

ДУБЧАК ЛЕСЯ

Західноукраїнський національний університет

<https://orcid.org/0000-0003-3743-2432>e-mail: dlo@wunu.edu.ua

НЕЧІТКЕ КЛАСТЕРНЕ МОДЕЛЮВАННЯ ЕНЕРГЕТИЧНОГО ВИХОДУ БІОГАЗУ

У статті розглянуто застосування методу нечіткої кластеризації Fuzzy C-means для аналізу впливу сировинного складу на енергетичний вихід біогазових установок. Запропоновано підхід до нормалізації вхідних даних, розрахунку функцій належності та прогнозування кількості біогазу за допомогою зваженого середнього значення виходу по кластерах. У роботі використано дані про масу компонентів сировини (гній, зерно, буряк, кукурудза, послід) та результуючу кількість виробленої енергії. Результати демонструють ефективність запропонованої моделі в умовах невизначеності, зокрема при відсутності точного історичного аналога. Методика дозволяє виявляти оптимальні комбінації сировини, оцінювати потенціал нових сумішей та ідентифікувати неефективні режими роботи установок. Це відкриває перспективи інтеграції моделі в інтелектуальні системи моніторингу та адаптивного керування процесом біогазогенерації.

Ключові слова: нечітка кластеризація, біогазова установка, Fuzzy C-means, сировина, відновлювана енергетика.

DUBCHAK LESIA

West Ukrainian National University

FUZZY CLUSTER MODELING OF BIOGAS ENERGY OUTPUT

This study presents a fuzzy clustering-based approach to modeling and predicting the energy output of biogas plants depending on the composition of organic feedstock. The main objective of the research was to apply the Fuzzy C-means (FCM) clustering algorithm to categorize various feedstock combinations and use these cluster affiliations to forecast biogas yield. The dataset used in the study includes normalized values for several types of biomass—namely, corn silage, sugar beet residue, grain, manure, and animal slurry—measured in kilograms, with the corresponding output energy expressed in kWh. The input data underwent min-max normalization, ensuring that all feature values were scaled into the [0, 1] interval. Missing or infinite values were replaced by feature-wise mean values. After preprocessing, the FCM algorithm was applied in the MATLAB environment, and the dataset was clustered into three groups based on the feedstock composition. The centroids of the clusters corresponded to normalized average energy outputs of 0.5459, 0.4204, and 0.5224 respectively. The fuzzy membership values allowed each sample to partially belong to multiple clusters, capturing the intrinsic vagueness of the feedstock properties. Using these memberships, a fuzzy phase-based model was developed to estimate the biogas output for any given input combination by computing a weighted average of the cluster outputs. The predictive performance of the model, evaluated using root mean square error and mean absolute error, achieved values of 0.35 and 0.23 respectively in normalized units, indicating satisfactory accuracy in modeling the relationship between feedstock composition and biogas energy yield. Moreover, the fuzzy approach enables adaptability and robustness in cases where the feedstock composition deviates from previously observed combinations, allowing real-time evaluation even in the absence of identical historical samples. The research also emphasizes the potential of fuzzy clustering to detect suboptimal or anomalous operating conditions through shifts in cluster memberships over time. This opens up opportunities for integrating the model into real-time control and monitoring systems of biogas facilities. Despite its advantages, the method remains sensitive to the initial number of clusters and requires complete input matrices, which should be addressed in future studies through hybrid fuzzy-neural or deep learning enhancements.

Keywords: fuzzy clustering, biogas plant, Fuzzy C-means, raw materials, renewable energy.

Стаття надійшла до редакції / Received 24.10.2025

Прийнята до друку / Accepted 15.11.2025

Постановка проблеми у загальному вигляді

та її зв'язок із важливими науковими чи практичними завданнями

Біогазові установки відіграють важливу роль у сучасній енергетиці як технологія, що дозволяє перетворювати органічні відходи на корисну енергію [1, 2]. Ці установки функціонують на основі процесу анаеробного зброджування, під час якого мікроорганізми у безкисневому середовищі розкладають органічну речовину з утворенням біогазу — суміші метану, вуглекислого газу та незначної кількості домішок, таких як сірководень. Отриманий біогаз може бути використаний для генерації електроенергії та тепла, або очищений до якості природного газу. Крім цього, побічним продуктом зброджування є дигестат — органічне добриво, придатне для аграрного використання.

