

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ  
ТЕРНОПІЛЬСЬКИЙ НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ  
ІМЕНІ ІВАНА ПУЛЮЯ

Кваліфікаційна наукова праця  
на правах рукопису

**ВОЛОЩУК АНДРІЙ ВОЛОДИМИРОВИЧ**

УДК:519.2:004.8:621.3

**ДИСЕРТАЦІЯ**

**МОДЕЛІ ТА МЕТОДИ АНАЛІЗУ ДАНИХ ЕЛЕКТРОСПОЖИВАННЯ В  
РОЗПОДІЛЕНИХ ЕЛЕКТРОЕНЕРГЕТИЧНИХ СИСТЕМАХ**

122 – Комп'ютерні науки

12 – Інформаційні технології

Подається на здобуття ступеня доктора філософії

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

\_\_\_\_\_ /А.В. Волощук/

Наукові керівники:  
кандидат технічних наук, доцент  
Осухівська Галина Михайлівна,  
доктор технічних наук, професор  
Сверстюк Андрій Степанович

Тернопіль – 2026

## АНОТАЦІЯ

*Волощук А.В.* Моделі та методи аналізу даних електроспоживання в розподілених електроенергетичних системах. – Кваліфікаційна наукова праця на правах рукопису.

Дисертація на здобуття ступеня доктора філософії за спеціальністю 122 – Комп’ютерні науки. Тернопільський національний технічний університет імені Івана Пулюя, Тернопіль, 2026.

У дисертаційній роботі розв’язано актуальне науково-прикладне завдання аналізу даних електроспоживання на основі моделі у вигляді періодично корельованих випадкових процесів (ПКВП) у поєднанні із методами машинного навчання (МН) для класифікації режимів електроспоживання в розподілених електроенергетичних системах.

У вступі обґрунтовано актуальність теми дисертації, визначено зв’язок роботи з науковими темами, сформульовано мету і завдання дослідження, визначено об’єкт, предмет і методи дослідження, наведено загальну характеристику дисертації, окреслено наукову новизну отриманих результатів та їхнє практичне значення, а також розкрито питання апробації результатів дисертації та їх висвітлення в наукових працях.

У першому розділі виконано аналіз сучасного стану електроенергетичних систем в умовах розвитку розподіленої генерації та інтеграції відновлюваних джерел енергії в концепції Energy 4.0, зокрема в контексті синхронізації ОЕС України з континентальною мережею ENTSO–E, що підвищує вимоги до оперативності та надійності аналізу даних електроспоживання. Розглянуто існуючі підходи до аналізу даних електроспоживання: статистичні моделі часових рядів, спектральні методи, методи машинного навчання та нейромережеві моделі, здійснено аналітичне оцінювання їх обмежень для класифікації режимів електроспоживання. Визначено, що основними прикладними завданнями є прогнозування електроспоживання, виявлення

аномальних та пікових режимів, а також класифікація збалансованих і незбалансованих режимів електроспоживання вузлів розподіленої електроенергетичної системи. Обґрунтовано, що адекватна модель сигналу електроспоживання має одночасно враховувати стохастичну природу та повторювану (циклічну) структуру добового циклу.

У другому розділі розглянуто клас математичних моделей, що одночасно описують стохастичну природу та повторювану структуру сигналу електроспоживання: модель циклічного випадкового процесу (ЦВП), яка дозволяє отримати інформацію про моменти вищих порядків та міжциклові кореляції, та модель періодично корельованого випадкового процесу з періодичністю математичного сподівання та автоковаріаційної функції. Обґрунтовано вибір моделі ПКВП як основної для аналізу даних електроспоживання. Детально розглянуто два методи аналізу ПКВП: компонентний метод, що базується на розкладі автоковаріаційної функції у ряд Фур'є за гармоніками періоду, та синфазний метод, що виконує когерентне усереднення реалізацій циклу у часовій області. Проведено порівняльний аналіз методів та обґрунтовано вибір компонентного методу як основного для формування інформативних ознак.

У третьому розділі наведено результати експериментального дослідження реальних даних електроспоживання приватного підприємства за 365 діб (погодинна дискретизація). Підтверджено багаторівневу повторювану структуру електроспоживання з домінуванням добового періоду. На основі компонентного методу ПКВП сформовано набір із 13 інформативних ознак (9 компонентних та 4 часових), статистичний аналіз підтвердив їх дискримінантну здатність. Серед ознак визначено чотири найбільш значущі: середній рівень енергії (`time_mean`), інтенсивність флуктуацій після критичної точки (`skew_after_45`), часова локалізація точки перелому (`tangent_45`) та стандартне відхилення до точки перелому (`std_before_45`). Проведено порівняльний аналіз ефективності 14 моделей машинного навчання. Найвищий F1-score (0,9877) та Recall=1,0000

досягнуто для трьох моделей: Extra Trees, KNN ( $k=5$ ) та SVM (RBF), за критерієм крос-валідаційної стабільності перевагу має Extra Trees (CV-F1=0,9826). Результати оцінювання за комплексом метрик якості (AUC, MCC, Kappa) підтверджують надійну роздільну здатність та узгодженість класифікації для всіх 14 моделей.

У четвертому розділі розроблено програмний модуль, що реалізує повний цикл аналізу даних електроспоживання – від опрацювання вхідних даних методами ПКВП до класифікації режиму електроспоживання з візуалізацією результатів. Запропоновано підхід до адаптивного вибору оптимального протоколу передавання даних (MQTT, CoAP, HTTPS) на основі аналізу мережевих параметрів методами машинного навчання, що забезпечує своєчасність надходження телеметричної інформації у розподілених електроенергетичних системах. За результатами порівняльного аналізу чотирьох моделей машинного навчання найкращу точність класифікації протоколу забезпечила модель MLP Neural Network (F1=0,9020).

*Ключові слова:* електроспоживання, електроенергетична система, аналіз, метод, випадковий процес, статистичні оцінки, математична модель, періодично корельований випадковий процес, математичне сподівання, кореляційна функція, інформативні ознаки, машинне навчання, штучний інтелект, класифікація, протоколи передавання даних.

## ABSTRACT

*Voloshchuk A.V.* Models and Methods for Electricity Consumption Data Analysis in Distributed Electrical Power Systems. – Qualifying scientific work on manuscript rights.

Dissertation for the degree of Doctor of Philosophy in Specialty 122 – Computer Science. Ternopil Ivan Puluj National Technical University, Ternopil, 2026.

The dissertation addresses a relevant scientific and applied problem of electricity consumption data analysis based on the model of periodically correlated random processes (PCRP) combined with machine learning (ML) methods for classifying electricity consumption modes in distributed electrical power systems.

The introduction substantiates the relevance of the research topic, establishes its connection with current scientific programs, formulates the aim and objectives, defines the object and subject of the study, describes the research methods, provides a general overview of the dissertation, highlights the scientific novelty and practical significance of the obtained results, and presents information on their approbation and dissemination.

The first chapter analyzes the current state of electrical power systems amid the development of distributed generation and the integration of renewable energy sources within the Energy 4.0 concept, particularly in the context of synchronization of the Ukrainian Integrated Power System with the ENTSO-E continental network, which imposes higher requirements on the timeliness and reliability of electricity consumption data analysis. Existing approaches to electricity consumption data analysis are reviewed: statistical time series models, spectral methods, machine learning and neural network models; an analytical assessment of their limitations for electricity consumption mode classification is performed. The main applied tasks identified include electricity consumption forecasting, detection of anomalous and peak modes, and classification of balanced and unbalanced electricity consumption modes of nodes in a distributed electrical power system. It is substantiated that an

adequate model of the electricity consumption signal must simultaneously account for both the stochastic nature and the repetitive (cyclic) structure of the daily cycle.

The second chapter examines a class of mathematical models that simultaneously describe the stochastic nature and repetitive structure of the electricity consumption signal: the cyclic random process (CRP) model, which provides information about higher-order moments and inter-cycle correlations, and the periodically correlated random process model with periodicity in the mean and autocovariance function. The selection of the PCRPM model as the primary model for electricity consumption data analysis is justified. Two PCRPM analysis methods are examined in detail: the component method, based on Fourier series expansion of the autocovariance function over period harmonics, and the synphase method, which performs coherent averaging of cycle realizations in the time domain. A comparative analysis is performed, justifying the selection of the component method as the primary approach for generating informative features.

The third chapter presents experimental results on real hourly electricity consumption data from a private enterprise over 365 days. The multi-level repetitive structure of consumption with a dominant daily cycle is confirmed. Using the PCRPM component method, a set of 13 informative features (9 component and 4 temporal) is formed; statistical analysis confirms their discriminative power. Four most significant features are identified: mean energy level (`time_mean`), fluctuation intensity after the critical point (`skew_after_45`), temporal localization of the inflection point (`tangent_45`), and signal standard deviation before the inflection point (`std_before_45`). A comparative evaluation of 14 machine learning models is conducted. The highest F1-score (0.9877) and Recall=1.0000 were achieved by three models: Extra Trees, KNN (k=5), and SVM (RBF), Extra Trees demonstrates the best cross-validation stability (CV-F1=0.9826). Evaluation results across the full set of quality metrics (AUC, MCC, Kappa) confirm reliable discriminative power and classification consistency for all 14 models.

The fourth chapter describes the developed software module that implements the full analysis pipeline – from PCRP-based input data processing to electricity consumption mode classification with result visualization. An adaptive approach for selecting the optimal data transmission protocol (MQTT, CoAP, HTTPS) based on network parameter analysis using machine learning is proposed, ensuring timely delivery of telemetry data in distributed electrical power systems. A comparative analysis of four machine learning models showed that the MLP Neural Network achieved the best protocol classification accuracy (F1=0.9020).

*Keywords:* electricity consumption, electrical power system, analysis, method, random process, statistical estimates, mathematical model, periodically correlated random process, mean value, correlation function, informative features, machine learning, artificial intelligence, classification, data transmission protocols.

## СПИСОК ПУБЛІКАЦІЙ ЗДОБУВАЧА ЗА ТЕМОЮ ДИСЕРТАЦІЇ

*Наукові праці, в яких опубліковано основні наукові результати дисертації:*

1. Voloshchuk A.<sup>1</sup>, Osukhivska H.<sup>2</sup>, Khvostivskyi M.<sup>3</sup>, Sverstiuk A.<sup>4</sup>. Application of periodically correlated stochastic processes for forecasting electricity consumption. *Measuring and Computing Devices in Technological Processes*. 2025. no. 3. P. 393–403. DOI: <https://doi.org/10.31891/2219-9365-2025-83-48>. (<sup>1</sup> – обґрунтував застосування синфазного методу ПКВП для аналізу сигналів електроспоживання, провів обчислювальні експерименти та підготував рукопис; <sup>2</sup> – здійснювала наукове керівництво, формулювала завдання дослідження та брала участь у написанні статті; <sup>3</sup> – брав участь в обговоренні результатів та оформленні публікації; <sup>4</sup> – здійснював наукове керівництво та брав участь у плануванні експериментів).
2. Voloshchuk A.<sup>1</sup>, Osukhivska H.<sup>2</sup>, Khvostivskyi M.<sup>3</sup>, Sverstiuk A.<sup>4</sup>. Analysis of electricity consumption using the component method of periodically correlated random processes. *Computer Systems and Information Technologies*. 2025. no. 3. P. 74–82. DOI: <https://doi.org/10.31891/csit-2025-3-8>. (<sup>1</sup> – запропонував процедуру виділення інформативних ознак із результатів компонентного методу ПКВП, провів експериментальне дослідження та підготував рукопис; <sup>2</sup> – здійснювала наукове керівництво та формулювала постановку завдання; <sup>3</sup> – брав участь в обговоренні методології та оформленні публікації; <sup>4</sup> – здійснював консультації щодо інтерпретації результатів класифікації, брав участь в обговоренні результатів та перевірці рукопису).
3. Voloshchuk A.<sup>1</sup>, Osukhivska H.<sup>2</sup>. Adaptive multi–protocol communication for energy systems. *Scientific journal of the Ternopil national technical university*. 2025. Vol. 119, no. 3. P. 97–106. DOI: [https://doi.org/10.33108/visnyk\\_tntu2025.03.097](https://doi.org/10.33108/visnyk_tntu2025.03.097). (<sup>1</sup> – розробив метод

- адаптивного вибору протоколу передавання даних на основі ансамблю алгоритмів машинного навчання, реалізував систему динамічного перемикавання протоколів та підготував рукопис; <sup>2</sup> – здійснювала наукове керівництво, брала участь у формулюванні завдань та написанні статті).*
4. Voloshchuk A.<sup>1</sup>, Osukhivska H.<sup>2</sup>. Application of a Cyclic Stochastic Process Model for Analyzing Electricity Consumption. *Computer–Integrated Technologies: Education, Science, Production*. 2025. no. 61. P. 35–42. DOI: <https://doi.org/10.36910/6775–2524–0560–2025–61–05>. (<sup>1</sup> – провів верифікацію моделі електроспоживання у вигляді циклічного випадкового процесу на реальних даних погодинного споживання, та підготував рукопис; <sup>2</sup> – здійснювала наукове керівництво, брала участь у формулюванні завдань та написанні статті).
5. Волощук, А.В.<sup>1</sup>, Осухівська, Г.М.<sup>2</sup>, Паляниця, Ю.Б.<sup>3</sup>, Луцків, А.М.<sup>4</sup>, Валяшек, В.Б.<sup>5</sup>. Визначення збалансованості електроенергетичних систем методами машинного навчання на основі інформативних ознак отриманих компонентним методом. *Вісник Національного університету водного господарства та природокористування*. 2025. Т. 4, № 112. С. 148–161. DOI: <https://doi.org/10.31713/vt4202514>. (<sup>1</sup> – запропонував і реалізував метод виділення інформативних ознак на основі компонентного методу ПКВП, провів порівняльний аналіз алгоритмів машинного навчання для класифікації режимів електроспоживання та підготував рукопис; <sup>2</sup> – здійснювала наукове керівництво та брала участь у написанні статті; <sup>3,4,5</sup> – брали участь в обговоренні методології, результатів та оформленні публікації).

*Наукові праці, які засвідчують апробацію матеріалів дисертації:*

1. Tymoshchuk D., Voloshchuk A., Sverstiuk A., Osukhivska H., Bahrii-Zaiats O. Machine learning-based information technology for analyzing energy peaks in

- power grid balancing. *Proceedings of the 2nd International Workshop on Advanced Applied Information Technologies (AdvAIT 2025)*. Khmelnytskyi, Ukraine – Žilina, Slovakia, 2025. Vol. 4163. P. 204–220. URL: <https://ceur-ws.org/Vol-4163/paper18.pdf>.
2. Voloshchuk A., Osukhivska H., Khvostivskyi M., Sverstiuk A., Khvostivska L. Component method for analyzing the energy consumption signal as a periodically correlated random process. *Proceedings of the International Workshop on Information Technologies: Theoretical and Applied Problems (ITTAP 2025)*. Ternopil, Ukraine – Opole, Poland, 2025. Vol. 4146. P. 128–137. URL: <https://ceur-ws.org/Vol-4146/paper13.pdf>.
  3. Voloshchuk A., Velychko D., Osukhivska H., Palamar A. Computer system for energy distribution in conditions of electricity shortage using artificial intelligence. *Proceedings of the 2nd International Workshop on Computer Information Technologies in Industry 4.0 (CITI 2024)*. Ternopil, Ukraine, 2024. Vol. 3742. P. 66–75. URL: <https://ceur-ws.org/Vol-3742/paper5.pdf>.
  4. Voloshchuk A., Osukhivska H. Artificial intelligence technologies in power grids. *Modern Trends of Social Transformations of Society in Conditions of Sustainable Development : abstracts of VIII Intern. Sci. and Pract. Conf., Stockholm, Sweden, February 24–26, 2025*. Stockholm, 2025. P. 210.
  5. Волощук А., Осухівська Г. Архітектура системи енергетичного підприємства для отримання даних про споживання електроенергії. *Матеріали XI наук.-техн. конф. «Інформаційні моделі, системи та технології»*, м. Тернопіль, 13–14 груд. 2023 р. Тернопіль : ТНТУ ім. Івана Пулюя, 2023. С. 140.
  6. Осухівська Г., Волощук А. Технології передавання та опрацювання даних в комп'ютеризованих системах обліку електроенергії. *Матеріали X наук.-техн. конф. «Інформаційні моделі, системи та технології»*, м. Тернопіль, 7–8 груд. 2022 р. Тернопіль : ТНТУ ім. Івана Пулюя, 2022. С. 136.

## ЗМІСТ

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ ТА СКОРОЧЕНЬ.....	13
ВСТУП.....	16
РОЗДІЛ 1. ОГЛЯД СУЧАСНИХ МЕТОДІВ ОПРАЦЮВАННЯ ТА АНАЛІЗУ ДАНИХ ЕЛЕКТРОСПОЖИВАННЯ В РОЗПОДІЛЕНИХ ЕЛЕКТРОЕНЕРГЕТИЧНИХ СИСТЕМАХ.....	25
1.1. Аналіз сучасного стану електроенергетичних систем та проблеми аналізу даних електроспоживання.....	25
1.2. Аналіз існуючих моделей та методів опрацювання даних електроспоживання.....	31
1.3. Методи машинного навчання для аналізу даних електроспоживання.....	37
1.4. Висновки до розділу 1.....	45
РОЗДІЛ 2. МОДЕЛІ ТА МЕТОДИ АНАЛІЗУ ДАНИХ ЕЛЕКТРОСПОЖИВАННЯ.....	48
2.1. Аналіз процесу електроспоживання.....	48
2.2. Модель циклічного випадкового процесу для аналізу даних електроспоживання.....	53
2.3. Математична модель періодично корельованого випадкового процесу для аналізу даних електроспоживання.....	62
2.4. Синфазний метод аналізу сигналів електроспоживання на основі моделі ПКВП.....	68
2.5. Компонентний метод ПКВП для аналізу сигналів електроспоживання.....	82
2.6. Висновки до розділу 2.....	90
РОЗДІЛ 3. КЛАСИФІКАЦІЯ РЕЖИМІВ ЕЛЕКТРОСПОЖИВАННЯ В РОЗПОДІЛЕНИХ ЕЛЕКТРОЕНЕРГЕТИЧНИХ СИСТЕМАХ МЕТОДАМИ МАШИНОГО НАВЧАННЯ НА ОСНОВІ ІНФОРМАТИВНИХ ОЗНАК ОТРИМАНИХ КОМПОНЕНТНИМ МЕТОДОМ.....	93
3.1. Формування інформативних ознак на основі компонентного методу аналізу ПКВП.....	93
3.2. Статистичний аналіз виділених ознак та їх дискримінантної здатності.....	102
3.3. Застосування методів машинного навчання та оцінка їх ефективності для класифікації режимів електроспоживання.....	114
3.4. Висновки до розділу 3.....	132
РОЗДІЛ 4. РОЗРОБКА ТА РЕАЛІЗАЦІЯ ПРОГРАМНОГО МОДУЛЯ АНАЛІЗУ ДАНИХ ЕЛЕКТРОСПОЖИВАННЯ.....	134
4.1. Обґрунтування архітектури та вибору програмних засобів.....	134
4.2. Адаптивний вибір оптимального протоколу передавання даних на основі методів машинного навчання.....	139
4.3. Програмний модуль аналізу даних електроспоживання та класифікації режимів електроспоживання.....	150
4.4. Висновки до розділу 4.....	155
ВИСНОВКИ.....	157
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ.....	160

ДОДАТКИ.....	178
ДОДАТОК А Список публікацій здобувача за темою дисертації.....	178
ДОДАТОК Б Довідки/акти впровадженнь результатів дисертаційних досліджень .....	182
ДОДАТОК В Результати аналізу даних електроспоживання районів синфазним та компонентним методами ПКВП.....	187
ДОДАТОК Г Фрагмент програмного коду розробленого ПЗ для вибору оптимального протоколу .....	193
ДОДАТОК Д Фрагмент програмного коду розробленого ПЗ для класифікації режимів електроспоживання.....	198
ДОДАТОК Е Фрагмент програмного коду з середовища MATLAB.....	203

## ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ ТА СКОРОЧЕНЬ

АВП – автоматизований вибір протоколу

ЕЕС – електроенергетична система

МН – машинне навчання

ОЕС – об'єднана енергосистема України

ПКВП – періодично корельований випадковий процес

ЦВП – циклічний випадковий процес

РЕЕС – розподілена електроенергетична система

СЕС – сонячна електростанція

Accuracy – точність класифікації

Precision – точність (частка правильно визначених позитивних спостережень)

Recall – повнота (частка правильно визначених позитивних спостережень серед усіх реально позитивних)

AdaBoost (англ. Adaptive Boosting) – адаптивний бустинг

API (англ. Application Programming Interface) – інтерфейс програмування застосунків

ARIMA (англ. Autoregressive Integrated Moving Average) – авторегресійна інтегрована модель ковзного середнього

AUC (англ. Area Under the Curve) – площа під ROC-кривою

CNN (англ. Convolutional Neural Network) – згорткова нейронна мережа

CoAP (англ. Constrained Application Protocol) – протокол передавання даних для обмежених пристроїв

CV (англ. Coefficient of Variation) – коефіцієнт варіації

CV-F1 – крос-валідаційний F1-score

DT (англ. Decision Tree) – дерево рішень

ENTSO-E – Європейська мережа операторів системи передавання електроенергії

ET (англ. Extra Trees) – екстремальні дерева

F1-score – гармонійне середнє точності та повноти

FFT (англ. Fast Fourier Transform) – швидке перетворення Фур'є

GAN (англ. Generative Adversarial Network) – генеративно-змагальна мережа

GB (англ. Gradient Boosting) – градієнтний бустинг

GRU (англ. Gated Recurrent Unit) – рекурентна мережа з керованими блоками

HTTPS (англ. HyperText Transfer Protocol Secure) – захищений протокол передавання даних

IoT (англ. Internet of Things) – інтернет речей

KDE (англ. Kernel Density Estimation) – оцінка густини ядра

kNN (англ. k-Nearest Neighbors) – метод k найближчих сусідів

Kur – коефіцієнт ексцесу

LDA (англ. Linear Discriminant Analysis) – лінійний дискримінантний аналіз

LightGBM (англ. Light Gradient Boosting Machine) – легкий градієнтний бустинг

LR (англ. Logistic Regression) – логістична регресія

LSTM (англ. Long Short-Term Memory) – мережа з довгою короткостроковою пам'яттю

MCC (англ. Matthews Correlation Coefficient) – коефіцієнт кореляції Метьюса

MDI (англ. Mean Decrease in Impurity) – середнє зменшення домішок

ML (англ. Machine Learning) – машинне навчання

MLP (англ. Multilayer Perceptron) – багатошаровий перцептрон

MLR (англ. Multiple Linear Regression) – множинна лінійна регресія

MQTT (англ. Message Queuing Telemetry Transport) – протокол обміну повідомленнями типу publish/subscribe

NB (англ. Naive Bayes) – наївний байєсівський класифікатор

PI (англ. Permutation Importance) – важливість ознак методом перестановки

PMU (англ. Phasor Measurement Unit) – пристрій вимірювання фазорів

RBF (англ. Radial Basis Function) – радіально-базисна функція

RF (англ. Random Forest) – випадковий ліс

RMSE (англ. Root Mean Square Error) – середньоквадратична похибка

ROC (англ. Receiver Operating Characteristic) – характеристична крива приймача

SARIMA (англ. Seasonal Autoregressive Integrated Moving Average) – сезонна авторегресійна інтегрована модель ковзного середнього

SCADA (англ. Supervisory Control and Data Acquisition) – система диспетчерського керування та збору даних

SHAP (англ. SHapley Additive exPlanations) – метод пояснення моделей

$S_k$  – коефіцієнт асиметрії

SMOTE (англ. Synthetic Minority Oversampling Technique) – синтетична надвибірка меншини

SSA (англ. Singular Spectrum Analysis) – метод сингулярного спектрального аналізу

SVM (англ. Support Vector Machine) – метод опорних векторів

XGBoost (англ. Extreme Gradient Boosting) – екстремальний градієнтний бустинг

$\xi(\omega, t)$  – випадковий (стохастичний) процес

$m_\xi(t)$  – математичне сподівання

$r_\xi(t, s)$  – кореляційна функція

$T(t, n)$  – функція ритму

$\hat{B}_k(u)$  – оцінка кореляційних компонент  $k$ -ї гармоніки

$\hat{b}_\xi(t, u)$  – оцінка автоковаріаційної функції

$\xi_k(t)$  – стаціонарна компонента  $k$ -го порядку

$\xi^0(t)$  – центрований сигнал

$N$  – кількість реалізацій циклу

$u$  – часовий зсув

$k$  – номер гармоніки

$x_{45}$  – позиція точки зростання сигналу

## ВСТУП

**Актуальність теми.** Стабільна робота електроенергетичних систем є одним із ключових факторів забезпечення сталого економічного розвитку та надійного енергопостачання в умовах активної цифрової трансформації енергетики.

Децентралізація генерації та зростання кількості активних споживачів ускладнюють структуру та режими функціонування сучасних електроенергетичних систем. Особливої актуальності ці завдання набули у зв'язку із синхронізацією ОЕС України з континентальною мережею ENTSO–E, що супроводжується підвищенням вимог до оперативності опрацювання даних електроспоживання на всіх рівнях електроенергосистеми.

Електроенергетичні системи належать до об'єктів критичної інфраструктури, тому своєчасне виявлення незбалансованих режимів електроспоживання є ключовою передумовою реагування на відхилення у функціонуванні системи. Це зумовлює потребу в автоматизованих методах класифікації режимів електроспоживання вузлів розподіленої електроенергетичної системи.

Додатковим викликом є зростання обсягів даних електроспоживання, що передаються до центрального вузла системи, оскільки втрати даних у нестабільних каналах зв'язку безпосередньо впливають на повноту інформації та якість класифікації режимів.

Пошук ефективних рішень для надійного функціонування таких систем базується на фундаментальних результатах наукової спільноти, проте практична реалізація інтелектуального опрацювання та класифікації ускладнюється ризиками втрати даних у нестабільних каналах зв'язку та потребою у нових методах своєчасної класифікації режимів електроспоживання (збалансований/незбалансований). Наведене зумовлює необхідність розроблення спеціалізованого підходу, який поєднує математичні моделі аналізу даних, методи машинного навчання (МН) та штучного інтелекту (ШІ) і адаптивні

механізми комунікації для підвищення надійності функціонування електроенергетичної системи у критичних режимах.

Проблемами забезпечення надійності та ефективності функціонування електроенергетичних об'єктів займалися Кириленко О.В., Курінний Е.Г., Шидловський А.К., Стогній Б.С., Кулик В.В. та інші. Розвиток цих напрямів у контексті забезпечення надійності та підвищення гнучкості інтелектуальних мереж Smart Grid висвітлено у працях Буткевича О.Ф., Жаркіна А.Ф., Попова В.А. та Блінова І.В. Питаннями математичного моделювання та оптимізації режимів локальних електроенергетичних систем з відновлюваними джерелами енергії присвячено роботи Лежнюка П.Д., Рубаненка О.Є., Гунько І.О., Кулика В.В. та ін. Задачі забезпечення стійкості ОЕС України в умовах синхронізації з ENTSO–Е досліджуються у працях Кузнецова В.Г., Кучанського В.В. та Тугая Ю.І. Питання функціонування розподілених електричних мереж та підвищення енергетичної ефективності локальних систем електропостачання розглянуто у роботах Розена В.П., Давиденко Л.В., Волинця В.І., Добровольської Л.Н. та ін.

Розробку математичних моделей в електроенергетиці на основі теорії випадкових процесів розвинуто у працях Марченка Б.Г., Щербака Л.М., Приймака М.В., Мацюка О.В., Фриза М.Є., Бабака В.П., Яворського І.М., Литвиненка Я.В., Марценка С.В., Готовича В.А. та інших.

У міжнародному науковому контексті завдань аналізу даних та керування в розподілених електроенергетичних системах досліджуються за кількома ключовими напрямками. Статистичне прогнозування навантажень та ієрархічне моделювання часових рядів в енергетиці розвинуто у роботах T. Hong, R.J. Hyndman, S. Ben Taieb, J.W. Taylor та ін., проте запропоновані підходи переважно орієнтовані на детерміновані трендові та ієрархічні моделі без глибокого врахування стохастичної циклічної структури сигналу.

Архітектурні рішення розподілених Edge–Cloud обчислень для оперативного аналізу даних на периферії мережі запропоновано у працях S.

Dustdar, R. Buuya, W. Shi, C. Feng та ін., що формує інфраструктурну основу для реалізації модулів класифікації безпосередньо на рівні вузлів збору даних.

Проблеми інформаційно-комунікаційної взаємодії у мережах Smart Grid та забезпечення якості обслуговування розроблено у працях H.V. Poor, E. Hossain, V.C. Gungor, Z. Han та ін., які обґрунтували вимоги до протоколів передавання даних в умовах розподіленої генерації. Застосування методів машинного навчання для класифікації режимів та виявлення аномалій у Smart Grid досліджено у роботах P. Pinson, A. Khosravi, M.F. Guato Burgos, Y. Himeur та ін.

Попри значні результати у дослідженнях переважає емпіричний підхід до побудови моделей машинного навчання, що не враховує внутрішню стохастичну структуру сигналу електроспоживання. Стохастична природа електроспоживання, її одночасна повторюваність та стохастична мінливість у межах добового циклу розглядається або як джерело шуму (у класичних статистичних моделях), або як складність, що компенсується обсягом навчальних даних (у нейромережевих підходах).

Це обмежує інтерпретованість результатів класифікації режимів електроспоживання та ускладнює формування інформативного ознакового простору, стійкого до нестационарних умов функціонування мережі. Крім того, завдання класифікації стану системи та адаптивного вибору комунікаційного протоколу передавання даних у більшості робіт розглядаються ізольовано, хоча характеристики передавання та повнота вхідних даних безпосередньо впливають на точність класифікації. Зазначені проблеми особливо загострюються в умовах розвитку розподілених електроенергетичних систем з інтегрованими джерелами відновлюваної енергії та зростаючою кількістю активних споживачів, що підвищує вимоги до оперативності визначення режиму електроспоживання мережі.

Це зумовлює необхідність розроблення інтегрованого підходу, що поєднує математичні моделі для аналізу даних електроспоживання, методи машинного

навчання для класифікації режимів електроспоживання та адаптивні механізми керування передаванням даних у розподілених електроенергетичних системах, що і визначає актуальність теми дисертаційного дослідження.

**Зв'язок роботи з науковими програмами, планами, темами.**

Дисертаційна робота пов'язана з науковими дослідженнями відповідно до напряму наукової роботи міжнародної наукової школи «Моделювання та математичне забезпечення інтелектуалізованих інформаційних систем в медицині, техніці та економіці», яка функціонує на кафедрі комп'ютерних наук Тернопільського національного технічного університету імені Івана Пулюя. Також дисертаційна робота пов'язана з науково-дослідною роботою «Інтелектуальні моделі в кібер-фізичних системах медико-біологічних процесів», № держреєстрації 0125U000105 (Тернопільський національний медичний університет імені І.Я. Горбачевського МОЗ України), в рамках якої здобувачем застосовано математичний апарат теорії ПКВП для аналізу даних електроспоживання та розроблено методи формування інформативних ознак для класифікації режимів електроспоживання розподілених електроенергетичних систем.

**Мета і завдання дослідження.** Метою дисертаційної роботи є обґрунтування методів аналізу даних електроспоживання на основі моделі періодично корельованих випадкових процесів у поєднанні з методами машинного навчання для класифікації режимів електроспоживання в розподілених електроенергетичних системах.

Для досягнення поставленої мети у роботі необхідно розв'язати такі *завдання*:

1. Проаналізувати відомі моделі та методи аналізу даних електроспоживання в розподілених електроенергетичних системах.
2. Обґрунтувати математичну модель опрацювання показників електроспоживання, яка враховує релевантні властивості, притаманні досліджуванім даним.

3. Сформувати набір інформативних ознак для визначення режиму електроспоживання на основі обґрунтованої моделі та обраного методу аналізу даних.

4. На основі запропонованих інформативних ознак провести класифікацію режимів електроспоживання електроенергетичної системи за допомогою методів машинного навчання.

5. Провести порівняльний аналіз ефективності моделей машинного навчання для класифікації режимів електроспоживання на основі матриць плутанини, ROC-кривих та операційних характеристик.

6. Запропонувати та обґрунтувати підхід до адаптивного вибору оптимального протоколу передавання даних електроспоживання у розподілених електроенергетичних системах на основі аналізу мережевих параметрів із застосуванням методів машинного навчання та оцінити ефективність запропонованого підходу.

7. Розробити програмні модулі для класифікації режимів електроспоживання та адаптивного вибору оптимального протоколу передавання даних, з метою визначення поточного режиму електроенергетичної системи.

**Об'єктом дослідження.** Процеси аналізу даних електроспоживання в розподілених електроенергетичних системах.

**Предметом дослідження.** Моделі та методи аналізу даних електроспоживання для класифікації режиму електроспоживання розподіленої електроенергетичної системи.

**Методи дослідження.** При розв'язанні завдань дослідження використано методи математичного моделювання процесів електроспоживання на основі компонентного та синфазного методів, математичні моделі у вигляді періодично корельованих випадкових процесів та методи їх статистичного аналізу, методи системного аналізу, методи машинного навчання для класифікації режиму електроспоживання в розподілених електроенергетичних системах.

**Наукова новизна одержаних результатів:**

1. Вперше на основі компонентного методу аналізу періодично корельованого випадкового процесу запропоновано набір інформативних ознак для класифікації режимів електроспоживання вузлів розподіленої електроенергетичної системи, що на відміну від класичних статистичних підходів одночасно враховує стохастичну природу та повторювану структуру сигналу електроспоживання.

2. Вперше запропоновано метод класифікації режимів електроспоживання розподілених електроенергетичних систем на основі інформативних ознак, отриманих компонентним методом ПКВП, із застосуванням методів машинного навчання, що забезпечило можливість ідентифікації режимів електроспоживання вузлів розподіленої електроенергетичної системи.

3. Удосконалено підхід до вибору оптимального протоколу передавання даних електроспоживання в розподілених електроенергетичних системах з використанням методів машинного навчання на основі параметрів каналу передавання даних, що дало змогу забезпечити стабільність передавання даних.

4. Отримав подальший розвиток математичний апарат періодично корельованих випадкових процесів в частині його поєднання з методами машинного навчання для аналізу даних електроспоживання в розподілених електроенергетичних системах, що дало змогу забезпечити класифікацію режимів електроспоживання.

**Практичне значення отриманих результатів.** Розроблено програмні модулі для класифікації режимів електроспоживання та адаптивного вибору оптимального протоколу передавання даних з метою забезпечення визначення поточного режиму роботи електроенергетичної системи та автоматичного перемикання протоколів при зміні умов функціонування каналів зв'язку.

Використані методи на основі теорії періодично корельованих випадкових процесів та алгоритмів машинного навчання можуть бути впроваджені в енергетичних системах різного масштабу – від приватних домогосподарств до комунальних мереж та інтелектуальних енергетичних мереж. Практичне застосування результатів сприятиме підвищенню стійкості функціонування за умов нестійкого передавання даних, оптимізації використання комунікаційних ресурсів та підвищенню надійності функціонування системи. Отримані результати можуть бути інтегровані в існуючі програмно–апаратні комплекси та хмарні сервіси без зміни їх існуючої архітектури, що забезпечує масштабованість та гнучкість впровадження.

Результати дослідження можуть бути використані у навчальному процесі закладів вищої освіти за спеціальністю 122 «Комп'ютерні науки» при викладанні дисциплін, пов'язаних з аналізом даних, машинним навчанням та розподіленими системами, зокрема: «Методи та системи штучного інтелекту», «Інтелектуальний аналіз даних», «Моделювання систем», «Інформаційні технології моніторингу та аналізу даних».

Акти впровадження та довідка про використання результатів у науково-дослідній роботі наведені в Додатку Б.

Перспективним напрямом є також застосування запропонованого підходу у практичних завданнях аналізу інших енергетичних процесів циклічної природи, зокрема для моніторингу газоспоживання, теплоспоживання або водоспоживання промислових та комунальних об'єктів, що характеризуються добовою та сезонною повторюваністю статистичних характеристик і можуть бути описані моделлю ПКВП.

