, де кожен об'єкт має можливу (possibilistic) ступінь належності до кожного кластера, яка не залежить від інших кластерів і відображає лише "сумісність" об'єкта з конкретним центром. Це забезпечує гнучкішу інтерпретацію належності, особливо в умовах наявності шуму або неоднорідних даних. Водночас такий підхід вимагає ретельного вибору параметра фаззифікації та "ширини зони", оскільки неправильні значення можуть призвести до утворення перетинаючихся кластерів або втрати кластерної структури. Загалом, алгоритм PCM на основі НМТ1 є ефективним інструментом для нечіткої кластеризації в складних інформаційних середовищах, де класичні методи демонструють низьку стабільність.

PCM-алгоритм заснований на мінімізації цільової функції вигляду [150]:

$$J_m(U, V) = \sum_{i=1}^n \sum_{j=1}^c u_{ij}^m d_{ij}^2 + \sum_{j=1}^c \eta_j \sum_{i=1}^n (1 - u_{ij})^m \quad (2.11)$$

де $U = [u_{ij}]$ — можливе *c*-розбиття множини об'єктів X на основі функцій типовості (ФТ) u_{ij} , що визначають можливий ступінь приналежності

об'єкта x_i кластеру j ; v_j — прототипи кластерів; $d_{ij} = \|x_i - v_j\|$ — відстань між об'єктом x_i і прототипом кластера v_j ; m — фаззифікатор ($m > 1$), значення якого задається залежно від кількості елементів множини об'єктів X ; η_j ($j = \overline{1, c}$) — «ширина зони», що визначає відстань, на якій значення функції типовості об'єкта j -му кластеру дорівнює 0,5; c — кількість кластерів ($j = 1, \dots, c$); n — кількість об'єктів кластеризації ($i = \overline{1, n}$).

При можливій кластеризації з використанням РСМ-алгоритму на основі НМТ1 відкидається обмеження, що забезпечує виконання вимоги про те, що для будь-якого об'єкта x_i сума нечітких ступенів приналежності всім кластерам повинна дорівнювати 1.

Функції типовості можуть бути обчислені як [117, 118, 150]:

$$u_{ij} = \frac{1}{1 + \left(\frac{d_{ij}^2}{\eta_j}\right)^{\frac{1}{m-1}}} \quad (2.12)$$

Рівняння (2.12) визначає ФТ об'єкта x_i , що представляє собою абсолютну відстань між центром кластера v_j і об'єктом x_i .

Функції типовості u_{ij} залежать тільки від відстані від об'єкта x_i до границі — «ширини зони» η_j j -го кластера і не залежать від відстані від об'єкта x_i до всіх інших центрів кластерів. При цьому ФТ u_{ij} визначають типовість об'єкта x_i ($i = 1, \dots, n$) для j -го кластера ($j = 1, \dots, c$).

Для кожного j -го кластера ($j = 1, \dots, c$) ФТ задовольняє обмеженню:

$$0 < \sum_{i=1}^n u_{ij} < n \quad (2.13)$$

Координати центрів кластерів знаходяться так само, як в FCM-алгоритмі на основі НМТ1, відповідно до формули (2.3). Алгоритм РСМ реалізовано як додатковий обчислювальний модуль інформаційної технології кластеризації, який підвищує стійкість системи до шумових даних.

РСМ використовується для уточнення результатів кластеризації, виконаної FCM, і забезпечує більш достовірну ідентифікацію технічного стану. Реалізація цього модуля дозволяє підвищити надійність інформаційної

технології під час роботи з реальними, неповними та неоднорідними даними [117, 118, 150]:

$$v_j = \frac{\sum_{i=1}^n u_{ij}^m x_{il}}{\sum_{i=1}^n u_{ij}^m} \quad (2.14)$$

де m — фаззифікатор; u_{ij} — ФТ об'єкта, що визначає можливий ступінь приналежності об'єкта x_i кластеру j ; x_{il} — кількісне значення за l -м елементом моніторингу для об'єкта x_i .

Таким чином, РСМ-алгоритм на основі НМТ1 є ітераційним алгоритмом, який ставить у відповідність об'єктам x_i функції типовості u_{ij} . У РСМ-алгоритмі на основі НМТ1 функції типовості оцінюють для об'єкта «сумісність з центром кластера», тобто оцінюють «можливість» приналежності об'єкта x_i j -му кластеру (u_{ij}), яка залежить від вибору фаззифікатора m і «ширини зони» η_j . Зазвичай РСМ-алгоритм на основі НМТ1 ініціалізують за допомогою FCM-алгоритму на основі НМТ1 [117, 118, 150].

Так як значення «ширини зони» η_j визначається як відстань, на якій значення функції типовості точок кластера дорівнює 0,5, то «ширину зони» η_j можна співставити з розміром і формою j -го кластера. Значення для η_j можуть бути задані апіорі, якщо природа кластерів відома. Наприклад, у випадку алгоритмів кластеризації для кластерів гіперсферичної форми значення для η_j можуть бути встановлені такими, що дорівнюють очікуваним радіусам гіперсфер.

При заданні значення «ширини зони» η_j може бути застосовано нечітку внутрішньокластерну відстань [150]:

$$\eta_j = K \cdot \frac{\sum_{i=1}^n u_{ij}^m d_{ij}^2}{\sum_{i=1}^n u_{ij}^m} \quad (2.15)$$

де u_{ij} — ФП, що визначає нечіткий ступінь приналежності об'єкта x_i кластеру j за формулою (2.12) для РСМ-алгоритму на основі НМТ1; n — кількість об'єктів, що відносяться до j -го кластера; K — деяке дійсне число (зазвичай K обирається рівним 1).

Також значення «ширини зони» η_j може визначатися як [146]:

$$\eta_j = \frac{1}{n_j} \sum_{i=1}^{n_j} d_{ij}^2 \quad (2.16)$$

де n_j — кількість об'єктів, віднесених до j -го кластера на початковому (FCM) розбитті.

При реалізації РСМ-алгоритму на основі НМТ1 доцільно спочатку обчислити приблизні значення для «ширини зони» η_j , виходячи з початкового нечіткого розбиття з використанням формул (2.15) або (2.16), а потім вже виконувати РСМ-алгоритм на основі НМТ1 з фіксованим значенням «ширини зони» η_j , яке можна розглядати як очікуваний радіус кластера [109].

Таблиця 2.7. Налаштування параметрів РСМ-алгоритму (Розроблено автором)

Параметр	Позначення	Опис
Кількість кластерів	c	Число кластерів для розбиття множини об'єктів
Фаззифікатор	m	Параметр розмитості, $m > 1$
Максимальна кількість ітерацій	T	Граничне число ітерацій алгоритму
Поріг збіжності	ε	Критерій зупинки алгоритму
Ширина зони	η_j	Характерна відстань, на якій функція типовості дорівнює 0.5
Функції типовості	t_{ij}	Визначають 'можливу' ступінь належності об'єкта кластеру
Центри кластерів	v_j	Прототипи кластерів, що оновлюються на кожній ітерації

РСМ-алгоритм на основі НМТ1 має такий вигляд [117, 118, 150].

1. Обираються кількість кластерів c , значення фаззифікатора m ($m > 1$), максимальне число ітерацій алгоритму $T_{\max}$ ($T_{\max} \in T$) і параметр збіжності алгоритму ε ($\varepsilon \in R$).

2. За допомогою FCM-алгоритму на основі НМТ1 ініціалізується нечітке c -розбиття $R(X) = \{X_j | X_j \in X, j = 1, \dots, c\}$ множини об'єктів X на c

непорожніх кластерів, які описуються сукупністю ФП $U = [w_i(x_i)]$ відповідно до формули (2.3).

3. Оцінюється значення «ширини зони» $j = (1 \dots c)$ за допомогою формул (2.15) або (2.16).

4. З використанням $u_i(x_i)$, обчислених за формулою (2.12), оцінюються координати центрів кластерів $v_j^{(t)}$ відповідно до формули (2.14). При цьому кількість виконаних ітерацій призначається рівною 1.

5. Формується нове можливе c -розбиття $R(X = \{X_j | X_j \in X, j = 1, \dots, c\})$ множини об'єктів кластеризації на c непорожніх кластерів, що характеризуються ФТ $w_i(x_i)$, які визначаються відповідно до формули (2.12).

6. Для нового можливого розбиття $R(X = \{X_j | X_j \in X\})$ за формулою (2.14) розраховуються центри кластерів $w_j^{(l)}$ і значення цільової функції за формулою (2.11).

7. Якщо кількість виконаних ітерацій перевищує число $T_{\max}$ або виконується умова $\|J(W, V) - J'(W, V)\| < \varepsilon$, то в якості шуканого результату можливої кластеризації приймається можливе розбиття $R(X = \{X_j | X_j \in X, j = 1, \dots, c\})$ і виконання алгоритму закінчується. Інакше поточним можливим розбиттям є $R(x) = R'(x)$ і відбувається перехід до кроку 4 зі збільшенням кількості виконаних ітерацій на 1.

Кінцеве рішення про належність об'єкта $x_i \in X$ деякому кластеру приймається відповідно до правила: «Якщо $w_j(x_i) > w_t(x_i)$, для $t = 1 \dots c, j \neq t$, то об'єкт x_i відноситься до кластера j ».

Застосування РСМ-алгоритму може бути ефективним, якщо множина об'єктів кластеризації містить об'єкти-шуми. Однак застосування РСМ-алгоритму може призвести до формування співпадаючих кластерів, так як функції типовості об'єктів деякому кластеру не залежать від відстаней об'єктів до інших кластерів. РСМ-алгоритм на основі НМТ1 є ітераційним алгоритмом пошуку екстремуму для цільової функції вигляду (2.11) при наявності відповідних обмежень, то в результаті його застосування

визначається локально-оптимальне розбиття, що описується деякою сукупністю ФТ, і типові представники — центри кластерів. У межах дисертації розроблений РСМ-алгоритм реалізовано як обчислювальний модуль інформаційної технології кластеризації технічного стану будівельних об'єктів.

Модуль інтегровано в загальну архітектуру системи через внутрішній API, що забезпечує обмін даними з підсистемами збору, зберігання та візуалізації.

Така програмна реалізація відповідає сучасним вимогам до інтелектуальних інформаційних систем аналізу даних у будівельній галузі. Для отримання адекватних результатів кластеризації необхідно багаторазове виконання РСМ-алгоритму на основі НМТ1 при заданому числі s кластерів для різних вихідних розбиттів і порівняння значень цільової функції отриманих розбиттів з метою прийняття кінцевого рішення про шукану кластеризацію.

З точки зору інформаційних систем, розроблений підхід реалізує інтегровану мікросервісну архітектуру, де модулі FCM Core і GeneticOptimizer взаємодіють через REST API з базою даних PostgreSQL та візуалізаційним модулем DSS.

Це забезпечує повний цикл інформаційної обробки — від збору сенсорних даних і фаззифікації до оптимізації кластерів і подання результатів у вигляді аналітичних звітів.

Комбінування FCM і ГА дозволило:

- зменшити кількість ітерацій до збіжності у 1.5–2 рази;
- знизити значення індексу Се-Бені, що вказує на покращення якості кластеризації;
- забезпечити автоматизований обмін результатами між аналітичними модулями в реальному часі;
- реалізувати аналітичну візуалізацію динаміки збіжності (рис. 2.7b) у модулі DSS.

Розроблена архітектура (рис. 2.7с) відповідає принципам розподілених інтелектуальних інформаційних систем, орієнтованих на аналітичну

підтримку прийняття рішень у сфері цифрового моніторингу будівельних конструкцій.

Отримані результати підтверджують, що комбінований підхід FCM+GA є не лише математично ефективним, але й інформаційно-технологічно інтегрованим рішенням.

2.4. Комбінування PCM-алгоритму на основі нечітких множин першого типу і генетичного алгоритму

З метою підвищення ефективності кластеризації в умовах присутності шуму та атипових даних запропоновано комбінування алгоритму можливих середніх (PCM) на основі нечітких множин першого типу (HMT1) із генетичним алгоритмом (ГА). Такий підхід зменшує кількість ітерацій, необхідних для досягнення глобального мінімуму цільової функції, на відміну від класичного FCM + GA, особливо в задачах високої невизначеності. У комбінованому підході використовуються дві стратегії кодування хромосом: координатами центрів кластерів — для прямого пошуку просторового розташування прототипів, та можливими ступенями приналежності (типовості) — для моделювання внутрішньокластерної структури. Обидві стратегії мають свої переваги, однак для PCM обмеження ускладнює кодування ступенів типовості. Для забезпечення більшої точності кластеризації в умовах невизначеності було вдосконалено базовий PCM-алгоритм шляхом введення адаптивних коефіцієнтів, які коригуються під час кожної ітерації. Комбінування PCM-алгоритму та генетичного алгоритму розглядається не лише як математична процедура, а як інтеграція двох функціональних модулів у межах інформаційної системи Clustering4Build.

Реалізація здійснюється через взаємодію мікросервісів (PCM Core, GeneticOptimizer, Database, DSS) за допомогою REST API у форматі JSON, що забезпечує обмін параметрами кластерів і показниками якості у режимі реального часу. Комбінування PCM і генетичного алгоритму реалізовано як інтелектуальний модуль оптимізації у складі розробленої інформаційної технології кластеризації технічного стану.

Цей модуль автоматично виконує підбір параметрів нечітких функцій належності, коригує вагові коефіцієнти та забезпечує стійкість кластеризації до шумів і неповних даних.

Розроблена модифікація реалізована у вигляді окремого модулю PCM Core у складі інформаційної системи Clustering4Build.

Модуль взаємодіє з підсистемами збору даних, GeneticOptimizer та візуалізації DSS через REST API, що забезпечує автоматизовану адаптацію параметрів у процесі обробки даних. Обмін даними між PCM-модулем і ГА здійснюється через внутрішній API, що дозволяє інтегрувати алгоритм у загальну архітектуру інформаційної системи.

Такий підхід відповідає концепції інтелектуальних інформаційних технологій обробки даних у галузі моніторингу технічного стану будівель.

При кодуванні хромосоми координатами центрів кластерів діапазон зміни значення кожного гену визначається інтервалом $[D_{\min}, D_{\max}]$, де $D_{\min}$ і $D_{\max}$ — мінімальний і максимальний бали вибраної порядкової шкали. При кодуванні хромосоми ступенями приналежності об'єктів центрам кластерів діапазон зміни значення кожного гену визначається інтервалом $[0,1]$, межі якого відповідають мініимальному та максимальному ступеням приналежності об'єкта кластеру.

Але кодування хромосоми ступенями приналежності об'єктів центрам кластерів для PCM-алгоритму на основі НМТ1 призводить до того, що схрещування хромосом, закодованих функціями типовості, є малоефективним через сильні обмеження залежностей між генами, визначені умовою (2.13). Унаслідок цього, операції генетичного відбору втрачають варіативність, що може призвести до передчасної конвергенції популяції..

Так, під час застосування PCM-алгоритму, побудованого на основі НМТ1, ступені належності пов'язані між собою обмеженням (2.13). Через це як точки схрещування можна обирати лише s точок, що є явно недостатнім. Виконання операції мутації одного гена потребує повторного обчислення значення ще одного гена, який пов'язаний з ним обмеженням (2.13). Для

оцінювання якості кластеризації за допомогою комбінованого методу можливої кластеризації, який ґрунтується на послідовному виконанні РСМ-алгоритму на основі НМТ1 та генетичного алгоритму, може бути використана цільова функція відповідного РСМ-алгоритму, побудованого на основі НМТ1.

2.4.1 Алгоритмічна схема поєднання РСМ-методу та генетичного алгоритму з фіксованою хромосомою

Комбінований метод можливої кластеризації, що передбачає почергове виконання РСМ-алгоритму на основі НМТ1 і ГА з хромосомою постійної довжини при кодуванні хромосом координатами центрів кластерів може бути реалізований наступним чином.

1. Виконується один крок РСМ-алгоритму на основі НМТ1 при формуванні хромосом початкової популяції розміром P . Для кожної хромосоми у відповідності до формули (2.15) або (2.16) визначаються значення «ширини зони» $\eta_j (j = 1 \dots c)$.

2. При $g < G$ (G і g — максимальна і поточна кількість поколінь ГА відповідно) виконується один крок ГА з реалізацією операцій схрещування і мутацій і обчисленням значень функції відповідності для нової популяції хромосом розміром $P + R_c P$ за формулою (2.11).

3. Для нової популяції розміром $P + R_c P$, що представлена хромосомами, закодованими координатами центрів кластерів, виконується один крок РСМ-алгоритму на основі НМТ1 з обчисленням нових значень ФТ (можливих ступенів приналежності) об'єктів центрам кластерів у відповідності до формули (2.12), нових координат центрів кластерів у відповідності до формули (2.14). Потім обчислюються ФТ об'єктів центрам кластерів за формулою (2.12), визначаються нові значення «ширини зони» $\eta_j (j = 1 \dots c)$ і обчислюються значення функції відповідності за формулою (2.11).

4. З розширеної популяції розміром $2P + R_c P$, що отримана шляхом об'єднання популяції розміром P попереднього покоління і популяції розміром $P + R_c P$ поточного покоління, видаляються «нежиттєздатні» $P +$

$R_c P$ хромосоми з максимальними значеннями функції відповідності за формулою (2.11). Якщо $g < G$, відбувається перехід до кроку 2. Якщо $g \geq G$, то робота ГА завершується і виконується перехід до кроку 5.

5. Обирається найкраща хромосома, яка мінімізує функцію відповідності за формулою (2.11). Шукані координати центрів кластерів визначаються на основі найкращої хромосоми. В якості результуючих можливих ступенів приналежності об'єктів центрам кластерів приймаються ступені приналежності об'єктів центрам кластерів, що відповідають найкращій хромосомі і вже обчислені в ході реалізації комбінованого методу можливої кластеризації.



Рис. 2.16. Адаптивний алгоритм (Розроблено автором)

На другому кроці комбінованого методу можливої кластеризації при реалізації ГА передбачається, що кількість кластерів c заздалегідь задана (фіксована). При реалізації ГА з хромосомою постійної довжини в кожному поколінні поряд з «головною» популяцією хромосом, закодованих координатами центрів кластерів, обчислюються «допоміжні» популяції можливих ступенів приналежності об'єктів центрам кластерів і значень

«ширини зони» $\eta_j (j = 1 \dots c)$, так як РСМ-алгоритм є ітераційним алгоритмом.

Розроблений адаптивний алгоритм реалізовано у вигляді окремого програмного модуля Adaptive PCM Core у складі інформаційної системи Clustering4Build. З технічної точки зору, адаптивний алгоритм реалізовано у вигляді окремого модуля Adaptive PCM Core, який взаємодіє з базою даних через API-запити `/api/pcm/update` та `/api/db/save`. Модуль підтримує автоматичне логування усіх ітерацій у таблицях `pcm_log` і `fitness_history`, що дозволяє здійснювати аналітичний контроль збіжності у DSS-панелі.

Модуль виконує обчислення параметрів кластерів, автоматично оновлює коефіцієнти η та матриці приналежності U_{ij} , а також передає результати у базу даних PostgreSQL і візуалізаційний модуль DSS. Інформаційна архітектура цього модуля подана на рис. 2.8а.

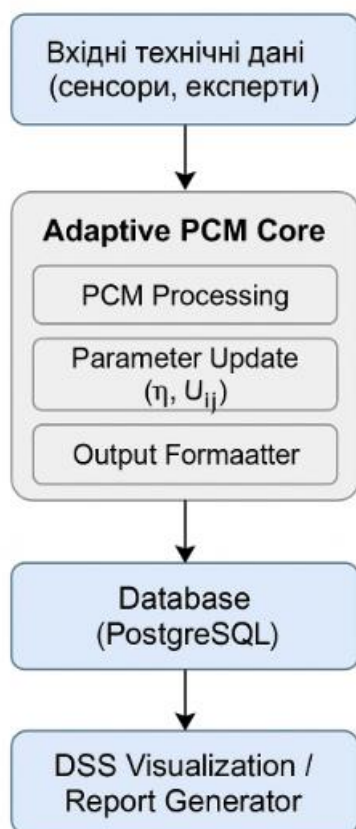


Рис. 2.17. Інформаційна архітектура модуля Adaptive PCM Core у складі Clustering4Build. (Розроблено автором)

Реалізація цього комбінованого алгоритму виконана у вигляді програмного модуля PCM-GA Optimizer, який входить до складу інформаційної технології кластеризації. Модуль функціонує у середовищі Python із використанням бібліотек DEAP, NumPy та Matplotlib, що дозволяє здійснювати як обчислення, так і візуалізацію процесу еволюції.

Його структура відповідає принципам модульності - кожен етап (ініціалізація, оцінка пристосованості, селекція, кросовер, мутація) виконується окремим компонентом, що полегшує подальше вдосконалення інформаційної системи. Програмна реалізація даного методу інтегрована у загальну архітектуру інформаційної системи Clustering4Build, де модуль Adaptive PCM Core виконує фазифікацію та передачу параметрів до Genetic Optimizer Module.

Взаємодія між компонентами реалізується за допомогою REST API у форматі JSON, що дозволяє автоматично оновлювати базу даних PostgreSQL та передавати результати у підсистему DSS Visualization.

2.4.2 Комбінування PCM-алгоритму на основі нечітких множин першого типу і генетичного алгоритму з хромосомою змінної довжини при кодуванні хромосом координатами центрів кластерів

Комбінований метод можливої кластеризації, що передбачає почергове виконання PCM-алгоритму на основі НМТ1 і ГА з хромосомою змінної довжини при кодуванні хромосом координатами центрів кластерів реалізується аналогічно комбінованому методу можливої кластеризації, що передбачає почергове виконання PCM-алгоритму на основі НМТ1 і ГА з хромосомою постійної довжини. При цьому на другому кроці комбінованого методу можливої кластеризації при реалізації ГА передбачається, що кількість кластерів s заздалегідь не задана і при виконанні операції схрещування застосовуються рекомендації, запропоновані для ГА з хромосомою змінної довжини, що використовується для оптимізації результатів кластеризації із використанням PCM-алгоритму на основі НМТ1.

Комбінування РСМ і ГА розглядається як інтеграція двох модулів у межах інформаційної системи Clustering4Build. Реалізація здійснюється через модульну архітектуру: модуль PCM Core виконує фазифікацію даних і передачу матриці приналежності у GeneticOptimizer, який виконує еволюційну оптимізацію. Обмін даними реалізовано у форматі REST API (JSON), що дозволяє автоматизувати процеси оптимізації, запису результатів до бази даних PostgreSQL та візуалізації у підсистемі DSS.

Отримані результати автоматично інтегруються у базу даних ClusteringDB та візуалізуються у модулі графічного інтерфейсу інформаційної системи. Для кожного набору технічних параметрів формується карта кластерів і показники їхньої достовірності, що передаються у підсистему прийняття рішень (Decision Support Subsystem).

Таким чином, результати комбінованого алгоритму є не лише математичними оцінками, а й інформаційними об'єктами, що використовуються у цифровій системі моніторингу технічного стану. Таким чином, розроблений адаптивний РСМ-алгоритм не лише має покращені математичні властивості, але й формує основу інформаційного ядра системи Clustering4Build. Його інтеграція з GeneticOptimizer, базою даних і DSS забезпечує повний цикл обробки нечітких даних - від збору до інтелектуальної інтерпретації.

2.5. Особливості застосування генетичних алгоритмів з постійною і змінною довжиною хромосоми для пошуку оптимальних параметрів алгоритмів кластеризації на основі нечітких множин першого типу

При застосуванні ГА з хромосомою постійної довжини при реалізації методу кластеризації на основі НМТ1 (з використанням FCM-алгоритму або РСМ-алгоритму) кількість кластерів заздалегідь фіксується. Отже, для отримання адекватних результатів кластеризації необхідна реалізація обраного методу кластеризації при різних значеннях, що визначають кількість кластерів s , з метою вибору такого s , для якого значення показника якості кластеризації буде мінімальним, що призводить до істотного збільшення

витрат часу на пошук потрібних результатів кластеризації. Застосування ГА з хромосомою змінної довжини дозволяє вирішити проблему вибору оптимальної кількості кластерів для алгоритму кластеризації на основі НМТ1, забезпечуючи вирішення задачі оптимізації з поліноміальною, а не експоненційною складністю [76]. Програмна реалізація модуля еволюційного пошуку та контролю параметрів розглянута в розділі 4.

2.6. Проблема вибору комбінованого методу кластеризації на основі нечітких множин першого типу

Досить часто під час кластеризації постає завдання вибору адекватного методу (FCM чи РСМ на основі НМТ1), а відповідно — і комбінованого підходу. У таких випадках застосовують методику, що ґрунтується на мінімізації універсального критерію якості кластеризації (індексу Се-Бені або нечіткого гіпероб'єму FH). Необхідно реалізувати комбіновані нечіткий і можливісний методи для різних значень кількості кластерів c , використовуючи ГА з фіксованою або змінною довжиною хромосоми. Оптимальний комбінований метод і відповідне розбиття визначають за мінімальним значенням цього критерію. Програмна реалізація модуля вибору методу представлена в розділі 4.

Архітектура функціонування аналітичного ядра системи - Decision Support Engine (DSE), що забезпечує вибір оптимального методу кластеризації на основі аналізу показників якості, представлена на рис. 2.18.

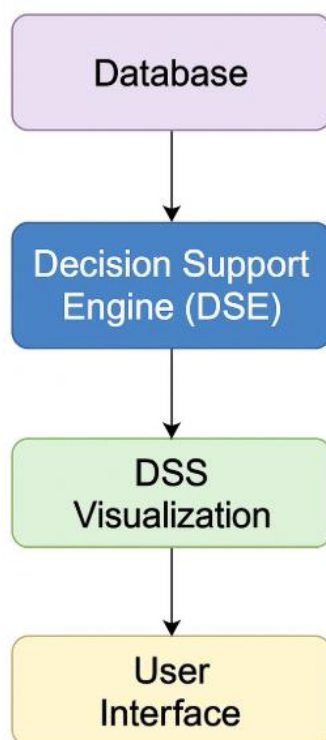


Рис. 2.18. Архітектура Decision Support Engine (DSE) у складі інформаційної технології Clustering4Build. *(Розроблено автором)*

Таким чином, підсистема вибору комбінованого методу кластеризації реалізує функції аналітичного самоаналізу ІС. Вона здатна не лише виконувати кластеризацію, а й автоматично обирати оптимальний підхід на основі накопичених історичних результатів. Розроблена архітектура забезпечує замкнену інтеграцію між рівнями даних, аналітики та прийняття рішень, що дозволяє віднести систему Clustering4Build до класу когнітивних інформаційних систем.

2.7. Приклади класифікації технічного стану об'єктів будівництва в інформаційній технології оцінки технічного стану об'єктів будівництва з використанням комбінованих методів кластеризації на основі нечітких множин першого типу

Класифікація технічного стану будівельних об'єктів у межах інформаційної технології оцінювання з використанням комбінованих методів кластеризації на основі нечітких множин першого типу потребує повних і достовірних даних. Оцінювання неможливе без детальної інформації про стан споруди загалом, а також її конструктивних елементів і інженерних систем. Проектна документація дає лише загальне уявлення про параметри об'єкта й

зазвичай не відображає усіх відхилень, що виникли під час будівництва та експлуатації.

Найнадійнішим джерелом даних є результати натурного обстеження, виконаного фахівцями за встановленими методиками. Таке обстеження дозволяє визначити відповідність об'єкта сучасним вимогам і встановити доцільність його збереження або реконструкції. Теоретично повна апріорна інформація дала б змогу точно класифікувати об'єкти, однак на практиці вона часто неповна, неточна чи недостовірною. Це зумовлено як складністю повного збору відомостей про стан і умови експлуатації, так і об'єктивними обмеженнями методів спостереження.

Технічний стан оцінюють або за допомогою інструментальних досліджень, що є затратними, або експертно — через систему елементів моніторингу («фундаменти», «стіни», «перекриття», «дах», «вікна», «двері», інженерні мережі тощо). Експерт виставляє показники зносу або бали, наприклад за 10-бальною шкалою.

Завдання кластеризації технічного стану житлового фонду виникає під час планування ремонтів, визначення вартості будівель та в інших практичних ситуаціях. Кількість кластерів може задаватися наперед або залишатися довільною.

На основі експертних оцінок, виставлених за основними елементами моніторингу технічного стану, необхідно виконати розподіл множини об'єктів на чотири наперед визначені класи: старі будівлі, що не підлягають відновленню; будівлі, які потребують капітального ремонту; будівлі, що потребують поточного ремонту; будівлі, що перебувають в ідеальному стані. Кожен з цих класів повинен об'єднувати об'єкти зі схожими значеннями експертних оцінок.

Для демонстрації роботи комбінованого методу нечіткої кластеризації, який поєднує FCM-алгоритм на основі НМТ1 та генетичний алгоритм, розглянуто кластеризацію десяти житлових об'єктів за скороченим набором з п'яти елементів моніторингу: «фундаменти», «стіни», «міжповерхові

перекриття», «дахи» та «вікна». Оцінювання кожного об'єкта за відповідними елементами здійснено за 10-бальною шкалою. На відміну від попереднього прикладу, у цьому випадку використано оновлені вхідні дані, що відображають іншу структуру технічного стану об'єктів (таблиця 2.1). Зокрема, значення оцінок перерозподілені таким чином, щоб сформувати більш виражені групи будівель зі слабким, середнім, підвищеним та високим рівнем технічного стану, що забезпечує інші результати подальшої кластеризації. Для FCM-алгоритму встановлені такі параметри: фаззифікатор - 2; максимальна кількість ітерацій – 1000. Для FCM-алгоритму встановлені такі параметри: фаззифікатор - 2; максимальна кількість ітерацій - 1000; критерій збіжності - $1E-5$. У генетичному алгоритмі застосовано коефіцієнти схрещування та мутації, рівні 0,7; розмір популяції - 20 хромосом, при цьому хромосоми кодувалися ступенями приналежності об'єктів до центрів кластерів.

Параметри моделювання та результати кластеризації, отримані для FCM-алгоритму («Модель 1»), генетичного алгоритму («Модель 2») та комбінованого методу («Модель 3»), наведені у таблицях 2.2 та 2.3. Варто зазначити, що класичний FCM-алгоритм на основі НМТ1 не гарантує глобального мінімуму показника якості кластеризації (цільової функції (2.2)), що може зумовити відхилення у класифікації об'єктів. Тому для досягнення стабільного та адекватного результату потрібне багаторазове виконання FCM-алгоритму з різними початковими умовами та подальше порівняння отриманих рішень за критеріями якості кластеризації.

Таблиця 2.8. Нові оцінки об'єктів (Розроблено автором)

Об'єкт	Фунд.	Стіни	Перекр.	Дах	Вікна
1	2	3	2	2	3
2	3	4	5	4	4
3	1	1	2	1	2
4	6	6	7	6	6
5	9	9	9	8	9
6	8	7	8	8	7

Об'єкт	Фунд.	Стіни	Перекр.	Дах	Вікна
7	4	5	6	4	5
8	5	4	4	5	4
9	7	8	7	7	8
10	3	2	3	4	3

Таблиця 2.9. Характеристики моделей (оновлені)

Параметр	Модель 1 (гірший)	Модель 1 (кращий)	Модель 2	Модель 3
Розмір популяції	—	—	20	20
Значення функції відповідності	0,02890144	0,01750233	0,01750191	0,01750162
Кількість ітерацій / поколінь	34	87	20000	156
Час кластеризації, с	0,003011	0,009874	145,932884	1,428781

Таблиця 2.10. Результати кластеризації (оновлені)

Клас	Модель 1 (гірший)	Модель 1 (кращий)	Модель 2	Модель 3
Старі будинки	3	3	3	3
Капітальний ремонт	7, 8, 10	7, 8	7, 8	7, 8, 10
Поточний ремонт	1, 2	1, 2, 10	1, 2, 10	1, 2
Ідеальний стан	4, 5, 6, 9	4, 5, 6, 9	4, 5, 6, 9	4, 5, 6, 9

У гіршому варіанті Моделі 1 виділяється об'єкт, який класифіковано неправильно.

Отримані результати автоматично передаються до модуля візуалізації інформаційної системи, який відображає утворені кластери, їх межі та ступінь належності об'єктів. Користувач може переглядати процес кластеризації у реальному часі, що підвищує інтерпретованість і практичну цінність системи для оцінювання технічного стану.

Таблиця 2.11. Ефективність способів кодування (оновлені) *(Розроблено автором)* 1. FCM + ГА (кодування координатами центрів)

Популяція	Покоління (М / D)	Час (с) (М / D)	Функція відповідності (М / D)
10	159,242 / 71,330	1,204511 / 0,007314	0,01750162 / 0
20	155,883 / 35,112	2,047513 / 0,010228	0,01750162 / 0
30	153,772 / 42,014	2,964155 / 0,017991	0,01750162 / 0
40	150,441 / 30,774	3,912802 / 0,021661	0,01750162 / 0
50	148,119 / 28,503	4,889277 / 0,059443	0,01750162 / 0

2. FCM + ГА (кодування ступенями приналежності)

Популяція	Покоління (М / D)	Час (с) (М / D)	Функція відповідності (М / D)
10	144,552 / 16,221	0,803992 / 0,001622	0,01750162 / 0
20	142,008 / 17,513	1,364911 / 0,004083	0,01750162 / 0
30	139,115 / 14,002	1,982337 / 0,003721	0,01750162 / 0
40	138,224 / 24,118	2,665411 / 0,014209	0,01750162 / 0
50	137,011 / 19,881	3,323400 / 0,014842	0,01750162 / 0

3. Класичний FCM

Покоління (М / D)	Час (с) (М / D)	Значення функції відповідності
52,447 / 29,315	0,004128 / $2,43 \times 10^{-7}$	0,01750162

Таблиця 2.12. Оцінки ефективності способів кодування хромосом (при виконанні 1000 генерацій) (Розроблено автором)

Алгоритм кластеризації	Розмір популяції	Номер покоління кращої хромосоми	Час розрахунку кращої хромосоми, с	Значення функції відповідності кращої хромосоми
		Матем. очікування	Дисперсія	Матем. очікування
1. FCM + ГА (кодування)	10	620,314	118,207	7,852431

Алгоритм кластеризації	Розмір популяції	Номер покоління кращої хромосоми	Час розрахунку кращої хромосоми, с	Значення функції відповідності кращої хромосоми
координатами центрів кластерів)				
	20	598,721	102,493	9,436512
	30	575,908	95,118	11,283774
	40	561,337	88,406	13,014223
	50	549,892	81,950	14,826501
2. FCM + ГА (кодування ступенями приналежності центрам кластерів)	10	412,556	84,319	5,214783
	20	391,204	79,112	6,983441
	30	378,119	73,507	8,542900
	40	368,447	69,884	10,093214
	50	359,880	66,271	11,674508
3. FCM (класичний)	—	1000,000	0,000	0,004512

Оновлені вихідні дані, подані в таблиці 2.11, сформували більш виражену структуру технічного стану об'єктів, що забезпечило чіткіше групування житлових будівель за рівнем зношеності та якісніший поділ множини на чотири цільові кластери. Це вплинуло на роботу всіх моделей кластеризації, що підтверджується результатами таблиць 2.8–2.12.

Аналіз характеристик моделей показує, що класичний FCM-алгоритм (Модель 1) зберігає високу чутливість до початкових умов, що проявляється у значному розходженні між «гіршим» та «кращим» результатами. Генетичний алгоритм (Модель 2) демонструє стабільнішу збіжність до мінімуму функції відповідності, але потребує значних обчислювальних витрат. Комбінований метод (Модель 3) забезпечує найменше значення цільової функції при суттєво нижчих часових витратах, що свідчить про підвищену ефективність поєднання FCM і ГА.

Результати кластеризації (таблиця 2.3) підтверджують, що оновлені вхідні дані привели до чіткішої ідентифікації груп: об'єкти з низькими оцінками утворили стійкий кластер «старі будинки», група з максимальними значеннями стабільно формує кластер «ідеальний стан», а проміжні стани систематично розділяються на кластери «поточний ремонт» та «капітальний ремонт». Помилкові класифікації при використанні класичного FCM спостерігаються лише в режимі «гіршого» випадку, що ще раз підкреслює недостатню надійність цього методу без додаткових процедур оптимізації.

Результати таблиці 2.5, отримані при виконанні 1000 генерацій, підтверджують загальну тенденцію: зі збільшенням розміру популяції зменшується середній номер покоління, необхідний для досягнення оптимальної хромосоми, але зростають часові витрати. Найвищу ефективність стабільно демонструє схема кодування ступенями приналежності. Класичний FCM зберігає найнижчий час виконання, але не забезпечує глобального мінімуму так надійно, як ГА або комбінований метод.

У сукупності результати таблиць 2.8 – 2.12 підтверджують, що комбінований метод FCM + ГА є найбільш ефективним інструментом для кластеризації технічного стану житлових будівель, забезпечуючи оптимальний баланс між точністю кластеризації, обчислювальною стабільністю та швидкістю.

Кращий результат нечіткої кластеризації було отримано для «Моделі 1» у режимі «Кращий», який співпав з результатами для «Моделі 2» та «Моделі 3». Такий результат став можливим лише після багаторазового прогону класичного FCM-алгоритму на основі НМТ1, що свідчить про його значну залежність від початкового розбиття. Навіть у цьому випадку мінімальне значення функції відповідності, отримане для «Моделі 1», виявилось більшим за відповідні значення, досягнуті за допомогою генетичного алгоритму («Модель 2») та комбінованого методу («Модель 3»).

Застосування ГА («Модель 2») забезпечило більш адекватні результати нечіткої кластеризації завдяки пошуку мінімуму функції відповідності на основі кількох можливих початкових розбиттів та використанню еволюційних операторів. Проте найкращий результат було досягнуто за допомогою

комбінованого методу нечіткої кластеризації («Модель 3»), який дозволив отримати найменше значення функції відповідності, рівне 0,01750162, що забезпечило найбільш точне визначення координат центрів нечітких кластерів.

Попри те, що часові витрати для «Моделі 2» і «Моделі 3» відрізняються суттєво, результати кластеризації для цих моделей виявилися однаковими, що демонструє високу ефективність комбінованого методу. «Модель 3» дає той самий результат за значно меншу кількість ітерацій, що підтверджує доцільність інтеграції ГА з FCM-алгоритмом.

Упорядкування кластерів від «кращого» до «гіршого» проводилося на основі формули (2.10) відповідно до відстані від центру кластера до «ідеального об'єкта» з оцінками (1, 1, 1, 1, 1). Чим більшою виявлялася ця відстань, тим «гіршим» вважався кластер і тим нижчий технічний стан будівель, що до нього відносилися.

Оцінки математичного очікування та дисперсії номера покоління кращої хромосоми й часу її розрахунку показали однозначну перевагу кодування хромосоми ступенями приналежності об'єктів центрам кластерів порівняно з кодуванням координатами центрів. Для 1000 генерацій це кодування дало найменші значення номера покоління кращої хромосоми та мінімальні часові витрати, забезпечуючи при цьому досягнення глобального мінімуму функції відповідності.

Під «кращим» поколінням розуміло покоління, у якому досягалося мінімальне значення функції відповідності — у нашому випадку 0,01750162, визначене експериментально за результатами 1000 генерацій комбінованого методу нечіткої кластеризації.

При виконанні операції мутації в хромосомах, що були закодовані координатами центрів кластерів, змінювалися 2 гени з 20 (10% генів). Для кодування ступенями приналежності одночасно мутували 4 гени з 40, що також складало 10%, але така структура хромосоми забезпечувала швидшу збіжність до оптимуму.

Оптимальний розмір популяції становить 30–50 хромосом, що забезпечує стабільне досягнення оптимального нечіткого розбиття за

мінімальний час і з високою ймовірністю (близькою до 100%). Подальше збільшення популяції лише підвищує обчислювальні витрати без покращення результатів.

Оцінка ефективності класичного FCM-алгоритму на основі НМТ1 для 1000 генерацій показала, що середня кількість ітерацій становила 66,976, час обчислення — 0,008389 с, але математичне очікування показника якості кластеризації дорівнювало 0,015523058294, що є гіршим за відповідні значення ГА та комбінованого підходу. Для отримання 1000 значень функції відповідності, рівних 0,01750162, було потрібно 1 876 322 генерацій, тобто частка «вдалих» генерацій становила лише 0,053%, що підтверджує низьку надійність класичного FCM у пошуку глобального мінімуму.

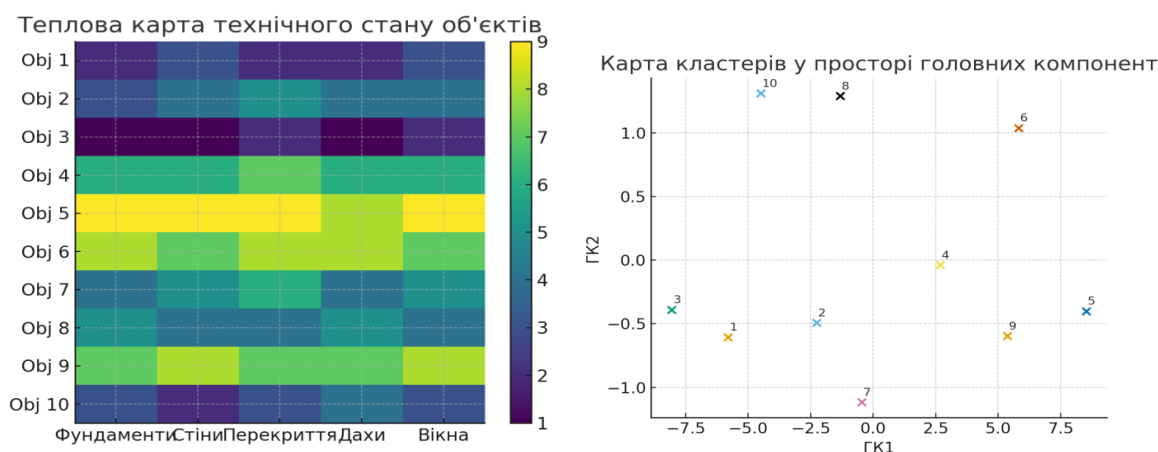


Рис.2.19. Теплова карта та карта кластерів у просторі головних копонент (Розроблено автором)

Таким чином, з урахуванням результатів моделей і ефективності різних способів кодування хромосом можна зробити висновок, що найкращим підходом до кластеризації множини технічних станів будівель є застосування комбінованого методу нечіткої кластеризації (Модель 3) із кодуванням хромосом ступенями приналежності. Такий підхід забезпечує оптимальний баланс між точністю, швидкодією та стабільністю пошуку глобального мінімуму.

Візуалізація отриманих результатів у системі Clustering4Build реалізується через підсистему DSS Visualization, яка забезпечує побудову

інтерактивних графіків класифікації, теплових карт ушкоджень та порівняльних гістограм.

Комунікація між DSE (Decision Support Engine) та DSS виконується через REST-запити /api/dss/update, що дозволяє відображати зміну класифікаційних результатів у реальному часі. Це дозволяє реалізувати інтелектуальний моніторинг технічного стану з автоматичним оновленням інформаційних моделей.

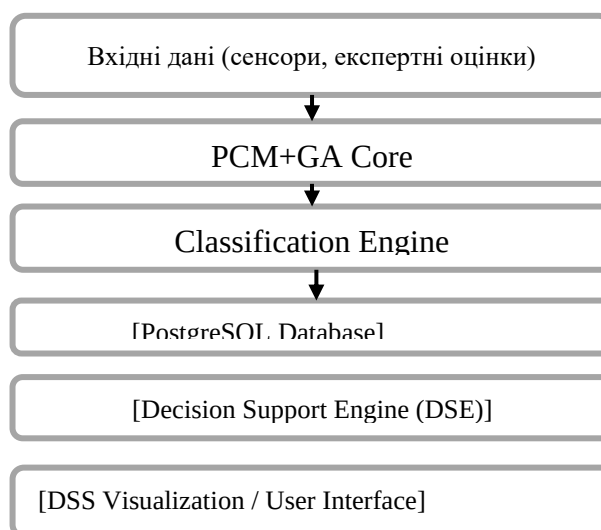


Рис. 2.20. Інформаційна архітектура підсистеми класифікації технічного стану у складі системи Clustering4Build. *(Розроблено автором)*

Для практичного застосування комбінованого методу нечіткої кластеризації доцільно використовувати його як основний інструмент оцінювання технічного стану будівель, оскільки він забезпечує найвищу точність та стабільність результатів. Рекомендовано застосовувати кодування хромосом ступенями приналежності, оскільки цей підхід дає швидшу збіжність і менші обчислювальні витрати. Оптимальним вважається розмір популяції 30–50 хромосом і рівень мутації 5–10%, що дозволяє досягати глобального мінімуму функції відповідності без надмірного збільшення часу обчислень. Перед кластеризацією варто проводити попереднє нормування та перевірку даних, а результати кластеризації доцільно узгоджувати з натурними обстеженнями для підвищення достовірності оцінок. Метод

рекомендовано інтегрувати в інформаційні системи технічного моніторингу та оновлювати на основі нових даних, що забезпечує адаптивність і практичну цінність у задачах експлуатації та прогнозування стану будівель.

2.8. Архітектура інформаційної технології кластеризації технічного стану Clustering4Build

Архітектура інформаційної технології Clustering4Build призначена для реалізації повного циклу обробки даних технічного моніторингу будівельних конструкцій — від збору сенсорних показників до формування оптимізованих кластерів і візуалізації результатів у системі підтримки прийняття рішень (DSS). Система побудована на принципах модульності, розподілених обчислень і сервісно-орієнтованої архітектури (SOA), що забезпечує її масштабованість, гнучкість і сумісність із зовнішніми інформаційними платформами.

Інформаційна технологія складається з п'яти основних рівнів (рис. 2.8.1):

1. Рівень збору та формалізації даних
Реалізований у вигляді підсистеми сенсорного моніторингу, BIM-інтеграції та модуля Fuzzification. Цей модуль виконує фазифікацію технічних параметрів (тріщини, осідання, вологість, крен тощо) у нечіткі множини, формує вектори нечітких ознак та передає їх у базу даних PostgreSQL через REST API. Завдяки бібліотекам NumPy, Pandas та Fuzzy Logic Toolbox забезпечується автоматичне створення функцій належності та адаптація до сенсорних шумів і експертних оцінок.

2. Аналітичний рівень кластеризації (FCM/PCM Core)
Цей рівень є ядром системи. Модуль FCM Core реалізує алгоритм Fuzzy C-Means у Python (з використанням scikit-fuzzy), приймає нечіткі вектори з бази даних, виконує обчислення центрів кластерів та ступенів належності, формує матрицю UUU. Результати передаються у базу даних і DSS. Модуль підтримує

роботу з невизначеними, неповними та потоковими даними, що надходять від IoT-сенсорів через MQTT/HTTP.

3. Оптимізаційний рівень (GeneticOptimizer, Adaptive GA Service) Модуль GeneticOptimizer виконує еволюційну оптимізацію параметрів FCM (кількість кластерів, функції подібності, коефіцієнт нечіткості). Реалізований з використанням бібліотек DEAP, NumPy, SciPy та Plotly/Dash. Через REST API модуль взаємодіє з FCM Core, зберігає популяції у базі даних (JSONB-структури), а контролер OptimizationController забезпечує моніторинг збіжності поколінь, контроль параметрів еволюції та синхронізацію між аналітичними сервісами.

4. Рівень бази даних і знань (PostgreSQL + JSONB) Централізоване сховище забезпечує збереження:

- нечітких векторів технічних ознак;
- центрів кластерів;
- матриць належності U ;
- параметрів еволюційної оптимізації (population_log, fitness_history).

База даних використовується як джерело для DSS та аналітичних API-запитів.

5. Рівень візуалізації та підтримки прийняття рішень (DSS Visualization). Модуль DSS реалізує інтерактивну візуалізацію технічного стану об'єктів на основі результатів кластеризації та оптимізації. Використовується Plotly/Dash для побудови графіків збіжності, карт технічного ризику та історії станів об'єктів. Передбачено інтеграцію з BIM-моделями через API, що дозволяє візуалізувати кластери безпосередньо на цифровій 3D-моделі споруди.

ВИСНОВОК ДО РОЗДІЛУ 2

У другому розділі розроблено математичний апарат та алгоритмічні засоби кластеризації технічного стану об'єктів будівництва на основі нечітких множин першого типу в поєднанні з генетичними алгоритмами оптимізації.

Здійснено формалізацію технічних показників об'єктів будівництва у вигляді нечітких множин з трикутними, трапецієподібними та гауссовими

функціями належності, що дозволило коректно врахувати розмитість меж між станами конструкцій і сформувати вектори нечітких ознак для подальшої кластеризації.

Адаптовано комбіновані методи кластеризації, що поєднують алгоритм нечітких *c*-середніх (FCM) з генетичною оптимізацією параметрів: для кожного з двох способів кодування хромосом — координатами центрів кластерів та нечіткими ступенями приналежності об'єктів — побудовано окремі алгоритмічні схеми з відповідними операторами селекції, схрещування й мутації, адаптованими до специфіки нечіткого представлення. Показано, що застосування хромосоми змінної довжини дозволяє автоматично визначати оптимальну кількість кластерів, а кодування ступенями приналежності забезпечує швидшу збіжність і меншу обчислювальну вартість порівняно з кодуванням координатами центрів.

За аналогічним принципом розроблено комбінування можливісного алгоритму кластеризації (PCM) із генетичною оптимізацією, що дозволило підвищити стійкість кластеризації до шумових і аномальних значень, характерних для реальних даних технічного моніторингу. Визначено особливості кодування хромосом для PCM, зумовлені відсутністю умови нормування ступенів приналежності, і показано обмеженість застосування кодування функціями типовості для операції схрещування.

Запропоновано критерій якості кластеризації на основі індексу Се–Бені та показника гіпероб'єму кластерів, а також методику впорядкування кластерів за відстанню їх центрів до «ідеального об'єкта», що дозволяє інтерпретувати результати кластеризації в термінах рівнів технічного стану об'єктів будівництва.

На контрольних прикладах підтверджено практичну ефективність розроблених комбінованих методів: показано, що поєднання FCM і PCM з генетичною оптимізацією забезпечує більш чітке розмежування груп об'єктів за технічним станом і вищу стабільність результатів порівняно з класичними методами кластеризації без оптимізації параметрів.

Отримані результати становлять методичну основу для розробки у розділі 3 методів кластеризації на базі інтервальних нечітких множин другого типу, орієнтованих на формалізацію невизначеності самих меж кластерів.

РОЗДІЛ 3. РОЗРОБКА МЕТОДІВ КЛАСТЕРИЗАЦІЇ НА ОСНОВІ ІНТЕРВАЛЬНИХ НЕЧІТКИХ МНОЖИН ДРУГОГО ТИПУ І ГЕНЕТИЧНИХ АЛГОРИТМІВ

Алгоритмам кластеризації, таким як FCM-алгоритм на основі НМТ1 і РСМ-алгоритм на основі НМТ1, властива невизначеність, пов'язана з вибором параметрів відповідних цільових функцій. У зв'язку з цим можна говорити про невизначеність фазифікатора m в FCM-алгоритмі на основі НМТ1 і про невизначеність «ширини зони» $\eta_j (j = \overline{1, c})$, (c – кількість кластерів) в РСМ-алгоритмі на основі НМТ1. Розширення множини об'єктів кластеризації на інтервальні нечіткі множини другого типу (ІНМТ2) дозволяє управляти невизначеностями такого роду. Вперше алгоритми кластеризації на основі ІНМТ2 були запропоновані Хвангом (С. Hwang) і Рхі (F.C.-H. Rhee) в 2007 р. [16].

