Міністерство освіти і науки України
Харківський національний університет імені В. Н. Каразіна
Навчально-науковий інститут «Українська інженерно-педагогічна академія»
Кафедра електротехніки та електроенергетики

КВАЛІФІКАЦІЙНА РОБОТА

магістра

на тему

ДОСЛІДЖЕННЯ ПРОЦЕСІВ ІДЕНТИФІКАЦІЇ ТОПОЛОГІЇ «КОРИСТУВАЧ-ФІДЕР» У НИЗЬКОВОЛЬТНИХ ЖИТЛОВИХ ЕЛЕКТРИЧНИХ МЕРЕЖАХ ПРИ КЛАСТЕРНОМ ОБ'ЄДНАННІ

(тема кваліфікаційної роботи)

Виконав: студент 2 маг. курсу, групи ДЕБ-24мг
спеціальності: 141 Електроенергетика, електротехніка та електромеханіка
_____/ Юрій ЛОЙКО
(підпис) (ім'я та прізвище)
Керівник _____/ Павло ВАСЮЧЕНКО
(підпис) (ім'я та прізвище)
Рецензент _____/ Олег ПОДОЛЯК
(підпис) (ім'я та прізвище)

«До захисту допущено»

Завідувач кафедри _____/Артем ЧЕРНЮК
(підпис) (ім'я та прізвище)
Секретар ЕК _____/Ігор КИРИСОВ
(підпис) (ім'я та прізвище)

Харків – 2025 рік

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ХАРКІВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
ІМЕНІ В. Н. КАРАЗІНА

Навчально – науковий інститут Українська інженерно-педагогічна академія
Кафедра електротехніки та електроенергетики
Спеціальність 141 Електроенергетика, електротехніка та електромеханіка
Освітньо-професійна програма Енергетична безпека

ЗАТВЕРДЖУЮ
Завідувач кафедри

_____
(підпис)

к.т.н., доцент Артем ЧЕРНЮК
(науковий ступінь, вчене звання, ім'я, прізвище)

«___» _____2025 р.

ЗАВДАННЯ
на кваліфікаційну роботу дипломну роботу
другого (магістерського) рівня вищої освіти

здобувачу вищої освіти Юрію ЛОЙКУ

1. Тема «ДОСЛІДЖЕННЯ ПРОЦЕСІВ ІДЕНТИФІКАЦІЇ ТОПОЛОГІЇ «КОРИСТУВАЧ-ФІДЕР» У НИЗЬКОВОЛЬТНИХ ЖИТЛОВИХ ЕЛЕКТРИЧНИХ МЕРЕЖАХ ПРИ КЛАСТЕРНОМ ОБ'ЄДНАННІ»
затверджена наказом по академії № 4801-5/3664 від «06» жовтня 2025 р.

2. Термін здачі закінченої роботи « 10 » 12 2025 р.

3. Вихідні дані до роботи/проєкту: Рівень напруги, умови експлуатації, характеристика струмоведучих частин, типи кабельних ліній, метод апарат дослідження, завдання, параметри обладнання

4. Зміст роботи/проєкту (перелік питань, які належить розробити):
Вступ. Перелік умовних позначень. ІДЕНТИФІКАЦІЯ ТОПОЛОГІЇ «КОРИСТУВАЧ–ЖИВИЛЬНИК» У НИЗЬКОВОЛЬТНИХ ЖИТЛОВИХ ЕЛЕКТРИЧНИХ МЕРЕЖАХ ПРИ КЛАСТЕРНОМУ ОБ'ЄДНАННІ. ТЕМАТИЧНЕ ДОСЛІДЖЕННЯ. ДОСЛІДЖЕННЯ ПОХИБКИ. Висновки. Список посилань

5.Перелік графічного матеріалу (презентаційний матеріал):
Слайди презентації

6. Консультант:

Розділ	Консультант	Підпис, дата		Оцінка (бали)
		Завдання видав	Завдання прийняв	

7. Дата видачі завдання «12» жовтня 2025 р.

Керівник _____ Павло ВАСЮЧЕНКО
(підпис) (ім'я, прізвище)

Завдання прийняв до виконання _____ Юрій ЛОЙКО
(підпис) (ім'я, прізвище)

КАЛЕНДАРНИЙ ПЛАН-ГРАФІК
виконання кваліфікаційної дипломної роботи

№ з/п	Назва етапів роботи та питань, які мають бути розроблені відповідно до завдання		Термін виконання	Позначки керівника про виконання завдань
1	Визначення актуальності теми дослідження		02.10.25 – 05.10.25	
2	Проведення аналізу наукової літератури		06.10.25 – 15.10.25	
3	Виконання першого розділу дипломної роботи		16.10.25 – 29.10.25	
4	Виконання другого та третього розділу дипломної роботи		30.10.25 – 26.11.25	
5	Оформлення пояснювальної записки		27.11.25 – 07.12.25	

Студент (ка) _____ Юрій ЛОЙКО
(підпис) (ім'я, прізвище)

РЕФЕРАТ

Магістерська робота виконана на 66 стр., містить 11 таблиці, 14 рисунків, що характеризують результати дослідження, використано 53 літературних посилань.

Об’єкт дослідження – низьковольтні житлові електричні мережі.

Предм дослідження – процеси ідентифікації топології «користувач-фідер».

Тема магістерської роботи – Дослідження процесів ідентифікації топології «користувач-фідер» у низьковольтних житлових електричних мережах при кластерном об’єднанні.

У цьому дослідженні розроблена керована даними структура для ідентифікації топології зв'язку між користувачем та фідером у низьковольтних житлових електричних мережах з використанням вимірювань навколишнього середовища (струму та напруги) від інтелектуальних лічильників.

Структура спочатку отримує необроблений набір даних за допомогою вейвлет-шумозаглушення, зменшення розмірності на основі аналізу головних компонент та вилучення часових ознак на основі глибокого навчання. Крім того, також застосовується підхід до виявлення аномалій на основі глибокого навчання. Для ідентифікації зв'язку між користувачем та фідером використовуються сім алгоритмів кластеризації, а потім результати об'єднуються за допомогою стратегії об'єднання результатів для подальшого підвищення точності ідентифікації.

Експерименти на трьох реальних житлових електричних мережах показують, що запропонований підхід значно перевершує результати, отримані методом одного кластеризації, та результати, отримані простим об'єднанням на основі голосування. Запропонований підхід досягає точності ідентифікації до 88% у розглянутих тематичних дослідженнях. Також проводяться дослідження похибки для підтвердження важливості кожного модуля в запропонованій структурі.

КЛЮЧОВІ СЛОВА: ТОПОЛОГІЯ «КОРИСТУВАЧ–ФІДЕР»; ЖИТЛОВА ЕЛЕКТРОМЕРЕЖА; ВИЯВЛЕННЯ АНОМАЛІЙ; КЛАСТЕРИЗАЦІЯ; ОБ'ЄДНАННЯ РЕЗУЛЬТАТІВ, ЕФЕКТИВНІСТЬ, ТЕХНІКО-ЕКОНОМІЧНІ ПОКАЗНИКИ

ABSTRACT

The master's work is completed on 66 pages, contains 11 tables, 14 figures characterizing the results of the research, 53 literary references are used.

The object of the study is low-voltage residential electrical networks.

The subject of the study is the processes of identifying the topology of the “user-feeder” in low-voltage residential electrical networks under cluster aggregation.

In this study, a data-driven framework is developed for identifying the topology of the connection between the user and the feeder in low-voltage residential electrical networks using environmental measurements (current and voltage) from smart meters.

The framework first obtains a raw data set using wavelet denoising, dimensionality reduction based on principal component analysis, and temporal feature extraction based on deep learning. In addition, a deep learning-based anomaly detection approach is also applied. Seven clustering algorithms are used to identify the connection between the user and the feeder, and then the results are combined using a result aggregation strategy to further improve the identification accuracy.

Experiments on three real residential power grids show that the proposed approach significantly outperforms the results obtained by the single clustering method and the results obtained by simple voting-based aggregation. The proposed approach achieves identification accuracy of up to 88% in the considered case studies. Ablation studies are also conducted to confirm the importance of each module in the proposed structure.

KEYWORDS: USER-FEEDER TOPOLOGY; RESIDENTIAL POWER GRID; ANOMALIES DETECTION; CLUSTERING; ASSIGNMENT OF RESULTS, EFFICIENCY, TECHNICAL AND ECONOMIC INDICATORS

ЗМІСТ

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ 8
ВСТУП 9
РОЗДІЛ 1 ІДЕНТИФІКАЦІЯ ТОПОЛОГІЇ «КОРИСТУВАЧ–ЖИВИЛЬНИК» У НИЗЬКОВОЛЬТНИХ ЖИТЛОВИХ ЕЛЕКТРИЧНИХ МЕРЕЖАХ ПРИ КЛАСТЕРНОМУ ОБ'ЄДНАННІ 11
1.1. Загальні відомості 11
ВИСНОВКИ ДО РОЗДІЛУ 1 13
РОЗДІЛ 2. ПОВ'ЯЗАНА РОБОТА 14
2.1. Методи ідентифікації топології на основі суміжних матриць 14
2.2. Методи ідентифікації топології на основі даних (машинного навчання) 16
2.3. Стан досліджень виявлення аномалій 17
ВИСНОВКИ ДО РОЗДІЛУ 2 18
РОЗДІЛ 3 ЗАПРОПОНОВАНА МЕТОДОЛОГІЯ 20
3.1. Огляд запропонованої структури 20
3.2. Попередня обробка на основі вейвлетів та зменшення розмірності 22
3.3. Методи виявлення аномалій 25
3.3.1. Ізоляційний ліс для виявлення аномалій 25
3.3.2. DeepSVDD для виявлення аномалій 26
3.4. Алгоритми кластеризації 29
3.4.1. DBSCAN для кластеризації 29
3.4.2. SOM для кластеризації 31
3.4.3. Визначення номера кластера 33
3.5. Стратегії кластеризації та ф'южну 34
3.5.1. Об'єднання кластеризації на основі KMeans 35
3.5.2. Об'єднання кластеризації на основі голосування 37
ВИСНОВКИ ДО РОЗДІЛУ 3 39
РОЗДІЛ 4. ТЕМАТИЧНЕ ДОСЛІДЖЕННЯ 40

4.1. Експериментальна установка 40
4.2. Продуктивність у реальних житлових електричних мережах 41
ВИСНОВКИ ДО РОЗДІЛУ 4 47
РОЗДІЛ 5. ДОСЛІДЖЕННЯ ПОХИБКИ 49
5.1. Налаштування похибки 49
5.2. Результати похибки 50
5.3. Порівняння витрат часу 53
ВИСНОВКИ ДО РОЗДІЛУ 5 55
ВИСНОВКИ Й РЕКОМЕНДАЦІЇ 57
СПИСОК ПОСИЛАНЬ 58

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ

ADN – активні розподільчі мережі

AMI – розширена інфраструктура вимірювань

DBSCAN – просторова кластеризація додатків з шумом на основі щільності

DeepSVDD – глибокий опис опорних векторів даних

DWT – дискретне вейвлет-перетворення

CEVR – коефіцієнт кумулятивної поясненої дисперсії

IQR – метод міжквартильного діапазону

HC – ієрархічна кластеризація

LSTM – довга короткострокова пам'ять

LOF – локальний фактор викиду

ML – машинне навчання

CNN – нейронна мережа

PCA – аналіз головних компонентів

PMU – Phasor Measurement Unit

RDQDA – дискримінантний аналіз зі зменшеною розмірністю

RNN – рекурентні нейронні мережі

SCADA – система диспетчерський контроль та збір даних

SVM – методи опорних векторів

SOM – самоорганізуюча карта

SVDD – методом опорних векторів даних

ВСТУП

Протягом останніх кількох десятиліть швидкий розвиток енергетичної галузі призвів до вищих вимог до експлуатаційної безпеки, гнучкості та економічної ефективності [1]. Як наслідок, розподільчі мережі повинні мати покращену спостережуваність та керованість, що є важливим для безпечної та ефективної роботи всієї енергосистеми [2]. Як життєво важливий вузол, що з'єднує кінцевих користувачів та джерела енергії, розподільча енергосистема відіграє вирішальну роль у забезпеченні ефективної роботи енергосистем [3]. Наразі велика кількість інтелектуальних лічильників широко розгорнута в низьковольтних розподільчих мережах по всьому світу [4]. Ці лічильники зазвичай реєструють дані про напругу, струм та потужність з інтервалом у 15 хвилин на стороні користувача, що призводить до величезної кількості даних, які потребують поглибленого аналізу та аналізу [5].

Більше того, порівняно з мережами передачі, розподільчі мережі мають складніші та динамічніші топологічні структури. Розподільча мережа зазвичай стикається з частішими змінами структури, оскільки для підвищення її операційної гнучкості та надійності було розгорнуто велику кількість комутаторів [6]. Наприклад, сучасні активні розподільчі мережі можуть зазнавати понад десяти перемикань протягом одного дня [7, 8]. Крім того, зі зростаючою інтеграцією розподілених джерел енергії та керованих навантажень, розподільчі мережі повинні постійно враховувати новозбудовані джерела або навантаження [9, 10]. Тому в розподільчій енергетичній мережі щодня можуть відбуватися часті топологічні зміни.

У цьому дослідженні розроблена керована даними структура для ідентифікації топології зв'язку між користувачем та фідером у низьковольтних житлових електричних мережах з використанням вимірювань навколишнього середовища (струму та напруги) від інтелектуальних лічильників. Структура спочатку отримує необроблений набір даних за допомогою вейвлет-шумозаглушення, зменшення розмірності на основі аналізу головних компонент

та вилучення часових ознак на основі глибокого навчання. Крім того, також застосовується підхід до виявлення аномалій на основі глибокого навчання. Для ідентифікації зв'язку між користувачем та фідером використовуються сім алгоритмів кластеризації, а потім результати об'єднуються за допомогою стратегії об'єднання результатів для подальшого підвищення точності ідентифікації. Експерименти на трьох реальних житлових електричних мережах показують, що запропонований підхід значно перевершує результати, отримані методом одного кластеризації, та результати, отримані простим об'єднанням на основі голосування. Запропонований підхід досягає точності ідентифікації до 88% у розглянутих тематичних дослідженнях. Також проводяться дослідження похибки для підтвердження важливості кожного модуля в запропонованій структурі.

Об’єкт дослідження – низьковольтні житлові електричні мережі.

Предм дослідження – процеси ідентифікації топології «користувач-фідер».

Тема магістерської роботи – Дослідження процесів ідентифікації топології «користувач-фідер» у низьковольтних житлових електричних мережах при кластерном об’єднанні.

РОЗДІЛ 1 ІДЕНТИФІКАЦІЯ ТОПОЛОГІЇ «КОРИСТУВАЧ–ЖИВИЛЬНИК» У НИЗЬКОВОЛЬТНИХ ЖИТЛОВИХ ЕЛЕКТРИЧНИХ МЕРЕЖАХ ПРИ КЛАСТЕРНОМУ ОБ'ЄДНАННІ

1.1. Загальні відомості

Як показано на рисунку 1.1, точна ідентифікація топології «користувач–живильник» у низьковольтній житловій електромережі є важливою для експлуатації та планування системи, наприклад, калібрування втрат лінії, виявлення крадіжки електроенергії та реагування на попит. На ранніх стадіях отримання топологічної інформації розподільчої мережі покладалося на ручне переміщення кожного «вузла» та «вимикача» в мережі. Цей підхід вимагає величезної кількості апаратного обладнання, що робить його трудомістким та громіздким, і часто призводить до значних помилок. Персонал з обслуговування мережі також стикається зі зростаючим навантаженням на ідентифікацію топології мережі. У відповідь на це енергетичні компанії та дослідники розробили ефективніші методи ідентифікації. Система SCADA (диспетчерський контроль та збір даних) широко застосовується для моніторингу та ідентифікації топології мережі. AMI (розширена інфраструктура вимірювань) та інтелектуальні лічильники були швидко впроваджені в усій системі. Багатовимірні електричні вимірювання з високою роздільною здатністю, зібрані інтелектуальними лічильниками, містять приховані експлуатаційні характеристики компонентів мережі, що зменшує потребу в людській праці для ідентифікації топології. Сьогодні темпи впровадження інтелектуальних лічильників досягли дуже високого рівня в багатьох країнах, що робить інтелектуальний аналіз даних інтелектуальних лічильників основним напрямком досліджень в аналізі розподільчих мереж [11, 12].