**Особистий внесок здобувача.** Наукові результати, викладені в дисертаційній роботі, отримані здобувачем особисто та є оригінальними. Здобувачем обґрунтовано доцільність застосування математичного апарату ПКВП для аналізу даних електроспоживання та обрано компонентний метод як основний для формування інформативних ознак класифікації. Самостійно

сформовано набір інформативних ознак на основі компонентного методу ПКВП, проведено їх статистичний аналіз та оцінено дискримінантну здатність для завдань класифікації режимів електроспоживання. Здобувачем розроблено метод адаптивного вибору оптимального протоколу передавання даних електроспоживання на основі методів машинного навчання та обґрунтовано параметри його реалізації. Самостійно реалізовано та експериментально перевірено програмний модуль аналізу даних електроспоживання, що інтегрує математичний апарат ПКВП та методи машинного навчання в єдиному циклі опрацювання даних.

У працях, опублікованих із співавторами, внесок здобувача зазначено у списку публікацій. Науковим керівникам – кандидату технічних наук, доценту Осухівській Г.М., та доктору технічних наук, професору Сверстюку А.С. – належать постановка дослідження та наукове консультування здобувача в процесі виконання роботи.

**Апробація результатів дисертації.** Наукові та практичні результати дисертаційного дослідження доповідалися та обговорювалися на міжнародних та всеукраїнських конференціях, зокрема на:

1. X науково-технічній конференції «Інформаційні моделі, системи та технології», 7–8 грудня 2022 р., Тернопільський національний технічний університет імені Івана Пулюя, м. Тернопіль, Україна.
2. XI науково-технічній конференції «Інформаційні моделі, системи та технології», 13–14 грудня 2023 р., Тернопільський національний технічний університет імені Івана Пулюя, м. Тернопіль, Україна.
3. 2nd International Workshop on Computer Information Technologies in Industry 4.0 (CITI-2024), 12–14 червня 2024 р., м. Тернопіль, Україна.
4. VIII International Scientific and Practical Conference «Modern Trends of Social Transformations of Society in Conditions of Sustainable Development», 24–26 лютого 2025 р., м. Стокгольм, Швеція.

5. ITTAP'2025: The 5th International Conference on Information Technologies (ITTAP-2025), 22–24 жовтня 2025 р., м. Тернопіль, Україна; м. Опольє, Польща.

6. 2nd International Workshop on Advanced Applied Information Technologies (AdvAIT-2025), 5 грудня 2025 р., м. Хмельницький, Україна.

В цілому, робота доповідалася на науковому семінарі в Тернопільському національному технічному університеті імені Івана Пулюя (м. Тернопіль).

**Публікації.** Основні результати дисертації висвітлено у 11 наукових працях, серед яких 5 статей у наукових фахових виданнях України та 6 публікацій у матеріалах міжнародних і всеукраїнських наукових конференцій, з яких 3 індексуються наукометричною базою Scopus.

**Структура та обсяг дисертації.** Дисертація складається з вступу, чотирьох розділів, висновків, списку використаних джерел та додатків. Загальний обсяг дисертації становить 213 сторінок, у тому числі: 145 сторінок основного тексту, 81 рисунок, 31 формула, 7 таблиць, 140 джерел, 6 додатків.

## **РОЗДІЛ 1. ОГЛЯД СУЧАСНИХ МЕТОДІВ ОПРАЦЮВАННЯ ТА АНАЛІЗУ ДАНИХ ЕЛЕКТРОСПОЖИВАННЯ В РОЗПОДІЛЕНИХ ЕЛЕКТРОЕНЕРГЕТИЧНИХ СИСТЕМАХ**

### **1.1. Аналіз сучасного стану електроенергетичних систем та проблеми аналізу даних електроспоживання**

Сучасна електроенергетика перебуває на етапі комплексної трансформації в межах парадигми Energy 4.0, у якій енергетичні системи розглядаються як складні кіберфізичні об'єкти з інтегрованими інформаційно–комунікаційними технологіями та інтелектуальними системами керування [1, 2]. Ця трансформація зумовлена декількома глобальними тенденціями, зокрема впровадженням відновлюваних джерел енергії (ВДЕ) у загальну електроенергетичну систему, розширенням мережі розподіленої генерації, збільшенням кількості активних споживачів, здатних не лише споживати, а й генерувати електроенергію [3, 4]. Концепція Energy 4.0 передбачає перехід від централізованої моделі керування до розподіленої інтелектуальної системи, де рішення приймаються на різних рівнях ієрархії на основі даних, отриманих у режимі реального часу [4, 5].

Прогнози розвитку IoT-інфраструктури свідчать про стійку тенденцію до зростання кількості підключених пристроїв. Зокрема, за прогнозами IoT Analytics, до 2030 року кількість IoT-пристроїв становитиме близько 39 млрд, а до 2035 року – перевищить 50 мільярдів [6]. Значна частина цих пристроїв інтегрується в електроенергетичні системи. Це підвищує вимоги до роботи електроенергосистеми, стан якої можна оцінити за результатами аналізу даних електроспоживання.

При цьому такий аналіз є важливим для визначення режиму електроспоживання розподілених електроенергетичних систем, виявлення аномалій та передаварійних станів, оптимізації режимів генерації та споживання,

керування попитом (demand response), підвищення енергоефективності, планування розвитку мережевої інфраструктури, а також забезпечення надійності та стійкості функціонування енергосистем загалом [7, 8].

Під розподіленою електроенергетичною системою (РЕЕС) розуміють тип електроенергетичної системи, в якій джерела генерації електричної енергії розміщені безпосередньо поблизу споживачів або інтегровані в розподільчі мережі, а не сконцентровані лише на великих централізованих електростанціях. На відміну від традиційної централізованої моделі, де енергія передається на великі відстані від генерації до споживача, у РЕЕС генерація є децентралізованою, що сприяє підвищенню гнучкості, ефективності і надійності системи.

У межах даного дослідження під стабільною роботою електроенергетичної системи розуміється її функціонування у збалансованому режимі, за якого статистичні характеристики даних електроспоживання зберігають характерну циклічну структуру та не містять ознак суттєвих відхилень режиму електроспоживання. Такий стан системи характеризується стабільністю статистичних характеристик та створює передумови для своєчасного виявлення переходу до незбалансованих режимів.

Водночас децентралізація генерації породжує низку нових технічних викликів: забезпечення надійності функціонування, оптимізація режимів роботи мереж та ефективна інтеграція відновлюваних джерел енергії. Зазначені проблеми досліджуються у працях вітчизняних вчених, зокрема у галузі інтелектуальних електроенергетичних систем [9], систем накопичення енергії та відновлюваної генерації [10], оптимізації режимів розподілених мереж [11] та енергетичної ефективності локальних систем електропостачання [12].

Зростання кількості відновлюваних джерел створює додаткове навантаження на лінії електропередавання та вимагає перегляду методів оцінки їх пропускної здатності [13, 14], а також нових підходів до управління режимами

генерації та компенсації небалансів [15, 16]. Функціонування РЕЕС базується на інтеграції енергетичної та інформаційної інфраструктури, що забезпечує:

- збір і передавання даних у реальному часі;
- адаптивне керування режимами роботи;
- балансування попиту та пропозиції електроенергії.

Важливою особливістю таких систем є можливість роботи в різних режимах у складі об'єднаної енергосистеми чи в автономному режимі (мікромережі, island mode).

РЕЕС забезпечує інтеграцію відновлюваних джерел енергії та інтелектуальне керування навантаженням в умовах концепції Energy 4.0, що дозволяє зменшувати втрати при передаванні, підвищувати енергетичну незалежність та реалізовувати адаптивне керування навантаженням з урахуванням поточного стану мережі.

До складу системи входить сукупність взаємодіючих елементів, серед яких:

- відновлювані джерела енергії (сонячні, вітрові установки);
- локальні джерела генерації (мікротурбіни, дизель–генератори);
- системи накопичення енергії (акумулятори, ESS);
- споживачі різних типів (побутові, комерційні, промислові);
- зарядна інфраструктура електромобілів;
- інтелектуальні системи керування та моніторингу.

З розвитком цифрових технологій, IoT (англ. Internet of Things) РЕЕС трансформується у кіберфізичну систему, де фізичні процеси генерації та споживання електроенергії тісно пов'язані з алгоритмами аналізу даних і прийняття рішень. На рисунку 1.1 представлено загальну архітектуру розподіленої електроенергетичної системи (РЕЕС).

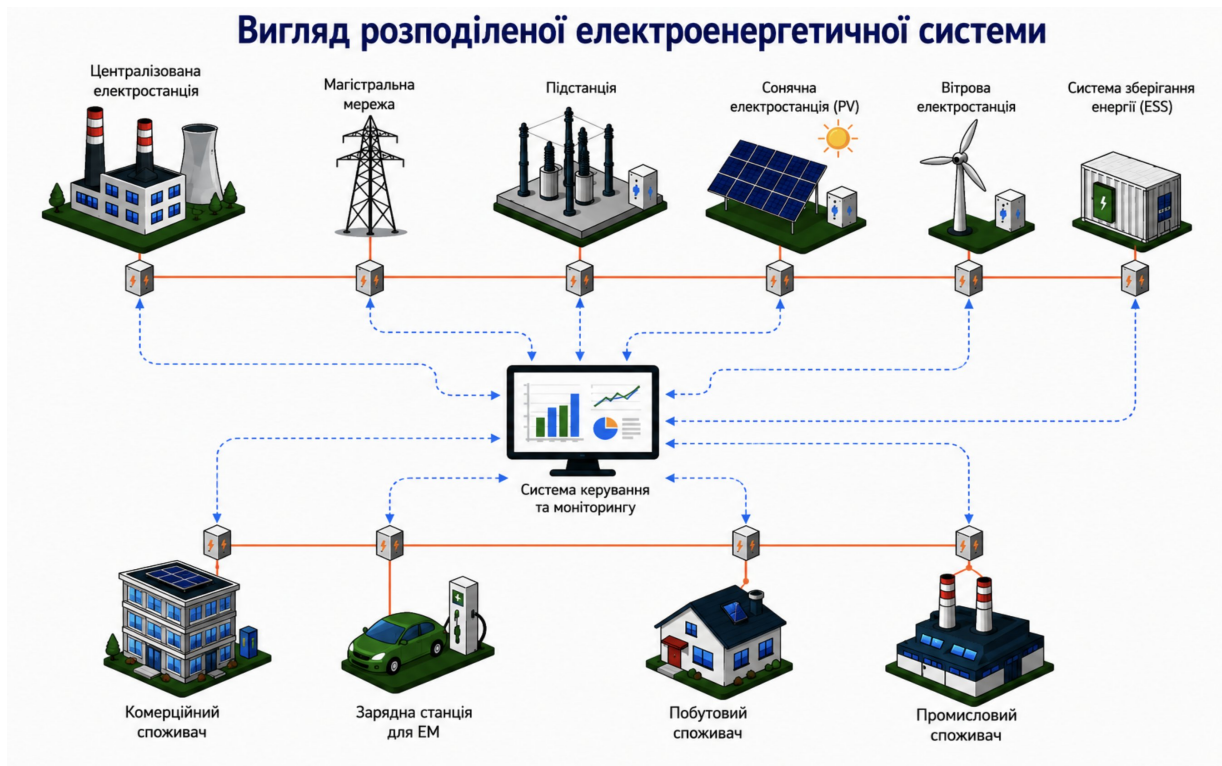


Рисунок 1.1 – Загальна архітектура розподіленої електроенергетичної системи

Відповідно до Національного плану дій з відновлюваної енергетики на період до 2030 року передбачено введення потужностей маневруючої генерації, що є необхідним для забезпечення стабільності роботи ОЕС України під час впровадження об'єктів відновлюваної енергетики [17]. Тому важливим завданням є аналіз даних електроспоживання вузлів розподіленої електроенергетичної системи, що використовуються як інформаційна основа для класифікації режимів роботи та прийняття рішень щодо керування системою.

Збалансованість електроенергетичної системи традиційно визначається як стан мережі, при якому забезпечується рівновага між генерованою та споживаною потужністю, стабільність частоти ( $50 \text{ Гц} \pm 0.2 \text{ Гц}$  для ОЕС України), підтримка номінальних рівнів напруги, а також симетричний розподіл навантаження між фазами у трифазних мережах. Оскільки прямиий контроль параметрів генерації є недоступним для вузлів розподіленої системи без власної генерації, фізичні відхилення від балансу можуть проявлятися через зміни

статистичної структури сигналу електроспоживання. Тому завданням оперативного керування є визначення поточного режиму електроспоживання вузла, тобто чи перебуває система у нормальному режимі функціонування, чи наближається до критичного стану.

Під режимом електроспоживання в контексті даного дослідження розуміється стан функціонування вузла розподіленої електроенергетичної системи, що визначається статистичними властивостями добового профілю сигналу – математичним сподіванням, дисперсією та кореляційною структурою. Розрізняють збалансований режим, якому відповідає стабільна циклічна структура добового профілю з низькою міжцикловою варіативністю, та незбалансований режим, що характеризується систематичними аномальними відхиленнями від нормальної циклічної структури та порушенням стабільності кореляційних характеристик. У роботі збалансований та незбалансований режими розглядаються як статистичні режими електроспоживання вузлів РЕЕС, що визначаються на основі аналізу стохастичної та повторюваної структури сигналу.

Критичними для керування енергосистемою також є пікові режими електроспоживання. У ці періоди зростає перевантаження електромереж та їх вузлів, збільшується дефіцит потужності, підвищується ризик порушення стабільності системи. Аналіз сучасних досліджень свідчить, що опис аномальних режимів електроспоживання потребує інших підходів, ніж аналіз типових режимів [18, 19, 20, 21, 22].

Використання даних електроспоживання дозволяє реалізувати моніторинг режиму роботи вузла без необхідності встановлення додаткових вимірювальних засобів, що суттєво спрощує практичне впровадження у розподілених електроенергетичних системах з великою кількістю територіально розосереджених вузлів споживання.

Класифікація режимів електроспоживання на збалансовані та незбалансовані має безпосереднє практичне значення для різних категорій

учасників енергоринку. Для оператора системи результати класифікації є підставою для превентивного введення резервних потужностей або перерозподілу навантаження до настання аварійного режиму. Для промислового споживача інформація про наближення системи до незбалансованого стану є підставою для підготовки до можливих обмежень. Для домогосподарства з власною СЕС така інформація може слугувати підставою для переходу в автономний режим роботи. Своєчасне виявлення змін у профілі електроспоживання дає змогу ідентифікувати відхилення параметрів електромережі (частоти, напруги) від номінальних значень та ризик переходу системи в аварійний стан.

Окремою категорією споживачів, для яких питання аналізу електроспоживання набуває дедалі більшої актуальності, є невеликі підприємства, фермерські господарства, приватні підприємства та об'єкти комунальної інфраструктури. Ці споживачі традиційно перебували поза фокусом уваги розробників систем енергоменеджменту, оскільки їхнє індивідуальне споживання є відносно невеликим. Однак у сукупності формують суттєву частку електроспоживання енергосистеми, а в умовах дефіциту генерації саме ця категорія найчастіше потрапляє під графіки обмежень. Для малого фермерського господарства, що використовує електричні насоси для зрошення, холодильне обладнання для зберігання продукції або системи мікроклімату для тваринницьких комплексів, непередбачуване відключення електроенергії може призвести до прямих економічних втрат, зокрема загибелі врожаю, псування продукції, падежу тварин тощо. Аналогічно для невеликого виробничого підприємства або майстерні – незаплановане знеструмлення означає зупинку технологічного процесу, брак продукції та зрив термінів виконання замовлень.

Зростання популярності приватних сонячних електростанцій та домашніх систем накопичення енергії створює додаткову потребу в інструментах аналізу та оптимізації електроспоживання на рівні окремого домогосподарства. Власник приватної СЕС зацікавлений у максимізації власного споживання згенерованої

електроенергії, оптимальному використанні акумуляторних батарей та мінімізації закупівель з загальної мережі. Для цього необхідні інструменти, що дають змогу аналізувати профіль споживання, прогнозувати генерацію та приймати рішення щодо режимів роботи обладнання. Існуючі на ринку системи для «розумного дому» зазвичай обмежуються простим моніторингом без глибокої аналітики, а професійні системи енергоменеджменту є надто дорогими та складними для індивідуального користувача.

Зважаючи на все вище сказане, існує об'єктивна потреба у розробленні спеціалізованих програмних рішень, які поєднують необхідний функціонал для аналізу, забезпечуючи гнучкість налаштування та незалежність. Такі системи мають бути масштабованими від рівня окремого домогосподарства до промислового підприємства, базуватися на відкритих стандартах, використовувати сучасні підходи до опрацювання даних та бути адаптованими до специфіки українського енергоринку.

## **1.2. Аналіз існуючих моделей та методів опрацювання даних електроспоживання**

Сучасні дослідження в галузі аналізу електроспоживання трансформуються від описового аналізу амплітудних характеристик до методів інтелектуального аналізу [21, 22]. Такий перехід обумовлений необхідністю врахування складної циклічної динаміки електроспоживання в умовах розвитку інтелектуальних енергосистем [23].

Розвиток математичного апарату для опрацювання даних електроспоживання характеризується переходом від класичних регресійних підходів до складних систем інтелектуального аналізу [24, 25, 26, 27, 28, 29, 30]. Як правило, регресійні моделі орієнтовані на розв'язання завдань оцінювання попиту на електроенергію в розподілених електроенергетичних системах, промислових установках та об'єктах різного призначення. Регресійні моделі

встановлюють залежність обсягу електроспоживання від зовнішніх пояснювальних змінних – температури повітря, дня тижня, виробничої активності, тарифних показників. Множинна регресія розширює лінійний підхід за рахунок одночасного врахування кількох факторів, що підвищує точність апроксимації середньострокових тенденцій споживання. Основним обмеженням є припущення лінійності зв'язків між змінними, яке в реальних умовах функціонування розподілених електроенергетичних систем часто не виконується через нелінійний характер реакції споживачів на тарифну політику та кліматичні чинники [27, 29].

Авторегресійні моделі широко застосовуються у завданнях короткострокового прогнозування електроспоживання в електроенергетичних мережах, забезпечуючи відтворення трендових та сезонних складових часового ряду електроспоживання на основі аналізу власної передісторії процесу, однак через припущення стаціонарності не здатні адекватно описувати процеси з періодично змінними статистичними характеристиками. Для подолання цього обмеження в завданнях енергетичної інформатики досліджуються властивості умовно–лінійних циклостаціонарних випадкових процесів, які враховують циклічну нестаціонарність даних електроспоживання [28].

Методи експоненційного згладжування застосовуються для середнього та довгострокового прогнозування обсягів електроспоживання в енергосистемах і на промислових об'єктах завдяки простоті реалізації та низькій обчислювальній складності [29, 30].

Експоненційне згладжування базується на присвоєнні попереднім спостереженням експоненційно спадаючих вагових коефіцієнтів, що забезпечує більший вплив актуальних даних на прогноз. Розширення Холта–Вінтерса враховує три компоненти часового ряду одночасно: рівень, тренд і сезонність, що робить його придатним для моделювання добових та тижневих циклів електроспоживання. Попри обчислювальну простоту та широке практичне застосування, метод орієнтований на відтворення середнього профілю

електроспоживання і не забезпечує адекватного опису різких пікових відхилень, характерних для реальних електроенергосистем [29, 30].

У класичних підходах для моделювання виступали статистичні моделі авторегресії та інтегрованого ковзного середнього, зокрема класи ARIMA та SARIMA [24, 25, 26]. Водночас застосування моделей класу ARIMA та SARIMA ґрунтується на припущенні слабкої стаціонарності процесу, що передбачає сталість математичного сподівання та дисперсії у часі. Для даних електроспоживання така умова, як правило, не виконується, оскільки дисперсія та кореляційні характеристики змінюються протягом добового циклу, що призводить до зниження точності опису пікових навантажень. У результаті такі моделі фактично інтерпретують пікове електроспоживання як випадкові відхилення, а не як закономірні прояви повторюваної структури процесу, що може призвести до некоректного класифікування режиму електроспоживання розподіленої електроенергосистеми. Наявність складних нелінійних зв'язків між погодними факторами і виробітком енергії з відновлюваних джерел знижує придатність класичних статистичних моделей, таких як множинна лінійна регресія (MLR), експоненційне згладжування (Holt–Winters) та ін., для точного опису цих процесів [24, 30].

Вагомий внесок у розроблення теоретичних основ та практичних методів математичного моделювання і статистичного опрацювання даних вимірювань в електроенергетиці внесли вчені Марченко Б.Г., Приймак М.В., Марценко С.В., Щербак Л.М., Готович В.А., Фриз М.Є. та ін. [31, 32, 33, 34, 35, 36, 37]. Переважна більшість зазначених досліджень та розроблених математичних моделей стосується штатного режиму процесу електроспоживання і не охоплює завдання класифікації збалансованого та незбалансованого режиму електроспоживання, що зумовлює актуальність подальших досліджень у цьому напрямі.

Методи спектрального аналізу на основі перетворення Фур'є забезпечують отримання усередненого спектра сигналу, однак не враховують зміну

спектральних характеристик у часі. У випадку електроспоживання це ускладнює виявлення нестабільних режимів, при яких амплітуди гармонічних компонент змінюються від циклу до циклу.

У роботі [35] авторами для аналізу процесу електроспоживання організації на річному проміжку спостережень застосовано метод «Гусениці–SSA», що дозволяє розкласти часовий ряд на трендову, коливальну та стохастичну компоненти, а також виявити моменти зміни статистичних характеристик процесу. Водночас зазначений підхід орієнтований на ретроспективний аналіз динаміки споживання окремої організації і не вирішує завдання класифікації поточного стану розподіленої електроенергосистеми як збалансованого або незбалансованого за ознаками пікового електроспоживання в режимі реального часу.

Серед методів, що застосовуються для математичного моделювання процесів електроспоживання варто виокремити умовні лінійні випадкові процеси та їх частинний випадок авторегресію з випадковими коефіцієнтами, застосування яких для оперативного прогнозування електроспоживання підприємств обґрунтовано в роботах [35, 36, 37]. Водночас для аналізу амплітудних характеристик якості електроенергії в електромережах запропоновано використовувати перетворення Гільберта, що дозволяє оцінювати змінювані в часі показники провалів напруги та шумових завад [38]. Хоча ці моделі дозволяють оцінити якість самого сигналу, вони не дозволяють оцінити збалансованість електроенергетичної системи. Обмеженість традиційних моделей проявляється у зниженні точності оцінювання режимів роботи системи, підвищенні ризику дисбалансів та неефективному використанні ресурсів. Також неточність прогнозування унеможливорює точне планування закупівель електроенергії на «ринку на добу наперед» (РДН) та внутрішньодобовому ринку (ВДР), що призводить до виникнення дисбалансів та економічних втрат [35, 36, 39].

Застосування апарату ПКВП до завдань аналізу електроспоживання обґрунтовано в роботах, присвячених стохастичній повторюваності фізичних і технічних процесів [40, 41, 42, 43, 44]. У межах цього підходу процес електроспоживання описується як стохастичний сигнал, статистичні характеристики якого (математичне сподівання та кореляційна функція) змінюються циклічно в часі з певним відомим або оцінюваним базовим періодом.

Продемонстровано, що реальні дані навантажень різних типів споживачів (промислових підприємств, транспортних вузлів, медичних установ) демонструють виражену нестационарну структуру: оцінка математичного сподівання є змінною в межах добового циклу функцією, дисперсія зростає у нічній зоні відносно денного рівня, а кореляційні компоненти мають коливний характер із загасанням зі збільшенням часового зсуву [40, 41, 42, 43]. Таким чином, стохастична природа електроспоживання не є «шумом навколо тренду» у традиційному сенсі, а є структурованою нестационарністю, прив'язаною до ритму доби. Характеристики кореляційних компонент є специфічними для кожного типу споживача і не можуть бути коректно усереднені без урахування цієї специфіки, що принципово відрізняє ПКВП-модель від стаціонарних та трендово-сезонних моделей. Математичний апарат, методи статистичного аналізу та практичні застосування ПКВП-моделі систематизовано у фундаментальній монографії [42].

Процеси електроспоживання характеризуються складною структурою, що поєднує регулярну повторювану складову та випадкову компоненту. Добові цикли мають виражену регулярну будову: ранковий пік електроспоживання, денний період стабільного споживання, вечірній максимум та нічний мінімум. Аналогічно спостерігаються тижневі цикли з відмінностями між робочими та вихідними днями, а також сезонні коливання, пов'язані з використанням додаткового обладнання для опалення взимку та кондиціонування влітку. Водночас амплітуда та форма добових графіків можуть суттєво варіювати під

впливом зовнішніх факторів: погодних умов, особливостей споживання під час масових заходів та свят, змін виробничого електроспоживання промислових споживачів, а також технічного стану енергосистеми [40, 41, 43].

На основі теорії періодичних лінійних випадкових процесів розроблено методику побудови прогнозних графіків енергонавантажень, що враховує циклічний характер споживання електроенергії [43]. Подальший розвиток цього напрямку пов'язаний з узагальненням ПКВП–моделі на випадок модульованих через період ймовірнісних характеристик, що дозволяє враховувати сезонну зміну параметрів добового циклу і забезпечує вищу точність моделювання процесів електроспоживання у різні пори року [44]. Це зумовлює специфічну природу електроспоживання, для якої характерне поєднання циклічності та стохастичності. Регулярна складова проявляється у повторюваності структури добових і сезонних графіків електроспоживання, тоді як випадкова компонента формується під впливом змінних зовнішніх факторів та індивідуальних особливостей режимів споживання [42, 43, 44, 45, 46, 47, 48].

Зазначені обмеження зумовлюють доцільність використання теорії ПКВП, яка дає змогу враховувати як періодичну структуру сигналу, так і мінливість його статистичних характеристик. Застосування математичного апарату енергетичної теорії стохастичних сигналів [49, 50], зокрема моделі ПКВП [43, 44, 45, 51], дозволяє реалізувати багатовимірне компонентне представлення процесу електроспоживання та забезпечити класифікацію режимів електроспоживання вузла розподіленої електроенергетичної системи.

Разом з тим, застосування виключно статистичних моделей для вирішення завдань класифікації є недостатнім з кількох причин. По-перше, такі моделі потребують явного задання математичної форми залежностей і обмежують адаптивність до змінних умов функціонування системи [24, 25, 30]. По-друге, визначення збалансованого та незбалансованого режиму є за природою завданням класифікації, для якої методи машинного навчання демонструють суттєво вищу ефективність порівняно з параметричними підходами [1, 8]. По-

третє, зростання обсягів телеметричних даних від IoT-пристроїв та інтелектуальних лічильників створює умови, за яких методи машинного навчання ефективно опрацьовують великі масиви різномірних даних [4, 6, 14].

Зазначене свідчить про те, що методи машинного навчання доцільно розглядати не як альтернативу статистичним моделям, а як їх доповнення, що забезпечує перехід від математичного опису структури процесу до оперативного прийняття рішень щодо стану електроенергетичної системи.

### **1.3. Методи машинного навчання для аналізу даних електроспоживання**

Паралельно з розвитком обчислювальних потужностей широкого розповсюдження набули методи штучного інтелекту, зокрема машинного навчання, які забезпечують виявлення прихованих закономірностей у великих масивах даних без явного задання математичної форми залежностей.

Залежно від обчислювальної складності, інтерпретованості та здатності до моделювання нелінійних залежностей методи машинного навчання традиційно поділяють на лінійні, ансамблеві, метричні та нейромережеві [52]. До лінійних належать логістична регресія, лінійний дискримінантний аналіз та наївний байєсівський класифікатор.

До ансамблевих методів належать алгоритми на основі дерев рішень, такі як: Decision Tree, Random Forest, Extra Trees, AdaBoost та Gradient Boosting. Random Forest і Extra Trees будують ансамблі паралельно на випадкових підвибірках даних та ознак, забезпечуючи стійкість до перенавчання та оцінку важливості ознак через критерій середнього зменшення домішок (MDI) [53, 54, 55]. AdaBoost і Gradient Boosting натомість використовують послідовну стратегію, де кожна наступна модель коригує помилки попередньої, що забезпечує вищу точність на складних нелінійних завданнях [56, 57].

Методи з високою обчислювальною складністю KNN, SVM з ядром RBF та багатосаровий перцептрон MLP забезпечують найвищу гнучкість при моделюванні нелінійних меж між класами [58]. KNN не має явної фази навчання, але потребує зберігання всієї вибірки в пам'яті, що обмежує його застосування в системах реального часу. SVM знаходить оптимальну гіперплощину максимального зазору у просторі перетворених ознак і демонструє стійку точність при обмежених даних.

Методи машинного навчання знайшли широке застосування в різних завданнях аналізу електроенергетичних систем, забезпечуючи високу точність при роботі з багатовимірними даними, зокрема для прогнозування електроспоживання та класифікації режимів роботи [53, 54, 59]. Для прогнозування електричного електроспоживання застосовуються ансамблеві алгоритми та рекурентні нейронні мережі, що забезпечують вищу точність порівняно з класичними статистичними підходами [53, 54, 57, 59]. MLP як апроксиматор здатний відтворювати довільні нелінійні залежності, однак потребує ретельного підбору архітектури та чутливий до ініціалізації ваг [60].

У роботі [61] розглянуто підхід до ймовірнісного прогнозування електроспоживання, що базується на побудові прогнозних інтервалів. Автори запропонували метод прогнозування квантилів на основі алгоритму Quantile Regression Forest, удосконаленого шляхом застосування процедури відбору ознак (Recursive Feature Elimination). На відміну від традиційного підходу до формування інтервалів прогнозу, який базується на прямому виборі квантилів (Direct Prediction Interval Construction), у роботі запропоновано метод визначення квантилів (Quantile Determination), що забезпечує підвищення надійності прогнозних інтервалів за рахунок узгодження фактичної ймовірності покриття з заданим рівнем довіри.

Розвиваючи ідею аналізу часових рядів, дослідники у [62] представили гібридну архітектуру, що поєднує згорткові нейронні мережі (CNN) та LSTM для прогнозування електроспоживання в житлових будинках. Такий підхід

дозволяє ефективно вилучати як локальні шаблони, так і часові залежності в даних споживання. Для оцінювання стійкості електроенергетичних систем у перехідних режимах роботи [63] запропоновано гібридний підхід на основі XGBoost та методу моделювання взаємодій між ознаками.

У даному підході XGBoost використовується для автоматичного формування просторово-часових ознак шляхом перетворення вимірювальних даних у розріджене ознакове представлення, тоді як класифікація стану системи здійснюється методом моделювання попарних взаємодій між ознаками через їх факторне розкладення. Запропонований метод забезпечує високу точність та швидкодію, що дозволяє використовувати його для оцінювання стабільності в режимі реального часу навіть за умов високої розмірності даних та наявності шуму.

Застосування графових нейронних мереж для моделювання топологічних властивостей енергомереж [64] дозволяє враховувати просторові взаємозв'язки між різними вузлами системи та виявляти приховані патерни в розподілі навантаження, що особливо важливо для аналізу каскадних відмов та планування розвитку мережі. Авторами [65] запропоновано підхід до виявлення аномалій у розподілених енергосистемах на основі автоенкодерів та федеративного навчання, що забезпечує децентралізоване навчання моделі без передавання приватних даних споживачів до центрального сервера

Сучасні дослідження демонструють ефективність ансамблевих методів машинного навчання, зокрема XGBoost та Random Forest, для завдань аналізу складних даних та класифікації в умовах дисбалансу класів. Застосування методів балансування даних, таких як SMOTE, ADASYN та інші, дозволяє суттєво підвищити точність моделей та покращити їх здатність до виявлення рідкісних подій [66]. У роботі [67] розглянуто підхід до класифікації режимів електроспоживання на основі порівняльного аналізу ансамблевих алгоритмів машинного навчання. Запропонований метод включає оптимізацію

гіперпараметрів моделей машинного навчання та забезпечує ефективну обробку незбалансованих вибірок даних електроспоживання.

У дослідженні [68] представлено систематизований огляд застосування методів штучного інтелекту для оцінювання динамічної стійкості електроенергетичних систем. Зокрема, розглядаються підходи, що базуються на машинному та глибокому навчанні, у межах яких завдання оцінювання стійкості формалізується як завдання бінарної класифікації (стійкий/нестійкий режим). При цьому використовуються як дані вимірювань (WAMS/PMU), так і результати імітаційного моделювання, що дозволяє враховувати різні типи збурень та режимів роботи енергосистеми.

Для подальшого підвищення точності виявлення аномалій у статті [69] автори розробили модель Transformer-GAN. Архітектура поєднує модуль Transformer, який використовує механізм самоуваги для захоплення довгострокових залежностей, та генеративно-змагальну мережу (GAN), де генератор вивчає нормальні патерни даних. Систематичний огляд застосування глибокого навчання для intelligent demand response [70] демонструє ефективність DL-методів для прогнозування електроспоживання та керування попитом в реальному часі, що є ключовим для балансування інтелектуальних енергетичних мереж (Smart Grid) та інтеграції відновлюваних джерел енергії. Застосування методів МН для оперативного керування електроспоживанням [71] дозволяє підвищити гнучкість та ефективність функціонування розподілених електроенергетичних систем в умовах нестабільного електроспоживання.

Для ідентифікації аномальних режимів та передаварійних станів доцільно використовувати комбінований підхід: дерева рішень дозволяють оперативно класифікувати відомі типи критичних ситуацій [72], тоді як автокодувальники забезпечують виявлення прихованих відхилень у споживанні електроенергії через аналіз похибок реконструкції даних [73].

Попри високу апроксимаційну здатність щодо нелінійних залежностей, методи на основі Random Forest та локальних регресійних моделей [53, 54, 74,

75] функціонують переважно як емпіричні моделі, що ускладнює інтерпретацію отриманих результатів з точки зору фізичних механізмів формування електроспоживання. Крім того, такі підходи орієнтовані на аналіз окремих часових інтервалів і не забезпечують повного врахування довгострокових кореляційних та циклічних властивостей процесу електроспоживання. Функціонування системи передбачає обов'язковий етап препроцесингу даних, що є критично важливим в умовах сучасних електроенергетичних систем. Втрата або спотворення даних, спричинені нестабільністю інформаційно-комунікаційної інфраструктури, ускладнюють адекватне оцінювання стану мережі та можуть призводити до як технічних, так і економічних наслідків [76].

На основі проведеного аналізу систематизовано основні методи машинного навчання, що можуть бути використані для класифікації станів електроенергетичної системи, їх наведено у таблиці 1.1.

Таблиця 1.1

Методи машинного навчання для класифікації станів електроенергетичної системи

Метод МН	Обчислювальні вимоги	Переваги у завданнях енергетики	Недоліки у завданнях енергетики
<b>Logistic Regression</b>	Дуже низькі	Найвища інтерпретованість; швидка робота в реальному часі.	Не вловлює складні нелінійні взаємозв'язки
<b>LDA</b>	Низькі	Ефективний при нормальному розподілі ознак.	Чутливий до аномальних викидів у даних Smart Meter.
<b>Naive Bayes</b>	Дуже низькі	Висока стійкість до нерелевантних ознак; не потребує багато пам'яті.	Припущення про незалежність ознак часто не справджується
<b>KNN (k=5)</b>	Високі (інференс)	Не потребує етапу навчання; добре кластеризує типові добові профілі.	Повільна робота при великих обсягах наборів даних.