Біогазові установки мають низку переваг [3]. Вони дозволяють зменшити залежність від викопного палива та забезпечують локальну енергетичну незалежність сільських господарств або комунальних підприємств. Такі системи сприяють скороченню викидів парникових газів, ефективно утилізують органічні відходи, запобігаючи забрудненню довкілля, а також дають змогу отримати додатковий прибуток у вигляді електроенергії, тепла та органічних добрив. Проте біогазові комплекси мають і свої обмеження. Насамперед це висока вартість будівництва та необхідність кваліфікованого обслуговування. Окрім того, економічна доцільність залежить від безперебійного постачання сировини та стабільної роботи системи. При порушенні санітарних вимог чи неефективному управлінні можливі негативні екологічні наслідки.

До біогазових установок як сировину подаються різні види органічних залишків. Найпоширенішими є гній великої рогатої худоби, свиней, пташиний послід, а також силос кукурудзи, залишки цукрового буряка,

зернові культури, харчові та сільськогосподарські відходи. Часто застосовується співферментація кількох типів сировини з метою підвищення енергетичного виходу та оптимізації мікробіологічного середовища у ферментері [4]. Тому дослідження виходу біогазу з відповідної сировини є актуальною задачею.

Аналіз досліджень та публікацій

У порівнянні з традиційними методами класифікації та регресійного прогнозування (наприклад, жорсткою кластеризацією k-середніх, лінійною регресією або деревами рішень), використання нечіткої кластеризації Fuzzy C-Means забезпечує суттєві переваги при роботі з технологічними даними біогазових установок. Зокрема, вона дає змогу ефективно моделювати перехідні та змішані стани сировини, що особливо актуально в умовах варіативного хімічного складу органічних відходів та нестабільності технологічного процесу [5-8].

Жорсткі методи класифікації зазвичай призводять до втрати інформації через необхідність жорсткого віднесення кожного зразка до одного кластеру. Натомість нечітка кластеризація дозволяє відобразити розмиту природу сировинних характеристик, тим самим покращуючи точність прогнозування без необхідності складного параметричного налаштування. Хоча точність моделі FCM дещо поступається високоточним нейромережевим підходам (наприклад, ANFIS або LSTM), її інтерпретованість, стабільність та простота реалізації роблять її конкурентоспроможною в умовах обмежених обчислювальних ресурсів або при нестачі даних.

У дослідженнях [9] та [10], ключова увага зосереджена на застосуванні нечітких кластеризаційних алгоритмів у сфері відновлюваної енергетики, зокрема для аналізу даних біоенергетичних систем та покращення діагностики технічних процесів. Основна мета обох робіт — подолання проблеми невизначеності у складних системах на основі нечіткої логіки, яка дозволяє враховувати нечіткість, багатозначність і неточність вихідних даних.

У статті Zenani et al. [11], представлено огляд способів застосування нечіткої логіки до біоенергетичних технологій. Автори наголошують, що процеси отримання біоенергії (наприклад, газифікація, піроліз, анаеробне зброджування) характеризуються значною кількістю змінних та непевністю. Нечітка логіка дозволяє моделювати такі системи на основі експертних знань і нечітких правил, що суттєво покращує точність прийняття рішень. Зокрема, дослідження вказують, що комбінація нечіткої логіки з іншими методами (наприклад, оптимізацією, машинним навчанням або моделюванням сценаріїв) дає змогу значно підвищити ефективність таких систем. Також наголошено, що нечітке моделювання особливо корисне для вибору сировини, оцінки продуктивності і прогнозування виробництва енергії з біомаси.