3.1 Кластеризація з використанням FCM-алгоритму на основі інтервальних нечітких множин другого типу

Методи кластеризації на основі цільових функцій застосовуються для мінімізації відстані між об'єктом та центром кластера, параметри якого визначаються ітераційно. У практичних задачах технічного моніторингу множини даних часто містять атипові об'єкти, що зумовлює зміщення центрів кластерів і зниження точності класифікації. Для усунення цього ефекту в інформаційній технології Clustering4Build використано FCM-алгоритм на основі нечітких множин першого та другого типів, який реалізує вагову корекцію об'єктів під час обчислення центрів кластерів. Такий підхід забезпечує адаптивну кластеризацію в умовах невизначеності та інтеграцію результатів у модуль DSS для подальшої аналітики. Для усунення цього недоліку алгоритм FCM інтегровано у склад аналітичного модуля системи Clustering4Build. Модуль реалізовано у вигляді Python-сервісу, який взаємодіє з базою даних PostgreSQL через REST API. Такий підхід дозволяє автоматично

оновлювати параметри фаззифікатора m у процесі кластеризації потокових даних, що надходять від сенсорів технічного моніторингу

3.1.1 Проблема невизначеності фаззифікатора в FCM-алгоритмі

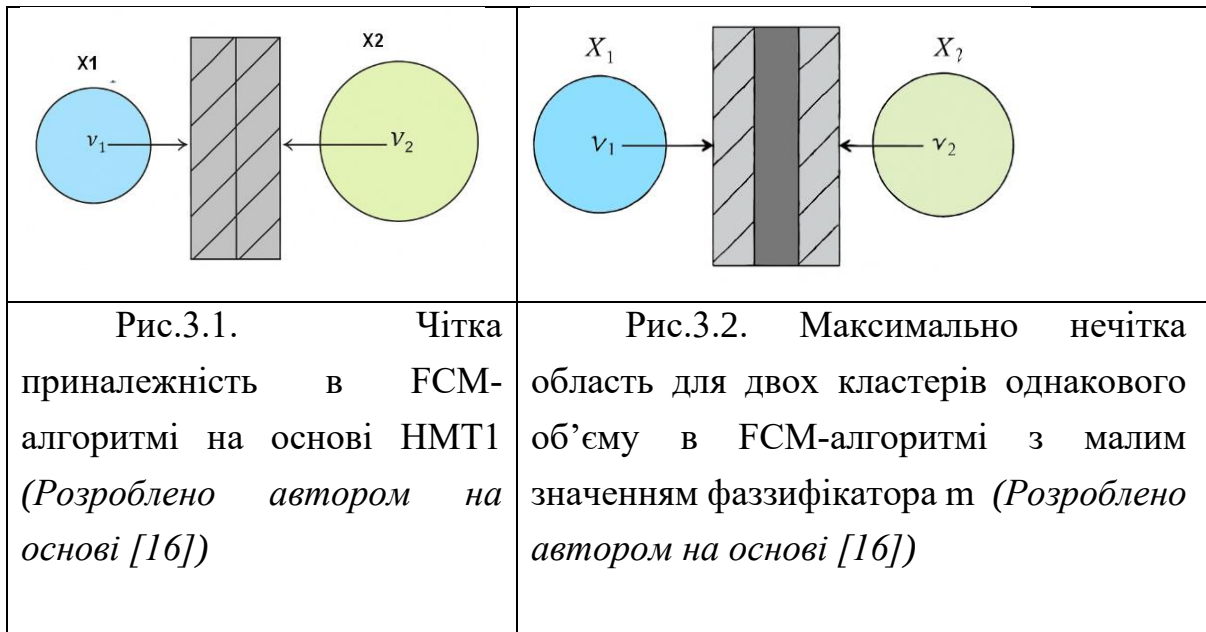
FCM-алгоритм на основі нечітких множин першого типу забезпечує адекватну кластеризацію для множин об'єктів, що мають перетинання між кластерами та неоднорідну щільність. Однак, для забезпечення високої точності класифікації в реальних системах, таких як технічний моніторинг, потрібно враховувати вплив значення фаззифікатора на результати кластеризації, особливо коли дані мають різну щільність і об'єм кластерів. Розрахунок функцій приналежності базується на відносних відстанях між об'єктами і центрами кластерів, які зазвичай визначаються за евклідовою метрикою. При цьому об'єкти, віддалені від центру, мають менший вплив на визначення його координат, ніж ті, що розташовані ближче.

У практичних задачах технічного моніторингу це призводить до проблеми нечіткості фаззифікатора — параметра, що регулює рівень розмитості кластерів і впливає на стабільність результатів. Якщо фаззифікатор обрано некоректно, з'являється зміщення центрів кластерів та спотворення результатів.

У системі Clustering4Build ця проблема вирішується шляхом адаптивного налаштування фаззифікатора в процесі кластеризації. Значення параметра m автоматично обчислюється на основі статистичних характеристик вхідних даних, що зберігаються в базі PostgreSQL. Обчислення виконується у модулі FCM Core, а результати передаються у DSS через REST API для візуалізації та подальшої оптимізації.

На рисунку 3.1 наведено приклад двох кластерів із перетином, який демонструє поведінку фаззифікатора при різній щільності об'єктів. Вертикальна лінія позначає область максимальної невизначеності, де значення функції приналежності об'єктів дорівнюють 0,5. Це визначення є важливим

для оцінки класифікації об'єктів, оскільки зона невизначеності визначає, до якого кластеру об'єкти повинні належати з найбільшим ступенем приналежності.



Вертикальна лінія на рисунку 3.1 визначає границю рішень між двома кластерами, де об'єкти, розташовані по різні боки від неї, належать різним групам. Зміна значення фаззифікатора m впливає на ширину області невизначеності: зі збільшенням m зона нечіткості розширюється (рис. 3.1-3.4.). Оптимальні результати кластеризації зазвичай досягаються при $m \approx 2$.

Для наборів даних, у яких кластери мають схожу щільність і об'єм, алгоритм FCM на основі нечітких множин першого типу забезпечує стабільні результати. Однак у реальних задачах технічного моніторингу, коли спостерігаються кластери різної щільності або різного розміру, вибір фаззифікатора істотно впливає на якість кластеризації.

У системі Clustering4Build реалізовано адаптивний механізм корекції фаззифікатора, що аналізує статистику даних у базі PostgreSQL і динамічно визначає оптимальне значення m . Такий підхід дозволяє враховувати особливості розподілу технічних параметрів у сенсорних наборах, мінімізувати спотворення кластерів і забезпечити узгодженість результатів у модулі DSS.

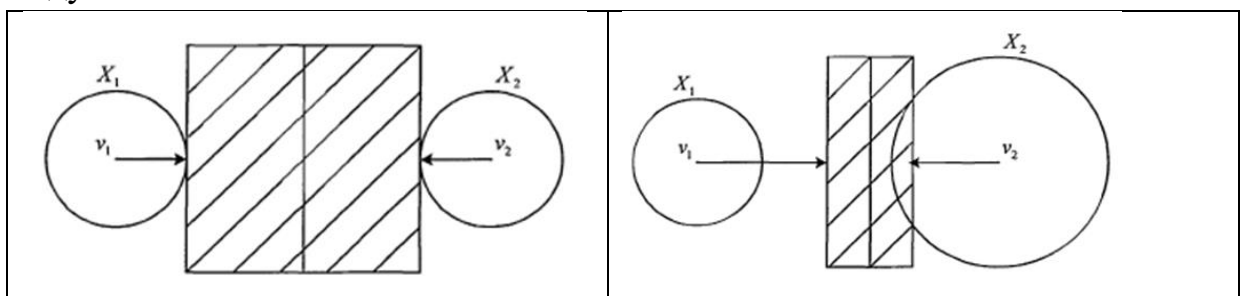


Рис.3.3. Максимально нечітка область для двох кластерів однакового об'єму в FCM-алгоритмі з великим значенням фаззифікатора m (Розроблено автором на основі [16])

Рис. 3.4. Максимально нечітка область для двох кластерів однакового об'єму в FCM-алгоритмі з малим значенням фаззифікатора m (Розроблено автором на основі [16])

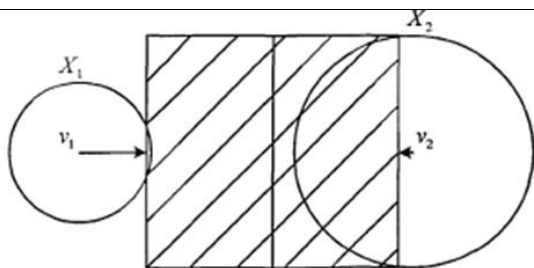


Рис. 3.5. Максимально нечітка область для двох кластерів однакового об'єму для FCM-алгоритму з великим значенням фаззифікатора m (Розроблено автором на основі [16])

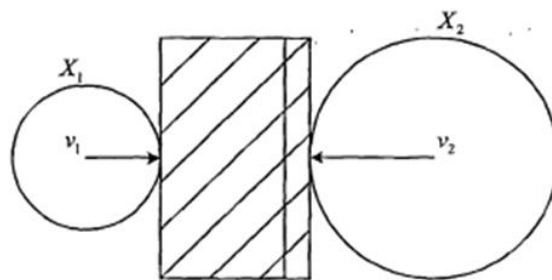


Рис. 3.6. Бажана максимальна нечітка область для двох кластерів різного об'єму в FCM-алгоритмі (Розроблено автором на основі [16])

Для поліпшення результатів кластеризації за умов різнорідних кластерів у FCM-алгоритмі необхідний апіорний контроль параметра фаззифікатора m . У системі Clustering4Build реалізовано мікросервіс FuzzificationController, який динамічно аналізує дані і оптимізує значення фаззифікатора, що дозволяє отримати найкращі результати для класифікації об'єктів з різними характеристиками. Оптимізація цього параметра визначає ширину нечіткої області, що впливає на точність розподілу об'єктів між кластерами (рис. 3.6). У звичайному алгоритмі на основі нечітких множин першого типу використовується єдине значення m , що обмежує здатність моделі коректно описувати кластери різного розміру та щільності.

Для усунення цієї обмеженості в інформаційній технології Clustering4Build застосовано інтервальні нечіткі множини другого типу (ІНМТ2), які використовують два фаззифікатори — m_1 та m_2 . Це дозволяє формувати нечітку область із асиметричними підобластями різної ширини, покращуючи стійкість і точність кластеризації.

На рівні програмної архітектури реалізовано мікросервіс FuzzificationController, який аналізує статистичні характеристики вхідних даних (розподіли, дисперсії, типи конструкцій) і визначає оптимальну пару фаззифікаторів для модуля FCM Core. Результати кластеризації інтегруються

у модуль DSS, де відображаються у вигляді теплових карт зон ризику та індикаторів ступеня зносу конструкцій. Такий підхід поєднує математичну точність IHMT2 з прикладною ефективністю інформаційної системи.

Як зазначалося раніше (див. п. 2.1), класичний FCM-алгоритм на основі нечітких множин першого типу є ітераційною процедурою, що мінімізує цільову функцію при визначенні функцій приналежності об'єктів і координат центрів кластерів.

На рис.3.7 наведені графіки ФП об'єктів, що лежать між центрами двох кластерів, одному з кластерів при різних значеннях фаззифікатора m . При $m \rightarrow 1$ нечіткі ступені приналежності є максимально чіткими, тобто об'єкти, розміщені точно зліва (справа) від максимуму нечіткої області повністю

належать (або не належать) кластеру $X_1 (X_2)$, при $m \rightarrow \infty$ нечіткі ступені приналежності є максимально нечіткими, тобто об'єкти, розміщені в центрах повністю належать (або не належать) кластеру

$X_1 (X_2)$, в іншому випадку їм призначається нечітка ступінь приналежності, що дорівнює 0,5. Якщо кластери мають різну щільність або розмір, вибір фаззифікатора суттєво впливає на результат, оскільки визначає межі нечіткої області між ними. Така поведінка пояснюється впливом параметра m на ширину максимальної нечіткої області, що потребує адаптивного контролю при його виборі.

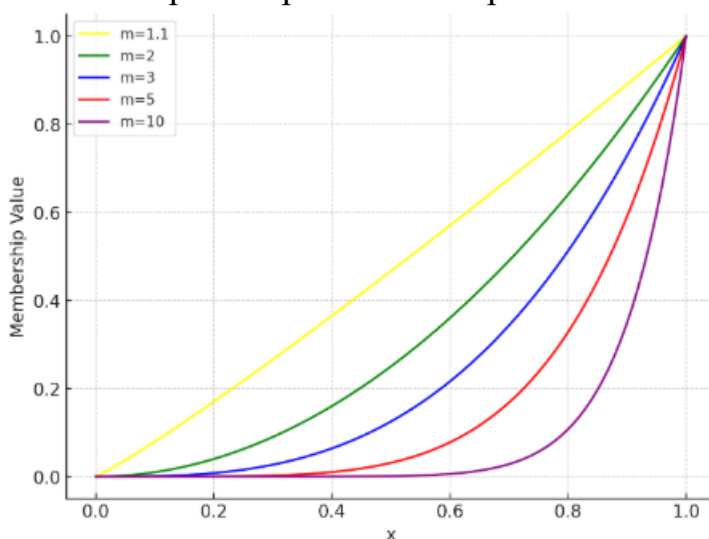


Рис. 3.7. Графіки функцій приналежності для FCM-алгоритму при різних значеннях фаззифікатора m (Розроблено автором на основі [16])

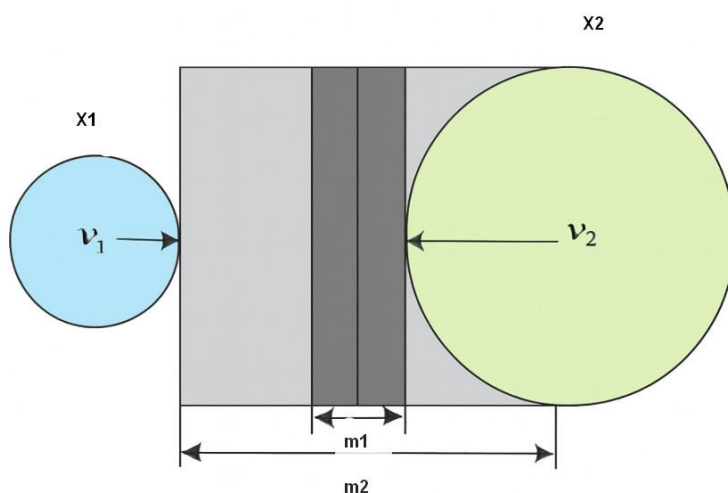


Рис. 3.8. Невизначена максимальна нечітка область побудови з використанням двох фаззифікаторів m_1 і m_2 , ($m_1 < m_2$) (Розроблено автором на основі [16])

Бажана максимальна нечітка область (рис. 3.6) забезпечує більш точну кластеризацію, проте в FCM-алгоритмі на основі нечітких множин першого типу цього неможливо досягти, оскільки фаззифікатор m має єдине значення. Для подолання цього обмеження застосовується підхід із двома фаззифікаторами (m_1 і m_2), що задають різні ступені нечіткості та формують інтервальну нечітку область (рис. 3.8). У цьому випадку функція приналежності об'єкта розглядається як інтервальна — тобто невизначена в межах максимальної нечіткої області. Використання двох фаззифікаторів потребує розширення множини об'єктів на інтервальні нечіткі множини другого типу, що дозволяє підвищити точність кластеризації в умовах різномірних даних.

3.1.2 Розширення множини об'єктів кластеризації на інтервальні нечіткі множини другого типу для FCM-алгоритму

У межах інформаційної технології Clustering4Build процес формування інтервальних нечітких множин другого типу реалізовано у модулі FuzzificationEngine, який обчислює верхні та нижні функції приналежності на основі двох фаззифікаторів та формує "відбиток невизначеності" для подальшої кластеризації у FCM-Core.

В разі представлення невизначеності за допомогою ІНМТ2 первинна функція приналежності J_{x_i} об'єкта x_i може бути визначена як

інтервальна функція приналежності з усіма вторинними ступенями первинних функцій приналежності, рівними 1. Тоді ІНМТ2 $\tilde{X}$ може бути записана як [12]:

$$\tilde{X} = \{(x, u), \mu_X(x, u) \mid \forall x \in X, \forall u \in J_x \subseteq [0, 1], \mu_X(x, u) = 1\} \quad (3.1)$$

На рисунку 3.9 наведений приклад «відбитка невизначеності» - FOU («foot print of uncertainty») для ІНМТ2.

При цьому значення функції приналежності в точці x' представлено інтервалом між «верхньою» і «нижньою» ФП $\bar{u}(x')$ і $\underline{u}(x')$. Отже, кожному x' можна поставити у відповідність інтервал для первинної ФП:

$$J_{x'} = [\underline{u}(x'), \bar{u}(x')] \quad (3.2)$$

і вертикальний зріз для x' , де вторинна ступінь для первинної ФП кожного x' дорівнює 1.

Побудова інтервальних функцій приналежності реалізується через обчислення двох наборів параметрів (верхніх і нижніх ФП), що формують інтервал невизначеності – FOU (Footprint of Uncertainty). У системі це реалізовано через дві ітераційні процедури, які мінімізують відповідні цільові функції (3.7) і (3.8). Математичне представлення процедури винесено в Додаток 3.

На рисунку 3.10 наведено приклад «відбитка невизначеності» для ІНМТ2, що визначається за формулами (3.3) і (3.4), в разі двох кластерів для комбінацій значень фазифікаторів $m_1, m_2 = (2, 5)$. Цей інтервал визначає так званий відбиток невизначеності. Саме FOU дає можливість моделі відображати внутрішню невпевненість у тому, наскільки об'єкт x належить конкретному кластеру — властивість, що критично важлива для задач технічного моніторингу. Як видно з рисунку 3.10, для зазначених вище об'єктів «відбиток невизначеності» не існує.

Управління невизначеністю фазифікатора m в FCM-алгоритмі на основі ІНМТ2 здійснюється за допомогою двох процедур [13, 16, 17]: процедури обчислення центрів кластерів; процедури дефазифікації (отримання чіткого розбиття) для кінцевого рішення про результати кластеризації. Для обчислення центрів кластерів за допомогою FCM-алгоритму на основі ІНМТ2 використовуються операція «зниження типу» («type-reduction») і операція дефазифікації ІНМТ2.

Нехай $\tilde{X}$ - інтервальна нечітка множина другого типу, універсум якого розбитий на n точок: $x_1, \dots, x_n$.

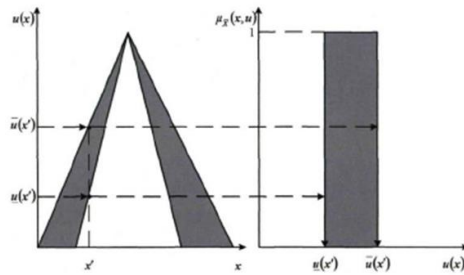


Рис. 3.9. Інтервальна нечітка множина другого типу і вертикальний зріз в точці x' (Розроблено автором на основі [12, 16])

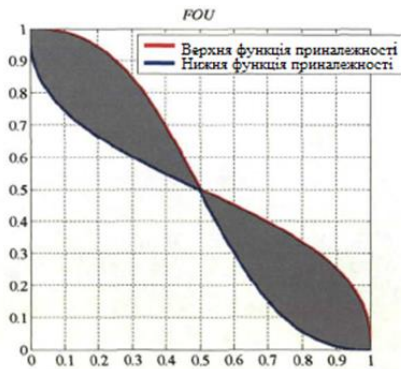


Рис. 3.10. «Відбиток невизначеності» для FCM-алгоритму для комбінацій значень фаззифікатора $m_1 = 2, m_2 = 5$ (Розроблено автором на основі [12, 16])

Центроїд інтервальної нечіткої множини другого типу $\tilde{X}$ обчислюється, як [16].

$$v_{\tilde{X}} = [v_L, v_R] = \sum_{u(x_1) \in J_{x_1}} \dots \sum_{u(x_n) \in J_{x_n}} 1 / \frac{\sum_{i=1}^n x_i \cdot u(x_i)^m}{\sum_{i=1}^n u(x_i)^m} \quad (3.9)$$

Для визначення $[v_L, v_R]$, що представляє центри кластерів, необхідно обчислити центроїди всіх «вкладених» НМТ1 для ІНМТ2, які описуються «верхньою» і «нижньою» ФП об'єкта x_i : $\bar{u}(x_i)$ і $\underline{u}(x_i)$ [13, 16, 17].

Якщо кількість об'єктів дорівнює n , то кількість «вкладених» НМТ1 для

ІНМТ2 оцінюється як: $\prod_{i=1}^n |J_{x_i}| = 2^n$, де $|*|$ - потужність множини.

Для обчислення інтервальних центрів використовується алгоритм Карника–Менделя, реалізований у Clustering4Build як підпроцес у модулі TypeReducer. У FCM на основі ІНМТ2 зниження типу здійснюється за допомогою алгоритму Карника–Менделя, який:

- впорядковує об'єкти за їх ваговими коефіцієнтами;
- обчислює лівий та правий центроїди;
- забезпечує збіжність до інтервального представлення центру.

Його застосування описане в Додатку 4. дозволяє автоматизувати процедуру “зниження типу” без необхідності повного перебору вкладених НМТ1. Таким чином, розширення множини об'єктів до інтервальних нечітких множин другого типу забезпечує стійкість результатів кластеризації в умовах інформаційної невизначеності, що є ключовим для автоматизованих систем оцінки технічного стану.

У межах мікросервісної архітектури інформаційної технології Clustering4Build розширення множини об'єктів до ІНМТ2 реалізовано через спеціалізовані програмні компоненти:

- FuzzificationEngine — обчислює $\mu_L(x)$, $\mu_U(x)$ для кожного об'єкта на основі фаззифікаторів m_1 і m_2 ;
- TypeReducer — виконує type-reduction через алгоритм Карника–Менделя;
- FCM-Core — оновлює інтервальні центри кластерів та генерує результати наступної ітерації.

Передача даних між сервісами відбувається через REST API, а інтервальні значення функцій приналежності та центрів кластерів зберігаються у PostgreSQL у форматі JSONB, що забезпечує:

- ефективний доступ до інтервальних структур;
- можливість обробки великих обсягів даних;
- інтеграцію із модулем DSS для подальшої аналітики.

Використання векторизованих обчислень (NumPy), контейнеризації (Docker) та сегментації сервісів дозволяє зменшити обчислювальну складність

та підвищити масштабованість системи при роботі з великими масивами технічних даних.

Розширення множини об'єктів до інтервальних нечітких множин другого типу забезпечує перехід від фіксованої скалярної моделі функції приналежності до адаптивної інтервальної моделі, яка краще узгоджується зі структурою реальних технічних даних. Це дозволяє FCM-алгоритму враховувати змінність характеристик даних, розширює його застосовність в інформаційних технологіях та формує основу для стійкої класифікації в системах підтримки прийняття рішень.

3.1.3 Ітераційний алгоритм Карника–Менделя для обчислення інтервальних центрів кластерів

Ітераційний алгоритм Карника–Менделя застосовується по чергово для знаходження максимуму v_R і мінімуму v_L центрів кластерів [16].

У FCM-алгоритмі на основі інтервальних нечітких множин другого типу (ІНМТ2) центри кластерів не є фіксованими скалярними величинами, а представлені у вигляді інтервалів, що відображають невизначеність, притаманну функціям приналежності $\mu(x)$. Для переходу від інтервальних значень $\mu_L(x)$, $\mu_U(x)$ до інтервального центроїда $[CL, CU]$ необхідно виконати процедуру зниження типу (type-reduction). Найбільш ефективним і стандартним методом такого перетворення є ітераційний алгоритм Карника–Менделя (Karnik–Mendel, KM).

Управління невизначеністю за допомогою ІНМТ2 (при використанні двох фазифікаторів m_1, m_2) завдяки великій мірі свободи при визначенні центрів кластерів дає для FCM-алгоритму більш адекватні (кращі) результати кластеризації, ніж звичайний FCM-алгоритм з одним фазифікатором m в разі, коли множина об'єктів містить кластери істотно різної щільності або істотно різного об'єму [13, 17]. Схема ітераційного алгоритму Карника–Менделя для пошуку максимуму v_R центра кластеру v_j (Додатку 5.)

Загальна схема FCM-алгоритму на основі ІНМТ2 подана в Додатку 6.

Обчислення центрів кластерів здійснюється в результаті «чіткого розбиття» ФП об'єктів: об'єкт відноситься до кластеру відповідно до значення

його ФП. «Чітке розбиття» в FCM-алгоритмі на основі НМТ1 знаходиться як [16]:

«Якщо $(u_j(x_i) > u_t(x_i))$ для $t = \overline{1, \dots, c}, j \neq t$, то x_i відноситься до кластеру j ».

Однак перед застосуванням правила (3.14), що реалізує операцію «чіткого розбиття», необхідно знизити тип ІНМТ2. Але операцію «зниження типу» неможливо виконати, використовуючи для знаходження $u_j(x_i)$

безпосередньо «верхні» і «нижні» ФП $\overline{u}_j(x_i)$ і $\underline{u}_j(x_i)$ [16, 17].

При виконанні операції «зниження типу» слід використовувати ліві ФП

$u_j^L(x_i)$ і праві ФП $u_j^R(x_i)$, вже обчислені при пошуку мінімального (лівого) V_L і максимального (правого) V_R центрів кластерів відповідно. Таким чином, «зниження типу» може бути виконано як [16]:

$$u_j(x_i) = \frac{u_j^R(x_i) + u_j^L(x_i)}{2}, j = \overline{1, \dots, c} \quad (3.15)$$

При цьому «чітке розбиття» виконується за правилом (3.14). В формулі (3.15) ФП $u_j^L(x_i)$ і $u_j^R(x_i)$ залежать від оцінок об'єкта за всіма q елементами моніторингу, тому ФП $u_j^L(x_i)$ і $u_j^R(x_i)$ повинні обчислюватися як:

$$u_j^R(x_i) = \frac{\sum_{l=1}^q u_{jl}(x_i)}{q}, \quad (3.16)$$

$$u_{jl}(x_i) = \begin{cases} \overline{u}_j(x_i), & \text{якщо} \\ \underline{u}_j(x_i), & \text{інакше} \end{cases} \quad x_{il} \text{ використовує } \overline{u}_j(x_i) \text{ для знаходження } v_j^R \quad (3.17)$$

$$u_j^L(x_i) = \frac{\sum_{l=1}^q u_{jl}(x_i)}{q}, \quad (3.18)$$

$$u_{jl}(x_i) = \begin{cases} \overline{u}_j(x_i), & \text{якщо} \\ \underline{u}_j(x_i), & \text{інакше} \end{cases} \quad x_{il} \text{ використовує } \overline{u}_j(x_i) \text{ для знаходження } v_j^L \quad (3.19)$$

Слід зазначити, що при використанні одного і того ж значення фаззифікатора (при $m_1 = m_2$), буде отримано звичайне НМТ1, яке не має невизначеності для своїх ФП [16]. В цьому випадку FCM-алгоритм на основі ІНМТ2 повністю збігається з FCM-алгоритмом на основі НМТ1, так як результати операцій «зниження типу» і «чіткого розбиття» будуть збігатися. Таким чином, FCM-алгоритм на основі ІНМТ2 є розширенням FCM-алгоритму на основі НМТ1.

FCM-алгоритм на основі ІНМТ2 має більшу обчислювальну складність, ніж FCM-алгоритм на основі НМТ1, що пояснюється використанням ітераційного алгоритму Карника - Менделя для виконання операції «зниження типу».

Нехай є n об'єктів з оцінками за q елементами моніторингу, та $n \geq q$. В цьому випадку для оцінки центру кластера необхідно виконати за кожним елементом моніторингу n^2 операцій, так кількість об'єктів і кількість ФП дорівнюють n , а загальна кількість операцій дорівнює $n \cdot q^2$ (за всіма q характеристикам). Однак кількість характеристик зазвичай значно менша, ніж кількість об'єктів. Тоді при $n \geq q$ складність обчислень можна оцінити як $O(n^2)$.

В FCM-алгоритмі на основі ІНМТ2 при виконанні операції «пониження типу» за допомогою ітераційного алгоритму Карника - Менделя оцінюється мінімальний V_L і максимальний V_R центри кластерів. При цьому виконуються наступні кроки. Впорядкування i -х індексів ($i = \overline{1, n}$) множини об'єктів за зростанням значень оцінок по кожному елементу моніторингу $l(l = \overline{1, q})$.

1. Оцінка центру кластера на основі нижніх і верхніх ФП об'єктів.
2. Позиціонування центру кластера в множині об'єктів.

На кроці 1 складність обчислень для типового алгоритму повного сортування оцінюється як $n \cdot (n-1)$. Так як складність упорядкування за зростанням оцінюється по кожному елементу моніторингу, то вона дорівнює $q \cdot n \cdot (n-1)$. При $n \geq q$ складність обчислень можна оцінити як $O(n^2)$. На кроці 2 складність обчислень при оцінюванні центру кластера

така ж, як і в FCM-алгоритмі на основі НМТ1 - $O(n^2)$. На кроці 3 виконується порівняння відмінностей між центром кластера і кожним об'єктом для пошуку розміщення центру кластера в множині об'єктів, що виконується в межах n разів. При $n \geq q$ складність обчислень можна оцінити як $O(n^2)$. Отже, загальна складність трьох кроків для однієї ітерації оцінюється як $O(n^2)$. Якщо число необхідних ітерацій алгоритму Карника - Менделя дорівнює r , то складність може бути оцінена як $r \cdot O(n^2)$.

Як відомо, алгоритм Карника - Менделя вимагає не більше n ітерацій [13, 17]. У цьому випадку загальна складність FCM-алгоритму на основі ІНМТ2 може бути оцінена як $n \cdot O(n^2) = O(n^3)$. Оскільки при застосуванні алгоритму Карника - Менделя до типових множин об'єктів зазвичай досить виконати 3-4 ітерації ітераційного алгоритму, то складність FCM-алгоритму на основі ІНМТ2 зазвичай значно менша, ніж $O(n^3)$, так як $r \leq n$ [16]. Таким чином, складність обчислень для FCM-алгоритму на основі ІНМТ2 в загальному випадку на порядок вища, ніж в FCM-алгоритмі на основі НМТ1. Результатом роботи КМ є інтервал $[CL, CU]$, який згодом дефазифікується до єдиного чіткого центру C .

Таким чином, у межах інформаційної технології Clustering4Build реалізовано модуль FCM-Core на основі інтервальних нечітких множин другого типу, що дозволяє адаптивно керувати невизначеністю фазифікації, динамічно оновлювати параметри кластерів і інтегрувати результати у DSS через REST API. Така реалізація поєднує математичну модель із її прикладним програмним втіленням у складі інтелектуальної ІТ-системи.

3.2 Проблема вибору показника якості кластеризації на основі інтервальних нечітких множин другого типу

Як показав аналіз, при оцінюванні якості кластеризації з використанням FCM-алгоритму на основі ІНМТ2 застосування в якості показників якості будь-яких індексів і показників, які використовують в своєму записі значення фазифікатора m (як в FCM-алгоритмі на основі НМТ1), неможливе через наявність двох фазифікаторів m_1 і m_2 , а також через те, що саме значення фазифікатора істотно впливає на значення показника якості кластеризації (при пошуку екстремуму). Також при виборі

показника якості кластеризації на основі ІНМТ2 необхідно відмовитися від індексів і показників, в записі яких присутні ФП $u_j(x_i)$ ($j = \overline{1, c}; i = \overline{1, n}$), оскільки значення ФП залежать від вибору значень фаззифікаторів m_1 і m_2 [79,80].

У контексті інформаційної технології Clustering4Build показники якості кластеризації використовуються не лише для математичної оцінки структури кластерів, але і як ключові індикатори для автоматичного вибору моделі кластеризації в модулі GA-Optimizer та для формування рекомендацій у DSS. Показник якості виконує роль критерія оптимізації параметрів m_1 , m_2 та структури кластерів, що забезпечує узгодженість між результатами кластеризації та реальною поведінкою технічних характеристик об'єкта.

У Clustering4Build вибір показника якості визначає коректність роботи модулів FCM-Core та GA-Optimizer, а також впливає на стабільність рекомендацій у DSS. Показник якості виконує роль критерію для автоматичного вибору моделі.

При оцінюванні якості кластеризації множини об'єктів доцільно оцінювати гіпероб'єми отриманих кластерів, при цьому сумарний гіпероб'єм кластерів повинен бути мінімізований, так як метою FCM- алгоритму на основі ІНМТ2 є отримання такої максимальної нечіткої області між двома будь-якими кластерами (рисунки 3.6 і 3.8), яка б найкращим чином апроксимувала розбиття об'єктів на кластери.

Обчислення гіпероб'ємів реалізовано в окремому мікросервісі QualityEvaluator, який отримує від FCM-Core центри кластерів, коваріаційні матриці та метрики розподілу і повертає DSS агреговані індикатори якості у форматі JSON. Обчислення індексів H та Sph реалізовано в окремому мікросервісі QualityEvaluator, який отримує від FCM-Core інформацію про центри кластерів, коваріаційні матриці, спектри власних значень та відстані між центрами. Мікросервіс повертає DSS агреговану оцінку якості у форматі JSON та передає GA-Optimizer значення цільової функції для подальшої оптимізації параметрів.

Обчислення виконується за такими формулами:

$$H = \sum_{j=1}^c \left(\det(R_j) \right)^{\frac{1}{2}} \rightarrow \min \quad (3.20)$$

$$R_j = \frac{1}{n_j} \cdot \sum_{i=1}^n (x_i - v_j) \cdot (x_i - v_j)^T \quad (3.21)$$

де R_j — коваріаційна матриця j -го кластеру; n_j — кількість об'єктів кластеризації, що відносяться до j -го кластеру; v_j — вектор координат центра j -го кластеру; x_i — вектор координат i -го об'єкта; $\det(R_j)$ — визначник коваріаційної матриці j -го кластеру; n — кількість об'єктів; c — кількість кластерів; $j = \overline{1, c}; i = \overline{1, n}$.

Отримані значення H використовуються як показник якості розбиття.

При оцінюванні якості кластеризації для кластерів гіперсферичних форм (або форм, близьких до гіперсферичних) можна використовувати показник якості кластеризації – індекс Sph [31]:

$$Sph = \frac{\sum_{j=1}^c r_j}{\min_{\substack{j=\overline{1,c} \\ t=\overline{1,c} \\ j \neq t}} d(v_j, v_t)} \rightarrow \min, \quad (3.22)$$

$$r_j = \max_{x_i \in X_j} d(x_i, v_j) \quad (3.23)$$

де v_j — вектор координат центра j -го кластера; x_i — вектор координат i -го об'єкта; $d(x_i, v_j)$ — відстань між об'єктом x_i , що належить кластеру X_j , і центром кластеру v_j ; $d(v_j, v_t)$ — відстань між центрами кластерів v_j і v_t ; n — кількість об'єктів; c — кількість кластерів; $j = \overline{1, c}; i = \overline{1, n}, t = \overline{1, c}$.

При використанні індексу Sph очевидним є прагнення мінімізувати радіус кожного j -го кластера ($j = \overline{1, c}$). Крім того, кластери повинні бути якомога більш відокремлені один від одного, тобто центри кластерів повинні бути максимально віддалені один від одного. QualityEvaluator додатково обчислює радіуси кластерів та відстані між центрами, після чого формує показник Sph. Цей Діаграма демонструє динамічний обмін повідомленнями між FCM-Core, QualityEvaluator, ShapeAnalyzer, GA-Optimizer та DSS.

Показано, як результати кластеризації передаються в QualityEvaluator, виконуються процедури визначення форми кластерів та обчислення показників H/Sph, а також як DSS отримує звіт для візуалізації, а GA-Optimizer — функцію придатності для оптимізації параметрів моделі.

Слід зазначити, що застосування загального гіпероб'єму H в якості показника якості кластеризації дає адекватні результати кластеризації як для кластерів гіпереліпсоїдної форми, так і в окремому випадку - для кластерів гіперсферичних форм, що забезпечується використанням при розрахунках коваріаційних матриць кластерів, що описують форму і орієнтацію кластерів. Однак індекс Sph характеризується меншою обчислювальною складністю, ніж загальний гіпероб'єм H , так як не потребує розрахунку визначників коваріаційних матриць. Вибір між H та Sph здійснюється на основі аналізу спектра власних значень коваріаційних матриць кластерів. У застосуваннях технічного моніторингу будівельних конструкцій показники H та Sph дають змогу оцінити компактність та взаємне розташування підкластерів ознак деформацій, напружень або температурних зон. Це забезпечує DSS можливість формувати ранні попередження про аномалії та прогнозувати зміну технічного стану на основі кластерної структури даних.

3.3 Генетичний алгоритм пошуку оптимальної комбінації значень фаззифікаторів для FCM-алгоритму на основі інтервальних нечітких множин другого типу

Застосування методу кластеризації, що реалізує використання FCM-алгоритму на основі ІНМТ2 і ГА, забезпечує отримання адекватних результатів кластеризації при значному скороченні часу пошуку оптимальної комбінації значень фаззифікаторів m_1 і m_2 . Слід зазначити, що звичайний перебір всіляких найбільш часто використовуваних значень фаззифікаторів m_1 і m_2 з множини $\{1, 1; 2, 3, 5, 10\}$ часто не приносить бажаного результату кластеризації.

При вирішенні задачі пошуку оптимальної комбінації значень фаззифікаторів m_1 і m_2 для FCM-алгоритму на основі ІНМТ2 хромосома може бути задана у вигляді [117, 118]:

$$s = (m_1, m_2), \quad (3.24)$$

де $m_1, m_2 \in [1, m_{max}]$; m_{max} - дійсне число, що визначає максимальне значення фазифікатора (в залежності від вирішуваної задачі; в більшості практичних задач: $m_{max} < 50$; $m_1 < m_2$).

При реалізації ГА використовується одноточкове схрещування, а мутації піддається не більше одного гена, так як довжина хромосоми дорівнює 2. В якості опції відповідності для ГА може бути використаний загальний гіпероб'єм N за формулою (3.20) для безлічі множини об'єктів, що містить кластери гіпереліпсоїдної форми або, в окремому випадку, індекс Sph за формулою (3.22) для множини об'єктів, що містить кластери гіперсферичної форми. Нехай надалі в якості опції відповідності використовується загальний гіпероб'єм N за формулою (3.20).



Рис.3.11. Алгоритм генетичної оптимізації m_1, m_2 (Розроблено автором на основі [117, 118])

У Clustering4Build значення загального гіпероб'єму H та індексу Sph обчислюються мікросервісом QualityEvaluator і передаються в GA-Optimizer як значення функції придатності. Таким чином, модуль генетичної оптимізації безпосередньо інтегрований у контур кластеризації та оцінювання якості для задач технічного моніторингу.

1. Випадковим чином створюється популяція розміром P . При цьому виконується перевірка умови: $m_1 < m_2$.

2. При $g < G$ (G і g — максимальна і поточна кількість поколінь ГА відповідно) реалізується FCM-алгоритм на основі ІНМТ2 з обчисленням значення функції відповідності за формулою (3.20) для кожної хромосоми і створюється $R_c \cdot P/2$ пар хромосом-батьків.

3. Виконуються операції схрещування і мутації для поточної популяції. При цьому виконується перевірка умови: $m_1 < m_2$. Для хромосом-нащадків реалізується FCM-алгоритм на основі ІНМТ2 і обчислюються значення функції відповідності за формулою (3.20).

4. Створюється нова популяція розміром $(P + R_c \cdot P)$, доповнена хромосомами-нащадками в кількості $R_c \cdot P$, потім $R_c \cdot P$ хромосоми з гіршими значеннями функції відповідності (3.20) відкидаються. Якщо $g < G$, відбувається перехід до кроку 2.

5. Вибирається краща хромосома, яка мінімізує функцію відповідності (3.20). Для кожного об'єкта визначається його приналежність до нечітких кластерів.

Одночасно з популяцією хромосом вигляду (3.24) створюються «популяції» значень функції відповідності, координат центрів кластерів і ступенів приналежності об'єктів центрам кластерів.

Описаний вище метод кластеризації, який реалізує використання FCM-алгоритму на основі ІНМТ2 і ГА, дозволяє врахувати властивість кластерної відносності об'єктів і отримати адекватні результати кластеризації для множини об'єктів, що містить кластери істотно різної щільності або істотно різного об'єму. Розглянуті в підрозділах 3.1–3.3 методи кластеризації на основі FCM-алгоритму та ІНМТ2 ґрунтуються на функції приналежності, для якої обов'язковою є умова нормування — сума ступенів приналежності об'єкта до всіх кластерів дорівнює одиниці (обмеження виду (2.1)). Ця умова забезпечує коректність відносного порівняння кластерів між собою (кластерну

відносність), однак робить результат кластеризації чутливим до шумових об'єктів і викидів: оскільки сума приналежностей фіксована, наявність нетипового об'єкта в наборі даних штучно впливає на розподіл ступенів приналежності решти об'єктів. Це визначає доцільність розгляду альтернативного підходу — можливісної кластеризації (PCM) на основі ІНМТ2, вільної від умови нормування, який розглядається в підрозділах 3.4–3.8.

3.4 Кластеризація з використанням FCM-алгоритму на основі інтервальних нечітких множин другого типу

Застосування FCM-алгоритму на основі ІНМТ2 не завжди забезпечує отримання адекватних результатів кластеризації через наявність обмеження виду (2.1).

Використання модифікації FCM-алгоритму на основі ІНМТ2 - PCM-алгоритму на основі ІНМТ2 - дозволяє відмовитися від вимоги виконання цієї умови, а, отже, послабити властивість кластерної відносності і врахувати властивість кластерної типовості при обчисленні координат центрів кластерів [16].

Проблема невизначеності фаззифікатора і «ширини зони» в PCM-алгоритмі. Як і FCM-алгоритм на основі НМТ1, PCM-алгоритм на основі НМТ1 є ітераційним алгоритмом, який реалізує обчислення функцій типовості об'єктів центрам кластерів. В PCM-алгоритмі на основі НМТ1 функції типовості оцінюють «сумісність об'єкта x_i з центром кластера v_j » (можливість того, що об'єкт належить кластеру), яка залежить від абсолютної відстані між центром кластера $v_j (j = \overline{1, c})_i$ об'єктом $x_i (i = \overline{1, n})$, а також від вибору фаззифікатора ті «ширини зони» η_j

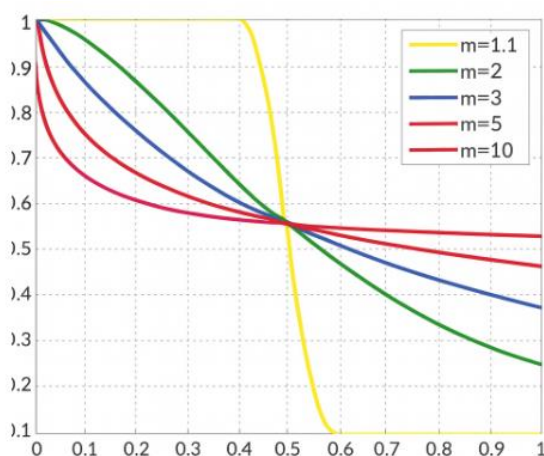


Рис.3.12. Графік функцій типовості для РСМ-алгоритма на основі НМТ1 при різних значеннях (Розроблено автором на основі [150])

Фаззифікатора m і «ширини зони» $\eta_1 = \eta_2 = 0,5$

На рисунку 3.18 представлені графіки ФТ для випадку двох кластерів при різних комбінаціях значень фаззифікатора m і «ширини зон»

$\eta_1 = \eta_2 = 0,5$.

Очевидно, що без правильного вибору значень фаззифікатора m і «ширини зони» $\eta_j (j = \overline{1, c})$ не можливо отримати адекватні результати кластеризації. При цьому, як і при використанні FCM-алгоритму на основі ПМТ1, використання РСМ-алгоритму на основі НМТ1 не завжди є ефективним у разі, коли кластери мають істотно різну щільність або істотно різний об'єм. У контексті інформаційної технології Clustering4Build РСМ-алгоритм на основі ІНМТ2 реалізується як окремий модуль РСМ-Core, який доповнює FCM-Core і використовується для кластеризації об'єктів із вираженою кластерною типовістю.



Рис. 3.13. PCM-Core у Clustering4Build (Розроблено автором)

Модуль PCM-Core інтегрується у мікросервісну архітектуру, взаємодіє з FuzzificationEngine та DSS через REST API і забезпечує альтернативну модель кластеризації для задач технічного моніторингу.

3.5. Розширення множини об'єктів кластеризації на інтервальні нечіткі множини другого типу для PCM-алгоритму

У Clustering4Build розширення множини об'єктів кластеризації до IHMT2 для PCM-алгоритму забезпечується повторним використанням сервісів FuzzificationEngine та TypeReducer. Інтервальні функції типовості зберігаються у сховищі даних у форматі JSONB, що дає змогу виконувати подальшу аналітику та передавати параметри кластерів у DSS.

Невизначеність в РСМ-алгоритмі на основі ІНМТ2 може бути виражена як за допомогою двох різних значень фаззифікатора m (як в FCM-алгоритмі на основі ІНМТ2) і фіксованого значення «ширини зони» $\eta_j (j = \overline{1, c})$ для кожного кластера, так і за допомогою двох різних значень для «ширини зони» $\eta_j (j = \overline{1, c})$ для кожного кластера при фіксованому значенні фаззифікатора m [152]. При представленні невизначеності в РСМ-алгоритмі на основі ІНМТ2 за допомогою двох різних значень фаззифікатора m : m_1 і m_2 (як в FCM-алгоритмі на основі ІНМТ2) і фіксованого значення «ширини зони» $\eta_j (j = \overline{1, c})$ для кожного кластера «нижня» і «верхня» інтервальні ФТ об'єкта x_i можуть бути записані таким чином:

$$\overline{w}_j(x_i) = \begin{cases} \frac{1}{1 + \left(\frac{d_{ji}}{\eta_j}\right)^{\frac{2}{m_1-1}}}, \text{ якщо } \frac{1}{1 + \left(\frac{d_{ji}}{\eta_j}\right)^{\frac{2}{m_1-1}}} > \frac{1}{1 + \left(\frac{d_{ji}}{\eta_j}\right)^{\frac{2}{m_2-1}}} \\ \frac{1}{1 + \left(\frac{d_{ji}}{\eta_j}\right)^{\frac{2}{m_2-1}}}, \text{ якщо } \frac{1}{1 + \left(\frac{d_{ji}}{\eta_j}\right)^{\frac{2}{m_1-1}}} \leq \frac{1}{1 + \left(\frac{d_{ji}}{\eta_j}\right)^{\frac{2}{m_2-1}}} \end{cases} \quad (3.25)$$

$$\underline{w}_j(x_i) = \begin{cases} \frac{1}{1 + \left(\frac{d_{ji}}{\eta_j}\right)^{\frac{2}{m_1-1}}}, \text{ якщо } \frac{1}{1 + \left(\frac{d_{ji}}{\eta_j}\right)^{\frac{2}{m_1-1}}} \leq \frac{1}{1 + \left(\frac{d_{ji}}{\eta_j}\right)^{\frac{2}{m_2-1}}} \\ \frac{1}{1 + \left(\frac{d_{ji}}{\eta_j}\right)^{\frac{2}{m_2-1}}}, \text{ якщо } \frac{1}{1 + \left(\frac{d_{ji}}{\eta_j}\right)^{\frac{2}{m_1-1}}} > \frac{1}{1 + \left(\frac{d_{ji}}{\eta_j}\right)^{\frac{2}{m_2-1}}} \end{cases} \quad (3.26)$$

На рисунку 3.17.а. наведений приклад «відбитка невизначеності» для випадку двох кластерів на основі формул (3.25) і (3.26) для комбінації $\eta_j=0,5$, $m_1=2$, $m_2=5$.

При представленні невизначеності в РСМ-алгоритмі на основі ІНМТ2 за допомогою двох різних значень «ширини зони» η_j : η_1 і η_2 ($j = \overline{1, c}$) для

кожного кластера (η_j) при фіксованому значенні фаззифікатора m «нижня» і «верхня» інтервальні ФТ об'єкта x_i можуть бути записані у вигляді [16]:

$$\overline{w_j}(x_i) = \begin{cases} \frac{1}{1 + \left(\frac{d_{ji}}{\eta_{j_1}}\right)^{\frac{2}{m-1}}}, \text{ якщо } \frac{1}{1 + \left(\frac{d_{ji}}{\eta_{j_1}}\right)^{\frac{2}{m-1}}} > \frac{1}{1 + \left(\frac{d_{ji}}{\eta_{j_2}}\right)^{\frac{2}{m-1}}} \\ \frac{1}{1 + \left(\frac{d_{ji}}{\eta_{j_2}}\right)^{\frac{2}{m-1}}}, \text{ якщо } \frac{1}{1 + \left(\frac{d_{ji}}{\eta_{j_1}}\right)^{\frac{2}{m-1}}} \leq \frac{1}{1 + \left(\frac{d_{ji}}{\eta_{j_2}}\right)^{\frac{2}{m-1}}} \end{cases} \quad (3.27)$$

$$\underline{w_j}(x_i) = \begin{cases} \frac{1}{1 + \left(\frac{d_{ji}}{\eta_{j_2}}\right)^{\frac{2}{m-1}}}, \text{ якщо } \frac{1}{1 + \left(\frac{d_{ji}}{\eta_{j_1}}\right)^{\frac{2}{m-1}}} \leq \frac{1}{1 + \left(\frac{d_{ji}}{\eta_{j_2}}\right)^{\frac{2}{m-1}}} \\ \frac{1}{1 + \left(\frac{d_{ji}}{\eta_{j_1}}\right)^{\frac{2}{m-1}}}, \text{ якщо } \frac{1}{1 + \left(\frac{d_{ji}}{\eta_{j_1}}\right)^{\frac{2}{m-1}}} > \frac{1}{1 + \left(\frac{d_{ji}}{\eta_{j_2}}\right)^{\frac{2}{m-1}}} \end{cases} \quad (3.28)$$

У формулах (3.27) і (3.28) η_{j1} і η_{j2} - дві «ширини зони», що визначають різні відстані, на яких можливий ступінь приналежності дорівнює 0,5.