## Продовження таблиці 1.1

Метод МН	Обчислювальні вимоги	Переваги у завданнях енергетики	Недоліки у завданнях енергетики
<b>Decision Tree</b>	Низькі	Легко візуалізувати логіку прийняття рішень (дерево рішень).	Схильність до перенавчання на дрібних флуктуаціях електроспоживання.
<b>Random Forest</b>	Середні	Висока стійкість до шуму; ефективна обробка великої кількості ознак.	Потребує більше пам'яті для зберігання великої кількості дерев.
<b>Extra Trees</b>	Середні	Менша дисперсія помилок порівняно з Random Forest; швидше навчання.	Менша інтерпретованість через випадковий вибір порогів розщеплення.
<b>AdaBoost</b>	Середні	Концентрується на "важких" для класифікації випадках (перехідні режими).	Чутливий до зашумлених даних та аномалій.
<b>XGBoost</b>	Високі	Висока точність класифікації режимів; вбудований механізм відбору ознак; стійкість до дисбалансу класів.	Потребує ретельного підбору гіперпараметрів; схильний до перенавчання без регуляризації.
<b>Gradient Boosting</b>	Високі	Найвища точність; гнучкість у налаштуванні через learning rate.	Тривалий час навчання; складність підбору гіперпараметрів.
<b>LightGBM</b>	Середні	Швидке навчання на великих масивах телеметричних даних; ефективна обробка числових ознак.	Чутливий до малих вибірок; може пропускати рідкісні аномальні режими.
<b>SVM (RBF kernel)</b>	Високі	Ефективний у просторах високої розмірності (багато компонент Фур'є).	Потребує ретельної нормалізації даних; складно масштабується.
<b>MLP Neural Net</b>	Високі	Здатність до самостійного виявлення глибоких закономірностей у даних.	Потребує великих навчальних вибірок; ефект «чорної скриньки».

Аналіз наведених підходів показує, що у більшості з них ознаковий простір формується безпосередньо з вхідних даних вимірювань або їх амплітудно-частотних перетворень без урахування стохастично-повторюваної структури сигналу електроспоживання.

Незважаючи на високу ефективність сучасних методів машинного та глибокого навчання, більшість із них орієнтована на емпіричне виявлення залежностей у даних без урахування внутрішньої структури процесів електроспоживання. Зокрема, такі підходи не враховують фізичну природу електроспоживання, що є характерною особливістю енергетичних систем. Це обмежує їх здатність до адекватного моделювання залежностей, знижує інтерпретованість результатів та потребує значних обсягів навчальних даних. Крім того, моделі машинного навчання демонструють чутливість до змін режимів функціонування електроенергосистеми, що ускладнює їх застосування в умовах нестаціонарності та високої варіабельності електроспоживання.

Методи МН застосовуються для завдань прогнозування електроспоживання, виявлення аномалій та оперативного керування в енергетиці, що підтверджує їх доцільність у завданнях аналізу електроспоживання. Проте їхню ефективність обмежується тим, що вони здебільшого працюють із даними без попереднього виділення інформативних ознак та не враховують внутрішню стохастично-повторювану структуру сигналу. Це зумовлює доцільність використання математичного апарату, який дозволить сформувати інформативні ознаки для класифікації режимів електроспоживання електроенергосистеми.

Порівняльну характеристику методів аналізу та опрацювання даних електроспоживання наведено в таблиці 1.2.

Таблиця 1.2

Порівняльна характеристика методів аналізу та опрацювання даних  
електроспоживання

Методи/ моделі	Переваги	Недоліки	Придатність для аналізу електроспоживання
<b>Статистичні моделі часових рядів</b>	Простота реалізації, висока інтерпретованість	Припускають стаціонарність; не враховують циклічну нестационарність	Обмежена (добре працюють лише для стабільних режимів)
<b>Періодично корельовані випадкові процеси (ПКВП)</b>	Враховують повторювальну нестационарність; адекватно описують стохастичну структуру	Складність оцінювання параметрів; потребують спеціалізованого математичного апарату	Дуже висока (природно моделюють повторюваність електроспоживання)
<b>Циклічні випадкові процеси (ЦВП)</b>	Описують циклічні сигнали як з регулярним, так і з нерегулярним ритмом; циклічна структура ймовірнісних характеристик; ЦВП як частковий випадок включає періодичний випадковий процес	Потребують попередньої сегментації процесу на цикли та зони; складність оцінювання функції ритму	Середня (апробовані для газоспоживання та ЕКГ-сигналів; для електроспоживання потребують адаптації моделі)
<b>Методи машинного навчання</b>	Висока точність апроксимації нелінійних залежностей	Ігнорують глибокі часові кореляції між циклами; потребують великих даних	Висока (особливо для класифікації в реальному часі)
<b>Методи глибокого навчання</b>	Моделюють складні часові залежності; враховують довгі послідовності	Надмірне згладжування екстремальних значень; проблема "чорної скриньки"	Висока (але вимагають GPU; неефективні для виділення екстремумів)

У зв'язку з цим актуальним є розроблення підходів, які поєднують можливості машинного навчання з методами, що враховують специфіку електроспоживання.

Одним із таких підходів є використання моделей періодично корельованих випадкових процесів, які дозволяють враховувати стохастичність та повторюваність даних електроспоживання. Поєднання ПКВП із методами машинного навчання створює передумови для класифікації режимів електроспоживання, зменшення вимог до обсягу навчальних даних та підвищення інтерпретованості отриманих результатів.

У межах даної роботи методи машинного навчання розглядаються не як самостійна альтернатива математичному моделюванню, а як інструмент класифікації режимів електроспоживання на основі інформативних ознак, сформованих із застосуванням стохастичних моделей.

#### **1.4. Висновки до розділу 1**

За результатами аналітичного огляду сучасного стану розподілених електроенергетичних систем та методів опрацювання даних електроспоживання сформульовано такі висновки.

1. На основі проведеного аналізу сучасного стану розподілених електроенергетичних систем встановлено, що їх розвиток відбувається в умовах концепції Energy 4.0, що характеризується децентралізацією генерації, інтеграцією відновлюваних джерел енергії та широким використанням інформаційно–комунікаційних технологій. Визначено, що в таких умовах своєчасне визначення режиму електроспоживання вузлів розподіленої електроенергетичної системи є необхідною умовою забезпечення стабільності та надійності електропостачання.

2. Проведено аналіз підходів до опрацювання даних електроспоживання та встановлено, що основними прикладними напрямками є

прогнозування електроспоживання, виявлення аномалій у профілях споживання та оцінювання режимів роботи вузлів мережі. Обґрунтовано, що важливою серед них є класифікація режимів електроспоживання електроенергетичної системи, яка забезпечує оперативне реагування на відхилення від штатного режиму роботи, особливо в умовах пікових навантажень та дії дестабілізуючих факторів.

3. Проведено порівняльний аналіз відомих моделей та методів опрацювання даних електроспоживання в розподілених електроенергетичних системах. Встановлено, що класичні статистичні моделі, трендові, ARIMA, SARIMA та їх модифікації ефективно описують регулярну складову процесу електроспоживання, однак для повноцінного врахування одночасно стохастичної природи сигналу та його повторюваної циклічної структури потребують подальшого розвитку та узагальнення.

4. Встановлено, що для адекватного аналізу режимів роботи електроенергетичних систем, зокрема для виявлення аномальних і пікових режимів та класифікації збалансованих і незбалансованих режимів електроспоживання, необхідно враховувати одночасно повторювану та стохастичну природу процесу електроспоживання. Це зумовлює доцільність використання моделі, яка дозволяє описати як детерміновану циклічну складову, так і стохастичну мінливість характеристик сигналу.

5. Проведено аналіз сучасних методів машинного навчання, зокрема ансамблевих, метричних та нейромережових підходів, які забезпечують високу точність моделювання нелінійних залежностей та ефективно застосовуються для завдань прогнозування і класифікації в енергетиці. Водночас встановлено, що більшість таких методів мають емпіричний характер, застосовуються безпосередньо до вхідних даних вимірювань без попереднього врахування внутрішньої стохастично-періодичної структури процесу електроспоживання та характеризуються обмеженою інтерпретованістю результатів, що ускладнює ідентифікацію причин відхилення системи від збалансованого стану.

6. Встановлено, що існуючі підходи до аналізу даних електроспоживання здебільшого розглядають стохастичну природу процесу та його повторювану структуру окремо. Поєднання математичного апарату періодично корельованих випадкових процесів із методами машинного навчання дозволяє врахувати обидва аспекти одночасно, що підвищує ефективність аналізу режимів роботи вузлів електроенергетичних систем при визначенні їх збалансованості.

7. На основі проведеного аналізу сформульовано науково–прикладну проблему, що полягає у розробленні методів аналізу даних електроспоживання, які поєднують математичний апарат періодично корельованих випадкових процесів та методи машинного навчання для підвищення точності класифікації режимів електроспоживання розподілених електроенергетичних систем та забезпечення адаптивного керування процесами передавання даних.

## РОЗДІЛ 2. МОДЕЛІ ТА МЕТОДИ АНАЛІЗУ ДАНИХ ЕЛЕКТРОСПОЖИВАННЯ

### 2.1. Аналіз процесу електроспоживання

Процес електроспоживання характеризується стохастичною та вираженою багатомасштабною повторюваною структурою, яка формується внаслідок накладання різних періодичних ритмів функціонування електроенергетичної системи. Аналіз реальних даних показує, що електроспоживання містить декілька вкладених часових масштабів, кожен з яких відображає окремі закономірності функціонування системи [77].

Зокрема, у структурі сигналу можна виділити такі основні періодичні компоненти:

- добову ( $T = 24$  години), що визначається ритмами життєдіяльності споживачів, графіками роботи підприємств тощо;
- тижневу ( $T = 168$  годин), пов'язану з відмінностями між робочими та вихідними днями;
- місячну ( $T \approx 30$  днів), що відображає організаційні та експлуатаційні фактори;
- сезонну ( $T \approx 365$  днів), обумовлену кліматичними умовами.

Короткострокові коливання (добові) модулюються довгостроковими змінами середнього рівня електроспоживання, що свідчить про ієрархічну структуру процесу [78]. Така багатомасштабність зумовлює необхідність використання моделей, здатних одночасно враховувати повторюваність і стохастичну мінливість сигналу.

Для верифікації запропонованої математичної моделі та методів аналізу використано реальні погодинні дані електроспоживання приватного підприємства, що спеціалізується на виробництві та обробці насінневої продукції.

Графік електроспоживання демонструє виражену сезонну складову, характерну для підприємств агропромислового комплексу: знижений рівень споживання у зимово-весняний період відповідає міжсезонню, коли функціонують переважно системи зберігання та адміністративна інфраструктура, поступове зростання середнього споживання починаючи з травня та досягнення піку у серпні-вересні відповідає активній фазі збирання врожаю та інтенсивного завантаження сушарок, сепараторів і транспортного обладнання. Плавний спад у жовтні-грудні відображає завершення сезону збирання та обробки врожаю. Протягом усього року зберігається стійка добова циклічність із характерними піками у робочий час та нічними мінімумами, що наочно підтверджується графіками на рисунках 2.1–2.4.

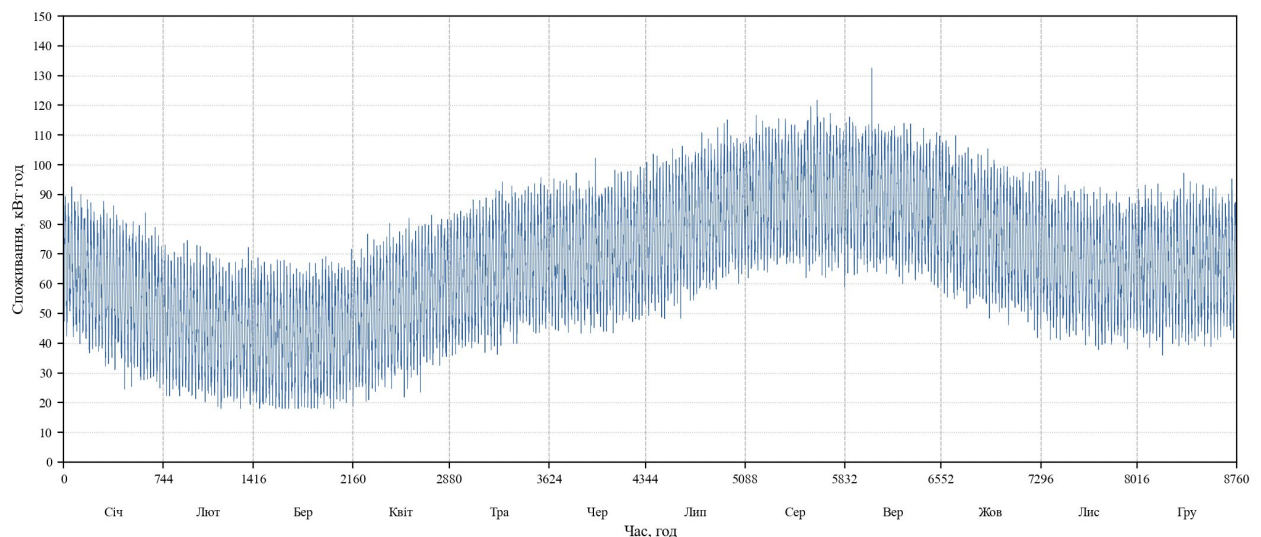


Рисунок 2.1 – Графік електроспоживання приватного підприємства збалансованого режиму в річному інтервалі.

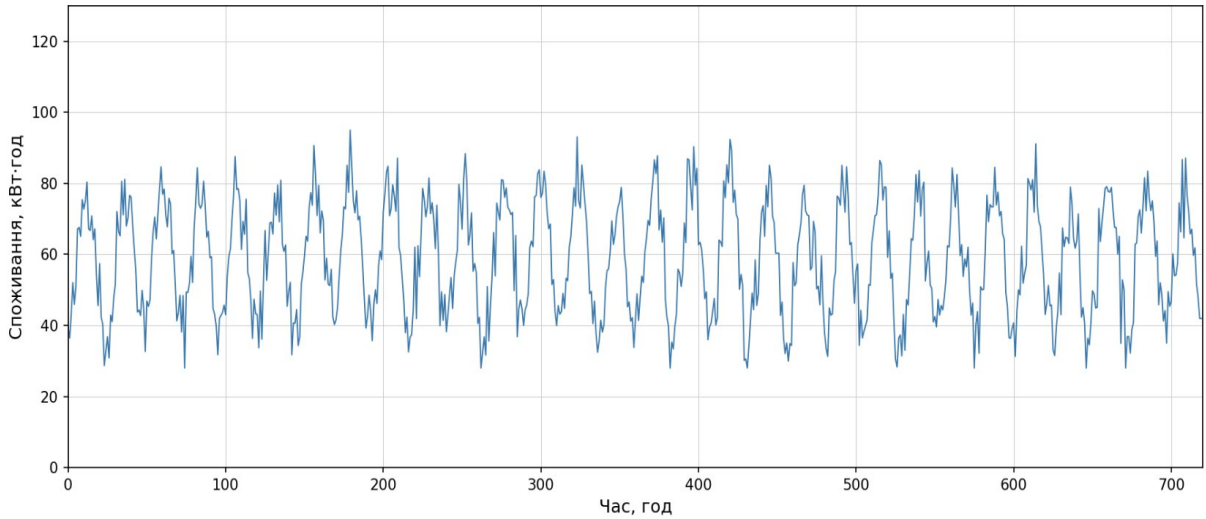


Рисунок 2.2 – Графік електроспоживання приватного підприємства збалансованого режиму в місячному інтервалі.

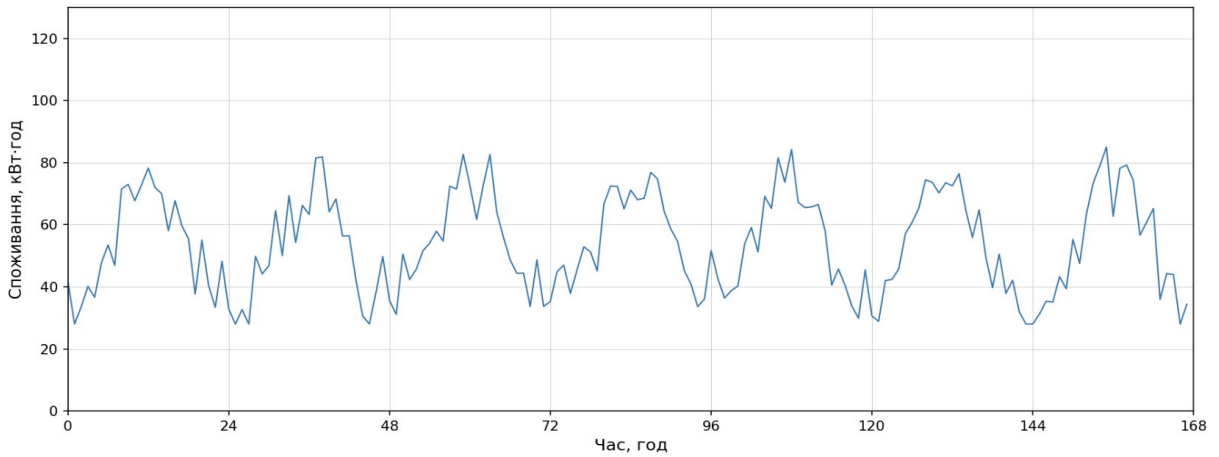


Рисунок 2.3 – Графік електроспоживання приватного підприємства збалансованого режиму в тижневому інтервалі.

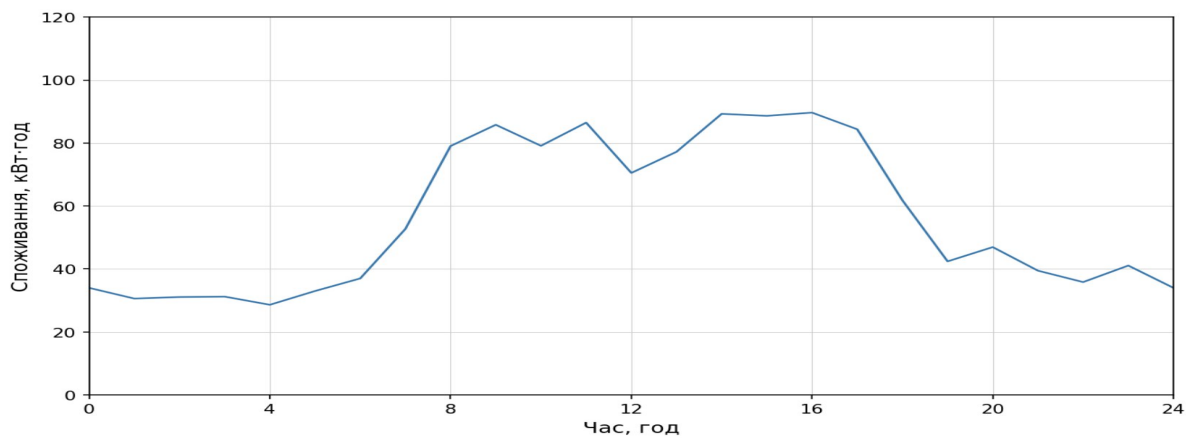


Рисунок 2.4 – Графік електроспоживання приватного підприємства збалансованого режиму в денному інтервалі.

При роботі електроенергетичної системи інколи виникають певні збої в її роботі внаслідок невідповідності між обсягами виробництва та споживання електричної енергії, що супроводжується відхиленням частоти та напруги від номінальних значень, зростанням споживання на елементи мережі та підвищенням ризику порушення стійкості функціонування системи, тобто її розбалансуванням.

Такий стан формується під впливом сукупності взаємопов'язаних факторів, серед яких важливу роль відіграє варіабельність відновлюваних джерел енергії, генерація яких залежить від погодних умов і має стохастичний характер. Додатковим чинником є нерівномірність добового графіка електроспоживання, що супроводжується формуванням пікових періодів споживання та різкими змінами режимів роботи мережі.

Стан збалансованості електроенергетичної системи визначається як режим її функціонування, за якого забезпечується відповідність між обсягами генерації та споживання електричної енергії з урахуванням технологічних втрат, а також підтримуються нормативні значення частоти та напруги в мережі. У класичному підході збалансованість системи безпосередньо пов'язана з рівновагою активної потужності, порушення якої призводить до відхилення частоти від номінального значення та зміни режимів напруги в вузлах електроенергетичної мережі.

Незбалансований стан характеризується виникненням дефіциту або надлишку генерації відносно поточного рівня споживання, що проявляється у динамічних відхиленнях параметрів системи та може призводити до передаварійних або аварійних режимів її функціонування. Особливої критичності такі стани набувають у періоди пікових навантажень, коли резерви регулювання є обмеженими.

У даній роботі режим електроспоживання електроенергетичної системи розглядається з позицій аналізу даних електроспоживання. На відміну від традиційного підходу, що базується на безпосередньому контролі фізичних

параметрів системи, запропоновано інтерпретувати збалансованість як стан, що відображається у статистичних характеристиках сигналу електроспоживання. Такий підхід дозволяє здійснювати оцінку режиму функціонування енергосистеми на основі аналізу її споживчої поведінки. Зокрема, збалансований стан відповідає стабільній періодичній структурі сигналу електроспоживання з відтворюваними статистичними характеристиками, тоді як незбалансований режим проявляється у порушенні цієї структури, зростанні варіативності та появі аномальних відхилень.

Це створює передумови для використання методів машинного навчання з метою класифікації станів електроенергетичної системи. Приклад електроспоживання при незбалансованості електроенергетичної системи наведено на рисунку 2.5.

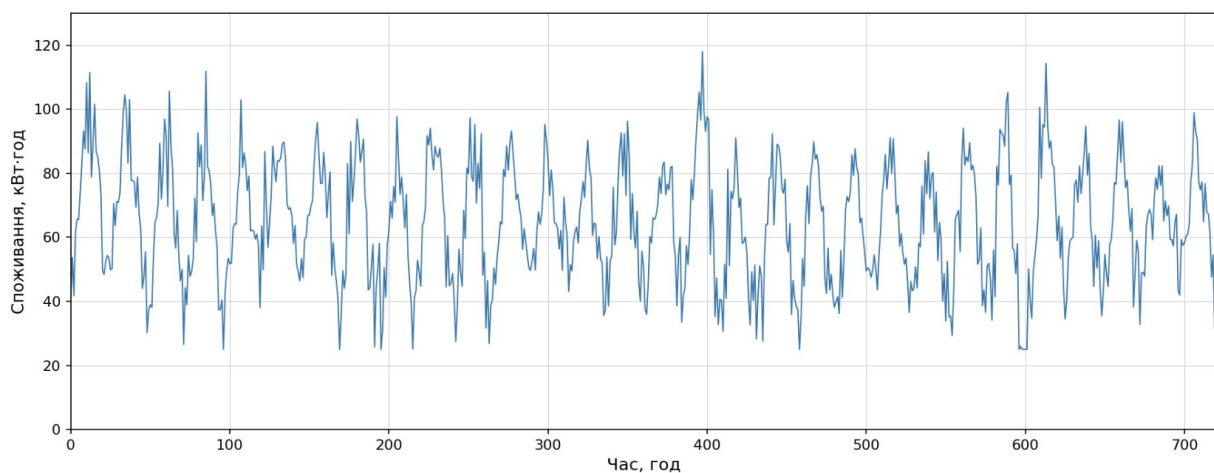


Рисунок 2.5 – Графік електроспоживання приватного підприємства в місячному інтервалі для незбалансованого режиму

Графік характеризується значною міжцикловою варіативністю амплітуди та нестабільністю пікових навантажень. Статистичні характеристики процесу – дисперсія та кореляційні компоненти змінюються від циклу до циклу, що свідчить про підвищену стохастичну мінливість параметрів електроспоживання та відповідає незбалансованому режиму роботи системи (Клас 2). На графіку збалансованого режиму

спостерігаються стабільні добові цикли з чітко вираженими піками електроспоживання, що повторюються кожні 24 години. Низька міжциклова варіативність амплітуди та стабільна форма добового профілю підтверджують передбачувану динаміку електроспоживання, що відповідає збалансованому режиму роботи системи (Клас 1).

Враховуючи виражену періодичну нестационарність енергетичного електроспоживання, математична модель сигналу  $\xi(t)$  формалізується у вигляді адитивної композиції детермінованої та стохастичної складових:

$$\xi(t) = \xi_s(t) + \xi_p(t), t \in \mathbf{R} \quad (2.1)$$

де  $\xi_s(t)$  центрована стохастична компонента, яка відображає випадкові варіації потужності, обумовлені внутрішньою нестабільністю системи;

$\xi_p(t)$  детермінована періодична компонента, що описує сталі режими функціонування об'єкта та базові добові ритми електроспоживання ( $T_k = 24$  год). Саме ця властивість зумовлює доцільність використання моделей, в яких поєднується стохастичність та повторюваність [79, 80, 81, 82, 83].

Зокрема, до таких моделей належать періодично корельовані випадкові процеси, циклічні випадкові процеси, які, своєю чергою, дозволяють враховувати періодичну змінність статистичних характеристик сигналу електроспоживання. Це забезпечує більш адекватне відображення реальної структури процесу електроспоживання порівняно з класичними стаціонарними моделями та створює основу для формування інформативних ознак з метою подальшої класифікації режимів роботи електроенергетичної системи.

## **2.2. Модель циклічного випадкового процесу для аналізу даних електроспоживання**

Процес електроспоживання характеризується двома фундаментальними властивостями, що визначають вибір адекватної математичної моделі. З одного боку, спостерігається виражена повторюваність структури сигналу: добові цикли електроспоживання відтворюються з характерними піками ранкового та вечірнього споживання і нічним мінімумом. З іншого боку, параметри кожного окремого циклу: амплітуда піків, тривалість фаз підвищеного електроспоживання, форма профілю варіюють випадковим чином під впливом погодних умов, поведінки споживачів та режимів роботи обладнання. Процес є одночасно циклічним за структурою та стохастичним за характером реалізації кожного окремого циклу.

Зазначена подвійна природа процесу зумовлює необхідність застосування математичної моделі, що органічно поєднує циклічність і стохастичність без зведення жодної з цих властивостей до другорядної. Серед існуючих класів стохастичних моделей циклічний випадковий процес (ЦВП) є найбільш узагальненим підходом, що дозволяє описувати сигнали як з регулярним, так і з нерегулярним ритмом розгортання циклів у часі [79, 80].

Апарат ЦВП знаходить широке застосування у різних галузях науки та техніки, де досліджувані сигнали мають циклічну стохастичну природу. Зокрема, у матеріалознавстві модель ЦВП застосовується для аналізу впорядкованого рельєфу поверхні матеріалів після лазерної обробки [84]. У біомедичній діагностиці моделі циклічних випадкових процесів використовуються для опису електрокардіосигналів та інших фізіологічних сигналів зі змінним ритмом [85, 86, 87, 88, 89]. Також цей математичний апарат використовується при аналізі процесу споживання газу [90], описі економічних процесів [91] та ін.

Формально циклічний випадковий процес визначається як стохастичний процес  $\xi(\omega, t)$ ,  $t \in R$ ,  $\omega \in \Omega$  для якого існує функція ритму  $T(t, n)$ , що описує варіацію відстані між однофазовими значеннями різних циклів і для якого вектори значень процесу у відповідних фазових точках різних циклів є

стохастично еквівалентними [85, 92]. Ключовим поняттям моделі є цикл мінімальний відрізок реалізації, що містить повну послідовність фаз процесу, та функція ритму, що визначає тривалість циклів і може бути як сталою, так і змінною.

Для ЦВП із зонно-часовою структурою цикл розбивається на зони, що відповідають різним режимам функціонування процесу, і статистичні оцінки обчислюються окремо для кожної зони. Оцінки математичного сподівання та кореляційної функції ЦВП є незміщеними та узгодженими за умови ергодичності процесу та достатньої кількості спостережених циклів.

Однією із таких широко відомих моделей є циклічно випадкові процеси (ЦВП) [89, 90]. Використання вищезгаданої моделі дозволяє розширити кількість методів аналізу, переходячи від спектрального розкладу до безпосереднього дослідження ймовірнісної структури сигналу в часовій області [80, 90].

Такий підхід забезпечує всебічне дослідження електроспоживання, оскільки дозволяє оцінювати не лише загальні тренди, а й стохастичну мінливість характеристик безпосередньо для кожної фази (години) добового циклу. У межах даної моделі електроспоживання розглядається як випадковий процес, для якого статистичні характеристики змінюються циклічно в часі з періодом  $T = 24$  години. Це дозволяє формалізувати «подвійну стохастичність» сигналу, випадковість значень та випадковість параметрів самого циклу. Однією з базових характеристик циклічного випадкового процесу є те, що математичне сподівання повторюється циклічно, яке відображає усереднений профіль електроспоживання в межах одного циклу [92] (рис. 2.7).

Реалізація статистичної оцінки середнього рівня електроспоживання  $m_{\xi_{\omega}}(t)$  здійснюється шляхом усереднення за однією реалізацією, де тривалість періоду дорівнює 24 години:

$$\hat{m}_{\xi_i}(t) = \frac{1}{M} \sum_{i=1}^{M-1} \xi_{i_{\omega}}(t + T(t,n)), i = \overline{1, M} \quad (2.2)$$

де  $\hat{m}_{\xi_i}$  – статистична оцінка математичного сподівання сигналу електроспоживання;

$(t)$  – реалізація сигналу електроспоживання для фіксованої фази циклу;

$T$  – період циклу (для добового електроспоживання  $T=24$  години);

$M$  – кількість циклів у вибірці;

$n$  – номер дискретного відліку;

$i = \overline{1, M}$  – номер реалізації (компоненти) процесу;

$T(t, n)$  – функція ритму

Застосування такого усереднення відповідає процедурі когерентного усереднення та дозволяє виділити базовий добовий профіль електроспоживання. Непостійність оцінки математичного сподівання за аргументом часу є прямим емпіричним підтвердженням нестационарності процесу та циклічної змінності його статистичних характеристик [76]. На рисунку 2.6 наведено дані електроспоживання приватного підприємства.

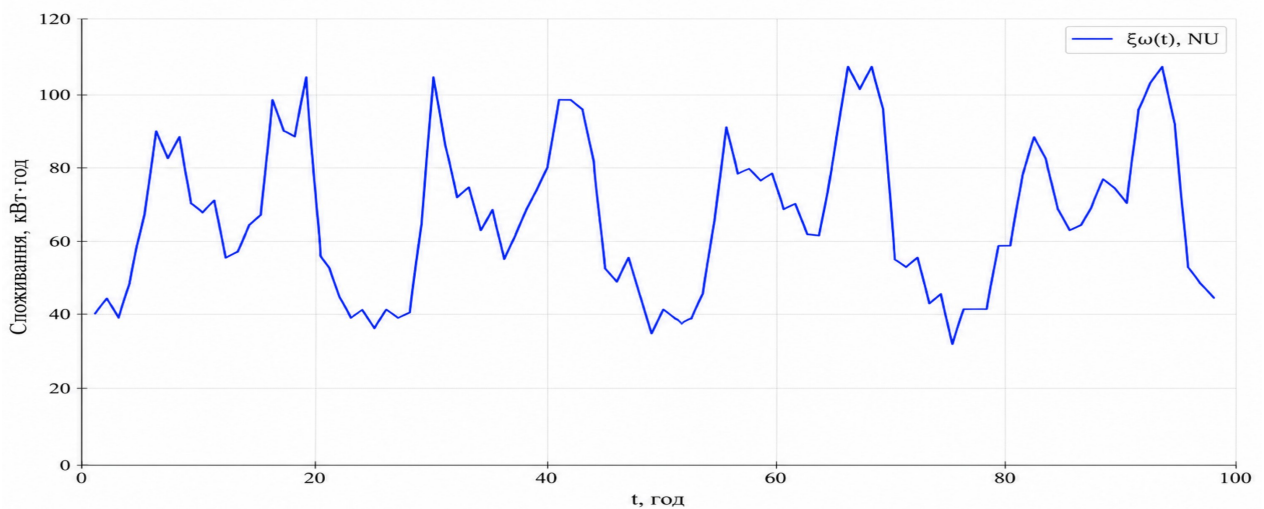


Рисунок 2.6 – Дані електроспоживання приватного підприємства за чотири повні доби

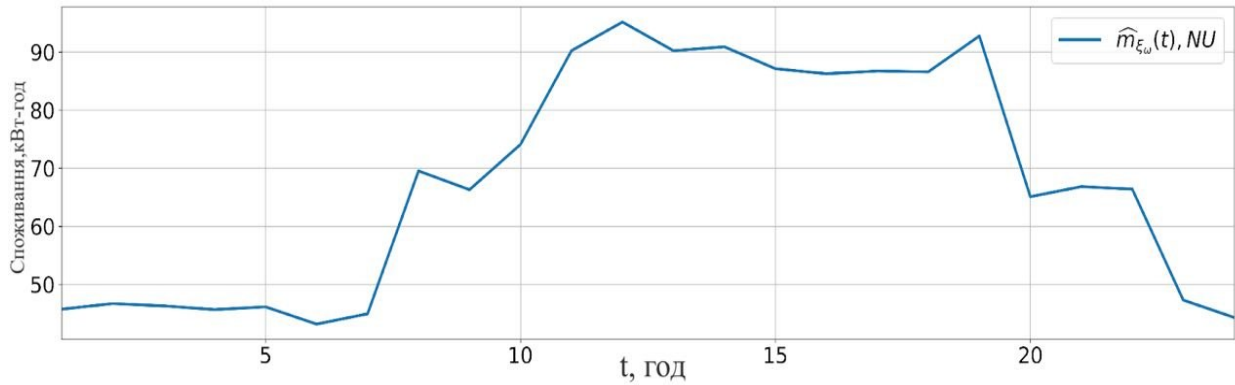


Рисунок 2.7 – Графік реалізації статистичних оцінок математичного сподівання сигналу електроспоживання приватного підприємства

Наступною важливою характеристикою циклічного випадкового процесу є циклічна дисперсія, яка характеризує інтенсивність випадкових флуктуацій електроспоживання в межах добового циклу. Статистична оцінка дисперсії визначається як:

$$\hat{d}_{\xi_i}(t) = \frac{1}{M} \cdot \sum_{i=1}^{M-1} \left[ \xi_{i\omega}(t + T(t,n)) - \hat{m}_{\xi_i}(t + T(t,n)) \right]^2, \quad i = \overline{1, M}, \quad (2.3)$$

Періодична змінність дисперсії свідчить про нерівномірну стохастичну активність процесу впродовж добового циклу: мінімальні значення відповідають стабільним нічним режимам, тоді як максимальні значення спостерігаються у пікові години електроспоживання, що характеризуються підвищеною невизначеністю.

Для більш повного опису одновимірних розподілів значень електроспоживання в межах циклу доцільним є аналіз початкових моментів вищих порядків.

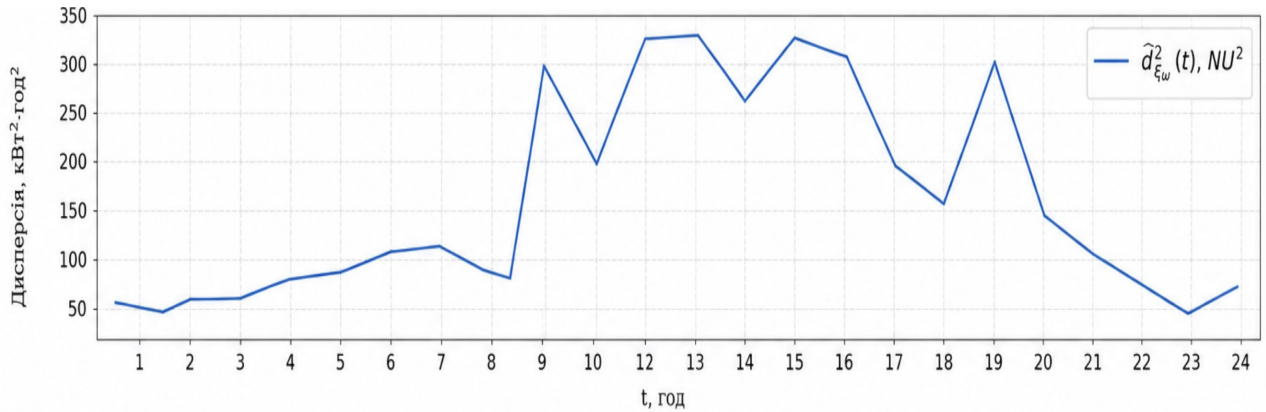


Рисунок 2.8 – Графік реалізацій статистичних оцінок дисперсії сигналу електроспоживання

Статистична оцінка початкового моменту  $k$ -го порядку  $(\hat{m}_{\xi_i})^k(t)$  для кожної реалізації  $\xi_i(\omega, t)$  циклічного випадкового процесу визначається співвідношенням:

$$(\hat{m}_{\xi_i})^k(t) = \frac{1}{M} \sum_{i=1}^{M-1} \xi_{i\omega}^k(t + T(t, n)), i = \overline{1, M}, \quad (2.4)$$

Отримані оцінки початкових моментів є періодичними функціями часу, що відрізняє циклічний випадковий процес від стаціонарних моделей, у яких відповідні моменти є сталими величинами.