З іншого боку, робота Li et al. [12] зосереджена на вдосконаленні алгоритму нечіткої кластеризації c-середніх (FCM) для діагностики несправностей трансформаторів за допомогою аналізу розчинених газів. Пропонується модифікація класичної функції приналежності, яка замінює традиційне обернено-пропорційне відношення на експоненційну форму. Це усуває проблему локальних екстремумів, що характерні для класичного підходу, і забезпечує монотонне зменшення приналежності зі збільшенням відстані до центру кластера. Таким чином, підвищується точність класифікації навіть при значному рівні шуму чи перешкод у даних.

Обидва дослідження демонструють, що нечітка кластеризація має суттєві переваги: вона дозволяє працювати з неточною, розмитою або частково доступною інформацією, моделює людське міркування і легко адаптується до зміни умов. Основним недоліком є підвищена чутливість до початкових параметрів (у класичних FCM), складність у виборі параметрів (кількість кластерів, функція належності), а також підвищені обчислювальні витрати при інтеграції з іншими методами.

У контексті біогазових установок ці підходи можуть бути використані для класифікації сировини за потенціалом виробництва енергії, прогнозування обсягів метану, адаптивного керування процесами ферментації або оптимізації параметрів для досягнення максимальної ефективності. Таким чином, нечітка кластеризація виступає універсальним інструментом аналізу в біоенергетиці, особливо у випадках, коли традиційні статистичні методи не справляються з неповнотою чи невизначеністю даних.

Кластеризація за методом Fuzzy C-means (FCM) є ефективним інструментом для виявлення прихованих закономірностей між типами сировини та виходом енергії на біогазових установках. Завдяки здатності цього методу формувати нечіткі кластери, де кожен зразок може належати до кількох груп одночасно, забезпечується висока гнучкість в обробці варіативних і частково невизначених даних.

Формулювання цілей статті

Метою роботи є: дослідження виходу біогазу відповідно до використаної сировини на основі нечіткої кластеризації. На основі отриманих кластерів можна прогнозувати вихід енергії для нових комбінацій сировини шляхом обчислення зваженого середнього енергетичного виходу для всіх кластерів, до яких належить зразок. Це дозволяє здійснювати оцінку продуктивності установки навіть у разі відсутності точного історичного аналога конкретного складу сировини.

Кластеризація також дозволяє оптимізувати подачу сировини шляхом виявлення тих комбінацій, які забезпечують максимальний вихід. Крім того, відстеження зміни кластерної приналежності з часом дає можливість виявляти відхилення в роботі установки, пов'язані з погіршенням якості сировини чи технічними несправностями.

Виклад основного матеріалу

Перед проведенням кластерного аналізу методом нечіткого розподілу (Fuzzy C-means) необхідно забезпечити адекватну попередню обробку вхідних даних, зокрема їх нормалізацію. Це дозволяє уникнути впливу різної шкали ознак на результати кластеризації та забезпечити стабільність алгоритму.

У даному дослідженні було використано міні-максну нормалізацію, що зводить усі значення кожного параметра до інтервалу $[0, 1]$. Цей метод зберігає відносні пропорції між значеннями та є оптимальним для нечітких моделей.

Для кожного числового елемента x_{ij} вхідної матриці X , де $i = 1, 2, \dots, N$ – номер зразка, а $j = 1, 2, \dots, M$ – номер ознаки, нормалізоване значення x'_{ij} обчислюється за формулою:

$$x'_{ij} = \frac{x_{ij} - \min(x_j)}{\max(x_j) - \min(x_j)},$$

де $\min(x_j)$ – мінімальне значення j -ї ознаки серед усіх ознак, $\max(x_j)$ – максимальне значення j -ї ознаки.