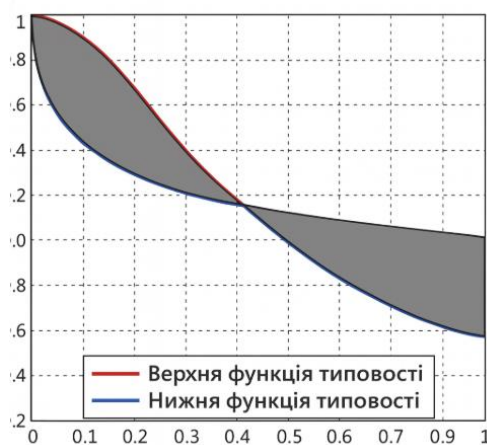


Рис. 3.14.а. «Відбиток невизначеності» для РСМ-алгоритму на основі ІНМТ2 (Розроблено автором на основі [16, 150])

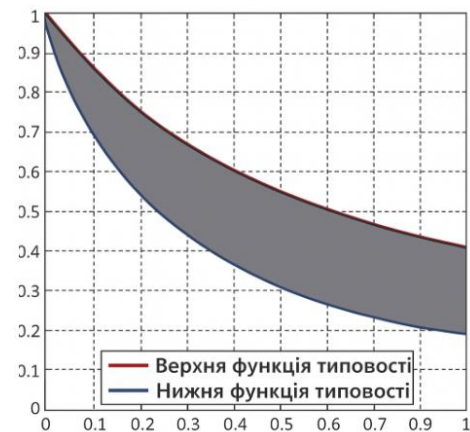


Рис. 3.14.б. «Відбиток невизначеності» для РСМ-алгоритма на основі ІНМТ2 (Розроблено автором на основі [16, 150])

На рисунку 3.14.б наведений приклад «відбитка невизначеності» для випадку двох кластерів на основі формул (3.27) і (3.28) для комбінації $m=3$, $\eta_{j1}=0,25$ і $\eta_{j2}=0,75$ комбінацій $\eta_j=0,5$, $m_1=2$, $m_1=5$.

Використання різних значень для фаззифікатора m визначає різні цільові функції в РСМ-алгоритмі на основі ІНМТ2 при $m=m_1$ і $m=m_2$ [16]:

$$J_{m_1}(W, V) = \sum_{j=1}^c \sum_{i=1}^n (w_j(x_i))^{m_1} \cdot d_{ji}^2 + \sum_{j=1}^c \eta_j^2 \cdot \sum_{i=1}^n (1 - w_j(x_i))^{m_1} \quad (3.29)$$

$$J_{m_2}(W, V) = \sum_{j=1}^c \sum_{i=1}^n (w_j(x_i))^{m_2} \cdot d_{ji}^2 + \sum_{j=1}^c \eta_j^2 \cdot \sum_{i=1}^n (1 - w_j(x_i))^{m_2} \quad (3.30)$$

Використання різних значень «ширини зони» визначає різні цільові функції в РСМ-алгоритмі на основі ІНМТ2 при

$$\eta_j (j = \overline{1, c})$$

$$\eta_j = \eta_{j1} \quad \eta_j = \eta_{j2}$$

$$J_{\eta_1}(W, V) = \sum_{j=1}^c \sum_{i=1}^n (w_j(x_i))^{\eta_1} \cdot d_{ji}^2 + \sum_{j=1}^c \eta_j^2 \cdot \sum_{i=1}^n (1 - w_j(x_i))^{\eta_1} \quad (3.31)$$

$$J_{\eta_2}(W, V) = \sum_{j=1}^c \sum_{i=1}^n (w_j(x_i))^{\eta_2} \cdot d_{ji}^2 + \sum_{j=1}^c \eta_j^2 \cdot \sum_{i=1}^n (1 - w_j(x_i))^{\eta_2} \quad (3.32)$$

Обчислення ФТ $w_j(x_i)$ і координат центрів кластерів реалізується з використанням ітераційного алгоритму Карника - Менделя (п. 3.1.3).

Обчислення ФТ і координат центрів кластерів реалізується з використанням ітераційного алгоритму Карника-Менделя (п. 3.1.3). Пошук оптимальних параметрів РСМ-алгоритму на основі ІНМТ2 виконується за допомогою генетичного алгоритму [117, 118].

Значення функції придатності формується на основі індексів якості, обчислених сервісом QualityEvaluator, що забезпечує узгодженість між якістю кластеризації та вимогами системи технічного моніторингу.

3.6 Генетичний алгоритм пошуку оптимальної комбінації значень фаззифікаторів, що реалізують управління невизначеністю, і значень «ширини зони» для РСМ-алгоритму на основі інтервальних нечітких множин другого типу

Метод кластеризації з використанням РСМ-алгоритму на основі ІНМТ2 з невизначеними значеннями фаззифікаторів і ГА дозволяє значно скоротити час пошуку оптимальної комбінації значень параметрів алгоритму кластеризації і забезпечити отримання адекватних результатів кластеризації [117, 118]. При цьому кластеризація об'єктів виконується з урахуванням

властивості кластерної типовості. Це забезпечує автоматизований пошук параметрів алгоритму кластеризації без участі користувача.

Нехай в РСМ-алгоритмі на основі ІНМТ2 для кожного кластера задається єдине значення «ширини зони» $\eta_j (j = \overline{1, c})$, а для фаззифікатора m визначається комбінація значень: m_1 і m_2 . Для пошуку оптимальної комбінації значень фаззифікаторів m_1 , m_2 , і оптимальних значень «ширини зони» $\eta_j (j = \overline{1, c})$ може бути використаний ГА з хромосомою виду:

$$s = (m_1, m_2, \eta_1, \dots, \eta_c), \quad (3.33)$$

де $m_1, m_2 \in (1, m_{\max}]$; $m_{\max}$ — деяке дійсне число, що визначає максимальне значення фаззифікатора; $m_1 < m_2$; η_j — «ширина зони» j -го кластера ($j = \overline{1, c}$); $\eta_j \in [\eta_j^{\min}, \eta_j^{\max}]$; $\eta_j^{\min} > 0$; $\eta_j^{\max} > 0$; $\eta_j^{\min} < \eta_j^{\max}$; $\eta_j^{\max} \leq \eta_{\max}$; $\eta_j^{\min} < \eta_j^{\max}$; $\eta_{\max}$ — деяке дійсне число, що визначає максимальне значення «ширини зони».

В якості функції відповідності для ГА в загальному випадку — для множини об'єктів, що містить кластери гіпереліпсоїдної форми — може використовуватися загальний гіпероб'єм H за формулою (3.20), а в окремому випадку — для множини об'єктів, що містить кластери гіперсферичної форми, — індекс Sph за формулою (3.22). Надалі під функцією відповідності буде розумітися загальний гіпероб'єм H за формулою (3.20), якщо не обумовлюється інше. Значення загального гіпероб'єму H та індексу Sph в інформаційній технології Clustering4Build обчислюються мікросервісом QualityEvaluator і передаються в GA-Optimizer як значення функції придатності. Таким чином, оцінка якості кластеризації виконується централізовано і є узгодженою для всіх алгоритмів кластеризації, що застосовуються в системі.

Тоді генетичний алгоритм має вигляд.

1. Випадковим чином створюється популяція розміром P . При цьому для першого і другого генів виконується перевірка умови: $m_1 < m_2$.

2. При $g < G$ (G і g — максимальна і поточна кількість поколінь ГА відповідно) реалізується РСМ-алгоритм на основі ІНМТ2 з обчисленням значення функції відповідності за формулою (3.20) для кожної хромосоми і створюється $R_c \cdot P / 2$ пар хромосом-батьків.
3. Виконуються операції схрещування і мутації для поточної популяції. При цьому для першого і другого генів виконується перевірка умови:
 $m_1 < m_2$ Для хромосом-нащадків реалізується РСМ-алгоритм на основі ІНМТ2 і обчислюються значення функції відповідності за формулою (3.20).
4. Створюється нова популяція розміром $(P + R_c \cdot P)$, доповнена хромосомами-нащадками в кількості $R_c \cdot P$, потім $R_c \cdot P$ хромосоми з гіршими значеннями функції відповідності за формулою (3.20) відкидаються. Якщо $g < G$, відбувається перехід до кроку 2.

Обирається краща хромосома, яка мінімізує функцію відповідності за формулою (3.20). Для кожного об'єкта визначається його приналежність до кластерів.

Одночасно з популяцією хромосом виду (3.33) існують «популяції» значень функції відповідності, координат центрів кластерів і ступенів приналежності об'єктів центрам кластерів.

Даний метод кластеризації і відповідний йому ГА можуть бути модифіковані за рахунок фіксації значень фазифікаторів m_1 і m_2 . Наприклад, в якості значень фазифікаторів m_1 і m_2 , що реалізують управління невизначеністю, для РСМ-алгоритму на основі ІНМТ2 можуть бути використані значення фазифікаторів m_1 і m_2 , отримані за допомогою FCM-алгоритму на основі ІНМТ2. Тоді при ініціалізації РСМ-алгоритму на основі ІНМТ2 слід використовувати координати центрів кластерів, обчислені за допомогою FCM-алгоритму на основі ІНМТ2. У Clustering4Build така модифікація реалізується як режим повторного використання параметрів, отриманих для FCM-алгоритму на основі ІНМТ2. Найкращі значення фазифікаторів зберігаються в ConfigStore і можуть застосовуватися при ініціалізації РСМ-алгоритму, що скорочує час оптимізації та забезпечує узгодженість між різними режимами кластеризації.

При фіксованих значеннях фазифікаторів m_1 і m_2 хромосома може бути представлена у вигляді:

$$s = (\eta_1, \dots, \eta_c) \quad (3.34)$$

де η_j - «ширина зони» j -го кластера ($j = \overline{1, c}$); $\eta_j \in [\eta_j^{\min}, \eta_j^{\max}]$; $\eta_j^{\min} > 0$; $\eta_j^{\max} > 0$; $\eta_j^{\min} < \eta_{\max}$; $\eta_j^{\max} \leq \eta_{\max}$; $\eta_j^{\min} < \eta_j^{\max}$; $\eta_{\max}$ - деяке дійсне число, що визначає максимальне значення «ширини зони».

При цьому в ряді випадків вдається поліпшити результати кластеризації за рахунок обліку властивості кластерної типовості, а розмірність оптимізаційної задачі зменшується на 2, так як довжина хромосоми виду (3.34) дорівнює $2c$.

Модифікований ГА реалізується так само, як і наведений вище ГА за винятком тих кроків, де для генів, що визначають значення фаззифікаторів m_1 і m_2 , необхідно виконувати перевірку умови: $m_1 < m_2$, оскільки значення фаззифікаторів фіксовані.

3.7. Генетичний алгоритм пошуку оптимальної комбінації значення фаззифікатора і значень «ширини зони», що реалізують управління невизначеністю, для РСМ-алгоритму на основі інтервальних нечітких множин другого типу

Метод кластеризації з використанням РСМ-алгоритму на основі ІНМТ2 з невизначеністю значень «ширини зони» і ГА дозволяє значно скоротити час пошуку оптимальної комбінації значень параметрів алгоритму кластеризації і забезпечити отримання адекватних результатів кластеризації [117, 118]. При цьому, як і в п. 3.6, кластеризація об'єктів виконується з урахуванням властивості кластерної типовості. У Clustering4Build цей метод реалізує автоматизований підбір параметрів m і ширин зон.

Нехай в РСМ-алгоритмі на основі ІНМТ2 значення фаззифікатора m є єдиним, а «ширина зони» визначається комбінацією значень: η_{j_1} і

η_{j_2} $j = \overline{1, c}$. Для пошуку оптимальної комбінації значення фаззифікатора m і значень «ширини зони» η_{j_1} і η_{j_2} $j = \overline{1, c}$ в РСМ- алгоритмі на основі ІНМТ2 слід використовувати ГА з хромосомою виду:

$$s = (m, \eta_{11}, \eta_{12}, \dots, \eta_{c1}, \eta_{c2}), \quad (3.35)$$

де m - фаззифікатор; η_{jh} - «ширина h -ої зони» ($h = \overline{1,2}$) j -го кластера $j = \overline{1,c}$; $\eta_{j1} < \eta_{j2}$.

У Clustering4Build хромосоми такого виду зберігаються у JSON-конфігураціях ConfigStore, що дозволяє відтворювати проходження алгоритму та аналізувати ефективність параметрів.

Діапазон зміни значення фаззифікатора m може бути визначений у вигляді: $(1, m_{\max}]$, де $m_{\max}$ - дійсне число, що визначає максимальне значення фаззифікатора. Діапазон зміни значень «ширини зони j -го кластера $j = \overline{1,c}$ може бути заданий довільно у вигляді інтервалу дійсних чисел $[\eta_j^{\min}, \eta_j^{\max}]$, де $\eta_j^{\min} > 0$; $\eta_j^{\max} > 0$; $\eta_j^{\min} < \eta_{\max}$; $\eta_j^{\max} \leq \eta_{\max}$; $\eta_j^{\min} < \eta_j^{\max}$; $\eta_{\max}$ - максимальне значення «ширини зони». При наявності попередньої інформації про природу кластерів діапазони зміни значень «ширини зони», $[\eta_j^{\min}, \eta_j^{\max}]$ $j = \overline{1,c}$ можуть бути задані апріорі. Використання при завданні діапазонів зміни значень «ширини зони» $[\eta_j^{\min}, \eta_j^{\max}]$ $j = \overline{1,c}$ формул типу (2.15) або (2.16) є недоцільним внаслідок зазвичай великих значень фаззифікатора m , а також неможливості визначення за допомогою цих двох значень «ширини зони».

В якості функції відповідності для ГА в загальному випадку - для множини об'єктів, що містить кластери гіпереліпсоїдної форми - може використовуватися загальний гіпероб'єм H за формулою (3.20), а в окремому випадку - для множини об'єктів, що містить кластери гіперсферичної форми, - індекс Sph за формулою (3.22). Надалі під функцією відповідності буде розумітися загальний гіпероб'єм H за формулою (3.20), якщо не обумовлюється інше. Значення індексів H та Sph обчислюються мікросервісом QualityEvaluator і повертаються PCM-GA-Optimizer у режимі функції придатності, що забезпечує узгодженість всієї системи кластеризації.

Тоді генетичний алгоритм має вигляд [117, 118].

1. Випадковим чином створюється популяція розміром P . При цьому для всіх генів, окрім першого, виконується перевірка умови: $\eta_{j1} < \eta_{j2}$ ($j = \overline{1,c}$)
2. При $g < G$ (G і g — максимальна і поточна кількість поколінь ГА відповідно) реалізується РСМ-алгоритм на основі ІНМТ2 з обчисленням

значення функції відповідності за формулою (3.20) для кожної хромосоми і створюється $RcP/2$ пар хромосом-батьків.

3. Виконуються операції схрещування і мутації для поточної популяції. При цьому для всіх генів, окрім першого, виконується перевірка умови: $\eta_{j1} < \eta_{j2}$, ($j=1, c$)

4. Для хромосом-нащадків реалізується РСМ-алгоритм на основі ІНМТ2 і обчислюється значення функції відповідності за формулою (3.20).

5. Створюється нова популяція хромосом $P+RcP$ (), доповнена хромосомами-нащадками в кількості RcP , потім RcP хромосоми з гіршими значеннями функції відповідності за формулою (3.20) відкидаються. Якщо $g < G$, відбувається перехід до кроку 2.

6. Вибирається краща хромосома, яка мінімізує функцію відповідності за формулою (3.20). Для кожного об'єкта визначається його приналежність до кластерів.

Одночасно з популяцією хромосом виду (3.35) існують «популяції» значень функції відповідності, координат центрів кластерів і ступенів приналежності об'єктів центрам кластерів.

Даний метод кластеризації і відповідний йому ГА можуть бути модифіковані за рахунок фіксації значення фазифікатора m . Діаграма відображає послідовність етапів генетичного алгоритму, реалізованого для підбору параметрів РСМ-алгоритму в середовищі інформаційної технології Clustering4Build. Послідовно показано ініціалізацію популяції, оцінювання хромосом за функцією відповідності, виконання селекції, схрещування та мутації, а також відбір оптимальної хромосоми. Діаграма демонструє ітераційний характер процесу та його узгодження з обчислювальним циклом РСМ-модуля. У Clustering4Build передбачено режим фіксації параметра m , що дає можливість скоротити розмірність оптимізації та прискорити збіжність алгоритму на великих дата-сетах.

Тоді при фіксованому значенні фазифікатора m хромосома може бути представлена у вигляді:

$$s = (\eta_{11}, \eta_{12}, \dots, \eta_{c1}, \eta_{c2}) \quad (3.36)$$

де η_{jh} - «ширина h -ої зони» ($h = \overline{1, 2}$) j -го кластера ($j = \overline{1, c}$); $\eta_{j1} < \eta_{j2}$.

При цьому в ряді випадків вдається поліпшити результати кластеризації за рахунок обліку властивості кластерної типовості, а розмірність оптимізаційної задачі зменшується на 1, так як довжина хромосоми виду (3.36) дорівнює $2c$.

Модифікований ГА реалізується так само, як і наведений вище ГА, але значення фазифікатора m вважається фіксованим.

Автоматизований вибір оптимальної моделі кластеризації (FCM2 або PCM2) здійснюється за індексами якості H та Sph , обчисленими за формулами (3.20)–(3.23).

Алгоритм кластеризації, що має мінімальне значення обраного індексу якості, автоматично позначається як пріоритетний та передається до підсистеми DSS для подальшого аналізу й відображення користувачеві.

У системі Clustering4Build результати обчислень обох алгоритмів зберігаються у DataRepository у структурованому вигляді, що забезпечує їх подальше використання в аналітичних модулях. Підсистема DSS отримує комплексний звіт про оцінку якості та порівняння моделей, що містить значення індексів H і Sph , різницю між алгоритмами, рекомендацію щодо вибору моделі та візуалізацію структури кластерів. Такий підхід забезпечує прозорість процесу вибору моделі, відтворюваність результатів та інтеграцію підсистеми прийняття рішень у загальну архітектуру інформаційної технології.

Порівняльна характеристика розглянутих підходів. Розглянуті в підрозділах 3.6 і 3.7 підходи до формалізації невизначеності PCM-алгоритму на основі IHMT2 відрізняються тим, який саме параметр моделі приймається як джерело невизначеності — фазифікатор m чи «ширина зони» η_j .

Підхід із двома фазифікаторами m_1 , m_2 (підрозділ 3.6) доцільно застосовувати, коли невизначеність пов'язана насамперед зі ступенем «розмитості» переходу між кластерами — тобто коли невідомо, наскільки різко чи плавно має змінюватися ступінь типовості об'єкта при віддаленні від центру кластера. Цей підхід також забезпечує методологічну єдність із FCM-алгоритмом на основі IHMT2 (підрозділи 3.1–3.3), де подвійний фазифікатор є єдиним доступним засобом формалізації невизначеності.

Підхід із двома значеннями «ширини зони» η_{j1}, η_{j2} (підрозділ 3.7) доцільно застосовувати, коли невизначеність пов'язана не зі ступенем розмитості, а з невизначеністю самого просторового «радіуса» кластера —

тобто коли невідомо, наскільки широкою є зона типовості кожного окремого кластера. Цей параметр є специфічним саме для РСМ-алгоритму і не має аналога у FCM, оскільки функція типовості (на відміну від функції приналежності) не нормується і залежить від «ширини зони» безпосередньо.

У межах кожного з двох підходів фіксація одного з параметрів (m та m — у підрозділі 3.7, або, аналогічно, окремого фазифікатора — у 3.6) є доцільною тоді, коли попередній аналіз даних показує низьку чутливість якості кластеризації до цього параметра: фіксація зменшує розмірність простору пошуку ГА на одиницю та прискорює збіжність за рахунок незначного звуження простору можливих розв'язків.

Остаточний вибір між чотирма розглянутими варіантами (два фазифікатори; два фазифікатори з фіксованою шириною зони; дві ширини зони; дві ширини зони з фіксованим фазифікатором) виконується на основі порівняння значень цільової функції (індексу Се–Бені чи гіпероб'єму H), отриманих для кожного варіанта на конкретній вибірці даних технічного моніторингу.

3.8 Проблема вибору методу кластеризації на основі нечітких множин першого типу або інтервальних нечітких множин другого типу в інформаційних технологіях оцінки технічного стану об'єктів будівництва

При виборі методу кластеризації на основі НМТ1 або ІНМТ2 може бути використана наступна методика.

У Clustering4Build вибір між кластеризацією на основі НМТ1 та ІНМТ2 здійснює мікросервіс ModelSelector, який аналізує параметри структурованих даних, гіпероб'єми, щільність об'єктів та індекси якості, отримані від QualityEvaluator.

Якщо для попередньо сформованих кластерів з мінімальним і максимальним гіпероб'ємами $h_{\min}$ і $h_{\max}$ відповідно виконується співвідношення:

$$h_{\max} / h_{\min} \geq 1,5 \quad (3.37)$$

де $h_{\min} = \min_{j=1,c} (h_j)$, $h_{\max} = \max_{j=1,c} (h_j)$, $h_j = (\det(R_j))^{1/2}$ - гіпероб'єм j -го кластера

($j = \overline{1, c}$), R_j — обчислюється за формулою (3.21), то слід визнати, що кластери мають «істотно різний» об'єм і застосовувати для кластеризації такої множини об'єктів методи кластеризації на основі ІНМТ2.

У разі, коли для попередньо сформованих кластерів з мінімальним і максимальним гіпероб'ємами h_{min} і h_{max} співвідношення (3.37) не виконується, то доцільно оцінити попередню кількість N_j об'єктів в кожному j -му кластері ($j = \overline{1, c}$). У Clustering4Build цей аналіз виконується ShapeAnalyzer, який формує спектр щільності для кожного кластера та передає його ModelSelector для порівняльної оцінки моделей.

Якщо для попередньо сформованих кластерів з мінімальною і максимальною кількістю об'єктів N_{min} і N_{max} відповідно виконується співвідношення:

$$N_{max} / N_{min} \geq 1,5, \quad (3.38)$$

де $N_{min} = \min_{j=1, c} (N_j)$, $N_{max} = \max_{j=1, c} (N_j)$, N_j - попередня кількість в j -му кластері ($j = \overline{1, c}$), то слід визнати, що кластери мають «істотно різну» потужність (і, отже, «істотно різну» щільність) і застосовувати для кластеризації такої множини об'єктів методи кластеризації на основі ІНМТ2.

У разі рівномірно сформованих кластерів пріоритет надається моделям на основі НМТ1, оскільки вони забезпечують менші обчислювальні витрати за збереження достатньої точності.

Процедура вибору між НМТ1 та ІНМТ2 може виконуватися також на основі індексу сферичності Sph, який обчислюється мікросервісом QualityEvaluator згідно з формулою (3.22). Порівняння значень цього індексу для альтернативних моделей дозволяє ModelSelector визначити модель, що найбільш повно відповідає геометричним характеристикам поточного розподілу даних, і передати відповідну рекомендацію до підсистеми DSS.

3.9. Приклади класифікації технічного стану об'єктів будівництва в інформаційних технологіях оцінки технічного стану об'єктів будівництва з використанням методів кластеризації на основі інтервальних нечітких множин другого типу

Розглянемо приклад кластеризації множини об'єктів на два кластери істотно різного об'єму і істотно різної щільності в двомірному просторі, що

відповідає кластеризації даних за двома елементами оцінювання. При цьому припустимо, що кластеризація виконується, наприклад, за елементами оцінювання «фундаменти» і «стіни» за 100- бальною шкалою (чим вищий бал, тим гірше, тим вищий відсоток зносу).

Попередньо покажемо, що в разі, коли кластери мають ідентичні потужність і об'єм (щільність), тобто ідентичну структуру, кластеризація з використанням FCM-алгоритму на основі НМТ1 при будь-яких значеннях фаззифікатора m (наприклад, з множини $\{1, 1; 2, 3, 5, 10\}$) дає результати, що збігаються з очікуваними, а вибір значення фаззифікатора m надає тільки незначний вплив на розташування центрів кластерів, самі ж результати кластеризації (розбиття на кластери) залишаються незмінними.

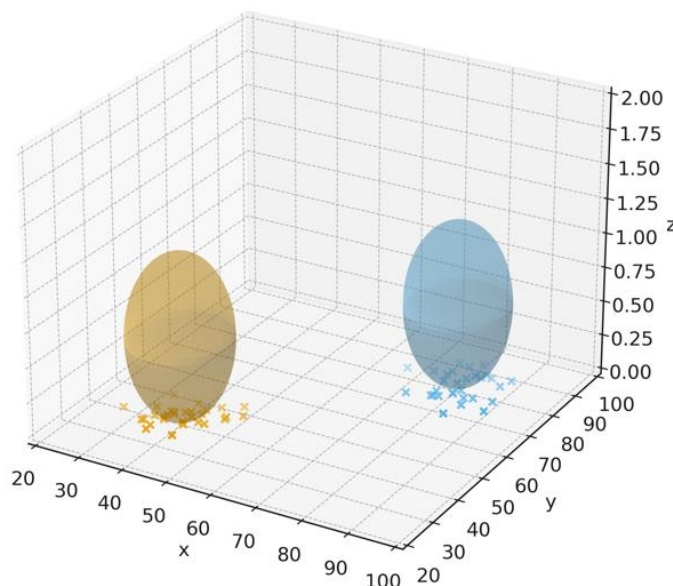


Рис. 3.15. Два кластери ідентичної структури (Розроблено автором)

На рисунку наведено тривимірну візуалізацію двох кластерів ідентичної структури у вигляді еліпсоїдних областей високої щільності. Еліпсоїди відображають наближені ізолінії однакової густини розподілу об'єктів у просторі ознак, що дозволяє наочно порівняти компактність, форму та взаємне розташування кластерів. Такий формат подання є зручним для аналізу геометричних властивостей кластерів у рамках інформаційної технології Clustering4Build. Нехай, крім того, кожний i -й об'єкт другого кластера ($i = \overline{1, 25}$) отриманий з відповідного i -го об'єкта першого кластера ($i = \overline{1, 25}$) зі зсувом по кожній координаті на число 34.

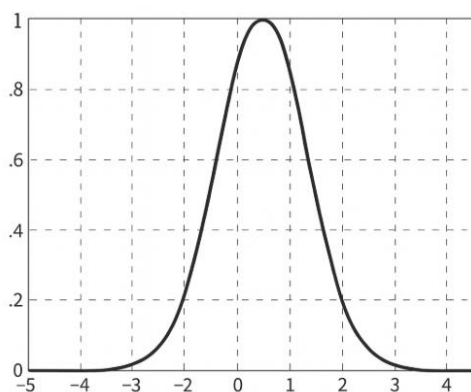


Рис. 3.16. Графічний вид нормалізованого нормального закону розподілу випадкової величини x з параметрами $\sigma = 0$ і $\delta = 1$ (Розроблено автором на основі [52])

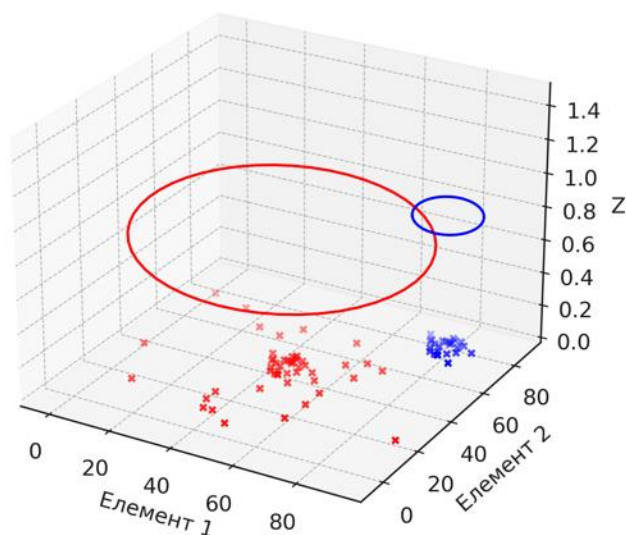


Рис. 3.17. Два кластери різного об'єму і різної щільності (Розроблено автором на основі [32])

Очевидно, що центри першого та другого кластерів повинні бути близькі до точок з координатами (44,44) (78,78) відповідно.

На рисунку подано просторову візуалізацію двох кластерів, сформованих на основі різних параметрів нормального закону розподілу. Перший кластер складається з 50 об'єктів, згенерованих за нормальними розподілами $N(44,16)$ та $N(44,841)$, що забезпечує його більшу дисперсію та відповідно більший об'єм. Другий кластер містить 25 об'єктів, отриманих на основі розподілу $N(78,16)$ і характеризується більшою щільністю. Еліпсоїдні поверхні апроксимують рівні густини та демонструють геометричні відмінності між кластерами у просторі озна

Нові кластери різної потужності і різного об'єму представлені на рисунку 3.21. У таблиці 3.1 наведені оцінки за елементами оцінювання об'єктів кластеризації.

Таблиця 3.1. Оцінки об'єктів *(Розроблено автором)*

Кластер	Об'єкт	Елемент 1	Елемент 2
Перший кластер (перші 25 об'єктів)	1	42	47
	2	37	49
	3	45	38
	4	45	38
	5	39	46
	6	49	42
	7	49	47
	8	44	47
	9	45	47
	10	45	49
	11	43	47
	12	47	49
	13	42	39
	14	53	44
	15	43	43
	16	44	38
	17	48	45
	18	44	40
	19	44	50
	20	41	41
	21	45	46
	22	39	45
	23	47	40
	24	50	35
	25	41	44
Кластер	Об'єкт	Елемент 1	Елемент 2
Другий кластер (25 об'єктів)	26	76	81
	27	71	83
	28	79	72
	29	79	72
	30	73	80
	31	83	76
	32	83	81
	33	78	81
	34	79	81
	35	79	83
	36	77	81

	37	81	83
	38	76	73
	39	87	78
	40	77	77
	41	78	72
	42	82	79
	43	78	74
	44	78	84
	45	75	75
	46	79	80
	47	73	79
	48	81	74
	49	84	69
	50	75	78
Кластер	Об'єкт	Елемент 1	Елемент 2
Перший кластер (додаткові об'єктів) 25	51	59	53
	52	93	20
	53	61	34
	54	25	62
	55	55	67
	56	15	71
	57	10	15
	58	43	50
	59	44	51
	60	35	15
	61	40	10
	62	2	75
	63	56	40
	64	70	55
	65	65	47
	66	61	26
	67	45	28
	68	64	57
	69	60	16
	70	37	67
	71	33	60
	72	35	20
	73	1	36
	74	37	10
	75	47	5

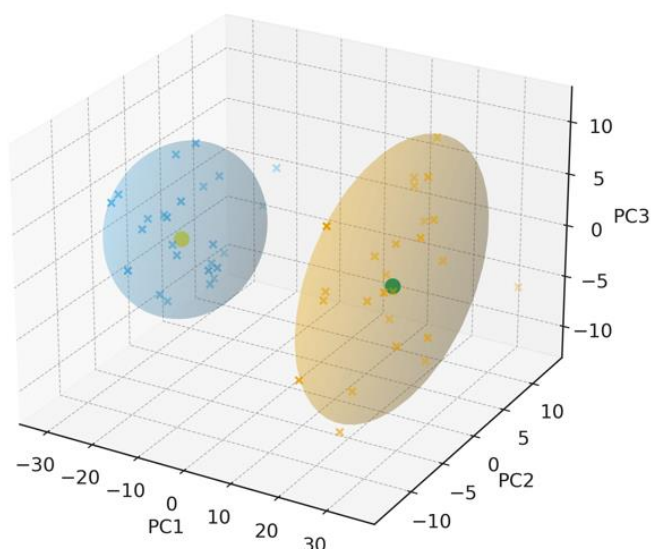


Рис. 3.18. Результати кластеризації множини з 50 об'єктів з використанням чіткого алгоритму с-середніх. (Розроблено автором)

Центри кластерів: Перший кластер:(44,44,43,84) Другий кластер: (78,44,77,84) Значення загального гіпероб'єму $H=29,173818$.

На рисунку 3.18 наведено результат кластеризації множини з 50 об'єктів, що утворюють два кластери ідентичної структури, з використанням чіткого алгоритму с-середніх. На рисунках 3.30-3.35 показані результати кластеризації множини з 50 об'єктів, що утворюють два кластери ідентичної структури, з використанням FCM-алгоритму на основі НМТ1 для значень фаззифікатора $m \in \{1,1; 2, 3, 5, 10\}$. При цьому результати з використанням класичного чіткого алгоритму с-середніх повністю збіглися з результатами кластеризації з використанням FCM-алгоритму на основі НМТ1 при $m = 1,1$. Як і очікувалося, у всіх цих випадках помилка кластеризації дорівнює нулю.

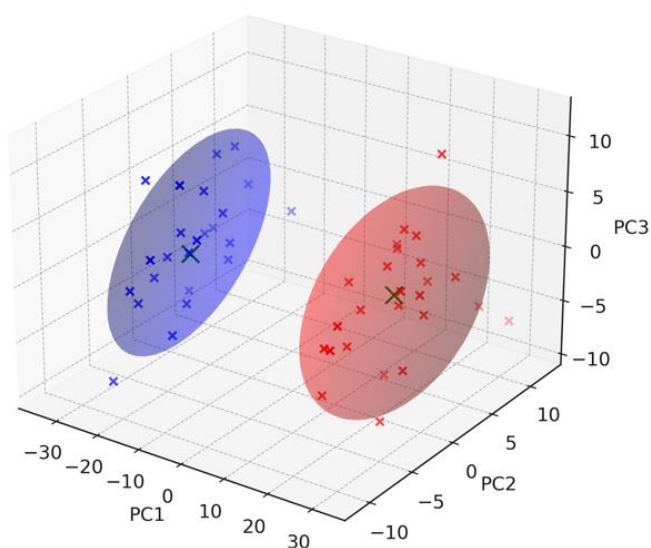


Рис.3.19. Результати кластеризації множини з 50 об'єктів для FCM-алгоритму на основі НМТ1 при $m=1,1$. *(Розроблено автором)*

Центри кластерів: Перший кластер:(44,44,43,84) Другий кластер: (78,44,77,84) Значення нечіткого загального гіпероб'єму $FH=29,173818$

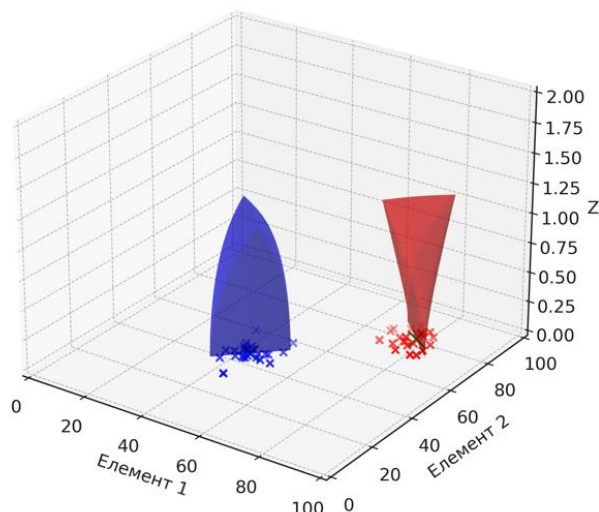


Рис. 3.20. Результати кластеризації множини з 50 об'єктів для FCM-алгоритму на основі НМТ1 при $m=2$. Центри кластерів: Перший кластер:(44,42990, 43,856397). Другий кластер: (78,426186, 78,845372). Значення нечіткого загального гіпероб'єму $FH=29,504266$ *(Розроблено автором)*

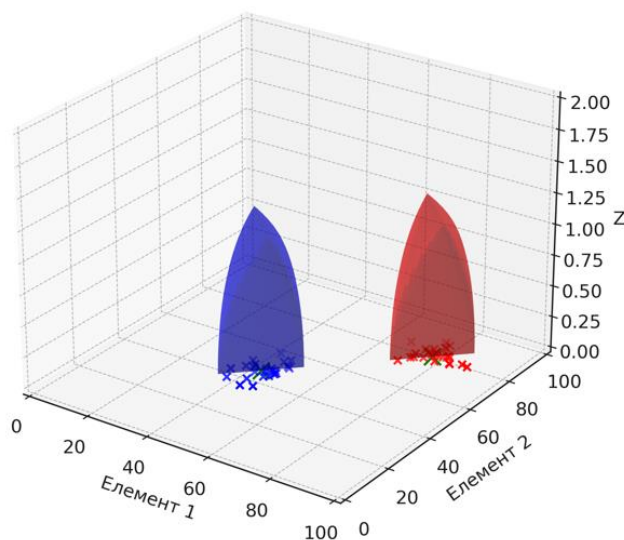


Рис.3.21. Результати кластеризації множини з 50 об'єктів для FCM-алгоритму на основі НМТ1 при $m=3$ Центри кластерів: Перший кластер:(44,368657, 43,947474) Другий кластер: (78,347867, 77,951054)

Значення нечіткого загального гіпероб'єму $FH=31,046304$ *(Розроблено автором)*

На рисунку 3.21 зображено дві групи об'єктів у просторі перших двох ознак з тривимірною інтерпретацією їхніх областей високої щільності у вигляді еліпсоїдних поверхонь. Центри кластерів, знайдені FCM-алгоритмом при $m = 3$, позначені зеленими маркерами. Представлене зображення ілюструє збереження послідовної структури кластерів при зміні фазифікатора m .

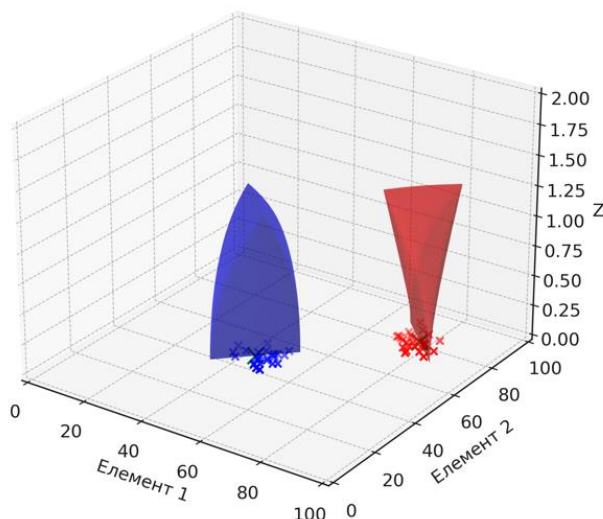


Рис.3.22. Результати кластеризації множини з 50 об'єктів для FCM-алгоритму на основі НМТ1 при $m=5$. (Розроблено автором)

На рис.3.22 наведено результати роботи FCM-алгоритму на основі нечітких множин типу 1 для значення фазифікатора $m = 5$. Зеленими позначено центри кластерів з координатами (44,562768; 45,397612) та (78,201091; 79,439204). Еліпсоїдні поверхні відображають області підвищеної густини та дозволяють візуально оцінити зміну форми кластерів порівняно з випадками $m = 2$ та $m = 3$. Значення нечіткого загального гіпероб'єму для даного розбиття становить $FH = 32,9472671$.

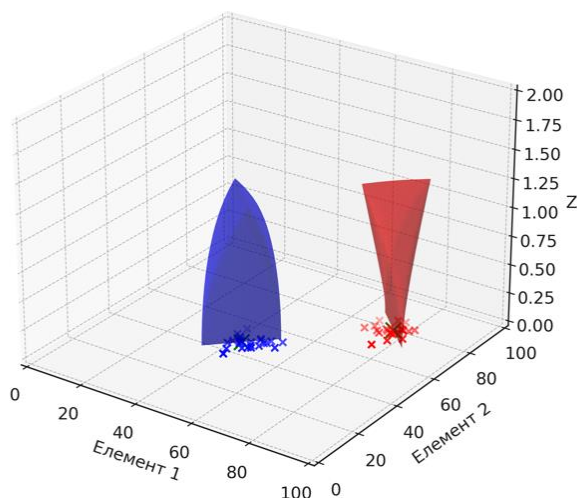


Рис. 3.23. Результати кластеризації множини з 50 об'єктів для FCM-алгоритму на основі НМТ1 при $m=10$ Центри кластерів: Перший кластер: (44,243620, 44,477169). Другий кластер: (76,987948, 77,028715). Значення нечіткого загального гіпероб'єму $FH=29,790622$ (Розроблено автором)

Необхідно зазначити, що застосування FCM-алгоритму на основі НМТ1 в даному випадку не потрібно, так як використання чіткого алгоритму c -середніх вже забезпечило адекватні результати кластеризації. Однак, отримані при використанні FCM-алгоритму на основі НМТ1 результати кластеризації множини з 50 об'єктів, що утворюють два кластери ідентичною структури, буде цікаво порівняти з результатами кластеризації множини з 75 об'єктів, що утворюють два кластери істотно різного об'єму і істотно різної щільності з використанням цього ж алгоритму.

Як видно з рисунків 3.19-3.23, всі об'єкти належним чином розбиваються на два кластери, хоча в кожному випадку розташування центрів кластерів незначно змінюється. Слід зазначити, що при $m=10$ спостерігається зміна «дзеркальності» відображення другого центру кластера щодо першого (при дотриманні «дзеркальності» відображення об'єктів першого кластера в другий кластер). При пошуку оптимального результату кластеризації з використанням FCM-алгоритму на основі НМТ1 при фіксованому значенні m був застосований комбінований метод нечіткої кластеризації на основі НМТ1, запропонований в розділі 2.

Для оцінки якості кластеризації як показник якості кластеризації - функції відповідності - використовувався загальний гіпероб'єм H за формулою (3.20) (для чіткого алгоритму c -середніх і FCM-алгоритму на основі ІНМТ2) і нечіткий загальний гіпероб'єм FH (2.13) (для FCM-алгоритму на основі НМТ1). На рисунках 3.33 і 3.34-3.38 наведені результати кластеризації множини з 75 об'єктів, що утворюють два кластери істотно різного об'єму і істотно різної щільності з використанням відповідно чіткого алгоритму c -середніх і FCM-алгоритму на основі НМТ1 для значень фаззифікатора m з множини $\{1,1; 2, 3, 5, 10\}$.

Як показав аналіз, у всіх випадках помилково відносяться до другого кластеру 3 об'єкти («об'єкт 50», «об'єкт 64» і «об'єкт 68»). Об'єкти першого кластера, помилково віднесені до другого кластеру, позначені на рисунках квадратними маркерами (як і всі об'єкти другого кластера).

На рисунку 3.24 показаний результат кластеризації множини з 75 об'єктів, що утворюють два кластери істотно різного об'єму і істотно різної щільності, з використанням FCM-алгоритму на основі ІНМТ2 для оптимальної комбінації значень фаззифікаторів m_1 і m_2 ($m_1=123,566384$ і $m_2=128,971359$), знайденої за допомогою ГА. В цьому випадку помилково віднесено до другому кластеру тільки один об'єкт («об'єкт 64»). Значення загального гіпероб'єму H склало 258,809334.

Слід зазначити, що кластеризація множини з 75 об'єктів, що утворюють два кластери істотно різного об'єму і істотно різної щільності, з використанням FCM-алгоритму на основі НМТ1 для значень фаззифікаторів m_1 і m_2 окремо дає помилку вищу. На рисунках 3.24 і 3.25 наведені результати кластеризації для FCM-алгоритму на основі НМТ1 для значень фаззифікаторів $m = m_1$ і $m = m_2$ відповідно.

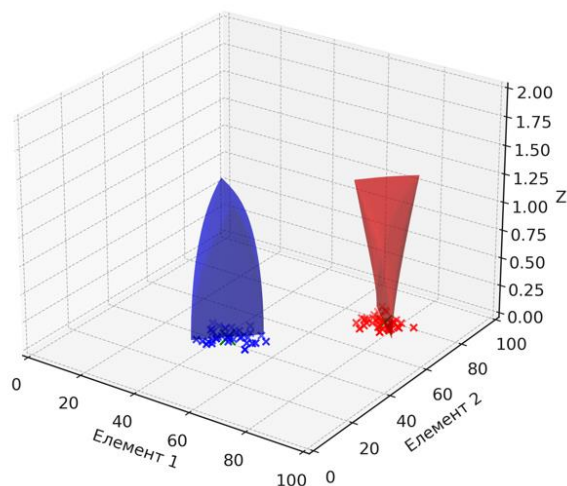


Рис. 3.24. Результати кластеризації множини з 75 об'єктів з використанням чіткого алгоритму с-середніх. Центри кластерів: Перший кластер:(43,061986, 40,579363) Другий кластер: (76,672041, 76,092026) Значення загального гіпероб'єму $H=261,801277$ (Розроблено автором)

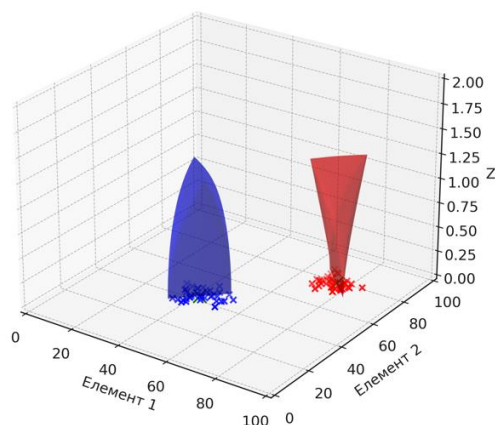


Рис. 3.25. Результати кластеризації множини з 75 об'єктів для FCM-алгоритму на основі НМТ1 при $m=1,1$. Центри кластерів: Перший кластер: (44,24861, 41,563893). Другий кластер: (77,939080, 77,793565). Значення нечіткого загального гіпероб'єму $FH=266,029983$ (Розроблено автором)

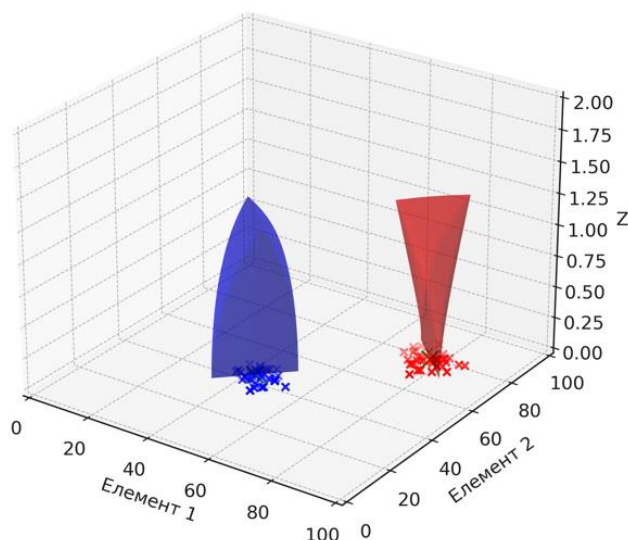


Рис. 3.26. Результати кластеризації множини з 75 об'єктів для FCM-алгоритму на основі НМТ1 при $m=2$. Центри кластерів: Перший кластер: (43,703771, 41,473404). Другий кластер: (77,695098, 76,771038). Значення нечіткого загального гіпероб'єму $FH=256,802848$ (Розроблено автором)

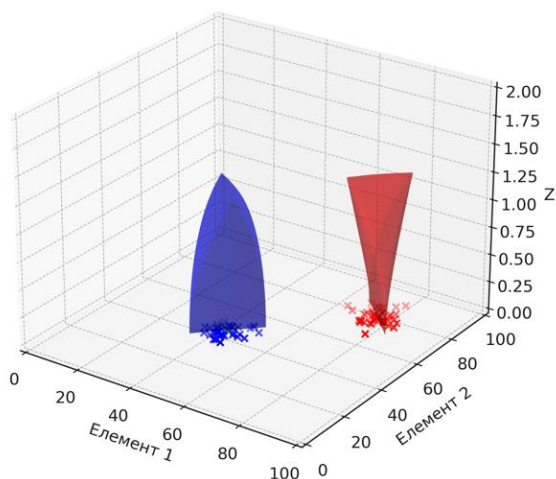


Рис. 3.27. Результати кластеризації множини з 75 об'єктів для FCM-алгоритму на основі НМТ1 при $m=3$. Центри кластерів: Перший кластер: (43,714562, 42,942246). Другий кластер: (76,873765, 76,673330). Значення нечіткого загального гіпероб'єму $FH=214,993413$ (Розроблено автором)

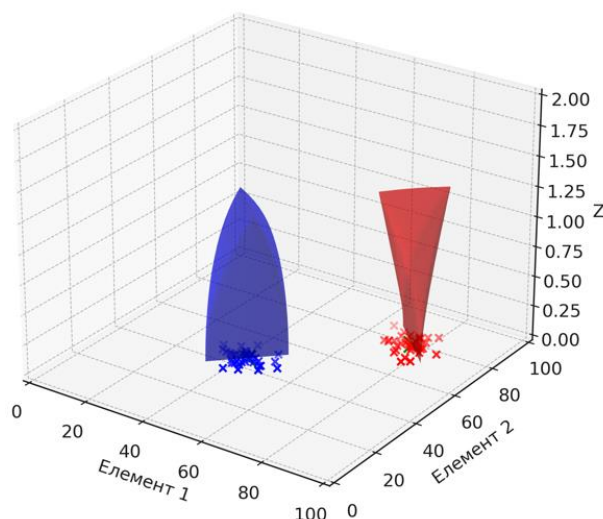


Рис. 3.28. Результати кластеризації множини з 75 об'єктів для FCM-алгоритму на основі НМТ1 при $m=5$. Центри кластерів: Перший кластер: (43,080655, 42,951077). Другий кластер: (76,749605, 77,068242). Значення нечіткого загального гіпероб'єму $FH=175,020556$ (Розроблено автором)

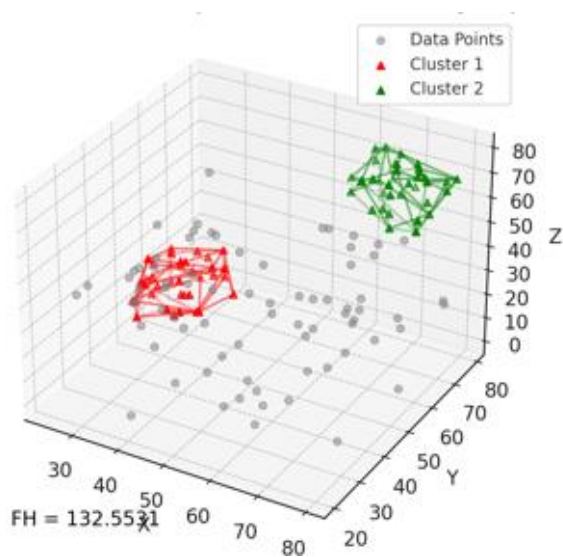


Рис.3.29. Результати кластеризації множини з 75 об'єктів для FCM-алгоритму на основі НМТ1 при $m=10$. Центри кластерів: Перший кластер: (44,022553, 38,011096). Другий кластер: (75,392049, 74,970995). Значення нечіткого загального гіпероб'єму $FH=132,553058$ (Розроблено автором)

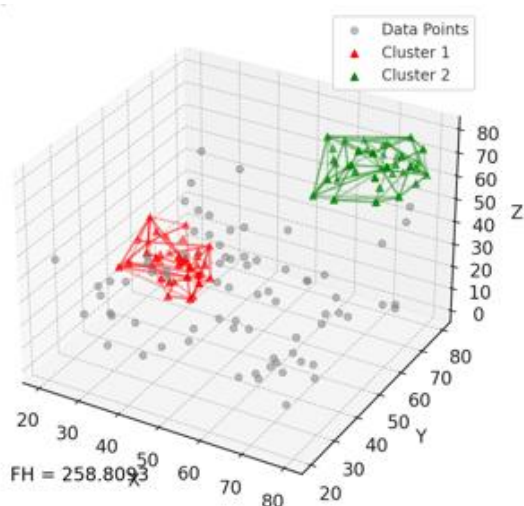


Рисунок 3.30. Результати кластеризації множини з 75 об'єктів для FCM-алгоритму на основі ІНМТ2 для комбінації $m_1 = 123,566384$, $m_2 = 128,971359$. Центри кластерів: Перший кластер: (4,500384, 42,229942). Другий кластер: (78,999999, 79,999999). Значення загального гіпероб'єму $H = 258,809334$ (Розроблено автором)

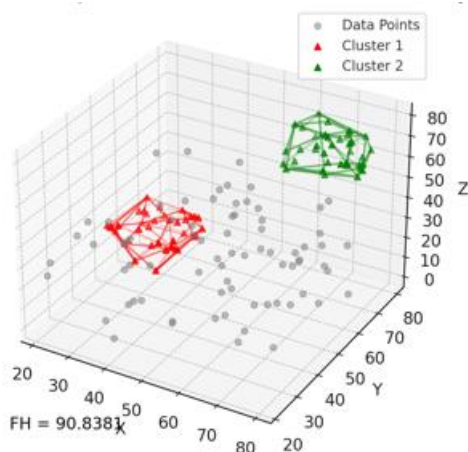


Рис. 3.31. Результати кластеризації множини з 75 об'єктів для FCM-алгоритму на основі НМТ1 для комбінації $m_1 = m_2 = 123,566384$. Центри кластерів: Перший кластер: (41,010356, 41,124315). Другий кластер: (75,002471, 74,0183144). Значення нечіткого загального гіпероб'єму $FH = 90,838124$ (Розроблено автором)

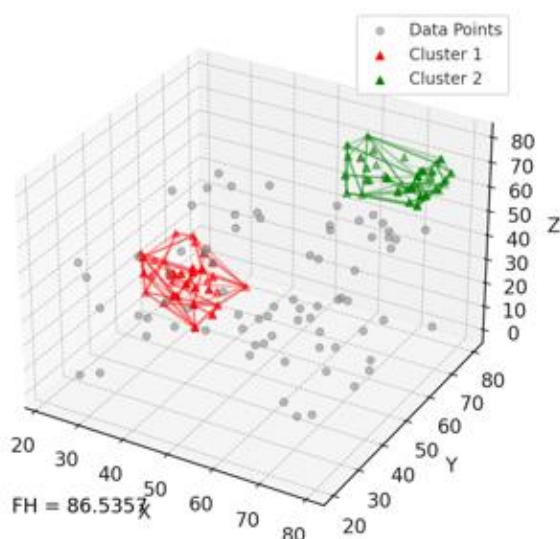


Рис.3.32. Результати кластеризації множини з 75 об'єктів для FCM-алгоритму на основі НМТ1 для комбінації $m_1 = m_2 = 128,971359$. Центри кластерів: Перший кластер: (42,979472, 42,982356). Другий кластер: (77,337608, 74,360781). Значення нечіткого загального гіпероб'єму $FH=86,535730$ (Розроблено автором)

В ідеальному випадку, якби координати центрів кластерів були б рівні (44,44) (78,78), значення загального гіпероб'єму H склало б 255,329649.