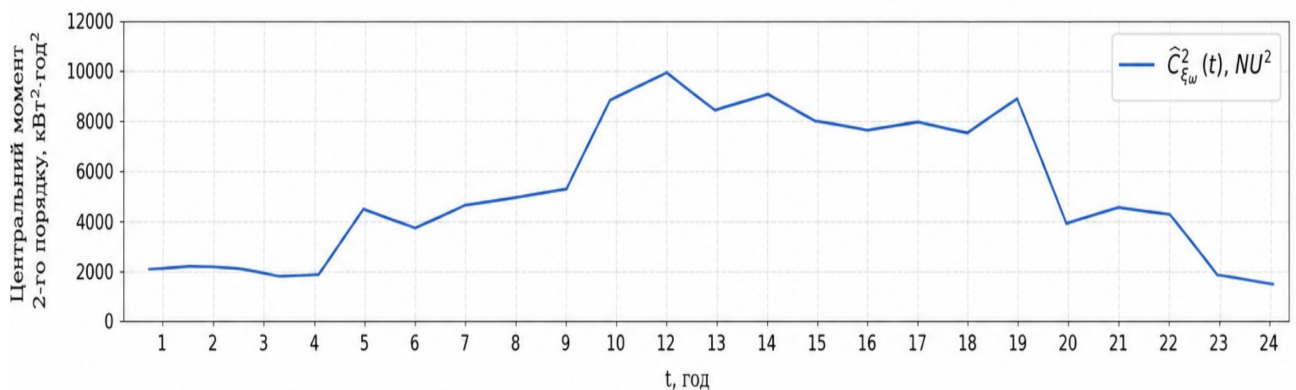


Рисунок 2.9 – Графік реалізації статистичної оцінки функції центрального моменту другого порядку

Аналіз моментів другого та третього порядків дозволяє оцінити зміну енергії флуктуацій і асиметрії розподілу електроспоживання залежно від фази добового циклу.

Для усунення впливу середнього рівня та детальнішого аналізу стохастичних властивостей процесу використовуються центральні моменти вищих порядків. Емпірична оцінка центрального моменту  $k$ -го порядку  $(\hat{d}_{\xi_i})^k(t)$  для компоненти процесу  $\xi_i(\omega, t)$ , визначається як:

$$(\hat{d}_{\xi_i})^k(t) = \frac{1}{M-1} \cdot \sum_{i=1}^{M-1} [\xi_{i\omega}(t + T(t, n)) - \hat{m}_{\xi_i}(t + T(t, n))]^k, i = \overline{1, M}, \quad (2.5)$$

Відповідно до теорії циклічних випадкових процесів статистичні характеристики ЦВП – математичне сподівання та автоковаріаційна функція – є  $T$ -періодичними функціями часу, що відображають залежність стохастичних властивостей процесу від фази добового циклу.

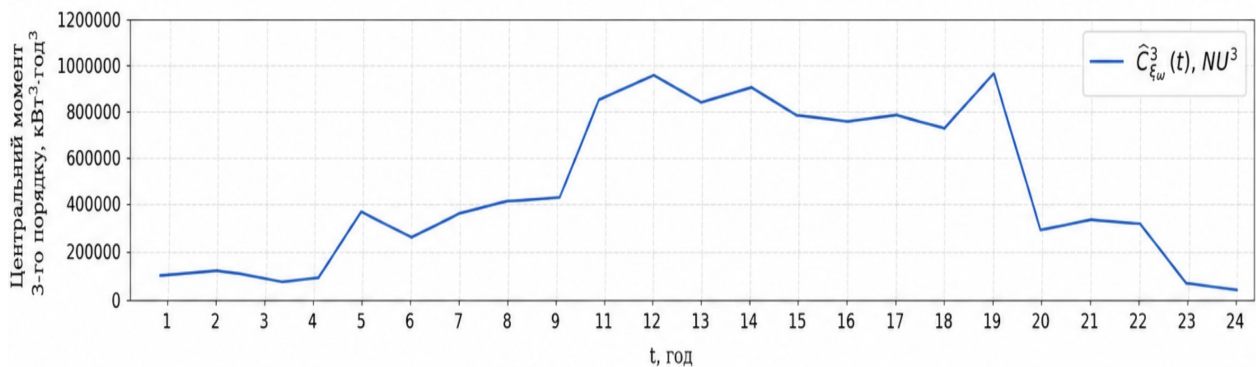


Рисунок 2.10 – Графік реалізації статистичної оцінки функції центрального моменту третього порядку

Для дослідження часових залежностей між значеннями електроспоживання у різні моменти часу використовується автоковаріаційна функція, яка відображає лінійну статистичну залежність між значеннями

процесу в різних фазах циклу. Емпірична оцінка автоковаріаційної функції визначається співвідношенням:

$$\hat{C}_{p_{\xi_{i_1}, \dots, \xi_{i_k}}}(t_1, \dots, t_k) = \frac{1}{M-M_1+1} \cdot \sum_{i=1}^{M-M_1} \left[ \xi_{i_1 \omega}^{R_1}(t_1 + T(t_{1,n})) \cdot \dots \cdot \xi_{i_k \omega}^{R_k}(t_k + T(t_{k,n})) \right], i_k = \overline{1, M}, \quad (2.6)$$

Аналіз автоковаріаційної функції виявляє характерні піки, зокрема позадіагональні максимуми, що свідчать про наявність міжциклових кореляцій і ефекту «пам'яті» системи щодо добового циклу.

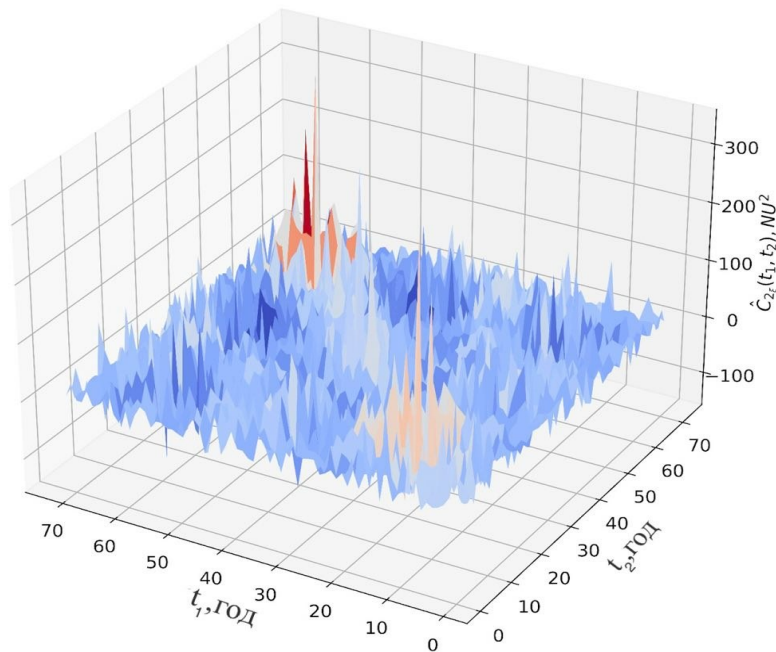


Рисунок 2.11 – Графік реалізації статистичних оцінок автоковаріаційної функції

Для нормування кореляційних залежностей використовується автокореляційна функція, емпірична оцінка якої має вигляд:

$$\hat{R}_{p_{\xi_{i_1}, \dots, \xi_{i_k}}}(t_1, \dots, t_k) = \frac{1}{M-M_1} \cdot \sum_{i=1}^{M-M_1} \left( \xi_{i_1 \omega}(t_1 + T(t_{1,n})) - \hat{m}_{\xi_{i_1}}(t_1 + T(t_{1,n})) \right)^{r_1} \cdot \dots \cdot \left( \xi_{i_k \omega}(t_k + T(t_{k,n})) - \hat{m}_{\xi_{i_1}}(t_k + T(t_{k,n})) \right)^{r_k}, (i)_k = \overline{1, M}, \quad (2.7)$$

Аналіз автокореляційної функції підтверджує циклічно повторювану структуру часових залежностей та узгоджується з концепцією циклічного випадкового процесу.

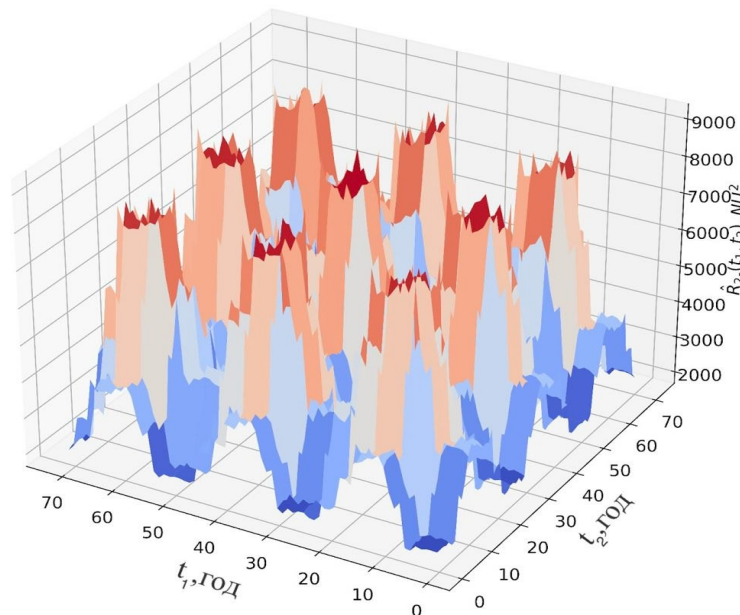


Рисунок 2.12 – Графік реалізації статистичних оцінок автокореляційної функції

Результати аналізу моментних та кореляційних характеристик експериментальних даних переконливо підтверджують циклічну стохастичну природу електроспоживання [48].

Апарат ЦВП у найбільш повному обсязі застосовується тоді, коли функція ритму є невідомою і потребує окремої ідентифікації ритмічної структури.

Змінність ритму циклів є визначальною характеристикою біомедичних сигналів, зокрема електрокардіосигналів, де тривалість серцевого циклу варіює залежно від функціонального стану організму [87, 88, 89].

Оцінювання параметрів ЦВП передбачає обов'язковий етап сегментації реалізації та ідентифікації ритмічної функції процедури [92, 93], реалізація яких ускладнюється в умовах зашумленості вимірювальних даних та наявності пропусків, а ритмо-адаптивні методи характеризуються суттєво вищою обчислювальною складністю порівняно з методами аналізу ПКВП. Оскільки добовий цикл електроспоживання є детермінованим і відомим наперед, період  $T_k = 24$  год задається явно на етапі збору даних, це усуває необхідність ідентифікації ритмічної функції і безпосередньо забезпечує застосованість апарату ПКВП без додаткових обчислювальних витрат.

### **2.3. Математична модель періодично корельованого випадкового процесу для аналізу даних електроспоживання**

ПКВП належать до класу нестационарних стохастичних процесів, для яких основні статистичні характеристики математичне сподівання та кореляційна функція є періодичними функціями часу з фіксованим періодом  $T_k$  [49, 50, 51].

Формально ПКВП визначаються як випадкові процеси другого порядку, кореляційна структура яких залежить не лише від величини затримки в часі, а й від абсолютного моменту часу, що дозволяє детально описувати циклічну мінливість процесу. Однією з ключових переваг моделей ПКВП є те, що вони дають змогу оцінювати статистичні характеристики математичне сподівання, дисперсію, коваріаційні функції на основі реальних даних без необхідності мати нескінченні послідовності спостережень. Вони створюють математичне підґрунтя для спектрального аналізу та побудови лінійних оцінювачів параметрів процесу [49, 50, 51].

Застосування ПКВП як базової моделі створює методологічні передумови для побудови формалізованих процедур аналізу, які дають змогу відокремити детерміновані закономірності від стохастичних флуктуацій та забезпечити підвищення точності роботи в умовах нестационарності.

Апарат ПКВП широко застосовується в різних галузях науки та техніки, наприклад, для аналізу стану механічних систем [94], вібраційної діагностики [95, 96], біомедичної діагностики сигналів [97, 98, 99, 100], телекомунікаціях [101, 102] та ін.

У межах цієї моделі електроспоживання розглядається як процес, у якому повторюваність проявляється не лише у значеннях сигналу, а й у періодичній зміні його статистичних характеристик, зокрема математичного сподівання, дисперсії та кореляційної функції.

Формально ПКВП визначаються як випадкові процеси другого порядку, для яких виконується умова періодичності статистичних характеристик [49, 50, 51]:

$$m_{\xi}(t) = m_{\xi}(t + T_k), \quad r_{\xi}(t, s) = r_{\xi}(t + T_k, s + T_k), \quad t, s \in \mathbf{R} \quad (2.8)$$

де  $T_k$  – період кореляції сигналу ( $T_k = 24$  год для добового циклу електроспоживання).

На відміну від стаціонарних процесів, у яких кореляційні властивості залежать лише від часового зсуву, у ПКВП кореляційна структура залежить також від абсолютного моменту часу всередині періоду, що відображає періодичну змінність статистичних характеристик сигналу [49, 50, 51].

Така математична структура природно відображає ієрархічні циклічні патерни реальних електроенергетичних сигналів: добові коливання споживання електроенергії, тижневі цикли промислової активності та сезонні варіації енергобалансу.

Для підтвердження відповідності реального сигналу електроспоживання умовам періодичності, на рисунках 2.13–2.16 наведено сигнали та статистичні характеристики для збалансованого та незбалансованого режимів електроспоживання електроенергетичної системи. На рисунку 2.13 наведено реалізацію сигналу для збалансованого режиму. На відміну від незбалансованого, добовий профіль відтворюється з помітно більшою стабільністю від реалізації до реалізації – форма кривої зберігається, а амплітуда коливань є меншою.



Рисунок 2.13 – Сигнал електроспоживання приватного підприємства у збалансованому режимі

Реалізацію сигналу електроспоживання для незбалансованого режиму наведено на рисунку 2.14. Сигнал демонструє періодичну структуру з характерним нічним мінімумом та денно-вечірнім зростанням електроспоживання. При цьому амплітуда піків суттєво відрізняється від реалізації до реалізації, а форма добового профілю не є стабільною, що є ознакою значної стохастичної мінливості процесу.



Рисунок 2.14 – Сигнал електроспоживання приватного підприємства у незбалансованому режимі

Добовий профіль математичного сподівання  $\hat{m}_\xi(t)$  для обох режимів наведено на рисунку 2.15. Обидва профілі демонструють періодичну структуру з мінімумом у нічні години та зростанням електроспоживання протягом дня. Незбалансований режим характеризується нижчим нічним рівнем та різкішим зростанням з ранку з досягненням суттєво більшого вечірнього максимуму. Збалансований режим має вищий стабільний нічний рівень та плавніший добовий профіль з меншою амплітудою піку.

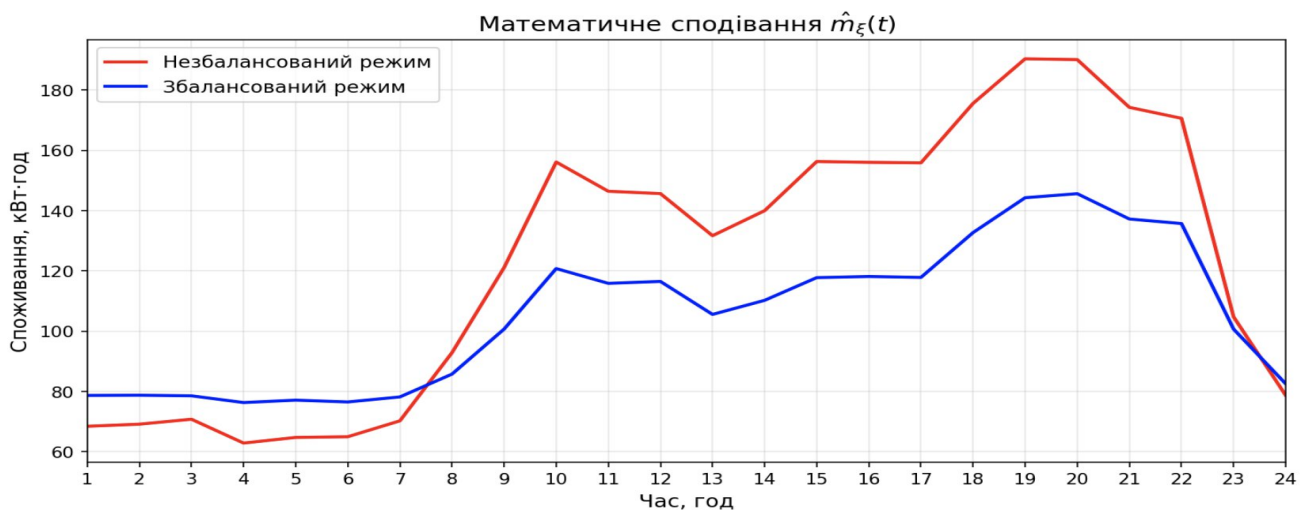


Рисунок 2.15 – Графік статистичної оцінки математичного сподівання сигналу електроспоживання для збалансованого та незбалансованого режиму

Оцінку дисперсії  $\hat{D}_\xi(t)$  наведено на рисунку 2.16. При збалансованому режимі дисперсія залишається рівномірно малою протягом усього добового циклу, що свідчить про стабільну повторюваність форми споживання від реалізації до реалізації. При незбалансованому режимі дисперсія суттєво перевищує значення збалансованого впродовж усієї доби та демонструє нерівномірний характер з вираженим піком у вечірні години. Залежність кореляційної функції від абсолютного моменту часу всередині періоду підтверджує періодичність статистичних характеристик сигналу електроспоживання та правомірність його моделювання у класі ПКВП [49, 50, 51].

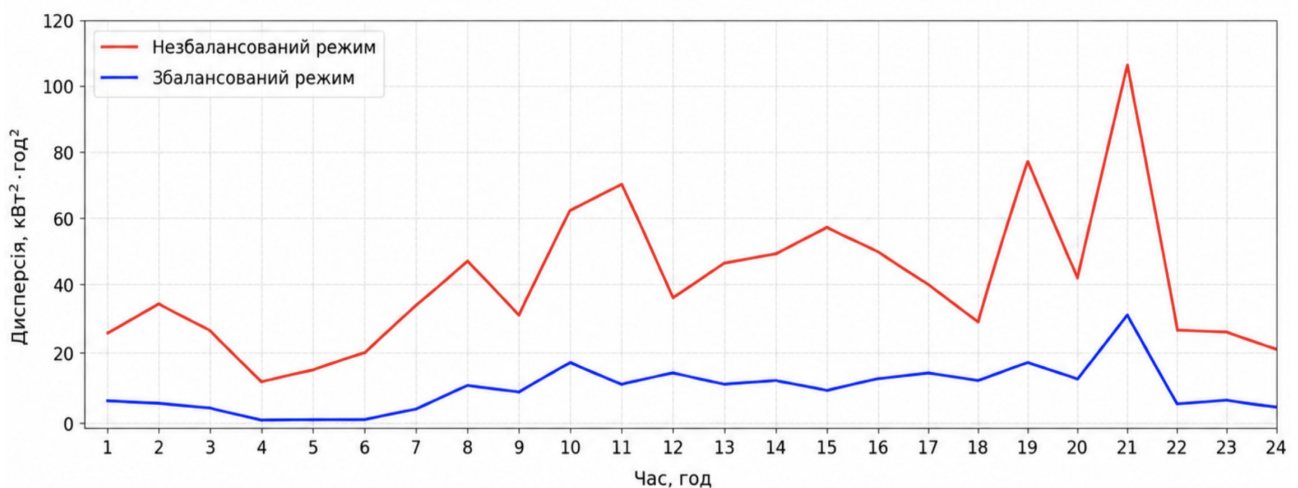


Рисунок 2.16 – Графік статистичної оцінки дисперсії сигналу електроспоживання для збалансованого та незбалансованого режиму

ПКВП дозволяє представити нестационарний процес через еквівалентні стаціонарні структури, що узгоджується з фізичною природою енергетичних процесів та дозволяє формалізувати їх нестационарність без втрати інформації про регулярну повторюваність. На відміну від стаціонарних випадкових процесів, для ПКВП повторюваність проявляється не в самих реалізаціях сигналу, а в періодичній зміні його ймовірнісних характеристик [43, 51, 83].

Процес електроспоживання характеризується двома визначальними властивостями, що обумовлюють вибір ПКВП як базової математичної моделі: періодичною змінністю моментних характеристик та залежністю кореляційної структури від абсолютного моменту часу всередині періоду [49, 50, 51]. Застосування цієї моделі дозволяє сформулювати набір інформативних характеристик, які містять інформацію про особливості зміни параметрів електроспоживання в часі [46, 47].

Зокрема, періодичне математичне сподівання фіксує базовий профіль електроспоживання, періодична дисперсія та коефіцієнт варіації відображають динаміку стохастичної мінливості, а кореляційні компоненти розкривають внутрішню інерційну структуру процесу. У межах моделі ПКВП сигнал електроспоживання  $\xi(t)$  представляється у вигляді суми стаціонарних компонент із періодичною змінною:

$$\xi(t) = \sum_{k \in \mathbf{Z}} \xi_k(t) e^{i \frac{2\pi k}{T_k} t}, \quad t \in \mathbf{R}, \quad k \in \mathbf{N}, \quad (2.9)$$

де  $\xi_k(t)$  стохастична складова структури сигналу електроспоживання, представлена як стаціонарно корельовані процеси (стаціонарні компоненти)  $e^{i \frac{2\pi k}{T_k} t}$  періодична (циклічна) складова сигналу електроспоживання з параметром добового періоду  $T_k = 24$  години;  $k$  – номер стаціонарної компоненти.

Таке подання дозволяє інтерпретувати електроспоживання як стохастичні коливання з повторюваною структурою, де регулярність проявляється у періодичності статистичних характеристик.

Отже, в рамках енергетичної теорії стохастичних сигналів [47, 48] процес електроспоживання розглядається як періодично корельований випадковий процес, що характеризується періодично змінною кореляційною функцією та відповідною їй матрицею взаємних спектральних густин потужності. Такий

підхід дозволяє кількісно оцінити енергетичний внесок кожної гармонічної складової у загальну структуру сигналу та виявити періодичні закономірності, що є прихованими при використанні класичних стаціонарних моделей. В свою чергу, наведені результати слугують теоретичним підґрунтям для подальшого застосування ПКВП-апарату в завданнях аналізу електроспоживання.

Зокрема, вони підтверджують, що добова ритмічність і стохастична мінливість параметрів циклу є не взаємовиключними, а взаємодоповнювальними властивостями одного і того ж сигналу, моделювання яких в єдиних теоретичних рамках є необхідною умовою адекватного аналізу режимів електроспоживання електроенергетичної системи.

#### **2.4. Синфазний метод аналізу сигналів електроспоживання на основі моделі ПКВП**

Методи аналізу періодично корельованих випадкових процесів можна класифікувати за способом представлення процесу на синфазні, компонентні, кореляційні та спектральні підходи [43, 50].

Для дослідження структури сигналу та виявлення міжциклових кореляційних залежностей у даному підрозділі застосовується синфазний метод аналізу періодично корельованих стохастичних процесів, який дозволяє перейти від аналізу усереднених характеристик до дослідження узгодженості фаз добових реалізацій та їх взаємозв'язків у часі.

Синфазний (когерентний) метод ПКВП базується на чотирьох послідовних операціях: центруванні сигналу, формуванні синфазних вибірок шляхом групування відліків однакових фаз різних циклів, обчисленні оцінки коваріаційної функції  $\hat{b}_\xi(t, u)$  та перетворенні Фур'є для отримання кореляційних компонент  $\hat{B}_k(u)$  [99, 100]. На відміну від компонентного методу, оцінювання  $\hat{b}_\xi(t, u)$  формується не з усього часового запису, а з синфазних вибірок окремих циклів.

Суть даного методу полягає у поданні вихідного випадкового процесу у вигляді сукупності реалізацій, синхронізованих за фазою періоду. Для цього часову вісь розбивають на інтервали довжиною, що дорівнює періоду  $T_k$ , після чого значення процесу, що відповідають однаковим фазовим моментам у різних періодах, об'єднуються в одну вибірку. Таким чином формується фазове представлення процесу, у якому статистичні характеристики оцінюються окремо для кожної фази.

У межах синфазного підходу оцінювання математичного сподівання здійснюється як середнє значення по всіх реалізаціях, що відповідають фіксованій фазі. Для кожного моменту часу  $t$  всередині базового циклу формується вибірка значень з  $N$  різних реалізацій циклу, після чого обчислюється середнє значення цієї вибірки як оцінка математичного сподівання  $\hat{m}_\xi(t)$ . Аналогічно визначаються інші статистичні характеристики, зокрема дисперсія та кореляційна функція, які при цьому набувають періодичної залежності від часу.

Ключовою особливістю синфазного методу є можливість врахування або ігнорування взаємкореляційних зв'язків між гармонічними компонентами. Врахування таких зв'язків дозволяє виявляти приховані закономірності структури процесу, тоді як їх ігнорування забезпечує отримання згладжених оцінок основних тенденцій.

Оцінка математичного сподівання сигналу електроспоживання визначається за виразом:

$$\hat{m}_\xi(t) = \frac{1}{N} \sum_{k=0}^{N-1} \xi(t + kT_k), \quad (2.10)$$

Математичне сподівання відображає середній рівень електроспоживання для кожної фази добового циклу та характеризує детерміновану складову процесу.

На основі отриманої оцінки виконується центрування сигналу – вилучення детермінованої складової з метою виділення стохастичних відхилень від типового добового профілю:

$$\overset{0}{\xi}(t) = \xi(t) - \widehat{m}_{\xi}(t), \quad (2.11)$$

Центрування дозволяє усунути періодичну середню складову сигналу та виділити випадкові відхилення від типового добового профілю електроспоживання. Оцінка коваріаційної функції визначається за виразом:

$$\widehat{b}_{\xi}(t, u) = \frac{1}{N} \sum_{k=0}^{N-1} \overset{0}{\xi}(t + u + kT_k) \overset{0}{\xi}(t + kT_k), \quad (2.12)$$

де  $k$  – номер циклу (реалізації);  $u$  – часовий зсув;  $N$  – кількість реалізацій циклу;  $T_k = 24$  год.

На основі отриманої оцінки коваріаційної функції обчислюються кореляційні компоненти шляхом перетворення Фур'є:

$$\widehat{B}_k(u) = \frac{1}{T} \int_0^T \widehat{b}_{\xi}(t, u) \exp\left(-ik \frac{2\pi}{T_k} t\right) dt, \quad (2.13)$$

де  $k$  – номер гармоніки;  $T_k = 24$  год добовий період.

Оцінки  $\widehat{B}_k(u)$  є незміщеними та асимптотично ефективними, що забезпечує статистичну надійність подальшого аналізу [103, 104].

У таких умовах методи, що ґрунтуються виключно на часовому або компонентному усередненні, не дають змогу повністю описати внутрішню структуру процесу, оскільки ігнорують інформацію про фазову узгодженість

між циклами. Синфазний метод, на відміну від класичних підходів, забезпечує аналіз сигналу в координатах фаз базового циклу, що дозволяє досліджувати взаємозв'язки між реалізаціями на однакових фазових інтервалах доби.

Для дослідження структури сигналу застосовується синфазний метод аналізу ПКВП, який дозволяє досліджувати процес в координатах фаз базового циклу та забезпечує оцінювання коваріаційної функції. На рисунку 2.17 наведено алгоритм реалізації синфазного методу аналізу періодично корельованого випадкового процесу.

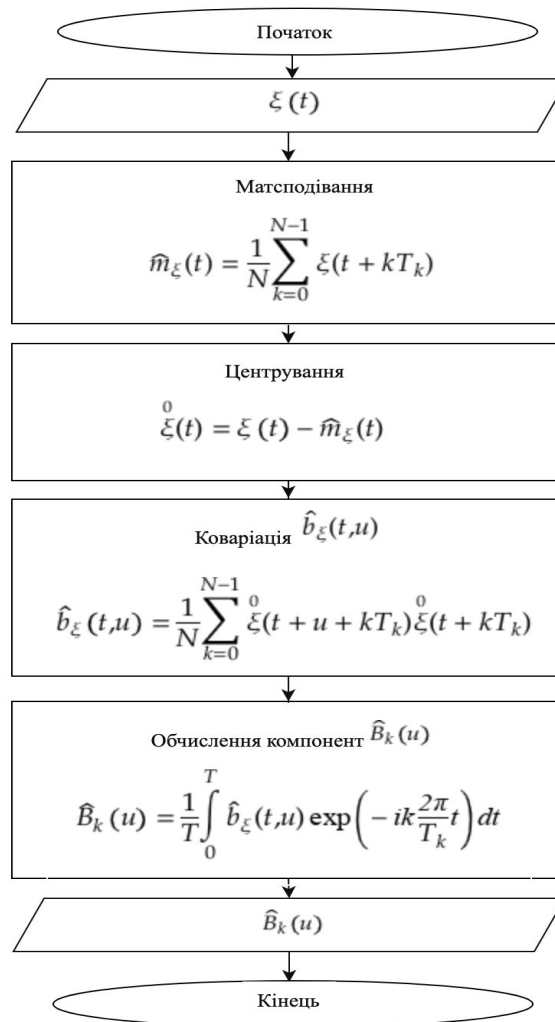


Рисунок 2.17 – Алгоритм реалізації синфазного методу аналізу ПКВП

На першому етапі сигнал електроспоживання подається у вигляді періодично корельованого випадкового процесу  $\xi(t)$ . Далі обчислюється

математичне сподівання  $\hat{m}_\xi(t)$  яке характеризує усереднений рівень електроспоживання в межах періоду спостереження. Після цього визначається оцінка коваріаційної функції  $\hat{b}_\xi(t, u)$  – що відображає взаємозв'язок між значеннями сигналу в різні моменти часу.

На основі отриманої оцінки коваріаційної функції обчислюються компоненти  $\hat{B}_k(u)$  шляхом спектрального перетворення. Отримані компоненти характеризують циклічну структуру електроспоживання та використовуються для формування інформативних ознак і подальшого аналізу режимів роботи електроенергетичної системи.

Теоретичне обґрунтування ефективності синфазного методу підтверджується його здатністю оптимально оцінювати параметри. Синфазний метод забезпечує статистичне оцінювання кореляційних компонентів центрованого сигналу електроспоживання на основі параметричних коваріаційних статистик, що дозволяє отримати інформативний опис структури процесу для кожної фази добового циклу. Це створює передумови для більш точного аналізу структури електроспоживання та підвищення якості класифікації режимів електроспоживання в розподілених електроенергетичних системах.

Процедура ідентифікації стану рівня електроспоживання енергетичної системи на основі кореляційних компонент використовує критерій, що ґрунтується на усередненні значень компонентів за часовими зсувами та компонентами. Оцінки кореляційних компонентів та їх усереднених реалізацій, отримані синфазним методом, є незміщеними та асимптотично ефективними [103, 104], що забезпечує високу точність моделювання та прогнозування. Це особливо важливо для розподілених електроенергетичних систем.

Статистика синфазного методу дозволяє виявляти приховані періодичні патерни в структурі електроспоживання, які не можуть бути отримані класичними методами аналізу часових рядів. Кореляційні компоненти передають інформацію про взаємозв'язки між різними фазами добового циклу.

Синфазний метод з урахуванням взаємкореляційних зв'язків дозволяє більш повно врахувати внутрішню структуру процесу електроспоживання, зокрема взаємозв'язки між різними фазами добового циклу. Результати застосування синфазного методу для збалансованого та незбалансованого режимів наведено на рис. 2.18–2.27.

Для підтвердження практичної застосовності запропонованого підходу на реальних даних різних об'єктів електроенергетичної інфраструктури проведено аналіз даних електроспоживання із застосуванням синфазного та компонентного методів ПКВП, результати наведено у додатку В. Для детальнішого аналізу просторово-часової структури процесу використано тривимірне представлення результатів синфазного методу.

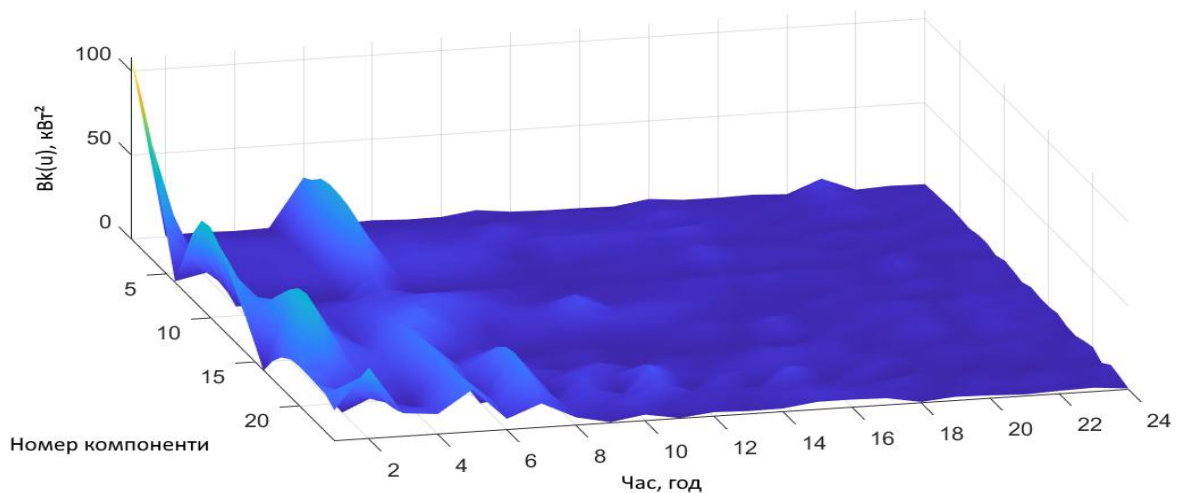


Рисунок 2.18 – Тривимірне представлення синфазного методу з урахуванням крос–кореляційних зв'язків збалансований режим (Клас 1)

На рис. 2.18 представлено тривимірне представлення синфазного методу з урахуванням крос-кореляційних зв'язків. Поверхня має виражену структуру із концентрацією високих значень у початкових компонентах та окремих часових інтервалах, що відображає наявність стійких кореляційних залежностей у процесі електроспоживання. На рис. 2.19 наведено тривимірне представлення синфазного методу без урахування крос-кореляційних зв'язків. Отримана

поверхня є більш згладженою та менш структурованою, що свідчить про втрату інформації про взаємозв'язки між компонентами процесу.

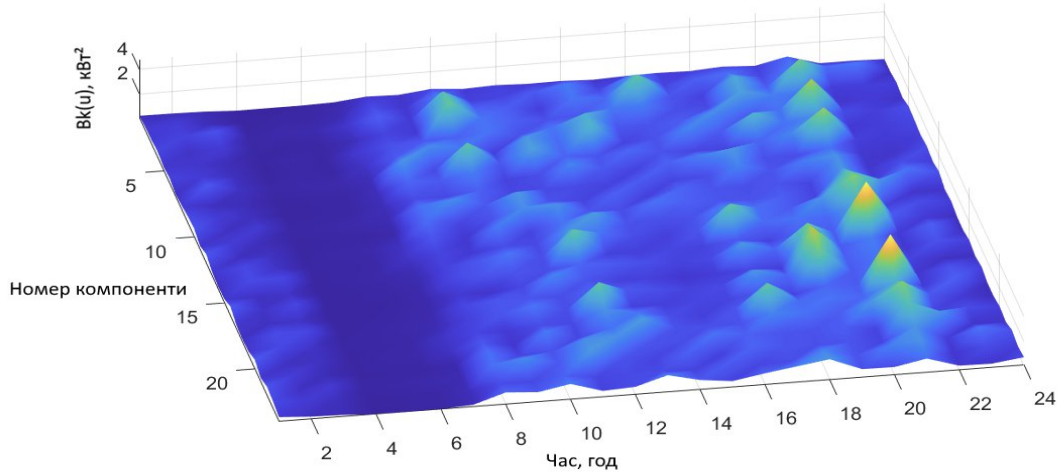


Рисунок 2.19 – Тривимірне представлення синфазного методу без урахування крос–кореляційних зв'язків збалансований режим (Клас 1)

Додаткове тривимірне представлення (рис. 2.20) підтверджує відсутність чітко вираженої структури та демонструє більш випадковий характер розподілу значень.