У реальних даних зустрічаються пропущені (NaN) або нескінченні (Inf) значення, які можуть порушити коректність нормалізації та кластеризації. Їх обробка здійснюється наступним чином: спочатку для кожної ознаки обчислюється середнє значення $\bar{x}_{ij}$ (без урахування NaN або Inf), а потім кожне NaN або Inf у відповідному стовпчику замінюється на це середнє:

$$x_{ij} = \begin{cases} \bar{x}_{ij}, & \text{якщо } x_{ij} \in \{NaN, Inf\} \\ x_{ij}, & \text{інакше} \end{cases}.$$

Таким чином, нормалізація даних дозволяє забезпечити масштабну однорідність ознак, знизити ризик зміщення центроїдів кластерів та гарантувати ефективність подальшого аналізу методом Fuzzy C-means.

Після нормалізації вхідних параметрів, які описують характеристики сировини (зокрема, маса гною, силосу, зерна, буряка тощо), було виконано кластеризацію із застосуванням алгоритму нечіткої кластеризації C-середніх. На відміну від традиційної жорсткої кластеризації, у FCM кожен об'єкт може належати до кількох кластерів одночасно з різним ступенем належності. Це дає змогу враховувати розмитість меж між станами технологічного процесу, особливо важливу у випадку неоднорідної біосировини.

Задамо вхідні значення у вигляді матриці:

$$X = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & x_{2n} \\ \dots & \dots & \dots & \dots \\ x_{N1} & x_{N2} & \dots & x_{Nn} \end{bmatrix},$$

де $X \in R^{N \times n}$ – нормалізована матриця вхідних даних з N об'єктів по n ознак.

Функція належності визначає, наскільки сильно певний вхідний зразок x_i пов'язаний із кластером j .

Вона розраховується за формулою:

$$\mu_{ij} = (\sum_{k=1}^c (\frac{\|x_i - v_j\|}{\|x_i - v_k\|})^{\frac{2}{m-1}})^{-1}.$$

Наприклад, при наявності трьох кластерів сировини — низькопотенційної, середньої та високопродуктивної — один і той самий зразок може одночасно на 60% належати до першого класу, на 30% — до другого і на 10% — до третього. Це дозволяє будувати адаптивні моделі прогнозування виходу енергії, які враховують не тільки самі значення вхідних параметрів, але і їх контекстну приналежність до певних станів.

Далі алгоритм FCM мінімізує функцію знаходження центру кластера:

$$J_m = \sum_{i=1}^N \sum_{j=1}^c \mu_{ij}^m \cdot \|x_i - v_j\|^2,$$

де μ_{ij} – ступінь належності об'єкта x_i до кластера j ;

m – коефіцієнт розмитості;

$v_j \in R^n$ – центр кластера j ;

c – кількість кластерів;

$\|\cdot\|$ – Евклідова норма.

Результатом кластеризації є матриця належностей $M = [\mu_{ij}]$ та центрів кластерів $V = \{v_1, v_2, \dots, v_c\}$.

На основі отриманих ступенів належності було реалізовано нечітке фазове моделювання виходу енергії, тобто, прогноз біогазу визначався як зважене середнє виходів по всіх кластерах з урахуванням їх належностей:

$$\hat{y}_i = \sum_{j=1}^c u_{ij} \cdot \bar{y}_j,$$

де $\hat{y}_i$ – прогнозований вихід біогазу для i -го зразка, $\bar{y}_j$ – середній вихід біогазу у j -му кластері.

Вхідні дані, які опрацьовувались включають кількість (у кілограмах) кукурудзи, буряка, зерна, посліду та гною, які використовуються на біогазовому підприємстві. Вихідна змінна — це кількість виробленої енергії (у кВт·год).

Проведено Fuzzy C-mean кластеризацію вхідних даних в середовищі Matlab. Результат кластеризації показав, що опрацьовані дані розділено на 3 кластери (рис.1), які перетинаються.