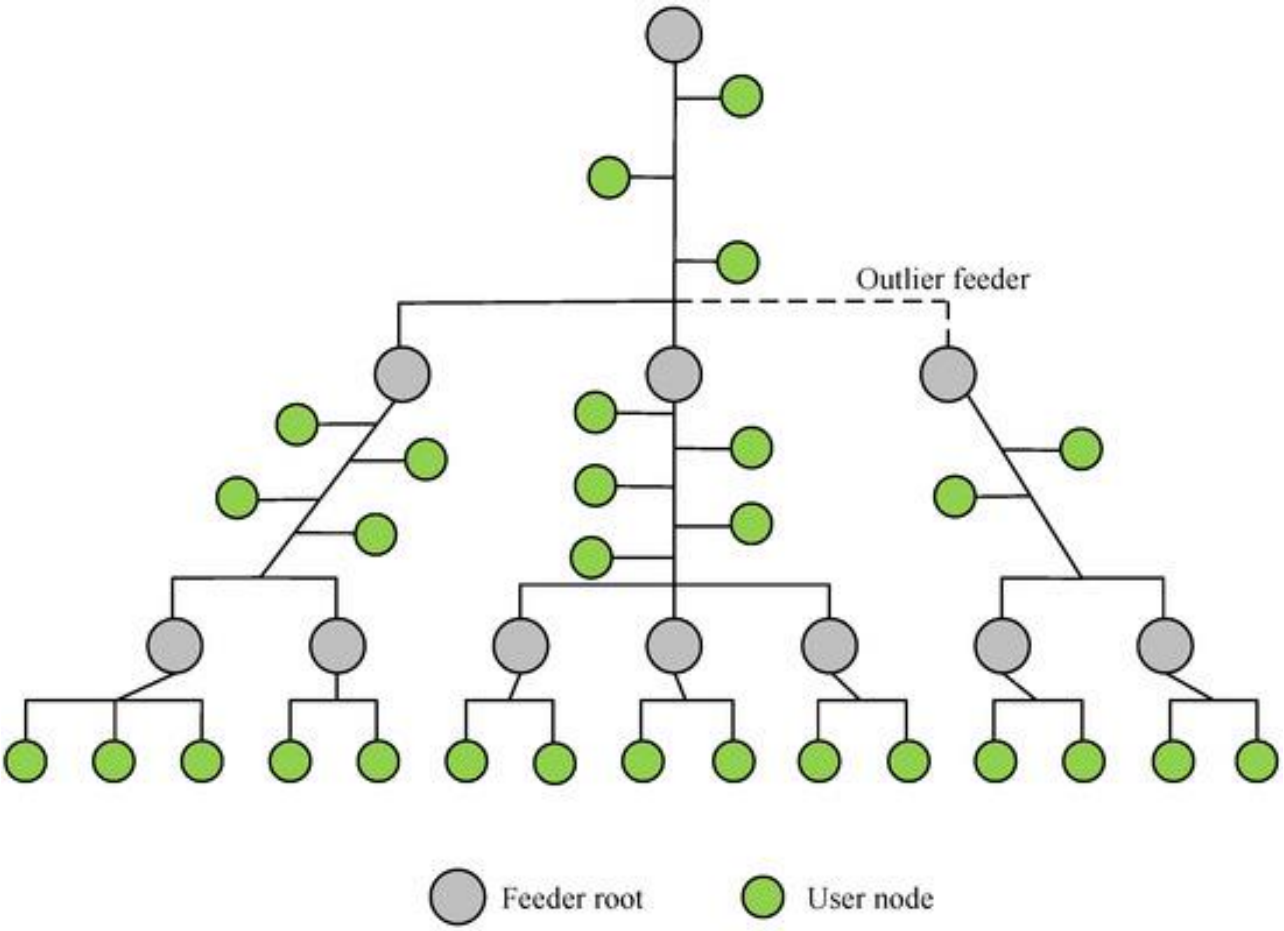

Рисунок 1.1. – Приклад топології «користувач–фідер» у низьковольтній розподільчій мережі електропостачання

Однак, через великий обсяг та високу розмірність даних інтелектуальних лічильників, вилучення репрезентативних ознак залишається серйозною проблемою; наприклад, деякі існуючі методи можуть демонструвати низьку продуктивність у виключенні вузлів користувачів поза межами зацікавленої області розподілу. Крім того, традиційні підходи до ідентифікації на основі класифікації сильно залежать від відомих міток певних «середніх» вузлів. Зрештою, їхня обчислювальна складність значно зростає з масштабом системи [13, 14].

ВИСНОВКИ ДО РОЗДІЛУ 1

Щоб усунути обмеження традиційних методів ідентифікації топології користувач-фідер, особливо при обробці шумних даних, змішаних із зовнішніми вимірюваннями фідерів (тобто викидами), у цьому дослідженні пропонується комплексна структура, яка спирається лише на вимірювання з низькою роздільною здатністю, неінтрузивні (навколишні) вимірювачі: спочатку проводиться попередня вейвлет-обробка та зменшення розмірності, потім вилучення ознак часових рядів; потім виконується виявлення викидів (аномалій); і, нарешті, застосовуються кілька моделей базової кластеризації. Для підвищення точності ідентифікації топології розроблено стратегію об'єднання результатів для об'єднання кількох результатів базової кластеризації в один з покращеною точністю кластеризації.

РОЗДІЛ 2. ПОВ'ЯЗАНА РОБОТА

У таблиці 2.1. наведено короткий огляд методів виведення топології розподільчої мережі за останні роки, а короткі описи деяких згаданих публікацій наведено в наступних абзацах. Крім того, в кінці цього розділу наведено короткий вступ до сучасного стану виявлення аномалій.

2.1. Методи ідентифікації топології на основі суміжних матриць

Матриця суміжності — це поширений метод, що використовується в теорії графів для опису графових структур. Вона має вигляд двовимірної матриці A, де

$$A_{ij} = \begin{cases} 1, & \text{if there exists a direct edge between node } i \text{ and node } j \\ 0, & \text{otherwise} \end{cases} \qquad (2.1.)$$

Порівняно з методами пошуку на основі дерев, матриця суміжності забезпечує більш прямий спосіб вираження структурних зв'язків в енергетичних системах і особливо підходить для алгоритмічної реалізації. Однак, у випадку великомасштабних і складних структур вузлів у сучасних розподільчих мережах, розмірне розширення матриць суміжності може призвести до низької обчислювальної ефективності.

Таблиця 2.1. – Зведення методів виведення топології розподільчої мережі.

Посилання.	Вимірювання	Підхід	Основна ідея	Результат
[4]	$\|V\|$, KHI	Аналіз гармонійних спотворень	Гармонічні закономірності поширення	Топологія
[5]	μПМУ/$\|PQ\|$	Узагальнений розподіл значень	Помилка реконструкції стиснення	Станів завантаження
[7]	μПМУ/ $\|V\|$, $\|I\|$, θ_V	Глибоке навчання	Оцінка матриці зв'язності	Топологічні переходи
[10]	μПМУ/$\|I\|$	Багатоагентна система	Локалізація несправності, ізоляція	Самовідновлювальна топологія
[2]	$\mu PMU/\|PQ\|$	Потік потужності + MLP	Матриця чутливості	Деревоподібна топологія
[15]	$\|V\|\|I\|\|PF\|$	Графове навчання	Топологія Штейнера	Топологія
[12]	$\|V\|$, $\|I\|$	Прогнозування на основі даних	Машинне навчання	Автобусне сполучення
[16]	μPMU/|V|,|I|	Опукла оптимізація	Кореляція напруги	Топологія
[17]	$\|PQ\|$, $\|V\|$, $\|I\|$	NV-DSSE	Залишки, оптимальне узгодження	Стан лінії/комутатора
[18]	μПМУ/ $\|V\|$, $\|I\|$, θ_V	MILNP + bigM	Гармонійні шляхи	Гармонічні струми
[19]	$\|V\|$, $\|I\|$	Глибоке навчання	Прогнозний аналіз залишків	Оцінка топології
[20]	μПМУ/$\|V\|$, $\|PQ\|$	Регресія на основі даних	Рівняння потоку потужності	Параметр лінії
[21]	μПМУ/|B|	Аналіз часових рядів	Методи машинного навчання	Топологія
[22]	μПМУ/$\|V\|$	Кластеризація	Інформація про топологію IEC 61850 SCL	Лінійна топологія
[23]	|PQ|,|V|,|I|	НЛП	Двоступеневий параметр	Топологія
[24]	μПМУ/$\|V\|$, θ_V	Ймовірнісна оцінка стану	Дивергенція KL	Стан лінії/комутатора
[25]	μПМУ/$\|V\|$, θ_V	Баєсівська мережа + граф	Ймовірність MAP	Топологія
[26]	$\|V\|$	Модель збурень підпростору	Співвідношення екстремальних власних значень	Топологічні зміни
[27]	μПМУ/$\|V\|\theta_V$	Ймовірнісна графічна модель	Умовна незалежність	Топологія

Щоб вирішити цю проблему, у [16] для ідентифікації топології було введено орієнтовану матрицю суміжності шляхом аналізу тенденцій падіння напруги на вузлах та побудови матриці суміжності на основі напруги. Це дозволяє одночасно ідентифікувати топологію та зміни напрямку потоку потужності. З точки зору оптимізації, у [17] було запропоновано вдосконалений метод множення матриць суміжності, який вводить розріджені елементи підматриці під час змін топології для зменшення складності. Крім того, було запропоновано метод ідентифікації топології на основі матриці довжини шляху, який покращує розуміння та візуалізацію зв'язності вузлів шляхом включення метрик найкоротшого шляху та операцій з матрицею min-plus [18].

2.2. Методи ідентифікації топології на основі даних (машинного навчання)

З широким розгортанням вимірювальних приладів та зростаючим накопиченням даних у розподільчих мережах, методи ідентифікації топології на основі даних поступово стали поширеними. На відміну від традиційних підходів, заснованих на моделях, методи на основі даних менше залежать від повних параметрів системи та демонструють більшу гнучкість та адаптивність, особливо в сценаріях з частими змінами в активній розподільчій мережі (ADN) або в середовищах, де вимірювання є шумними [19].

Машинне навчання (ML) – це підклас підходів, керованих даними, який спирається на навчання на основі даних для вивчення закономірностей та виконання завдань прогнозування. В енергетичних системах ML використовується для класифікації топологій систем на основі історичних робочих станів або даних часових рядів. Вхідні дані, такі як напруга, струм та потужність, можуть бути подані в навчені моделі для визначення зв'язку між вузлами та визначення станів комутаторів або лінійних з'єднань. Наприклад, було запропоновано двоетапний метод неконтрольованої кластеризації напруги

для класифікації вузлів за рівнями напруги та пов'язування топології, керованої напругою, з відомими моделями [20]. У [21] 15-хвилинний часовий ряд напруги використовується для вилучення ознак та ідентифікації топологій, поєднуючи статистичну класифікацію ознак «один до багатьох» та дискримінантний аналіз зі зменшеною розмірністю (RDQDA) для класифікації вузлів.

Глибоке навчання – це підклас підходів машинного навчання, який будує багатошарові мережі для автоматичного вилучення ознак з великомасштабних даних, фіксуючи приховані структури та складні шаблони. Наприклад, згорткова нейронна мережа (CNN) у поєднанні з методами опорних векторів (SVM) була застосована для ідентифікації топологічних перемикань у розподільчих мережах. Модель автоматично вивчала локальні шаблони у формах хвиль та досягла високоточного виявлення перемикань [19].

У таксономії методів машинного навчання самонавчання стосується навчання, яке не потребує маркованих даних і спирається виключно на розподіл і структуру даних, і підходить для дослідницького аналізу або завдань кластеризації. У розподільчих мережах самонавчання може допомогти виявити приховані групи вузлів або зміни в топології шляхом аналізу кластерів напруги, тенденцій потоків потужності або сигналів комутації. Наприклад, для ідентифікації топологій у низьковольтних житлових електричних мережах було запропоновано метод, заснований на файлах конфігурації SCL IEC 61850 та алгоритмах кластеризації KMeans [22]. Метод інтегрує просторові та електричні характеристики, покращуючи ідентифікацію регіональних зон електропостачання та відстеження топологічних оновлень.

2.3. Стан досліджень виявлення аномалій

Виявлення аномалій стосується ідентифікації точок даних, які суттєво відрізняються від більшості даних або певних закономірностей [28]. Ці аномалії можуть свідчити про несправності датчиків, помилки зв'язку, зміни в топології

системи або структурні аномалії. У процесі ідентифікації топології виявлення аномалій може підвищити точність і надійність ідентифікації. Крім того, фільтрація аномальних даних допомагає підвищити точність навчальних даних для моделей штучного інтелекту.

Ранні методи виявлення аномалій переважно спиралися на статистичні методи. Наприклад, Пірсон запропонував аналіз головних компонентів (PCA) для визначення напрямку, який найкраще відображає варіацію в даних, тим самим зменшуючи розмірність та візуалізуючи важливі закономірності [29]. Для стандартизації даних та виявлення викидів були розроблені статистичні методи на основі Z-показника [30]. Метод міжквартильного діапазону (IQR) також широко використовується завдяки своїй простоті та ефективності у виявленні аномалій. У [31] було запропоновано двоетапний метод неконтрольованої кластеризації напруги для класифікації вузлів за рівнями напруги та пов'язування топології, керованої напругою, з відомими моделями.

Оскільки кластеризація та методи на основі відстані набули популярності, для виявлення локальних аномалій щільності були розроблені такі алгоритми, як локальний фактор викиду (LOF) [32]. DBSCAN (просторова кластеризація додатків з шумом на основі щільності) був запроваджений для виявлення кластерів довільної форми та ідентифікації точок шуму на основі різниці щільності [33].

ВИСНОВКИ ДО РОЗДІЛУ 2

З розвитком технологій великих даних та штучного інтелекту, методи на основі машинного навчання стали мейнстрімом. Глибокі нейронні мережі продемонстрували сильні можливості узагальнення, а CNN, рекурентні нейронні мережі (RNN) та методи опорних векторів широко застосовуються в завданнях виявлення аномалій [34]. У деяких роботах використовуються дані зображень PMU (Phasor Measurement Unit) у поєднанні з CNN та RNN для високоточного

виявлення в різних сценаріях [35]. Гібридні підходи, що інтегрують розподілені фреймворки на основі Spark з SVM та мережами GRU з акцентом на увагу, показали значні покращення у захопленні часових ознак, тоді як двопотокові LSTM-моделі ще більше підвищили точність ідентифікації топології з використанням даних, що не належать до PMU [36].

Дослідницькі зусилля з виявлення аномалій в енергетичних системах зосереджені у двох напрямках: методи на основі моделей, що спираються на попередні знання про систему, та методи, що базуються на даних, що використовують польові вимірювання. Останній підхід, що підтримується великими реальними даними, стає популярним у сферах діагностики несправностей та інтелектуального прийняття рішень [37].

РОЗДІЛ 3. ЗАПРОПОНОВАНА МЕТОДОЛОГІЯ

3.1. Огляд запропонованої структури

У цьому розділі представлено огляд запропонованої структури для ідентифікації топології та виявлення аномалій у мережах розподілу електроенергії, як показано на рисунку 3.1.