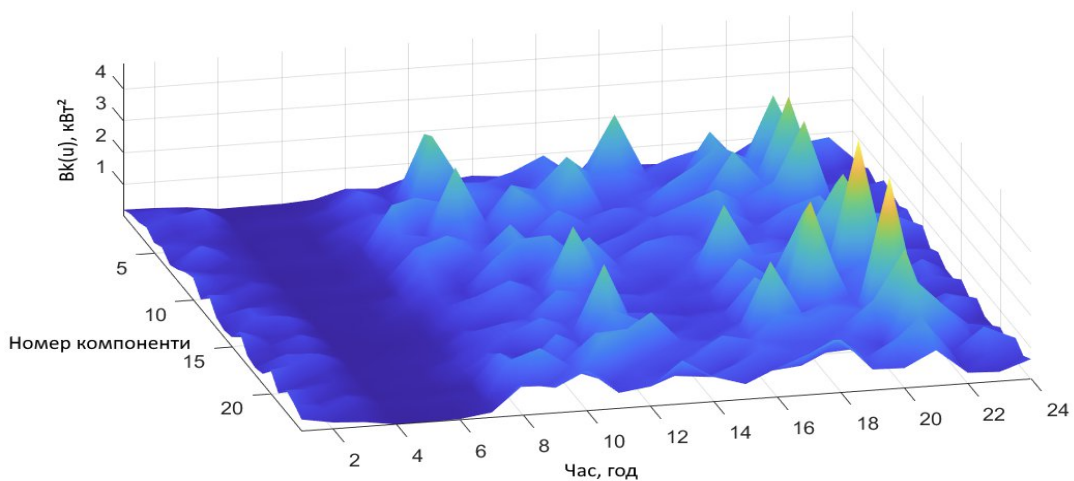


Рисунок 2.20 – Альтернативне тривимірне представлення синфазного методу без урахування крос–кореляційних зв'язків збалансований режим (Клас 1)

Аналіз тривимірних представлень компонентних коефіцієнтів виконується із застосуванням двох процедур когерентного усереднення. Перша процедура усереднення за компонентами дозволяє отримати узагальнену характеристику кореляційної структури як функцію часового зсуву. Для кожного значення зсуву  $u$  обчислюється середнє значення компонентних коефіцієнтів по всіх гармоніках:

$$M_u \{ \hat{B}_k(u) \} = \frac{1}{U_{max}} \sum_{u=1}^{U_{max}} \hat{B}_k(u), \quad u = \overline{1, U_{max}}, \quad k = \overline{1, K_{max}} \quad (2.14)$$

де  $k$  – номер кореляційної компоненти;

$u$  – часовий зсув;

$U_{max}$  – максимальна довжина часового зсуву;

$K_{max}$  – максимальна кількість компонент сигналу електроспоживання.

Результат усереднення за компонентами показує, як змінюється середня інтенсивність кореляційних зв'язків зі збільшенням часового зсуву. Для збалансованого стану ця функція має квазістаціонарний характер з незначними флуктуаціями. Для незбалансованого спостерігається виражене монотонне зростання, що свідчить про наявність сильних довгострокових кореляцій. Друга процедура усереднення за часовим зсувом характеризує середній вклад кожної гармоніки по всіх часових зсувах:

$$M_k \{ \hat{B}_k(u) \} = \frac{1}{K_{max}} \sum_{u=1}^{K_{max}} \hat{B}_k(u), \quad u = \overline{1, U_{max}}, \quad k = \overline{1, K_{max}} \quad (2.15)$$

де  $k$  – номер кореляційної компоненти;

$u$  – часовий зсув;

$U_{max}$  – максимальна довжина часового зсуву;

$K_{max}$  – максимальна кількість компонент сигналу електроспоживання.

Результат усереднення за часовим зсувом дозволяє ідентифікувати домінуючі гармоніки циклу. Для збалансованого стану розподіл вкладів за

компонентами є відносно рівномірним. Для незбалансованого виділяються окремі домінуючі гармоніки, пов'язані з піковими режимами споживання.

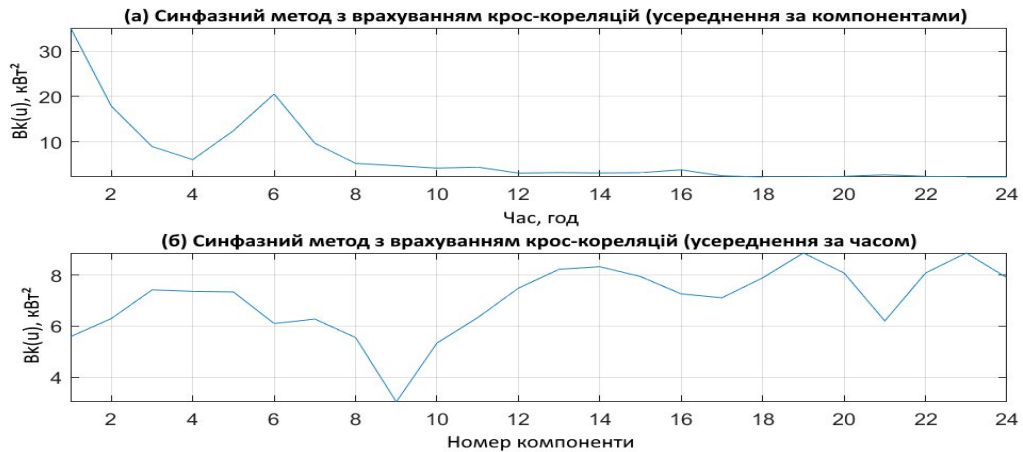


Рисунок 2.21 – Синфазний метод з урахуванням крос–кореляцій збалансований режим (Клас 1):

- а) усереднення за компонентами;
- б) усереднення за часовим зсувом

На рисунку 2.21 представлено результати синфазного методу з урахуванням крос-кореляційних зв'язків при усередненні за компонентами та за часом для збалансованого режиму.

Залежність кореляційної функції (рис. 2.21а) характеризується спаданням зі збільшенням часового зсуву та наявністю локальних максимумів що відповідають характерним фазам підвищеного електроспоживання – це свідчить про короткочасну пам'ять процесу та виражену добову ритмічність. Усереднення за часом (рис. 2.21б) виявляє поступове зростання значень по мірі збільшення номера компоненти з окремими локальними коливаннями що підтверджує розподіленість кореляційної енергії між гармонічними складовими та стабільність структури збалансованого режиму.

На рис. 2.22 наведено результати синфазного методу без урахування взаємкореляційних зв'язків. У цьому випадку залежності мають значно меншу

амплітуду та більш випадковий характер, що свідчить про втрату інформації щодо взаємозв'язків між компонентами процесу.

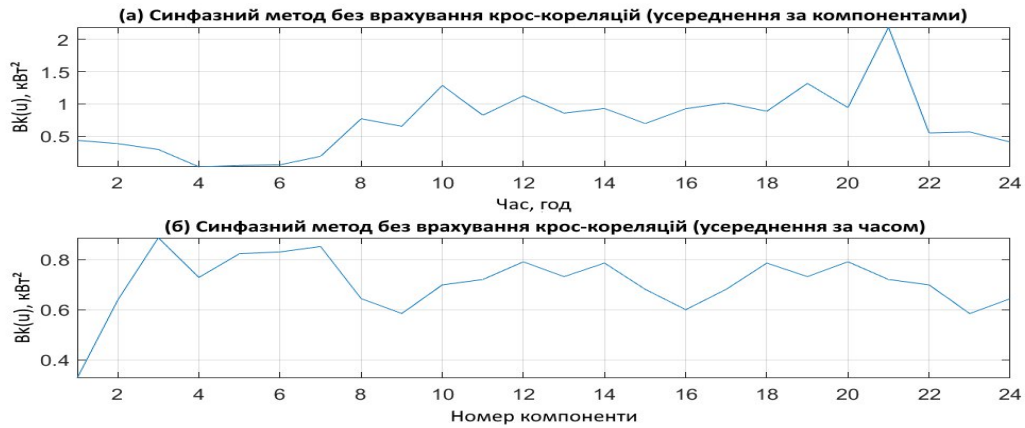


Рисунок 2.22 – Синфазний метод без урахуванням крос–кореляцій збалансований режим (Клас 1):

- а) усереднення за компонентами;
- б) усереднення за часовим зсувом

Графік усереднення за часом (рис. 2.22б) характеризується більш рівномірним, але менш інформативним розподілом значень, що підтверджує зниження ефективності аналізу без урахування крос-кореляцій.

На рис. 2.23 представлено тривимірне представлення з урахуванням крос-кореляційних зв'язків. Поверхня має складну структуру з наявністю локальних піків та різких змін, що відображає нестабільність процесу та значні коливання електроспоживання.

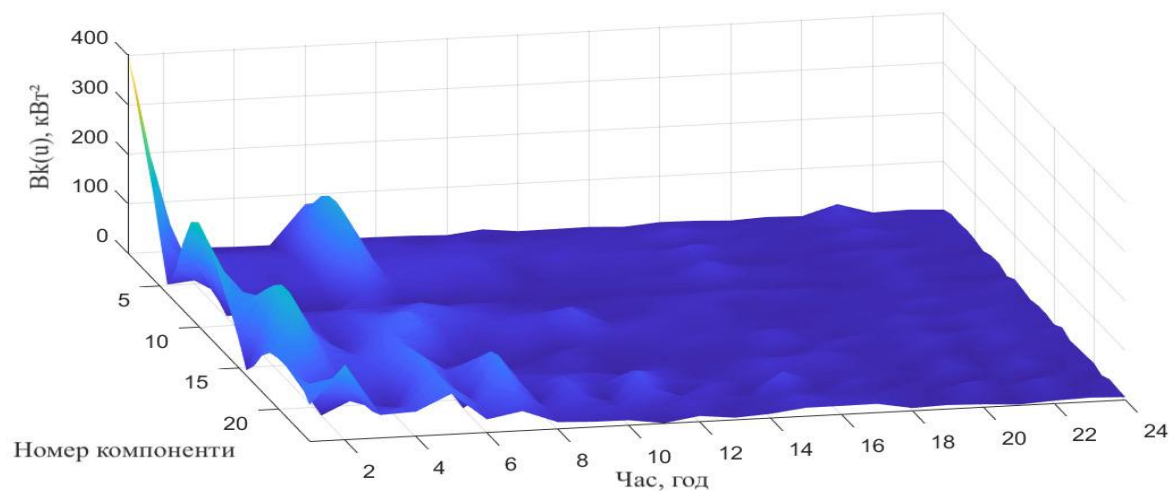


Рисунок 2.23 – Тривимірне представлення синфазного методу з урахуванням крос-кореляцій незбалансований режим (Клас 2)

На рис. 2.24 наведено тривимірне представлення без урахування крос-кореляцій. У цьому випадку поверхня є більш розмитою та характеризується відсутністю чітко виражених закономірностей.

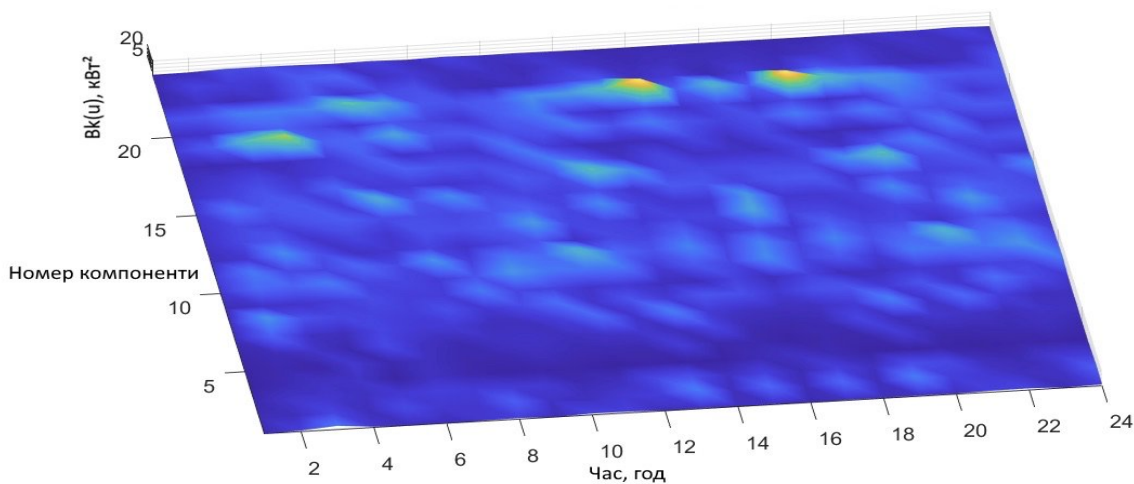


Рисунок 2.24 – Тривимірне представлення синфазного методу без урахування крос-кореляцій незбалансований режим (Клас 2)

Додаткове тривимірне представлення (рис. 2.25) демонструє значну кількість локальних максимумів, що підтверджує випадковий характер розподілу значень у незбалансованому режимі.

Тривимірні поверхні кореляційних компонент для незбалансованого стану демонструють суттєво відмінну картину порівняно зі збалансованим режимом. Кількісне оцінювання цих відмінностей здійснюється із застосуванням критеріїв когерентного усереднення, визначених виразами (2.14) та (2.15).

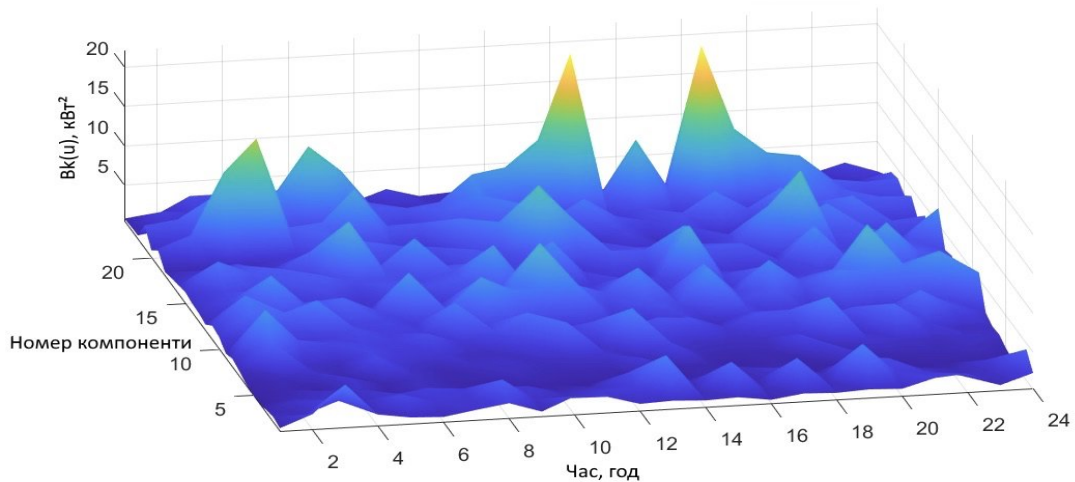


Рисунок 2.25 – Альтернативне тривимірне представлення синфазного методу без урахування крос–кореляцій незбалансований режим (Клас 2)

Натомість незбалансований режим характеризується підвищеною варіативністю, наявністю різких змін та меншою передбачуваністю. Врахування крос–кореляційних зв’язків дозволяє частково відновити структуру процесу, однак вона залишається значно менш впорядкованою порівняно зі збалансованим режимом.

На рис. 2.26 представлено результати синфазного аналізу з урахуванням крос–кореляційних зв’язків. Залежність кореляційної функції  $B_k(u)$  від часу (рис. 2.26а) характеризується більш різкими коливаннями та наявністю виражених локальних максимумів, що свідчить про нестабільність процесу електроспоживання.

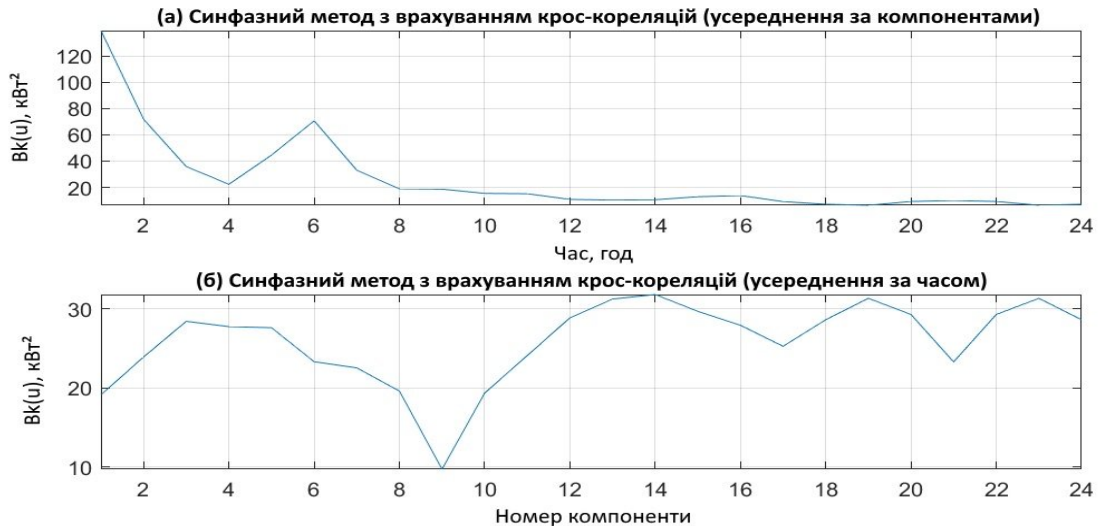


Рисунок 2.26 – Синфазний метод з урахуванням крос-кореляцій незбалансований режим (Клас 2):  
 а) усереднення за компонентами;  
 б) усереднення за часовим зсувом

Усереднення за часом (рис. 2.26б) демонструє нерівномірний розподіл енергії між компонентами, що вказує на наявність домінуючих складових та порушення збалансованості структури процесу.

На рис. 2.27 наведено результати синфазного методу без урахування крос-кореляційних зв'язків. У цьому випадку залежності мають ще більш випадковий характер, а амплітуди є значно нижчими, що свідчить про втрату інформації щодо взаємозв'язків між компонентами.

Графік усереднення за часом (рис. 2.27б) характеризується хаотичним розподілом значень без чітко вираженої структури. Порівняння результатів для збалансованого та незбалансованого режимів показує, що у збалансованому випадку спостерігається чітка детермінована повторюваність ймовірнісних характеристик, що свідчить про стабільність ритміки процесу.

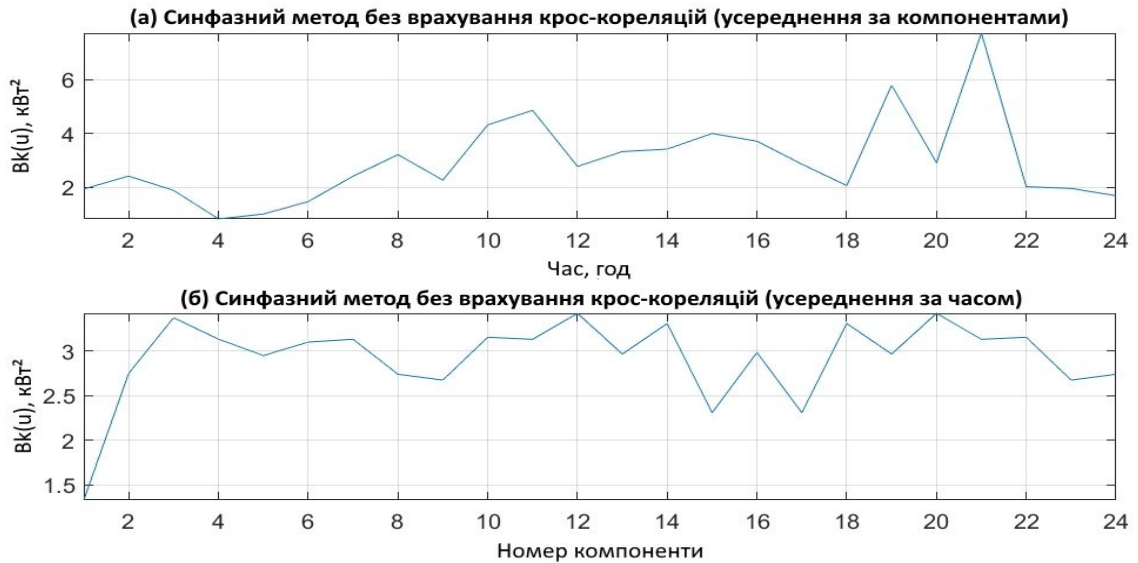


Рисунок 2.27 – Синфазний метод без врахування крос-кореляцій  
незбалансований режим (Клас 2):  
а) усереднення за компонентами;  
б) усереднення за часовим зсувом

Кореляційні функції в такому режимі мають сталу структуру в межах кожного циклу, що дозволяє з високою точністю ідентифікувати компоненти циклічного тренду. Натомість у незбалансованому режимі відбувається порушення фазової синхронності та спотворення ймовірнісної структури, що проявляється у підвищеній дисперсії та виникненні локальних аномалій, які важко піддаються апроксимації класичними методами.

Отже, синфазний метод дозволяє виявити відмінності у структурі електроспоживання для різних режимів та може бути використаний як інструмент дослідження та попереднього опрацювання даних. Синфазний метод забезпечує аналіз фазової структури сигналу електроспоживання та дозволяє виявляти міжциклові кореляційні залежності, однак має обмеження щодо обсягу необхідних даних та чутливості до пропусків у часових рядах

Характерною особливістю процесів електроспоживання є те, що за наявності стабільної добової періодичності параметри циклу амплітуда, форма та фазове положення не залишаються фіксованими від реалізації до реалізації.

Поведінкові чинники, вплив зовнішніх умов та інерційність енергетичної системи призводять до виникнення фазових зсувів між окремими добовими реалізаціями сигналу.

Подальший аналіз доцільно проводити із застосуванням методів, орієнтованих на узагальнення отриманих характеристик та побудову ефективних моделей. Отримані оцінки періодичних статистичних характеристик дають змогу перейти від аналізу безпосередніх часових реалізацій електроспоживання до аналізу їх структурних властивостей. Такий підхід забезпечує більш стійке представлення процесу в умовах шумів, пропусків даних та нестационарності.

Принципова відмінність синфазного методу полягає у фазовому синхруванні відліків сигналу в межах часового вікна, що забезпечує безпосереднє оцінювання ймовірнісних характеристик у часовій області. На противагу компонентному підходу, орієнтованому на частотну область, синфазний метод дозволяє уникнути втрати інформації про локальні часові особливості та нестационарні ефекти всередині періоду ритміки [102, 103].

Тому незважаючи на ефективність синфазного методу для аналізу структури електроспоживання, він має ряд обмежень, що необхідно враховувати при практичному застосуванні. Метод потребує достатньо тривалого інтервалу спостереження для отримання статистично спроможних оцінок. При добовому періоді необхідна вибірка, що охоплює щонайменше кілька десятків реалізацій, що обмежує його застосування в системах оперативного моніторингу.

Суттєвим обмеженням є також чутливість методу до точності визначення базового періоду  $T_k$  навіть незначна похибка у його оцінці призводить до десинхронізації фаз між реалізаціями та спотворення оцінок статистичних характеристик [105]. Особливою проблемою є відсутність безпосередньої спектральної інтерпретації отриманих характеристик, на відміну від компонентного методу, що працює у частотній області та дозволяє кількісно оцінити внесок кожної гармонічної складової у загальну структуру процесу.

Крім того, в умовах пропусків даних, спричинених нестабільністю комунікаційної інфраструктури, синфазне усереднення по фазових точках стає нерівномірним, що знижує точність оцінок та порушує умови їх незміщеності. Зазначені обмеження обумовлюють доцільність поєднання синфазного методу з компонентним підходом, що забезпечує взаємне доповнення аналізу у часовій та частотній областях відповідно для повноцінного дослідження структури процесу електроспоживання.

## **2.5. Компонентний метод ПКВП для аналізу сигналів електроспоживання**

Практична реалізація ПКВП пов'язана з оцінюванням характеристик різними методами: синфазним (когерентним), компонентним, методом найменших квадратів і лінійної фільтрації [106, 107, 108]. У даному підрозділі детально розглядається компонентний метод як основний інструмент формування ознакового простору для класифікації, який базується на розкладанні автоковаріаційної функції у ряди Фур'є та оцінюванні відповідних коефіцієнтів. Компонентний метод перемножує сигнал із його зсуненою копією та комплексною гармонікою базової частоти, що дозволяє виділяти окремі спектральні компоненти та визначати які гармоніки формують повторюваність сигналу електроспоживання. На відміну від синфазного підходу, компонентний метод використовує весь часовий запис реалізації [47, 105, 106, 107], що забезпечує вищу точність оцінювання за швидких змін коваріаційної структури всередині періоду та зумовлює його доцільність як основного інструменту формування ознакового простору для класифікації режимів електроспоживання розподілених електроенергетичних систем. Асимптотична незміщеність та узгодженість оцінок досягаються за умови ергодичності процесу та достатньої довжини вибірки [105, 107].

У завданнях аналізу електроенергетичних даних компонентний метод дозволяє ефективно моделювати стохастичну циклічність. Компонентний метод забезпечує точнішу оцінку кореляційної структури сигналу електроспоживання порівняно з традиційними стаціонарними моделями, що зумовлює його доцільність для класифікації режимів електроспоживання розподілених електроенергетичних систем.

Для ПКВП кореляційна структура є періодичною за аргументом  $t$  з періодом  $T_k$ , що відображає повторювану зміну статистичних властивостей процесу протягом добового інтервалу. Саме ця повторюваність є ключовою властивістю, яка відрізняє ПКВП від стаціонарних випадкових процесів та дозволяє застосувати спектральне розкладання.

На основі отриманої оцінки автокореляційної функції обчислюються оцінки кореляційних компонент шляхом перетворення Фур'є з базовою частотою  $\frac{2\pi}{T_k}$ .

$$\hat{B}_k(u) = \frac{1}{T} \int_0^T \xi^0\{t\} \xi^0(t+u) \exp\left(-ik \frac{2\pi}{T_k} t\right) dt, \quad (2.16)$$

де  $k$  – номер гармоніки;  $T_k = 24$  год – добовий період.

Оцінки кореляційних компонентів  $\hat{B}_k(u)$  є незміщеними та характеризують зміни у часовій структурі процесу електроспоживання [103, 104].

Фізичний зміст компонентів  $\hat{B}_k(u)$  полягає в тому, що вони характеризують вклад окремих циклічних складових у кореляційну структуру електроспоживання. Нульова компонента ( $k = 0$ ) відповідає стаціонарній частині кореляційної функції, яка описує усереднений рівень кореляції без урахування циклічних змін. Перша гармоніка ( $k = 1$ ) відображає основний добовий ритм споживання, тоді як вищі гармоніки описують більш локальні дестабілізуючі

чинники електроспоживання, пов'язані з піковими режимами, короткочасними сплесками споживання та нестабільністю генерації.

Алгоритм реалізації компонентного методу включає чотири послідовні етапи, визначення періоду  $T_k$  обчислення математичного сподівання  $\hat{m}_\xi(t)$ , центрування сигналу та обчислення кореляційних компонентів  $\hat{B}_k(u)$ . Узагальнену схему реалізації алгоритму наведено на рисунку 2.28.

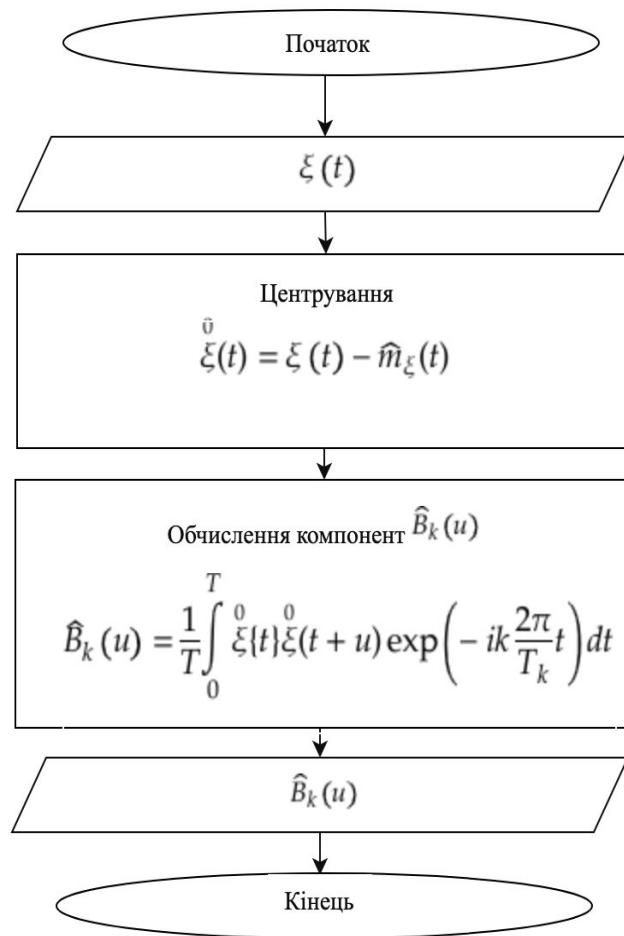


Рисунок 2.28 – Алгоритм реалізації компонентного методу аналізу ПКВП

Тим самим, компонентне розкладання забезпечує перехід від двовимірної функції  $\hat{b}_\xi(t,u)$ , що залежить від двох безперервних аргументів, до набору одновимірних функцій  $\hat{B}_k(u)$ , параметризованих дискретним індексом  $k$ . Це суттєво спрощує аналіз кореляційної структури та дозволяє ідентифікувати домінуючі циклічні складові. Безпосередній аналіз компонентів  $\hat{B}_k(u)$  є

складним завданням, оскільки вони залежать від номера компоненти  $k$  та часового зсуву  $u$ .

Для підвищення статистичної надійності оцінок, зменшення розмірності даних та виділення стійких закономірностей у роботі системи застосовано процедури когерентного усереднення.

Для сигналу електроспоживання як ПКВП компонентний метод із застосуванням когерентного усереднення забезпечує значне підвищення статистичної надійності оцінок. Математичні властивості таких процедур гарантують незміщеність оцінок (математичне сподівання оцінки дорівнює істинному значенню параметра), спроможність (збіжність оцінки до істинного значення при збільшенні обсягу вибірки) та ефективність у вигляді мінімальної дисперсії серед класу незміщених оцінок [105, 107, 108, 109].

Збалансований стан системи електроспоживання характеризується рівномірним розподілом електроспоживання протягом добового циклу, відсутністю різких пікових сплесків та стабільним балансом між генерацією та споживанням. Для даного класу було проведено компонентний аналіз із побудовою тривимірної поверхні компонентних коефіцієнтів та обчисленням результатів когерентного усереднення.

На рис. 2.29 представлено тривимірну поверхню компонентних коефіцієнтів  $\hat{B}_k(u)$  для збалансованого стану. Поверхня має характерну морфологію з переважно рівномірним розподілом енергії за компонентними та часовими зсувами у центральній частині.

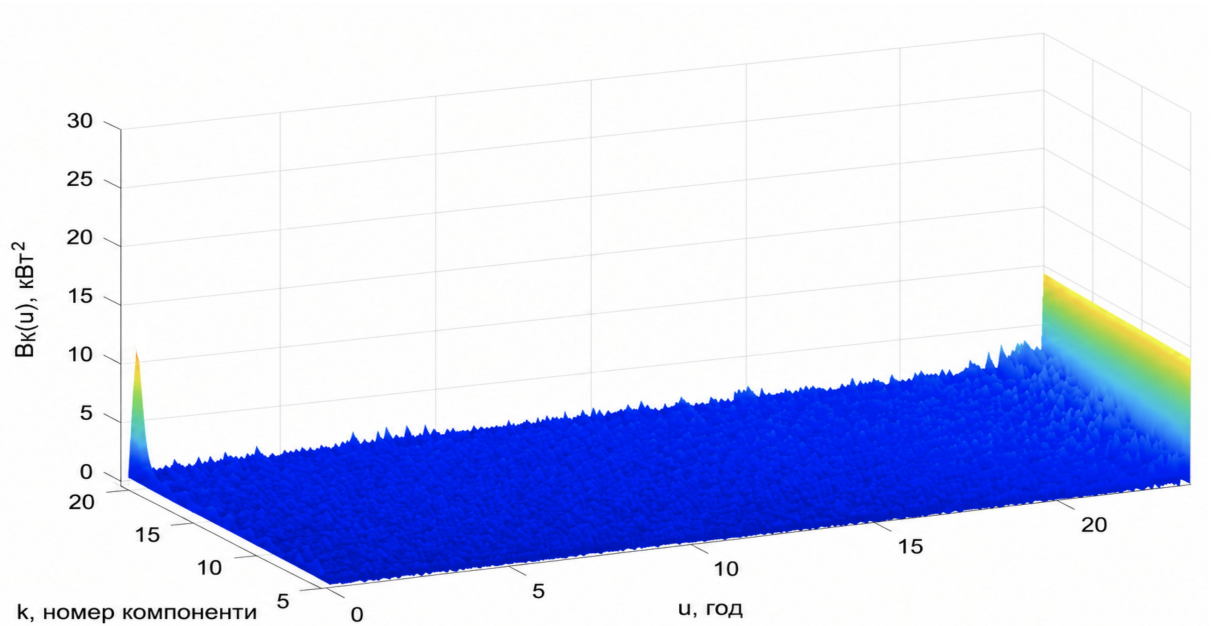


Рисунок 2.29 – Тривимірне представлення компонентного методу для збалансованого режиму електроспоживання (Клас 1)

Значення коефіцієнтів у зоні стабільності знаходяться на відносно низькому рівні, що свідчить про слабку циклічну модуляцію кореляційної структури. Виражений пік у крайній області часового зсуву (при великих значеннях  $u$ ) відповідає граничним ефектам оцінювання кореляційних функцій та є типовим артефактом кінцевої довжини реалізації.

Результати когерентного усереднення для збалансованого стану, отримані згідно з виразами (2.14) та (2.15), наведено на рис. 2.30. Графік усереднення за компонентами (рис. 2.30а) демонструє стабільний рівень кореляційних характеристик протягом більшої частини діапазону часових зсувів. Значення функції знаходяться переважно в межах  $0,5 - 1,0$  кВт<sup>2</sup> з незначними флуктуаціями, що підтверджує квазістаціонарний характер кореляційної структури збалансованого стану. Характерне зростання на великих зсувах ( $u > 600$  год) є типовим крайовим ефектом.