Середній вихід біогазу для кожного кластера (кВт·год, нормалізовано):

Кластер 1: 0.5459 (нормалізований)

Кластер 2: 0.4204 (нормалізований)

Кластер 3: 0.5224 (нормалізований)

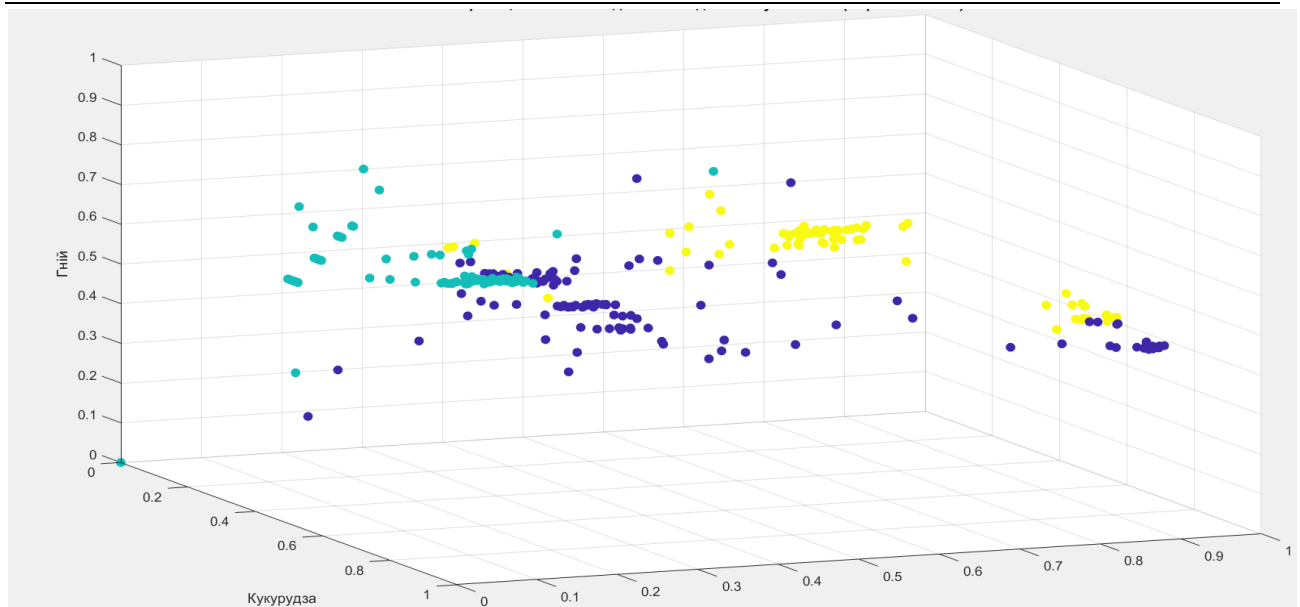


Рис.1. Кластеризація біогазових даних методом Fuzzy C-means

Для оцінки точності моделі було обчислено стандартні метрики:

- середню абсолютну помилку (MAE)

$$MAE = \frac{1}{N} \sum_{i=1}^N |\hat{y}_i - y_i|;$$

- середньоквадратичну помилку (RMSE)

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (\hat{y}_i - y_i)^2}.$$

У результаті проведеного моделювання було досягнуто задовільної точності прогнозування з показниками $RMSE = 0.35$ та $MAE = 0.23$ (у нормалізованих одиницях), що свідчить про ефективність застосованого підходу у нечіткому аналізі процесу біогазогенерації.

Незважаючи на свою ефективність, метод Fuzzy C-means має низку обмежень. Зокрема, він є чутливим до вибору кількості кластерів, потребує повної матриці вхідних даних без пропусків і погано справляється з високорозмірними просторами з великою кількістю шуму. Крім того, метод не враховує динаміку чи асоціативні властивості даних, які можуть бути важливими в умовах реального середовища.

Висновки

У роботі представлено математичну модель фазифікації біогазового процесу на основі нечіткої кластеризації. Проведено повний цикл обробки експериментальних даних: від нормалізації та фазової класифікації до розрахунку функцій належності та прогнозу виходу енергії. Результати підтвердили ефективність застосування FCM-методу для аналізу та прогнозування енергетичної продуктивності біогазових установок.