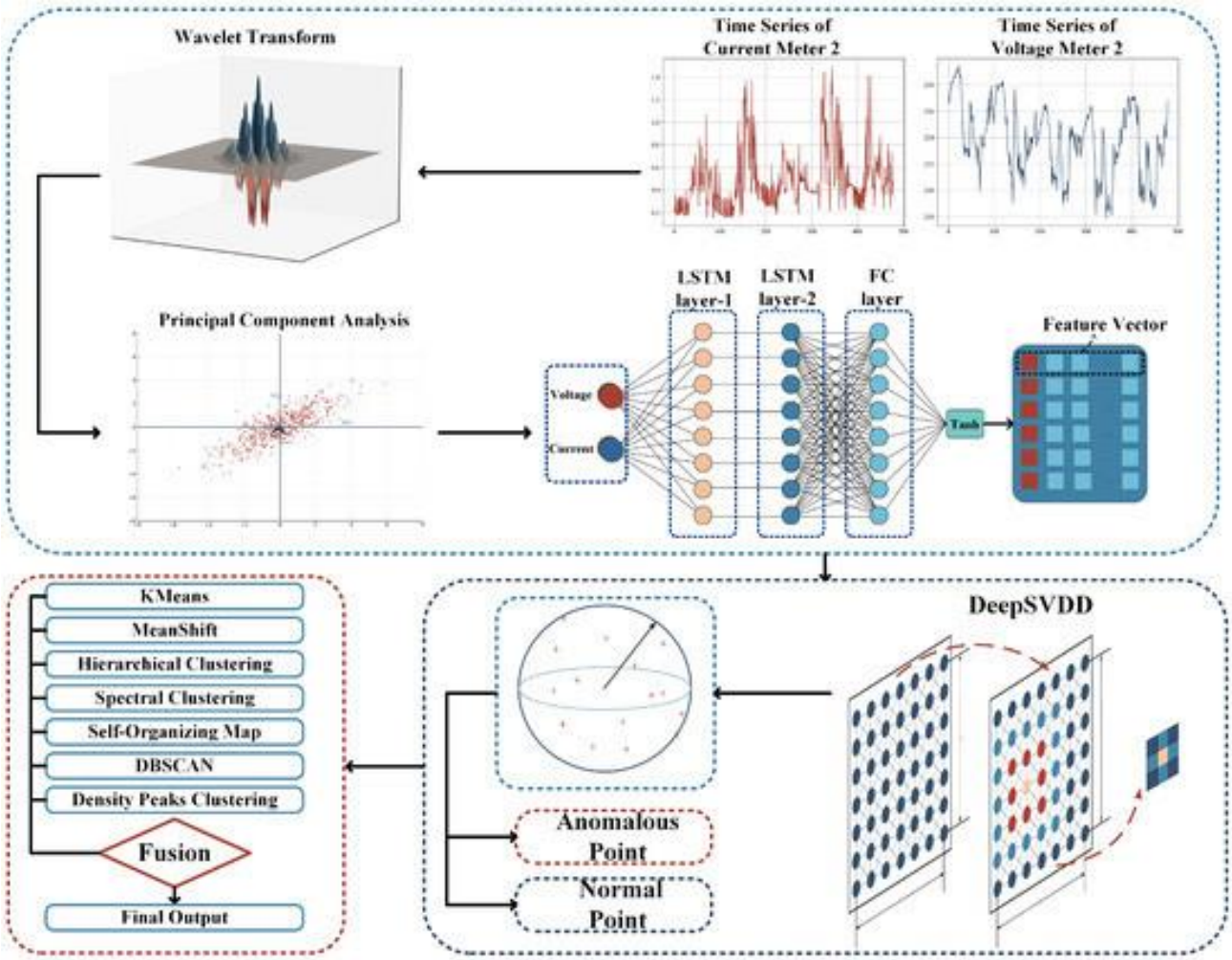

Рисунок 3.1. – Огляд запропонованої структури: PCA, LSTM, DeepSVDD, кластеризація та об'єднання

Фреймворк інтегрує кілька методів, включаючи PCA для зменшення розмірності, LSTM для вилучення ознак часових рядів, DeepSVDD (глибокий опис опорних векторів даних) для виявлення аномалій, сім алгоритмів кластеризації та стратегію об'єднання результатів для підвищення точності.

Головною метою є створення наскрізного конвеєра для видалення аномалій та кластеризації даних мережі розподілу електроенергії якомога точніше.

Процес починається з попередньої обробки необроблених даних з розподільчої мережі, включаючи дані про струм та напругу в житлових приміщеннях споживачів. Необроблені дані проходять мінімально-максимальну нормалізацію, щоб забезпечити однаковий масштаб усіх характеристик. Відсутні точки даних обробляються за допомогою методів імпутації. Нормалізовані дані потім готові до подальшого аналізу та обробки.

PCA застосовується для зменшення розмірності даних, зберігаючи при цьому найбільш значущу дисперсію. Для визначення кількості компонентів, які потрібно зберегти, як критерій зазвичай використовується коефіцієнт кумулятивної поясненої дисперсії (CEVR) [38]:

$$\eta_k = \frac{\sum_{i=1}^{k} \lambda_i}{\sum_{i=1}^{p} \lambda_i}, \qquad (3.1.)$$

де p – загальна кількість ознак, а k – кількість вибраних компонентів. Значенняηkвказує на частку загальної дисперсії, поясненої першими k компонентами. Кількість головних компонентів обрана таким чином, щоб зберегти понад 99% дисперсії даних. Цей крок допомагає зменшити складність даних, гарантуючи, що збережуться лише найважливіші ознаки. Кумулятивна пояснена дисперсія розраховується для PCA, щоб визначити оптимальну кількість компонентів, і результати показують, що 80 або більше головних компонентів достатньо для ефективного виявлення та класифікації аномалій.

Після PCA для вилучення часових ознак з даних використовується модель LSTM (довга короткострокова пам'ять). LSTM обрана завдяки своїй здатності фіксувати довгострокові залежності в послідовних даних, що є вирішальним для виявлення тенденцій та аномалій у залежних від часу вимірюваннях. Щоб уникнути подальшої втрати інформації після зменшення розмірності PCA,

вихідна розмірність кодера LSTM встановлюється відповідно до зменшеного простору ознак. Таким чином, LSTM зосереджується на фіксації часових залежностей без введення додаткового стиснення поза кроком PCA. Мережа LSTM використовує прихований розмір 64 та обробляє часові ознаки, моделюючи послідовність даних з кожної точки вимірювання. Після навчання модель LSTM видає стисле представлення часових шаблонів, яке потім використовується як вхідні дані для виявлення аномалій за допомогою DeepSVDD.

Для виявлення аномалій використовується DeepSVDD. DeepSVDD – це метод виявлення аномалій на основі глибокого навчання, який вивчає межу для відокремлення нормальних точок даних від аномальних. Він мінімізує відстань між точками даних та заздалегідь визначеним центром у просторі ознак.

Після виявлення та фільтрації аномалій решта даних проходить кластеризацію за допомогою семи різних алгоритмів кластеризації: DBSCAN, DPC (кластеризація піків щільності), HC (ієрархічна кластеризація), KMeans, MeanShift, SOM (самоорганізуюча карта) та Spectral Clustering. Кожен алгоритм виконує різний тип аналізу попередньо оброблених даних, використовуючи різні метрики та алгоритми відстані. Використання кількох методів кластеризації забезпечує надійну та надійну ідентифікацію закономірностей у даних.

Після кластеризації результати семи алгоритмів кластеризації об'єднуються за певними стратегіями об'єднання результатів, і отримується кінцевий результат ідентифікації топології. У наступних підрозділах буде детально розглянуто кожен модуль запропонованої структури.

3.2. Попередня обробка на основі вейвлетів та зменшення розмірності

Для покращення представлення часових сигналів струму та придушення високочастотного шуму ми застосовуємо попередню обробку сигналу на основі вейвлетів та реконструкції перед зменшенням розмірності. Цей етап попередньої

обробки служить для вилучення критичних структурних варіацій з часом, одночасно зменшуючи надлишковість, і застосовується до кожної послідовності вимірювань (напруги та струму) окремо.

Кожен оригінальний сигнал $\mathbf{x} = \{x_t\}_{t=1}^{T}$ розкладається за допомогою трирівневого дискретного вейвлет-перетворення (DWT) з вейвлетом Добеші-1 (db1), що дає ієрархічний набір коефіцієнтів апроксимації та деталізації:

$$\mathbf{x} \xrightarrow{\text{DWT}} \{\mathbf{A}_3, \mathbf{D}_3, \mathbf{D}_2, \mathbf{D}_1\} \qquad (3.2.)$$

тут, A_3 позначає коефіцієнти апроксимації на рівні 3, що враховують низькочастотні компоненти, тоді як $\mathbf{D}j$, де j = 1, 2, 3 представляють коефіцієнти деталізації на кожному рівні розкладання.

Щоб підкреслити домінантні структури сигналу та усунути флуктуації низької величини, ми застосовуємо нелінійне порогове регулювання до коефіцієнтів деталізації. Зокрема, для кожного рівня j ми обчислюємо 30-й процентиль. λj коефіцієнтів та встановити всі записи нижче цього порогу рівними нулю:

$$D_k^{(j)} = \begin{cases} 0, & D_k^{(j)} < \lambda_j \\ D_k^{(j)}, & \text{otherwise} \end{cases} \quad \text{where} \quad \lambda_j = \text{percentile}_{30}(\mathbf{D}_j) \qquad (3.3.)$$

Щоб видалити довгострокові тенденції та посилити тимчасові особливості, коефіцієнти апроксимації також встановлюються на нуль:

$$\mathbf{A}_3 = \mathbf{0} \qquad (3.4.)$$

Оброблені коефіцієнти потім використовуються для реконструкції сигналу за допомогою оберненого DWT, що призводить до отримання послідовності ознак з усунутим шумом.

$$\tilde{\mathbf{x}} = \mathrm{IDWT}\left(\mathbf{0}, \tilde{\mathbf{D}}_3, \tilde{\mathbf{D}}_2, \tilde{\mathbf{D}}_1\right) \qquad (3.5.)$$

Щоб забезпечити узгодженість між вибірками та зменшити артефакти на краях від перетворення, ми витягуємо сегмент фіксованої довжини з кожної реконструйованої послідовності (у цій статті, від часового індексу 100 до 200):

$$\tilde{\mathbf{x}}^{\mathrm{final}} = \tilde{\mathbf{x}}[100:200] \qquad (3.6.)$$

Після отримання всіх реконструйованих послідовностей ознак, ми упорядковуємо їх у матрицю ознак $\mathbf{X} \in \mathbb{R}^{N \times T}$, де N – кількість користувачів, а T – довжина послідовності. Ця матриця додатково обробляється за допомогою PCA для зменшення розмірності:

$$\mathbf{Z} = \mathbf{X}\mathbf{W} \qquad (3.7.)$$

тут, $\mathbf{W} \in \mathbb{R}^{T \times d}$ позначає проекційну матрицю, що складається з d головних компонентів, та $\mathbf{Z} \in \mathbb{R}^{N \times d}$ – результуюче низьковимірне представлення, яке використовується як вхідні дані для наступної моделі ідентифікації топології.

Цей конвеєр попередньої обробки на основі вейвлетів гарантує, що зберігаються лише найбільш інформативні та структурозберігаючі компоненти сигналу, тим самим покращуючи якість та інтерпретованість вхідних ознак.

3.3. Методи виявлення аномалій

У цьому дослідженніописано два підходи до виявлення аномалій: DeepSVDD [39] (прийнятий метод) та iForest (Ізоляційний ліс) [40] (метод, що використовується для порівняння).

3.3.1. Ізоляційний ліс для виявлення аномалій

Традиційні методи виявлення аномалій часто спираються на обчислення щільності, відстані або структури кластерування серед зразків. Натомість, основна ідея iForest полягає в тому, що аномалії більш схильні до ізоляції. Тобто, аномалії можна легше відокремити від решти даних з меншою кількістю розділень.

Враховуючи набір даних $X = \{x1, x2, \ldots, x\,n\} \subset \mathbb{R}^d$, кожне Дерево Ізоляції будується за допомогою таких кроків:

(1) Випадково вибрати індекс ознаки$j \in \{1, 2, \ldots, d\}$;

(2) Випадковим чином вибрати значення поділу $p \in [\min x_j, \max x_j]$ on the selected feature;

(3) Розподіліть дані наступним чином:

$$X_{\text{left}} = \{x \in X \mid x_j \leq p\}, \quad X_{\text{right}} = \{x \in X \mid x_j > p\};$$

(4) Повторюйте кроки 1–3 рекурсивно для кожного розділу, доки підмножина не міститиме один зразок або не буде досягнуто максимальної висоти дерева.

Кожне iTree записує довжину шляху $h\ (x)$ необхідно ізолювати кожну вибірку x. Після побудови T дерев середня довжина шляху розраховується наступним чином:

$$c(n) = 2H(n-1) - \frac{2(n-1)}{n}, \quad H(n) \approx \ln(n) + \gamma \qquad (3.8.)$$

де $\gamma \approx 0{,}5772$ – стала Ейлера-Маскероні.

На основі порогового значення θ, результат виявлення аномалії y_i для кожного зразка x_i визначається так:

$$y_i = \begin{cases} 1, & \text{if } s(x_i) > \theta \quad \text{(anomalous)} \\ 0, & \text{o therwise} \quad \text{(normal)} \end{cases} \qquad (3.9.)$$

Підсумовуючи, алгоритм «Ізоляційного лісу» ізолює точки даних, побудовуючи випадково розділені бінарні дерева. Він оцінює бал аномалії на основі того, наскільки легко виділити вибірку в просторі ознак. Порівняно з традиційними підходами на основі відстані або щільності, iForest демонструє вищу обчислювальну ефективність та стійкість у багатовимірних просторах, не покладаючись на припущення кластеризації.

У цьому дослідженні модель iForest використовується (як один із двох методів виявлення аномалій) для видалення прихованих аномалій та забезпечення додаткового, чистішого вхідного сигналу для методів кластеризації на пізнішому етапі.

3.3.2. DeepSVDD для виявлення аномалій

DeepSVDD – це метод наскрізного виявлення аномалій, який поєднує глибокі нейронні мережі з (SVDD). Метою є відображення вхідних даних у низьковимірний простір вбудовування за допомогою нейронної мережі та мінімізація відстаней між точками даних та заздалегідь визначеним центром, тим самим ефективно виявляючи аномалії у просторі представлення. Ілюстрація DeepSVDD представлена на рисунку 3.2.

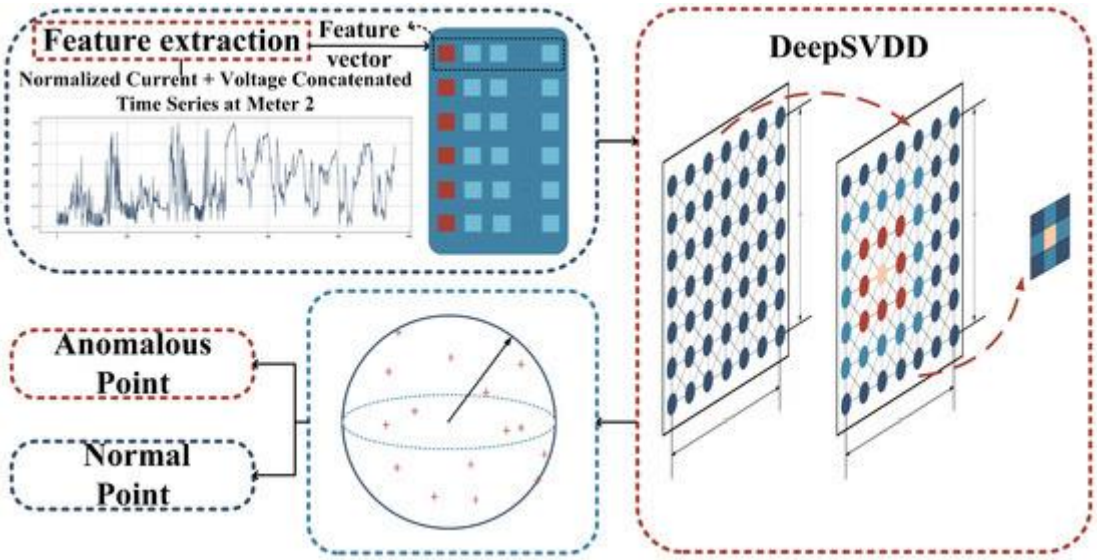

Рисунок 3.2. – Ілюстрація методу DeepSVDD

Основна ідея DeepSVDD полягає у вивченні трансформації $f(x; \theta)$: $\mathbb{R}d \rightarrow \mathbb{R}k$ що вбудовує вхід x у компактний простір таким чином, що нормальні дані залишаються близько до центруc $\in\mathbb{R}k$, тоді як аномалії є віддаленими. Вбудовування $z = f(x; \theta)$ засвоюється шляхом мінімізації мети:

$$\min_{\theta} \frac{1}{n} \sum_{i=1}^{n} \|f(x_i; \theta) - c\|^2 \qquad (3.10.)$$

Центр c зазвичай ініціалізується з використанням середнього значення вихідних даних початкових вагових коефіцієнтів мережі:

$$c = \frac{1}{n} \sum_{i=1}^{n} f(x_i; \theta_0) \qquad (3.11.)$$

Щоб зменшити перенавчання, додається член регуляризації:

$$\mathcal{L}_{\text{reg}} = \frac{1}{n} \sum_{i=1}^{n} \|f(x_i; \theta) - c\|^2 + \lambda \cdot \|\theta\|^2 \qquad (3.12.)$$

де λ є гіперпараметром регуляризації.