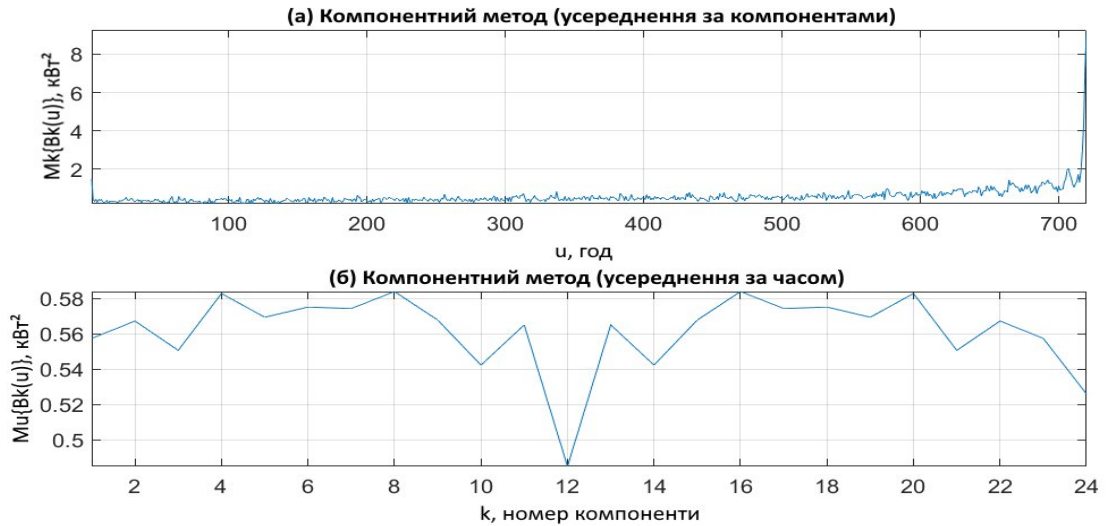


Рисунок 2.30 – Оцінки кореляційних компонент за компонентним методом для збалансованого режиму (Клас 1):  
 а) усереднення за компонентами;  
 б) усереднення за часовим зсувом

Графік усереднення за часовим зсувом (рис. 2.30б) показує рівень середнього вкладу кожної компоненти з відносно рівномірним розподілом по компонентах. Незначний провал вказує на локальне зменшення вкладу відповідної гармоніки, що може бути пов'язане з характерними добовими ритмами електроспоживання та 12-годинною періодичністю (два піки активності ранковий та вечірній). В цілому, рівномірність розподілу за компонентами є характерною ознакою збалансованого стану.

Незбалансований стан системи електроспоживання (Клас 2) характеризується нерівномірним розподілом електроспоживання, наявністю виражених пікових навантажень, значною міжцикловою варіативністю амплітуди сигналу та порушенням стабільності кореляційної структури добового профілю сигналу. Для даного класу компонентний аналіз демонструє суттєво відмінну картину порівняно зі збалансованим станом.

На рис. 2.31 представлено тривимірну поверхню компонентних коефіцієнтів для незбалансованого стану.

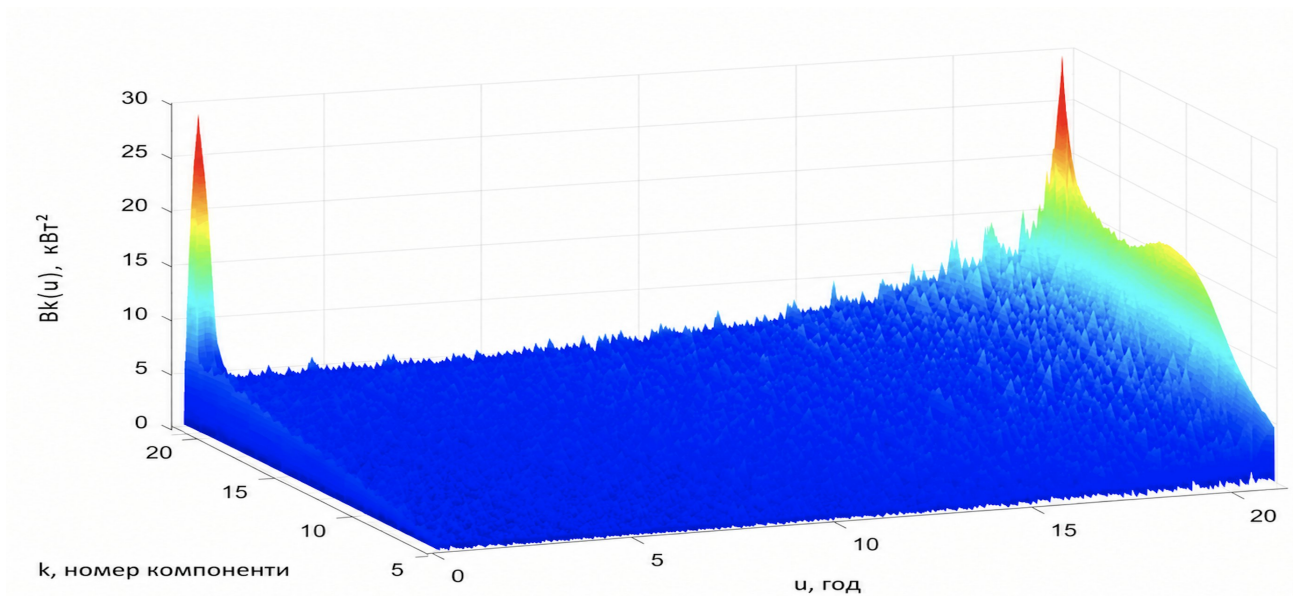


Рисунок 2.31 – Тривимірне представлення компонентного методу для незбалансованого режиму електроспоживання (Клас 2)

Для виявлення домінуючих гармонік та оцінювання довгострокових кореляцій незбалансованого стану застосовано процедури когерентного усереднення згідно з виразами (2.14) та (2.15), результати яких наведено на рис. 2.32 для незбалансованого стану.

Графік усереднення за компонентами (рис. 2.32а) виявляє принципово іншу поведінку порівняно зі збалансованим станом, спостерігається виражене монотонне зростання функції при збільшенні часового зсуву. Це свідчить про наявність сильних довгострокових кореляцій, характерних для незбалансованого режиму, коли піки електроспоживання мають стійкий вплив на подальшу динаміку системи та не затухають протягом значних часових інтервалів.

Графік усереднення за часовим зсувом (рис. 2.32б) показує середній рівень вкладу компонентів у діапазоні, що перевищує відповідні значення для збалансованого стану. Характерні піки вказують на домінування окремих гармонік циклу, пов'язаних із піковими режимами споживання [47, 110].

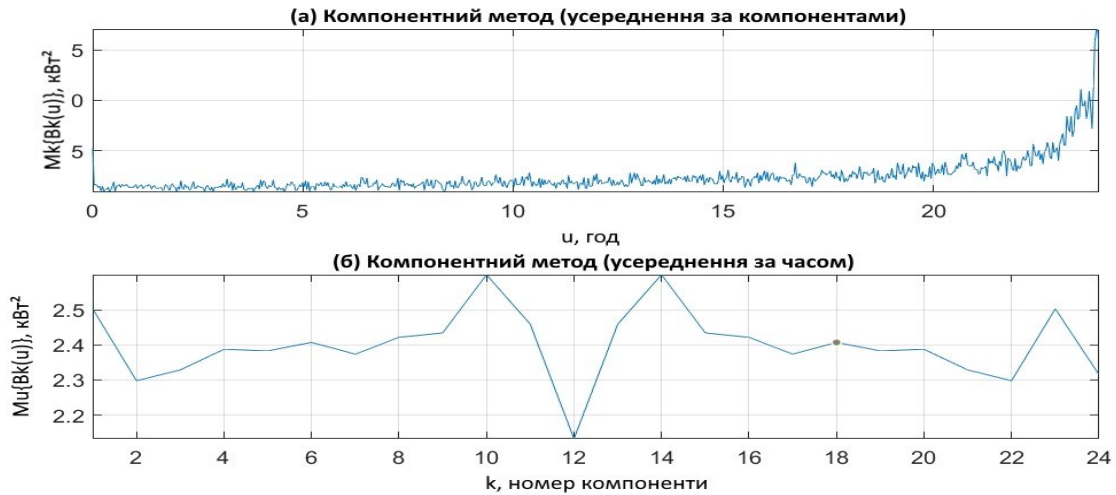


Рисунок 2.32 – Оцінки кореляційних компонент за компонентним методом для незбалансованого режиму (Клас 2):  
 а) усереднення за компонентами;  
 б) усереднення за часовим зсувом

На відміну від збалансованого стану, де розподіл за компонентами є рівномірним, тут спостерігається виражена нерівномірність із провалом та підйомом, що відображає складну структуру циклічних дестабілізуючих факторів.

Відповідно, теорія періодично корельованих випадкових процесів та компонентний метод аналізу формують математичний апарат для ідентифікації стохастичної структури електроенергетичних процесів і створюють методологічну основу для побудови систем моніторингу та класифікації режимів електроспоживання, поєднуючи із методами машинного навчання [47, 50, 106].

Додатково проведено аналіз даних електроспоживання трьох районів Тернопільської області із застосуванням синфазного та компонентного методів ПКВП з метою верифікації запропонованого підходу на реальних даних вузлів розподіленої електроенергетичної системи. Результати аналізу наведено у додатку В. Зокрема, використання компонентного методу аналізу ПКВП на етапі попереднього опрацювання даних (feature engineering) дозволяє сформувати компактний та інформативний вектор статистичних ознак, це сприяє

підвищенню точності та обчислювальної ефективності алгоритмів класифікації стану системи.

## 2.6. Висновки до розділу 2

1. На основі аналізу експериментальних даних встановлено, що процес електроспоживання є нестационарним, проте характеризується подвійною природою одночасно повторюваною (циклічною) за структурою та стохастичною за характером реалізації кожного окремого циклу, що зумовлює необхідність застосування математичних моделей здатних одночасно враховувати повторюваність і стохастичну мінливість сигналу електроспоживання.

2. Проаналізовано можливість застосування моделі циклічного випадкового процесу до опису сигналів з регулярним та нерегулярним ритмом розгортання циклів у часі. Встановлено, що використання моделі циклічного випадкового процесу є недоцільним, оскільки базовий добовий період є детермінованою величиною прив'язаною до усталених ритмів організації виробничої і побутової діяльності, що усуває необхідність ідентифікації функції ритму.

3. Обґрунтовано доцільність використання математичного апарату періодично корельованих випадкових процесів як базової моделі електроспоживання, яка забезпечує оптимальне співвідношення між адекватністю математичної моделі та практичною реалізованістю її методів аналізу. Повторюваність безпосередньо інтегрована у кореляційну структуру моделі через умову періодичності статистичних характеристик, що дозволяє формалізувати нестационарність процесу без надлишкової параметризації та мінімізація використання обчислювальних ресурсів.

4. Проведено аналіз даних електроспоживання синфазним методом оцінювання характеристик ПКВП та встановлено, що він дозволяє досліджувати

фазову структуру процесу та виявляти міжциклові кореляційні залежності. Водночас визначено, що метод має обмеження, пов'язані з чутливістю до фазових зсувів, необхідністю великого обсягу даних та обмеженим урахуванням міжфазових залежностей.

5. Проведено оцінювання статистичних характеристик процесу електроспоживання, зокрема математичного сподівання, дисперсії, моментів вищих порядків та кореляційних функцій, і встановлено, що їх періодична змінність підтверджує фазову залежність стохастичних властивостей сигналу та наявність міжциклових залежностей.

6. Порівняльний аналіз результатів для збалансованого та незбалансованого режимів електроспоживання підтвердив, що збалансований стан характеризується впорядкованою структурою кореляційних характеристик, тоді як незбалансований режим має підвищену варіативність та наявність локальних аномалій.

7. Обґрунтовано доцільність застосування компонентного методу аналізу ПКВП для формування інформативних ознак, що базується на розкладанні автоковаріаційної функції у ряди Фур'є та забезпечує повне врахування кореляційної структури процесу і ідентифікацію домінуючих циклічних складових.

8. Обґрунтовано доцільність використання характеристик, отриманих компонентним методом ПКВП, як інформативних ознак для класифікації режимів електроспоживання вузлів розподіленої електроенергетичної системи методами машинного навчання, що забезпечує перехід від аналізу часових реалізацій до структурованого представлення даних електроспоживання.

### **РОЗДІЛ 3. КЛАСИФІКАЦІЯ РЕЖИМІВ ЕЛЕКТРОСПОЖИВАННЯ В РОЗПОДІЛЕНИХ ЕЛЕКТРОЕНЕРГЕТИЧНИХ СИСТЕМАХ МЕТОДАМИ МАШИННОГО НАВЧАННЯ НА ОСНОВІ ІНФОРМАТИВНИХ ОЗНАК ОТРИМАНИХ КОМПОНЕНТНИМ МЕТОДОМ**

#### **3.1. Формування інформативних ознак на основі компонентного методу аналізу ПКВП**

Для класифікації режимів електроспоживання електроенергосистеми доцільним є поєднання математичного апарату ПКВП та методів машинного навчання, здійснивши перехід від аналізу часових реалізацій електроспоживання до структурованого представлення у вигляді інформативних ознак, сформованих на основі компонентного методу ПКВП [35, 47, 58, 108].

У межах аналізу даних енергоспоживання, представленого через матриці компонентного методу, кожен добуву реалізацію доцільно розглядати як реалізацію випадкового процесу з накладеною детермінованою трендовою складовою. Такий підхід дозволяє формалізувати структуру даних у вигляді композиції гладкої апроксимуючої функції та випадкових відхилень, що відображають фізичну неоднорідність системи.

Застосування двох схем агрегації кореляційних компонент усереднення за часовим зсувом та усереднення за номером гармоніки дозволяє оцінити як часову динаміку сумарної енергії сигналу, так і внесок окремих гармонічних складових у загальну кореляційну структуру [47, 108], з одного боку, аналіз часової еволюції сумарної енергії, а з іншого дослідження просторового розподілу енергетичних внесків між компонентами. Сформована навчальна вибірка (датасет) відображає статистичну структуру процесу електроспоживання у вигляді набору числових характеристик для збалансованого та незбалансованого режимів [66]. Для автоматизованого визначення режиму електроспоживання

(збалансований та незбалансований) використано методику вилучення ознак з матриць, отриманих компонентним методом.

Формування інформативних ознак здійснюється на основі статистичного аналізу процесів електроспоживання та компонентного представлення періодично нестаціонарних сигналів. Виділення кореляційних компонентів і оцінювання їх часової структури базуються на положеннях теорії лінійних періодичних процесів [111]. Для аналізу сформованих ознак та побудови моделей класифікації застосовуються сучасні методи машинного навчання, зокрема ансамблеві алгоритми, які демонструють високу ефективність у завданнях класифікації на основі багатовимірних ознакових просторів [112]. Окрім класичних ML-підходів, для аналізу часових рядів електроспоживання використовуються сучасні неймережеві архітектури, зокрема, гібридні Transformer-моделі [113].

Простір ознак, сформований на основі компонентного методу періодично корельованих випадкових процесів, є інформативним для розділення режимів електроспоживання та забезпечує високу точність класифікації незалежно від вибору алгоритму машинного навчання. Кожен фрейм даних (одна доба) опрацьовувався за двома схемами агрегації: усереднення за компонентами та усереднення за часом.



Рисунок 3.1 – Розподіл класів у сформованому датасеті

Загальна вибірка становила 365 діб: 206 діб у збалансованому режимі (Клас 1) та 159 діб у незбалансованому режимі (Клас 2) (рис. 3.1). На рисунку 3.2 наведено порівняння усереднених значень для збалансованого (Клас 1) та незбалансованого (Клас 2) режимів електроспоживання в єдиних координатах, що дозволяє безпосередньо оцінити кількісну та якісну відмінність між режимами.

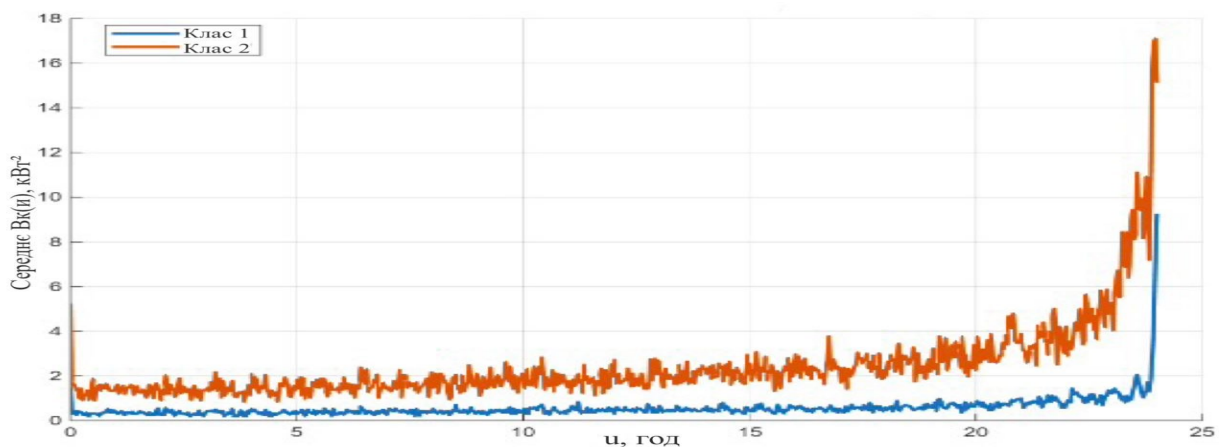


Рисунок 3.2 – Компонентний метод (усереднення за компонентами): для збалансованого (Клас 1) та незбалансованого (Клас 2) режимів

З рисунку 3.2 видно, що у діапазоні часових зсувів криві для двох режимів суттєво відрізняються за рівнем та характером поведінки. Для збалансованого режиму (Клас 1) функція залишається відносно стабільною на низькому рівні без вираженої тенденції до зростання, що свідчить про швидке згасання кореляційних зв'язків між відліками. Для незбалансованого режиму (Клас 2) спостерігається систематично вищий рівень функції з поступовим монотонним зростанням, що є ознакою накопичення довгострокових кореляційних залежностей, характерних для повторюваних аномальних навантажень.

Результатом компонентного методу є двовимірна матриця кореляційних компонент, де перший індекс відповідає номеру гармоніки, другий – часовому зсуву. Безпосереднє використання цієї матриці як вхідних даних для методів

машинного навчання є недоцільним через високу розмірність та відсутність інтерпретованості. Тому застосовуються дві процедури когерентного усереднення, які зводять двовимірний масив до двох одновимірних векторів, функції часового зсуву та функції номера компоненти. З кожного вектора виділяються числові характеристики, що формують вектор інформативних ознак.

Вектор функції часового зсуву характеризується зростаючим характером (рис. 3.2), що зумовлює вибір експоненціальної моделі для його апроксимації. Серед можливих монотонно зростаючих функцій експоненціальна модель є найпростішою двопараметричною, що адекватно описує спостережуваний характер кривої:

$$B(u) \approx ae^{bx}, \quad x \in [0,1], \quad (3.1)$$

де  $x \in [0,1]$  – нормований часовий зсув;  $a$  – початковий рівень кореляцій; параметр  $b$  – характеризує інтенсивність зростання кореляційних зв'язків, що безпосередньо характеризує режим електроспоживання. Вибір експоненційної моделі обґрунтований монотонним наростанням кореляцій з різким прискоренням у кінцевій частині інтервалу.

На основі результатів компонентного методу ПКВП сформовано ознаковий простір для класифікації режимів електроспоживання. За схемою усереднення за компонентами сформовано 9 ознак, за схемою усереднення за часовим зсувом 4 ознак, що формують вектор із 13 інформативних ознак. Такий підхід до багаторівневого формування інформативних компонентів узгоджується з сучасними концепціями інтелектуального аналізу в регіональних енергосистемах [47, 107].

Аналізуючи датасет, видно, що параметр  $a$  у збалансованому режимі (Клас 1) демонструє відносно вузький діапазон значень, що свідчить про стабільність початкового рівня сигналу. Така стабільність означає, що початковий стан

системи є близьким до рівноважного і не містить значних локальних перевантажень. Подібні закономірності статистичної поведінки є характерними для сучасних систем моніторингу енергетичних параметрів промислових підприємств [114].

У незбалансованому режимі (Клас 2), навпаки, параметр  $a$  характеризується значно більшою варіативністю: від надзвичайно малих величин, що фактично відповідають відсутності вираженого початкового рівня, до значень, що перевищують одиницю. Це свідчить про наявність сильних неоднорідностей у початковому розподілі енергії, коли окремі компоненти можуть домінувати або, навпаки, бути практично відсутніми [8].

Параметр  $b$ , який визначає швидкість експоненційного зростання, у даному датасеті виступає ключовим індикатором динамічної поведінки системи. Для збалансованого режиму його значення, як правило, лежать у межах  $1/4$ , що відповідає плавній, майже квазістаціонарній еволюції сигналу. Це означає, що система не демонструє різких змін у часі, а її енергетична структура змінюється поступово. У незбалансованому режимі спостерігається значно ширший спектр значень  $b$ , від дуже малих (близько  $(0,2)$ ), що відповідає майже стаціонарним сигналам, до надзвичайно великих (наприклад,  $b \approx 17$ , які відображають різкі експоненціальні сплески. Така поведінка свідчить про те, що незбалансований режим є внутрішньо неоднорідним і може включати як слабо змінні, так і різко зростаючі процеси. Як одну із метрик оцінювання якості моделі використовується коефіцієнт детермінації:

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}^{*i})^2}{\sum_{i=1}^N (y_i - \bar{y})^2}, \quad (3.2)$$

Аналіз коефіцієнта детермінації на основі датасету показує, що для обох класів значення  $R^2$  є відносно низькими (порядку  $10^{-3}$ - $10^{-1}$ ), що свідчить про обмежену здатність експоненціальної моделі повністю описати сигнал.

Це означає, що значна частина варіації сигналу обумовлена стохастичними компонентами. Водночас у незбалансованому режимі спостерігаються вищі значення  $R^2$ , що узгоджується з наявністю більш виражених трендів у частині сигналів. Середньоквадратична похибка представлених даних демонструє чітку залежність від класу і описується формулою:

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2}, \quad (3.3)$$

Для збалансованого режиму значення RMSE знаходяться приблизно на рівні (1,8–2,0), тоді як для незбалансованого режиму вони значно більші, досягаючи значень (7–9). Це прямо вказує на зростання амплітуди флуктуацій і, відповідно, на підвищену нестабільність енергетичного процесу. Особливої уваги заслуговує характеристика  $x_{45}$ , яка визначається з рівняння:

$$abe^{bx} = 1, \quad x_{45} = \frac{1}{b} \ln\left(\frac{1}{ab}\right), \quad (3.4)$$

Характеристика  $x_{45}$  відображає момент часу, коли швидкість зростання сигналу досягає критичного значення, де  $x_{45}$  позиція критичної точки максимальної швидкості зростання сигналу,  $N = 720$  загальна кількість відліків у межах одного періоду. У збалансованому режимі  $x_{45} \approx N$ , що означає відсутність різких змін протягом більшої частини інтервалу. У незбалансованому режимі  $x_{45} \ll N$  критична точка зміщується до початку інтервалу, що свідчить про раннє виникнення аномальних відхилень сигналу електроспоживання.

У незбалансованому режимі значення  $x_{45}$  можуть бути значно меншими (наприклад, 220 або навіть 1), що вказує на ранню появу різких змін і швидке наростання енергетичних піків. Після видалення експоненціального тренду стає можливим аналіз чисто стохастичної складової компонент:

$$y_{\text{det}}(x) = y(x) - ae^{bx}, \quad (3.5)$$

Стандартні відхилення до та після точки  $x_{45}$  демонструють суттєву різницю між режимами. У збалансованому випадку величини  $\sigma_{\text{before}}$  та  $\sigma_{\text{after}}$  є близькими, що підтверджує однорідність шуму. У незбалансованому режимі спостерігається різке зростання  $\sigma_{\text{after}}$ , що відображає посилення флуктуацій після досягнення критичної точки.

Коефіцієнти скошеності та ексцесу для ділянки після  $x_{45}$  дають змогу уточнити характер розподілу шуму. Для ілюстрації розподілу середніх значень компонентів електроспоживання при усередненні за часом розглянемо відповідні залежності для збалансованого та незбалансованого режимів (рис. 3.3).

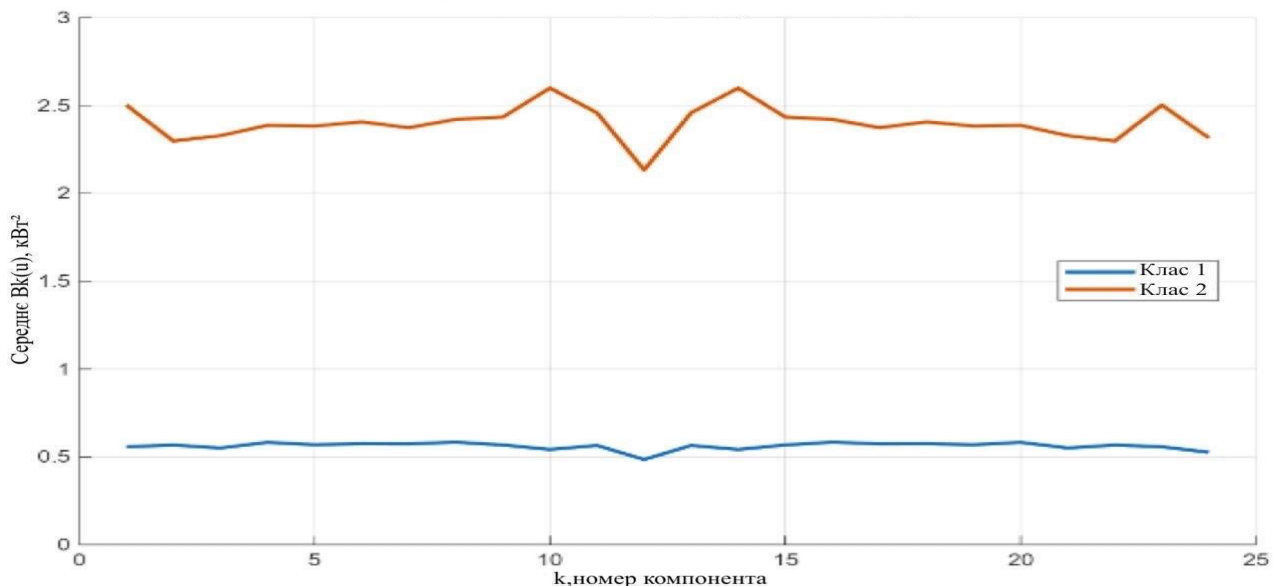


Рисунок 3.3 – Усереднені значення компонент електроспоживання для різних режимів (усереднення за часом)

У збалансованому режимі статистичні характеристики розподілу сигналу близькі до нормального розподілу, тоді як незбалансований режим

характеризується суттєвими відхиленнями від нього та підвищеною ймовірністю аномальних відхилень сигналу. У випадку усереднення за часом аналіз переходить у простір розподілу енергії між компонентами.

З рисунка 3.3 видно, що для незбалансованого режиму характерні значно вищі значення компонентів порівняно зі збалансованим, що свідчить про підвищений рівень електроспоживання та більшу варіативність процесу.

Крім того, у незбалансованому режимі спостерігаються локальні відхилення значень компонентів, що відображають наявність пікових навантажень та зміну режимів функціонування системи [107]. Для збалансованого режиму значення компонентів є більш стабільними та мають менший розкид, що відповідає рівномірному характеру електроспоживання. Аналіз усереднених компонентів підтверджує наявність суттєвих відмінностей між режимами та узгоджується з результатами, отриманими на основі статистичних ознак [108]. Середнє значення сигналу у незбалансованому режимі є систематично більшим, ніж у збалансованому, що відображає загальне підвищення рівня електроспоживання.

$$\mu = \frac{1}{N} \sum_{i=1}^N y_i \quad (3.6)$$

Стандартне відхилення також значно зростає (до значень близько 20), що вказує на сильну нерівномірність розподілу. Коефіцієнт асиметрії і коефіцієнт ексцесу у часово-усереднених даних демонструють більш складну поведінку. У збалансованому режимі вони залишаються відносно помірними, тоді як у незбалансованому режимі можливі як значні від'ємні значення скошеності, так і підвищений ексцес, що свідчить про асиметричний характер розподілу компонент з підвищеним ексцесом та наявністю екстремальних відхилень у пікові періоди електроспоживання.

На основі наведених міркувань сформовано набір з 13 ознак, 9 параметрів для випадку «усереднення за компонентами» та 4 для випадку «усереднення за часом». Детальний опис ознак наведено в таблиці 3.1.

Таблиця 3.1

## Перелік ознак компонентного методу та їх фізична інтерпретація

Ознака	Метод усереднення	Фізична інтерпретація
exp_a	По компонентах	Коефіцієнт амплітуди експоненціальної моделі, характеризує початковий рівень кореляційних зв'язків
exp_b	По компонентах	Показник інтенсивності експоненціального зростання або спадання кореляційних зв'язків
exp_rsquare	По компонентах	Коефіцієнт детермінації $R^2$ , якість апроксимації
exp_rmse	По компонентах	Середньоквадратична похибка між сигналом та моделлю
tangent_45	По компонентах	Індекс точки, де похідна експоненти = 1, момент різкого зростання
std_before_45	По компонентах	Стандартне відхилення сигналу до точки $x_{45}$
std_after_45	По компонентах	Стандартне відхилення сигналу після точки $x_{45}$
skew_after_45	По компонентах	Асиметрія розподілу залишків після точки $x_{45}$
kurt_after_45	По компонентах	Ексцес розподілу залишків після точки $x_{45}$
time_mean	По часу	Середнє значення компонент, відображає середній рівень електроспоживання
time_std	По часу	Стандартне відхилення, характеризує розкид значень компонент
time_skew	По часу	Асиметрія розподілу сигналу
time_kurt	По часу	Ексцес розподілу значень у часі

Таким чином, на основі компонентного методу періодично корельованих випадкових процесів сформовано набір із 13 інформативних ознак, який відображає як часову динаміку, так і просторові характеристики добових реалізацій електроспоживання.

### **3.2. Статистичний аналіз виділених ознак та їх дискримінантної здатності**

Сформований набір ознак є результатом переходу від часових реалізацій сигналу до його статистичного та структурного опису [115, 116], що дозволяє значно зменшити розмірність даних при збереженні їх інформативності. Для оцінювання дискримінантної здатності сформованого ознакового простору застосовано комплекс взаємодоповнюючих інструментів статистичного аналізу: оцінки густини розподілу на основі методу ядрового згладжування (KDE), діаграми розмаху та матрицю кореляцій Пірсона.

KDE-оцінки дозволяють кількісно та візуально оцінити ступінь перекриття розподілів класів C1 (збалансований) та C2 (незбалансований) для кожної ознаки окремо, що менше перекриття між розподілами, то вища індивідуальна дискримінантна здатність ознаки та її внесок у розділення класів. Діаграми розмаху доповнюють KDE-аналіз, надаючи порівняльну характеристику медіан, міжквартильних розмахів та викидів для кожного класу, що дозволяє виявити систематичні зсуви між класами та оцінити стійкість ознак до впливу аномальних спостережень.

Для дискримінативної здатності ознак побудовано графіки розподілів для кожної з 13 ознак окремо для збалансованого (C1) та незбалансованого (C2) класів (рис. 3.4).

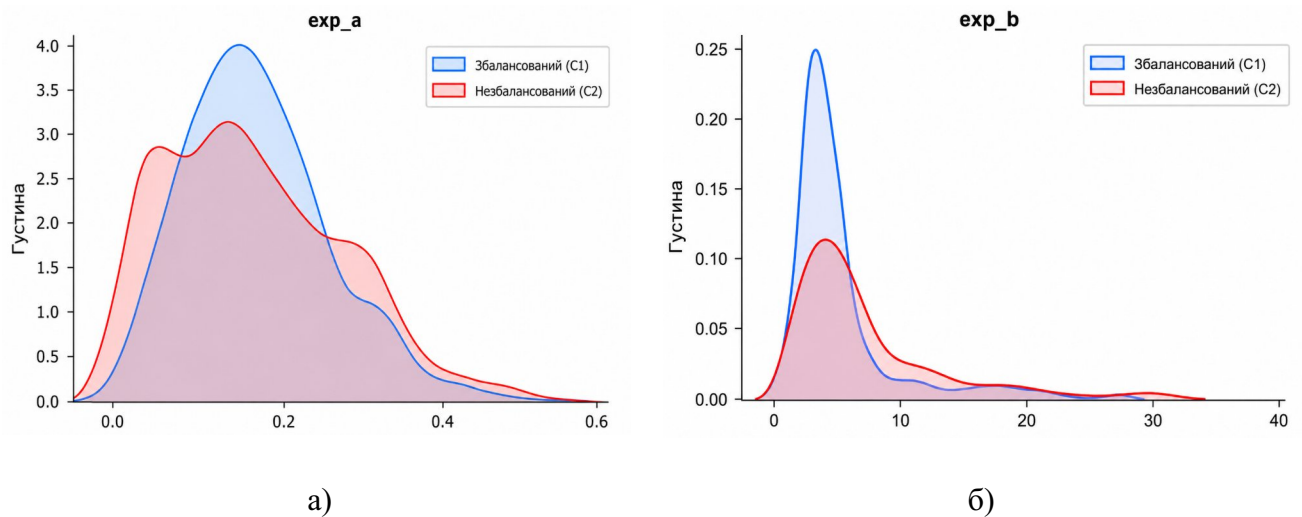


Рисунок 3.4 – Розподіли параметрів експоненціальної апроксимації:

а) exp\_a – початковий рівень кривої;

б) exp\_b – швидкість зростання

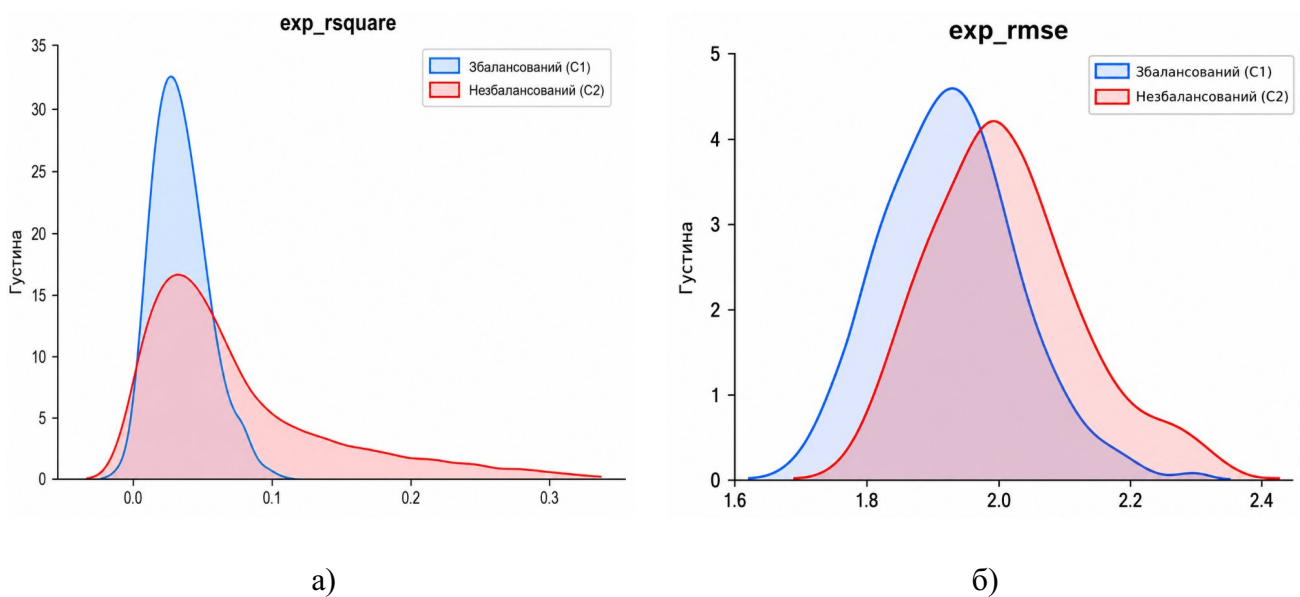


Рисунок 3.5 – Розподіли ознак якості апроксимації:

а) exp\_rsquare – наскільки добре модель описує криву;