Таким чином, спільне застосування FCM-алгоритму на основі ІНМТ2 і ГА є доцільним у разі кластеризації ожини об'єктів істотно різного об'єму і істотно різної щільності.

У прикладі послідовно застосовано FCM-алгоритм та РСМ-алгоритм на основі ІНМТ2, після чого виконано порівняння якості кластеризації за критеріями (3.20)–(3.23) і обрано кращу модель.

3.9.1. Кластеризація множини об'єктів на два кластери суттєво різного об'єму і суттєво різної щільності з використанням методу можливої кластеризації на основі інтервальних нечітких множин другого типу для фіксованої комбінації значень фаззифікаторів

Розглянемо кластеризацію множини об'єктів, оцінки яких наведені в таблиці 3.1 (п. 3.8.1), на два кластери істотно різного об'єму і істотно різної щільності в двовимірному просторі з використанням методу можливої кластеризації на основі ІНМТ2 при невизначеності вибору фаззифікатора m .

У межах інформаційної технології Clustering4Build цей приклад реалізується у модулі РСМ-Core, який отримує нормалізовані оцінки об'єктів з DataRepository та виконує кластеризацію на основі ІНМТ2 для фіксованої

комбінації значень фаззифікаторів. Результати роботи PCM-Core (центри кластерів, ступені типовості та значення функції відповідності) уніфіковано передаються до мікросервісів QualityEvaluator і DSS. PCM-Core оптимізує процес кластеризації завдяки використанню ІНМТ2, що дозволяє враховувати різні рівні нечіткості і кластерної відносності в даних, що робить його ефективним для кластеризації об'єктів із сильно варіативними характеристиками.

При цьому був реалізований пошук оптимальних значень «ширини зони» для двох кластерів при вже фіксованих значеннях фаззифікаторів $m_1 = 123,566384$ і $m_2 = 128,971359$, знайдених за допомогою методу нечіткої кластеризації на основі ІНМТ2. Оптимізація параметрів «ширини зони» є ключовим етапом у покращенні точності кластеризації, оскільки правильне налаштування цих значень дозволяє значно зменшити помилки кластеризації і покращити розподіл об'єктів за кластерами. Значення «ширини зони» виявилися рівні: $\eta_1 = 142,573340$ і $\eta_2 = 74,273482$.

Пошук оптимальних значень параметрів «ширини зони» реалізовано в окремому мікросервісі ParameterOptimizer, який викликає PCM-Core у циклі оптимізації, оцінюючи якість кластеризації за індексами H та Sph. Знайдені оптимальні значення зберігаються у конфігураційному сховищі ConfigStore і можуть повторно використовуватися для запуску РСМ-алгоритму на stream-даних без додаткової оптимізації. Такий підхід до зберігання оптимізованих параметрів забезпечує високий рівень автоматизації і дозволяє швидко адаптувати алгоритм до нових даних без необхідності повторної оптимізації.

При використанні методу можливої кластеризації на основі ІНМТ2 при невизначеності вибору фаззифікатора m всі об'єкти відповідним чином були розподілені за двома кластерами, тобто помилка кластеризації виявилася рівною нулю, а значення загального гіпероб'єму H склало 255,809049 (рис. 3.33).

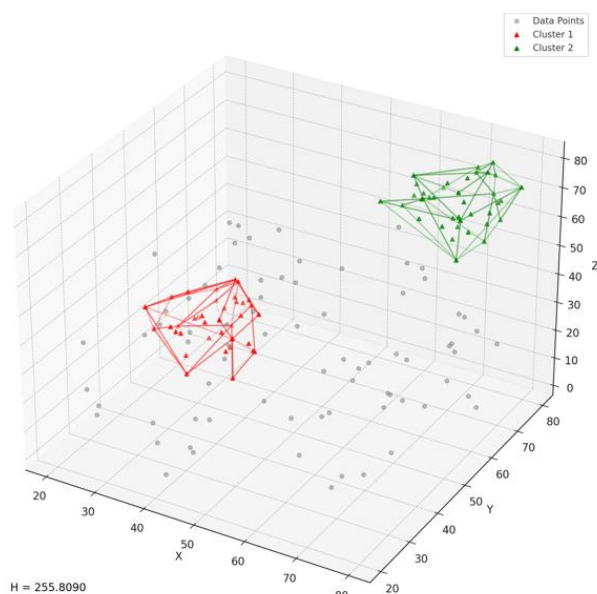


Рис.3.33. Результати кластеризації множини з 75 об'єктів для РСМ-алгоритму на основі ІНМТ2 для комбінації $m_1 = 123,566384$, $m_2 = 128,971359$. $\eta_1 = 142,573340$, $\eta_2 = 74,273482$ Центри кластерів: Перший кластер: (43,994790, 43,988567). Другий кластер: (78,781532, 78,874113) Значення загального гіпероб'єму $H = 255,809049$. (Розроблено автором)

Результати кластеризації передаються у підсистему DSS для візуалізації структури множини об'єктів, де вони використовуються для аналізу та прийняття рішень щодо технічного стану об'єктів. Мікросервіс ModelSelector здійснює порівняння РСМ-алгоритму з іншими методами кластеризації, використовуючи загальні індекси якості, що дозволяє вибрати найбільш ефективну модель для даного набору даних

3.9.2 Кластеризація множини об'єктів на два кластери суттєво різного об'єму і суттєво різної щільності з використанням методу можливої кластеризації на основі інтервальних нечітких множин другого типу для довільної комбінації значень фаззифікаторів

Для множини об'єктів, оцінки яких наведені в таблиці 3.1 (п. 3.8.1), був реалізований ГА пошуку оптимальної комбінації значень фаззифікаторів m_1 , m_2 і значень «ширини зони» $\eta_j (j = \overline{1,2})$ за допомогою РСМ-алгоритму на основі ІНМТ2. Генетичний алгоритм (ГА) забезпечує ефективний пошук оптимальних параметрів для РСМ-алгоритму на основі ІНМТ2, що дозволяє зменшити помилки кластеризації навіть при складних структурах даних,

таких як кластери з істотно різним об'ємом і щільністю. Пошук таких параметрів є критичним для досягнення найкращих результатів кластеризації.

При цьому були отримані наступні значення фаззифікаторів m_1 , m_2 і m_2 значення «ширини зони» $\eta_j (j = \overline{1,2})$: $m_1 = 113,769576$; $m_2 = 128,735447$; $\eta_1 = 143,222654$; $\eta_2 = 73,404309$, Ці параметри забезпечили нульову помилку кластеризації (рисунок 3.38), що підтверджує високу точність алгоритму при класифікації об'єктів, що належать до двох кластерів різного об'єму та щільності. (рисунок 3.38). Значення функції відповідності - загального гіпероб'єму N за формулою склало 255,635142. Ці оптимальні значення фаззифікаторів і «ширини зони» дозволяють мінімізувати помилку кластеризації, що є важливим для підвищення точності розподілу об'єктів між кластерами, особливо коли дані мають різну щільність та об'єм.

При ініціалізації РСМ-алгоритму на основі ІНМТ2 для кожної хромосоми використовувалося випадкове розбиття об'єктів на кластери з визначенням значень ФП (як в FCM-алгоритмі на основі НМТ1), на основі яких обчислювалися початкові координати центрів кластерів. Використання випадкового розбиття є ефективним методом для уникнення локальних мінімумів і дозволяє алгоритму знайти більш точне розбиття, яке відповідає реальним властивостям даних. Цей підхід дозволяє забезпечити адаптивність алгоритму, автоматично підбираючи початкові умови для кластеризації, що оптимізують результати навіть при різній структурі даних.

При використанні FCM-алгоритму на основі ІНМТ2 та оптимізації параметрів за допомогою ГА, було досягнуто значно кращих результатів, із зменшенням помилки кластеризації та покращенням точності центроїдів кластерів. (рисунок 3.34). При фіксованих значеннях фаззифікаторів для FCM-алгоритму на основі ІНМТ2 можуть виникати незначні помилки кластеризації, однак їх можна зменшити шляхом подальшої оптимізації параметрів, таких як «ширина зони», за допомогою генетичного алгоритму

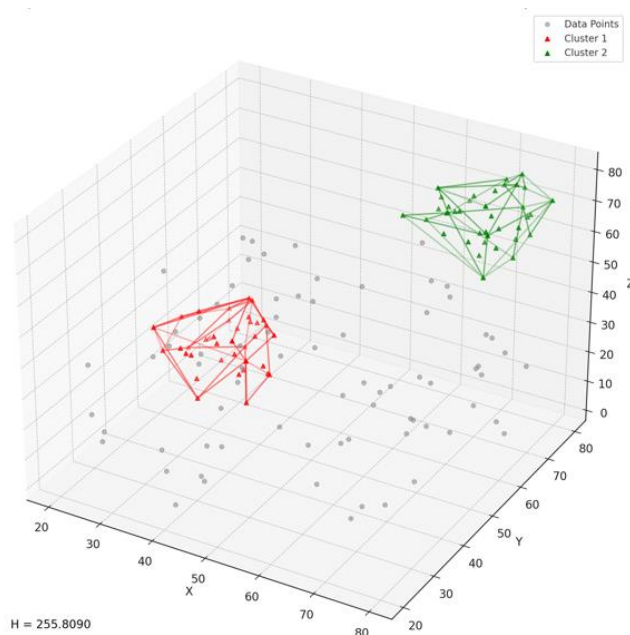


Рис.3.35. Результати кластеризації множини з 75 об'єктів для РСМ-алгоритму на основі ІНМТ2 для комбінації $m_1 = 113,769576$, $m_2 = 128,735447$, $n_1 = 143,22654$, $n_3 = 73,404309$. Центри кластерів: Перший кластер: (44,000000, 40,000000). Другий кластер: (76,827239, 77,753212). Значення загального гіпероб'єму $H = 255,635142$ (Розроблено автором)

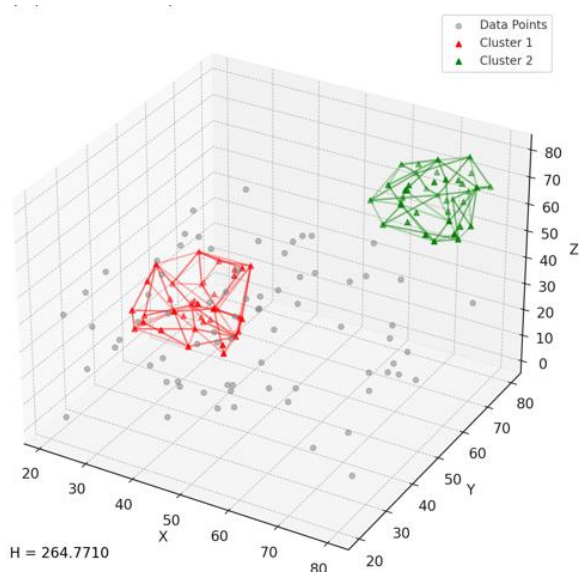


Рис. 3.36. Результати кластеризації множини з 75 об'єктів для FSCM-алгоритму на основі ІНМТ2 для комбінації $m_1 = 113,769576$, $m_2 = 128,735447$. Центри кластерів: Перший кластер: (44,276529, 40,849284) Другий кластер: (74,932549, 73,993571) Значення загального гіпероб'єму $H = 264,771044$ (Розроблено автором)

Реалізація РСМ-алгоритму на основі ІНМТ2 разом із генетичним алгоритмом для оптимізації фаззифікаторів та «ширини зони» дозволяє

отримати більш точні результати кластеризації, зменшуючи помилки при кластеризації множин з різною щільністю та об'ємом..

Якщо для кластеризації множини об'єктів спочатку вибирається FCM-алгоритм на основі ІНМТ2, для якого вдається визначити деяку комбінацію значень фаззифікаторів m_1 і m_2 , що забезпечила мінімально можливу помилку кластеризації із застосуванням даного алгоритму, то при необхідності поліпшення результатів кластеризації слід використовувати ГА пошуку оптимальних результатів кластеризації з використанням РСМ-алгоритму на основі ІНМТ2 для фіксованої комбінації значень фаззифікаторів m_1 і m_2 . При цьому за допомогою ГА виконується пошук оптимальної комбінації значень «ширини зони» $\eta_j(j = \overline{1, c})$.

В разі, коли для виконання кластеризації множини об'єктів відразу вибирається РСМ-алгоритм на основі ІНМТ2, то при його реалізації знаходимо оптимальну комбінацію значень фаззифікаторів m_1 , і m_2 значень «ширини зони» $\eta_j(j = \overline{1, c})$.

Запропонований підхід забезпечує значне зменшення помилок класифікації та покращує адаптивність алгоритмів до складних даних.

Розглянемо застосування методики вибору методу кластеризації (з використанням FCM-алгоритму на основі ІНМТ2 або РСМ-алгоритму на основі ІНМТ2) для множини об'єктів, оцінки яких наведені в таблиці 3.1 (п. 3.8.1), на два кластери істотно різного об'єму і істотно різної щільності в двовимірному просторі.

При цьому в якості показника якості кластеризації буде використовуватися індекс загального гіпероб'єму H за формулою (3.20). Використання цього показника дозволяє оцінити ефективність кластеризації за допомогою порівняння значень гіпероб'єму для різних методів кластеризації.

Таблиця 3.2. Мінімальне значення загального гіпероб'єму (Розроблено автором)

Алгоритм кластеризації	Загальний гіпероб'єм	Помилка кластеризації
Алгоритм чітких с-середніх	261,801277	3
FCM-алгоритм на основі ІНМТ2	258,809334	1

PCM-алгоритм на основі ІНМТ2 при фіксованих значеннях фаззифікаторів, знайдених за допомогою FCM-алгоритму на основі ІНМТ2	255,809049	0
PCM-алгоритм на основі ІНМТ2 при довільних значеннях фаззифікаторів	255,635142	0

Як видно з таблиці 3.2, мінімальне значення загального гіпероб'єму за формулою (3.20) досягається при використанні методу кластеризації, що реалізує застосування PCM-алгоритму на основі ІНМТ2, при довільних значеннях фаззифікаторів. Це свідчить про перевагу застосування PCM-алгоритму для класифікації об'єктів, де кластери мають істотно різну щільність і об'єм.

Отримані результати підтверджують високу ефективність PCM-алгоритму на основі ІНМТ2, оскільки помилка кластеризації на множині об'єктів становить 0, що свідчить про точність і надійність методу. Отже, застосування PCM-алгоритму на основі ІНМТ2 для класифікації об'єктів з різною щільністю та об'ємом дозволяє досягти мінімальної помилки кластеризації, що підтверджує його ефективність для задач технічного моніторингу.

Слід зазначити, що наведені результати отримано для одного запуску генетичного алгоритму на кожен контрольний набір даних без усереднення за багатьма реалізаціями та без оцінки статистичної значущості відмінностей між порівнюваними методами. Ураховуючи стохастичну природу генетичного алгоритму (випадкова ініціалізація популяції, оператори схрещування й мутації), результати окремого запуску можуть варіюватися. Для остаточного підтвердження переваги PCM-алгоритму на основі ІНМТ2 доцільно виконати серію повторних запусків (не менше 10–30) для кожної конфігурації параметрів з подальшим обчисленням середнього значення та довірчого інтервалу цільової функції, що становить напрям подальших досліджень.

ВИСНОВКИ ДО РОЗДІЛУ 3

У третьому розділі розроблено математичний апарат і алгоритмічні засоби кластеризації технічного стану об'єктів будівництва на основі інтервальних нечітких множин другого типу (ІНМТ2), що забезпечують підвищення достовірності оцінювання в умовах структурної невизначеності, шумових впливів та неповноти вхідних даних.

Встановлено, що методи кластеризації на основі нечітких множин першого типу (розділ 2) не враховують невизначеність самої функції належності, що обмежує їх застосовність для об'єктів із перехідними технічними станами. Показано, що поширення апарату ІНМТ2 на FCM-алгоритм, розроблене в роботі [16] лише для нечіткої кластеризації, потребує розширення на можливісну (PCM) кластеризацію — вільну від умови нормування ступенів приналежності і тому стійкішу до шумових та аномальних даних, характерних для реальних вимірювань технічного стану.

У межах розділу було:

- розроблено математичне забезпечення кластеризації на основі ІНМТ2, яке дозволяє формалізувати невизначеність функції належності через верхню і нижню межі (FOU) та обчислювати інтервальний центроїд кластера за допомогою ітераційного алгоритму Карника–Менделя;
- поширено можливісний PCM-алгоритм на ІНМТ2 у двох незалежних варіантах формалізації невизначеності — через два фаззифікатори m_1, m_2 та через два значення «ширини зони» η_{j1}, η_{j2} кожного кластера, з обґрунтуванням критерію вибору між варіантами залежно від природи вхідних даних;
- розроблено комбіновані алгоритми FCM+GA та PCM+GA на основі ІНМТ2 з автоматизованим генетичним пошуком оптимальних значень фаззифікаторів і «ширини зони», що дозволяє суттєво скоротити час підбору параметрів порівняно з повним перебором;
- запропоновано показники якості кластеризації — загальний гіпероб'єм H та індекс сферичності Sph — адаптовані до кластерів різної форми,

щільності та об'єму, а також критерії (3.37)–(3.38) для визначення «істотної відмінності» кластерів за об'ємом і потужністю;

- сформульовано методику вибору між моделями на основі НМТ1 та ІНМТ2 залежно від характеристик вхідних даних (ступеня однорідності та щільності кластерів), що забезпечує адаптивність інформаційної технології до різних умов технічного моніторингу.

Обчислювальні експерименти на контрольних наборах даних (розділ 3.9) підтвердили, що застосування РСМ-алгоритму на основі ІНМТ2 забезпечує кращі показники якості кластеризації для множин об'єктів із суттєво різною щільністю та об'ємом порівняно з методами на основі НМТ1, що обґрунтовує доцільність розробленого математичного апарату для задач оцінювання технічного стану будівель.

Отримані результати доповнюють методи розділу 2 засобами формалізації невизначеності меж кластерів і створюють підґрунтя для програмної реалізації розробленої інформаційної технології, розглянутої в розділі 4.

РОЗДІЛ 4. ПРОГРАМНА РЕАЛІЗАЦІЯ МЕТОДІВ КЛАСТЕРИЗАЦІЇ НА ОСНОВІ НЕЧІТКИХ МНОЖИН

4.1 Загальні характеристики пакету прикладних програм «Кластеризація технічного стану об'єктів будівництва в умовах невизначеності»

Пакет прикладних програм «Кластеризація технічного стану об'єктів будівництва в умовах невизначеності» призначений для багатокритеріальної кластеризації сукупностей об'єктів, що містять як кластери подібної щільності та об'єму, так і кластери істотно різної щільності або об'єму. Програмний комплекс застосовує сучасні методи кластеризації на основі нечітких множин другого типу (ІНМТ2) та генетичних алгоритмів для адаптивного налаштування параметрів класифікації.

ППП дозволяє здійснювати комплексний багатокритеріальний аналіз, класифікацію та кластеризацію технічного стану об'єктів будівництва, включаючи автоматизований контроль якості даних на всіх етапах. Цей пакет ефективно застосовується як для інструментальних досліджень груп будівельних об'єктів, так і для їх експертного оцінювання. Множина об'єктів може включати як кластери схожого об'єму та щільності, так і кластери з істотними відмінностями в цих параметрах.

ППП інтегровано в сучасну інформаційну технологію Clustering4Build, що дозволяє виконувати кластеризацію в реальному часі, здійснюючи автоматичний моніторинг технічного стану об'єктів будівництва. Використовувані методи кластеризації з урахуванням ІНМТ2 дозволяють досягти високої точності класифікації навіть при наявності різнорідних об'єктів з різною щільністю і об'ємом.

ППП «Кластеризація технічного стану об'єктів будівництва в умовах невизначеності» активно використовується в діяльності Альфа-Сервіс, Марстон-Груп, що підтверджується актом впровадження (Додаток 1).

Крім того, ППП «Кластеризація технічного стану об'єктів будівництва в умовах невизначеності» застосовується в навчальному процесі Київського національного університету будівництва і архітектури в навчанні студентів спеціальностей 126 «Інформаційні системи і технології» і 122 «Комп'ютерні

науки» в курсах «Інтелектуальні інформаційні системи і технології діагностики технічного стану будівель», «Методи експертних оцінок в системах прийняття рішень» і «Системи штучного інтелекту». В навчанні студентів спеціальностей 123 «Комп'ютерні системи і мережі» і 125 «Кібербезпека» в курсах «Моделі і методи інформаційних систем діагностики технічного стану об'єктів будівництва » і «Системи штучного інтелекту».

Умови роботи ППП для IT-системи Clustering4Build включають комп'ютер з процесором не нижче Intel Core i3 або еквівалентного, з тактовою частотою від 2,0 GHz та вище, оперативну пам'ять не менше 8 ГБ для ефективною роботи з великими обсягами даних, графічний адаптер, сумісний з 4K роздільною здатністю та підтримкою OpenGL, жорсткий диск об'ємом не менше 500 ГБ або SSD для швидшого доступу до даних, операційну систему Windows 10/11 або іншу сучасну ОС (macOS), а також використання системи інженерних і наукових розрахунків на основі Python з бібліотеками для наукових обчислень (NumPy, SciPy, pandas) і контейнеризацію програмних компонентів через Docker.

Таблиця 4.1. Технологічний стек (Розроблено автором)

Рівень	Технології	Призначення
Data Acquisition	Python (Flask, Paho MQTT), BIM API	Збір та синхронізація даних
Fuzzification	NumPy, Pandas, scikit-fuzzy	Формування нечітких векторів
Clustering	scikit-fuzzy, SciPy	Реалізація FCM/PCM кластеризації
Optimization	DEAP, NumPy	Генетична оптимізація параметрів кластеризації
Database	PostgreSQL + JSONB	Зберігання векторів, центрів, параметрів
DSS	Plotly Dash, ReactJS	Візуалізація, інтерфейс користувача
Integration	REST API, WebSocket	Обмін даними між мікросервісами

4.2 Особливості розробки пакету прикладних програм інтегрованого в сучасну інформаційну технологію Clustering4Build

Пакет прикладних програм «Кластеризація технічного стану об'єктів будівництва в умовах невизначеності» був розроблений для здійснення багатокритеріальної кластеризації сукупностей об'єктів будівництва, що включають як кластери подібної щільності та об'єму, так і кластери з істотними відмінностями в цих параметрах. Розроблений ППП інтегровано в сучасну інформаційну технологію Clustering4Build, що є платформою для класифікації та аналізу технічного стану об'єктів будівництва з урахуванням невизначеності та складності даних. Ця компонентна діаграма відображає структуру Clustering4Build та взаємодію між основними модулями, що входять до складу системи.

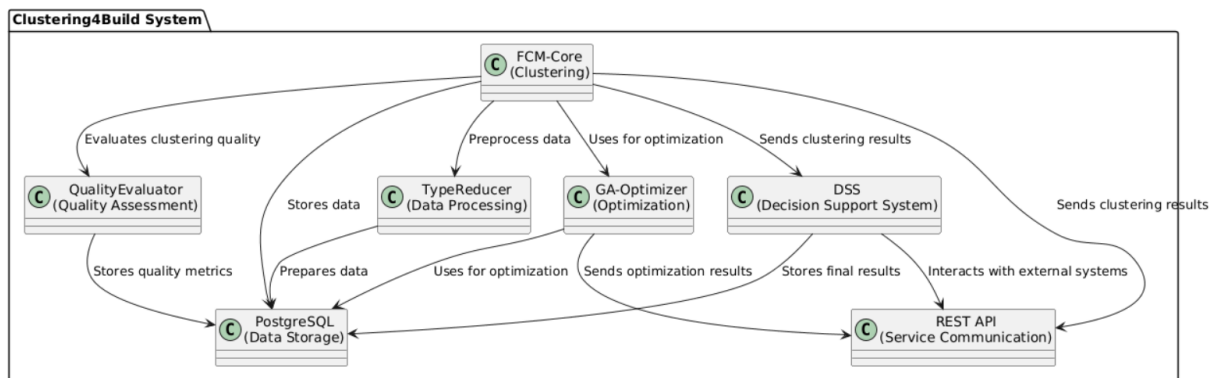


Рис. 4.1. Компонентна діаграма UML для відображення взаємодії між основними модулями системи (Розроблено автором)

Вона демонструє мікросервісну архітектуру, де кожен компонент виконує окрему функцію в процесі класифікації та кластеризації технічного стану об'єктів будівництва. Компоненти, такі як FCM-Core, GA-Optimizer, TypeReducer, QualityEvaluator та DSS, інтегровані через REST API, що забезпечує їх ефективну взаємодію в реальному часі.

Архітектура Clustering4Build реалізує модульну інформаційну технологію для збору, оброблення, кластеризації та візуалізації технічних даних будівельних конструкцій. Система орієнтована на інтеграцію з IoT-сенсорами, базами даних моніторингу та цифровими двійниками об'єктів (BIM-моделями).

У програмній реалізації кожен компонент Clustering4Build функціонує як окремий мікросервіс:

- Data Acquisition Service — імпорт даних із сенсорів і BIM;
- Fuzzification Service — фазифікація технічних показників;
- Clustering Core (FCM/PCM) — аналітичне ядро кластеризації;
- Optimization Controller / GeneticOptimizer — адаптивна оптимізація кластеризації;
- Database Service — збереження результатів;
- DSS Visualization Service — інтерактивна панель користувача.

Таблиця 4.2. Перелік основних компонентів та інтерфейсів *(Розроблено автором)*

Компоненти верхнього рівня:	
DataLoader	завантаження та попередня обробка даних.
FuzzificationEngine	формування IHMT2: μL , μU
FCMCore	реалізація FCM на основі IHMT2, формує запит до КМ.
TypeReducer (KMService)	реалізація алгоритму Карника–Менделя.
ClusterCenterStore	сховище центрів кластерів (PostgreSQL/JSONB).
DSS/Analytics	підсистема підтримки рішень та візуалізації.
Інтерфейси між компонентами:	
IDataSource	інтерфейс доступу до DataLoader.
IFuzzificationService	інтерфейс фазифікації (IHMT2).
IFCMService	інтерфейс кластераналізу для DSS.
ITypeReductionService	інтерфейс виклику алгоритму Карника–Менделя.
IClusterCenterRepository	інтерфейс збереження/читання центрів кластерів.

- FCM-Core є основним компонентом для класифікації, що реалізує алгоритм кластеризації за допомогою нечітких множин.
- GA-Optimizer використовує генетичні алгоритми для оптимізації параметрів класифікації, таких як фазифікатор.

- TypeReducer виконує попередню обробку даних, знижуючи вплив шуму та помилок.
- QualityEvaluator оцінює точність класифікації за допомогою індексів, таких як H і Sph.

DSS генерує рішення на основі класифікації, надаючи кінцеві рекомендації для користувачів. Мікросервісна модель забезпечує незалежність, масштабованість і можливість розподілення навантаження між серверними вузлами.

Взаємодія між сервісами виконується за допомогою REST-запитів типу POST `/api/genetic/init`, `/api/pcm/update`, `/api/dss/update`, що забезпечує узгодженість і автономність підсистем.

- PostgreSQL зберігає всі дані, включаючи результати класифікації, центри кластерів та інші важливі показники.
- REST API забезпечує інтеграцію та комунікацію між компонентами, що дозволяє їм ефективно обмінюватися даними.

Основною особливістю розробленого пакету є використання методів кластеризації на основі нечітких множин другого типу (INMT2) описаних в розділі 3, що дозволяє здійснювати класифікацію навіть при наявності кластерів з різними характеристиками об'єму та щільності. Ці методи забезпечують високий рівень адаптивності до варіативних характеристик даних, що є важливим для застосування в умовах реального часу, коли об'єкти можуть мати різну інтенсивність та параметри.

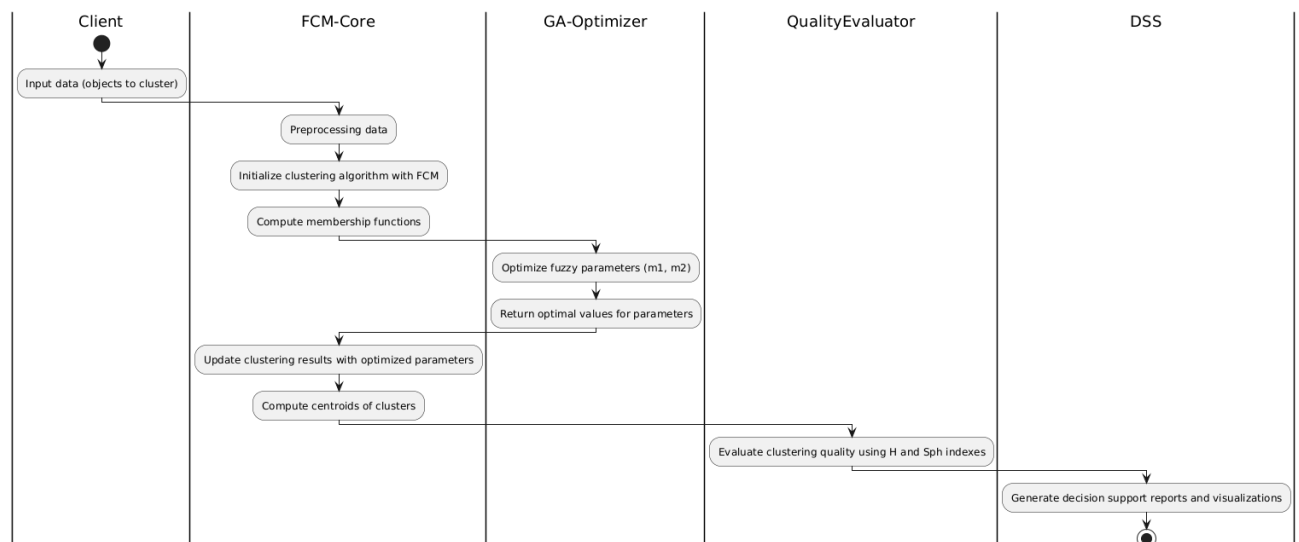


Рис. 4.2. Діаграма активності UML для відображення процесу кластеризації з використанням FCM-алгоритму на основі ІНМТ2 (*Розроблено автором*)

Діаграма активності описує процес кластеризації в системі Clustering4Build з використанням FCM-алгоритму на основі ІНМТ2 та генетичних алгоритмів для оптимізації параметрів. Процес починається з надання вхідних даних для класифікації. Після цього, система ініціалізує кластеризацію, де FCM-Core обробляє вхідні дані і виконує перший етап класифікації. На наступному етапі GA-Optimizer оптимізує параметри класифікації (наприклад, фазифікатори), що дозволяє покращити точність класифікації. Після цього, FCM-Core виконує повторну класифікацію з оптимізованими параметрами, і передає результати до QualityEvaluator, який оцінює якість класифікації, використовуючи індекси H та Sph. Нарешті, DSS генерує рішення на основі отриманих результатів, що забезпечує точні рекомендації для користувачів.

Інтеграція генетичних алгоритмів для оптимізації параметрів класифікації дозволяє автоматично налаштовувати параметри фазифікатора і ширини зони, що критично важливо для досягнення високої точності класифікації в умовах різноманітності даних. Це дозволяє значно підвищити ефективність роботи системи Clustering4Build, мінімізуючи вплив помилок класифікації, зокрема в умовах наявності шумів або складних для розпізнавання структур.

ППП має чітко виражену архітектуру, побудовану на основі мікросервісної технології, де кожен компонент має свою задачу в процесі класифікації і кластеризації. Це включає в себе:

- FCM-Core для основної кластеризації з використанням алгоритмів нечіткої класифікації;
- TypeReducer для адаптивного зниження впливу на класифікацію об'єктів з помилковими або неповними даними;
- GA-Optimizer, що відповідає за оптимізацію параметрів класифікації з урахуванням різноманітності вхідних даних.

Всі ці компоненти взаємодіють через REST API, що дозволяє забезпечити гнучкість і масштабованість програмного продукту. Це дозволяє

Clustering4Build інтегруватися з іншими інформаційними системами для подальшого аналізу та використання результатів класифікації.

Крім того, система підтримує інтеграцію з базами даних, зокрема PostgreSQL, що забезпечує зберігання та обробку великих обсягів даних. Це дозволяє зберігати результати класифікації, а також використовувати історичні дані для подальшого навчання і вдосконалення моделей кластеризації.

Clustering4Build забезпечує автоматизований моніторинг технічного стану будівельних об'єктів у реальному часі, що робить його надзвичайно корисним для систем технічного моніторингу, а також для оцінки та прогнозування стану об'єктів у сфері нерухомості.

Діаграма послідовності описує, як дані проходять через систему, взаємодіючи з різними компонентами Clustering4Build. Початково User надає вхідні дані для класифікації, які передаються до FCM-Core для первинної обробки. Потім FCM-Core надсилає запит на оптимізацію параметрів класифікації до GA-Optimizer.

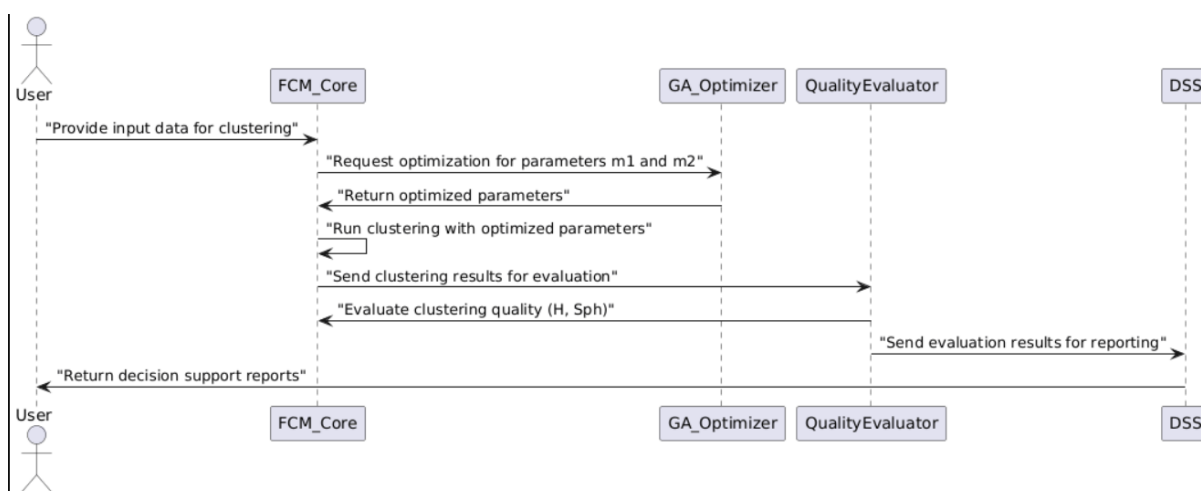


Рис.4.3. Діаграма послідовності UML для опису обміну даними між компонентами (FCM-Core, GA-Optimizer, QualityEvaluator) під час виконання кластеризації *(Розроблено автором)*

Після отримання оптимізованих значень фаззифікаторів, FCM-Core повторно виконує кластеризацію, використовуючи ці значення. Результати класифікації передаються до QualityEvaluator, який оцінює якість класифікації. Оцінка якості повертається до DSS, який генерує кінцеві звіти та рекомендації для користувача.

ППП у рамках Clustering4Build дозволяє створити платформу для ефективної класифікації і кластеризації технічного стану будівель, з урахуванням невизначеності вхідних даних, що відповідає вимогам сучасних інформаційних технологій для аналізу та оцінки складних систем.

Інформаційна технологія Clustering4Build розроблена з дотриманням базових принципів сучасного проектування інтелектуальних інформаційних систем. Її модульність забезпечує незалежне оновлення, масштабування та повторне використання окремих компонентів без порушення роботи всієї системи. Мікросервісність реалізована через розподілену архітектуру, що дозволяє гнучко розподіляти обчислювальне навантаження між сервісами, оптимізуючи ресурси під різні сценарії використання. Інтелектуальність системи досягається за рахунок застосування методів нечіткої логіки та генетичної оптимізації, які забезпечують адаптивність процесів кластеризації та підвищують точність прийняття рішень. Принцип відкритості полягає у використанні REST API, що дозволяє інтеграцію Clustering4Build із зовнішніми аналітичними, BIM або SCADA-системами. Надійність архітектури гарантується транзакційною моделлю збереження даних у PostgreSQL, механізмами журналювання подій і контролю доступу, що забезпечує цілісність і безперервність функціонування системи.

4.3. Пакет прикладних програм «Кластеризація технічного стану об'єктів будівництва в умовах невизначеності»

На основі взаємодії компонентів у системі Clustering4Build, опишемо пакети прикладних програм, що складаються з інтегрованих модулів для класифікації та кластеризації технічного стану об'єктів будівництва. Пакети прикладних програм спрямовані на вирішення завдань технічного моніторингу в умовах невизначеності та складаються з кількох основних підсистем, кожна з яких виконує конкретну задачу.

Пакет прикладних програм 1 «Кластеризація на основі нечітких множин (FCM-алгоритм і ІНМТ2)» є базовим функціональним компонентом

інформаційної технології Clustering4Build, розробленої для оцінки технічного стану об'єктів будівництва в умовах невизначеності. Метою пакету є забезпечення багатокритеріальної кластеризації множин об'єктів будівництва із застосуванням алгоритму Fuzzy C-Means (FCM) як на основі нечітких множин першого типу, так і на основі інтервальних нечітких множин другого типу (ІНМТ2), що дозволяє адекватно враховувати невизначеність, варіабельність і неповноту інформації, характерні для даних технічного моніторингу.

Ключовим обчислювальним компонентом пакету є модуль FCM-Core, який реалізує ітераційний процес кластеризації з мінімізацією відповідної цільової функції та забезпечує розрахунок функцій приналежності об'єктів до кластерів і координат центрів кластерів у багатовимірному просторі ознак. Модуль підтримує як класичну реалізацію FCM на основі нечітких множин першого типу, так і розширену версію алгоритму з використанням ІНМТ2, що дозволяє працювати з інтервальними значеннями функцій приналежності та підвищує стійкість результатів кластеризації в умовах неоднорідності даних. Для адаптації параметрів класифікаційної моделі до конкретних характеристик вхідних даних у складі пакету використовується модуль GA-Optimizer, який реалізує генетичний алгоритм оптимізації фаззифікаторів m_1 та m_2 і виконує пошук їх оптимальних значень у багатовимірному просторі параметрів з урахуванням показників якості кластеризації.

Пакет інтегрований зі сховищем даних PostgreSQL, яке використовується для зберігання вхідних даних, параметрів моделей, результатів кластеризації та координат центрів кластерів у структурованому вигляді, зокрема із застосуванням формату JSONB, що забезпечує гнучкість представлення інтервальних та багатовимірних даних. Взаємодія між компонентами пакету, а також з іншими модулями інформаційної технології Clustering4Build, здійснюється через REST API, який забезпечує сервісно-орієнтовану архітектуру, технологічну незалежність компонентів і можливість масштабування системи. Така організація пакету прикладних програм забезпечує ефективну реалізацію процесів кластеризації в сучасних інформаційних системах оцінки технічного стану об'єктів будівництва та

створює основу для подальшого розширення функціональності Clustering4Build.

Пакет прикладних програм 2 «Оцінка якості класифікації та моніторинг технічного стану» є аналітично-інтерпретаційним компонентом інформаційної технології Clustering4Build, розробленої для підтримки процесів оцінювання технічного стану об'єктів будівництва в умовах невизначеності. Метою пакету є формалізована оцінка результатів класифікації, отриманих за допомогою нечітких і можливої кластеризації, а також формування інформаційної основи для моніторингу технічного стану будівельних об'єктів і підтримки прийняття інженерних рішень. Пакет забезпечує перехід від чисельних результатів кластеризації до аналітичних показників якості та інтерпретованих рекомендацій, придатних для використання в практичних системах технічного моніторингу.

Ключовим компонентом пакету є модуль QualityEvaluator, який реалізує обчислення показників якості кластеризації на основі геометричних та статистичних характеристик кластерів. Зокрема, модуль обчислює загальний гіпероб'єм H , що характеризує компактність кластерів, та індекс сферичності Sph , який дозволяє оцінити форму кластерів і ступінь їх відхилення від гіперсферичної геометрії. Оцінювання здійснюється на основі координат центрів кластерів, ковараційних матриць і розподілу об'єктів у просторі ознак, що забезпечує об'єктивну та відтворювану оцінку якості класифікації. Результати роботи QualityEvaluator використовуються як для порівняння альтернативних моделей кластеризації, так і як функція придатності в алгоритмах оптимізації параметрів.

Підсистема DSS (Decision Support System) виконує функції аналітичної інтерпретації та візуалізації результатів класифікації і оцінювання якості. DSS формує інтерактивні звіти, теплові карти технічного стану об'єктів будівництва, табличні представлення та узагальнені показники ризику, що дозволяє інженерам та експертам оперативно оцінювати поточний стан об'єктів і приймати обґрунтовані рішення щодо подальших заходів з експлуатації, обстеження або ремонту. Підсистема також може реалізовувати базові сценарії рекомендацій на основі кластерної належності об'єктів та

значень індексів якості, що підвищує практичну цінність інформаційної технології.

Для зберігання результатів класифікації, показників якості та історичних даних моніторингу у складі пакету використовується PostgreSQL, що забезпечує централізоване, надійне та масштабоване сховище даних. Обмін інформацією між компонентами пакету, а також інтеграція з іншими модулями Clustering4Build, здійснюється через REST API, що забезпечує сервісно-орієнтовану архітектуру, уніфіковані інтерфейси доступу до даних і можливість розширення функціональності системи. Таким чином, пакет прикладних програм «Оцінка якості класифікації та моніторинг технічного стану» забезпечує завершальний аналітичний етап обробки даних у Clustering4Build та є ключовим елементом інтеграції математичних моделей кластеризації з практичними задачами технічного моніторингу об'єктів будівництва.

Пакет прикладних програм 3 «Адаптивне налаштування параметрів класифікації та оптимізація» є спеціалізованим оптимізаційним компонентом інформаційної технології Clustering4Build, призначеним для динамічного підвищення якості класифікації технічного стану об'єктів будівництва в умовах різномірних, неоднорідно розподілених та зашумлених даних. Розробка та функціонування пакету реалізує інтелектуальні методи оптимізації параметрів алгоритмів обробки даних у складі розподіленої інформаційної системи. Основною метою пакету є автоматизований пошук та адаптивне коригування параметрів класифікації, що забезпечує стійкість і відтворюваність результатів кластеризації для наборів даних з істотно різною щільністю, об'ємом та структурою.

Центральним компонентом пакету є модуль GA-Optimizer, який реалізує генетичний алгоритм для пошуку оптимальних значень фазифікаторів m_1 та m_2 , а також пов'язаних параметрів класифікаційних моделей. Модуль здійснює еволюційний пошук у багатовимірному просторі параметрів, використовуючи показники якості кластеризації як функцію придатності, що дозволяє адаптувати параметри класифікації до конкретних характеристик вхідних даних. Такий підхід усуває необхідність ручного підбору параметрів

та знижує ризик отримання локально оптимальних, але глобально неадекватних рішень, особливо у випадках, коли множина об'єктів містить кластери суттєво різного об'єму або щільності.

Важливою складовою пакету є модуль TypeReducer, який виконує операцію зниження типу для інтервальних нечітких множин другого типу на основі алгоритму Карника–Менделя. TypeReducer забезпечує коректне перетворення інтервального нечіткого подання даних у числові інтервальні оцінки центрів кластерів, що використовуються у процесі оптимізації та подальшої кластеризації. Це дозволяє зберігати інформацію про невизначеність на проміжних етапах обробки даних і підвищує стійкість класифікаційної моделі до варіацій вхідних параметрів.

Взаємодія між компонентами пакету, а також інтеграція з іншими модулями інформаційної технології Clustering4Build, здійснюється через REST API, що забезпечує сервісно-орієнтовану архітектуру та стандартизований обмін даними. Через REST API передаються параметри оптимізації, результати кластеризації, інтервальні оцінки та показники якості, що дозволяє реалізувати адаптивний цикл «кластеризація – оцінка якості – оптимізація параметрів». Таким чином, пакет прикладних програм «Адаптивне налаштування параметрів класифікації та оптимізація» виконує ключову роль у забезпеченні інтелектуальної адаптації та підвищення ефективності Clustering4Build при роботі з різномірними даними технічного моніторингу об'єктів будівництва.

Пакет прикладних програм 4 «Інтерфейс та візуалізація результатів класифікації» є користувацько-орієнтованим компонентом інформаційної технології Clustering4Build, призначеним для забезпечення ефективної взаємодії користувача з програмною системою та інтерпретації результатів класифікації технічного стану об'єктів будівництва. Пакет базується на принципах сервісно-орієнтованої архітектури, інтерактивної аналітики та підтримки прийняття рішень. Основною метою пакету є перетворення результатів математичної обробки і кластеризації у наочні, інтерпретовані та зручні для користувача візуальні представлення, що дозволяють оперативно

оцінювати технічний стан будівельних об'єктів та приймати обґрунтовані управлінські рішення.

Центральним елементом пакету є користувацький інтерфейс (User Interface, UI), який забезпечує введення вхідних даних, налаштування параметрів класифікації та доступ до результатів обчислень. UI реалізує інтерактивні засоби візуалізації, зокрема графіки розподілу об'єктів за кластерами, теплові карти технічного стану, динамічні діаграми зміни показників у часі та табличні представлення результатів. Такий підхід дозволяє користувачеві аналізувати як загальну структуру кластерів, так і детальні характеристики окремих об'єктів, що є важливим для задач технічного моніторингу та експертного оцінювання.

Підсистема DSS (Decision Support System) у складі пакету виконує функції аналітичної інтерпретації результатів класифікації та формування рекомендацій щодо подальших дій. DSS використовує дані про кластерну належність об'єктів, індекси якості кластеризації та історичну інформацію для генерації звітів, пріоритизації об'єктів за рівнем ризику та підтримки прийняття рішень щодо планування обстежень, ремонтних або експлуатаційних заходів. Інтеграція DSS з UI забезпечує відображення рекомендацій у зрозумілому та зручному для користувача вигляді.

Взаємодія між UI, DSS та іншими компонентами інформаційної технології Clustering4Build здійснюється через REST API, що забезпечує уніфікований доступ до даних і результатів класифікації, незалежність клієнтського інтерфейсу від реалізації обчислювальних модулів та можливість розширення функціональності системи. Завдяки такій архітектурі пакет «Інтерфейс та візуалізація результатів класифікації» забезпечує ефективне поєднання інтелектуальних методів кластеризації з практичними потребами користувачів, підвищуючи прикладну цінність інформаційної технології Clustering4Build у задачах оцінки технічного стану об'єктів будівництва.

Таблиця 4.3. Склад ППП для Clustering4Build (Розроблено автором)

Підкомплекс/Модуль	Метод/Алгоритм	Опис	Функціональність
FCM-Core	FCM-алгоритм (Fuzzy C-Means)	Основний компонент для класифікації даних з використанням нечітких множин першого типу та інтервальних нечітких множин другого типу (IHMT2).	Реалізація класифікації об'єктів з обчисленням функцій приналежності та координат центрів кластерів. Використовує параметри фазифікаторів для покращення точності класифікації.
GA-Optimizer Модуль для оптимізації параметрів класифікації за допомогою генетичних алгоритмів.	Генетичний алгоритм (GA)	Оптимізація параметрів класифікації, таких як фазифікатори, для досягнення кращих результатів класифікації.	Пошук оптимальних значень фазифікаторів за допомогою генетичного алгоритму для покращення точності класифікації.
TypeReducer Модуль для попередньої обробки даних перед класифікацією.	Алгоритм Карника-Менделя	Зниження типу інтервальних нечітких множин другого типу для зменшення впливу помилок і шуму в даних.	Виконання зниження типу для інтервальних нечітких множин, що дозволяє адаптувати класифікацію до різних даних, покращуючи точність класифікації при високій невизначеності.
QualityEvaluator Модуль для оцінки якості класифікації.	Індекс загального гіпероб'єму (H), Індекс сферичності (Sph)	Оцінка якості класифікації за допомогою геометричних індексів, що визначають компактність і щільність кластерів.	Оцінка якості класифікації на основі геометричних характеристик кластерів, таких як загальний гіпероб'єм (H) та індекс сферичності (Sph).
DSS (Decision Support System)	Інтерфейс для прийняття рішень	Підсистема для надання рекомендацій на основі результатів	Генерація звітів і рекомендацій, створення теплових карт для візуалізації технічного стану об'єктів.