Після навчання, оцінка аномалії для нового вхідного значення x обчислюється наступним чином:

$$s(x) = \|f(x; \theta^*) - c\|^2 \qquad (3.13.)$$

зразок класифікується як аномальний, якщо його відстань перевищує заздалегідь визначений поріг δ:

$$\text{label}(x) = \begin{cases} 1, & \text{if } s(x) > \delta \quad \text{(Anomalous)} \\ 0, & \text{otherwise} \quad \text{(Normal)} \end{cases} \tag{3.14.}$$

Робочий процес навчання та оцінювання DeepSVDD узагальнено в Алгоритмі 1, що забезпечує наскрізний конвеєр від вбудовування навчання до оцінювання аномалій. Відповідно до загальноприйнятої практики виявлення аномалій без нагляду, поріг визначення високого процентиляδчасто використовується для відокремлення нормальних зразків від аномальних [41]. Відповідно до цього принципу, ми встановлюємо порігδдо 99-го процентиля розподілу відстаней у наших експериментах. Аналогічно, центр c гіперсфери в DeepSVDD не вибирається довільно, а ініціалізується як емпіричне середнє значення всіх навчальних вибірок у початковому просторі вбудовування:

$$c = \frac{1}{N} \sum_{i=1}^{N} \phi(x_i; \theta_{\text{init}}) \tag{3.15.}$$

де $\phi(x_i; \theta_{\text{init}})$ позначає представлення вибірки$x i$за початкових параметрів мережі. Цей вибір мотивований тим фактом, що за фіксованого відображення оптимальний центр, який мінімізує середній квадрат відстані, є середнім значенням вбудовувань.

Algorithm 1 DeepSVDD-Based Anomaly Detection

Require: Dataset $\mathcal{X} = \{x_1, x_2, ..., x_n\}$, neural network $f(x; \theta)$, regularization weight λ
Ensure: Trained parameters θ^*, center c, threshold δ
1: Initialize θ_0 with random weights
2: Compute $c \leftarrow \frac{1}{n} \sum_{i=1}^{n} f(x_i; \theta_0)$
3: **repeat**
4: Compute loss $\mathcal{L} = \frac{1}{n} \sum_i \|f(x_i; \theta) - c\|^2 + \lambda \|\theta\|^2$
5: Update $\theta \leftarrow \theta - \eta \cdot \nabla_\theta \mathcal{L}$
6: **until** convergence
7: Set $\theta^* \leftarrow \theta$
8: Compute anomaly scores $s(x_i) = \|f(x_i; \theta^*) - c\|^2$
9: Determine threshold δ (e.g., using percentile of $s(x)$)

DeepSVDD особливо підходить для виявлення аномалій без вчителя у високовимірних часових рядах даних, як це трапляється в мережах розподілу

електроенергії. У цьому дослідженні він використовується після вилучення часових ознак за допомогою LSTM. Така фільтрація може допомогти підвищити узгодженість і точність подальших процесів кластеризації. Інтеграція нейронного вбудовування та гіперсферного навчання SVDD забезпечує модель потужною здатністю до узагальнення, що дозволяє їй виявляти складні аномальні закономірності в реальних даних вимірювань.

3.4. Алгоритми кластеризації

У цьому дослідженні для порівняння та об'єднання розглядається кілька алгоритмів кластеризації, включаючи KMeans, MeanShift, DBSCAN, DPC, Spectral Clustering, Hierarchical Clustering та Self-Organizing Map. Основна ідея та алгоритмічні деталі для DBSCAN та SOM описані нижче. Детальну інформацію про інші п'ять методів можна знайти в [42].

3.4.1. DBSCAN для кластеризації

DBSCAN – це метод кластеризації на основі щільності. Він групує точки даних, які розташовані щільно, позначаючи точки в областях з низькою щільністю як викиди. Алгоритм вимагає двох ключових параметрів: радіуса околиці ε та мінімальну кількість балів, для формування щільної області. Схематичне зображення DBSCAN зображено на рисунку 3.3.

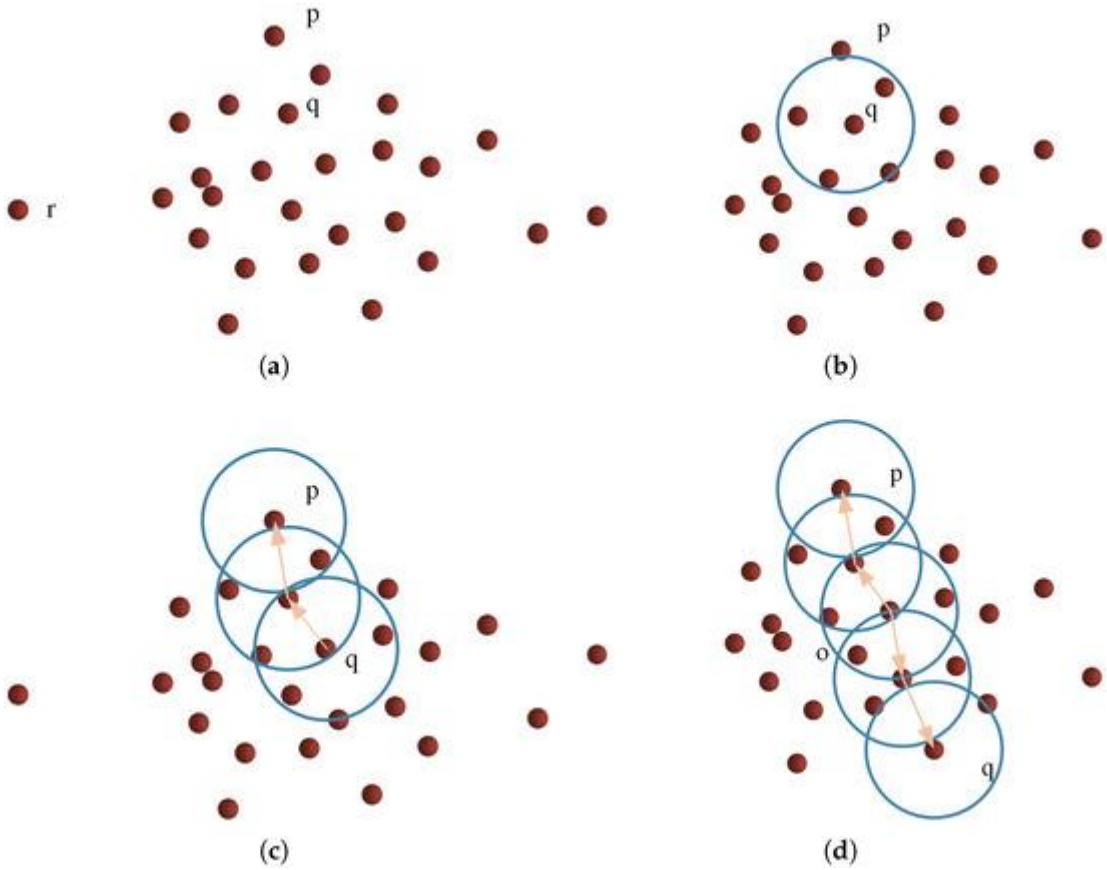

Рисунок 3.3. – Ілюстрація концепцій DBSCAN. (a) Різні типи точок: p – гранична точка, q – центральна точка, а r – точка шуму. (b) p є досяжною за щільністю з q, але q не є безпосередньо досяжною за щільністю з p. Це демонструє асиметрію досяжності за щільністю. (c) p є досяжною за щільністю з q, але q не є досяжною за щільністю з p, оскільки q не є центральною точкою. (d) p та q пов'язані за щільністю через ланцюг центральних точок (наприклад, o), що означає, що існує послідовність кроків досяжності за щільністю, що пов'язують їх разом

Враховуючи бал xi, той/та/те ε-сусідство визначається наступним чином:

$$N_\varepsilon(x_i) = \{x_j \mid \|x_j - x_i\| \leq \varepsilon\} \qquad (3.16.)$$

Якщо $|N\varepsilon(xi)| \geq MinPts$, тоді xi є ключовою точкою. Точка xj безпосередньо досяжна за щільністю з xi якщо $xj \in N\varepsilon(xi)$ і xi є ключовим моментом:

$x_j \in N_\varepsilon(x_i)$, xi є ядром точки (3.17.)

Потім для формування кластерів використовуються показники досяжності за щільністю та зв'язності за щільністю. DBSCAN починається з довільної точки та рекурсивно розширює кластер з основних точок та їхніх сусідніх точок, досяжних за щільністю.

Порівняно з традиційними методами, такими як K-середні, DBSCAN не вимагає попереднього визначення кількості кластерів і є більш стійким до викидів. Зазвичай він добре працює в сценаріях з неправильною формою кластерів або зашумленими даними.

3.4.2. SOM для кластеризації

SOM – це алгоритм ненавченої нейронної мережі, запропонований Теуво Кохоненом. Він зазвичай використовується для кластеризації даних, зменшення розмірності та завдань візуалізації. Метою SOM є відображення високовимірних нелінійних даних у низьковимірну (зазвичай 2D) сітку, зберігаючи при цьому топологічну структуру вихідних даних. Ілюстрацію методу SOM показано на рисунку 3.4.

Припустимо, що вхідний зразок $\epsilon x \in \mathbb{R}d$, а вихідний простір — це двовимірна сітка, що складається з m нейронів. Кожен нейрон i має ваговий вектор $wi \in \mathbb{R}d$ Процес навчання SOM включає такі кроки:

Спочатку виберіть нейрон, вага якого найближча до вхідної вибірки, як одиницю найкращого збігу (BMU):

$$i^* = \arg\min_i \|x - w_i\| \qquad (3.18.)$$

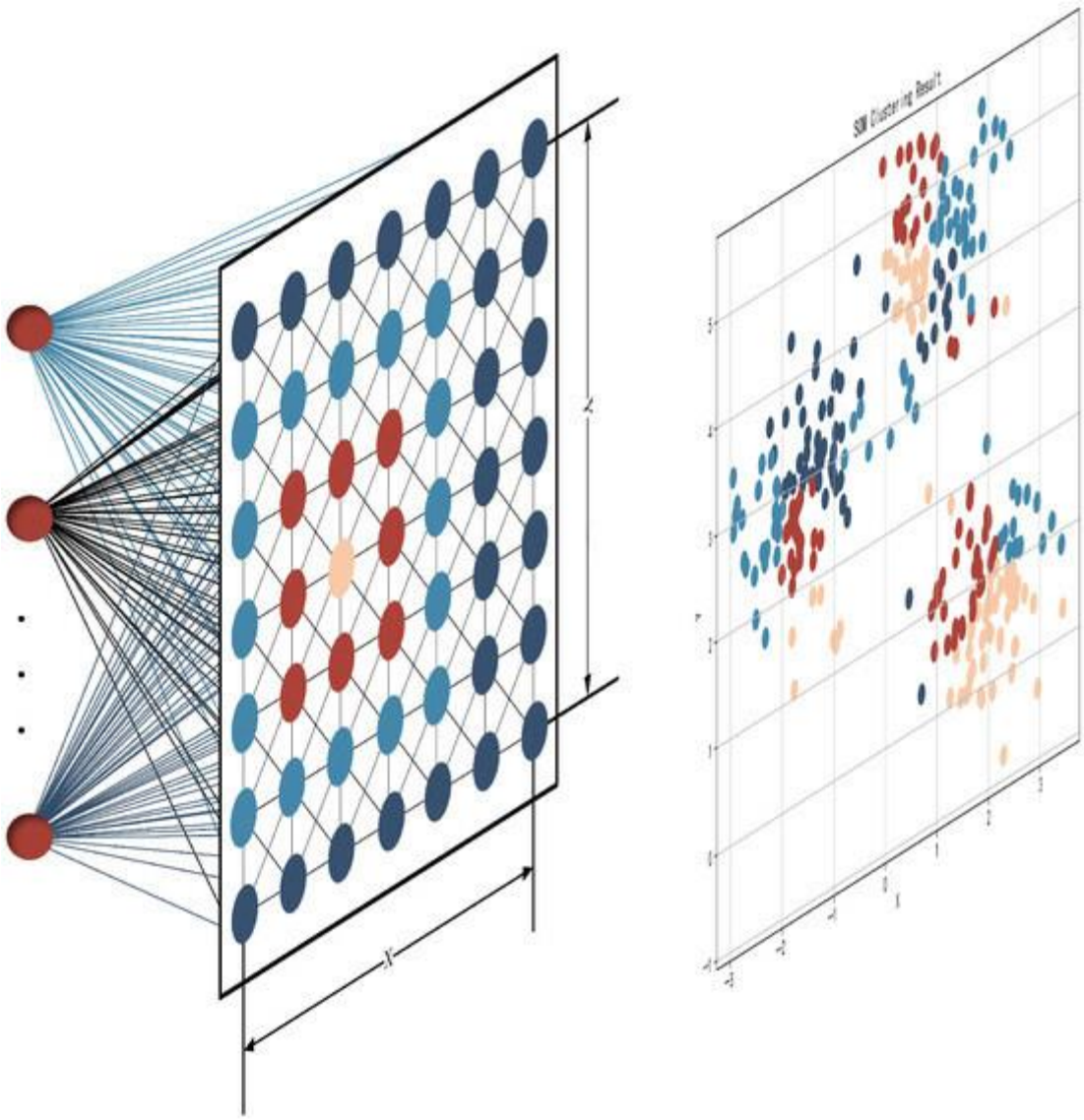

Рисунок 3.4. – Ілюстрація методу SOM

Потім оновіть ваги BMU та сусідніх нейронів:

$$w_i(t+1) = w_i(t) + \eta(t) \cdot h_{i,i^*}(t) \cdot (x - w_i(t)) \qquad (3.19.)$$

де $\eta(t)$– швидкість навчання, а hi, $i^*(t)$ – функція сусідства, що визначається як

$$h_{i,i^*}(t) = \exp\left(-\frac{\|r_i - r_{i^*}\|^2}{2\sigma^2(t)}\right) \qquad (3.20.)$$

Швидкість навчання та радіус околиці з часом зменшуються з такими функціями оновлення:

$$\eta(t) = \eta_0 \cdot \exp\left(-\frac{t}{\tau'}\right) \qquad (3.21.)$$

$$\sigma(t) = \sigma_0 \cdot \exp\left(-\frac{t}{\tau}\right) \qquad (3.22.)$$

Після навчання SOM формує візуальну топологічну карту, де кожен нейрон представляє кластер даних. Дані, призначені сусіднім нейронам, часто мають схожі топологічні ознаки.

3.4.3. Визначення номера кластера

Визначення кількості кластерів є критично важливим завданням у кластерному аналізі. Неправильний вибір може призвести до неправильної інтерпретації базової структури даних. Щоб визначити відповідну кількість кластерів, ми використовуємо дві метрики валідації: коефіцієнт силуету та індекс Калінського-Харабаша (CH).

Коефіцієнт силуету кількісно визначає, наскільки добре розділені кластери, вимірюючи подібність вибірки до власного кластера порівняно з іншими кластерами. Його можна розрахувати наступним чином:

$$s(i) = \frac{b(i) - a(i)}{\max a(i), b(i)} \qquad (3.23.)$$

де $a(i)$ – середня відстань від вибірки i до всіх інших точок у тому ж кластері, та $b(i)$– середня відстань до всіх точок найближчого кластера. Високе значенняs (i) вказує на те, що вибірка добре відповідає власному кластеру. Найкращий номер кластера вибирається для максимізації середнього балу Silhouette Score по всіх вибірках.

Індекс CH оцінює кластерну валідність на основі співвідношення міжкластерної дисперсії до внутрішньокластерної дисперсії. Він визначається як

$$CH(k) = \frac{SS_B/(k-1)}{SS_W/(N-k)} \qquad (3.24.)$$

де N – загальна кількість точок даних, k – кількість кластерів,SSB– це міжкластерна сума квадратів, а SS_W– це сума квадратів всередині кластера. Оптимальна кількість кластерів – це та, яка дає найвищий бал CH.