б) exp\_rmse – похибка апроксимації

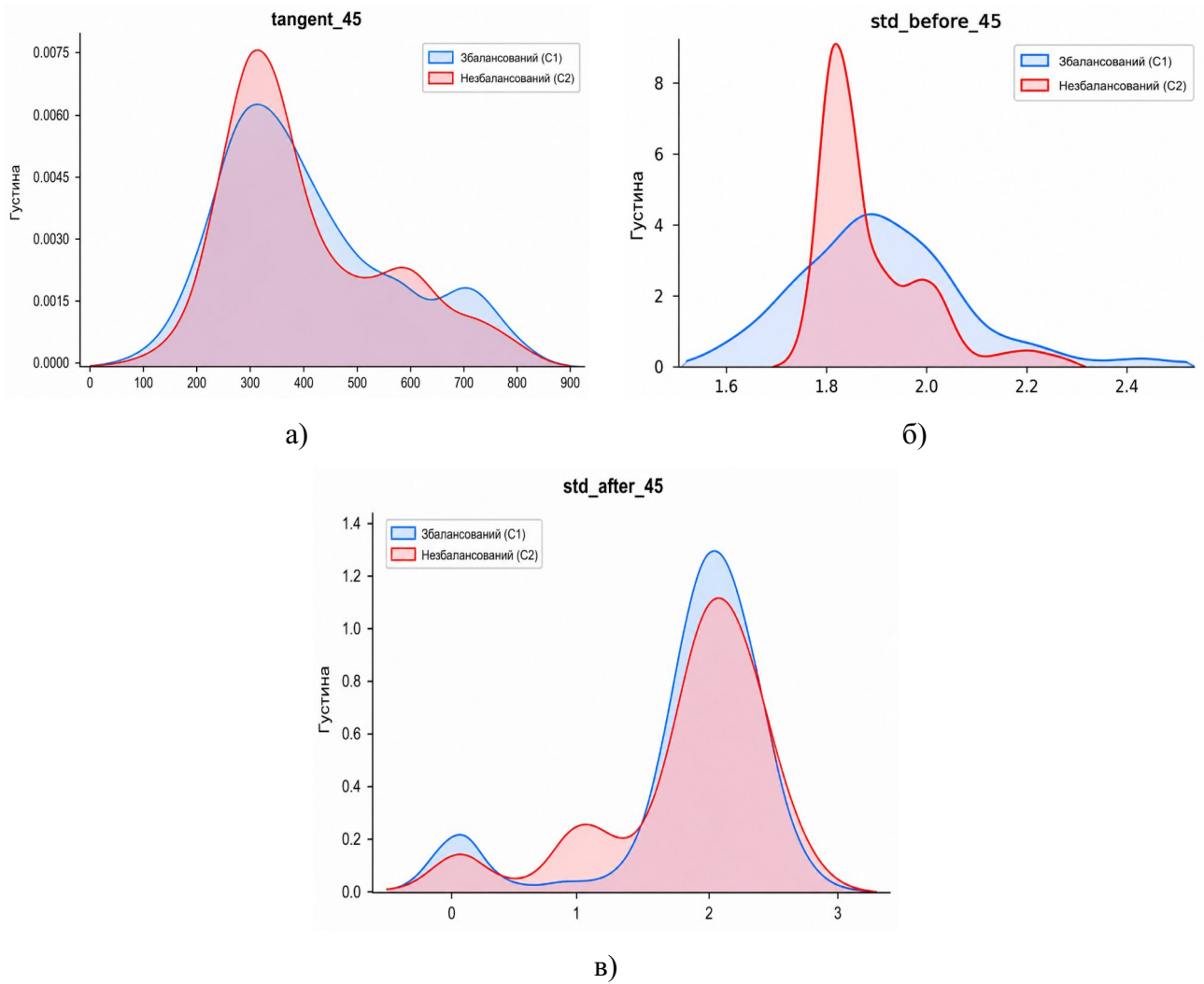


Рисунок 3.6 – Розподіли ознак точки критичного зростання та залишку:

а) `tangent_45` – точка де починається різке зростання;

б) `std_before_45` – розкид до цієї точки;

в) `std_after_45` – розкид після цієї точки

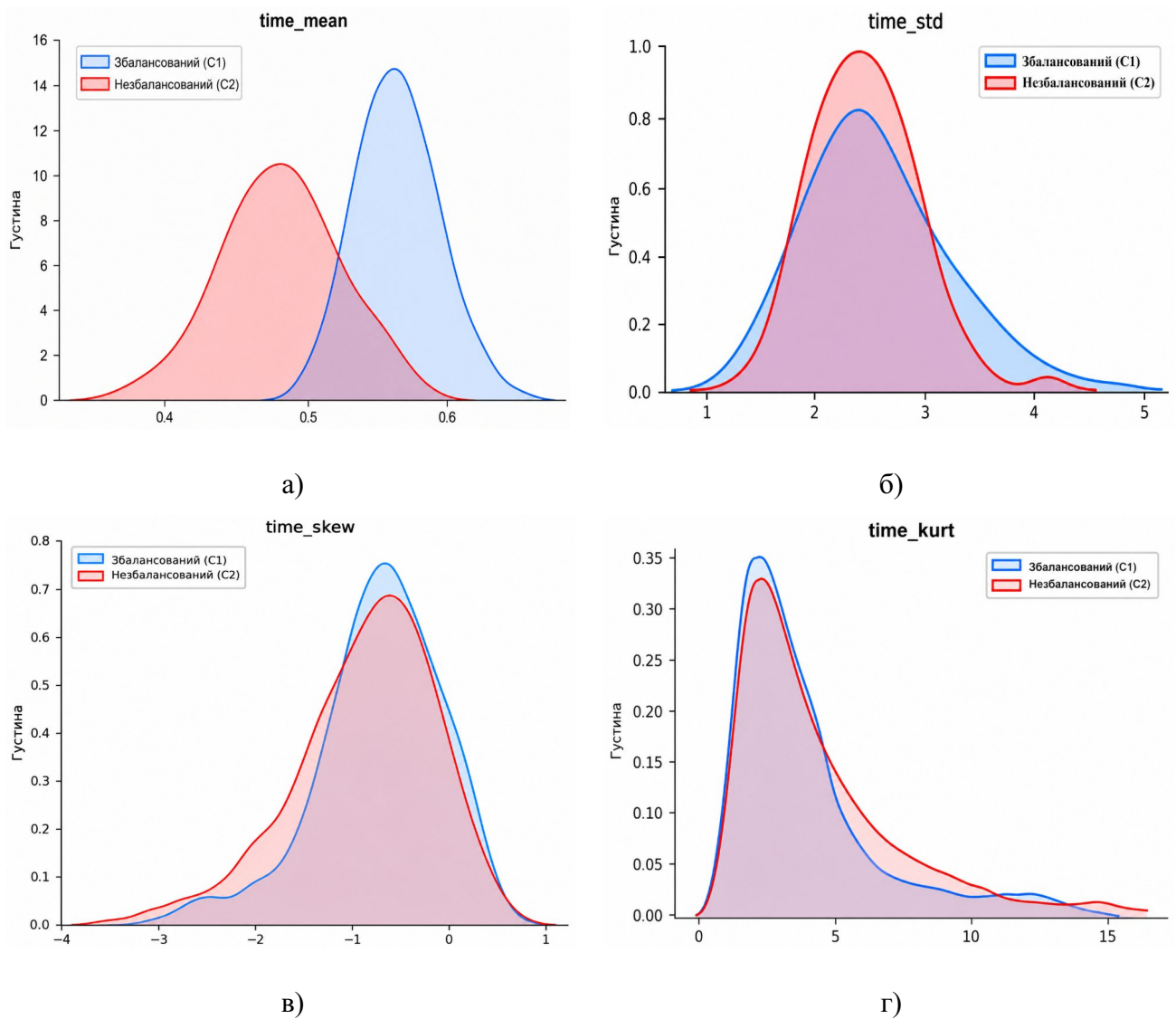


Рисунок 3.7 – Розподіли ознак схеми усереднення за часовим зсувом:

а) time\_mean – середній рівень;

б) time\_std – нерівномірність;

в) time\_skew – асиметрія;

г) time\_kurt – ексцес розподілу значень у часі

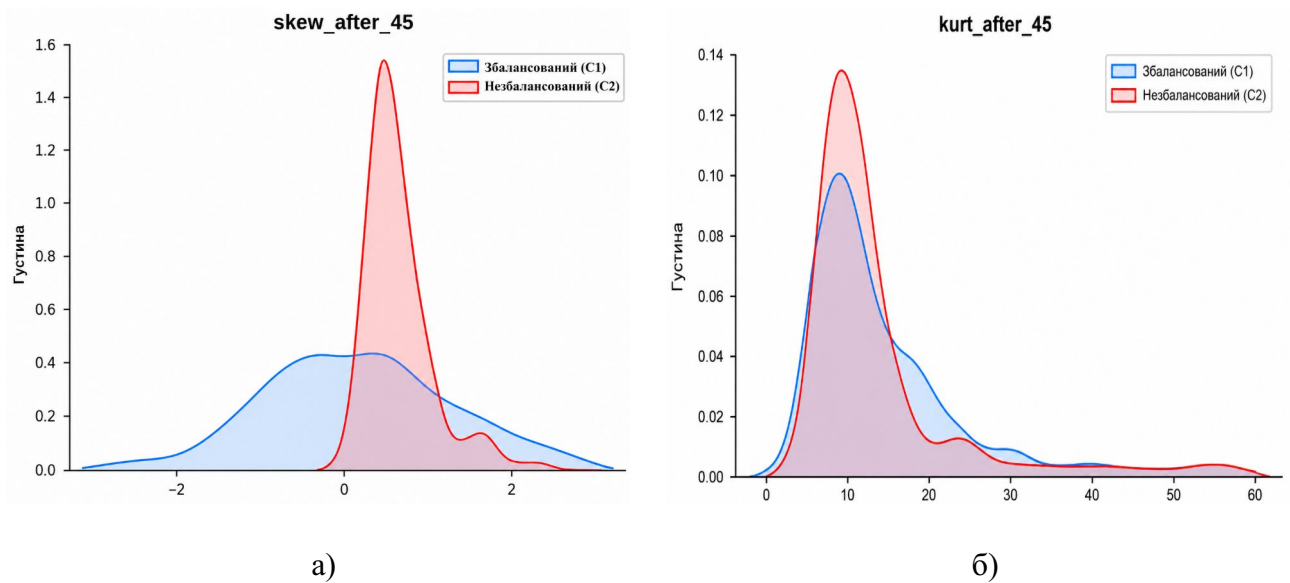


Рисунок 3.8 – Розподіли ознак форми залишку після точки  $x_{45}$ :

а) skew\_after\_45 – асиметрія після точки перегину;

б) kurt\_after\_45 – важкість хвостів

Чітке розділення розподілів підтверджує, що сформовані ознаки мають високу дискримінативну здатність, що є ключовою передумовою для ефективного застосування методів машинного навчання [111, 117].

Наведені розподіли відображають чітке розділення класів за більшістю ознак. Найбільш виражені відмінності спостерігаються для ознак `time_mean`, `time_std`, `exp_rmse` та `std_after_45`, для яких розподіли мають мінімальне перекриття. Це підтверджує інформативність зазначених ознак для завдань класифікації режимів електроспоживання [39, 40]. Додатково для порівняння медіан, кватилів та викидів побудовано діаграми розмаху розподілів (рисунок 3.9-3.12) [118, 119].

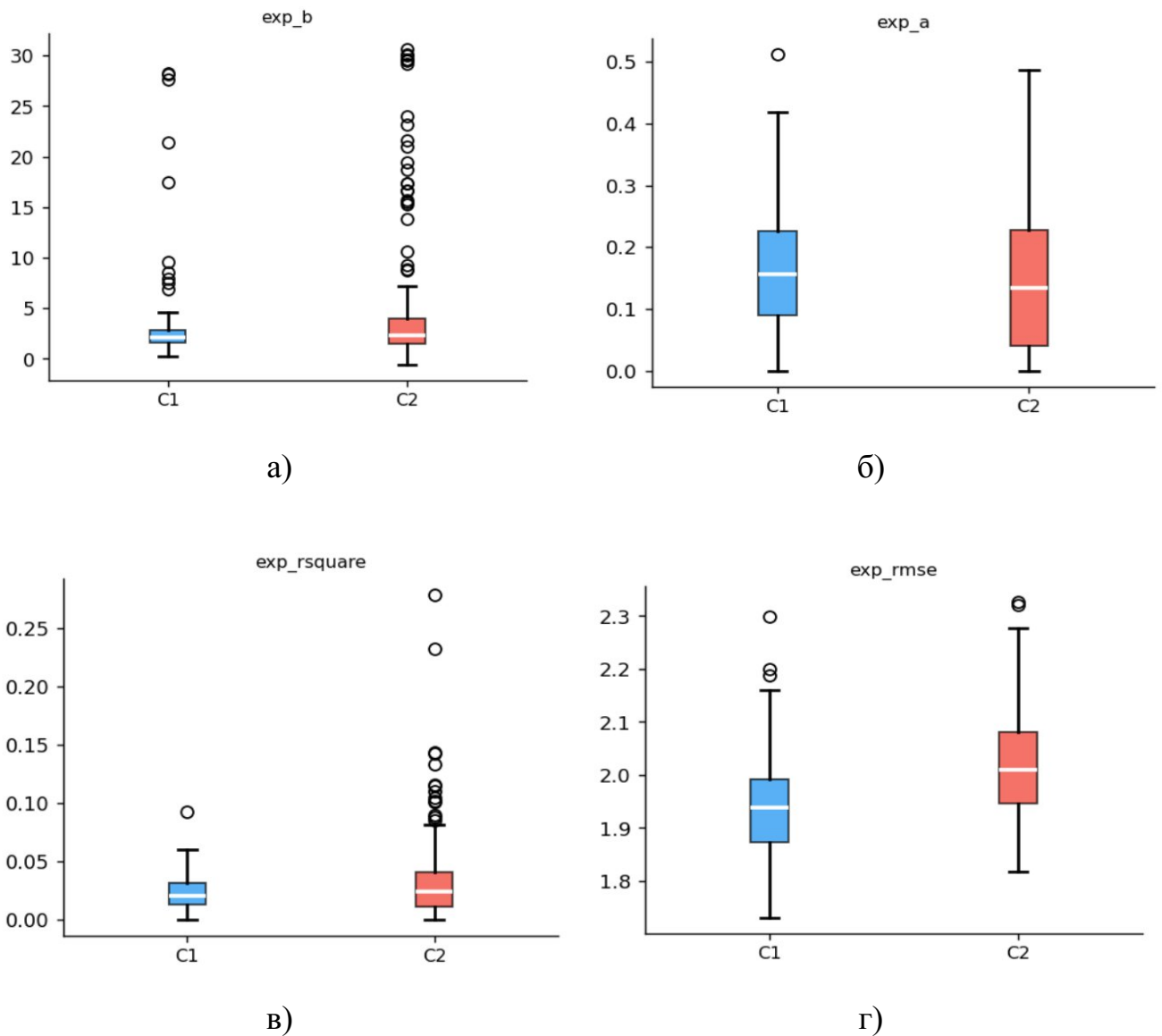


Рисунок 3.9 – Коробкові діаграми параметрів апроксимації для класів C1 та C2:

- а) exp\_b – швидкість зростання;
- б) exp\_a – початковий рівень;
- в) exp\_rsquare – коефіцієнт детермінації;
- г) exp\_rmse – середньоквадратична похибка

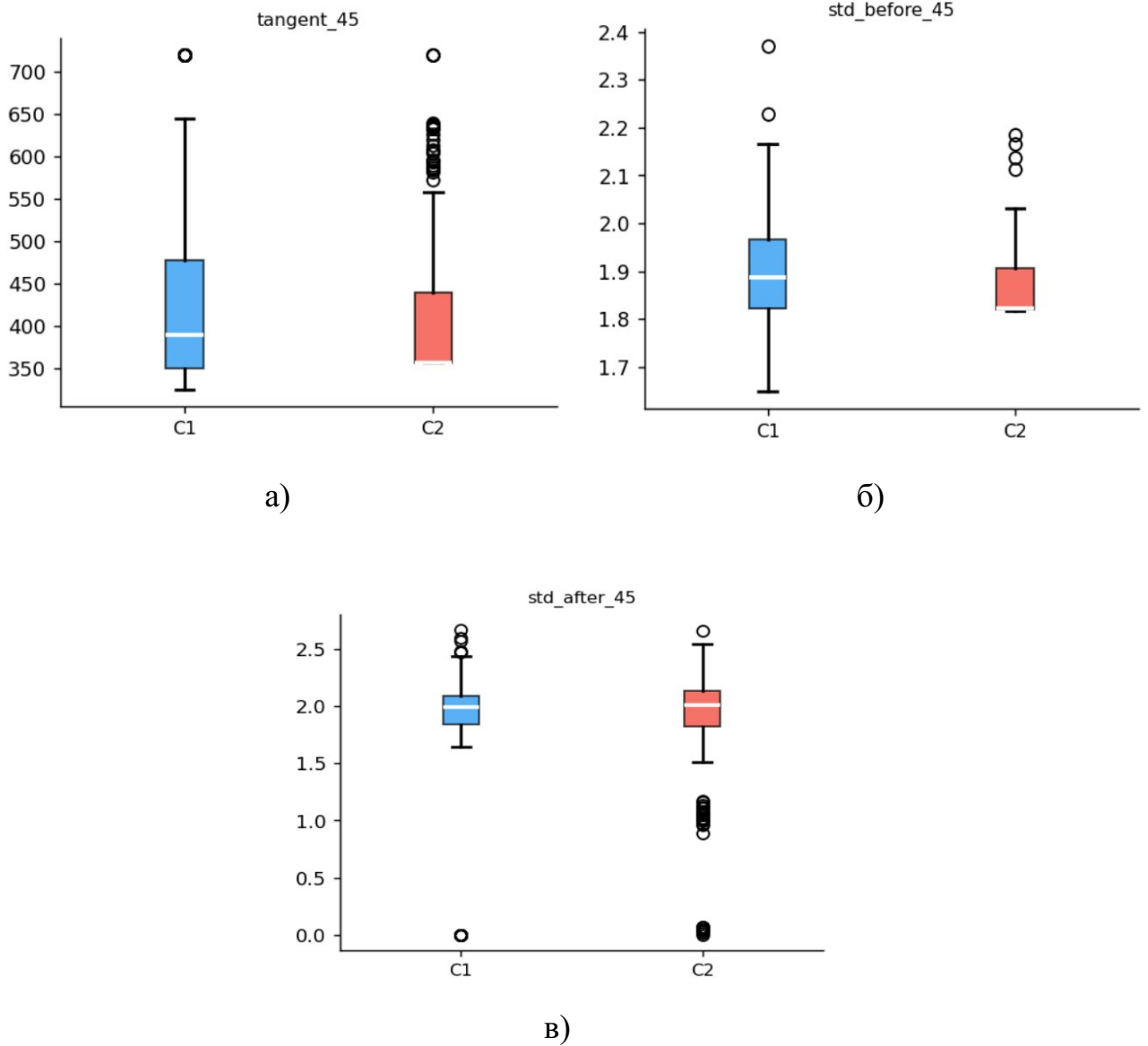


Рисунок 3.10 – Коробкові діаграми ознак  $x_{45}$  та залишку для класів C1 та C2:

- а) `tangent_45` – точка критичного зростання;
- б) `std_before_45` – розкид залишку до  $x_{45}$ ;
- в) `std_after_45` – розкид залишку після  $x_{45}$

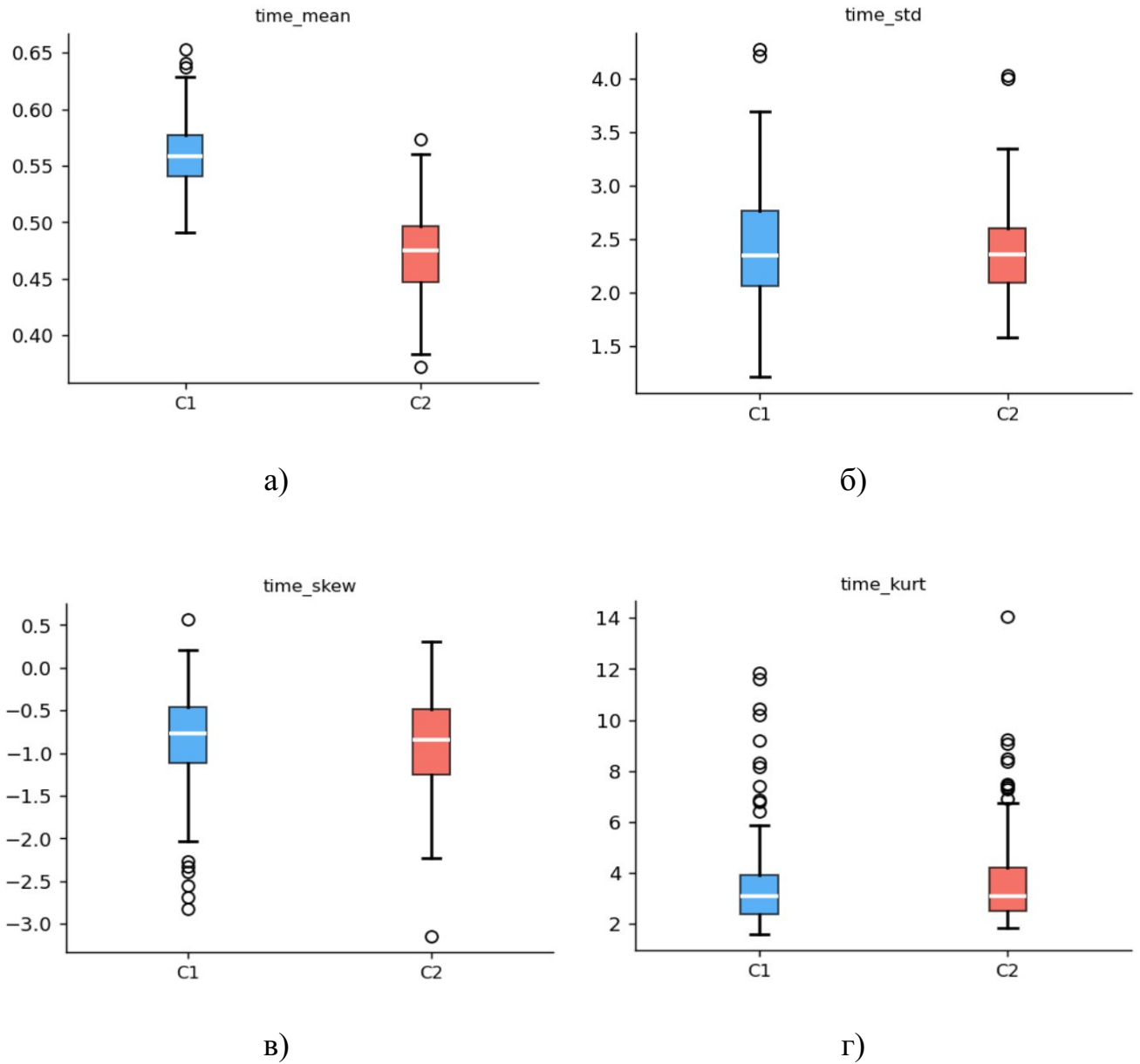


Рисунок 3.11 – Коробкові діаграми ознак усереднення за часовим зсувом для класів C1 та C2:

- а) time\_mean – середнє значення вектора;
- б) time\_std – стандартне відхилення вектора;
- в) time\_skew – асиметрія вектора;
- г) time\_kurt – ексцес вектора

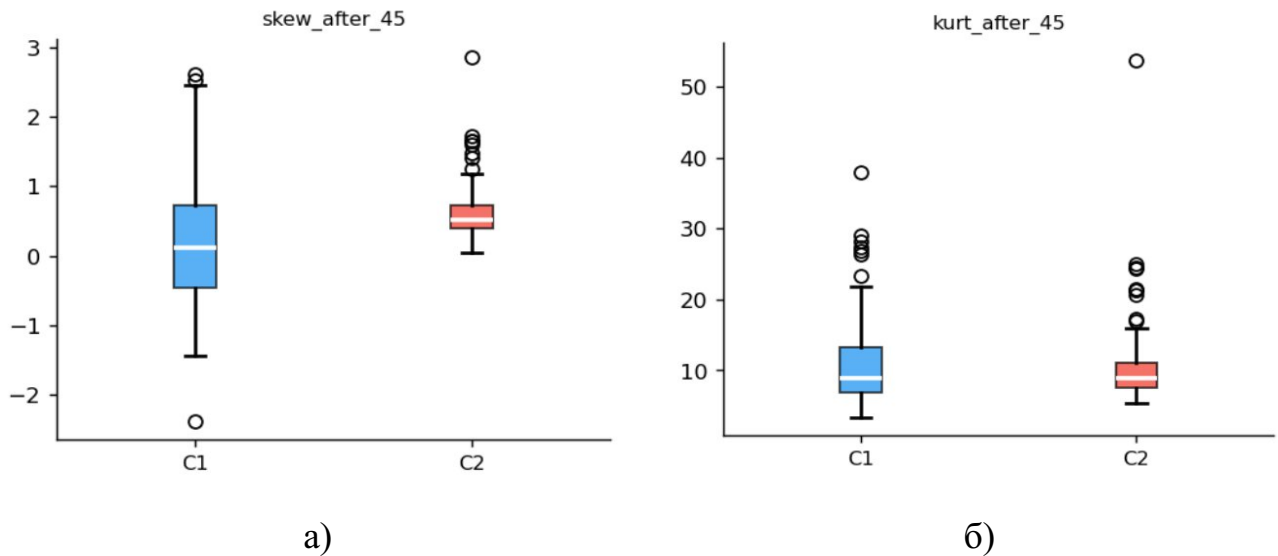


Рисунок 3.12 – Коробкові діаграми ознак форми залишку для класів C1 та C2:

а) skew\_after\_45 – асиметрія залишку після  $x_{45}$ ;

б) kurt\_after\_45 – ексцес залишку після  $x_{45}$

Наведені діаграми розподілу відображають відмінності у значеннях ознак для різних режимів електроспоживання. Зокрема, для ознак, що характеризують варіативність процесу ( $exp\_rmse$ ,  $std\_after\_45$ ,  $time\_std$ ), у незбалансованому режимі спостерігається значно більший розмах значень, що вказує на підвищену нестабільність електроспоживання.

Водночас ознака, що характеризує середній рівень процесу ( $time\_mean$ ), демонструє систематичне зміщення значень для різних класів, що відображає зміну загального рівня електроспоживання. Для ознак асиметрії та форми розподілу ( $skew\_after\_45$ ,  $kurt\_after\_45$ ) характерна підвищена варіативність у незбалансованому режимі, що свідчить про наявність імпульсних та екстремальних значень.

Окремо слід відзначити наявність викидів у незбалансованому режимі, що підтверджує підвищену стохастичність процесу електроспоживання. Для збалансованого режиму навпаки характерні більш компактні та стабільні розподіли ознак.

Отримані результати підтверджують інформативність сформованого ознакового простору та його придатність для завдань класифікації режимів електроспоживання. З метою виявлення мультиколінеарності між сформованими ознаками побудовано матрицю кореляцій Пірсона (рисунок 3.13).

Аналіз матриці дозволив ідентифікувати пари ознак з помірно високими коефіцієнтами кореляції, що свідчить про наявність часткової лінійної залежності між окремими параметрами ознакового простору [120].

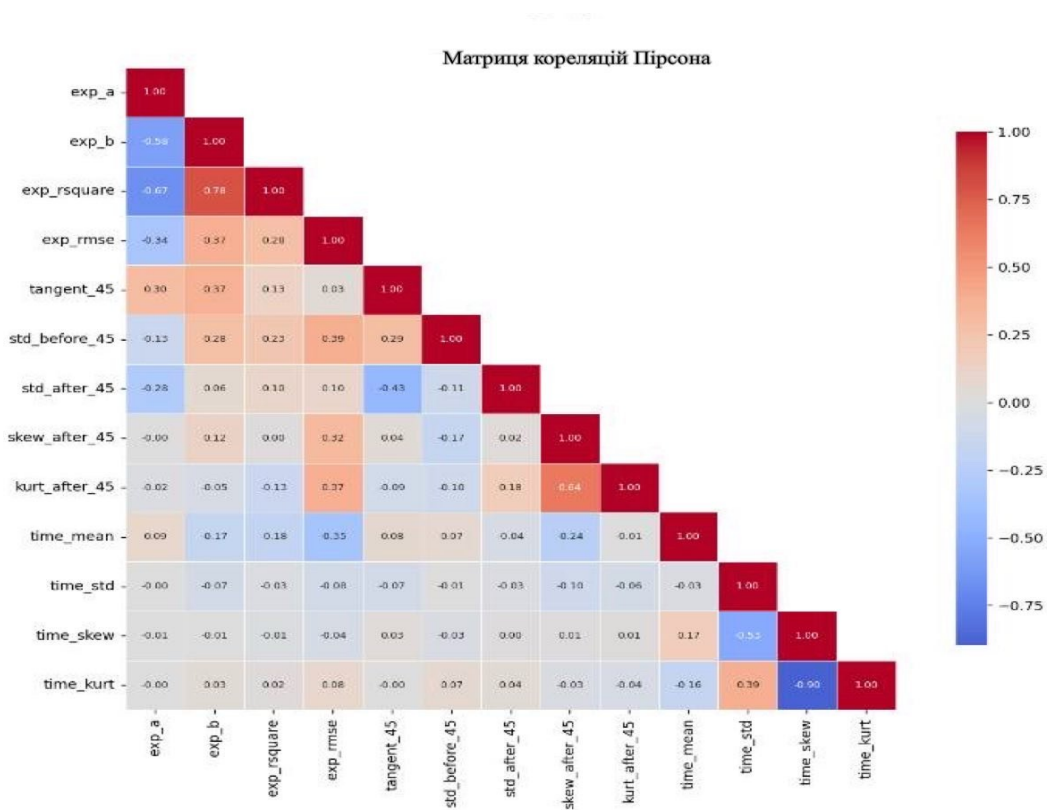


Рисунок 3.13 – Матриця кореляцій Пірсона для ознак компонентного методу

Для кількісної оцінки лінійних взаємозв'язків між статистичними ознаками компонентного методу використано коефіцієнт кореляції Пірсона, який для двох ознак  $x_i$  та  $x_j$  обчислюється за формулою:

$$r_{ij} = \frac{\sum_{k=1}^N (x_i^k - \bar{x}_i)(x_j^k - \bar{x}_j)}{\sqrt{\sum_{k=1}^N (x_i^k - \bar{x}_i)^2 \cdot \sum_{k=1}^N (x_j^k - \bar{x}_j)^2}} \quad (3.7)$$

де  $N$  кількість спостережень у вибірці;  $x_i^k, x_j^k$  значення  $k$ -го спостереження ознак  $x_i$  та  $x_j$  відповідно;  $\bar{x}_i, \bar{x}_j$  – вибіркові середні значення відповідних ознак.

Аналіз кореляційної матриці дозволяє виділити кілька груп взаємопов'язаних ознак. Найвищий від'ємний зв'язок спостерігається між `time_skew` та `time_kurt` ( $r=-0,90$ ), що відображає компенсаторний характер асиметрії та ексцесу часових характеристик сигналу. Між ознаками `exp_b` та `exp_rsquare` спостерігається помірно висока кореляція ( $r=0,78$ ), а `exp_a` корелює з обома від'ємно ( $r=-0,58$  та  $r=-0,67$  відповідно), що відображає взаємозв'язок параметрів апроксимаційної моделі.

Ознаки `skew_after_45` та `kurt_after_45` демонструють між собою помірну кореляцію ( $r=0,64$ ), що вказує на їх спільну чутливість до форми розподілу залишків. Ознака `time_std` має помірний від'ємний зв'язок з `time_skew` ( $r=-0,53$ ). Решта ознак – `exp_rmse`, `tangent_45`, `std_before_45`, `std_after_45`, `time_mean` – демонструють слабкі кореляції між собою та з іншими параметрами ( $r<0,40$ ), що свідчить про їх відносну незалежність та інформативність для завдань класифікації.

Для збалансованого режиму характерна компактна локалізація спостережень, тоді як для незбалансованого режиму спостерігається значно більша варіативність значень ознак. Така структура даних підтверджує високу дискримінаційну здатність відібраних ознак та обґрунтовує доцільність їх використання у завданнях класифікації режимів електроспоживання.

Отримані результати узгоджуються з моделлю ПКВП [43], оскільки демонструють періодичну зміну статистичних характеристик та підвищену варіативність у незбалансованих режимах. Для відображення відмінностей між класами побудовано усереднені оцінки кореляційних компонентів  $\hat{B}_k(u)$  з

довірчими інтервалами 95% окремо для збалансованого (С1) та незбалансованого (С2) режимів. Усереднена оцінка за компонентами обчислювалась як середнє значення кореляційних компонент по всіх гармоніках  $k$  для кожного фіксованого часового зсуву  $u$  згідно з виразом (2.14). Усереднена оцінка за часовим зсувом обчислювалась як середнє значення  $\hat{B}_k(u)$  по всіх значеннях  $u$  для кожної фіксованої гармоніки  $k$  згідно з виразом (2.15).

З метою порівняльного аналізу кореляційної структури збалансованого та незбалансованого режимів електроспоживання побудовано усереднені оцінки кореляційних компонентів  $\hat{B}_k(u)$  з довірчими інтервалами 95% за схемою усереднення за компонентами  $k$  (рисунки 3.14–3.15).

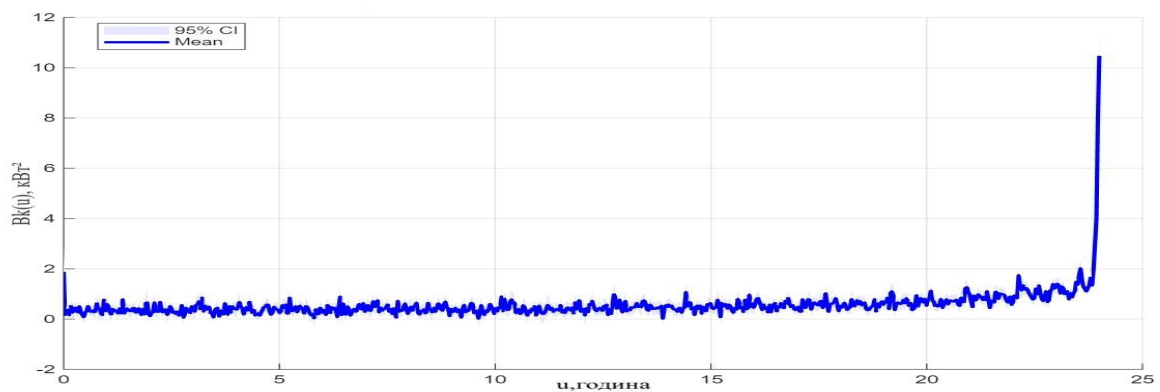


Рисунок 3.14 – Усереднені кореляційні компоненти з довірчим інтервалом 95% для збалансованого режиму (усереднення за компонентами)

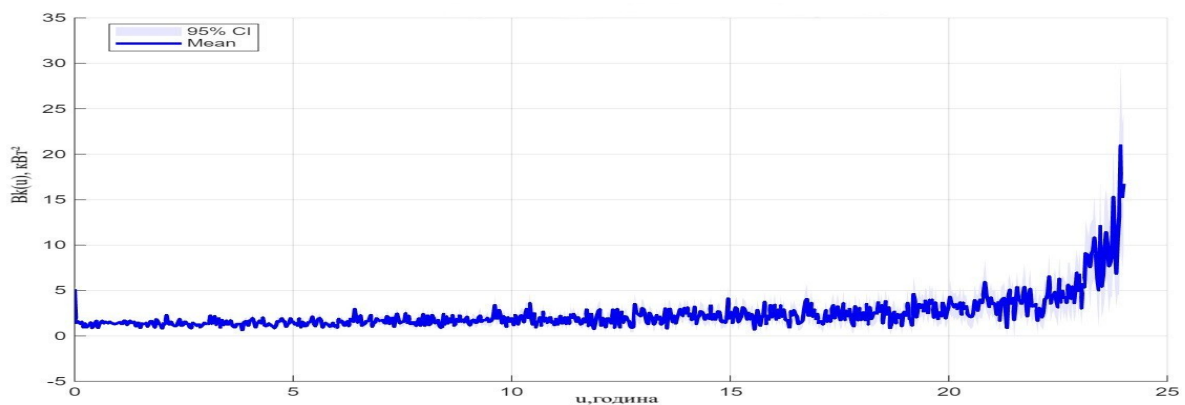


Рисунок 3.15 – Усереднені кореляційні компоненти з довірчим інтервалом 95% для незбалансованого режиму (усереднення за компонентами)

Для збалансованого режиму характерна відносно стабільна поведінка сигналу з незначним зростанням наприкінці інтервалу та вузькими довірчими інтервалами. В свою чергу порівняння профілів підтверджує, що незбалансований режим характеризується суттєво вищою міжцикловою варіативністю як при усередненні за компонентами, так і при усередненні за часом. На рисунку 3.16 наведено усереднений профіль електроспоживання з довірчим інтервалом 95% для збалансованого режиму при усередненні за часом.