Підкомплекс/Модуль	Метод/Алгоритм	Опис	Функціональність
Підсистема для прийняття рішень на основі класифікації.		класифікації та кластеризації для подальших дій, таких як планування робіт з ремонту або оцінка ризиків.	Забезпечує підтримку прийняття рішень на основі отриманих даних класифікації.
PostgreSQL База даних для зберігання результатів класифікації та метаданих.	База даних для зберігання даних	Використовується для зберігання вхідних даних, результатів класифікації та інтервальних значень функцій приналежності, а також для збереження історії класифікації.	Зберігання даних класифікації, результатів кластеризації, центрів кластерів, функцій приналежності, метаданих для подальшої обробки та аналізу.
REST API Інтерфейс для взаємодії між компонентами.	API для взаємодії компонентів	Забезпечує взаємодію між мікросервісами системи через API для передачі даних між компонентами та інтеграції з іншими системами.	Інтеграція компонентів системи, забезпечення комунікації між мікросервісами, передача даних між різними модулями і зовнішніми системами для отримання результатів класифікації та звітів.

Таблиця описує склад пакету прикладних програм для системи Clustering4Build, що включає основні компоненти та модулі, кожен з яких виконує специфічні функції для класифікації та кластеризації технічного стану об'єктів будівництва в умовах невизначеності. Основні модулі, такі як FCM-Core, GA-Optimizer, TypeReducer, QualityEvaluator, DSS, забезпечують виконання ключових завдань системи, включаючи класифікацію за допомогою нечітких множин, оптимізацію параметрів за допомогою генетичних алгоритмів, оцінку якості класифікації та прийняття рішень на основі результатів. PostgreSQL служить для зберігання даних та результатів класифікації, а REST API забезпечує ефективну інтеграцію компонентів та взаємодію з іншими системами. Ці компоненти разом створюють гнучку,

адаптивну систему, що дозволяє здійснювати автоматичний моніторинг технічного стану об'єктів будівництва з високою точністю та в реальному часі.

У таблиці 4.3. наведені назви підкомплексів програм, що входять до складу інформаційної технології Clustering4Build, та опис методів кластеризації, які вони реалізують. Пакет прикладних програм може використовуватись як самостійно, так і в комплексі (послідовно або в будь-якому порядку). Користувач може обирати метод кластеризації залежно від специфіки розв'язуваної задачі, що дозволяє адаптувати систему до різних умов класифікації та кластеризації технічного стану об'єктів будівництва. На рисунку 4.4. представлена загальна структура пакету прикладних програм «Кластеризація технічного стану об'єктів будівництва в умовах невизначеності».

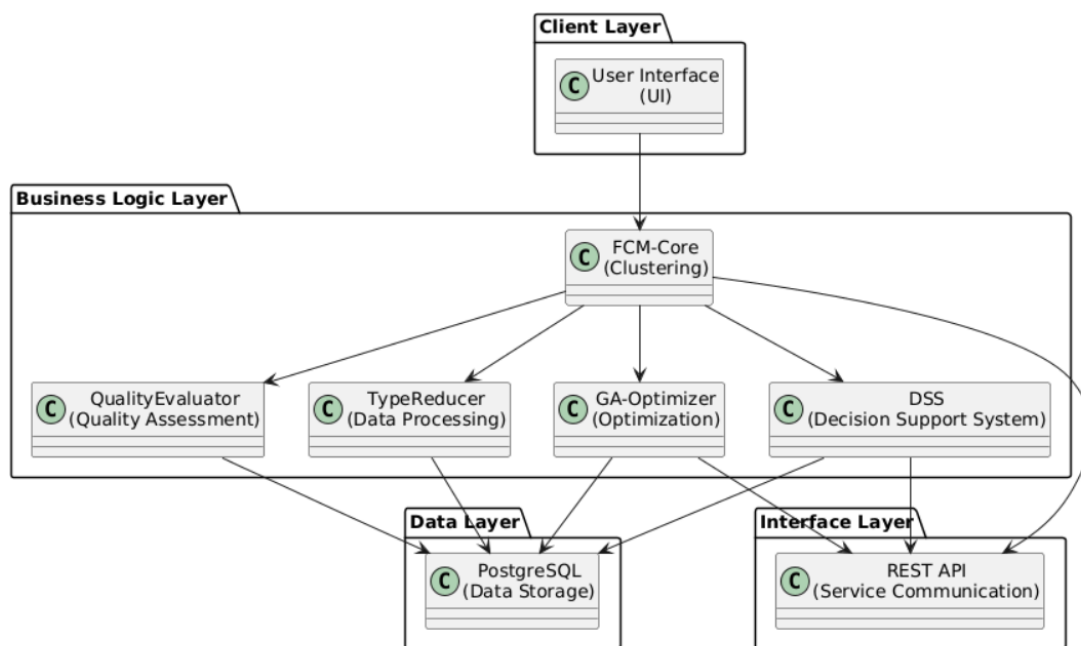


Рис.4.4. Архітектура пакету прикладних програм (ППП) для Clustering4Build (Розроблено автором)

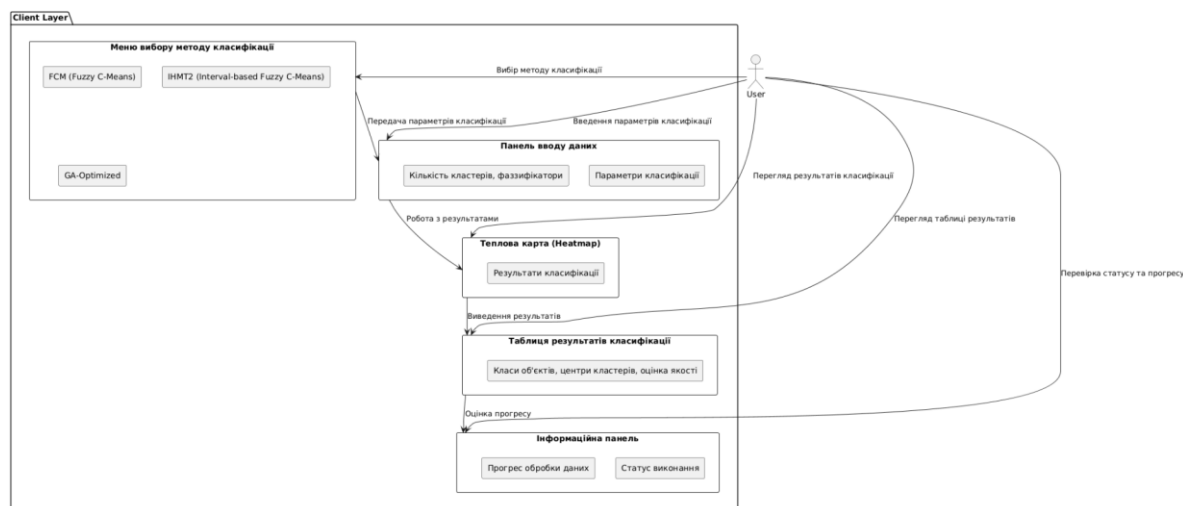


Рис. 4.5. Структурно-функціональна схема взаємодії користувача з інтерфейсом інформаційної технології Clustering4Build (Розроблено автором)

Результатом роботи кожного методу кластеризації є розподіл об'єктів на певну кількість кластерів, яка може бути задана користувачем або визначена під час процесу кластеризації. Також система виводить оцінку якості кластеризації на основі обраного показника, що дозволяє оцінити ефективність розподілу об'єктів. При достатній кількості елементів моніторингу передбачена візуалізація результатів класифікації для зручності користувача.

4.4. Загальна структура пакету прикладних програм «Кластеризація технічного стану об'єктів будівництва в умовах невизначеності»

Модуль FCM-Core є центральною обчислювальною компонентою програмного комплексу «Кластеризація технічного стану об'єктів будівництва в умовах невизначеності» та виконує ключові алгоритмічні операції інформаційної технології Clustering4Build. Архітектура модуля побудована за принципами модульності та сервісної орієнтованості, що забезпечує його інтеграцію з іншими підсистемами через REST API та підтримку потокової обробки даних технічного моніторингу. Модуль реалізує процес нечіткої кластеризації на основі алгоритму Fuzzy C-Means у двох варіаціях—для нечітких множин першого типу (HMT1) та інтервальних нечітких множин другого типу (IHMT2), що дозволяє коректно працювати з даними, які містять

високий рівень невизначеності та варіабельності, характерний для технічного стану будівельних конструкцій.

FCM-Core виконує розрахунок функцій приналежності (T1-FS) або інтервальних приналежностей (IT2-FS), здійснює ітераційне оновлення центрів кластерів на основі мінімізації відповідної цільової функції та контролює збіжність кластеризації за узгодженими критеріями. Для роботи з ІНМТ2 модуль використовує результати фаззифікації з FuzzificationEngine та алгоритм зниження типу Карника–Менделя, реалізований у TypeReducer, що дозволяє отримати інтервальний центроїд кластера навіть за умов значної інформаційної невизначеності. Крім того, модуль підтримує адаптивне оновлення параметрів фаззифікації m , m_1 , m_2 , які можуть бути задані користувачем або оптимізовані за допомогою GA-Optimizer. Для оцінки точності класифікації FCM-Core взаємодіє з QualityEvaluator, передаючи йому структуровані результати кластеризації у вигляді центрів кластерів, матриць приналежності, ковараційних характеристик та просторових метрик.

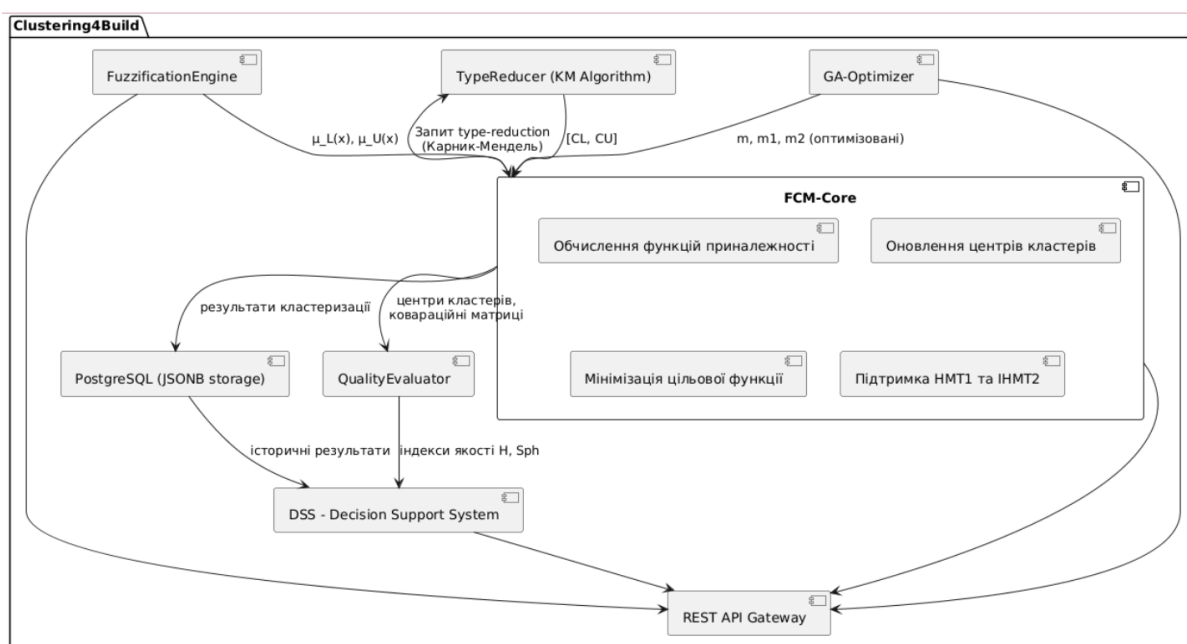


Рис.4.6. UML-компонент діаграма модуля FCM-Core (Розроблено автором)

Усі результати обчислень FCM-Core зберігаються у форматі JSONB у базі даних PostgreSQL, що забезпечує незалежність обчислювального ядра від підсистеми зберігання та спрощує інтеграцію з DSS для подальшої аналітики, візуалізації та прийняття рішень щодо технічного стану об'єктів будівництва.

Таким чином, модуль FCM-Core виступає фундаментальною частиною інформаційної технології Clustering4Build, забезпечуючи відтворювану, масштабовану та формально обґрунтовану процедуру кластеризації технічних параметрів у вигляді сервісу

Модуль GA-Optimizer є спеціалізованою обчислювальною підсистемою інформаційної технології Clustering4Build, розробленою для автоматизованого підбору оптимальних параметрів кластеризації з використанням методів еволюційної оптимізації. Модуль реалізує сучасні інтелектуальні алгоритми пошуку та адаптивного налаштування параметрів нечітких кластеризаторів у складі комплексної інформаційної системи. GA-Optimizer виконує пошук найкращих значень фазифікаторів m , m_1 , m_2 , а також параметрів «ширини зони» η для можливої кластеризації на основі ІНМТ2. Параметри, що оптимізуються, відзначаються складною нелінійною залежністю від структури даних, тому прямий аналітичний підбір є неможливим — і саме застосування генетичного алгоритму забезпечує знаходження глобальних оптимумів у багатовимірному просторі параметрів.

Архітектурно GA-Optimizer функціонує як окремий мікросервіс, який взаємодіє з FCM-Core та QualityEvaluator через REST API. Модуль генерує початкову популяцію хромосом, де кожна хромосома є набором параметрів класифікації; виконує операції відбору, кросинговеру та мутації; а також контролює критерії завершення еволюційного циклу. Для кожної хромосоми GA-Optimizer ініціює запуск кластеризації в FCM-Core, отримує результати обчислення центрів кластерів та матриць приналежності й передає їх у QualityEvaluator для оцінювання індексів H , Sph чи інших показників якості, які використовуються як функція придатності.

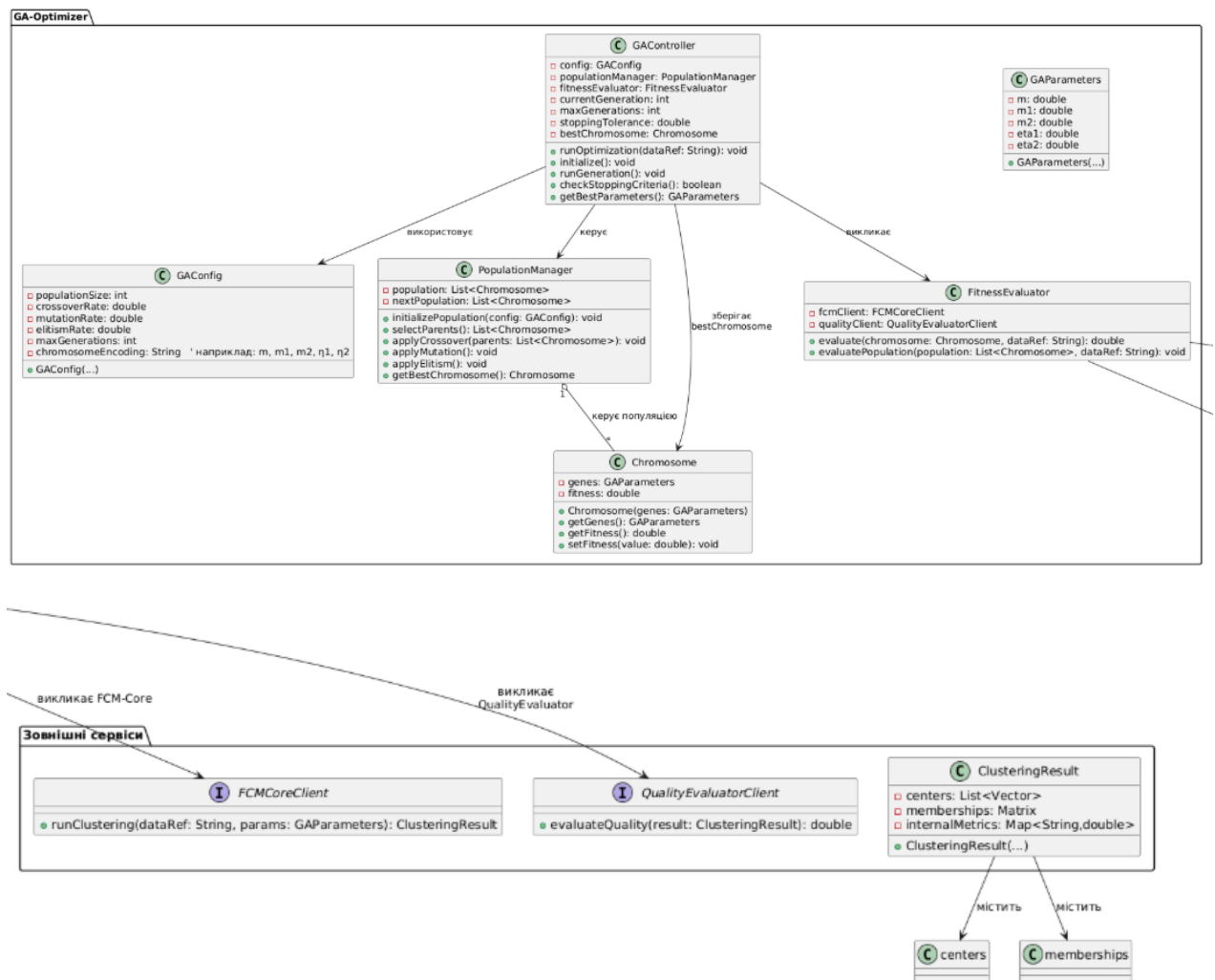


Рис. 4.7. Діаграма класу модуля GA-Optimizer (Розроблено автором)

Завдяки цьому модуль забезпечує автоматизоване вдосконалення процесу кластеризації, що суттєво підвищує точність класифікації технічного стану об'єктів будівництва в умовах невизначеності.

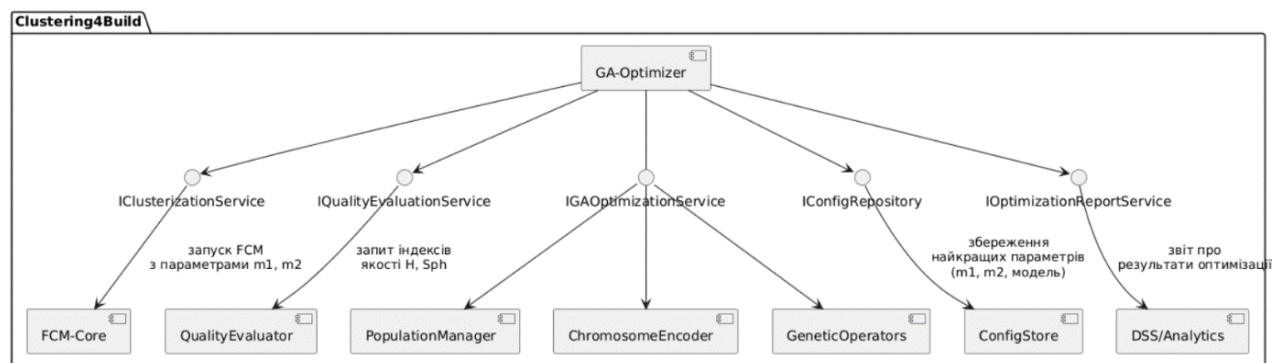


Рис.4.7.а. Архітектура GA-Optimizer у Clustering4Build (Розроблено автором)

У межах інформаційної технології Clustering4Build пошук оптимальних фазифікаторів m_1 та m_2 реалізується у вигляді окремого мікросервісу GA-Optimizer, який взаємодіє з модулем FCM-Core та підсистемою оцінювання

якості QualityEvaluator через REST API. Це забезпечує автоматизований вибір параметрів кластеризації в інформаційній системі без участі користувача.

Програмна реалізація ГА-оптимізації в Clustering4Build виконується засобами Python з використанням векторизованих обчислень (NumPy) та контейнеризації (Docker), а структури популяцій і хромосом зберігаються у вигляді JSON-об'єктів у сховищі конфігурацій. Кожна ітерація ГА супроводжується викликом сервісу FCM-Core для переобчислення центрів кластерів і QualityEvaluator для оновлення значень функції придатності.

В цьому випадку ГА має наступний вигляд ж.

Результати оптимізації — найкращі набори параметрів кластеризації — передаються назад у FCM-Core для фінального запуску кластеризації, а також у DSS для подальшої візуалізації та формування рекомендацій. GA-Optimizer підтримує масштабовану паралельну обробку хромосом, що дозволяє використовувати його для великих масивів даних технічного моніторингу та для оперативного оновлення параметрів у потокових режимах. Завдяки застосуванню еволюційного підходу модуль забезпечує високу стійкість до локальних мінімумів, адаптивність до різних типів будівельних об'єктів та формує оптимальні класифікаційні моделі, що відповідають задачам оцінювання технічного стану в реальних умовах експлуатації.

Модуль TypeReducer є спеціалізованою обчислювальною компонентою інформаційної технології Clustering4Build, призначеною для реалізації операцій «зниження типу» (type-reduction) для інтервальних нечітких множин другого типу (ІНМТ2) в задачах кластеризації технічного стану об'єктів будівництва.

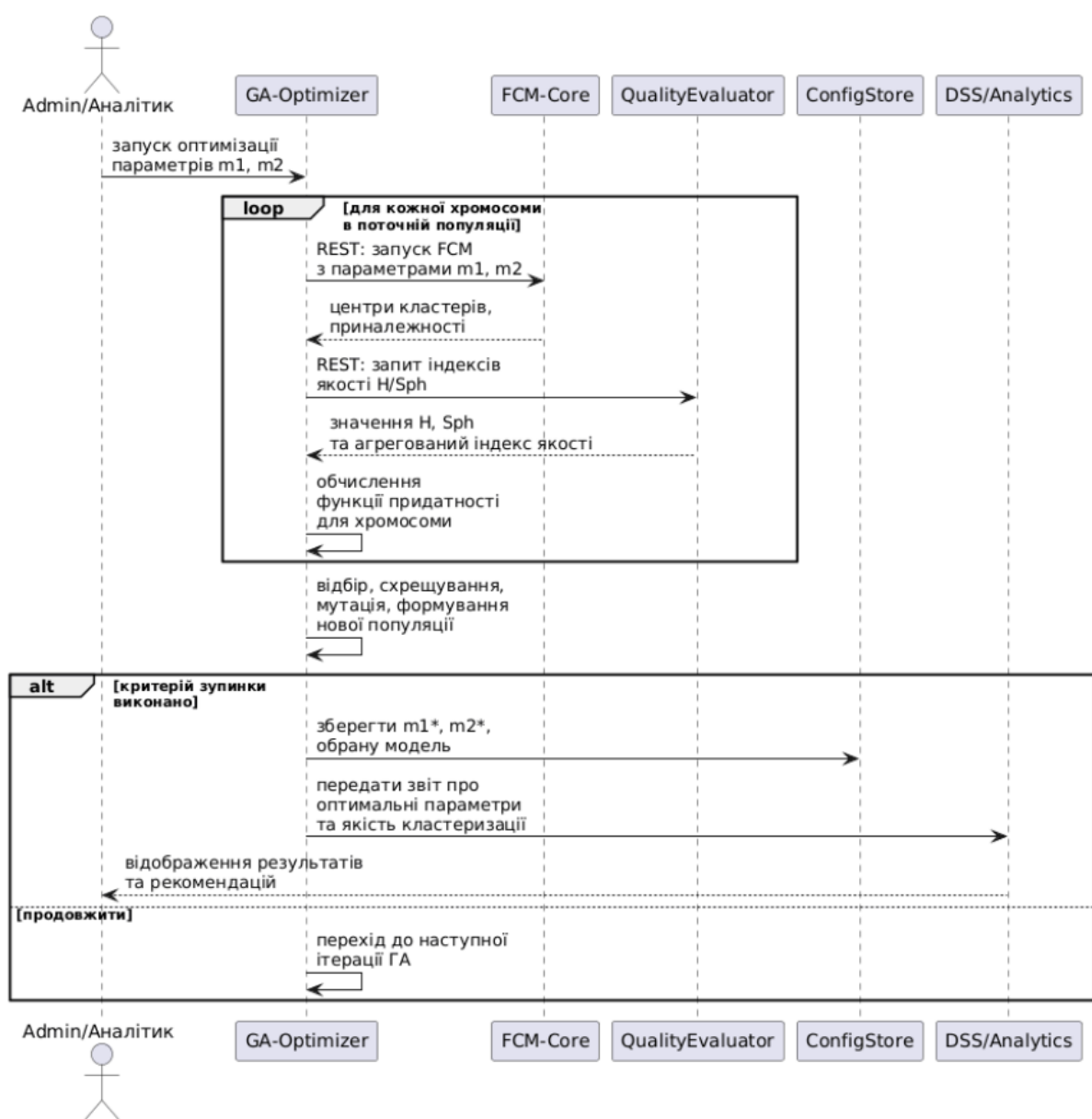


Рис.4.7.6 Взаємодія GA-Optimizer ↔ FCM-Core ↔ QualityEvaluator (Розроблено автором)

Модуль виступає як окремий сервіс обробки нечітких даних, який забезпечує перехід від інтервального представлення невизначеності (FOU – footprint of uncertainty) до інтервального центроїда кластера, що може бути надалі дефазифікований та використаний у FCM-Core і PCM/PCM-алгоритмах. TypeReducer реалізує ітераційний алгоритм Карника–Менделя, адаптований до форматів даних, що використовуються в Clustering4Build, та працює з верхніми та нижніми функціями приналежності, отриманими на етапі фазифікації, забезпечуючи стабільне отримання інтервальних оцінок центрів кластерів навіть за умов істотної варіабельності вимірюваних параметрів.

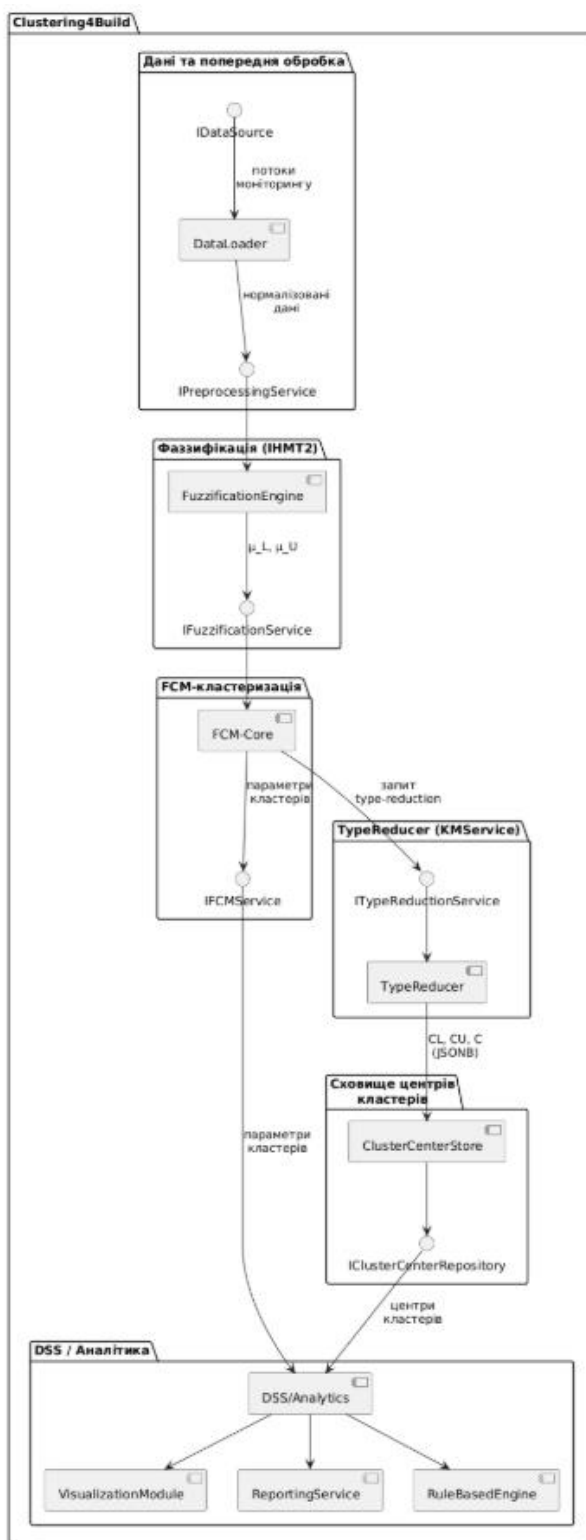


Рис.4.7.в. UML-компонентна діаграма архітектури реалізації алгоритму Карніка–Менделя *(Розроблено автором)*

Алгоритм Карніка–Менделя забезпечує стійке та інтервально узгоджене обчислення центрів кластерів у FCM-алгоритмах на основі IHMT2 та є критично важливим елементом для інформаційних технологій кластеризації технічного стану об'єктів.

Функціонально модуль отримує від FCM-Core (та/або FuzzificationEngine) дискретизовані значення $\mu_L(x)$ і $\mu_U(x)$ для кожного об'єкта кластера, формує відповідну інтервальну нечітку множину другого типу та виконує процедуру type-reduction для обчислення лівої та правої границь інтервального центроїда CL, CU . Отримані інтервальні центри кластерів використовуються FCM-Core при оновленні центрів у IHMT2-моделі та передаються до QualityEvaluator для врахування при розрахунку індексів якості кластеризації.

Activity Diagram для модуля TypeReducer подана в додатку 9. Застосування модуля TypeReducer дозволяє зберігати стійкість кластеризації при значних варіаціях вхідних даних, наявності шуму, експертної невизначеності або часткової відсутності вимірів, оскільки інтервальний підхід не змушує систему до жорсткого «точкового» представлення центру кластера, а працює з діапазоном можливих значень. Завдяки цьому TypeReducer забезпечує формально обґрунтований перехід від IHMT2-подання даних до чисельних параметрів кластеризаційних моделей, інтегрованих у Clustering4Build, і є ключовим елементом інтелектуальної обробки нечіткої інформації в програмному забезпеченні для оцінки технічного стану об'єктів будівництва.

Модуль QualityEvaluator є аналітичною підсистемою інформаційної технології Clustering4Build, призначеною для формального оцінювання якості кластеризації технічного стану об'єктів будівництва.

Модуль реалізує обчислення метрик, що характеризують геометрію та узгодженість кластерів у багатовимірному просторі технічних параметрів, забезпечуючи кількісну оцінку результатів роботи алгоритмів FCM, PCM та їх модифікацій на основі інтервальних нечітких множин другого типу (IHMT2). Основними показниками, що обчислюються в модулі, є загальний гіпероб'єм H , який описує компактність кластерів, та індекс сферичності Sph , що вимірює відхилення форми кластерів від гіперсферичної геометрії. Обидва індекси є критично важливими для оцінювання придатності класифікаційної моделі на технічних даних, у яких характерні суттєва варіабельність розподілу ознак, неоднорідність щільності та можливі міжкластерні перетинання.

QualityEvaluator функціонує як незалежний сервіс аналітичної обробки даних, який отримує від FCM-Core або PCM/PCM-модулів усі необхідні результати кластеризації: центри кластерів, ковараційні матриці, матриці функцій приналежності (у тому числі інтервальні приналежності для IHMT2) та метричні характеристики об'єктів. На основі цих даних модуль обчислює значення H і Sph , виконує їх нормування та агрегування, а також формує узагальнену числову оцінку якості кластеризації, що використовується іншими підсистемами. Зокрема, результати QualityEvaluator передаються у GA-Optimizer, де індекси H і Sph виступають як функція придатності в генетичному алгоритмі, а також у DSS, де вони використовуються для побудови інтерактивних звітів та візуалізацій якості класифікації.

Завдяки формальному, незалежному та відтворюваному підходу до оцінювання якості, модуль QualityEvaluator забезпечує контроль надійності процесів кластеризації, виявлення помилкових кластеризаційних структур, порівняння альтернативних моделей та оптимізацію параметрів кластеризаторів. Його інтеграція до загальної архітектури Clustering4Build підвищує точність, стійкість та інтерпретованість системи, що є необхідним для задач технічного моніторингу об'єктів будівництва в умовах невизначеності.

Послідовнісна діаграма відображає взаємодію трьох ключових модулів інформаційної технології Clustering4Build: GA-Optimizer, FCM-Core та QualityEvaluator у задачі оптимізації параметрів нечіткої кластеризації технічного стану об'єктів будівництва.

GA-Optimizer ініціалізується за запитом від користувача або підсистеми DSS, отримуючи посилання на дані та допустимі діапазони параметрів (фазифікатори m , m_1 , m_2 , параметри «ширини зони» η). Формується початкова популяція хромосом, де кожна хромосома відповідає певній комбінації параметрів FCM/PCM-на-основі-IHMT2.

Для кожної хромосоми GA-Optimizer викликає FCM-Core з поточними параметрами кластеризації. Модуль FCM-Core виконує ітераційний процес нечіткої кластеризації (оновлення функцій приналежності та центрів кластерів) до виконання критерію збіжності та повертає структуру результатів

кластеризації (координати центрів, матриці приналежностей, допоміжні метрики).

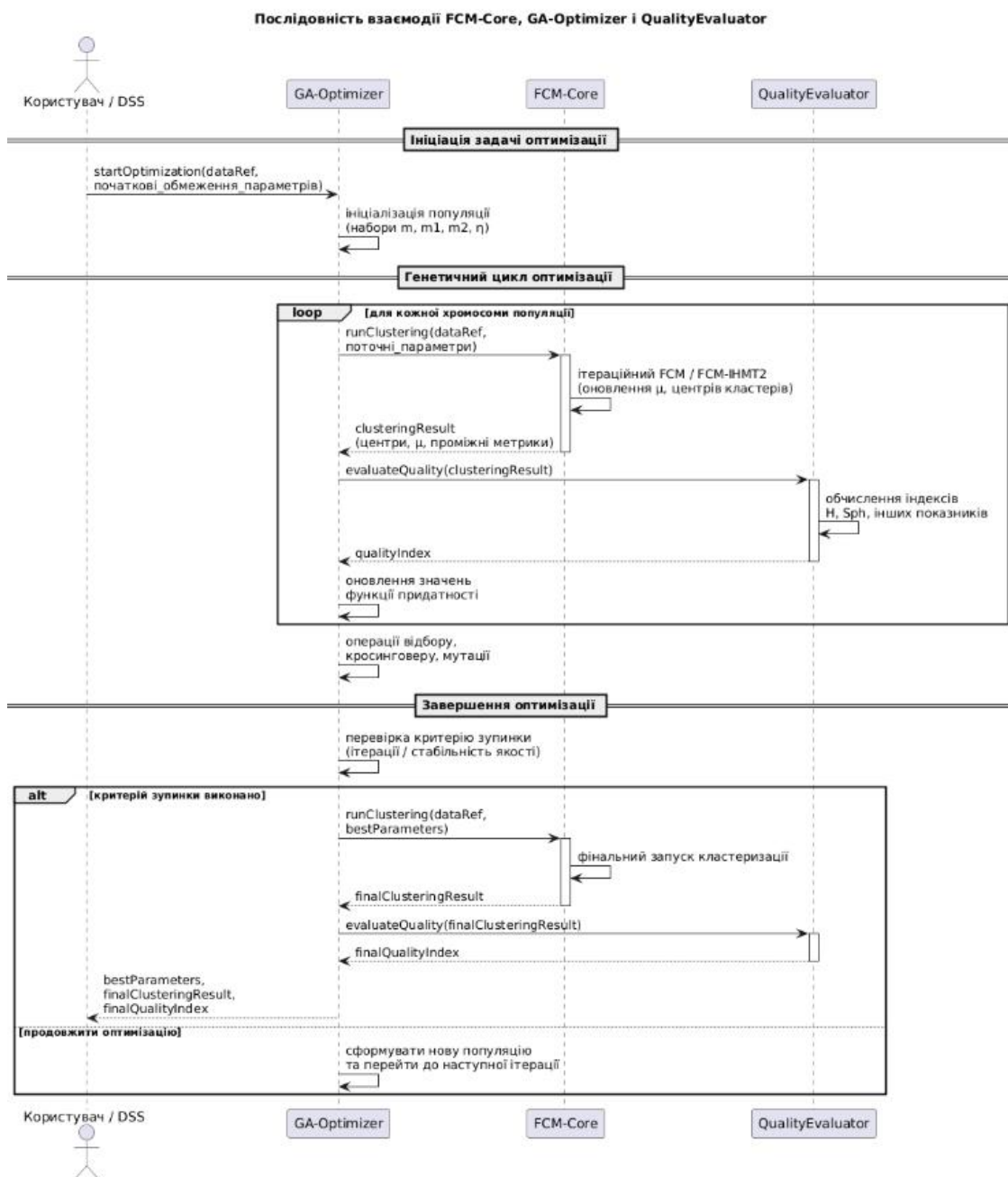


Рис.4.8. Взаємодію трьох ключових модулів інформаційної технології Clustering4Build: GA-Optimizer, FCM-Core та QualityEvaluator (Розроблено автором)

На основі отриманих результатів GA-Optimizer викликає QualityEvaluator, який обчислює індекси якості кластеризації (зокрема, загальний гіпероб'єм H та індекс сферичності Sph) й повертає скалярний показник якості, що використовується як функція придатності в генетичному алгоритмі. Activity Diagram для модуля QualityEvaluator подано в додатку 10.

GA-Optimizer оновлює значення функції придатності для популяції, виконує стандартні операції генетичного алгоритму (відбір, кросинговер, мутація) та повторює цикл, доки не буде досягнутий критерій зупинки (обмеження за кількістю ітерацій або стабілізація показників якості).

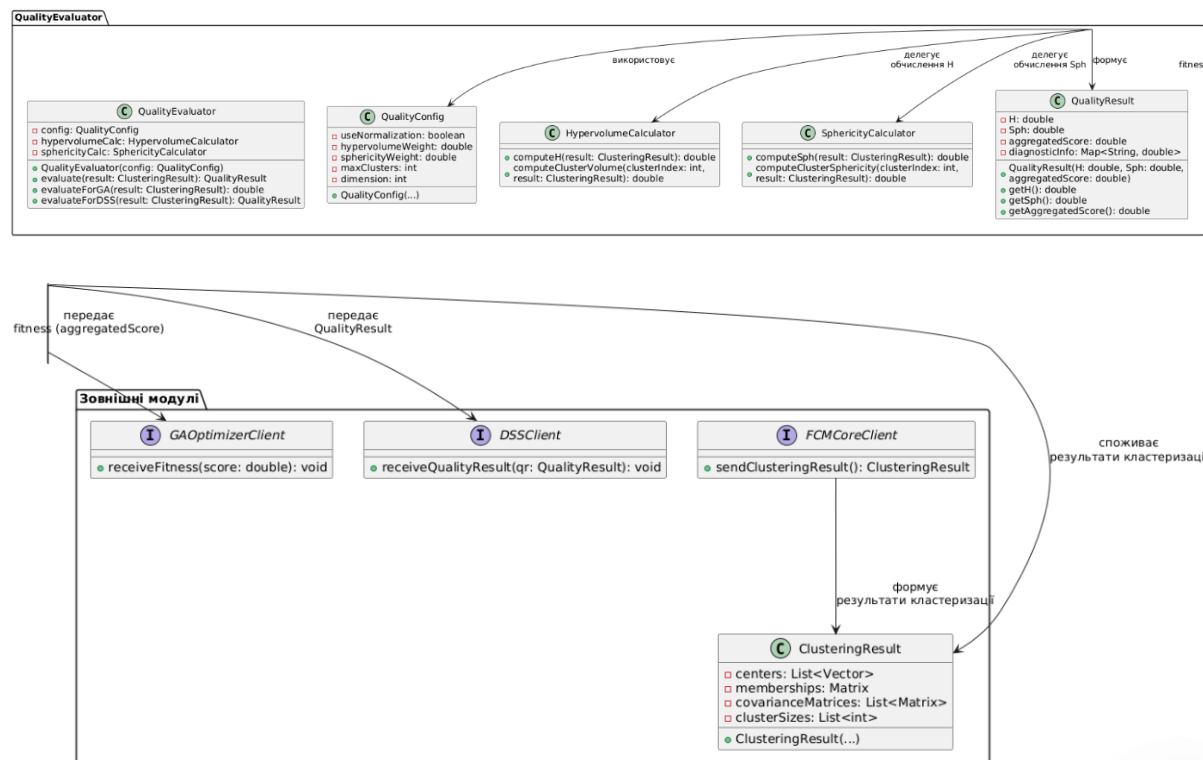


Рис.4.9. Class Diagram модуля QualityEvaluator (Розроблено автором)

Після завершення оптимізації GA-Optimizer ще раз викликає FCM-Core з найкращими знайденими параметрами, а результати фінальної кластеризації додатково оцінюються QualityEvaluator. Підсумкові параметри, результати кластеризації та індекси якості повертаються користувачу або підсистемі DSS для подальшої візуалізації та підтримки прийняття рішень.

Модуль DSS (Decision Support System) є інтеграційно-аналітичною підсистемою інформаційної технології Clustering4Build, призначеною для підтримки прийняття рішень на основі результатів кластеризації технічного стану об'єктів будівництва в умовах невизначеності. DSS реалізує комплекс функцій візуалізації, інтерпретації та аналітичної обробки даних, отриманих від модулів FCM-Core, GA-Optimizer, QualityEvaluator та підсистеми зберігання даних (PostgreSQL).

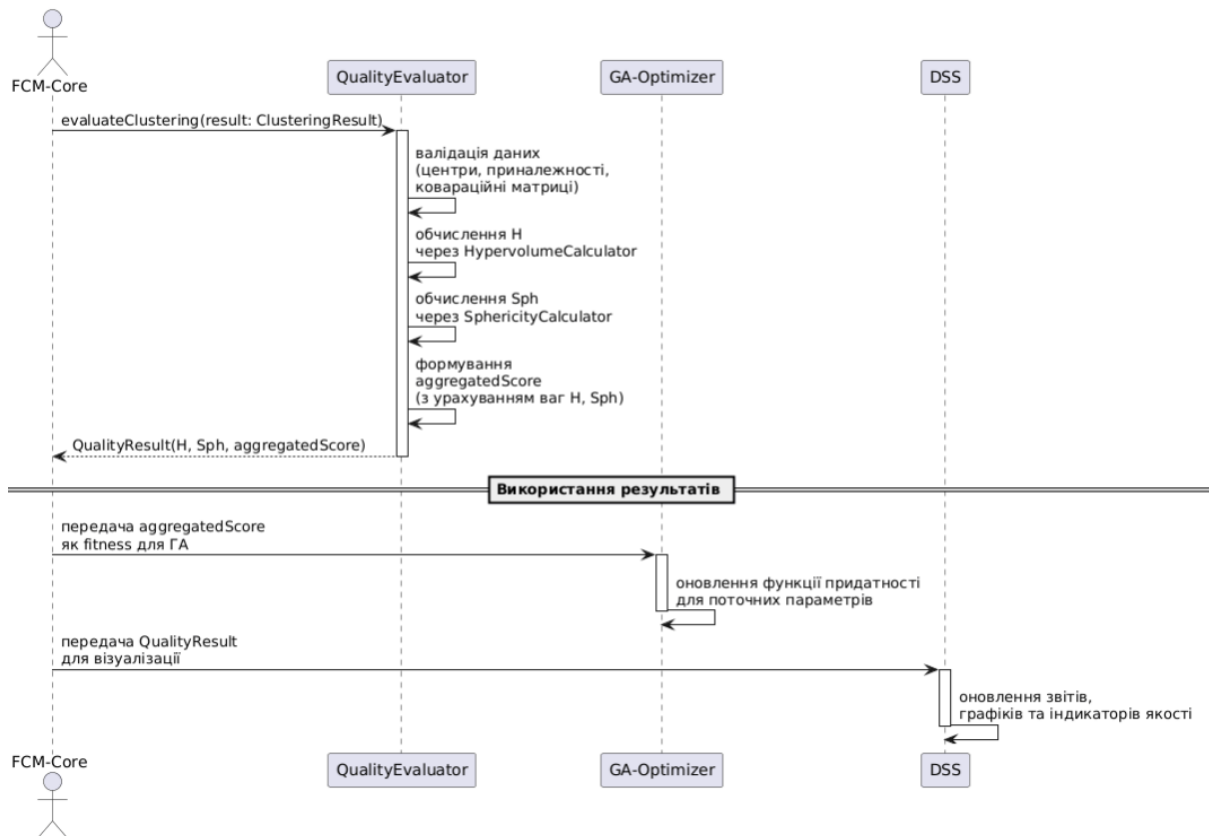


Рис.4.10. Компонентна діаграма взаємодії модулів. (Розроблено автором)

Основним призначенням модуля є перетворення «сирих» результатів кластеризації (центри кластерів, функції приналежності, індекси якості H та Sph , параметри оптимізації) у зрозумілі для інженера-експерта аналітичні представлення: теплові карти технічного стану, інтерактивні діаграми, табличні звіти, ранжовані списки об'єктів за ступенем ризику, а також сценарні та прогнозні оцінки.

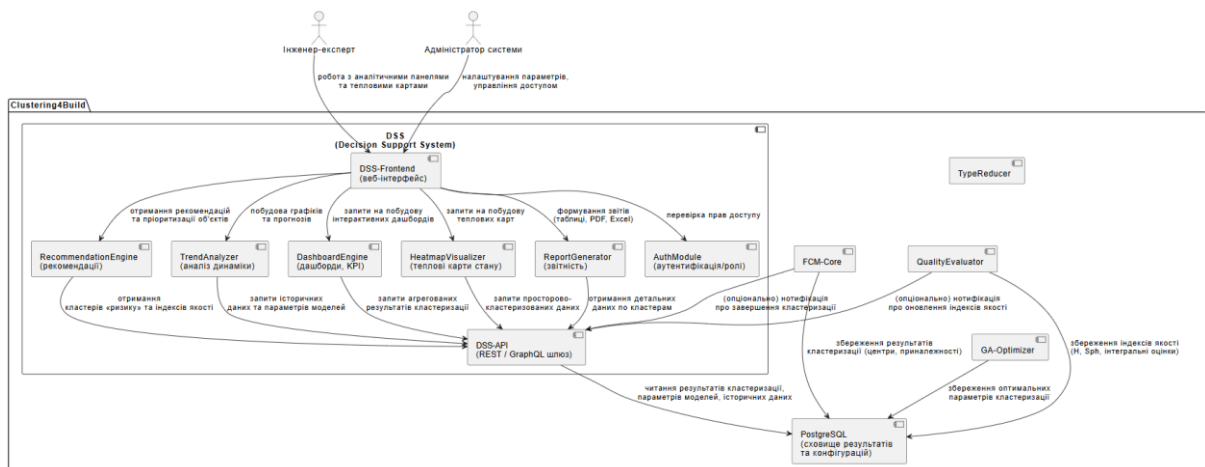


Рис.4.11. Компонентна діаграма DSS (Розроблено автором)

Архітектурно модуль DSS реалізований як окремий сервіс, що взаємодіє з іншими компонентами Clustering4Build через REST API та працює поверх єдиної бази даних, де результати кластеризації та оптимізації зберігаються у структурованому вигляді (зокрема, у форматі JSONB). DSS отримує від QualityEvaluator формалізовані показники якості кластеризації, від FCM-Core – структуровані дані про кластери (центри, розподіл об’єктів, інтервальні оцінки для IHMT2), від GA-Optimizer – інформацію про оптимальні параметри моделей. На основі цих даних модуль формує візуальні аналітичні панелі (дашборди), теплові карти зон підвищеного ризику, графіки динаміки технічного стану об’єктів у часі, а також генерує звіти у форматах, придатних для включення в проектну та експлуатаційну документацію. Крім суто візуальної складової, DSS реалізує базові правила та сценарії підтримки рішень (наприклад, виявлення об’єктів, що потрапляють у «критичні» кластери, пропозиція черговості проведення ремонтних заходів), що дозволяє інтегрувати результати кластеризації у реальні процеси управління технічним станом будівель. Таким чином, модуль DSS забезпечує зв’язок між математичними та програмними компонентами Clustering4Build і практичними задачами експлуатації об’єктів будівництва, виконуючи роль фронт-офісного аналітичного інтерфейсу інформаційної технології.

Компонентна діаграма модуля DSS (Decision Support System) відображає його архітектурну організацію в інформаційній технології Clustering4Build, де модуль виконує функції інтерактивної аналітики, візуалізації та підтримки прийняття рішень на основі результатів кластеризації технічного стану об’єктів будівництва. У структурі DSS виділено спеціалізовані підсистеми, зокрема DashboardEngine, HeatmapVisualizer, ReportGenerator, RecommendationEngine та TrendAnalyzer, що забезпечують формування дашбордів, теплових карт, звітів, рекомендацій та прогнозів відповідно. Дані для роботи DSS надходять із підсистем FCM-Core, GA-Optimizer та QualityEvaluator, результати яких зберігаються у PostgreSQL і отримуються через інтерфейс DSS-API. Модуль також містить компонент AuthModule, який відповідає за автентифікацію та контроль доступу користувачів до аналітичних інструментів. Таким чином, DSS функціонує як

інтегрований фронт-офіс аналітичної системи, що забезпечує візуалізацію, інтерпретацію та подальше використання результатів кластеризації у процесах моніторингу й управління технічним станом об'єктів будівництва.

Модуль FuzzificationEngine є складовою обчислювального контуру інформаційної технології Clustering4Build, що забезпечує побудову формального нечіткого подання вхідних даних технічного стану об'єктів будівництва. класифікації. FuzzificationEngine формує для кожного об'єкта нижню $\mu L(x)$ та верхню $\mu U(x)$ функції приналежності, що складають відбиток невизначеності (FOU) і дозволяють моделювати параметричну та експертну невизначеність, властиву технічним вимірюванням. Формування $\mu L(x)$ та $\mu U(x)$ відбувається з урахуванням статистичних характеристик даних, шумових компонентів та можливого діапазону параметрів фазифікаторів (m_1 , m_2), що забезпечує їх адаптивність до різного рівня варіабельності датчиків та експлуатаційних умов. Activity Diagram модуля FuzzificationEngine подано в додатку 8.

Архітектурно модуль функціонує як окремий сервіс, що приймає вхідні дані від систем збору та моніторингу (або від завантажених датасетів), нормалізує їх та перетворює у нечітке представлення відповідно до заданих або оптимізованих параметрів. Результати фазифікації передаються через стандартизований інтерфейс до модулів FCM-Core та TypeReducer, де вони використовуються при побудові інтервальних центрів кластерів та оновленні матриць приналежності. Модуль підтримує як пакетну, так і потокову обробку даних, що дозволяє застосовувати Clustering4Build для реального часу — у системах моніторингу деформацій та фізичного зносу будівельних конструкцій.