3.5. Стратегії кластеризації та ф'южну

Щоб зменшити невідповідності та нестабільність між різними алгоритмами кластеризації та отримати більш надійний результат ідентифікації топології, у цьому дослідженні:

- Застосовує стратегію об'єднання з урахуванням структури, що базується на кластеризації KMeans у просторі вбудовування міток;
- Реалізує стратегію агрегації на рівні рішень шляхом голосування зваженою більшістю для цілей порівняння.

Обидві стратегії спрямовані на інтеграцію результатів, отриманих за допомогою кількох базових кластеризацій. Їхні основні ідеї описані нижче.

$$l_i^{(k)} \in \mathcal{L}_k = \{0, 1, \ldots, K_k - 1\} \qquad (3.25.)$$

де Kk позначає кількість кластерів, повернутих k-м алгоритмом. Оскільки різні алгоритми кластеризації можуть створювати невідповідну кількість або порядок міток кластерів, ми перетворюємо набори міток на уніфіковане категоріальне кодування.

3.5.1. Об'єднання кластеризації на основі KMeans

Стратегія об'єднання кластерів на основі KMeans дозволяє спільне навчання кластерних структур у новому просторі, де згруповані зразки з подібними шаблонами багатовидових міток, як показано на рисунку 3.5.

Стратегія об'єднання кластерів на основі KMeans дозволяє спільне навчання кластерних структур у новому просторі, де згруповані зразки з подібними шаблонами багатовидових міток, як показано на рисунку 3.5

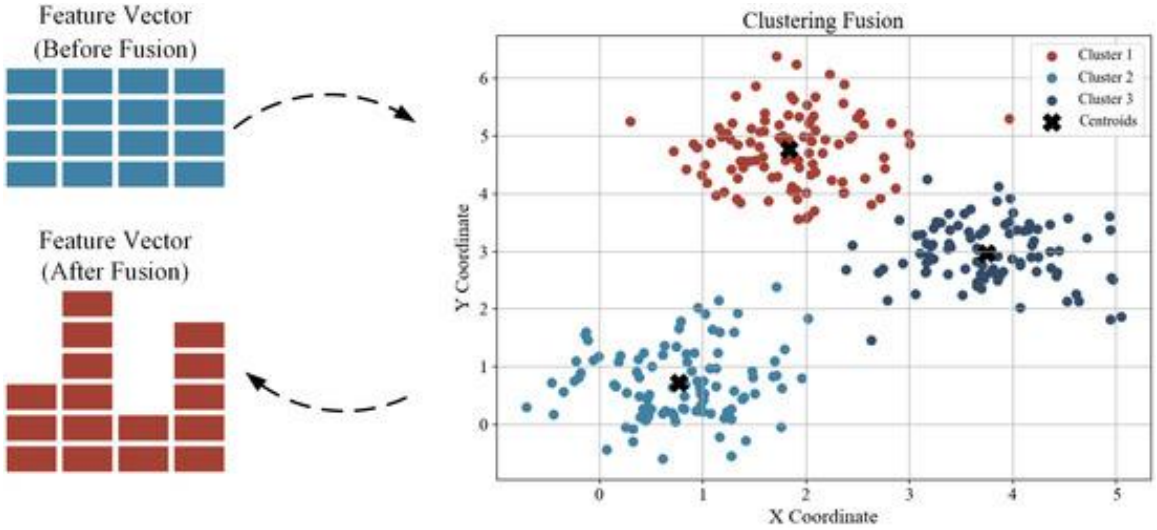

Рисунок 3.5. – Ілюстрація стратегії об'єднання кластерів на основі KMeans

Тут ми визначаємо вбудовування міток для кожного зразка шляхом об'єднання його категоріальних кластерних призначень з усіх базових алгоритмів:

$$\mathbf{l}_i = [l_i^{(1)}, l_i^{(2)}, ..., l_i^{(M)}]^\top \in \mathbb{R}^M \qquad (3.26.)$$

Шляхом об'єднання векторів міток усіх n вибірок ми будуємо матрицю вбудовування міток:

$$L = \left[\mathbf{l}_1, \mathbf{l}_2 \ldots, \mathbf{l}_n\right]^{\top} \in \mathbb{R}^{n \times M} \qquad (3.27.)$$

Потім ми виконуємо другий раунд кластеризації KMeans на L, щоб отримати остаточні мітки $\tilde{l}i$:

$$L = \left[\mathbf{l}_1, \mathbf{l}_2 \ldots, \mathbf{l}_n\right]^\top \in \mathbb{R}^{n \times M} \qquad (3.28.)$$

де Cj – j-й кластер, аμjє його центроїдом у просторі міток. Остаточне призначення таке

$$\min_{\{C_j\}_{j=1}^K} \sum_{j=1}^{K} \sum_{\mathbf{l}_i \in C_j} \|\mathbf{l}_i - \boldsymbol{\mu}_j\|_2^2 \qquad (3.29.)$$

Псевдокод показано в Алгоритмі 2, який підсумовує кроки для побудови матриці міток та виконання KMeans Fusion.

$$\hat{l}_i = \arg\min_j \|\mathbf{l}_i - \boldsymbol{\mu}_j\|_2 \qquad (3.30.)$$

Псевдокод показано в Алгоритмі 2, який підсумовує кроки для побудови матриці міток та виконання KMeans Fusion.

Algorithm 2 KMeans-Based Clustering Fusion Strategy

Require: Base cluster labels $l_i^{(k)}$ for $i = 1, \ldots, n, k = 1, \ldots, M$; number of fused clusters K
Ensure: Final label $\hat{l}_i$ for each sample
1: **for** $i = 1$ to n **do**
2: $\quad \mathbf{l}_i \leftarrow [l_i^{(1)}, l_i^{(2)}, \ldots, l_i^{(M)}]$
3: **end for**
4: $L \leftarrow$ matrix formed by stacking all $\mathbf{l}_i$
5: Apply KMeans to L with K clusters
6: **for** $i = 1$ to n **do**
7: $\quad \hat{l}_i \leftarrow$ assignment of $L[i, :]$ to nearest centroid
8: **end for**
9: **return** $\{\hat{l}_i\}_{i=1}^n$

3.5.2. Об'єднання кластеризації на основі голосування

У цьому дослідженні стратегія об'єднання на основі голосування реалізована виключно для цілей порівняння. Вона агрегує результати кластеризації баз за допомогою аналізу частоти міток. Як показано на рисунку 3.6. цей підхід отримує остаточну оцінку міток за допомогою консенсусного моделювання.

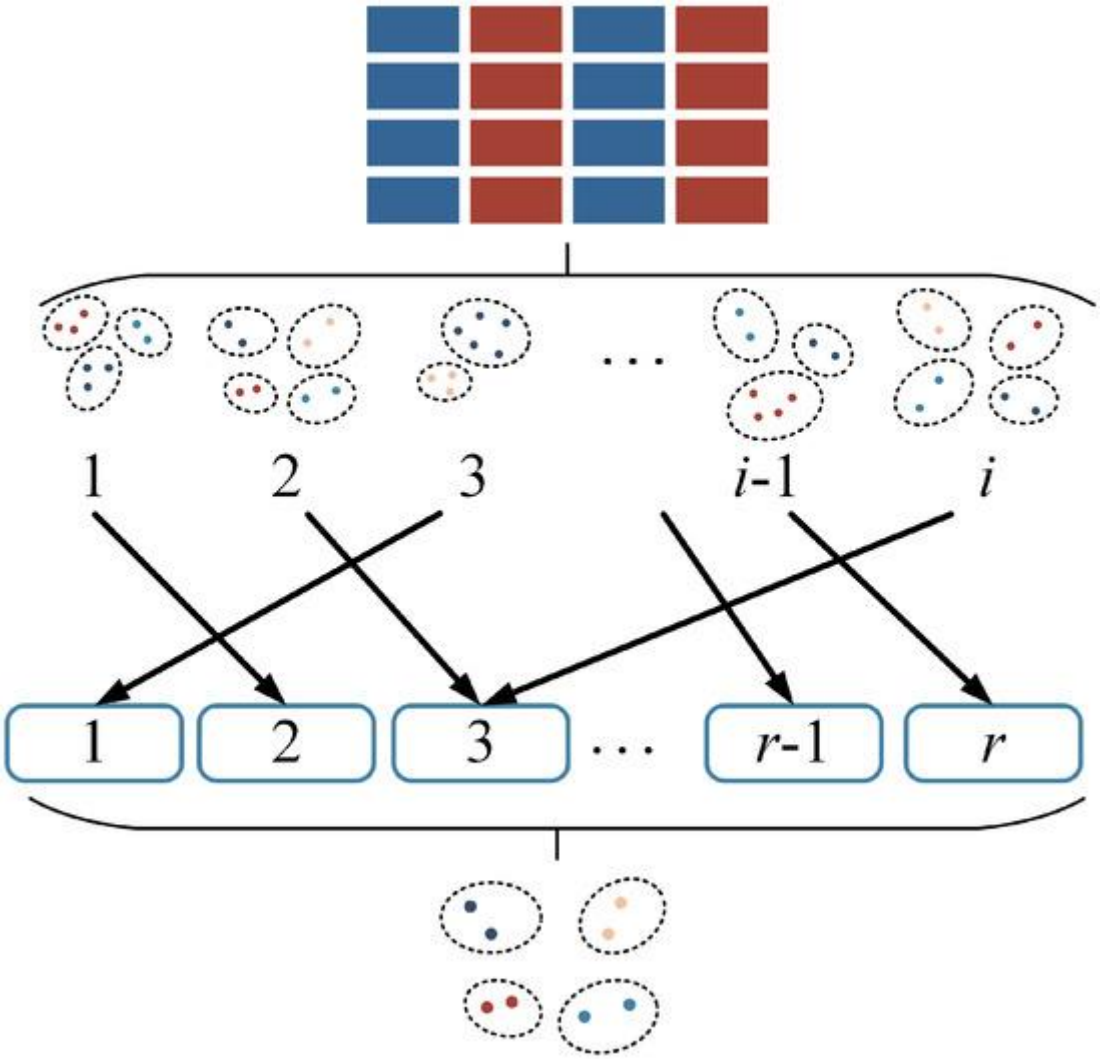

Рисунок 3.6. – Ілюстрація стратегії об'єднання кластерів на основі голосування

Остаточне маркування для кожного зразка визначається

$$\hat{l}_i = \arg\max_{c \in \mathcal{L}} \sum_{k=1}^{M} \mathbb{I}(l_i^{(k)} = c) \qquad (3.31.)$$

де $\mathscr{L}$ позначає об'єднання всіх наборів міток, а$\mathbb{I}(\cdot)$ є індикаторною функцією.

Щоб врахувати різну надійність різних алгоритмів кластеризації, ми визначаємо вектор ваг:

$$\mathbf{w} = [w_1, w_2, \ldots, w_M]^\top \in \mathbb{R}^M, \quad \text{with} \quad \sum_{k=1}^{M} w_k = 1 \tag{3.32.}$$

Тоді правило зваженого злиття має вигляд

$$\hat{l}_i = \arg\max_{c \in \mathcal{L}} \sum_{k=1}^{M} w_k \cdot \mathbb{I}(l_i^{(k)} = c) \tag{3.33.}$$

Псевдокод для цієї процедури голосування зваженою більшістю представлено в алгоритмі 3.

Algorithm 3 Voting-Based Clustering Fusion Strategy

Require: Base cluster labels $l_i^{(k)}$ for $i = 1, \ldots, n, k = 1, \ldots, M$; weight vector $\mathbf{w}$
Ensure: Final label $\hat{l}_i$ for each sample
1: **for** $i = 1$ to n **do**
2: Initialize vote dictionary $v_c \leftarrow 0$ for all $c \in \mathcal{L}$
3: **for** $k = 1$ to M **do**
4: $v_{l_i^{(k)}} \leftarrow v_{l_i^{(k)}} + w_k$
5: **end for**
6: $\hat{l}_i \leftarrow \arg\max_c v_c$
7: **end for**
8: **return** $\{\hat{l}_i\}_{i=1}^n$

ВИСНОВКИ К РОЗДІЛУ 3

Зрештою, загальна блок-схема запропонованої нами структури зображена на рисунку 3.7.

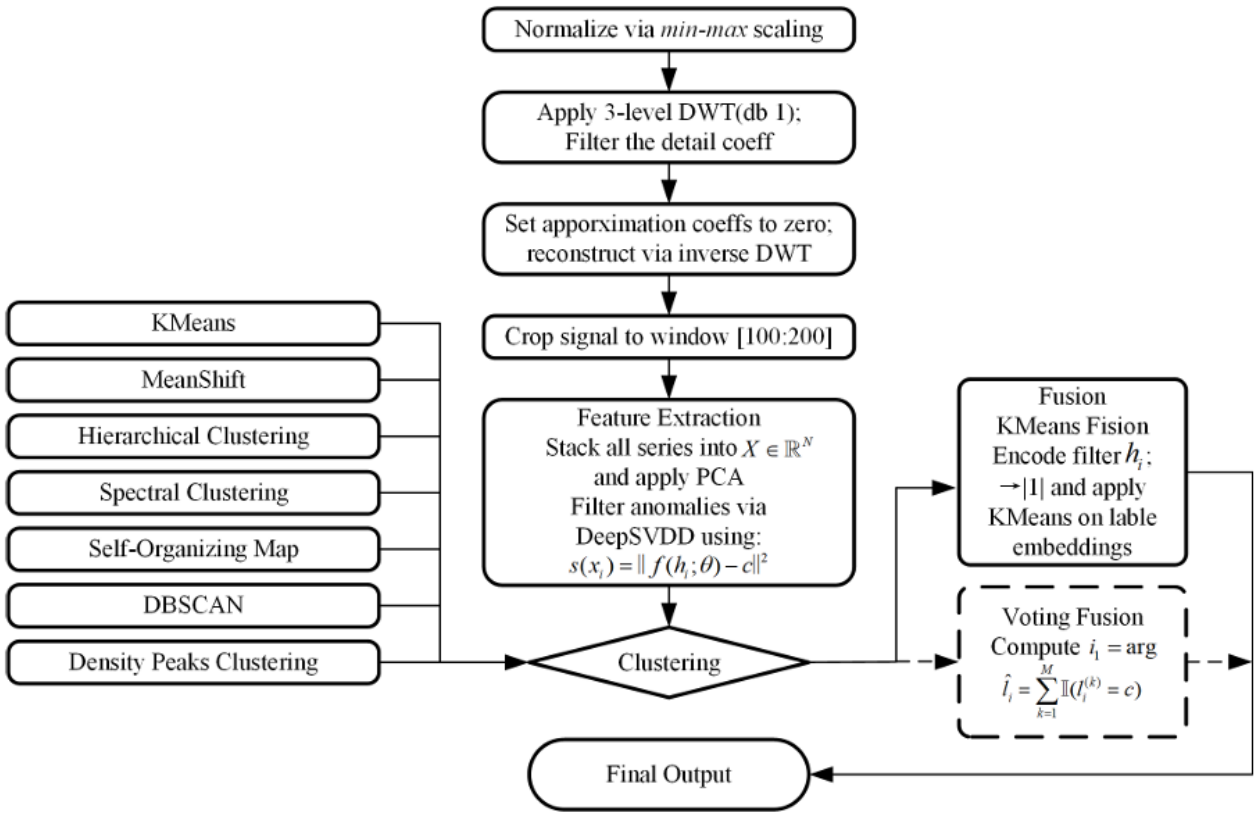

Рисунок 3.7. – Блок-схема запропонованої структури ідентифікації топології

РОЗДІЛ 4. ТЕМАТИЧНЕ ДОСЛІДЖЕННЯ

4.1. Експериментальна установка

Ці експерименти базуються на реальних даних, зібраних з трьох реальних житлових електричних мереж (позначених як Сітка 1, 2 та 3). Для Сіток 1 та 2 дані записуються з 4 жовтня 2025 року по 8 жовтня 2025 року, тоді як Сітка 3 охоплює довший період з 1 січня 2025 року по 15 березня 2025 року. Всі набори даних включають вимірювання струму та напруги житлових споживачів, причому 96 зразків записуються на день з інтервалом 15 хвилин. Сітка 1 має загалом 84 користувачів, які розташовані на двох фідерах. Сітка 2 має 89 користувачів, розподілених по п'яти фідерах. У Сітці 3 19 користувачів розподілені між трьома фідерами. Хоча дані спочатку записуються щодня з інтервалом 15 хвилин, показання лічильника в деякі дні є неповними або пошкодженими. Тому ми проводимо процедуру очищення даних (наприклад, видалення порожніх записів). Потім ми вирівнюємо час початку та час завершення вимірювань кожного дня, зберігаючи лише мінімальний загальний часовий діапазон. Зрештою, кожен день може мати однакову довжину (наприклад, 84) дійсних точок вимірювання.