Отримані метрики точності (RMSE, MAE) свідчать про добру узгодженість прогнозу з реальними даними, незважаючи на нечіткий і варіативний характер сировини. У подальших дослідженнях планується поєднання даного підходу з методами гібридного нечіткого моделювання (наприклад, Fuzzy-BSB або Fuzzy-ANFIS) з метою покращення точності та адаптивності в динамічних умовах експлуатації.

Запропонована модель нечіткої фазифікації виробничого процесу біогазових установок може бути інтегрована до систем моніторингу та оптимізації ферментаційних процесів. Вона дозволяє класифікувати технологічний стан на основі поточних даних про сировину, здійснювати прогноз виходу біогазу з урахуванням фазових змін, виявляти потенційно критичні або неефективні режими роботи в реальному часі, а також підвищити керування процесу в умовах невизначеності, без необхідності використання великих обсягів історичних даних.

Це відкриває можливості для створення інтелектуальних систем керування установками, що адаптуються до змін у складі відходів та забезпечують стабільну генерацію енергії.

Література

1. Конечний Р. Економічна ефективність застосування біогазових установок. Технічне забезпечення інноваційних технологій в агропромисловому комплексі, 2021, 556.
2. Dubchak L. Modern renewable energy sources and methods for detecting their defects. Computer Systems and Information Technologies, 2024, №2, pp. 21–26. <https://doi.org/10.31891/csit-2024-2-3>
3. Деев Д.В. Автоматизована система керування біогазових установок в охороні навколишнього середовища. In: The XXIV International Scientific and Practical Conference «Modern technologies among us in the

environment», June 17-19, 2024, Rome, Italy. 419 p.

4. Паламаренко Я.В.. Оцінка ефективності функціонування біогазових установок: вітчизняний та закордонний досвід. *Problems of Economy*, 2023, 57.3.

5. Петров Д., Бойчук А. (2022). Метод кластеризації даних на основі алгоритму обходу графа. *Herald of Khmelnytskyi National University. Technical Sciences*, 307(2), 87-91. <https://doi.org/10.31891/2307-5732-2022-307-2-87-91>

6. Zaitsev, D., Filimonchuk, T., Huk, A., & Maistrenko, H. (2024). Огляд засобів ефективної сегментації зображень з використанням методів кластеризації даних. *Системи управління, навігації та зв'язку. Збірник наукових праць*, 1(75), ст.77-81.

7. Варчук, Д. А., Пришляк, О. В., Виноградова, Д. Р., Демент'єв, Р. В., & Марценюк, М. О. (2024). Математичне забезпечення для нечіткої кластеризації даних в паралельній системі управління базами даних. *СІТ'2024, Тернопіль*, 2 травня 2024

8. Dyvak, Mykola, Volodymyr Manzhula, Andriy Melnyk, Bohdan Rusyn, and Iryna Spivak. 2024. "Modeling the Efficiency of Biogas Plants by Using an Interval Data Analysis Method" *Energies* 17, no. 14: 3537. <https://doi.org/10.3390/en17143537>

9. Cinar, Samet, Senem Onen Cinar, Nils Wiecek, Ihsanullah Soho, and Kerstin Kuchta. 2021. "Integration of Artificial Intelligence into Biogas Plant Operation" *Processes* 9, no. 1: 85. <https://doi.org/10.3390/pr9010085>

10. Bukhtoyarov, Vladimir, Vadim Tynchenko, Kirill Bashmur, Oleg Kolenchukov, Vladislav Kukartsev, and Ivan Malashin. 2025. "Fuzzy Neural Network Applications in Biomass Gasification and Pyrolysis for Biofuel Production: A Review" *Energies* 18, no. 1: 16. <https://doi.org/10.3390/en18010016>

11. Zenani, Sibabalwe, KeChrist Obileke, Odilo Ndiweni, and Patrick Mukumba. 2025. "A Review of the Application of Fuzzy Logic in Bioenergy Technology" *Processes* 13, no. 7: 2251. <https://doi.org/10.3390/pr13072251>

12. Li E, Wang L, Song B, Jian S. Improved Fuzzy C-Means Clustering for Transformer Fault Diagnosis Using Dissolved Gas Analysis Data. *Energies*. 2018; 11(9):2344. <https://doi.org/10.3390/en11092344>

References

1. Konechny R. Ekonomichna efektyvnist' zastosuvannya biohazovykh ustanovok. *Tekhnichne zabezpechennya innovatsiynykh tekhnolohiy v ahropromyslovomu kompleksi*, 2021, 556.