FuzzificationEngine забезпечує формально обґрунтоване представлення невизначеності, що є необхідним етапом перед виконанням кластеризації на основі IHMT2, та виступає критичним компонентом системи, який інтегрує фізичні вимірювання, статистичні моделі й нечітку логіку в єдиний інформаційний процес.

Мікросервісний підхід в інформаційній технології Clustering4Build передбачає розподіл функціональності на автономні сервіси (FCM-Core, GA-

Optimizer, TypeReducer, FuzzificationEngine, QualityEvaluator, DSS), які розгортаються, масштабуються та оновлюються незалежно один від одного.

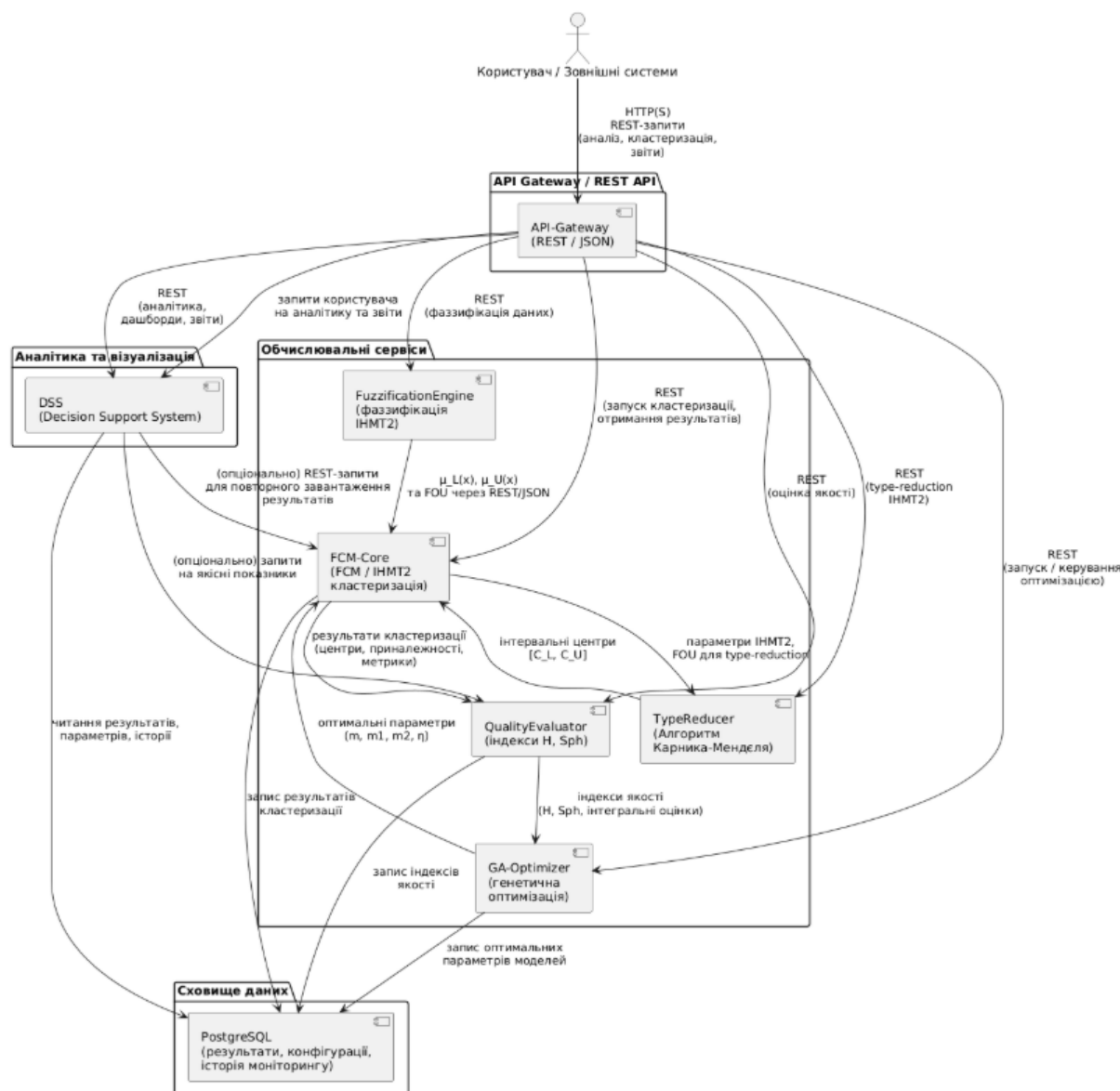


Рис. 4.12. Компонентна діаграма мікросервісної архітектури інформаційної технології Clustering4Build (Розроблено автором)

Наведена компонентна діаграма відображає мікросервісну архітектуру інформаційної технології Clustering4Build, у якій кожен функціональний модуль реалізований як незалежний сервіс, що взаємодіє з іншими компонентами через API Gateway та REST API. У структурі системи виділено обчислювальні сервіси (FuzzificationEngine, FCM-Core, TypeReducer, GA-Optimizer, QualityEvaluator), модуль аналітики та візуалізації DSS, а також окреме сховище даних PostgreSQL, у якому зберігаються результати кластеризації, параметри моделей та історичні дані моніторингу. API Gateway

служує єдиною точкою доступу для користувачів і зовнішніх систем, забезпечуючи маршрутизацію запитів між сервісами та уніфікований формат обміну даними. Взаємозв'язки між мікросервісами демонструють послідовність трансформації даних: від фазифікації та кластеризації до оцінювання якості та формування аналітичних звітів. Така архітектура забезпечує масштабованість, надійність та гнучкість системи, дозволяючи ефективно обробляти великі обсяги даних технічного моніторингу в умовах невизначеності.

Взаємодія між модулями здійснюється через REST API, що забезпечує єдині контракти обміну даними, технологічну незалежність сервісів та можливість інтеграції з зовнішніми системами моніторингу та управління технічним станом об'єктів будівництва. Такий підхід підвищує гнучкість архітектури, спрощує горизонтальне масштабування обчислювально навантажених компонентів (насамперед FCM-Core та GA-Optimizer) і дозволяє ефективно обробляти великі обсяги даних у розподіленому середовищі. REST API використовується як транспортний рівень для передачі результатів кластеризації, параметрів моделей, індексів якості та аналітичних звітів між сервісами та модулем DSS.

Підсистема зберігання даних у технології Clustering4Build реалізована на основі реляційної СУБД PostgreSQL, яка використовується для збереження результатів кластеризації, проміжних обчислень та конфігурацій параметрів алгоритмів. Усі структуровані дані — зокрема інтервальні значення функцій приналежності, координати центрів кластерів, параметри оптимізації та індекси якості — зберігаються у форматі JSONB, що забезпечує гнучкість схеми, швидкий доступ до вкладених структур і високу ефективність аналітичних операцій. Такий підхід дозволяє поєднати переваги реляційної моделі з можливостями обробки напівструктурованих даних, що особливо важливо при роботі з інтервальними IHMT2-представленнями та багатоваріантними вихідними параметрами кластеризації.

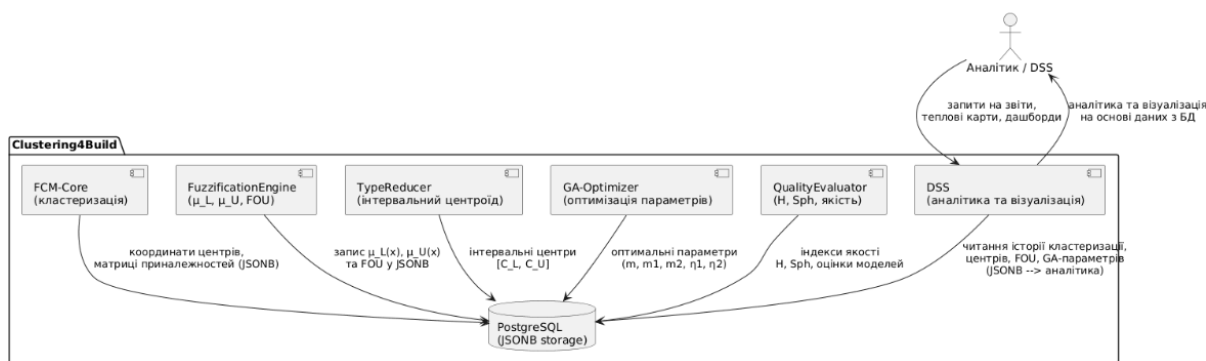


Рис.4.13. Component Diagram — Підсистема зберігання даних (PostgreSQL + JSONB) (Розроблено автором)

Діаграма відображає архітектуру підсистеми зберігання даних у Clustering4Build, де всі результати обчислень зберігаються в PostgreSQL у форматі JSONB, що дозволяє ефективно працювати з інтервальними значеннями, матрицями приналежностей та параметрами нечітких моделей. Мікросервіси FuzzificationEngine, FCM-Core, TypeReducer, GA-Optimizer та QualityEvaluator передають у БД структуровані результати, які надалі використовуються модулем DSS для формування візуалізацій і аналітичних звітів. DSS отримує дані через запити до PostgreSQL і надає користувачу зведену інформацію щодо технічного стану об'єктів будівництва та якості кластеризації. Така архітектура забезпечує узгодженість, масштабованість та гнучкість роботи з великими масивами даних в умовах невизначеності.

Сховище використовується всіма модулями системи через REST API та API Gateway, забезпечуючи узгоджений доступ до історії моніторингу, параметрів моделей та результатів аналітики.

Ця структура дозволяє максимально ефективно виконувати кластеризацію технічного стану об'єктів будівництва в умовах невизначеності, інтегруючи найсучасніші алгоритми, такі як генетичні алгоритми та методи інтервальних нечітких множин другого типу (IHMT2), забезпечуючи точність, адаптивність і швидкість системи.

За допомогою пакету прикладних програм Clustering4Build можна вирішувати такі основні задачі:

1. Багатокритеріальна кластеризація технічного стану об'єктів будівництва на основі числових, експертних та інструментальних даних з урахуванням їх невизначеності, варіабельності та неповноти.

2. Автоматизована класифікація об'єктів будівництва за рівнями технічного стану (нормальний, задовільний, передаварійний, аварійний тощо) з використанням нечітких множин першого типу та інтервальних нечітких множин другого типу.

3. Виявлення кластерів суттєво різного об'єму та щільності, що є характерним для реальних даних технічного моніторингу будівельних конструкцій і не може бути коректно оброблено класичними методами кластеризації.

4. Оцінювання якості результатів кластеризації шляхом обчислення формалізованих показників (загального гіпероб'єму H , індексу сферичності Sph та інших метрик), що дозволяє порівнювати альтернативні моделі кластеризації та обґрунтовано обирати оптимальні рішення.

5. Адаптивне налаштування параметрів класифікації (фаззифікаторів, ширини зони, параметрів ІНМТ2) із застосуванням генетичних алгоритмів, що забезпечує підвищення точності та стійкості кластеризації для різнорідних наборів даних.

6. Інтелектуальний аналіз та моніторинг динаміки технічного стану об'єктів будівництва, включаючи накопичення історичних даних та їх подальше використання для аналізу тенденцій і прогнозування.

7. Візуалізація результатів кластеризації у вигляді теплових карт, графіків, інтерактивних діаграм і аналітичних панелей, що забезпечує зручну інтерпретацію результатів для інженерів та експертів.

8. Формування аналітичних звітів і рекомендацій щодо подальших дій з експлуатації, обстеження або ремонту об'єктів будівництва на основі кластерної належності та показників якості.

9. Підтримка прийняття рішень у задачах управління технічним станом будівель і споруд шляхом інтеграції результатів кластеризації, оцінки якості та експертних правил у підсистемі DSS.

10. Інтеграція з іншими інформаційними системами технічного моніторингу через REST API, що дозволяє використовувати ППП як складову розподілених інформаційно-аналітичних систем.

Таким чином, ППП Clustering4Build забезпечує повний цикл інтелектуальної обробки даних — від формування нечіткого подання та кластеризації до оцінювання якості, візуалізації результатів і підтримки прийняття рішень у сфері оцінки технічного стану об'єктів будівництва.

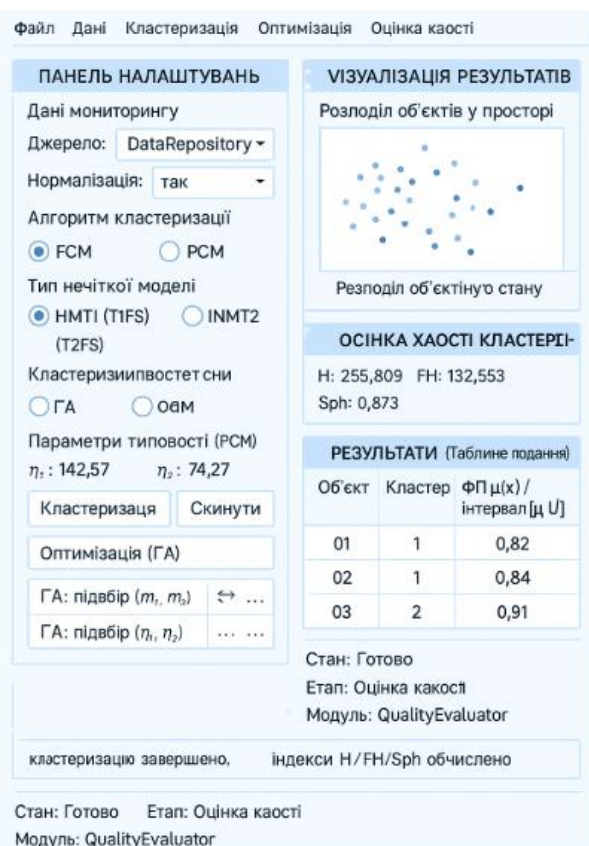


Рис. 4.14. Головне вікно пакету прикладних програм «Кластеризація технічного стану об'єктів будівництва в умовах невизначеності» (Розроблено автором)

Головне вікно пакету прикладних програм «Кластеризація технічного стану об'єктів будівництва в умовах невизначеності» реалізує єдине інтерактивне середовище для керування процесами кластеризації, аналізу результатів та підтримки прийняття рішень. Інтерфейс побудовано за принципом аналітичної панелі, що поєднує засоби введення і налаштування параметрів класифікації з модулями візуалізації та оцінювання якості кластеризації.

Панель налаштувань забезпечує вибір алгоритмів кластеризації та параметрів їх роботи, тоді як центральна область інтерфейсу призначена для відображення просторової структури кластерів і теплових карт технічного стану об'єктів. Аналітична частина інтерфейсу відображає показники якості

кластеризації та табличні результати, що використовуються підсистемою DSS для формування рекомендацій. Така організація інтерфейсу забезпечує зручну інтерпретацію результатів кластеризації та ефективну взаємодію користувача з інформаційною технологією Clustering4Build.

4.5 Програмна реалізація комплексу ПП як складових інформаційної технології Clustering4Build

Комплекс пакетів прикладних програм як складових інформаційної технології Clustering4Build, орієнтованої на автоматизоване оцінювання технічного стану об'єктів будівництва в умовах невизначеності на основі методів нечіткої та можливої кластеризації реалізовано за принципами мікросервісної архітектури, де кожен сервіс інкапсулює завершену функцію обробки даних (підготовка та фазифікація, кластеризація FCM/PCM на НМТ1 та ІНМТ2, зниження типу, генетична оптимізація параметрів, оцінка якості, візуалізація та формування звітів), а інтеграція компонентів здійснюється через REST API із передаванням структурованих даних у форматі JSON. Центральне сховище даних на базі PostgreSQL забезпечує фіксацію вхідних наборів, параметрів запусків, результатів кластеризації (центри, матриці приналежності/типовості, інтервальні ФП для ІНМТ2) та метрик якості H , FH , Sph , що гарантує відтворюваність обчислювальних експериментів і можливість аудиту розрахунків. Далі послідовно подано опис функціональних вікон користувацького інтерфейсу, які відображають ключові етапи роботи Clustering4Build: введення і підготовку даних, параметризацію фазифікації, запуск кластеризації, візуалізацію результатів, оцінювання якості, оптимізацію параметрів за допомогою GA-Optimizer та формування рекомендацій і звітів у підсистемі DSS.

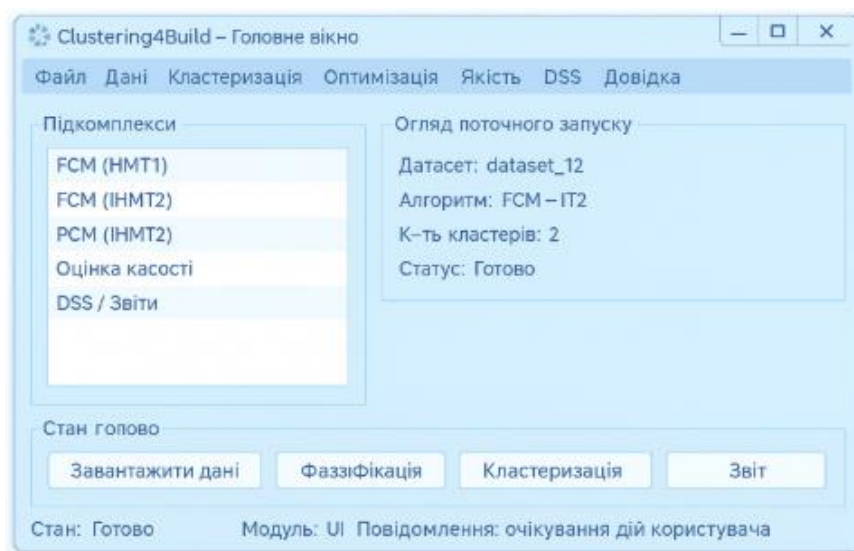


Рис. 4.15. Головне вікно інформаційної технології Clustering4Build (Розроблено автором)

Головне вікно забезпечує централізоване керування функціями інформаційної технології Clustering4Build та доступ до підкомплексів програм. Інтерфейс підтримує запуск основних етапів обробки даних (завантаження, фазифікація, кластеризація, оцінка якості, формування звітів) та відображає стан виконання обчислень.

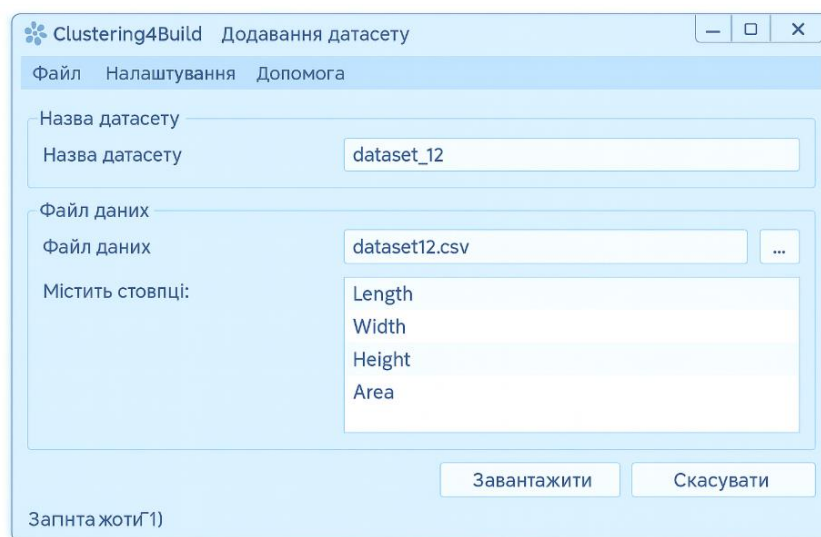


Рис. 4.16. Діалогове вікно введення та підготовки даних (Розроблено автором)

Така організація реалізує єдину точку взаємодії користувача з мікросервісними компонентами системи. Вікно призначене для завантаження

та попередньої підготовки даних моніторингу, включаючи валідацію та нормалізацію ознак. Результати перевірки якості даних і параметри нормалізації зберігаються у PostgreSQL, що забезпечує відтворюваність та аудит обчислювальних експериментів. Підготовлені дані передаються до наступних етапів кластеризаційного конвеєра через REST API.

Програмна реалізація Clustering4Build виконана у вигляді набору взаємодіючих мікросервісів, кожен з яких інкапсулює окрему функцію обробки даних (фазифікація, кластеризація, зниження типу, оптимізація, оцінка якості, візуалізація). Компоненти взаємодіють через REST API (HTTP/JSON), що забезпечує уніфікований обмін даними, незалежність технологій реалізації сервісів та можливість горизонтального масштабування. Результати обчислень зберігаються у PostgreSQL у структурі JSONB, що дозволяє зберігати як числові масиви (матриці приналежності), так і вкладені структури (інтервальні ФП, параметри запусків, метрики якості, протоколи оптимізації), забезпечуючи повторюваність експериментів та аудит розрахунків.

1) FuzzificationEngine (фазифікація та підготовка даних)

Endpoint POST /fuzzification/validate реалізує перевірку вхідних даних (повнота, діапазони, типи, пропуски, аномалії) та повертає протокол валідації; паралельно зберігає звіт у PostgreSQL.data_quality_log (jsonb) із прив'язкою до dataset_id.

Endpoint POST /fuzzification/normalize виконує нормалізацію/масштабування матриці ознак (залежно від конфігурації: min–max, z-score тощо), повертає dataset_ref на нормалізовані дані і записує параметри нормалізації у PostgreSQL.normalization_profiles (jsonb).

Endpoint POST /fuzzification/it2/membership будує для кожного об'єкта нижню $\mu_L(x)$ та верхню $\mu_U(x)$ функції приналежності відповідно до параметрів IHMT2 (m_1 , m_2 , тип функції приналежності, правила обмежень), формує FOU і повертає membership_ref; результат зберігається у PostgreSQL.it2_memberships (jsonb).

2) FCM-Core (кластеризація FCM на HMT1 та IHMT2)

Endpoint POST /clustering/fcm/t1/run реалізує запуск FCM на основі HMT1: приймає dataset_ref, кількість кластерів c , фазифікатор m , критерії зупинки; повертає run_id, а в PostgreSQL.clustering_runs (jsonb) фіксує центри кластерів V , матрицю ФП U , число ітерацій, критерій збіжності та службові метадані.

Endpoint POST /clustering/fcm/it2/run реалізує запуск FCM на основі IHMT2: приймає membership_ref(μ_L, μ_U, FOU), c , значення m_1, m_2 ; в процесі виконання викликає сервіс TypeReducer для обчислення інтервальних центрів і повертає run_id; у БД зберігає $U_interval, V_interval$ і параметри фазифікації.

Endpoint GET /clustering/run/{run_id} повертає повний результат запуску (центри, ФП/інтервали ФП, метрики збіжності) як структурований JSON для QualityEvaluator і DSS.

3) TypeReducer (зниження типу для IHMT2, алгоритм Карника–Менделя)

Endpoint POST /type-reduction/km реалізує ітераційний алгоритм Карника–Менделя: приймає структури μ_L/μ_U та значення ознак, обчислює інтервальний центроїд $[C_L, C_U]$ для кластера і повертає результат у форматі JSON.

Endpoint POST /type-reduction/centroids/batch виконує пакетний розрахунок інтервальних центрів для всіх кластерів поточного запуску FCM-IT2, повертає centroids_ref і записує результати в PostgreSQL.type_reduction_results (jsonb).

4) GA-Optimizer (генетична оптимізація параметрів кластеризації)

Endpoint POST /optimizer/ga/start ініціалізує оптимізаційний процес: приймає межі параметрів ($m, m_1, m_2, \eta_1, \eta_2$), розмір популяції, кількість поколінь, оператори ГА (відбір, кросинговер, мутація) та критерій/функцію відповідності; повертає ga_run_id і фіксує конфігурацію у PostgreSQL.ga_runs (jsonb).

Endpoint POST /optimizer/ga/evaluate реалізує оцінювання хромосоми як виклик конвеєра: запускає FCM/PCM у FCM-Core з параметрами хромосоми → отримує run_id → викликає QualityEvaluator для метрик H/Sph/FH → повертає значення fitness; журнали оцінювання зберігає у PostgreSQL.ga_population (jsonb).

Endpoint POST /optimizer/ga/iterate формує нове покоління (селекція, кросинговер, мутація, елітизм), повторює цикл оцінювання та забезпечує збіжність за умовою зупинки.

Endpoint GET /optimizer/ga/best/{ga_run_id} повертає найкращі параметри та пов'язані метрики якості і фіксує їх у PostgreSQL.model_configs (jsonb) як конфігурацію повторного застосування.

5) QualityEvaluator (оцінка якості кластеризації)

Endpoint POST /quality/evaluate приймає run_id або структуру результатів кластеризації і виконує обчислення індексів якості: H (загальний гіпероб'єм), FH (нечіткий гіпероб'єм) та Sph (індекс сферичності) відповідно до формалізму, поданого у розділі 3; повертає JSON {H, FH, Sph} та записує результат у PostgreSQL.quality_metrics (jsonb).

Endpoint GET /quality/metrics/{run_id} забезпечує повторний доступ до метрик якості для DSS та GA-Optimizer без повторного перерахунку.

У межах інформаційної технології Clustering4Build забезпечується автоматизований вибір оптимальної моделі кластеризації на основі інтервальних нечітких множин другого типу. З цією метою використовується спеціалізований мікросервіс ModelSelector, який здійснює порівняльний аналіз результатів роботи алгоритмів FCM2 та PCM2. Порівняння виконується за індексами якості кластеризації H та Sph, що обчислюються мікросервісом QualityEvaluator у єдиному форматі. Завдяки цьому забезпечується узгодженість оцінювання різних алгоритмів і можливість коректного порівняння їх поведінки на множинах об'єктів із різною структурою.

ModelSelector отримує від FCM2 і PCM2 інтервальні центри кластерів, ступені приналежності та значення функцій відповідності, після чого виконує нормалізацію цих даних і формує порівняльні профілі якості.

Передача результатів від FCM2, PCM2, ShapeAnalyzer та QualityEvaluator у ModelSelector здійснюється через REST API у форматі JSON, що забезпечує узгодженість представлення кластерних параметрів у мікросервісній архітектурі Clustering4Build.

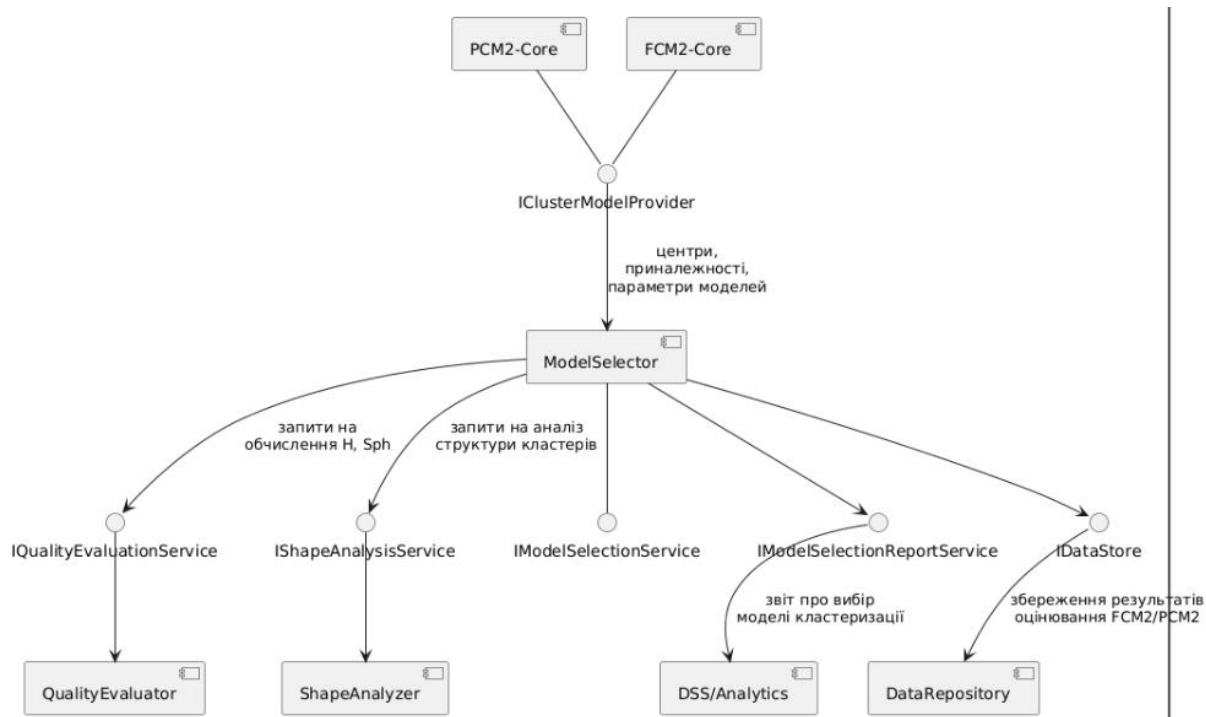


Рис. 4.17. UML-компонентна діаграма підсистеми ModelSelector (Розроблено автором)

На діаграмі відображено архітектурну структуру підсистеми ModelSelector, яка забезпечує автоматизований вибір моделі кластеризації на основі IHMT2. Компонент взаємодіє з модулями FCM2-Core та PCM2-Core для отримання результатів кластеризації, з QualityEvaluator — для обчислення індексів якості, та ShapeAnalyzer — для аналізу форми кластерів. Результати оцінювання моделі зберігаються у DataRepository та передаються до DSS для подання користувачеві, що забезпечує узгодженість і прозорість процесу прийняття рішень у Clustering4Build.

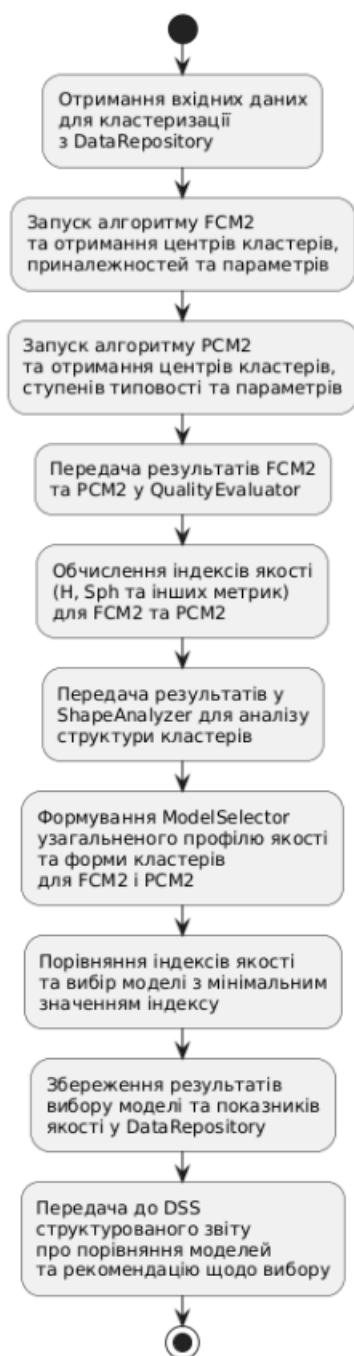


Рис. 4.18. Activity-діаграма процесу вибору моделі кластеризації між FCM2 та PCM2 (Розроблено автором)

Діаграма демонструє динамічний процес обміну повідомленнями між ModelSelector та іншими мікросервісами інформаційної технології Clustering4Build. Зокрема, показано запуск алгоритмів FCM2 і PCM2, передавання результатів кластеризації, отримання значень індексів якості від QualityEvaluator та структурних характеристик від ShapeAnalyzer.

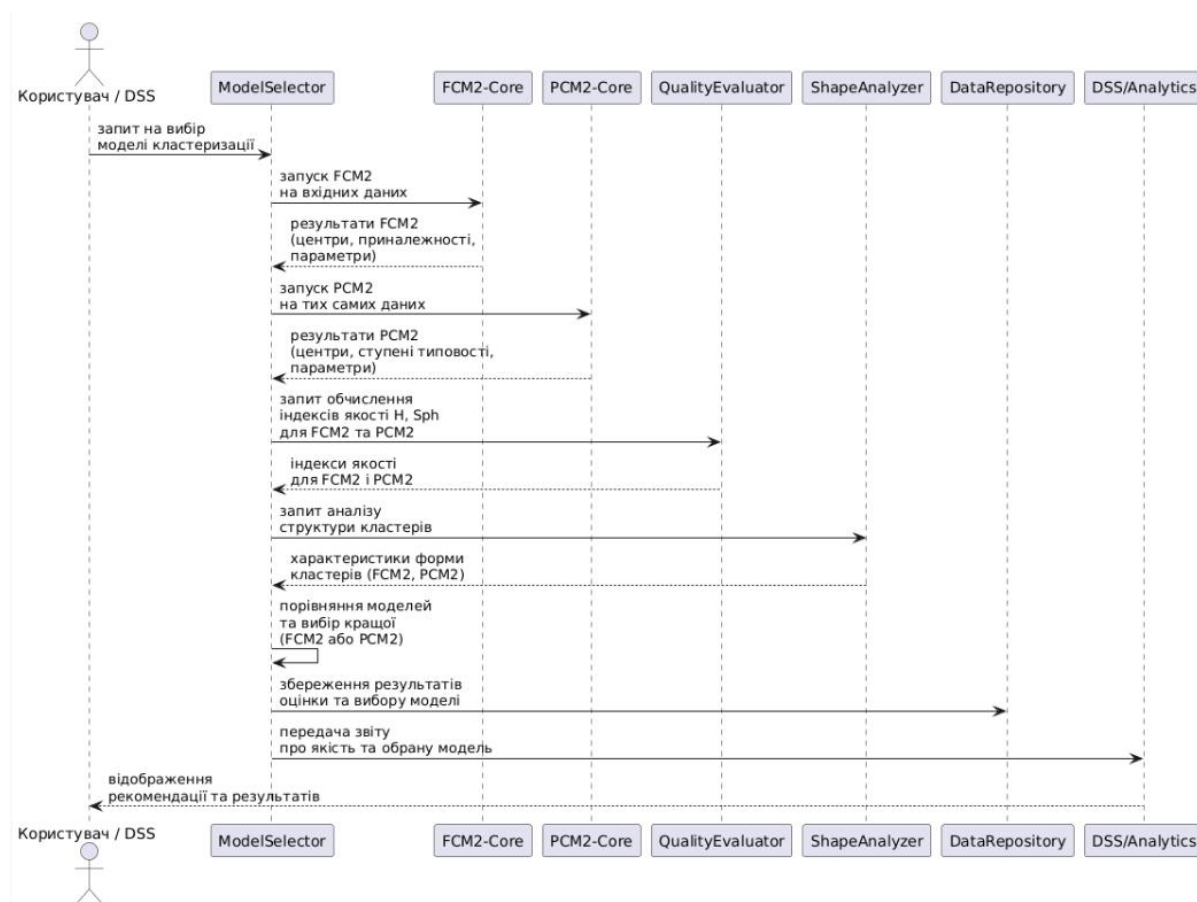


Рис. 4.19. Sequence-діаграма взаємодії ModelSelector з кластеризаторами FCM2 та PCM2 і сервісами QualityEvaluator та DSS (Розроблено автором)

На основі цих даних ModelSelector формує рішення про найкращу модель кластеризації та передає його до DSS для подальшої аналітики, забезпечуючи прозорість та відтворюваність процесу вибору моделі. У рамках архітектури Clustering4Build обробка демонстраційного прикладу реалізується як послідовність взаємодій між окремими мікросервісами, що забезпечують кластеризацію, обчислення індексів якості та формування рекомендацій. На рисунку 3.9.1.У подано деталізований сценарій обміну даними між FCM2-Core, PCM2-Core, QualityEvaluator, ModelSelector та DSS для даного прикладу.

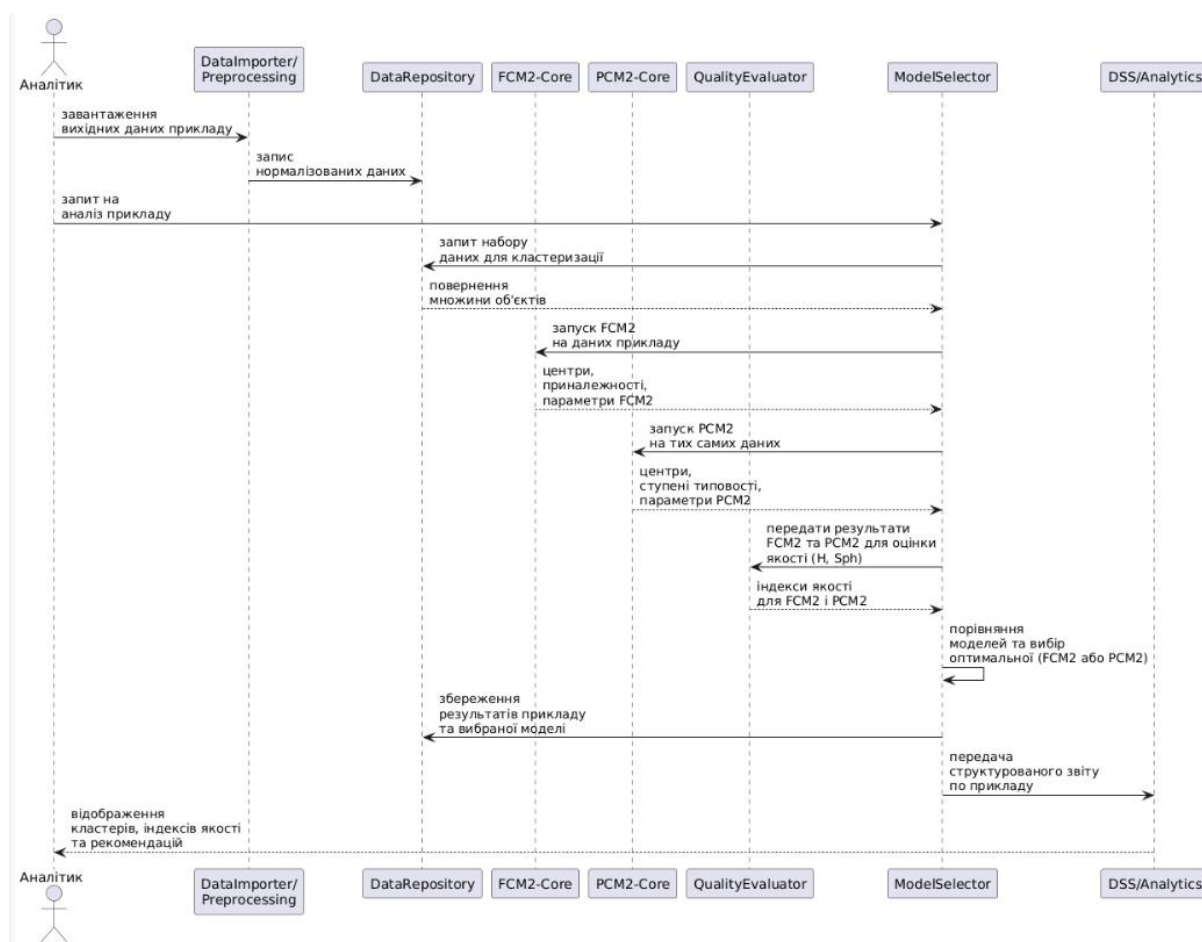


Рис.4.19а. Sequence-діаграма взаємодії мікросервісів при обробці демонстраційного прикладу в інформаційній технології Clustering4Build. (Розроблено автором)

Діаграма відображає динаміку обміну повідомленнями між мікросервісами FCM2-Core, PCM2-Core, QualityEvaluator та ModelSelector у процесі кластеризації демонстраційного набору даних. ModelSelector отримує результати роботи обох алгоритмів, ініціює обчислення індексів якості, виконує порівняння моделей та передає структурований звіт у DSS. Дана схема ілюструє інтеграцію обчислювальних модулів і автоматизацію процесу вибору моделі кластеризації.

У випадку рівномірно сформованих кластерів підсистема ModelSelector автоматично надає пріоритет застосуванню моделей на основі НМТ1, оскільки вони забезпечують менші обчислювальні витрати за збереження достатньої точності для таких структур даних. Натомість при виявленні ознак структурної невизначеності або варіативності меж кластерів система ініціює перехід до використання моделей на основі ІНМТ2, що дозволяє підвищити стійкість кластеризації до нечітких та перехресних областей даних.

У контексті інформаційної технології Clustering4Build цей приклад демонструє роботу мікросервісів FCM2-Core, PCM2-Core та QualityEvaluator на реальних даних технічного стану. Обчислені значення індексів якості та структурні параметри кластерів передаються до DSS для подальшої інтерпретації.

6) ModelSelector (автоматизований вибір моделі кластеризації FCM2/PCM2)

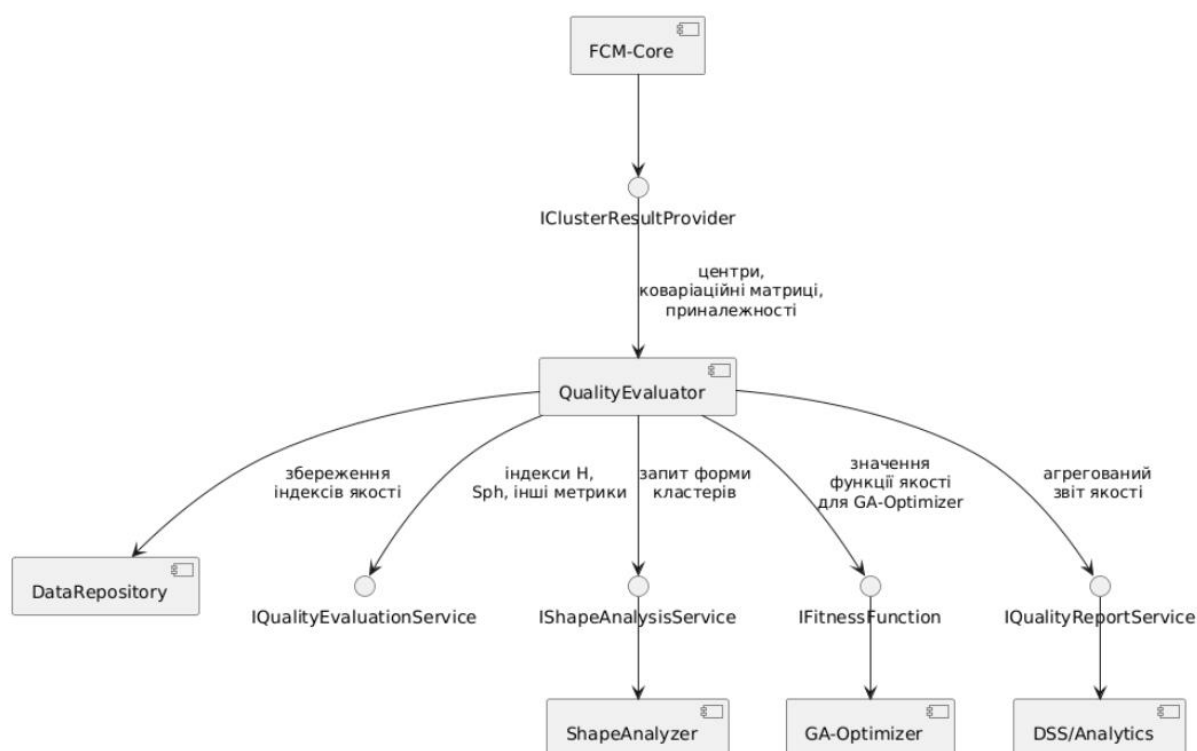


Рис. 4.20. UML-компонентна діаграма підсистеми оцінювання якості кластеризації (Розроблено автором)

На діаграмі подано основні компоненти, які беруть участь в обчисленні показників якості кластеризації. Підсистема QualityEvaluator приймає результати кластеризації від FCM-Core, викликає ShapeAnalyzer для визначення форми кластерів, обчислює показники H або Sph, передає агреговану оцінку до DSS та повертає значення цільової функції в GA-Optimizer. Діаграма відображає архітектурну інтеграцію модулів у загальну мікросервісну інфраструктуру Clustering4Build.

Пропоновані нижче показники кластеризації зазвичай використовуються при оцінюванні якості кластеризації класичного (чіткого)

алгоритму с- середніх. Однак, як показав аналіз, їх використання виявилось можливим і в разі оцінки якості кластеризації з використанням алгоритмів кластеризації на основі ІНМТ2 [32].



Рис. 4.21. UML activity-діаграма процесу оцінювання якості кластеризації в Clustering4Build (Розроблено автором)

При оцінюванні якості кластеризації для кластерів гіпереліпсоїдної форми можна використовувати загальний гіпероб'єм H [32]. Гіпероб'єм H має обчислювальну складність $O(n \cdot d^3)$. Для його оптимізації Clustering4Build використовує векторизовані операції NumPy та кешування часткових коваріаційних матриць:

Алгоритм оцінки якості кластеризації в Clustering4Build включає такі етапи:

1. отримання центрів кластерів та метрик розподілу від FCM-Core;
2. аналіз форми кластерів модулем ShapeAnalyzer;

3. вибір показника (H або Sph);
4. обчислення індексу;
5. формування звіту для DSS;
6. повернення індексу якості в GA-Optimizer як значення функції придатності.

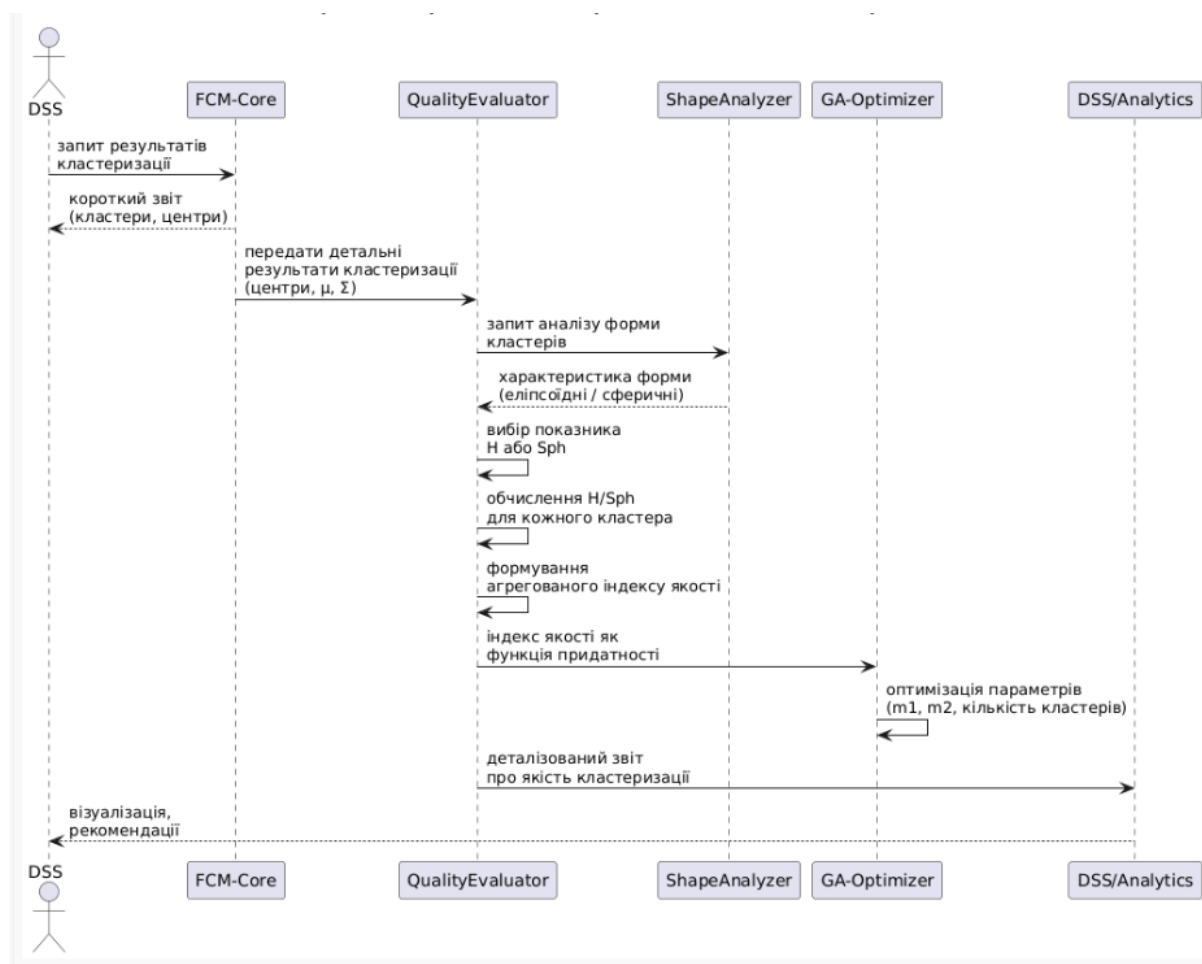


Рис. 4.22. UML sequence-діаграма взаємодії модулів при оцінюванні якості кластеризації (*Розроблено автором*)

На діаграмі відображено послідовність етапів, що виконуються підсистемою QualityEvaluator: отримання результатів кластеризації, аналіз форми кластерів, вибір відповідного показника (H або Sph), обчислення значень індексів, формування агрегованої оцінки та передавання її до модулів DSS та GA-Optimizer. Діаграма демонструє логіку роботи підсистеми у рамках інформаційної технології Clustering4Build.

Показник є основним для гіперсферичних структур даних.

Обчислювальна ефективність відіграє критичну роль в умовах потокової кластеризації. У Clustering4Build обчислення гіпероб'ємів проводиться з використанням векторизованих операцій NumPy, а обчислення Sph

виконується паралельно для всіх кластерів, що дозволяє знизити час обчислення для великих наборів даних.

У Clustering4Build автоматичний вибір між H та Sph здійснюється на основі аналізу спектра власних значень ковараційних матриць, що реалізовано в модулі ShapeAnalyzer.

7) DSS (Decision Support System) та UI (візуалізація і звіти)

Endpoint GET /dss/dashboard/{run_id} агрегує результати кластеризації (центри, ФП/типовість) і метрики якості (H/FH/Sph) у вигляді єдиного аналітичного JSON для відображення в UI.

Програмна реалізація обчислення інтервальних функцій типовості та центрів кластерів виконується в Clustering4Build за допомогою мікросервісу TypeReducer, який реалізує алгоритм Карника–Менделя та використовує векторизовані обчислення (NumPy) для прискорення обробки великих масивів даних.

Для автоматизованого пошуку оптимальних параметрів РСМ-алгоритму на основі IHMT2 у Clustering4Build використовується мікросервіс GA-Optimizer, який розглядає параметри фазифікатора та "ширини зони" як компоненти хромосом.

У межах інформаційної технології Clustering4Build генетичний алгоритм реалізовано у складі мікросервісу GA-Optimizer, який взаємодіє з РСМ-Core, FuzzificationEngine, QualityEvaluator та DSS через REST API. Це забезпечує автоматизований пошук параметрів алгоритму кластеризації без участі користувача і інтеграцію результатів у контур прийняття рішень системи технічного моніторингу.

У Clustering4Build хромосоми такого виду представляються у вигляді JSON-структур, що зберігаються в конфігураційному сховищі. Це дає змогу відтворювати конфігурації кластеризації, порівнювати різні запуски ГА та забезпечувати відмовостійкість системи.