Нормалізація даних виконується за такою формулою:

$$x' = \frac{x - \min(X)}{\max(X) - \min(X)} \qquad (4.1)$$

де x — це вихідна точка даних, аx'– це перетворене значення.

Після нормалізації до кожної послідовності застосовується процес шумозаглушення на основі вейвлетів для покращення якості сигналу та придушення високочастотного шуму.

Конфігурації для семи алгоритмів кластеризації наведено в таблиці 4.1 .

Таблиця 4.1. Конфігурація алгоритмів кластеризації.

Алгоритм	**Деталі конфігурації**
KMeans	Кількість кластерів: 2 (Сітка 1), 5 (Сітка 2); випадкова ініціалізація
Середній зсув	Пропускна здатність оцінюється на основі щільності даних
DBSCAN	ε-радіус та мінімальні точки, встановлені на основі розподілу даних
ЦОД	Порогові значення, що базуються на показниках локальної щільності та відстані
Спектральне кластеризування	Матриця Лапласа, що використовується для спектрального вбудовування
Ієрархічна кластеризація	Агломеративна кластеризація зі зв'язком Варда
ШОВ	Розмір сітки: 2 × 3; швидкість навчання: 0,5; 500 ітерацій

Експерименти проводяться з використанням бібліотек Python (версія 3.9), таких як scikit-learn (версія 1.3.1), Keras (версія 2.11.0) та MiniSom (версія 2.3.0) для побудови та навчання моделі. Використані апаратні характеристики такі:

- Процесор: Intel i7-11700;
- Оперативна пам'ять: 32 ГБ;
- Графічний процесор: NVIDIA RTX 3080.

4.2. Продуктивність у реальних житлових електричних мережах

Для оцінки ефективності запропонованого фреймворку було проведено серію експериментів на всіх наборах даних з використанням повного конвеєра, який включає зменшення розмірності на основі PCA, вилучення часових ознак на основі LSTM, виявлення аномалій DeepSVDD та кластеризаційне об'єднання.

Щоб вирішити проблему перестановки міток, властиву кластеризації без учителя, ми використовуємо угорський алгоритм [43] для оптимального

зіставлення передбачуваних міток кластерів.$\hat{y}$з істинними груповими (тобто фідерними) мітками y , що описується наступним чином:

(1)

Нехай n – кількість точок даних, а C – кількість кластерів. Ми визначаємо матрицю витрат$\mathbf{M} \in \mathbb{R}C \times C$, де кожен запис $\mathcal{M}ij$ представляє кількість точок з передбачуваною міткою i та міткою істинності j:

$$A = \frac{1}{n} \sum_{k=1}^{n} \mathbb{I}(\pi^*(\hat{y}_k) = y_k) \qquad (4.2.)$$

Мета полягає в тому, щоб знайти взаємно однозначне відображенняπ: { 1 , … , C } → { 1 , … , C }що максимізує повне збігання:

$$\max_{\pi \in \mathcal{P}} \sum_{i=1}^{C} \mathcal{M}_{i,\pi(i)} \qquad (4.3.)$$

де $\mathcal{P}$— множина всіх перестановок над C елементами. Цю задачу можна ефективно вирішити за допомогою угорського алгоритму.

(2)

Після отримання оптимального відображення π*, ми перепризначаємо кожну передбачену мітку $\hat{y}$ k доπ* ($\hat{y}k$) перед обчисленням точності кластеризації. Точність кластеризації обчислюється шляхом ділення кількості правильно зіставлених точок (тобто точок, чия передбачувана мітка відповідає мітці істинної позиції після перерозподілу) на загальну кількість точок у сітці. Формула для точності кластеризації A задається так:

$$A = \frac{1}{n} \sum_{k=1}^{n} \mathbb{I}(\pi^*(\hat{y}_k) = y_k) \qquad (4.4.)$$

де n – загальна кількість точок даних, а$\mathbb{I}$ (π*($\hat{y}k$) =yk) — це функція-індикатор, яка повертає 1, якщо передбачена мітка після перепризначення дорівнює справжній мітці, та 0 в іншому випадку.

Як показано в Таблиці 4.2. , точність кластеризації оцінюється для наборів даних з трьох енергетичних мереж. Серед усіх протестованих алгоритмів DBSCAN досяг найвищої точності на Сітці 1 (83%) та Сітці 2 (79%), тоді як DBSCAN та KMeans найкраще показали себе на Сітці 3 (73%). Ієрархічна кластеризація також дала конкурентні результати з точністю 77%, 81% та 69% на Сітці 1, Сітці 2 та Сітці 3 відповідно. На противагу цьому, SOM та Spectral Clustering продемонстрували відносно низьку продуктивність, особливо на Сітці 1. Загалом, ці результати підкреслюють мінливість продуктивності різних алгоритмів кластеризації.

Таблиця 4.2. – Точність кластеризації запропонованого методу за різними алгоритмами кластеризації

Метод кластеризації	Точність сітки 1	Точність сітки 2	Точність сітки 3
DBSCAN	83%	79%	73%
ЦОД	73%	68%	71%
Головний уповноважений	77%	81%	69%
KMeans	71%	75%	73%
Середній зсув	73%	66%	60%
ШОВ	28%	44%	53%
Спектральне кластеризування	60%	67%	53%

Потім застосовуються та порівнюються дві стратегії кластеризації, як показано в Таблиці 4.3. та на Рисунку 4.1. Стратегія кластеризації на основі KMeans досягла точності кластеризації 88%, 83%, та 79% на сітці 1, сітці 2 та сітці 3 відповідно, із середньою точністю 84%. На противагу цьому, проста стратегія, заснована на голосуванні, отримала точність 80%, 76%, та 74%, усереднення 77%. Ці результати демонструють, що метод об'єднання на основі KMeans послідовно перевершує підхід на основі голосування, пропонуючи чудову адаптивність до різних житлових енергетичних мереж. Конкретні

результати ідентифікації топології для Grid1, Grid2 та Grid3 представлені в Таблиці 4.4., Таблиці 4.5. та Таблиці 4.6. відповідно.

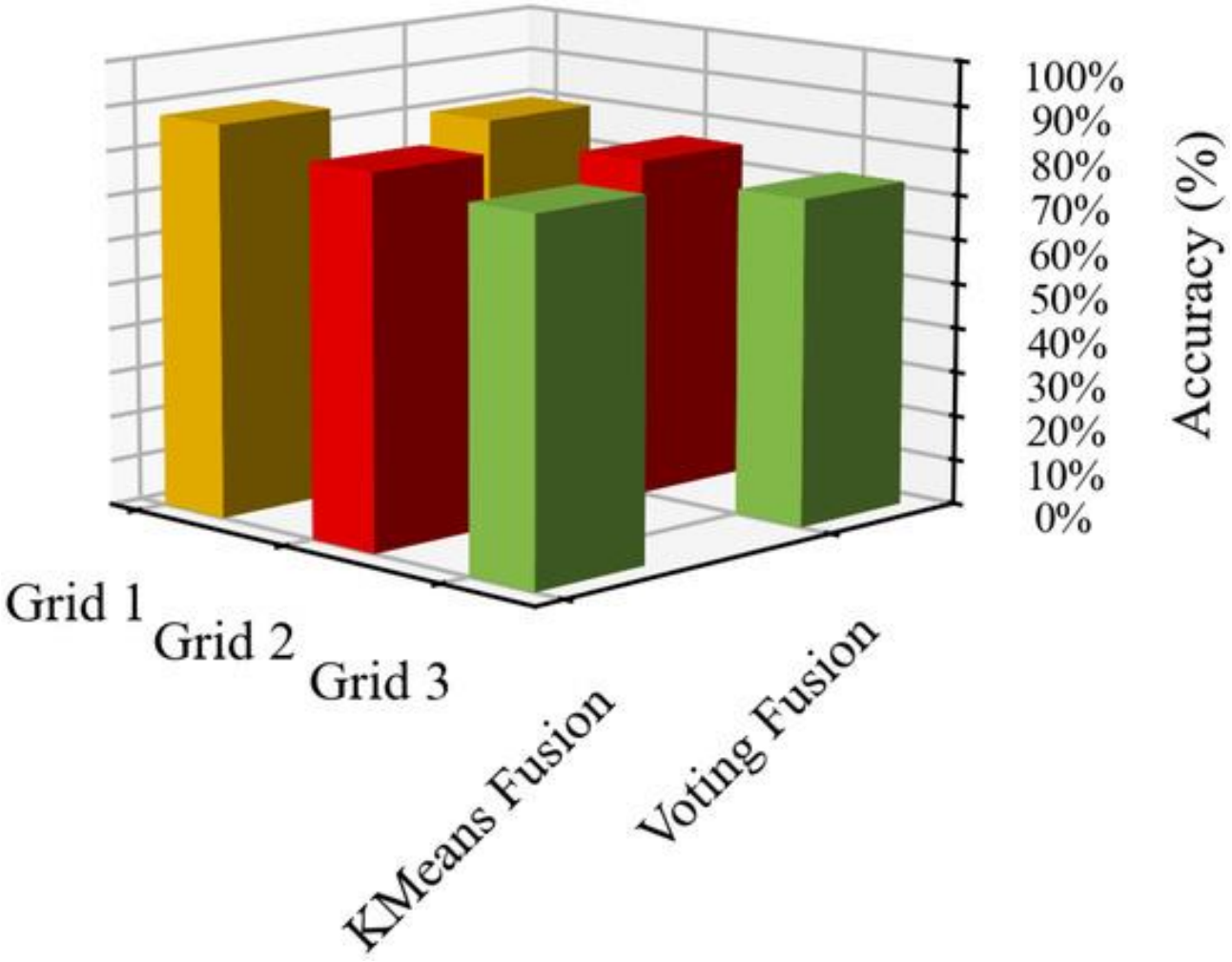

Рисунок 4.1. – Порівняння стратегій кластеризації з точки зору точності ідентифікації на сітці 1, сітці 2 та сітці 3. Кожна пара стовпців відповідає стратегії злиття (KMeans Fusion або Voting Fusion), причому кольори представляють різні сітки: зелений для сітки 1, жовтий для сітки 2 та червоний для сітки 3. Можна помітити, що KMeans Fusion послідовно досягає вищої точності, ніж Voting Fusion, на всіх трьох сітках

Таблиця 4.3. – Порівняння стратегій кластеризації з точки зору точності ідентифікації

Стратегія ф'южету	Точність сітки 1	Точність сітки 2	Точність сітки 3
Злиття KMeans	88 %	83 %	79 %
Голосування Fusion	80 %	76 %	74 %

Таблиця 4.4. – Порівняння між базовими даними та запропонованим методом у сітці 1

Ідентифікатор годівниці	Ідентифікатор користувача	
	Правдива відповідь	Запропонований метод
1	78, 79	18 років, 34, 74, 76, 77, 78,
2	2, 3, 4, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18 років, 19, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 80, 81, 82, 83, 84, 85	2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 2014, 15, 16, 17, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 35, 36, 37, 39, 40 41, 42, 43, 44, 45, 46 47, 48, 49, 50, 51, 52 54, 55, 56, 57, 58, 59 60, 62, 63, 64, 65, 66 67, 68, 69, 70, 71, 72 73, 75, 80, 81, 82, 83
Аномалія (зовнішні годівниці)	5	38, 53, 61, 84, 85

Таблиця 4.5. – Порівняння між базовими даними та запропонованим методом у таблиці 2

Ідентифікатор годівниці	Ідентифікатор користувача	
	Правдива відповідь	Запропонований метод
1	4, 5, 6, 7, 8, 9, 12, 19, 20, 21, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 60, 61, 62, 63, 64, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 89, 90	4, 5, 6, 8, 9, 17, 18 років, 19, 20, 21, 24, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 43, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 60, 61, 62, 64, 66, 67, 68, 69, 70, 71, 72, 73, 74, 76, 78, 79, 80, 81, 82, 83, 84, 85, 86, 89, 90
2	11, 14	11, 13, 14, 75
3	2, 3, 13, 16	2, 3, 13, 15, 16
4	15, 17, 18 років	7, 12, 31
5	87, 88	42, 44, 59, 65, 87, 88
Аномалія (зовнішні годівниці)	10, 22, 23, 59, 65	10, 22, 23, 45, 63, 77

Таблиця 4.6. – Порівняння між базовими даними та запропонованим методом у сітці 3

Ідентифікатор годівниці	Ідентифікатор користувача	
	Правдива відповідь	Запропонований метод
1	15, 16, 17, 18 років	15, 18 років
2	2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14	2, 3, 4, 5, 7, 8, 9, 10, 11, 12, 14, 17
Аномалія (зовнішні годівниці)	19, 20	6, 13, 19, 20

Для подальшого дослідження неправильно кластеризованих користувачів, часова мінливість їхніх послідовностей напруги та струму визначається кількісно. Для кожного лічильника m , послідовність напруги$vm(\ t\ )$і поточна серія$im(\ t\ )$ об'єднуються в один вектор:

$$\mathbf{x}_m \;=\; \left[\, v_m(1),\ldots,v_m(T_v),\; i_m(1),\ldots,i_m(T_i) \,\right]^{\top} \in \mathbb{R}^{T_v+T_i} \qquad (38)$$

Загальна мінливість у цьому комбінованому профілі потім вимірюється його дисперсією:

$$s_m^2 \;=\; \mathrm{Var}(\mathbf{x}_m) \;=\; \frac{1}{(T_v+T_i)-1}\sum_{k=1}^{T_v+T_i}\left(x_{m,k}-\bar{x}_m\right)^2, \quad \bar{x}_m = \frac{1}{T_v+T_i}\sum_{k=1}^{T_v+T_i} x_{m,k} \qquad (39)$$

Метри з дуже малими s^2_m демонструють слабку часову мінливість як напруги, так і струму, що робить їхні електричні сигнатури менш виразними. Отже, ці точки частіше плутають з іншими живильними елементами під час кластеризації та злиття, що пояснює значну частину залишкових помилок ідентифікації.

ВИСНОВКИ ДО РОЗДІЛУ 4

Як показано на рисунку 4.2 , часова дисперсія об'єднаних векторів напруги-струму значно менша для цих вимірювачів. Такі профілі з низькою дисперсією

означають, що відповідні електричні сигнатури менш виразні, що робить їх більш схильними до плутанини під час кластеризації та об'єднання.