2. Dubchak L. Modern renewable energy sources and methods for detecting their defects. *Computer Systems and Information Technologies*, 2024, №2, pp. 21–26. <https://doi.org/10.31891/csit-2024-2-3>

3. Dyeyev D.V. Avtomatyzovana systema keruvannya biohazovykh ustanovok v okhroni navkolysn'oho seredovyscha. In: *The XXIV International Scientific and Practical Conference «Modern technologies among us in the environment»*, June 17-19, 2024, Rome, Italy. 419 p.

4. Palamarenko YA.V.. Otsinka efektyvnosti funktsionuvannya biohazovykh ustanovok: vitchyznyany ta zakordonny dosvid. *Problems of Economy*, 2023, 57.3.

5. Petrov D., Boychuk A. (2022). Metod klasteryzatsiyi danykh na osnovi alhorytmu obkhodu hrafa. *Herald of Khmelnytskyi National University. Technical Sciences*, 307(2), 87-91. <https://doi.org/10.31891/2307-5732-2022-307-2-87-91>

6. Zaitsev, D., Filimonchuk, T., Huk, A., & Maistrenko, H. (2024). Ohlyad zasobiv efektyvnoyi sehmentatsiyi zobrazhen' z vykorystannyam metodiv klasteryzatsiyi danykh. *Systemy upravlinnya, navihatsiyi ta zv'yazku. Zbirnyk naukovykh prats'*, 1(75), st.77-81.

7. Varchuk, D. A., Pryshlyak, O. V., Vynogradova, D. R., Dement'yev, R. V., & Martsenyuk, M. O. (2024). Matematychnе zabezpechennya dlya nechitkoy klasteryzatsiyi danykh v paralel'niy systemi upravlinnya bazamy danykh. *SIT'2024, Ternopil'*, 2 travnya 2024

8. Dyvak, Mykola, Volodymyr Manzhula, Andriy Melnyk, Bohdan Rusyn, and Iryna Spivak. 2024. "Modeling the Efficiency of Biogas Plants by Using an Interval Data Analysis Method" *Energies* 17, no. 14: 3537. <https://doi.org/10.3390/en17143537>

9. Cinar, Samet, Senem Onen Cinar, Nils Wiecek, Ihsanullah Soho, and Kerstin Kuchta. 2021. "Integration of Artificial Intelligence into Biogas Plant Operation" *Processes* 9, no. 1: 85. <https://doi.org/10.3390/pr9010085>

10. Bukhtoyarov, Vladimir, Vadim Tynchenko, Kirill Bashmur, Oleg Kolenchukov, Vladislav Kukartsev, and Ivan Malashin. 2025. "Fuzzy Neural Network Applications in Biomass Gasification and Pyrolysis for Biofuel Production: A Review" *Energies* 18, no. 1: 16. <https://doi.org/10.3390/en18010016>

11. Zenani, Sibabalwe, KeChrist Obileke, Odilo Ndiweni, and Patrick Mukumba. 2025. "A Review of the Application of Fuzzy Logic in Bioenergy Technology" *Processes* 13, no. 7: 2251. <https://doi.org/10.3390/pr13072251>

12. Li E, Wang L, Song B, Jian S. Improved Fuzzy C-Means Clustering for Transformer Fault Diagnosis Using Dissolved Gas Analysis Data. *Energies*. 2018; 11(9):2344. <https://doi.org/10.3390/en11092344>