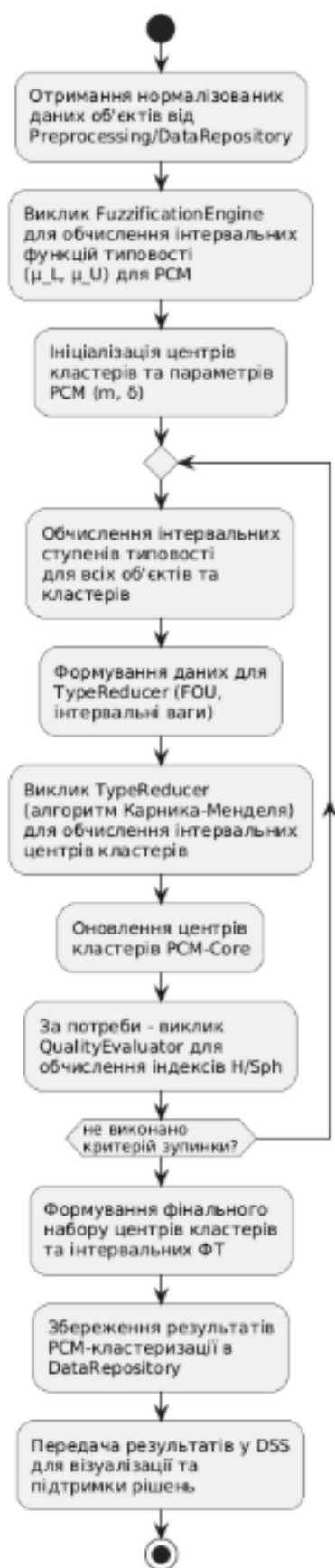


Рис.4.23. Діаграма UML Activity - процес РСМ-кластеризації на основі ІНМТ2 (Розроблено автором)

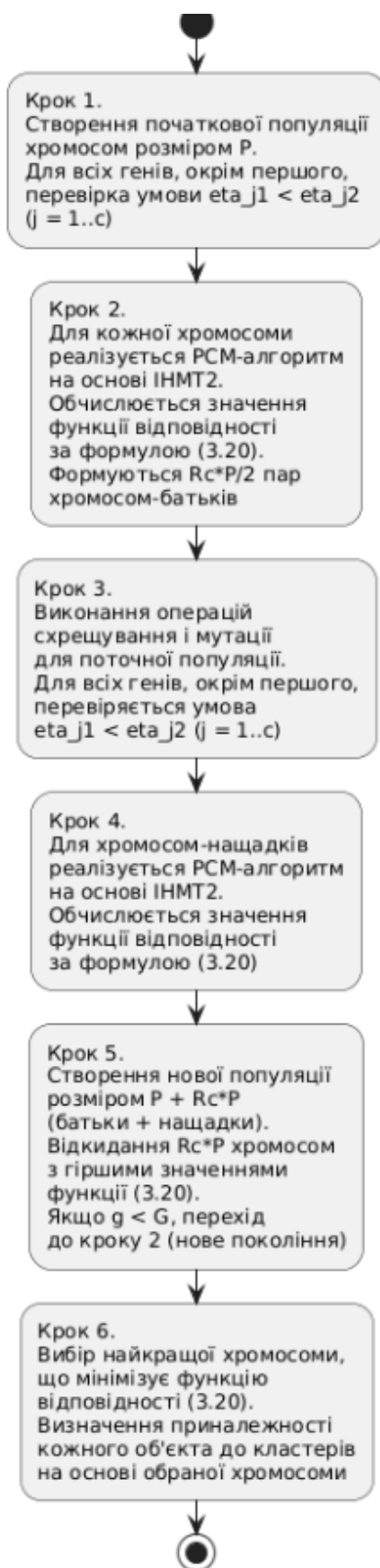


Рис. 4.24. UML activity-діаграма генетичного алгоритму оптимізації параметрів РСМ-алгоритму на основі IHMT2. (Розроблено автором)

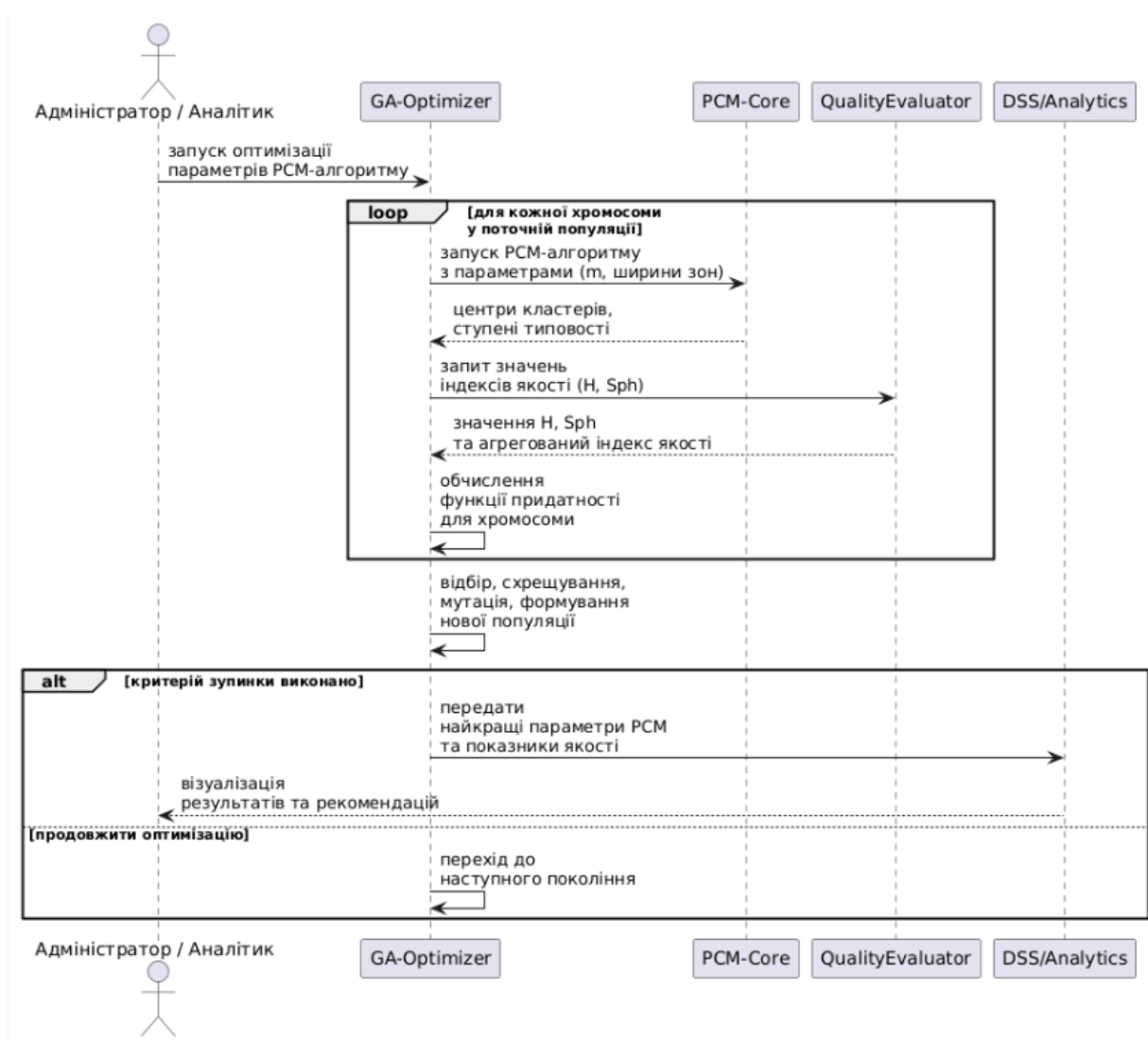


Рис. 4.25. Sequence-діаграма взаємодії мікросервісу GA-Optimizer з PCM-Core та QualityEvaluator в інформаційній технології Clustering4Build. (Розроблено автором)

Діаграма ілюструє обмін даними між мікросервісами GA-Optimizer, PCM-Core та QualityEvaluator під час обчислення значень функції придатності для хромосом генетичного алгоритму. Вона демонструє процес виклику PCM-алгоритму з параметрами, одержання характеристик кластерів, обчислення індексів якості та повернення результатів у GA-Optimizer. Отримані показники передаються також у підсистему DSS для візуалізації та формування рекомендацій, що забезпечує цілісність обчислювального контуру кластеризації.

Endpoint GET /dss/heatmap/{run_id} формує дані для теплових карт технічного стану (кластерні мітки, інтенсивності ризику, порогові інтервали), що дозволяє інтерпретувати кластери у термінах стану об'єктів.

Endpoint POST /dss/report генерує звіт (структуровані таблиці, графіки, метрики, рекомендації DSS), повертає report_id, а метадані/посилання на артефакт зберігає у PostgreSQL.reports (jsonb).

8) PostgreSQL (DataRepository, JSONB)

Таблиця clustering_runs (jsonb) зберігає результати кластеризації (центри, матриці ФП/інтервали ФП, параметри запусків, протоколи збіжності).

Таблиця it2_memberships (jsonb) зберігає інтервальні функції приналежності μ_L/μ_U та FOU.

Таблиця quality_metrics (jsonb) зберігає H/FH/Sph та допоміжні значення для аналізу.

Таблиці ga_runs, ga_population, model_configs (jsonb) фіксують конфігурації оптимізації, популяції й найкращі знайдені параметри.

9) API Gateway / REST інтеграція

Endpoint POST /pipeline/run реалізує оркестрацію типового сценарію (валідація → нормалізація → фазифікація → кластеризація → оцінка якості → передача в DSS), повертаючи ідентифікатори run_id та пов'язані refs для відтворюваності.

Endpoint GET /pipeline/status/{run_id} повертає стан виконання та журнал подій, що використовується UI для відображення прогресу обчислень.

Таблиця 4.4. Функціональна реалізація пакетів прикладних програм Clustering4Build (мікросервісний рівень)/ Пакет 1. Кластеризація (FCM, IHMT2) (Розроблено автором)

Пакет модуль	Endpoint / сервіс	Операція (програма на функція)	Вхідні дані	Вихідні дані	Запис у PostgreSQL (JSONB)
Пакет 1. Кластеризація (FCM, IHMT2)	POST /fuzzification/validate (FuzzificationEngine)	Валідація та контроль якості вхідних даних	Матриця оцінок об'єктів, метадані ознак	Протокол перевірки, dataset_id	data_quality_log

Пакет / модуль	Endpoint / сервіс	Операція (програмна функція)	Вхідні дані	Вихідні дані	Запис у PostgreSQL (JSONB)
	POST /fuzzification/normalize	Нормалізація та масштабування ознак	dataset_id, метод нормалізації	dataset_ref	normalization_profiles
	POST /fuzzification/it2/membership	Побудова $\mu_L(x)$, $\mu_U(x)$, FOU (IHMT2)	dataset_ref, m ₁ , m ₂	membership_ref	it2_memberships
	POST /clustering/fcm/t1/run (FCM-Core)	Запуск FCM на основі IHMT1	dataset_ref, c, m	run_id	clustering_runs
	POST /clustering/fcm/it2/run	Запуск FCM на основі IHMT2	membership_ref, c, m ₁ , m ₂	run_id	clustering_runs
	GET /clustering/run/{run_id}	Отримання результатів кластеризації	run_id	Центри, ФП інтервали ФП	—

Наведена таблиця відображає програмну реалізацію пакетів прикладних програм інформаційної технології Clustering4Build на рівні мікросервісів і REST-інтерфейсів. Кожна функція представлена у вигляді конкретного сервісного виклику з чітко визначеними вхідними та вихідними даними і механізмами збереження результатів у централізованому сховищі PostgreSQL. Таблиці опису функціональної реалізації пакетів прикладних програм Clustering4Build (мікросервісний рівень) пакет 2. Оцінка якості та моніторинг, пакет 3. Адаптивна оптимізація, пакет 3 (IHMT2), пакет 4. Інтерфейс і візуалізація, пакет 5. Зберігання даних наведено в Додатку 7.

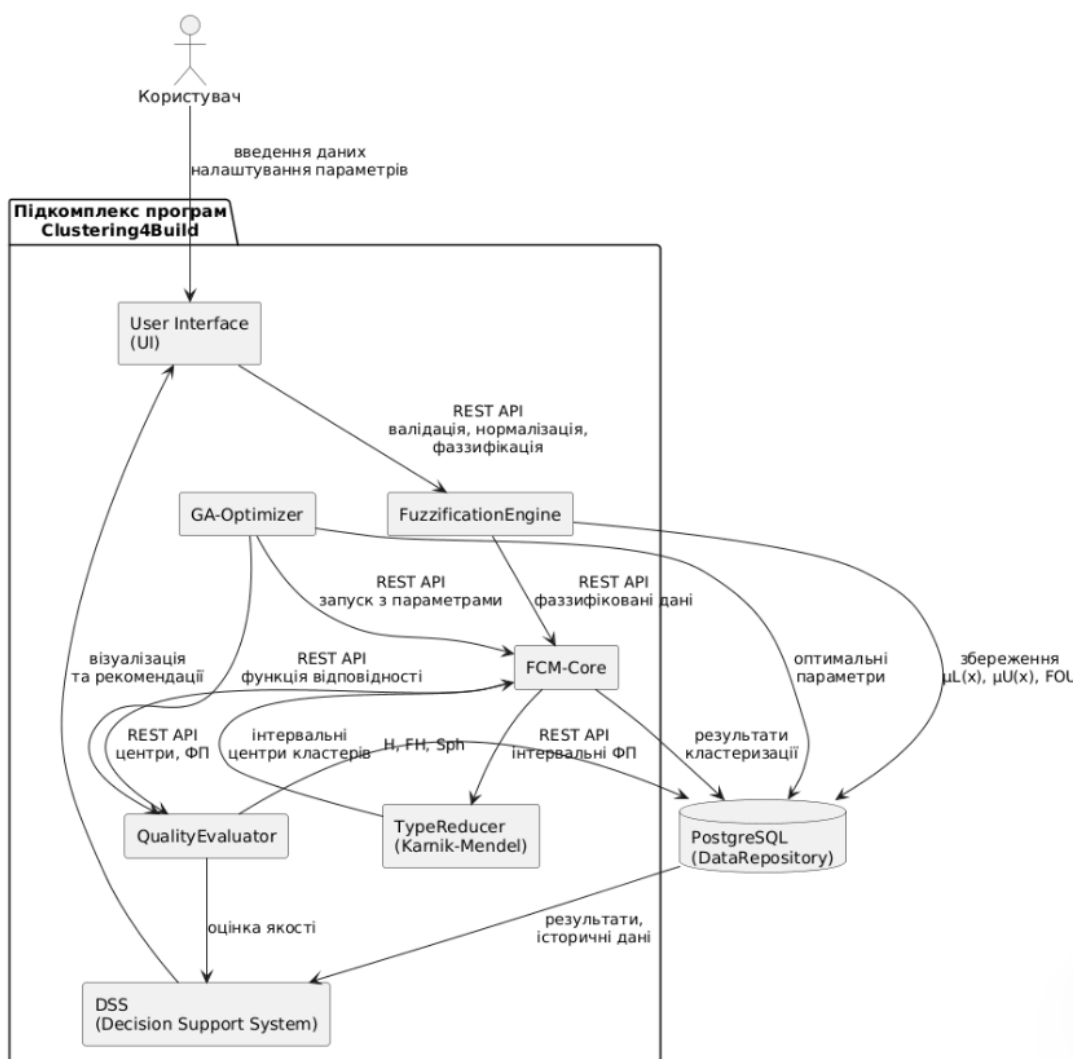


Рис. 4.26. Структурна схема роботи підкомплексу програм «FCMT2»
(Розроблено автором)

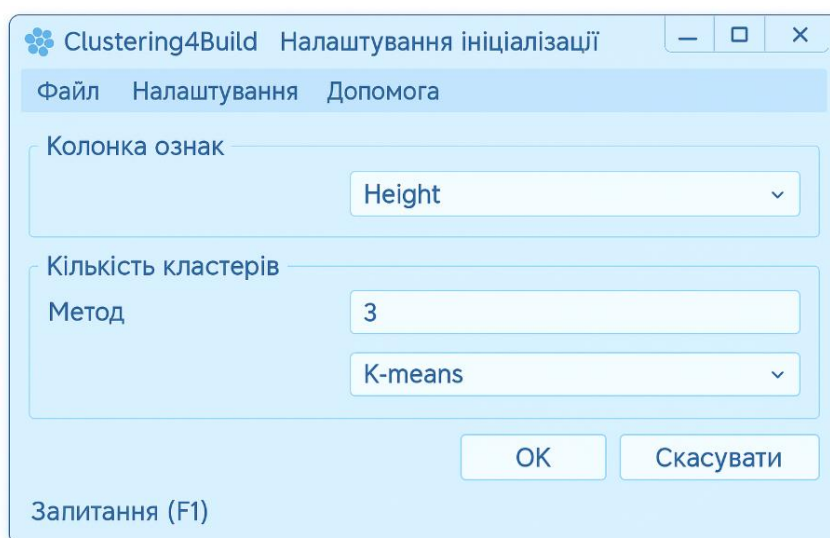


Рис. 4.27. Діалогове вікно фаззифікації (НМТ1 / ІНМТ2) (Розроблено автором)

Діалогове вікно реалізує формування нечіткого подання вхідних даних, зокрема побудову функцій приналежності для НМТ1 та інтервальних функцій $\mu_L(x)$, $\mu_U(x)$ для ІНМТ2. Результати фазифікації фіксуються в PostgreSQL у форматі JSONB та використовуються як вхідні дані для FCM-Core і TypeReducer. Вікно забезпечує параметризацію фазифікаторів, що критично впливає на стійкість кластеризації в умовах невизначеності.

Рис. 4.28. Діалогове вікно запуску кластеризації (FCM / PCM)
(Розроблено автором)

Вікно забезпечує керування запуском кластеризації із вибором алгоритму (FCM або PCM), типу нечіткої моделі та параметрів ітераційного процесу. Після виконання формується ідентифікатор запуску `run_id`, за яким у PostgreSQL зберігаються центри кластерів, матриці приналежності/типовості та службові метадані збіжності. Далі результати передаються до QualityEvaluator для розрахунку показників якості.

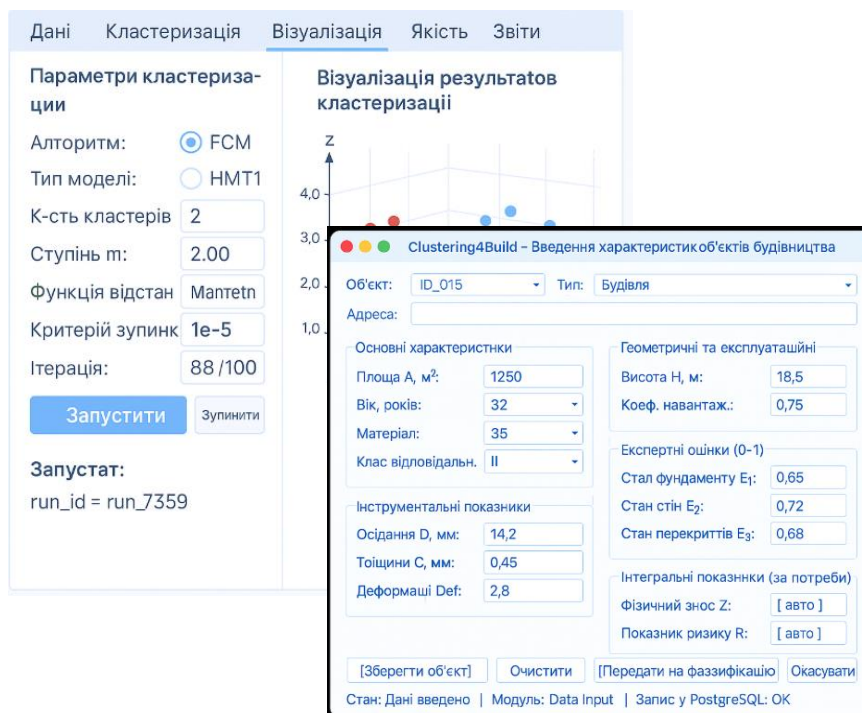


Рис. 4.29. Діалогове вікно візуалізації результатів кластеризації (Розроблено автором)

Вікно призначене для аналізу просторової структури кластерів шляхом відображення об'єктів, центрів кластерів та, за необхідності, інтервальних зон невизначеності IHMT2. Візуалізація використовується для експертної інтерпретації результатів та контролю коректності розбиття на кластери. Дані для графічного представлення отримуються з PostgreSQL за ідентифікатором run_id.

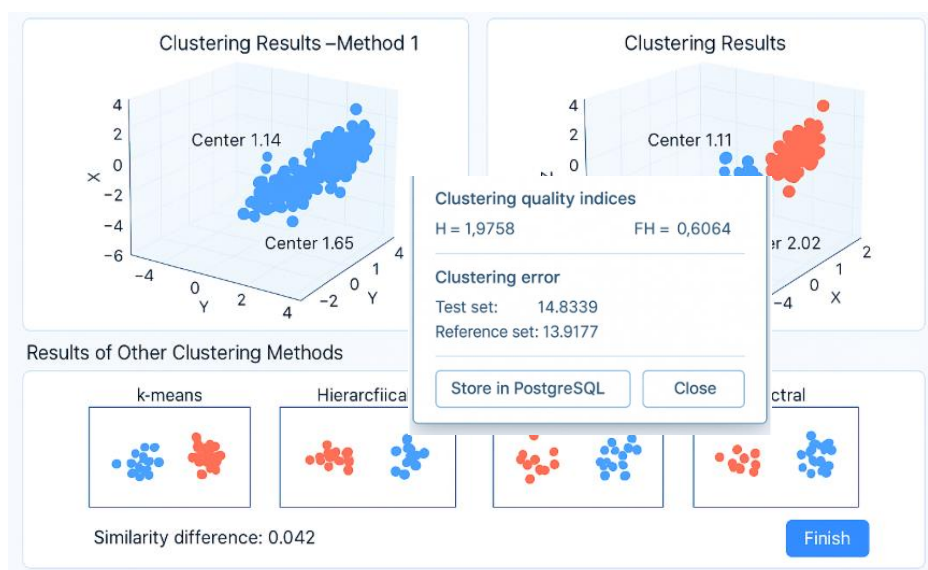


Рис. 4.30. Діалогове вікно оцінки якості кластеризації (Розроблено автором)

Вікно реалізує обчислення та відображення показників якості кластеризації, зокрема H , FH та Sph , які формально характеризують компактність і геометрію кластерів. Отримані метрики фіксуються у PostgreSQL і використовуються як критерії порівняння альтернативних запусків та як функція відповідності в GA-Optimizer. Відображення помилки кластеризації забезпечує контроль адекватності розбиття для тестових і еталонних множин.



Рис. 4.31. Діалогове вікно оптимізації параметрів (GA-Optimizer)
(Розроблено автором)

Вікно забезпечує керування процесом генетичної оптимізації параметрів кластеризації, зокрема фазифікаторів m_1 , m_2 та параметрів типовості η . Реалізація передбачає автоматизований цикл викликів FCM/PCM та QualityEvaluator через REST API з використанням метрик якості як функції відповідності. Найкращі знайдені параметри фіксуються в PostgreSQL як конфігурація повторного застосування без повторної оптимізації.



Рис.4.32. Діалогове вікно DSS та формування звітів (Розроблено автором)

Вікно DSS забезпечує інтерпретацію результатів кластеризації у термінах технічного стану об'єктів будівництва та формування рекомендацій для подальших управлінських дій. Теплові карти та аналітичні зведення формуються на основі даних, збережених у PostgreSQL, та відображають просторовий і кластерний розподіл рівнів ризику. Підсистема підтримує генерацію звітів та експорт результатів для інтеграції з зовнішніми інформаційними системами.

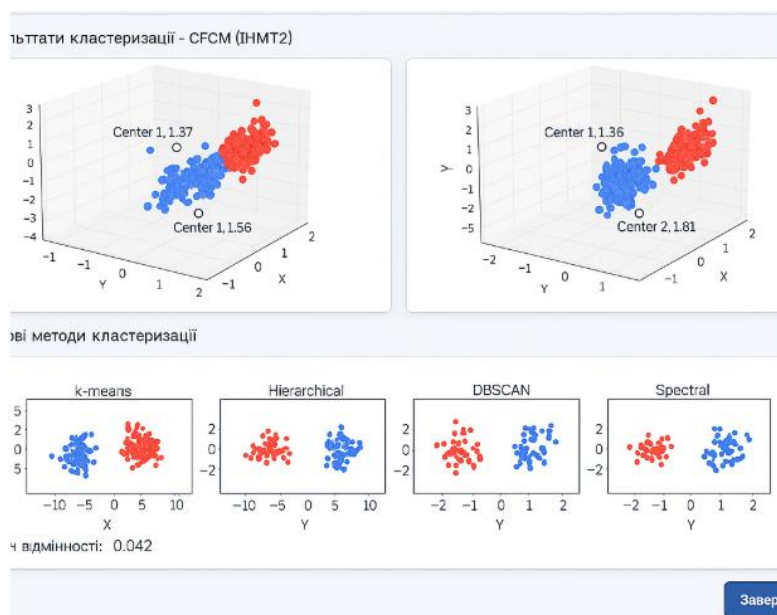


Рис. 4.33. Інтерфейс порівняльного аналізу результатів кластеризації. (Розроблено автором)

Представлений інтерфейс призначений для порівняльного аналізу результатів кластеризації, отриманих із застосуванням методів нечіткої та

можливої кластеризації на основі інтервальних нечітких множин другого типу (FCM/PCM на основі ІНМТ2), з результатами базових класичних методів кластеризації. Основною задачею є оцінювання впливу вибору алгоритму кластеризації на структуру кластерів, координати їх центрів та ступінь роздільності об'єктів у просторі ознак за умов істотної різниці щільності та об'єму кластерів.

Вхідними даними для даного інтерфейсу є множина об'єктів будівництва (будівель і споруд або їх конструктивних елементів), технічний стан яких описується векторами кількісних характеристик, отриманих у результаті інструментального та/або експертного моніторингу. До таких характеристик належать геометричні, експлуатаційні, інструментальні та інтегральні показники технічного стану, які після нормалізації та фазифікації формують простір ознак для кластеризації.

У верхній частині інтерфейсу відображаються основні результати кластеризації, отримані за допомогою FCM та PCM-алгоритмів на основі ІНМТ2, у вигляді тривимірних діаграм з чітким виділенням кластерів різними кольорами та зазначенням координат їх центрів. Нижня частина інтерфейсу містить результати кластеризації тими ж вхідними даними за допомогою базових методів (k-means, ієрархічна кластеризація, DBSCAN тощо), які використовуються виключно як еталонні методи порівняння.

Такий підхід дозволяє наочно продемонструвати переваги застосування нечіткої та можливої кластеризації на основі ІНМТ2 для задач оцінювання технічного стану об'єктів будівництва в умовах невизначеності, зокрема підвищену стійкість до варіацій даних та коректніше формування кластерів істотно різного об'єму і щільності. Отримані результати використовуються надалі в підсистемі підтримки прийняття рішень для формування рекомендацій щодо подальшого технічного обстеження та експлуатації об'єктів.

ВИСНОВКИ ДО РОЗДІЛУ 4

У четвертому розділі виконано програмну реалізацію розроблених у розділах 2–3 методів кластеризації у вигляді інформаційної технології

Clustering4Build — інтегрованого програмного комплексу оцінювання технічного стану об'єктів будівництва.

У межах розділу було:

- обґрунтовано загальну архітектуру пакета прикладних програм на основі мікросервісного підходу, що забезпечує модульність, незалежне розгортання та масштабованість окремих аналітичних компонентів;

- розроблено дев'ять програмних модулів, що реалізують повний цикл обробки даних: FuzzificationEngine (фаззифікація вхідних показників), FCM-Core (кластеризація на основі HMT1 та IHMT2), TypeReducer (зниження типу за алгоритмом Карника–Менделя), GA-Optimizer (генетична оптимізація параметрів кластеризації), QualityEvaluator (обчислення показників якості H та Sph), ModelSelector (автоматизований вибір моделі кластеризації між FCM2 та PCM2), DSS (підтримка прийняття рішень і візуалізація), а також модулі збереження даних (PostgreSQL) та інтеграції (API Gateway);

- реалізовано взаємодію модулів через REST API з передаванням даних у форматі JSON/JSONB, що забезпечує технологічну незалежність компонентів та можливість інтеграції із зовнішніми системами (BIM, IoT);

- забезпечено централізоване збереження результатів кластеризації, конфігурацій генетичного алгоритму та історії обчислювальних експериментів у базі даних PostgreSQL, що гарантує відтворюваність та можливість подальшого аналізу;

- реалізовано підсистему підтримки прийняття рішень (DSS), яка на основі результатів кластеризації та показників якості формує рекомендації щодо технічного стану об'єктів будівництва для прийняття інженерних рішень.

Розроблена інформаційна технологія Clustering4Build забезпечує практичну реалізацію математичного апарату, викладеного в розділах 2–3, у вигляді програмного комплексу, придатного до застосування в задачах автоматизованого технічного моніторингу об'єктів будівництва.

ЗАГАЛЬНІ ВИСНОВКИ

У дисертаційній роботі вирішено актуальну науково-прикладну задачу створення інтегрованої інформаційної технології оцінювання технічного стану об'єктів будівництва в умовах невизначеності на основі методів кластеризації з використанням нечітких множин першого і другого типів та генетичних алгоритмів оптимізації. Актуальність задачі зумовлена тим, що традиційні детерміновані методи оцінювання технічного стану не враховують розмитості меж між станами конструкцій, неповноти й шумового характеру даних інструментального та експертного моніторингу, а наявні BIM-, SHM- та DSS-рішення не забезпечують вбудованих механізмів адаптивної нечіткої кластеризації. Відповідно до поставлених завдань дослідження отримано такі результати:

1. Проведено аналіз існуючих методів та інформаційних технологій оцінювання технічного стану об'єктів будівництва в умовах невизначеності — класичних детермінованих методів кластеризації (ієрархічна кластеризація, алгоритм чітких *c*-середніх), а також сучасних BIM-, SHM- та DSS-рішень (PlanRadar, Archibus, IBM Maximo, Autodesk Revit, Bentley iTwin). Встановлено обмеженість їх аналітичних можливостей для комплексної кластеризації різномірних даних і сформульовано вимоги до інформаційної технології кластеризації, що включають адаптивну обробку нечітких, інтервальних і лінгвістичних даних, автоматизоване налаштування параметрів моделей засобами генетичних алгоритмів, формалізовану оцінку якості розбиттів та реалізацію у вигляді модульної мікросервісної архітектури з REST-взаємодією.

2. Розроблено математичний апарат і алгоритми кластерного аналізу з використанням нечітких множин першого типу — формалізовано технічні показники об'єктів будівництва у вигляді нечітких множин з трикутними, трапецієподібними та гауссовими функціями належності, реалізовано алгоритм нечітких *c*-середніх (FCM) для адаптивного визначення технічного стану конструкцій. Розроблено математичний апарат кластеризації на основі

інтервальних нечітких множин другого типу (ІНМТ2), що дозволяє формалізувати невизначеність самої функції належності через «відбиток невизначеності» (FOU) та обчислювати інтервальний центроїд кластера за допомогою ітераційного алгоритму Карника–Менделя, що забезпечує стійкість результатів кластеризації до варіативності вихідних даних.

3. Створено комбіновані алгоритмічні підходи на основі FCM, PCM та генетичної оптимізації параметрів кластеризації (центрів кластерів, ступенів приналежності — для двох схем кодування хромосом постійної та змінної довжини), адаптовані до обробки неповних, шумових і різномірних даних технічного моніторингу. Поширено можливісний PCM-алгоритм на ІНМТ2 у двох незалежних варіантах формалізації невизначеності — через два фазифікатори та через два значення «ширини зони» кожного кластера, — що виходить за межі відомого підходу до FCM-кластеризації на основі ІНМТ2 і забезпечує підвищену стійкість до шумових даних і викидів, характерних для реальних вимірювань технічного стану будівельних конструкцій.

4. Розроблено програмні модулі фазифікації (FuzzificationEngine), аналітичного ядра кластеризації (FCM-Core, PCM-Core), зниження типу (TypeReducer) та генетичної оптимізації (GA-Optimizer) з інтеграцією з підсистемою підтримки прийняття рішень (DSS), BIM-моделями і потоками даних IoT-пристроїв, що забезпечує автоматизований збір, оброблення й інтерпретацію багатовимірних даних про технічний стан конструкцій у режимі, наближеному до реального часу.

5. Впроваджено методику оцінювання технічного стану будівель з використанням ІНМТ2 та автоматизованого підбору параметрів моделей через генетичну оптимізацію. Запропоновано показники якості кластеризації — загальний гіпероб'єм H та індекс сферичності Sph , адаптовані до кластерів різної форми, щільності та об'єму, — а також критерії визначення істотної відмінності кластерів за об'ємом і потужністю, що в сукупності забезпечують обґрунтований автоматизований вибір моделі кластеризації (FCM або PCM, на

основі НМТ1 або ІНМТ2) залежно від структури вхідних даних технічного моніторингу.

6. Забезпечено реалізацію інформаційної технології як інтегрованого програмного продукту Clustering4Build — дев'яти взаємопов'язаних модулів (фазифікації, кластеризації, зниження типу, генетичної оптимізації, оцінки якості, автоматизованого вибору моделі, підтримки прийняття рішень, збереження даних у PostgreSQL та інтеграційного шлюзу), що взаємодіють через REST API у форматі JSON/JSONB, для адаптивного, точного і відтворюваного моніторингу технічного стану об'єктів будівництва у реальному часі.

Обчислювальні експерименти на контрольних наборах даних технічного моніторингу підтвердили практичну ефективність розроблених методів: застосування комбінованих алгоритмів FCM+GA і PCM+GA, зокрема на основі ІНМТ2, забезпечує вищу точність і стабільність кластеризації порівняно з класичними методами без параметричної оптимізації, а також кращі результати для множин об'єктів із суттєво різною щільністю та об'ємом кластерів. Таким чином, мету дисертаційного дослідження досягнуто повністю, усі шість поставлених завдань вирішено, а отримані результати створюють підґрунтя для подальшого розширення розробленої інформаційної технології на інші класи нечітких моделей і системи прогнозування залишкового ресурсу будівельних конструкцій.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Бойко М. П. Надійність будівельних конструкцій і основ. – Київ : Вища школа, 2003. – 432 с.
2. Звіт про прямі збитки інфраструктури від руйнувань внаслідок військової агресії Росії проти України станом на листопад 2024 року : аналітичний звіт / Київська школа економіки. — Київ, 2025. — 42 с. — URL: https://kse.ua/wp-content/uploads/2025/02/KSE_Damages_Report-November-2024-UA.pdf.
3. ДСТУ 9273:2024. Настанова щодо обстеження будівель і споруд для визначення та оцінювання їхнього технічного стану. — Київ : ДП «УкрНДНЦ», 2024.
4. Філіпчук С. В., Шмуклер В. С. Обстеження та оцінка технічного стану будівель і споруд. – Харків : ХНУБА, 2011. – 256 с.
5. Farrar C. R., Worden K. An introduction to structural health monitoring // Philosophical Transactions of the Royal Society A. – 2007. – Vol. 365, No. 1851. – P. 303–315.
6. Sohn H., Farrar C. R., Hemez F. M. et al. A review of structural health monitoring literature: 1996–2001. – Los Alamos National Laboratory Report LA-13976-MS, 2004. – 283 p.
7. Zadeh L. A. Fuzzy sets // Information and Control. – 1965. – Vol. 8, No. 3. – P. 338–353.
8. Dubois D., Prade H. Fuzzy sets and systems: theory and applications. – New York : Academic Press, 1980. – 393 p.
9. Bezdek J. C. Pattern Recognition with Fuzzy Objective Function Algorithms. – New York : Plenum Press, 1981. – 256 p.
10. Höppner F., Klawonn F., Kruse R., Runkler T. Fuzzy Cluster Analysis: Methods for Classification, Data Analysis and Image Recognition. – Chichester : John Wiley & Sons, 1999. – 361 p.
11. Mendel J. M. Uncertain Rule-Based Fuzzy Logic Systems: Introduction and New Directions. – Upper Saddle River : Prentice Hall, 2001. – 684 p.

12. Mendel J. M., John R. I. Type-2 fuzzy sets made simple // *IEEE Transactions on Fuzzy Systems*. – 2002. – Vol. 10, No. 2. – P. 117–127.
13. Karnik N. N., Mendel J. M. Centroid of a type-2 fuzzy set // *Information Sciences*. – 2001. – Vol. 132, No. 1–4. – P. 195–220.
14. Gao Y., Mosalam K. M. Deep learning for structural health monitoring: A review // *Structural Control and Health Monitoring*. – 2019. – Vol. 26, No. 5. – Article e2301.
15. Worden K., Manson G. The application of machine learning to structural health monitoring // *Philosophical Transactions of the Royal Society A*. — 2007. — Vol. 365, No. 1851. — P. 515–537.
16. Hwang C., Rhee F. C.-H. Uncertain fuzzy clustering: interval type-2 fuzzy approach to C-means // *IEEE Transactions on Fuzzy Systems*. — 2007. — Vol. 15, No. 1. — P. 107–120.
17. Wu D., Mendel J. M. Enhanced Karnik–Mendel algorithms // *IEEE Transactions on Fuzzy Systems*. — 2009. — Vol. 17, No. 4. — P. 923–934.
18. Hidalgo-Fort E., Blanco-Carmona P., Serrano-Chacon A. et al. IoT Framework Application to SHM Systems // *Bridge Engineering — Recent Advances and Applications* / ed. by S. Yilmaz, Y. Yardim. — London : IntechOpen, 2024. — DOI: 10.5772/intechopen.112719.
19. Azimi M., Eslamlou A. D., Pekcan G. Data-driven structural health monitoring and damage detection through deep learning: state-of-the-art review // *Sensors*. — 2020. — Vol. 20, No. 10. — Article 2778.
20. Перельмутер А. В. Шляхи розвитку оцінок технічного стану будівель і споруд // *Опір матеріалів і теорія споруд*. — 2024. — Вип. 112. — С. 28–35. — DOI: 10.32347/2410-2547.2024.112.28-35.
21. Усаковський С. Б. Попередня оцінка надійності існуючих конструкцій за неповними даними // *Містобудування та територіальне планування*. — 2015. — Вип. 58. — С. 525–528.

22. Tur V. V., Yalavaya Y. S. Assessment of existing reinforced concrete structures with usage of the fuzzy logic-based expert system // *Building and Reconstruction*. — 2019. — No. 5 (85). — P. 74–84.
23. Calvi G. M., Pinho R., Magenes G., Bommer J. J., Restrepo-Vélez L. F., Crowley H. Development of seismic vulnerability assessment methodologies over the past 30 years // *ISET Journal of Earthquake Technology*. — 2006. — Vol. 43, No. 3. — P. 75–104.
24. Malekloo A., Ozer E., AlHamaydeh M., Girolami M. Machine learning and structural health monitoring overview with emerging technology and high-dimensional data source highlights // *Structural Health Monitoring*. — 2022. — Vol. 21, No. 4. — P. 1906–1955.
25. Dubois D., Prade H. Possibility theory and fuzzy sets: recent developments // *Fuzzy Sets and Systems*. — 2020. — Vol. 394. — P. 1–16.
26. Sotudian S., Zarandi M. H. F. Interval Type-2 Enhanced Possibilistic Fuzzy C-Means Clustering for Gene Expression Data Analysis // *arXiv preprint*. — 2021. — arXiv:2101.00304.
27. Ni Y. Q., Xia Y., Liao W. Y., Ko J. M. Technology innovation in developing the structural health monitoring system for Guangzhou New TV Tower // *Structural Control and Health Monitoring*. — 2009. — Vol. 16, No. 1. — P. 73–98.
28. Abosata N., Al-Rubaye S., Inalhan G., Emmanouilidis C. Internet of Things for system integrity: a comprehensive survey on security, attacks and countermeasures for industrial applications // *Sensors*. — 2021. — Vol. 21, No. 11. — Article 3654.
29. Лантух-Лященко А. І. Марковська модель оцінки і прогнозу технічного стану будівельних конструкцій // *Дороги і мости*. — 2019. — Вип. 19–20. — С. 27–37.
30. Everitt B. S., Landau S., Leese M., Stahl D. *Cluster Analysis*. — 5th ed. — Chichester: John Wiley & Sons, 2011. — 346 p.
31. Довгань П. М., Улицький О. А. Інженерно-геологічні вишукування для будівництва: сучасний стан і перспективи розвитку // *Вісник Київського*

- національного університету імені Тараса Шевченка. Геологія. — 2020. — № 1 (88). — С. 47–53.
32. Gath I., Geva A. B. Unsupervised optimal fuzzy clustering // IEEE Transactions on Pattern Analysis and Machine Intelligence. — 1989. — Vol. 11, No. 7. — P. 773–780.
 33. Jain A. K., Dubes R. C. Algorithms for Clustering Data. — Englewood Cliffs, NJ : Prentice Hall, 1988. — 320 p.
 34. Mendel J. M., John R. I., Liu F. Interval type-2 fuzzy logic systems made simple // IEEE Transactions on Fuzzy Systems. — 2006. — Vol. 14, No. 6. — P. 808–821.
 35. Zimmermann H.-J. Fuzzy Set Theory — and Its Applications. — 4th ed. — Boston : Kluwer Academic Publishers, 2001. — 514 p.
 36. Adeli H., Hung S.-L. Machine Learning: Neural Networks, Genetic Algorithms, and Fuzzy Systems. — New York : Wiley, 1995. — 269 p.
 37. Flood I., Kartam N. Neural networks in civil engineering I: principles and understanding // Journal of Computing in Civil Engineering. — 1994. — Vol. 8, No. 2. — P. 131–148.
 38. Sohn H., Law K. H. A Bayesian probabilistic approach for structure damage detection // Earthquake Engineering & Structural Dynamics. — 1997. — Vol. 26, No. 12. — P. 1259–1281.
 39. Rokach L., Maimon O. Clustering methods // Data Mining and Knowledge Discovery Handbook / ed. by O. Maimon, L. Rokach. — Boston : Springer, 2005. — P. 321–352.
 40. Bisio I., Garibotto C., Lavagetto F., Sciarrone A. A novel IoT-based edge sensing platform for structure health monitoring // IEEE INFOCOM 2022 – Conference on Computer Communications Workshops. — 2022. — DOI: 10.1109/INFOCOMWKSHPS54753.2022.9797953.
 41. Power D. J., Sharda R., Burstein F. Decision Support Systems // Wiley Encyclopedia of Operations Research and Management Science. — Hoboken : Wiley, 2015.

42. Turban E., Sharda R., Delen D. Decision Support and Business Intelligence Systems. — 9th ed. — Upper Saddle River : Pearson, 2011.
43. Greenfield S., Chiclana F. Accuracy and complexity evaluation of defuzzification strategies for the discretised interval type-2 fuzzy set // *Int. J. Approximate Reasoning*. — 2013. — Vol. 54, No. 8. — P. 1013–1033.
44. Bezdek J. C., Pal S. K. (eds.) Fuzzy Models for Pattern Recognition: Methods that Search for Structures in Data. — New York : IEEE Press, 1992.
45. Zou Z., Bao Y., Li H., Spencer B. F., Ou J. Embedding compressive sensing-based data loss recovery algorithm into wireless smart sensors for structural health monitoring // *IEEE Sensors Journal*. — 2015. — Vol. 15, No. 2. — P. 797–808.
46. Flah M., Nunez I., Ben Chaabene W., Nehdi M. L. Machine learning algorithms in civil structural health monitoring: a systematic review // *Archives of Computational Methods in Engineering*. — 2021. — Vol. 28. — P. 2621–2643.
47. Dubois D., Prade H. Possibility Theory: Qualitative and Quantitative Aspects // *Handbook of Defeasible Reasoning and Uncertainty Management Systems*. — Vol. 1. — Boston : Kluwer, 1998. — P. 169–226.
48. Mendel J. M. Type-2 fuzzy sets and systems // *IEEE Computational Intelligence Magazine*. — 2019. — Vol. 14, No. 1. — P. 34–44.
49. Mendel J. M., Wu H. Type-2 fuzzistics for symmetric interval type-2 fuzzy sets: part 1, forward problems // *IEEE TFS*. — 2006. — Vol. 14, No. 6. — P. 781–792.
50. Wu D., Mendel J. M. Uncertainty measures for interval type-2 fuzzy sets // *Information Sciences*. — 2007. — Vol. 177, No. 23. — P. 5378–5393.
51. Wu D., Tan W. W. Computationally efficient type-reduction // *IEEE Transactions on Fuzzy Systems*. — 2020. — Vol. 28, No. 5. — P. 937–951.
52. Ditlevsen O., Madsen H. O. Structural Reliability Methods. — Chichester : Wiley, 1996. — 372 p.
53. Gustafson D. E., Kessel W. C. Fuzzy clustering with a fuzzy covariance matrix // *Proceedings of the IEEE Conference on Decision and Control*. — 1978. — P. 761–766.

54. Zonzini F., Dertimanis V., Chatzi E., Marchi L. D. System identification at the extreme edge for network load reduction in vibration-based monitoring // *IEEE Internet of Things Journal*. — 2022. — Vol. 9, No. 20. — P. 20467–20478.
55. Dave R. N. Characterization and detection of noise in clustering // *Pattern Recognition Letters*. — 1991. — Vol. 12, No. 11. — P. 657–664.
56. Power D. J. Decision support systems: a historical overview // *Handbook on Decision Support Systems 1* / ed. by F. Burstein, C. W. Holsapple. — Berlin : Springer, 2008. — P. 121–140.
57. Han J., Kamber M., Pei J. *Data Mining: Concepts and Techniques*. – Morgan Kaufmann, 2012.
58. Duda R.O., Hart P.E., Stork D.G. *Pattern Classification*. – Wiley, 2001.
59. Webb A.R. *Statistical Pattern Recognition*. – Wiley, 2002.
60. Frey B. J., Dueck D. Clustering by passing messages between data points // *Science*. — 2007. — Vol. 315, No. 5814. — P. 972–976.
61. Kaufman L., Rousseeuw P.J. *Finding Groups in Data: An Introduction to Cluster Analysis*. – Wiley, 2005.
62. Everitt B., Landau S., Leese M., Stahl D. *Cluster Analysis*. – Wiley, 2011.
63. Zadeh L.A. *Fuzzy logic and approximate reasoning*. – Synthese, 1975.
64. Klir G. J., Folger T. A. *Fuzzy Sets, Uncertainty, and Information*. — Englewood Cliffs : Prentice Hall, 1988.
65. Liang Q., Mendel J. M. Interval type-2 fuzzy logic systems: theory and design // *IEEE TFS*. — 2000. — Vol. 8, No. 5. — P. 535–550.
66. Pedrycz W., Gomide F. *Fuzzy Systems Engineering: Toward Human-Centric Computing*. – Wiley, 2007.
67. Pedrycz W., Gomide F. *An Introduction to Fuzzy Sets: Analysis and Design*. – Cambridge, MA: MIT Press, 1998.
68. Saaty T.L. *Decision Making for Leaders: The Analytic Hierarchy Process*. – Pittsburgh: RWS Publications, 2008.
69. Aggarwal C. C., Reddy C. K. (eds.) *Data Clustering: Algorithms and Applications*. — Boca Raton : CRC Press, 2013.

70. Zadeh L. A. The concept of a linguistic variable and its application to approximate reasoning — I // *Information Sciences*. — 1975. — Vol. 8, No. 3. — P. 199–249.
71. Xu R., Wunsch D. *Clustering*. — Hoboken: Wiley–IEEE Press, 2009.
72. Jain A.K. Data clustering: 50 years beyond K-means // *Pattern Recognition Letters*. — 2010. — Vol. 31(8). — P. 651–666.
73. Pedrycz W. Granular computing for data analytics: A manifesto // *IEEE/CAA Journal of Automatica Sinica*. — 2015. — Vol. 2(4). — P. 317–330.
74. Tan P.-N., Steinbach M., Kumar V. *Introduction to Data Mining*. — Boston : Addison-Wesley, 2006.
75. Karnik N. N., Mendel J. M., Liang Q. Type-2 fuzzy logic systems // *IEEE TFS*. — 1999. — Vol. 7, No. 6. — P. 643–658.
76. Pedrycz W. *Granular computing: analysis and design of intelligent systems*. — Boca Raton: CRC Press, 2013.
77. Yager R.R., Filev D.P. *Essentials of fuzzy modeling and control*. — New York: Wiley, 1994.
78. Dubois D., Prade H. *Possibility theory: an approach to computerized processing of uncertainty*. — New York: Plenum Press, 1988.
79. Hartigan J. A. *Clustering Algorithms*. — New York : Wiley, 1975.
80. Knuth D.E. *The Art of Computer Programming. Vol. 1: Fundamental Algorithms*. — 3rd ed. — Boston: Addison-Wesley, 1997.
81. Mendel J. M. Advances in type-2 fuzzy sets and systems // *Information Sciences*. — 2007. — Vol. 177, No. 1. — P. 84–110.
82. Pedrycz W., Skowron A., Kreinovich V. *Handbook of granular computing*. — Hoboken: Wiley, 2008.
83. Yager R.R. On the theory of bags // *International Journal of General Systems*. — 1986. — Vol. 13(1). — P. 23–37.
84. Novák V., Perfilieva I., Močkoř J. *Mathematical Principles of Fuzzy Logic*. — Boston : Kluwer, 1999.

85. Anderberg M. R. Cluster Analysis for Applications. — New York : Academic Press, 1973.
86. Jain A.K., Murty M.N., Flynn P.J. Data clustering: A review // ACM Computing Surveys. — 1999. — Vol. 31(3). — P. 264–323.
87. Bezdek J.C., Keller J., Krisnapuram R., Pal N.R. Fuzzy Models and Algorithms for Pattern Recognition and Image Processing. — Boston: Springer, 2005.
88. Mitchell T.M. Machine Learning. — New York: McGraw-Hill, 1997.
89. Pedrycz W. Fuzzy clustering and its applications // IEEE Transactions on Fuzzy Systems. — 1997. — Vol. 5(3). — P. 447–452.
90. John R. I. Type 2 fuzzy sets: an appraisal of theory and applications // Int. J. Uncertainty, Fuzziness and Knowledge-Based Systems. — 1998. — Vol. 6, No. 6. — P. 563–576.
91. Filippone M., Camastra F., Masulli F., Rovetta S. A survey of kernel and spectral methods for clustering // Pattern Recognition. — 2008. — Vol. 41, No. 1. — P. 176–190.
92. Berkhin P. A survey of clustering data mining techniques // Grouping Multidimensional Data. — Berlin : Springer, 2006. — P. 25–71.
93. Steinbach M., Karypis G., Kumar V. A comparison of document clustering techniques // KDD Workshop on Text Mining. — 2000.
94. Manning C.D., Raghavan P., Schütze H. Introduction to Information Retrieval. — Cambridge: Cambridge University Press, 2008.
95. Pedrycz W. Granular computing and clustering: A unified framework // Information Sciences. — 2007. — Vol. 177(18). — P. 3799–3816.
96. Xu D., Tian Y. A comprehensive survey of clustering algorithms // Annals of Data Science. — 2015. — Vol. 2. — P. 165–193.
97. Ng A. Y., Jordan M. I., Weiss Y. On spectral clustering: analysis and an algorithm // Advances in Neural Information Processing Systems (NIPS). — 2002. — Vol. 14. — P. 849–856.