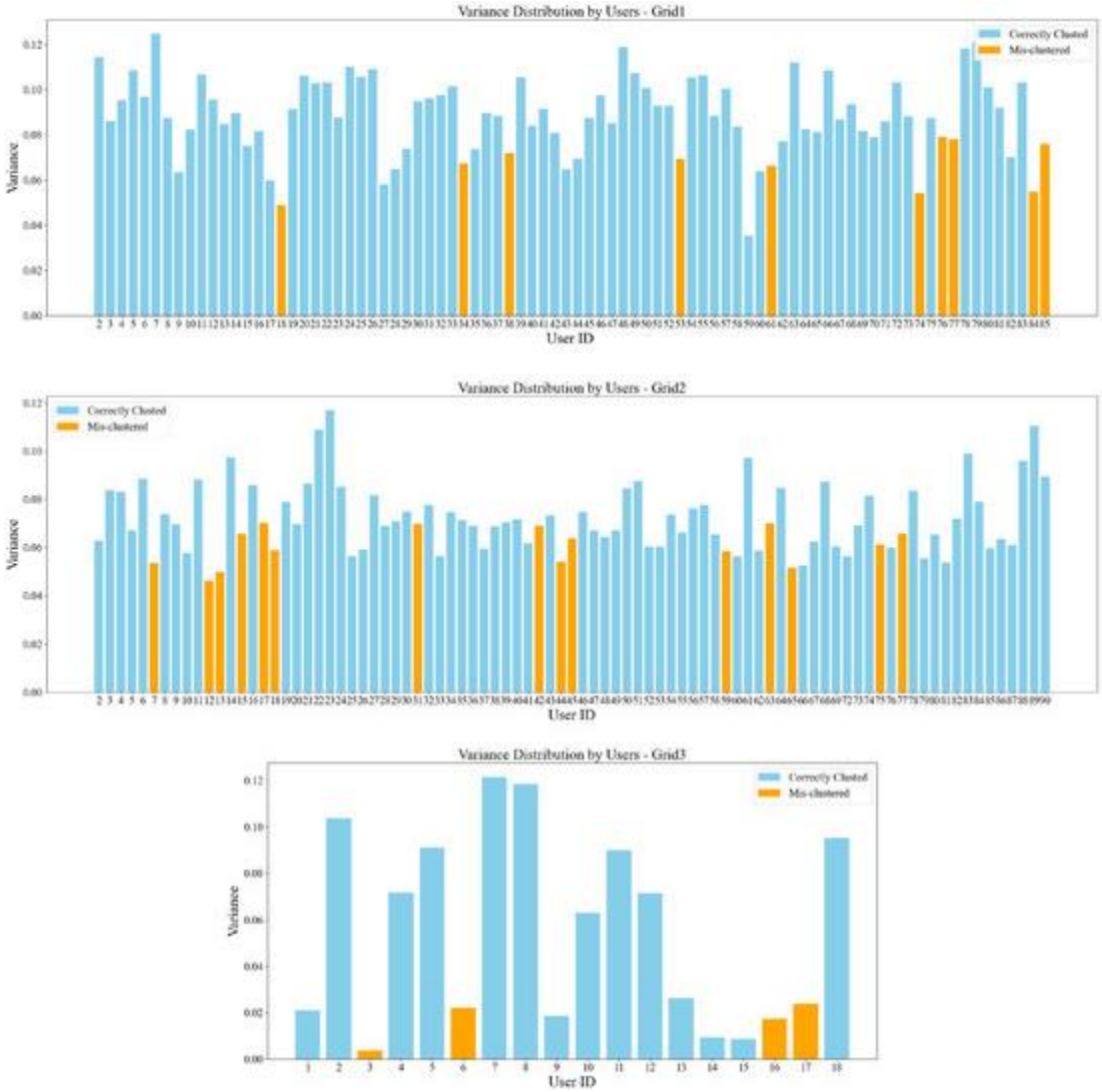

Рисунок 4.2. – Розподіл дисперсії об'єднаних векторів напруги-струму між користувачами в Сітці 1, Сітці 2 та Сітці 3, де середня дисперсія позначена червоною пунктирною лінією

РОЗДІЛ 5. ДОСЛІДЖЕННЯ ПОХИБКИ

Для всебічної оцінки внеску кожного модуля в запропоновану структуру ми проводимо експерименти з похибки за чотирма конфігураціями. Мета полягає в тому, щоб перевірити ефективність виділення часових ознак, виявлення аномалій та кластеризаційного об'єднання для точності ідентифікації топології.

5.1. Налаштування похибки

Розглядаються такі чотири налаштування:

(1)

PCA + iForest (без злиття) : PCA застосовується для зменшення розмірності, а аномалії видаляються за допомогою Ізоляційного лісу. Сім методів кластеризації використовуються окремо без будь-якого злиття.

(2)

PCA + LSTM + DeepSVDD (без злиття) : LSTM витягує часові ознаки з даних, відновлених за допомогою PCA. DeepSVDD використовується для виявлення аномалій. Кластеризація застосовується безпосередньо без злиття.

(3)

Сирі дані (без об'єднання) : Вихідні дані струму та напруги (по 480 точок кожна) використовуються для побудови 960-вимірних сирих ознак. Кластеризаційне об'єднання проводиться без PCA, LSTM або виявлення аномалій.

(4)

Повний фреймворк (PCA + LSTM + DeepSVDD + Fusion) : Повний фреймворк, що включає всі модулі.

5.2. Результати похибки

Результати похибки зведені в таблиці 5.1.

Таблиця 5.1. – Порівняння точності ідентифікації топології за різних налаштувань похибки

Експериментальна обстановка	Точність		
	Сітка 1	Сітка 2	Сітка 3
PCA + LSTM + iForest (без злиття)	–	–	–
DBSCAN	81%	74%	71%
ЦОД	57%	62%	69%
Головний уповноважений	59%	53%	53%
KMeans	61%	58%	73%
Середній зсув	80%	72%	73%
ШОВ	62%	47%	53%
Спектральне кластеризування	63%	43%	33%
PCA + LSTM + DeepSVDD (без злиття)	–	–	–
DBSCAN	83%	79%	73%
ЦОД	73%	68%	71%
Головний уповноважений	77%	81%	69%
KMeans	71%	75%	73%
Середній зсув	73%	66%	60%
ШОВ	28%	44%	53%
Спектральне кластеризування	60%	67%	53%
Необроблені дані (без об'єднання)	–	–	–
DBSCAN	61%	55%	53%
ЦОД	35%	32%	33%
Головний уповноважений	43%	28%	33%
KMeans	34%	24%	53%
Середній зсув	45%	26%	42%
ШОВ	30%	28%	20%
Спектральне кластеризування	59%	36%	27%
Повний фреймворк	–	–	–
Голосування Fusion	80%	76%	74%
Злиття KMeans	88%	83%	79%

З таблиці 5.1. можна зробити кілька ключових спостережень:

- Метод «Необроблені дані + злиття» забезпечує найнижчу точність. Вихідні дані містять шум та необроблені дані, що призводить до низької якості кластеризації навіть після злиття. Метод «MeanShift» у Grid 2 працює особливо погано, що підкреслює його нестабільність.
- PCA + iForest (без злиття) демонструє покращення для більшості методів. Ізоляційний ліс ефективно фільтрує аномалії, а PCA зменшує розмірність. Однак відсутність кластеризаційного злиття все ще призводить до варіацій, залежних від методу.
- PCA + LSTM + DeepSVDD (без об'єднання) досягає кращої загальної продуктивності. DeepSVDD забезпечує дрібнозернисте видалення аномалій, а LSTM фіксує часові закономірності. Тим не менш, продуктивність все ще варіюється між алгоритмами кластеризації через відсутність інтеграції на рівні міток.
- Повний фреймворк (запропонований нами фреймворк) дає найвищі та найстабільніші результати. Стратегія об'єднання покращує узагальнення та узгодженість між моделями кластеризації, тоді як PCA, LSTM та DeepSVDD доповнюють один одного в попередній обробці та розробці ознак.

Вищезазначені результати також зображено на рисунку 5.1. Як можна побачити з наведених тут результатів, кожен модуль у запропонованій системі відіграє вирішальну роль, і видалення будь-якого модуля може призвести до погіршення продуктивності.

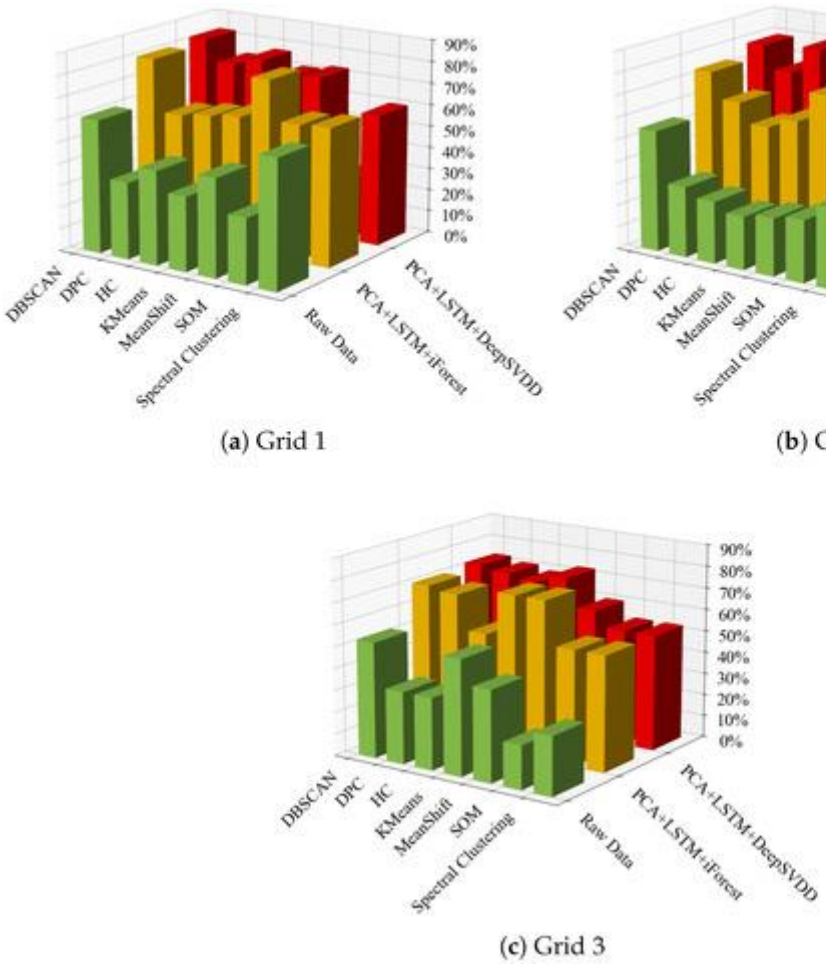

Рисунок 5.1. – Порівняння точності ідентифікації топології для різних конфігурацій злиття та алгоритмів кластеризації для трьох мереж розподілу електроенергії. Кожна 3D-гістограма ілюструє продуктивність необроблених даних, налаштувань PCA + LSTM + iForest та PCA + LSTM + DeepSVDD за різних методів кластеризації (наприклад, DBSCAN, KMeans, HC). Точність представлена висотою стовпчиків, і можна спостерігати постійні покращення із застосуванням вилучення ознак на основі LSTM та фільтрації аномалій DeepSVDD

Експериментальні результати для всіх трьох наборів даних сітки демонструють ефективність запропонованої системи ідентифікації топології на основі даних. Примітно, що Сітка 3, яка охоплює довший період моніторингу та демонструє складнішу поведінку навантаження, досягає подібних покращень до Сітки 1 та Сітки 2 при використанні повної системи. Стратегія об'єднання кластеризації на основі KMeans послідовно перевершує метод на основі

голосування, з підвищенням точності на 8%, 9% та 5% для Сіток 1, 2 та 3 відповідно. Ці результати свідчать про те, що вбудовування виходів кластеризації в єдиний простір міток дозволяє краще консенсусне моделювання в гетерогенних алгоритмах. Дослідження похибки додатково підтверджує, що виключення будь-якого основного модуля, зокрема часового моделювання на основі LSTM або фільтрації аномалій на основі DeepSVDD, призводить до суттєвого погіршення продуктивності. У Сітці 3 цей ефект особливо виражений, що підкреслює необхідність як глибокого вилучення ознак, так і придушення викидів.

5.3. Порівняння витрат часу

Результати витрат часу наведено в Таблиці 5.2. Як можна побачити, повний фреймворк вимагає найдовшого середнього часу виконання – 11,44 с, тоді як метод необроблених даних є найшвидшим – 5,14 с. Окрім повного фреймворку, результати в таблиці відповідають середньому часу виконання, отриманому для семи алгоритмів кластеризації, застосованих у дослідженні похибки. Методи PCA + LSTM + DeepSVDD (без злиття) та PCA + LSTM + iForest (без злиття) мають помірний час виконання – 9,21 та 8,08 с відповідно. Загалом, на **Рисунку 5.2** видно, що час виконання збільшується зі збільшенням кількості користувачів та складності сітки. Довший час виконання повного фреймворку можна пояснити головним чином включенням етапу навчання DeepSVDD. Мережа приймає як вхідні дані…$\mathbf{H} \in \mathbb{R}N \times d$матриця ознак, де d – це розмірність ознак, що виводиться LSTM. Процес навчання DeepSVDD включає багаторазове пряме та зворотне поширення, що робить цей етап порівняно трудомістким зі збільшенням N. У майбутній роботі основна увага буде зосереджена на подальшому підвищенні обчислювальної ефективності шляхом використання матричних операцій на основі GPU, оскільки етап навчання DeepSVDD може виграти від паралельного виконання.

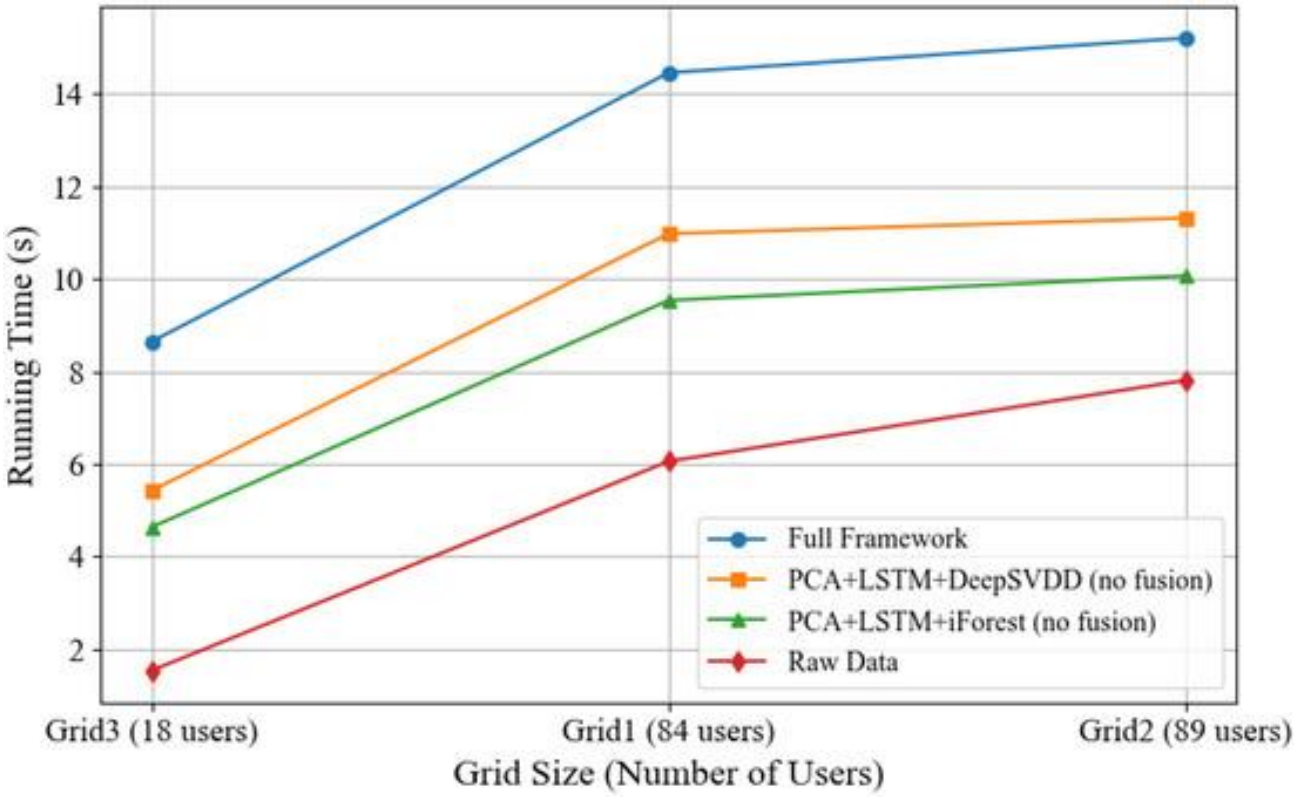

Рисунок 5.2. – Час виконання різних фреймворків за різних розмірів сітки (кількості користувачів)

Таблиця 5.2. – Порівняння середнього часу виконання програми за різними методами

Метод	Час виконання (с)			
	Сітка 1	Сітка 2	Сітка 3	Середній
Повний фреймворк	14.46	15.21	8.65	11.44
PCA + LSTM + DeepSVDD (без злиття)	10.98	11.32	5.34	9.21
PCA + LSTM + iForest (без злиття)	9.54	10.07	4.64	8.08
Необроблені дані	6.06	7.81	1,54	5.14

ВИСНОВКИ ДО РОЗДІЛУ 5

У реальних умовах впровадження інтелектуальних лічильників проблеми з якістю даних, такі як несправності датчиків, помилки зв'язку та відсутність записів, є неминучими та можуть негативно вплинути на продуктивність методів, керованих даними.