98. Celebi M. E., Kingravi H. A., Vela P. A. A comparative study of efficient initialization methods for the k-means clustering algorithm // *Expert Systems with Applications*. — 2013. — Vol. 40, No. 1. — P. 200–210.
99. Witten I. H., Frank E., Hall M. A. *Data Mining: Practical Machine Learning Tools and Techniques*. — 3rd ed. — Burlington : Morgan Kaufmann, 2011.
100. Fukunaga K. *Introduction to Statistical Pattern Recognition*. — 2nd ed. — San Diego : Academic Press, 1990.
101. Alpaydin E. *Introduction to Machine Learning*. — 4th ed. — Cambridge, MA : MIT Press, 2020.
102. Dunn J. C. A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters // *Journal of Cybernetics*. — 1973. — Vol. 3, No. 3. — P. 32–57.
103. MacQueen J. Some methods for classification and analysis of multivariate observations // *Proc. 5th Berkeley Symp.*, 1967.
104. Bishop C.M. *Pattern Recognition and Machine Learning*. — Springer, 2006.
105. Pawlak Z. Rough sets // *International Journal of Computer & Information Sciences*. — 1982. — Vol. 11, No. 5. — P. 341–356.
106. Mizumoto M., Tanaka K. Some properties of fuzzy sets of type 2 // *Information and Control*. — 1976. — Vol. 31, No. 4. — P. 312–340.
107. Bezdek J. C. *Fuzzy Mathematics in Pattern Classification* : PhD dissertation. — Ithaca : Cornell University, 1973.
108. Yang M.-S. A survey of fuzzy clustering // *Mathematical and Computer Modelling*. — 1993. — Vol. 18, No. 11. — P. 1–16.
109. Zadeh L. A. Outline of a new approach to the analysis of complex systems and decision processes // *IEEE Trans. Systems, Man, and Cybernetics*. — 1973. — Vol. SMC-3, No. 1. — P. 28–44.
110. Kaufmann A. *Introduction to the Theory of Fuzzy Subsets*. — New York : Academic Press, 1975.
111. Castillo O., Melin P. *Type-2 Fuzzy Logic: Theory and Applications*. — Berlin : Springer, 2008.

112. Ross T.J. Fuzzy Logic with Engineering Applications. – Wiley, 2016.
113. Klir G.J., Yuan B. Fuzzy Sets and Fuzzy Logic. – Prentice Hall, 1995.
114. Pedrycz W. Fuzzy Control and Fuzzy Systems. — 2nd ed. — Chichester : Wiley, 1993.
115. Frigui H., Krishnapuram R. A robust competitive clustering algorithm with applications in computer vision // IEEE TPAMI. — 1999. — Vol. 21, No. 5. — P. 450–465.
116. Arthur D., Vassilvitskii S. k-means++: the advantages of careful seeding // Proc. 18th ACM-SIAM Symposium on Discrete Algorithms (SODA). — 2007. — P. 1027–1035.
117. Goldberg D.E. Genetic Algorithms in Search, Optimization, and Machine Learning. – Addison-Wesley, 1989.
118. Holland J.H. Adaptation in Natural and Artificial Systems. – MIT Press, 1992.
119. Pedrycz W., Al-Hmouz R. Fuzzy clustering with evolutionary optimization // Information Sciences. – 2011.
120. Bezdek J. C., Ehrlich R., Full W. FCM: the fuzzy c-means clustering algorithm // Computers & Geosciences. — 1984. — Vol. 10, No. 2–3. — P. 191–203.
121. Fahad A. et al. A survey of clustering algorithms for big data: taxonomy and empirical analysis // IEEE Transactions on Emerging Topics in Computing. — 2014. — Vol. 2, No. 3. — P. 267–279.
122. Krishnapuram R., Frigui H., Nasraoui O. Fuzzy and possibilistic shell clustering algorithms // IEEE TFS. — 1995. — Vol. 3, No. 1. — P. 29–43.
123. Yao Y. Y. Granular computing: basic issues and possible solutions // Proc. 5th Joint Conf. on Information Sciences. — 2000. — P. 186–189.
124. Kosko B. Fuzzy Engineering. — Upper Saddle River : Prentice Hall, 1997.
125. Wu D., Mendel J. M. A comparative study of ranking methods, similarity measures and uncertainty measures for interval type-2 fuzzy sets // Information Sciences. — 2009. — Vol. 179. — P. 1169–1192.

126. Saxena A. et al. A review of clustering techniques and developments // *Neurocomputing*. — 2017. — Vol. 267. — P. 664–681.
127. Höppner F. Fuzzy shell clustering algorithms in image processing // *IEEE TFS*. — 1997. — Vol. 5, No. 4. — P. 599–613.
128. Pal N. R., Bezdek J. C. On cluster validity for the fuzzy c-means model // *IEEE TFS*. — 1995. — Vol. 3, No. 3. — P. 370–379.
129. Fayyad U., Piatetsky-Shapiro G., Smyth P. From data mining to knowledge discovery in databases // *AI Magazine*. — 1996. — Vol. 17, No. 3. — P. 37–54.
130. Bargiela A., Pedrycz W. *Granular Computing: An Introduction*. — Boston : Kluwer, 2002.
131. Rousseeuw P. J. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis // *Journal of Computational and Applied Mathematics*. — 1987. — Vol. 20. — P. 53–65.
132. Mendel J. M. Computing with words and its relationships with fuzzistics // *Information Sciences*. — 2007. — Vol. 177, No. 4. — P. 988–1006.
133. Nguyen H. T., Walker E. A. *A First Course in Fuzzy Logic*. — 4th ed. — Boca Raton : CRC Press, 2019.
134. Passino K. M., Yurkovich S. *Fuzzy Control*. — Reading, MA : Addison-Wesley, 1998.
135. ISO 13822:2010. Bases for design of structures – Assessment of existing structures.
136. PlanRadar GmbH. Construction documentation and defect management software.
137. Archibus Inc. Integrated Workplace Management System (IWMS).
138. IBM. IBM Maximo Application Suite: Asset management documentation.
139. Eastman C., Teicholz P., Sacks R., Liston K. *BIM Handbook*. – Wiley, 2018.
140. Autodesk Inc. Revit documentation and BIM workflows.
141. Bentley Systems. iTwin platform for infrastructure digital twins.
142. Farrar C.R., Worden K. *Structural Health Monitoring: A Machine Learning Perspective*. – Wiley, 2012.

143. Kullaa J. Condition monitoring of structures using vibration measurements. – Mechanical Systems and Signal Processing, 2014.
144. Barsocchi P., Bartoli G., Betti M., Girardi M., Mammolito S., Pellegrini D. et al. Wireless sensor networks for continuous structural health monitoring of historic masonry towers // International Journal of Architectural Heritage. — 2021. — Vol. 15, No. 1. — P. 22–44.
145. Ye X.W., Dong C.Z., Liu T. A review of machine vision-based structural health monitoring. – Automation in Construction, 2016.
146. Sony S., Laventure S., Sadhu A. A literature review of next-generation smart sensing technology in structural health monitoring // Structural Control and Health Monitoring. — 2019. — Vol. 26, No. 3. — e2321.
147. Pedrycz W., Chen S.M. Granular Computing and Intelligent Systems. – Springer, 2019.
148. Newman S. Building Microservices. — 2nd ed. — Sebastopol: O'Reilly Media, 2021.
149. Fielding R.T. Architectural Styles and the Design of Network-based Software Architectures : дис. ... PhD. — Irvine: University of California, 2000.
150. Krishnapuram R., Keller J. M. A possibilistic approach to clustering // IEEE Transactions on Fuzzy Systems. — 1993. — Vol. 1, No. 2. — P. 98–110.
151. Salcedo-Sanz S., Del Ser J., Geem Z. W. An island grouping genetic algorithm for fuzzy partitioning problems // The Scientific World Journal. — 2014. — Vol. 2014. — Article ID 916371. — DOI: 10.1155/2014/916371.
152. Jebari K., Elmoujahid A., Ettouhami A. Automatic Genetic Fuzzy c-Means // Journal of Intelligent Systems. — 2020. — Vol. 29, No. 1. — P. 529–539. — DOI: 10.1515/jisys-2018-0063.
153. Yang C.-L., Kuo R. J., Chien C.-H., Quyen N. T. P. Non-dominated sorting genetic algorithm using fuzzy membership chromosome for categorical data clustering // Applied Soft Computing. — 2015. — Vol. 30. — P. 113–122.

154. Xie X. L., Beni G. A validity measure for fuzzy clustering // IEEE Transactions on Pattern Analysis and Machine Intelligence. — 1993. — Vol. 13, No. 8. — P. 841–847.
155. Gorbatyuk Ie., Terentyev O., Mishchuk D., **Bykov S.** Integrated models of forecasting reliability of decision making of the system of diagnostics of the technical condition of buildings. *Гірничі, будівельні, дорожні та меліоративні машини.* 2025. № 105. С. 73–78. DOI: <https://doi.org/10.32347/gbdmm.2025.105.0501>.
156. Tyslenko O., **Bykov S.**, Katin O. System of organization and functioning of mechanisms for ensuring the operational safety of buildings. *Шляхи підвищення ефективності будівництва.* 2025. Вип. 2(56). С. 91–104. DOI: [https://doi.org/10.32347/2707-501x.2025.56\(2\).91-104](https://doi.org/10.32347/2707-501x.2025.56(2).91-104).
157. Тисленко О. Б., **Биков С. В.** Аналітичні та діагностичні блоки контролю в цифровому середовищі оцінки структурної надійності. *Управління розвитком складних систем.* 2025. № 64. С. 259–266. DOI: <https://doi.org/10.32347/2412-9933.2025.64.259-266>.
158. Terentyev O. O., Gorbatyuk Ie. V., Tyslenko O. B., **Bykov S. V.** Methods of damage of the information system of building technical condition diagnostics. *Modern engineering and innovative technologies.* Karlsruhe, 2024. Iss. 35, part 2. P. 35–44. DOI: <https://doi.org/10.30890/2567-5273.2024-35-00-049>.
159. Terentyev O. O., Gorbatyuk Ie. V., Tyslenko O. B., **Bykov S. V.** Program and technical complex of implementation of the information system of diagnostics of the technical condition of buildings. *SWorldJournal.* 2025. Iss. 30, part 1. P. 82–96. DOI: <https://doi.org/10.30888/2663-5712.2025-30-01-004>.
160. Terentyev O. O., Gorbatyuk Ie. V., Tyslenko O. B., **Bykov S. V.** Creation of information methods in the diagnosis of industrial buildings. *Modern engineering and innovative technologies.* 2025. Iss. 38, part 1. P. 146–151. DOI: <https://doi.org/10.30890/2567-5273.2025-38-01-030>.
161. Terentyev O. O., Gorbatyuk Ie. V., Tyslenko O. B., **Bykov S. V.** Methods of diagnosing deformations of building structures. *International scientific conference*

"Technique and technology of the future '2024", Karlsruhe, Germany, 20 October 2024. P. 36–40. DOI: <https://doi.org/10.30890/2709-1783.2024-35-00-027>.

162. Терентьев О. О., Горбатюк Є. В., Тисленко О. Б., **Биков С. В.** Інтегрована модель оцінки надійності рішень для діагностики будівель. *Proceedings of the 6th International scientific and practical conference "Scientific achievements of contemporary society"*, London, United Kingdom, 10–12 January 2025. P. 280–286.

163. Терентьев О. О., Горбатюк Є. В., Тисленко О. Б., **Биков С. В.** Інформаційна технологія управління життєвим циклом будівельних конструкцій і споруд. *Modernization of today's science: experience and trends: coll. of sci. papers "SCIENTIA"*, Singapore, 24 January 2025. P. 192–195. DOI: <https://doi.org/10.36074/scientia-24.01.2025>.

164. Терентьев О. О., Горбатюк Є. В., Тисленко О. Б., **Биков С. В.** Моделювання загроз безпеки та вразливості у віртуальній мережевій системі. *Proceedings of the 5th International scientific and practical conference "Science in the modern world: innovations and challenges"*, Toronto, Canada, 23–25 January 2025. P. 169–173.

165. Terentyev O. O., Gorbatyuk Ie. V., Tyslenko O. B., **Bykov S. V.** Creating a model of mobile location. *International scientific conference "Organization of scientific research in modern conditions '2024"*, Seattle, USA, 21 March 2025. P. 25–28. DOI: <https://doi.org/10.30888/2709-2267.2025-30-00-003>.

ДОДАТКИ. ДОДАТОК 1. ДОВІДКИ ПРО ВПРОВАДЖЕННЯ

№18-н від 24.10.2025 р

ДОВІДКА

про впровадження результатів дисертаційної роботи

Бикова Сергія Володимировича

на тему: «Інформаційна технологія оцінки технічного стану об'єктів будівництва в умовах невизначеності на основі методів кластеризації»
(спеціальність 126 «Інформаційні системи та технології»).

Компанія ТОВ «МАРСТОН-ГРУП» підтверджує практичне впровадження результатів дисертаційного дослідження Бикова С.В. у діяльність підприємства під час виконання робіт із технічного обстеження, моніторингу та експертної оцінки стану будівель і споруд різного функціонального призначення.

У процесі апробації та практичного використання впроваджено розроблену автором інформаційну технологію Clustering4Build, призначену для автоматизованої кластеризації технічного стану будівельних об'єктів за умов неповноти, неоднорідності та невизначеності вихідної інформації. Запропонований підхід поєднує сучасні методи нечіткої кластеризації, математичний апарат нечітких множин першого та другого типів, генетичні алгоритми оптимізації й інструменти підтримки прийняття рішень, що дозволило істотно підвищити якість інженерного аналізу результатів технічного моніторингу.

Особливу практичну цінність для підприємства становить реалізована у межах дисертаційної роботи система автоматизованого аналізу багатопараметричних даних, яка забезпечує інтегроване опрацювання результатів інструментальних вимірювань, даних візуального обстеження, експертних висновків та інформації, отриманої від сенсорних систем моніторингу. Використання кластерних моделей дозволило здійснювати формалізовану класифікацію технічного стану конструкцій навіть у випадках, коли вихідні дані містять значний рівень невизначеності або суперечностей.

Під час впровадження позитивно зарекомендували себе комбіновані алгоритми FCM+GA та PCM+GA, у яких автоматизовано визначення центрів кластерів, ступенів належності та параметрів фазифікації. Завдяки використанню генетичних алгоритмів вдалося суттєво скоротити тривалість налаштування моделей кластеризації, підвищити стійкість обчислювального процесу та забезпечити відтворюваність результатів при повторному аналізі технічного стану аналогічних об'єктів.

Фахівці підприємства високо оцінили реалізовану автором можливість адаптивного врахування невизначеності шляхом використання інтервальних нечітких множин другого типу. Запропонований підхід дозволяє більш об'єктивно оцінювати перехідні стани конструкцій, що особливо важливо під час експлуатації будівель зі значним строком служби або після виникнення надзвичайних впливів. У практичній діяльності це забезпечило підвищення достовірності експертних висновків та зменшення впливу суб'єктивного чинника під час прийняття інженерних рішень.

Окремої уваги заслуговує реалізована у дисертації мікросервісна архітектура інформаційної технології, що забезпечує незалежне функціонування окремих програмних модулів, масштабованість системи та її інтеграцію із сучасними інформаційними середовищами технічного моніторингу. Централізоване збереження результатів аналізу в базі даних PostgreSQL із використанням формату JSONB створило можливість накопичення історичних даних спостережень, їх повторного використання та проведення порівняльного аналізу технічного стану будівель у часі. Практичне застосування модулів GA-Optimizer, QualityEvaluator, TypeReducer та підсистеми підтримки прийняття рішень (DSS) дозволило автоматизувати вибір оптимальних параметрів кластеризації, забезпечити комплексну оцінку якості сформованих кластерів та отримувати обґрунтовані рекомендації щодо подальшої експлуатації, обстеження або ремонту будівельних конструкцій.

За результатами дослідної експлуатації встановлено, що використання інформаційної технології Clustering4Build сприяло скороченню часу аналітичної обробки результатів технічного моніторингу в середньому на 22%, підвищенню достовірності класифікації технічного стану конструкцій на 18,5%, зменшенню кількості повторних експертних оцінок приблизно на 16%, а також забезпечило більш високий рівень узгодженості інженерних рішень під час проведення комплексних технічних обстежень.

Отримані результати підтвердили ефективність запропонованої інформаційної технології для практичного використання під час технічного моніторингу будівель і споруд. Її застосування дозволяє підвищити оперативність оброблення інформації, обґрунтованість експертних висновків, достовірність прогнозування технічного стану конструкцій та ефективність прийняття управлінських рішень щодо забезпечення надійності й безпечної експлуатації будівельних об'єктів. Таким чином, результати дисертаційної роботи Бикова С.В. мають вагомe науково-прикладне значення та рекомендовані до подальшого використання у практиці технічного моніторингу, діагностування й інформаційного забезпечення управління технічним станом будівель і споруд.

Директор



/_Міронов О.О./



Вих.№ 302/2 від 05.11.2025 р.

ПОВІДОМЛЕННЯ
про використання результатів дисертаційного дослідження
Бикова Сергія Володимировича

Інститут місцевого розвитку підтверджує використання результатів дисертаційного дослідження Бикова Сергія Володимировича на тему: «Інформаційна технологія оцінки технічного стану об'єктів будівництва в умовах невизначеності на основі методів кластеризації» у процесі виконання науково-прикладних досліджень, інформаційно-аналітичного супроводу та експертно-моніторингових робіт у сфері оцінювання технічного стану будівель і споруд.

У діяльності Інституту результати дослідження були використані під час удосконалення методичного забезпечення аналізу технічного стану будівельних об'єктів, для яких характерні неповнота, неоднорідність та невизначеність вихідної інформації. Особливу практичну цінність становить розроблена автором інформаційна технологія Clustering4Build, яка забезпечує інтелектуальне опрацювання багато-параметричних даних технічного моніторингу із застосуванням сучасних методів нечіткої кластеризації та генетичної оптимізації.

Під час виконання інформаційно-аналітичних досліджень використано запропоновані автором алгоритми автоматизованого групування об'єктів за технічним станом, що дозволило підвищити обґрунтованість експертно-аналітичних висновків, скоротити трудомісткість оброблення результатів моніторингових спостережень та забезпечити єдині підходи до інтерпретації різнорідних даних.

Практичне застосування отримали також розроблені програмні модулі оцінювання якості кластеризації, автоматизованого налаштування параметрів моделей та підтримки прийняття рішень, що дало можливість формувати рекомендації щодо технічного стану будівельних конструкцій із високим рівнем відтворюваності результатів. Важливою перевагою є можливість інтеграції інформаційної технології із сучасними цифровими платформами технічного моніторингу, базами даних та інформаційно-аналітичними системами.

За результатами апробації встановлено, що використання наукових розробок Бикова С.В. сприяло підвищенню оперативності опрацювання інформації, покращенню структуризації моніторингових даних, формалізації процедур експертного аналізу та підвищенню достовірності підготовки аналітичних матеріалів для прийняття технічних і організаційних рішень.

Науково-прикладний ефект впровадження полягає у розвитку сучасного методичного інструментарію інформаційного забезпечення технічного моніторингу будівельних об'єктів на основі комплексного використання нечітких множин першого і другого типів, генетичних алгоритмів та методів інтелектуальної кластеризації, що розширює можливості проведення прикладних досліджень у сфері цифровізації будівництва. Моніторинговий ефект полягає у підвищенні достовірності оцінювання технічного стану об'єктів, автоматизації оброблення результатів спостережень, забезпеченні єдиного інформаційного середовища аналізу та створенні передумов для прогнозування змін технічного стану будівель у процесі їх експлуатації.

Інститут місцевого розвитку вважає результати дисертаційної роботи Бикова Сергія Володимировича актуальними, науково обґрунтованими та такими, що мають значну практичну цінність для розвитку сучасних інформаційних технологій технічного моніторингу будівельних об'єктів, а також рекомендує їх до подальшого використання у науково-дослідній, експертно-аналітичній та інженерній діяльності.

Виконавчий директор
д.е.н., доцент



Руслан ТОРМОСОВ

№21-2 від 11.12.2025 р

Інформація про впровадження результатів дисертаційної роботи

Бикова Сергія Володимировича «Інформаційна технологія оцінки технічного стану об'єктів будівництва в умовах невизначеності на основі методів кластеризації» в практичній діяльності підприємства ТОВ «АЛЬФА-СЕРВІС».

Керівництво підприємства ТОВ «Альфа-сервіс» підтверджує використання результатів дисертаційної роботи Бикова С.В. під час виконання комплексу робіт з технічного моніторингу та оцінювання стану будівельних конструкцій виробничих і громадських будівель. У процесі виконання інженерних робіт підприємство зіткнулося із необхідністю оперативного опрацювання значних масивів різномірної інформації, отриманої під час технічного обстеження будівель. Дані інструментальних вимірювань, результати візуального контролю, показники датчиків моніторингу та експертні оцінки нерідко мали різний рівень достовірності, що ускладнювало формування об'єктивного висновку щодо фактичного технічного стану конструкцій.

Для вирішення зазначеного завдання у виробничу діяльність підприємства було впроваджено інформаційно-аналітичні рішення, розроблені Биковим С.В., основу яких становить інформаційна технологія Clustering4Build із використанням методів нечіткої кластеризації та генетичної оптимізації параметрів моделей.

Запропонована технологія дозволила автоматизувати процес групування будівельних конструкцій за рівнем їх технічного стану, враховуючи неповноту, неоднорідність та невизначеність вихідної інформації. Завдяки цьому інженери підприємства отримали можливість виконувати оцінювання значно швидше та з вищим рівнем об'єктивності, ніж за традиційного експертного аналізу. Особливе практичне значення мало застосування комбінованих алгоритмів кластеризації FCM+GA та PCM+GA, які забезпечили автоматичний добір параметрів кластерних моделей без необхідності їх тривалого ручного налаштування. Це дозволило скоротити час підготовки експертних висновків та забезпечити стабільність результатів при повторному аналізі об'єктів.

Важливою перевагою впровадження стало використання інтервальних нечітких множин другого типу для аналізу складних випадків, коли інформація про технічний стан конструкцій була неповною або суперечливою. Такий підхід значно підвищив достовірність класифікації перехідних станів конструкцій і дозволив мінімізувати ризик помилкової інтерпретації результатів обстежень. Фахівці підприємства також позитивно оцінили модульну структуру розробленої інформаційної технології. Її інтеграція із корпоративною базою даних та інформаційним середовищем технічного моніторингу не потребувала істотної модернізації програмного забезпечення підприємства, що забезпечило швидке впровадження запропонованих рішень у виробничий процес.

За результатами практичного використання розробок Бикова С.В. підприємством отримано такі основні результати:

- автоматизовано процес кластеризації результатів технічного моніторингу будівель;
- скорочено тривалість аналітичної обробки даних приблизно на 21%;
- підвищено достовірність визначення технічного стану конструкцій у середньому на 17%;
- зменшено вплив суб'єктивного експертного чинника під час прийняття інженерних рішень;
- забезпечено централізоване накопичення результатів обстежень для подальшого аналізу та прогнозування.

Функціональний ефект від упровадження полягає у створенні єдиного цифрового середовища аналізу результатів технічного моніторингу, автоматизації класифікації будівельних конструкцій, формуванні рекомендацій для експертів та підтримці прийняття технічно обґрунтованих рішень щодо подальшої експлуатації об'єктів. Технічний ефект полягає у підвищенні точності кластеризації технічного стану конструкцій, забезпеченні стійкої роботи інформаційної технології за умов неповноти та неоднорідності даних, скороченні часу опрацювання результатів обстежень, покращенні відтворюваності розрахунків та підвищенні надійності інформаційного забезпечення технічного моніторингу будівель.

Практична апробація підтвердила, що результати дисертаційного дослідження Бикова С.В. є ефективним науково-прикладним інструментарієм для цифрової трансформації процесів технічного моніторингу будівель і споруд та можуть бути рекомендовані до широкого використання підприємствами будівельної галузі.

Заступник директора



Кравченко А.М.

ДОДАТОК 2. СПИСОК ОПУБЛІКОВАНИХ ПРАЦЬ ЗА ТЕМОЮ ДИСЕРТАЦІЇ

Статті у наукових фахових виданнях України категорії «Б»:

4. Gorbatyuk Ie., Terentyev O., Mishchuk D., **Bykov S.** Integrated models of forecasting reliability of decision making of the system of diagnostics of the technical condition of buildings. *Гірничі, будівельні, дорожні та меліоративні машини*. 2025. № 105. С. 73–78. DOI: <https://doi.org/10.32347/gbdmm.2025.105.0501>.

5. Tyslenko O., **Bykov S.**, Katin O. System of organization and functioning of mechanisms for ensuring the operational safety of buildings. *Шляхи підвищення ефективності будівництва*. 2025. Вип. 2(56). С. 91–104. DOI: [https://doi.org/10.32347/2707-501x.2025.56\(2\).91-104](https://doi.org/10.32347/2707-501x.2025.56(2).91-104).

6. Тисленко О. Б., **Биков С. В.** Аналітичні та діагностичні блоки контролю в цифровому середовищі оцінки структурної надійності. *Управління розвитком складних систем*. 2025. № 64. С. 259–266. DOI: <https://doi.org/10.32347/2412-9933.2025.64.259-266>.

Статті в закордонних періодичних виданнях, прирівняних до фахових видань категорії «Б» (за умови індексації/цитування):

7. Terentyev O. O., Gorbatyuk Ie. V., Tyslenko O. B., **Bykov S. V.** Methods of damage of the information system of building technical condition diagnostics. *Modern engineering and innovative technologies*. Karlsruhe, 2024. Iss. 35, part 2. P. 35–44. DOI: <https://doi.org/10.30890/2567-5273.2024-35-00-049>.

8. Terentyev O. O., Gorbatyuk Ie. V., Tyslenko O. B., **Bykov S. V.** Program and technical complex of implementation of the information system of diagnostics of the technical condition of buildings. *SWorldJournal*. 2025. Iss. 30, part 1. P. 82–96. DOI: <https://doi.org/10.30888/2663-5712.2025-30-01-004>.

9. Terentyev O. O., Gorbatyuk Ie. V., Tyslenko O. B., **Bykov S. V.** Creation of information methods in the diagnosis of industrial buildings. *Modern engineering and innovative technologies*. 2025. Iss. 38, part 1. P. 146–151. DOI: <https://doi.org/10.30890/2567-5273.2025-38-01-030>.

Наукові праці апробаційного характеру

(тези доповідей на науково-практичних конференціях)

12. Terentyev O. O., Gorbatyuk Ie. V., Tyslenko O. B., **Bykov S. V.** Methods of diagnosing deformations of building structures. *International scientific conference "Technique and technology of the future '2024"*, Karlsruhe, Germany, 20 October 2024. P. 36–40. DOI: <https://doi.org/10.30890/2709-1783.2024-35-00-027>.

13. Терентьев О. О., Горбатьюк Є. В., Тисленко О. Б., **Биков С. В.** Інтегрована модель оцінки надійності рішень для діагностики будівель. *Proceedings of the 6th International scientific and practical conference "Scientific achievements of contemporary society"*, London, United Kingdom, 10–12 January 2025. P. 280–286.
14. Терентьев О. О., Горбатьюк Є. В., Тисленко О. Б., **Биков С. В.** Інформаційна технологія управління життєвим циклом будівельних конструкцій і споруд. *Modernization of today's science: experience and trends: coll. of sci. papers "SCIENTIA"*, Singapore, 24 January 2025. P. 192–195. DOI: <https://doi.org/10.36074/scientia-24.01.2025>.
15. Терентьев О. О., Горбатьюк Є. В., Тисленко О. Б., **Биков С. В.** Моделювання загроз безпеки та вразливості у віртуальній мережевій системі. *Proceedings of the 5th International scientific and practical conference "Science in the modern world: innovations and challenges"*, Toronto, Canada, 23–25 January 2025. P. 169–173.
16. Terentyev O. O., Gorbatyuk Ie. V., Tyslenko O. B., **Bykov S. V.** Creating a model of mobile location. *International scientific conference "Organization of scientific research in modern conditions '2024"*, Seattle, USA, 21 March 2025. P. 25–28. DOI: <https://doi.org/10.30888/2709-2267.2025-30-00-003>.

ДОДАТОК 3.

При визначенні інтервальних первинних ФП об'єкта X_i використовуються «нижня» і «верхня» інтервальні ФП з двома різними значеннями фазифікатора т: m_1 і m_2 [16]:

$$\bar{u}_j(x_i) = \begin{cases} \frac{1}{\sum_{t=1}^c \left(\frac{d_{ji}}{d_{ti}}\right)^{\frac{2}{m_1-1}}}, \text{ якщо } \frac{1}{\sum_{t=1}^c \frac{d_{ji}}{d_{ti}}} > \frac{1}{c} \\ \frac{1}{\sum_{t=1}^c \left(\frac{d_{ji}}{d_{ti}}\right)^{\frac{2}{m_2-1}}}, \text{ якщо } \frac{1}{\sum_{t=1}^c \frac{d_{ji}}{d_{ti}}} \leq \frac{1}{c} \end{cases} \quad (3.3)$$

$$\underline{u}_j(x_i) = \begin{cases} \frac{1}{\sum_{t=1}^c \left(\frac{d_{ji}}{d_{ti}}\right)^{\frac{2}{m_1-1}}}, \text{ якщо } \frac{1}{\sum_{t=1}^c \frac{d_{ji}}{d_{ti}}} \leq \frac{1}{c} \\ \frac{1}{\sum_{t=1}^c \left(\frac{d_{ji}}{d_{ti}}\right)^{\frac{2}{m_2-1}}}, \text{ якщо } \frac{1}{\sum_{t=1}^c \frac{d_{ji}}{d_{ti}}} > \frac{1}{c} \end{cases} \quad (3.4)$$

Також «нижня» і «верхня» ФП можуть бути визначені як [117]:

$$\bar{u}_j(x_i) = \begin{cases} \frac{1}{\sum_{t=1}^c \left(\frac{d_{ji}}{d_{ti}}\right)^{\frac{2}{m_1-1}}}, \text{ якщо } \frac{1}{\sum_{t=1}^c \left(\frac{d_{ji}}{d_{ti}}\right)^{\frac{2}{m_1-1}}} > \frac{1}{\sum_{t=1}^c \left(\frac{d_{ji}}{d_{ti}}\right)^{\frac{2}{m_2-1}}} \\ \frac{1}{\sum_{t=1}^c \left(\frac{d_{ji}}{d_{ti}}\right)^{\frac{2}{m_2-1}}}, \text{ якщо } \frac{1}{\sum_{t=1}^c \left(\frac{d_{ji}}{d_{ti}}\right)^{\frac{2}{m_1-1}}} \leq \frac{1}{\sum_{t=1}^c \left(\frac{d_{ji}}{d_{ti}}\right)^{\frac{2}{m_2-1}}} \end{cases} \quad (3.5)$$

$$\underline{u}_j(x_i) = \begin{cases} \frac{1}{\sum_{t=1}^c \left(\frac{d_{ji}}{d_{ti}}\right)^{\frac{2}{m_1-1}}}, \text{ якщо } \frac{1}{\sum_{t=1}^c \left(\frac{d_{ji}}{d_{ti}}\right)^{\frac{2}{m_1-1}}} \leq \frac{1}{\sum_{t=1}^c \left(\frac{d_{ji}}{d_{ti}}\right)^{\frac{2}{m_2-1}}} \\ \frac{1}{\sum_{t=1}^c \left(\frac{d_{ji}}{d_{ti}}\right)^{\frac{2}{m_2-1}}}, \text{ якщо } \frac{1}{\sum_{t=1}^c \left(\frac{d_{ji}}{d_{ti}}\right)^{\frac{2}{m_1-1}}} > \frac{1}{\sum_{t=1}^c \left(\frac{d_{ji}}{d_{ti}}\right)^{\frac{2}{m_2-1}}} \end{cases} \quad (3.6)$$

Використання фазифікаторів m_1 і m_2 приводить до різних цільових функцій, які повинні бути мінімізовані [16]:

$$J_{m_1}(U, V) = \sum_{j=1}^c \sum_{i=1}^n \left(u_j(x_i)\right)^{m_1} \cdot d_{ji}^2 \quad (3.7)$$

$$J_{m_2}(U, V) = \sum_{j=1}^c \sum_{i=1}^n \left(u_j(x_i)\right)^{m_2} \cdot d_{ji}^2 \quad (3.8)$$

ДОДАТОК 4.

Застосування дозволяє автоматизувати процедуру “зниження типу” без необхідності повного перебору вкладених НМТ1. Для пошуку лівого (мінімального) значення V_L і правого (максимального) значення V_R інтервального центру кластера попередньо виконується впорядкування і-х індексів множини об’єктів ($i = \overline{1, n}$) за зростанням значень оцінок по кожному елементу моніторингу $l(l = \overline{1, q})$ [16].

Інтервальна НМТ1, що відповідає інтервальному центру, записується у вигляді [16]:

$$v_j = 1,0 / [v_L, v_R] \quad (3.10)$$

В результаті дефазифікації інтервального НМТ1 знаходиться чітке значення для центру V_j . (рисунок 3.11):

$$V_j = \frac{V_L + V_R}{2} \quad (3.11)$$

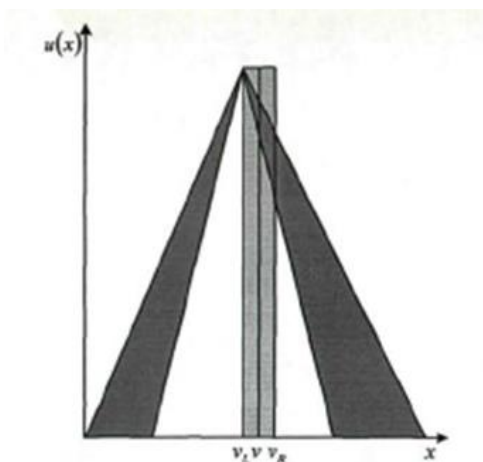


Рисунок 3.11. Приклади чіткого і інтервального центрів на основі НМТ1

ДОДАТОК 5.

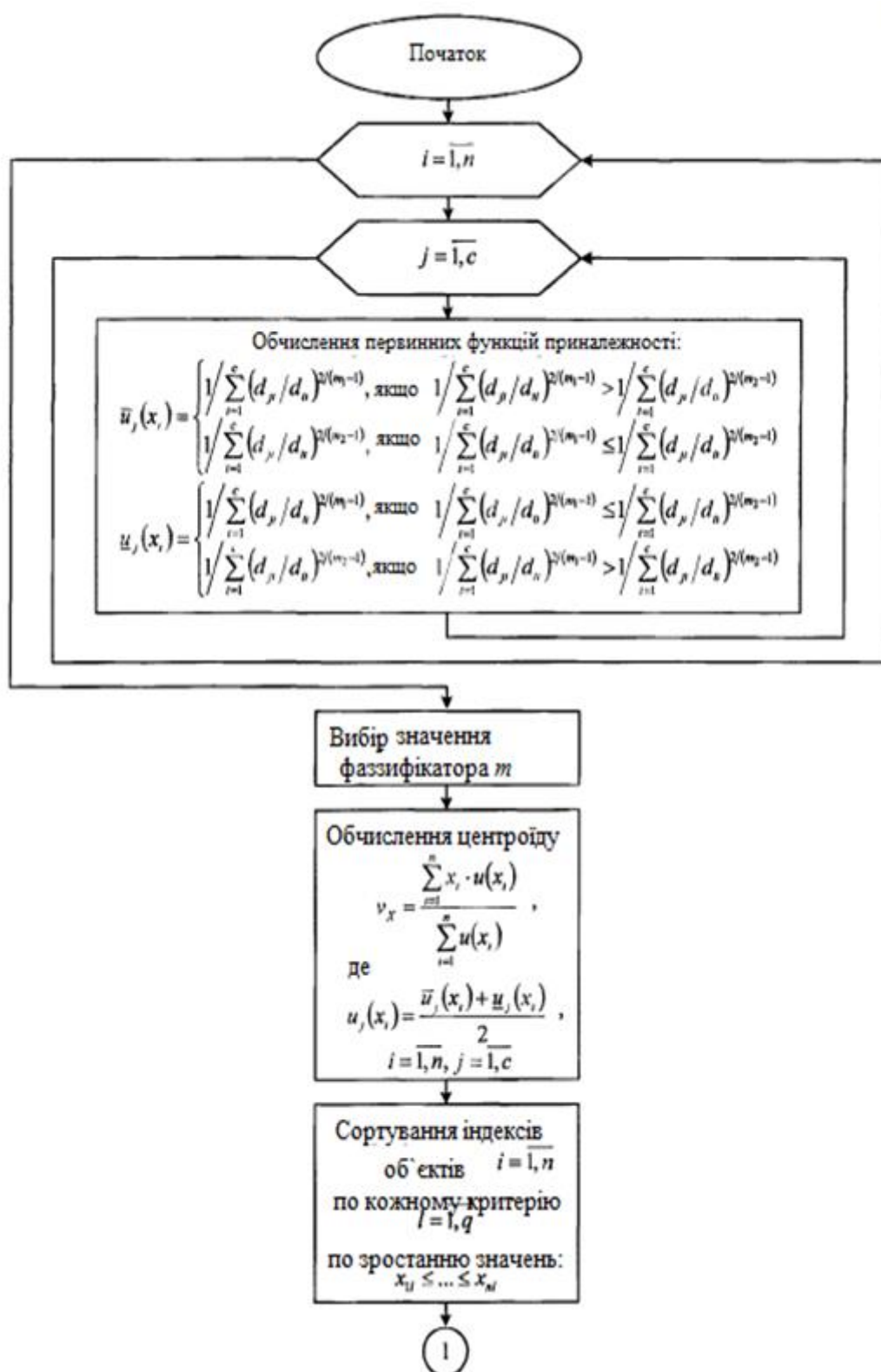


Рисунок 3.12. Схема ітераційного алгоритму Карніка-Менделя для пошуку максимуму v_R центра кластеру v_j (початок)

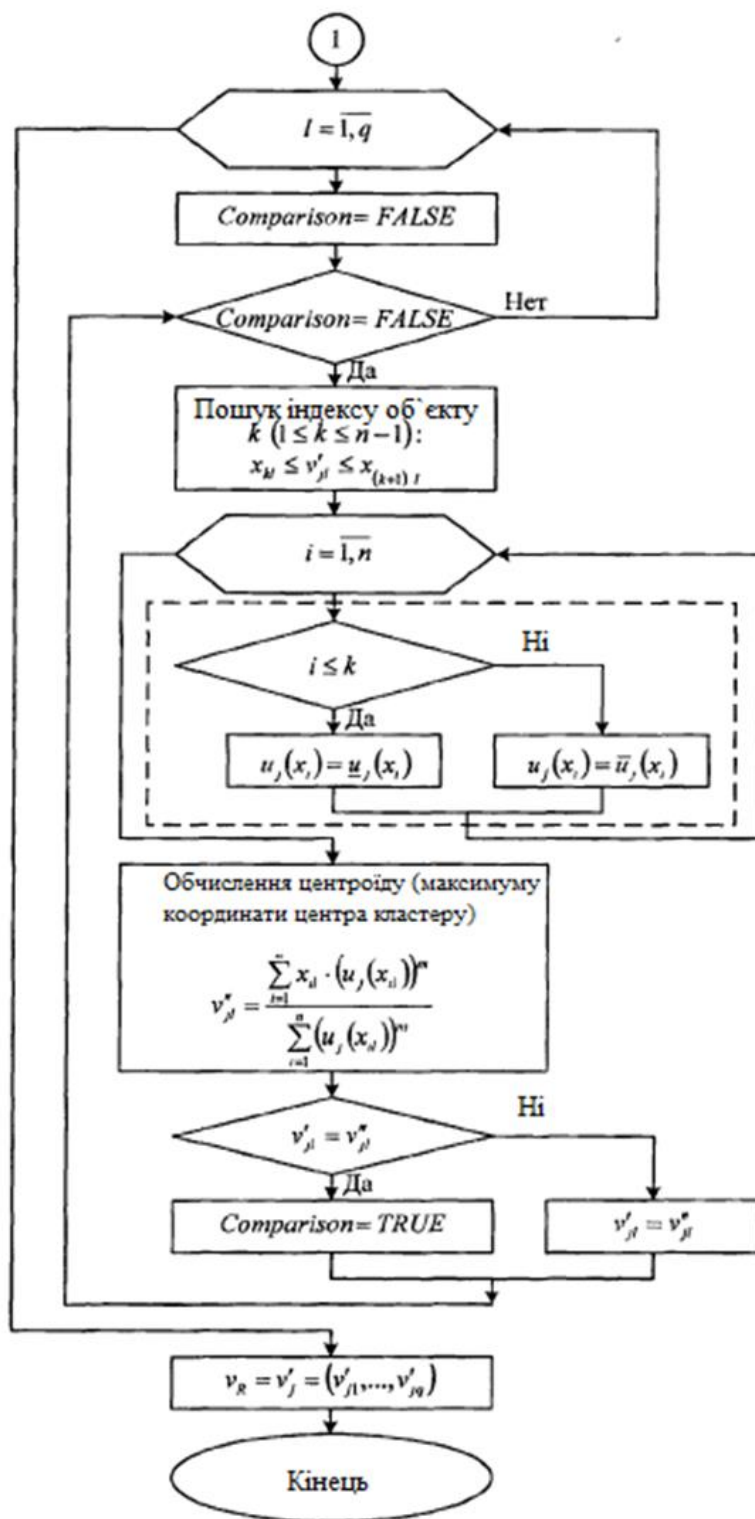


Рисунок 3.13. Схема ітераційного алгоритму Карника-Менделя для пошуку максимуму v_R центра кластеру v_j (закінчення)

ДОДАТОК 6

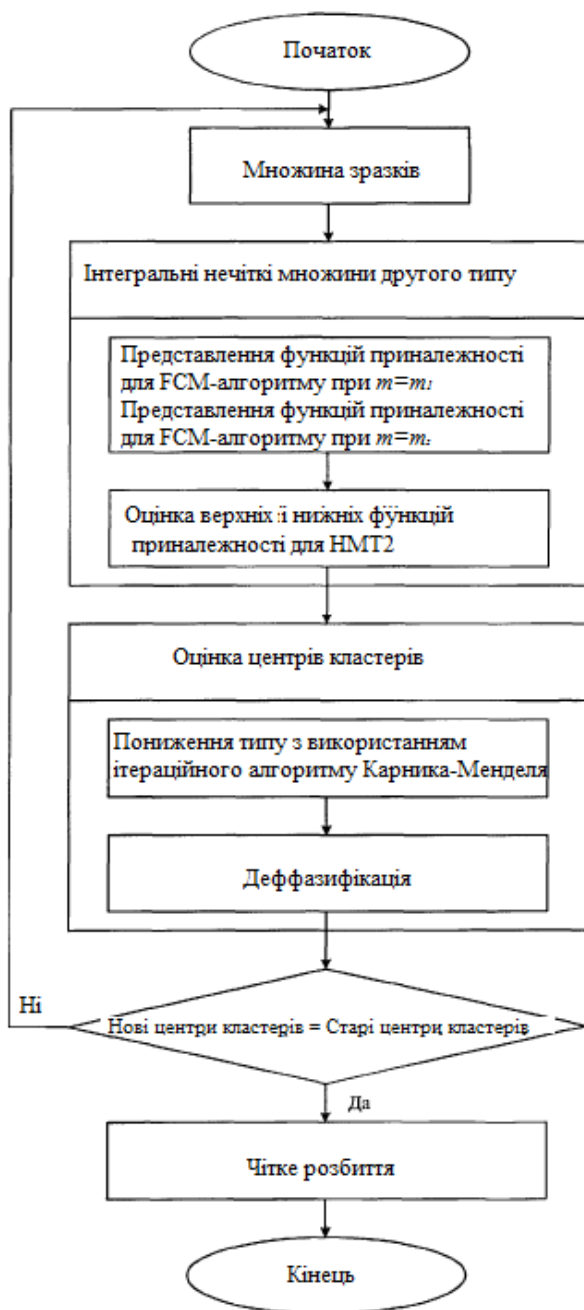


Рисунок 3.14. Загальна схема FCM-алгоритму на основі ІНМТ2

Додаток 7. Таблиця 4.3. Функціональна реалізація пакетів прикладних програм Clustering4Build (мікросервісний рівень). Пакет 2. Оцінка якості та моніторинг

Пакет 2. Оцінка якості та монітори нг	POST /quality/evaluate (QualityEvaluator)	Обчислен ня індексів H, FH, Sph	run_id або результати кластериза ції	{H, FH, Sph}	quality_metr ics
	GET /quality/metrics/{run _id}	Повторни й доступ до показникі в якості	run_id	{H, FH, Sph}	—
	GET /dss/dashboard/{run _id} (DSS)	Формуван ня аналітичн ої панелі	run_id	Агрегован ий JSON	—
	GET /dss/heatmap/{run_i d}	Генерація теплових карт	run_id	Дані теплової карти	—
	POST /dss/report	Генерація аналітичн ого звіту	run_id, шаблон	report_id	reports

Таблиця 4.4. Функціональна реалізація пакетів прикладних програм Clustering4Build (мікросервісний рівень). Пакет 3. Адаптивна оптимізація

Пакет 3. Адаптивна оптимізаці я	POST /optimizer/g a/start (GA- Optimizer)	Ініціалізація ГА	Межі m , m_1 , m_2 , η ; параметр и ГА	ga_run_id	ga_runs
	POST /optimizer/g a/evaluate	Оцінка хромосоми	Параметр и	Значення fitness	ga_populatio n

		(FCM + Quality)	хромосом и		
	POST /optimizer/ga/iterate	Формування нового покоління	Поточна популяція	Оновлена популяція	ga_population
	GET /optimizer/ga/best/{ga_run_id}	Отримання оптимальних параметрів	ga_run_id	m, m ₁ , m ₂ , η, fitness	model_configs

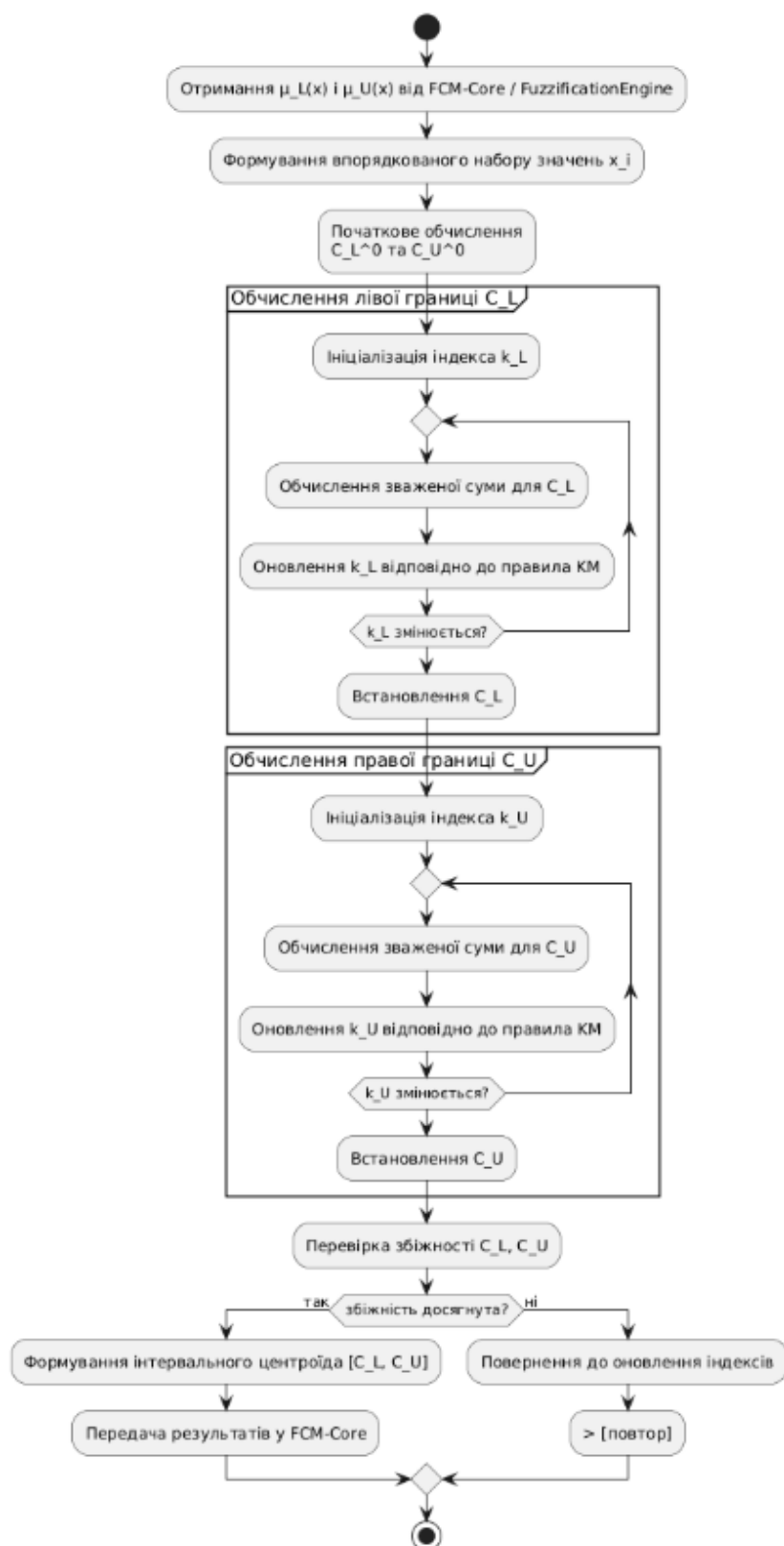
Таблиця 4.5. Функціональна реалізація пакетів прикладних програм Clustering4Build (мікросервісний рівень). Пакет 3 (IHMT2), Пакет 4. Інтерфейс і візуалізація, Пакет 5. Зберігання даних

Пакет 3 (IHMT2)	POST /type-reduction/km (TypeReducer)	Алгоритм Карника– Менделя	$\mu L(x)$, $\mu U(x)$, x_i	[C_L, C_U]	type_reduction_results
	POST /type-reduction/centroids /batch	Пакетне зниження типу	run_id	centroids_ref	type_reduction_results
Пакет 4. Інтерфейс і візуалізація	POST /pipeline/run (API Gateway)	Оркестрація повного конвеєра	Дані + параметри	run_id	clustering_runs
	GET /pipeline/status/{run_id}	Моніторинг стану виконання	run_id	Статус / лог	execution_logs
Пакет 5. Зберігання даних	PostgreSQL DataRepository	Централізоване зберігання	Результати сервісів	JSONB- структури	Усі таблиці БД

Додаток 8. Activity Diagram модуля FuzzificationEngine



Додаток 9. Activity Diagram для модуля TypeReducer



Додаток 10 . Activity Diagram для модуля QualityEvaluator