Щоб вирішити цю проблему, було розроблено три експерименти для моделювання збою передачі даних. Для кожного користувача 5%, 10% та 20% вибірок даних було випадковим чином видалено з відповідних часових рядів напруги та струму. Видалення проводилося незалежно для напруги та струму, дотримуючись простої стратегії випадкової вибірки без заміни, таким чином, щоб кожен запис мав однакову ймовірність бути виключеним.

Як показано в таблиці 5.3 , коли 20% вимірювань напруги та струму було випадковим чином видалено, точність ідентифікації топології за стратегіями Voting та KMeans Fusion демонструвала помірне зниження.

Таблиця 5.3. – Точність повної структури за різних показників втрати даних.

Втрата даних	Точність					
	Стратегія голосування			Стратегія KMeans		
	Сітка 1	Сітка 2	Сітка 3	Сітка 1	Сітка 2	Сітка 3
0%	80%	76%	74%	88%	83%	79%
5%	80%	76%	74%	88%	83%	79%
10%	76%	74%	74%	86%	82%	79%
20%	71%	70%	72%	75%	73%	72%

Як показано на рисунку 5.3 , точність стратегій Voting та KMeans Fusion поступово знижується зі збільшенням втрати даних. Тим не менш, точність фреймворку все ще залишається вище 70% для всіх сіток.

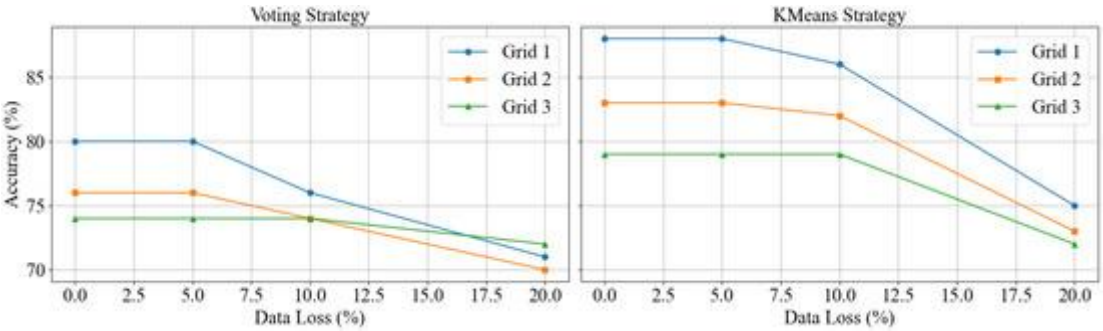

Рисунок 5.3. – Зміна точності стратегій Voting та KMeans Fusion за різних показників втрати даних

ВИСНОВКИ Й РЕКОМЕНДАЦІЇ

Забезпечення енергетичної безпеки є критично важливим завданням сучасної енергетики.

Це дослідження представляє модульну структуру для ідентифікації топології «користувач–живильник» у низьковольтних житлових електричних мережах з використанням даних часових рядів інтелектуальних лічильників навколишнього середовища. Запропонована структура поєднує вейвлет-шумозаглушення сигналів, PCA для зменшення розмірності, LSTM для захоплення часових залежностей, DeepSVDD для виявлення аномалій та стратегію об'єднання міток на основі KMeans для підвищення точності та надійності запропонованого методу. Експерименти на трьох реальних житлових електричних мережах підтверджують ефективність структури, де результати, отримані за допомогою стратегії об'єднання міток на основі KMeans, перевершують як окремі алгоритми кластеризації, так і стратегію об'єднання на основі голосування. Дослідження похибки підтверджує цінність кожного компонента, показуючи, що відсутність часового моделювання або фільтрації аномалій призводить до помітного погіршення продуктивності кластеризації. У майбутній роботі буде досліджено розширення запропонованої структури на багатофазне або графове моделювання розподільчих мереж, що розвиваються. Крім того, для подальшого підвищення точності результатів кластеризації будуть розглянуті такі підходи: (1) використання довших часових рядів; (2) застосування більш просунутих методів обробки сигналів, ніж вейвлет-методи (наприклад, емпіричне розкладання мод (EMD)) на етапі попередньої обробки даних.

СПИСОК ПОСИЛАНЬ

1. Конституція України: Конституція України; Верховна Рада України від 28.06.1996 № 254к/96–ВР // База даних «Законодавство України» / 45 Верховна Рада України. URL: https://zakon.rada.gov.ua/go/254%D0%BA/96–%D0%B2%D1%80
2. Омельченко Володимир. Сектор відновлюваної енергетики України до, під час та після війни. Центр Разумкова. 2022. URL: https://razumkov.org.ua/statti/sektor–vidnovlyuvanoyi–energetyky–ukrayiny–do–pidchas–ta–pislya–viyny
3. Про Концепцію діяльності органів виконавчої влади у забезпеченні енергетичної безпеки України : Постанова Кабінету Міністрів України від 19.01.1998 № 48 URL: https://zakon.rada.gov.ua/go/48–98–%D0%BF/
4. Про нафту і газ : Закон України від 12.07.2001 № 2665–III URL: https://zakon.rada.gov.ua/go/2665–14
5. Про Національний план дій з енергоефективності на період до 2030 року : Розпорядження Кабінету Міністрів України; План, Заходи від 12.2021 № 1803–р // База даних «Законодавство України» / Верховна Рада України. URL: https://zakon.rada.gov.ua/go/1803–2021–%D1%80
6. Про Національну комісію, що здійснює державне регулювання у сферах енергетики та комунальних послуг : Закон України від 22.09.2016 № 1540–VIII // База даних «Законодавство України» / Верховна Рада України. URL: https://zakon.rada.gov.ua/go/1540–19
7. Про ринок електричної енергії : Закон України від 13.04.2017 № 2019–VIII URL: https://zakon.rada.gov.ua/go/2019–19
8. Про схвалення Концепції впровадження “розумних мереж” в Україні до 2035 року : Розпорядження Кабінету Міністрів України; Концепція, План, Заходи від 14.10.2022 № 908–р // База даних «Законодавство України» / Верховна Рада України. URL: https://zakon.rada.gov.ua/go/908–2022–%D1%80 .

9. Про схвалення Плану розвитку системи передачі на 2021–2030 роки : Постанова; Нацком.енергетики, ком.послуг від 20.01.2021 № 57 // База даних «Законодавство України» / Верховна Рада України. URL: https://zakon.rada.gov.ua/go/v0057874–21
10. Паливно–енергетичний комплекс України на порозі третього тисячоліття [За заг. ред. А.К. Шидловського, М.П. Ковалка]. – К.: УЕЗ, – 2001. – 398 с.
11. Salman, U.T.; Shafiq, S.; Al-Ismail, F.S.; Khalid, M. A Review of Improvements in Power System Flexibility: Implementation, Operation and Economics. Electronics 2022, 11, 581
12. Shobole, A.A.; Wadi, M. Multiagent Systems Application for the Smart Grid Protection. Renew. Sustain. Energy Rev. 2021, 149
13. Hurayra, M.A.; Parvez, M.; Barua, R.; Haque, M.R.; Maruf, S.N.A.; Sahadat, M.A.; Alam, M.P.; Zuhaer, T. A Comprehensive Review: Distribution Network Management Strategy, Impact and Future. Eur. J. Appl. Sci. Eng. Technol. 2025, 3, 309–328
14. Othman, A.; Watson, N.R.; Lapthorn, A.; Mukhedkar, R. Utilising Smart-Meter Harmonic Data for Low-Voltage Network Topology Identification. Energies 2025, 18, 3333
15. Tong, X.; Kang, C.; Xia, Q. Smart Metering Load Data Compression Based on Load Feature Identification. IEEE Trans. Smart Grid 2016, 7, 2414–2422
16. Aschidamini, G.L.; da Cruz, G.A.; Resener, M.; Leborgne, R.C.; Pereira, L.A. A Framework for Reliability Assessment in Expansion Planning of Power Distribution Systems. Energies 2022, 15, 5073
17. Amoateng, D.O.; Yan, R.; Mosadeghy, M.; Saha, T.K. Topology Detection in Power Distribution Networks: A PMU Based Deep Learning Approach. IEEE Trans. Power Syst. 2021, 37, 2771–2782
18. Raghuvamsi, Y.; Teeparthi, K.; Kosana, V. A Novel Deep Learning Architecture for Distribution System Topology Identification with Missing PMU Measurements. Results Eng. 2022, 15

19. Tian, J.; Wang, P.; Zhu, D. Overview of Chinese New Energy Vehicle Industry and Policy Development. Green Energy Resour. 2024, 2, 100075
20. Qiu, W.; Sun, K.; Xiao, H. Advances in Urban Power Distribution System. Energies 2022, 15, 7329
21. IoTAnalytics. Smart Electricity Meter Market 2024: Global Adoption Landscape; IoT Analytics GmbH: Hamburg, Germany, 2024. Available online: https://iot-analytics.com (accessed on 1 September 2025)
22. Ye, X.; Zhang, Z.; Qiu, Y. Review of Application of High Frequency Smart Meter Data in Energy Economics and Policy Research. Front. Sustain. Energy Policy 2023, 2, 1171093
23. Wang,Y.; Chen, Q.; Hong, T.; Kang, C. Review of Smart Meter Data Analytics: Applications, Methodologies, and Challenges. IEEE Trans. Smart Grid 2018, 10, 3125–3148
24. Geth, F.; Vanin, M.; Van Hertem, D. Data Quality Challenges in Existing Distribution Network Datasets. arXiv 2023, arXiv:2308.00487
25. Pengwah, A.B.; Fang, L.; Razzaghi, R.; Andrew, L.L.H. Topology Identification of Radial Distribution Networks Using Smart Meter Data. IEEE Syst. J. 2022, 16, 5708–5718
26. Zhou, Y.; Li, Z.; Bao, X.; Mo, B. Convex Optimization Topology Identification of Distribution Network Under Distributed Conditions; College of Electrical Engineering, Shanghai University of Electric Power: Shanghai, China, 2023
27. Liu, L.; Zhao, X. Application of Improved Adjacency Matrix Multiplication in Distribution Network Flow Calculation. Energy Procedia 2012, 14, 983–989
28. Noschese, S.; Reichel, L. Network analysis with the aid of the path length matrix. Numer. Linear Algebra Appl. 2024, 95, 451–470
29. Kurup, A.; Summers, A.; Bidram, A.; Reno, M.J.; Martínez-Ramón, M. Ensemble models for circuit topology estimation, fault detection and classification in distribution systems. Sustain. Energy Grids Netw. 2023, 34, 101017

30. Zhang,J.; Wang, Y.; Weng, Y.; Zhang, N. Topology identification and line parameter estimation for non-PMU distribution network: Anumerical method. IEEE Trans. Smart Grid 2020, 11, 4440–4453
31. Francis, C.; Rao, V.; Trevizan, R.D.; Reno, M.J. Topology identification of power distribution systems using time series of voltage measurements. In Proceedings of the IEEE Power and Energy Conference at Illinois (PECI), Urbana, IL, USA, 1–2 April 2021; IEEE: Urbana, IL, USA, 2021; pp. 1–7.
32. Sun, L.; Chen, Y.; Du, Q.; Ding, R.; Liu, Z.; Cheng, Q. Topology identification of low-voltage power lines based on IEC 61850 and the clustering method. Energies 2023, 16, 1126
33. Wang,C.; Lou, Z.; Li, M.; Zhu, C.; Jing, D. Identification of Distribution Network Topology and Line Parameter Based on Smart Meter Measurements. Energies 2024, 17, 830
34. Li, X.; Han, B.; Li, G.; Luo, L.; Wang, K.; Jiang, X. Dynamic Topology Awareness in Active Distribution Networks Under DG Uncertainties Using GMM-PSEs and KL Divergence. IEEE Trans. Sustain. Energy 2021, 12, 2086–2096
35. Liu, Y.; Ren, P.; Zhao, J.; Liu, T.; Wang, Z.; Tang, Z. Real-Time Topology Estimation for Active Distribution System Using Graph-Bank Tracking Bayesian Networks. IEEE Trans. Ind. Inform. 2023, 19, 6127–6137
36. Zhou, N.; Luo, L.; Sheng, G.; Jiang, X. Power distribution network dynamic topology awareness and localization based on subspace perturbation model. IEEE Trans. Power Syst. 2019, 35, 1479–1488
37. Energies 2025, 18, 4908 26 of 26 27. Deka, D.; Talukdar, S.; Chertkov, M.; Salapaka, M.V. Graphical models in meshed distribution grids: Topology estimation, change detection & limitations. IEEE Trans. Smart Grid 2020, 11, 4299–4310
38. Chandola, V.; Banerjee, A.; Kumar, V. Anomaly Detection: A Survey. ACM Comput. Surv. 2009, 41, 15

39. Pearson, K. On lines and planes of closest fit to systems of points in space. Lond. Edinb. Dublin Philos. Mag. J. Sci. 1901, 2, 559–572
40. Montgomery, D.C.; Runger, G.C. Applied Statistics and Probability for Engineers, 5th ed.; Wiley: Hoboken, NJ, USA, 2012
41. Zhu,Y.; Lu, S. Load profile disaggregation by blind source separation: A wavelets-assisted independent component analysis approach. In Proceedings of the 2014 IEEE PES General Meeting | Conference & Exposition, National Harbor, MD, USA, 27–31 July 2014; IEEE: National Harbor, MD, USA, 2014; pp. 1–5
42. Tukey, J.W. Exploratory Data Analysis; Addison-Wesley: Reading, MA, USA, 1977
43. Ester, M.; Kriegel, H.-P.; Sander, J.; Xu, X. A density-based algorithm for discovering clusters in large spatial databases with noise. In Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining (KDD), Portland OR, USA, 2–4 August 1996; AAAI Press: Cambridge, MA , USA, 1996; pp. 226–231
44. Priyadarsini, M.; Sonekar, N. A CNN-based approach for anomaly detection in smart grid systems. Electr. Power Syst. Res. 2025, 238, 111077
45. Zhu,Y.; Lu, S. Image embedding of PMU data for deep learning towards transient disturbance classification. In Proceedings of the 2018 IEEE International Conference on Energy Internet (ICEI), Beijing, China, 21–25 May 2018; IEEE: Beijing, China, 2018; pp. 169–174
46. Tao, Y.; Yan, J.; Niu, E.; Zhai, P.; Zhang, S. An SVM-Based Anomaly Detection Method for Power System Security Analysis Using Particle Swarm Optimization and t-SNE for High-Dimensional Data Classification. Processes 2025, 13, 549
47. Zideh, M.J.; Chatterjee, P.; Srivastava, A.K. Physics-Informed Machine Learning for Data Anomaly Detection, Classification, Localization, and Mitigation: A Review, Challenges, and Path Forward. arXiv 2023, arXiv:2309.10788
48. Ma,Y.Z. ATutorial on Principal Component Analysis. Technical Report. arXiv 2014

49. Ruff, L.; Vandermeulen, R.; Görnitz, N.; Deecke, L.; Siddiqui, S.A.; Binder, A.; Müller, E.; Kloft, M. Deep One-Class Classifica tion. In Proceedings of the 35th International Conference on Machine Learning (ICML), Stockholm, Sweden, 10–15 July 2018; pp. 4393–4402
50. Liu, F. T.; Ting, K. M.; Zhou, Z.-H. Isolation Forest. In Proceedings of the 2008 Eighth IEEE International Conference on Data Mining (ICDM), Pisa, Italy, 15–19 December 2008; pp. 413–422
51. Chalapathy, R.; Chawla, S. Deep learning for anomaly detection: A survey. arXiv 2019, arXiv:1901.03407
52. scikit-learn: User Guide-Unsupervised Learning-Clustering. Available online: https://scikit-learn.org/stable/modules/ clustering (accessed on 1 September 2025)
53. Xie, J.; Girshick, R.; Farhadi, A. Unsupervised Deep Embedding for Clustering Analysis. In Proceedings of the 33rd International Conference on Machine Learning (ICML), New York, NY, USA, 19–24 June 2016; pp. 478–487.