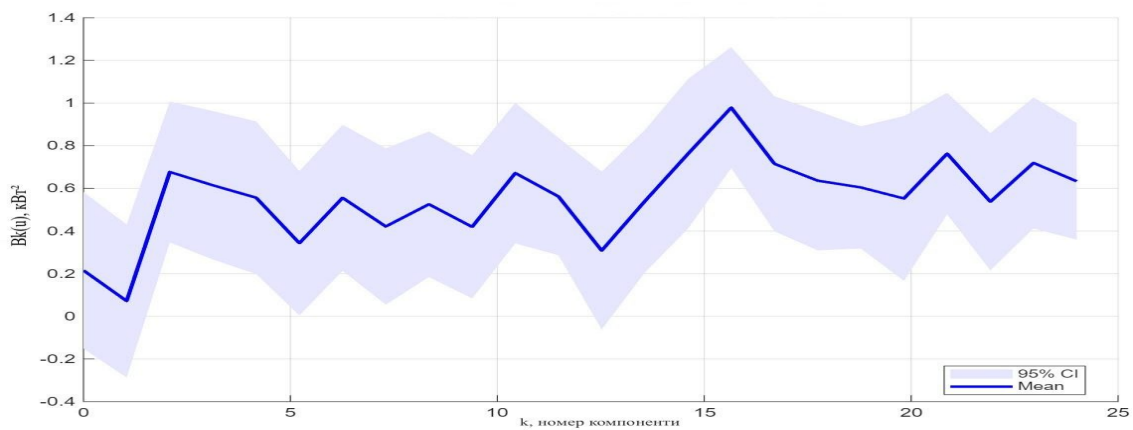


Рисунок 3.16 – Усереднені кореляційні компоненти з довірчим інтервалом 95% для збалансованого режиму (усереднення за часовим зсувом)

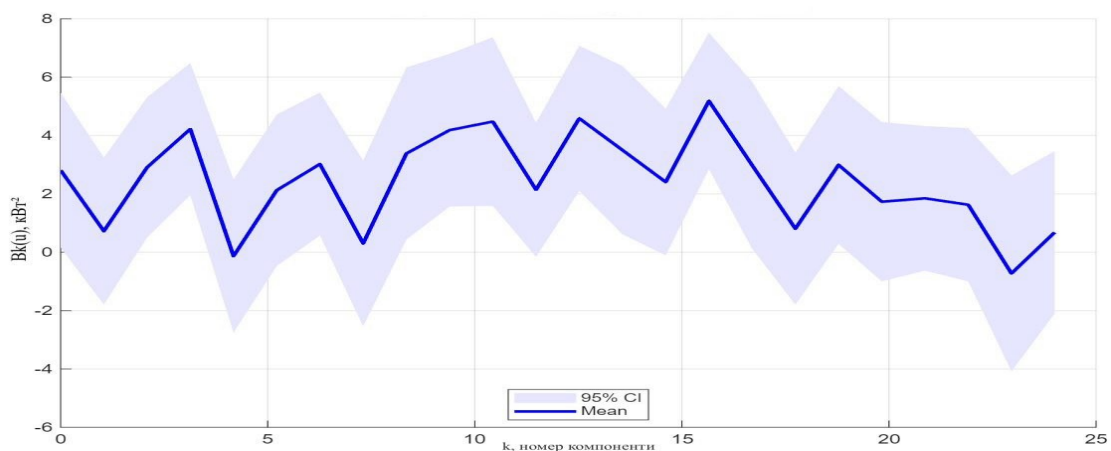


Рисунок 3.17 – Усереднені кореляційні компоненти з довірчим інтервалом 95% для незбалансованого режиму (усереднення за часовим зсувом)

Для збалансованого режиму характерні відносно рівномірні коливання значень із помірною варіативністю та вузькими довірчими інтервалами. Натомість для незбалансованого режиму спостерігається більший розкид значень і ширші довірчі інтервали, що свідчить про підвищену варіативність електроспоживання.

### **3.3. Застосування методів машинного навчання та оцінка їх ефективності для класифікації режимів електроспоживання**

Для оцінки дискримінативної здатності сформованого ознакового простору проведено експеримент із застосуванням таких методів машинного навчання [121]: лінійних (Logistic Regression, LDA), ймовірнісних (Naive Bayes), метричних (KNN), деревоподібних (Decision Tree, Random Forest, Extra Trees), бустингових (AdaBoost, Gradient Boosting, XGBoost, LightGBM), нейромережових (MLP), методу опорних векторів (SVM) та ансамблевих (Voting Ensemble).

Запропонований набір моделей представляє собою комплексну парадигму класифікаційних алгоритмів машинного навчання, що охоплює спектр від базових лінійних моделей до складних нелінійних ансамблів та штучних нейромереж, загалом використано 14 методів машинного навчання.

Для забезпечення коректної роботи алгоритмів машинного навчання та усунення впливу різниці в масштабах вимірювань сформованих інформативних ознак, на етапі попереднього опрацювання даних виконано їх нормалізацію. У роботі застосовано метод лінійного масштабування (Min-Max нормалізація), який відображає вихідні значення ознак у фіксований діапазон [0, 1]. Математично цей процес описується виразом:

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (3.8)$$

де  $X$  – вихідне значення ознаки,  $X_{min}$  та  $X_{max}$  – відповідно мінімальне та максимальне значення цієї ознаки у межах аналізованої вибірки, а  $X_{norm}$  – нормалізоване значення, що використовується для подальшого навчання методів машинного навчання.

Такий підхід дозволив зберегти відносні пропорції між значеннями та форму розподілу даних, повністю нівелювавши розбіжність у розмірностях між часовими характеристиками (що вимірюються в годинах або відносних одиницях точок перелому) та компонентними ознаками ПКВП (що відображають інтенсивність стохастичних флуктуацій сигналу). Це забезпечило однакову вагу всіх 13 ознак під час формування розділової гіперплощини та обчислення критеріїв стабільності в ансамблевих моделях (Extra Trees, Random Forest) та нейромережових.

Logistic Regression виступає класичним статистичним методом для оцінки ймовірностей належності до певного класу, де стандартний параметр регуляризації ( $C=1.0$ ) запобігає перенавчанню, а тисяча ітерацій гарантує стабільну збіжність оптимізаційного процесу. LDA (Лінійний дискримінантний аналіз) проектує багатовимірний простір ознак у простір меншої вимірності, аналітично максимізуючи відстань між центроїдами різних класів та мінімізуючи дисперсію всередині них. Naive Bayes з припущенням про гауссівський розподіл спирається на теорему Байєса із припущенням про стохастичну незалежність усіх предикторів. KNN ( $k=5$ ) здійснює локальну апроксимацію топології даних, класифікуючи нові вектори на основі мажоритарного голосування п'яти найближчих сусідів.

У свою чергу Decision Tree формує ієрархічні нелінійні правила прийняття рішень з обмеженням глибини до шести рівнів. Random Forest агрегує прогнози трьохсот незалежних дерев рішень, кожне побудоване на вибірці з повторенням із навчального набору [122]. Extra Trees додатково вносить стохастичність у процес розбиття вузлів [123]. AdaBoost послідовно навчає двісті базових

класифікаторів, адаптивно збільшуючи ваги хибно класифікованих спостережень. Gradient Boosting оптимізує функцію втрат шляхом ітеративного додавання дерев [124, 125], які апроксимують антиградієнт. XGBoost та LightGBM представляють оптимізовані реалізації градієнтного бустингу. Voting Ensemble об'єднує прогнози кількох кращих моделей.

SVM (RBF) використовує метод опорних векторів з радіальною базисною функцією як ядро ( $C=10$ ), що дозволяє неявно відобразити вхідні дані у нескінченновимірний простір [123]. MLP Neural Net являє собою багатопшаровий перцептрон з двома прихованими шарами на 128 та 64 нейрони з L2-регуляризацією ( $\alpha=10^{-4}$ ).

Для комплексної оцінки якості класифікації режимів електроспоживання використано систему метрик, що охоплює як класичні показники, так і кореляційні міри узгодженості. Усі метрики базуються на елементах матриці плутанини (confusion matrix), яка фіксує чотири типи результатів класифікації: істинно-позитивні (TP), істинно-негативні (TN), хибно-позитивні (FP) та хибно-негативні (FN).

Accuracy (точність класифікації) визначає загальну частку правильно класифікованих спостережень серед усіх елементів вибірки:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (3.9)$$

Ця метрика є інтуїтивно зрозумілою, проте може бути оманливою при незбалансованих класах, оскільки модель, яка завжди прогнозує мажоритарний клас, може демонструвати високу точність без реальної дискримінативної здатності.

Precision (точність, або позитивна прогностична цінність) оцінює частку дійсно позитивних результатів серед усіх спостережень, класифікованих як позитивні:

$$Precision = \frac{TP}{TP+FP}, \quad (3.10)$$

Високе значення Precision свідчить про низький рівень хибних спрацювань, що є критичним у завданнях, де помилкове віднесення до незбалансованого режиму може спричинити необґрунтоване втручання в роботу системи. Recall (повнота, або чутливість) визначає частку правильно ідентифікованих позитивних випадків серед усіх фактично позитивних:

$$Recall = \frac{TP}{TP + FN}, \quad (3.11)$$

Висока повнота гарантує, що більшість реальних випадків незбалансованого режиму буде виявлена, що є пріоритетним для систем раннього попередження про аномальне електроспоживання.

F1-score являє собою гармонічне середнє між Precision та Recall і забезпечує збалансовану оцінку якості класифікації, особливо при нерівномірному розподілі класів:

$$F1 = \frac{2 \cdot precision \times recall}{precision + recall}, \quad (3.12)$$

F1-score набуває значення 1 лише за умови одночасного досягнення ідеальних Precision та Recall, що робить цю метрику основним критерієм порівняння моделей у цьому дослідженні. Коефіцієнт кореляції Метьюза (MCC, Matthews Correlation Coefficient) є кореляційною мірою між спостережуваними та передбаченими класами [125], яка враховує усі чотири елементи матриці плутанини:

$$MCC = \frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}}, \quad (3.13)$$

де  $TP$  кількість правильно класифікованих позитивних прикладів;  $TN$  кількість правильно класифікованих негативних прикладів;  $FP$  кількість хибнопозитивних;  $FN$  кількість хибнонегативних.

МСС приймає значення від  $-1$  до  $+1$ , де  $+1$  відповідає ідеальній класифікації,  $0$  випадковому вгадуванню,  $-1$  повній інверсії класифікації. На відміну від Ассигасу, МСС є інформативним навіть при значному дисбалансі класів, оскільки збалансовано враховує помилки обох типів.

Коефіцієнт узгодженості Коена (Cohen's Kappa) оцінює ступінь узгодженості між передбаченою та фактичною класифікацією з поправкою на випадковий збіг:

$$k = \frac{p_o - p_e}{1 - p_e}, \quad (3.14)$$

де  $p_o$  – спостережувана частка узгоджених класифікацій (тобто Ассигасу), а  $p_e$  очікувана частка збігів за умови незалежності передбачень від істинних міток.

Значення  $k = 1$  відповідає ідеальній узгодженості,  $k = 0$  рівню випадкового класифікатора, а від'ємні значення свідчать про систематичну розбіжність між прогнозами та реальністю [127]. Основною метрикою обрано F1-score, оскільки вона враховує баланс між помилками першого та другого роду, що є критично важливим для завдань аналізу енергосистем.

Площа під ROC-кривою (AUC, Area Under the Receiver Operating Characteristic Curve) оцінює здатність моделі розрізняти класи незалежно від обраного порогу класифікації [126]. ROC-крива будується у координатах «чутливість (TPR) частка хибних спрацювань (FPR)» при варіюванні порогу від  $0$  до  $1$ .  $AUC = 1.0$  відповідає ідеальному розрізненню класів,  $AUC = 0,5$  випадковому класифікатору. У контексті класифікації режимів електроспоживання значення  $AUC > 0.99$ , отримані для більшості моделей,

свідчать про практично ідеальну роздільну здатність сформованого простору ознак.

Крос-валідаційне значення F1 (CV-F1) отримане методом стратифікованої k-кратної крос-валідації ( $k = 5$ ) і відображає стійкість та узагальнюючу здатність моделі на незалежних підвбірках. Суттєва розбіжність між тестовим F1 та CV-F1 може свідчити про перенавчання або нестійкість моделі. Обрана система метрик забезпечує всебічну оцінку моделей з різних перспектив Accuracy та F1 оцінюють загальну якість, MCC та Карра кореляційну узгодженість з поправкою на випадковість, AUC порогово-незалежну роздільну здатність, а CV-F1 стійкість до варіацій у даних.

Для оцінювання моделей машинного навчання датасет розділено на навчальну (273 зразки, 75%) та тестову (92 зразки, 25%). Результати оцінки на тестовій вибірці наведено в таблиці 3.2 та відсортовано за спаданням F1-score.

Таблиця 3.2

Порівняльний аналіз моделей машинного навчання за метриками якості

Моделі МН	Acc	Prec	Rec	F1	MCC	Карра	CV-F1
KNN (k=5)	0,9891	0,9756	1,0000	0,9877	0,9782	0,9779	0,9344
Extra Trees	0,9891	0,9756	1,0000	0,9877	0,9782	0,9779	0,9826
SVM (RBF)	0,9891	0,9756	1,0000	0,9877	0,9782	0,9779	0,9524
MLP Neural Net	0,9891	1,0000	0,9750	0,9873	0,9781	0,9778	0,9477
XGBoost	0,9783	0,9750	0,9750	0,9750	0,9558	0,9558	0,9574
Random Forest	0,9783	0,9750	0,9750	0,9750	0,9558	0,9558	0,9659
Voting Ensemble	0,9783	0,9750	0,9750	0,9750	0,9558	0,9558	0,9626
AdaBoost	0,9674	0,9744	0,9500	0,9620	0,9337	0,9335	0,9745
LightGBM	0,9674	0,9744	0,9500	0,9620	0,9337	0,9335	0,9743
Logistic Regression	0,9674	0,9744	0,9500	0,9620	0,9337	0,9335	0,9249

## Продовження таблиці 3.2

Моделі МН	Acc	Prec	Rec	F1	MCC	Kappa	CV-F1
Naive Bayes	0,9674	1,0000	0,9250	0,9610	0,9352	0,9331	0,8507
Decision Tree	0,9565	0,9737	0,9250	0,9487	0,9119	0,9110	0,9174
Gradient Boosting	0,9565	0,9737	0,9250	0,9487	0,9119	0,9110	0,9074
LDA	0,9457	0,9730	0,9000	0,9351	0,8905	0,8885	0,9191

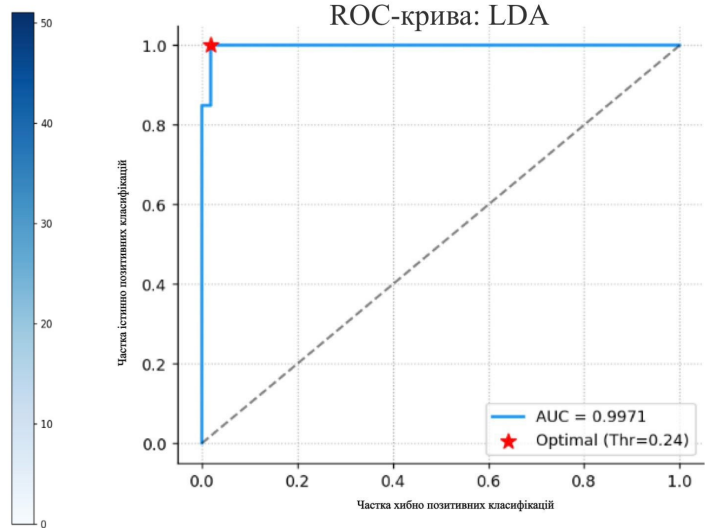
Отримані результати свідчать про те, що сформований ознаковий простір має високу роздільну здатність, що дозволяє досягати майже ідеальної класифікації незалежно від типу алгоритму.

Найкращі результати за F1-score (0,9877) показали три моделі: KNN (k=5), Extra Trees та SVM (RBF), які досягли Recall = 1,0 та Precision = 0,9756. За критерієм крос-валідаційної стабільності перевагу має Extra Trees (CV-F1 = 0,9826) порівняно з SVM (CV-F1 = 0,9524) та KNN (CV-F1 = 0,9344). MLP Neural Net також показав F1 = 0,9873, з значення Precision та Recall (0,9750), найнижчий результат показав LDA з F1 = 0,9351. Перевага Extra Trees за критерієм CV-F1 свідчить про вищу узагальнювальну здатність моделі на нових даних порівняно з KNN та SVM, що є визначальним для практичного застосування у системах моніторингу режимів електроспоживання. Найнижчу ефективність продемонструвала модель LDA (F1=0,9351), що свідчить про високу дискримінантну здатність сформованого ознакового простору навіть для лінійних класифікаторів.

Усі моделі демонструють високу якість класифікації з MCC та Карра вище 0,88 що свідчить про надійність виділених ознак для розрізнення режимів енергоспоживання. Крос-валідаційні оцінки підтверджують стійкість моделей, при цьому Extra Trees показує найкращу збіжність між тестовою та крос-валідаційною оцінкою.



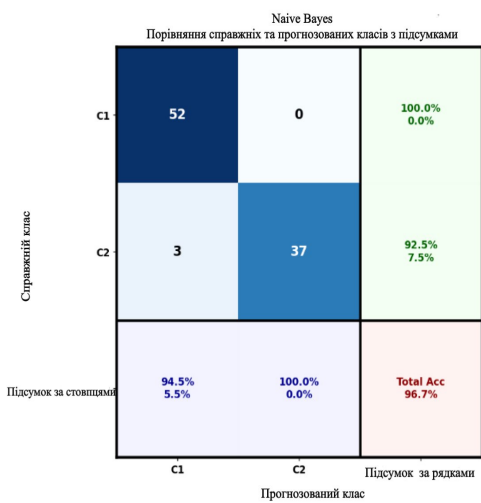
а)



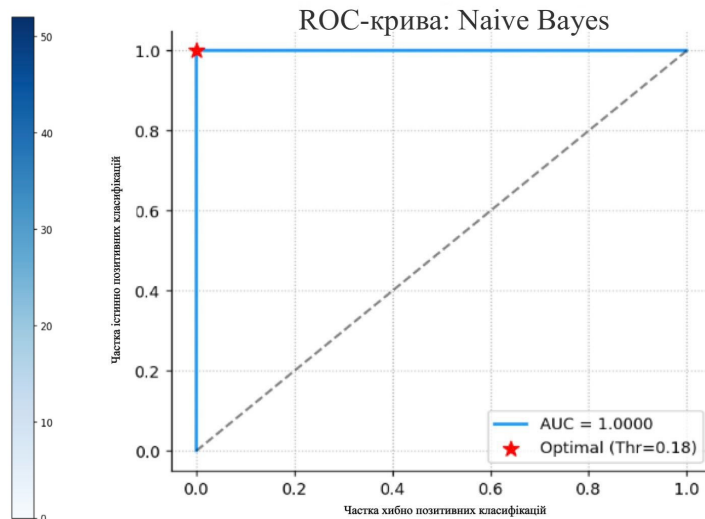
б)

Рисунок 3.18– Модель LDA:

а) матриця плутанини; б) ROC–крива



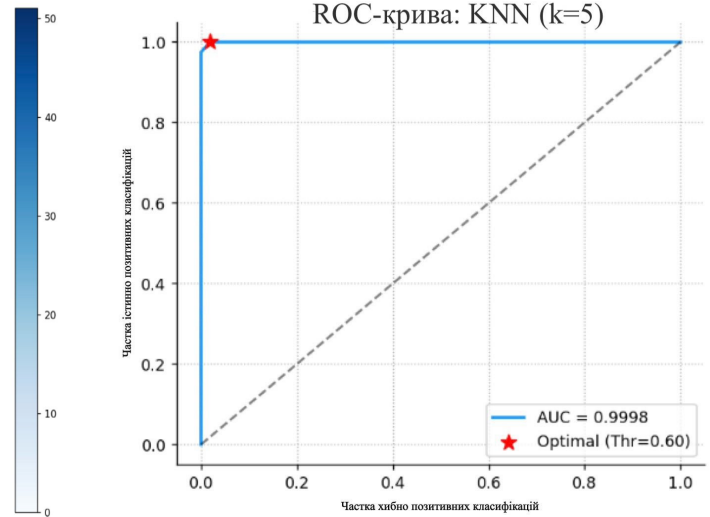
а)



б)

Рисунок 3.19 – Модель Naive Bayes:

а) матриця плутанини; б) ROC–крива

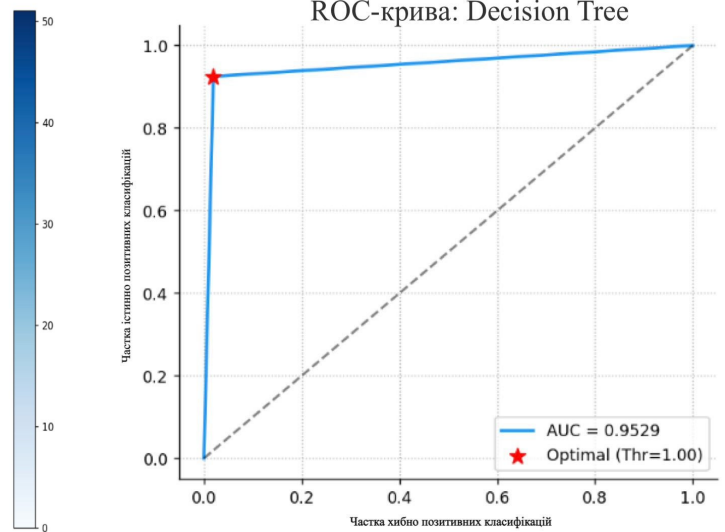
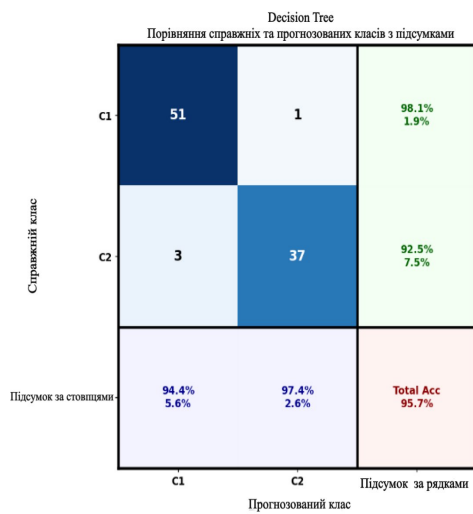


а)

б)

Рисунок 3.20 – Модель KNN(k=5):

а) матриця плутанини; б) ROC-крива

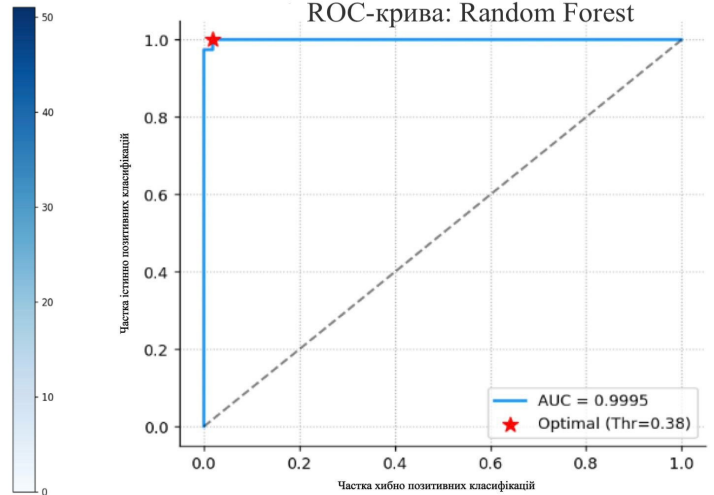
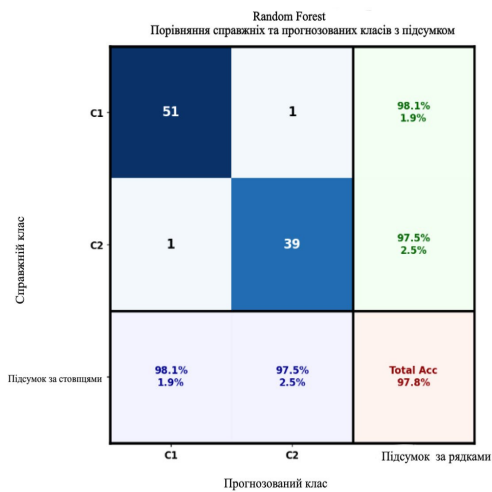


а)

б)

Рисунок 3.21 – Модель Decision Tree:

а) матриця плутанини; б) ROC-крива

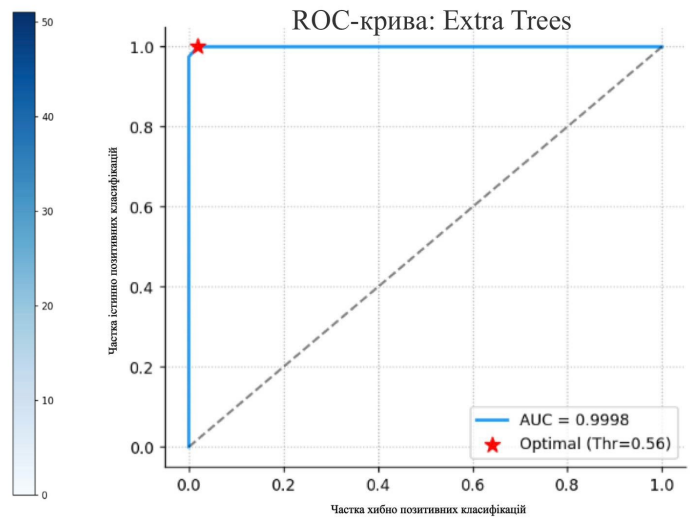
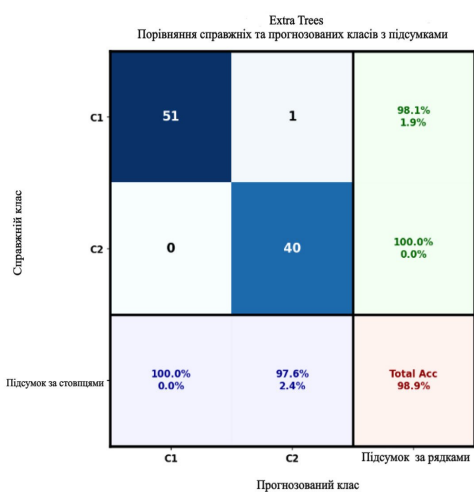


а)

б)

Рисунок 3.22 – Модель Random Forest:

а) матриця плутанини; б) ROC-крива



а)

б)

Рисунок 3.23 – Модель Extra Trees:

а) матриця плутанини; б) ROC-крива

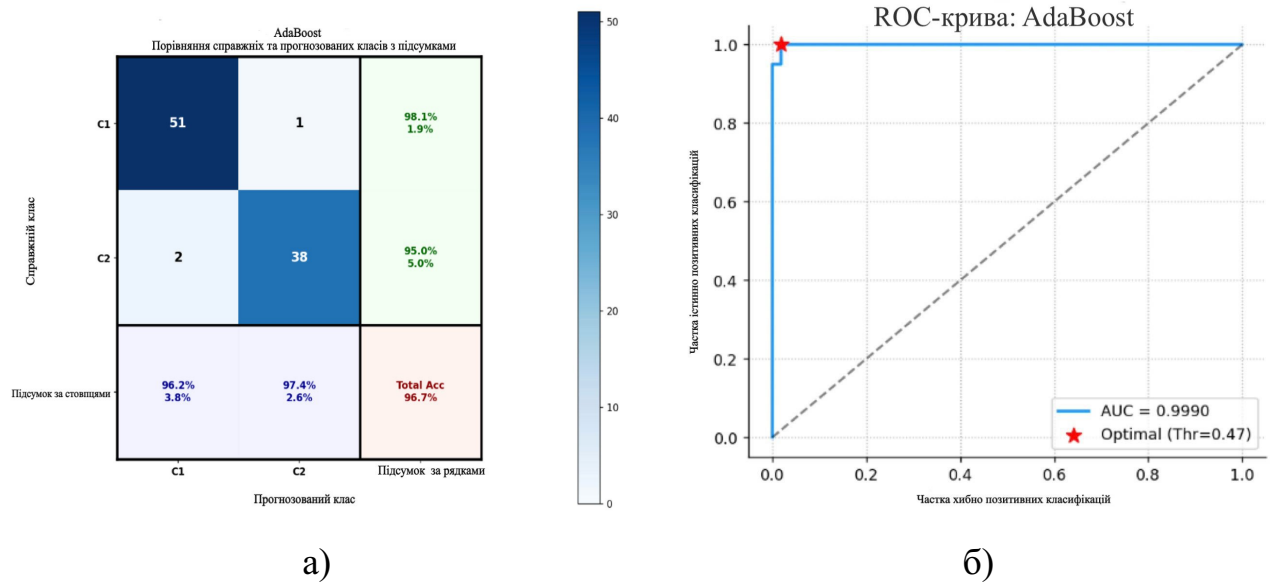


Рисунок 3.24 – Модель AdaBoost:

a) матриця плутанини; b) ROC-крива

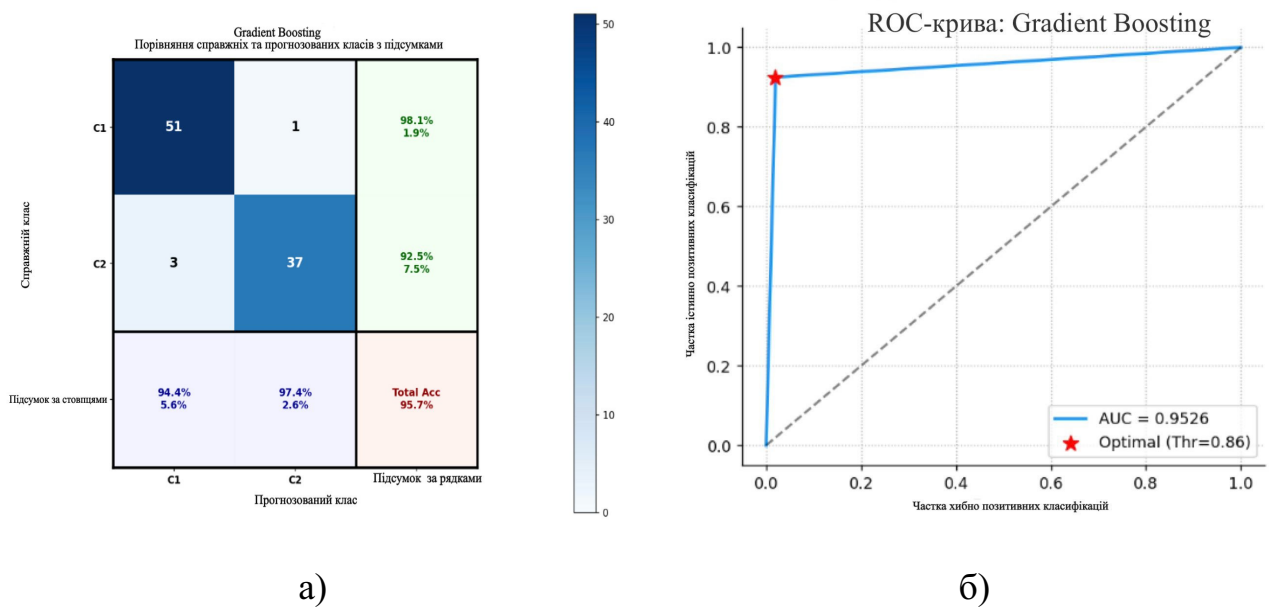
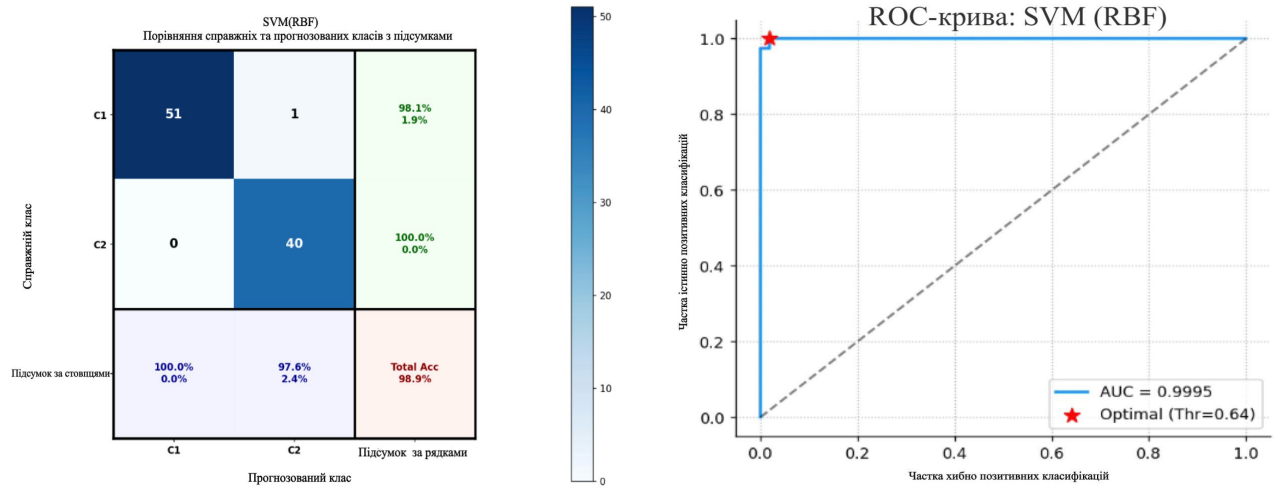


Рисунок 3.25 – Модель Gradient Boosting:

a) матриця плутанини; b) ROC-крива

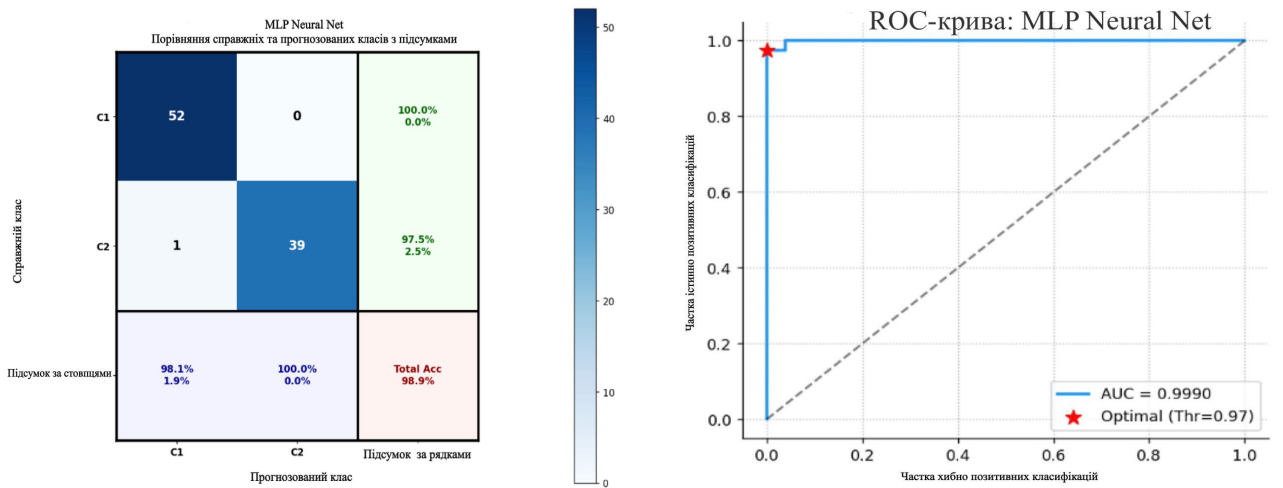


а)

б)

Рисунок 3.26 – Модель SVM(RBF):

а) матриця плутанини; б) ROC-крива

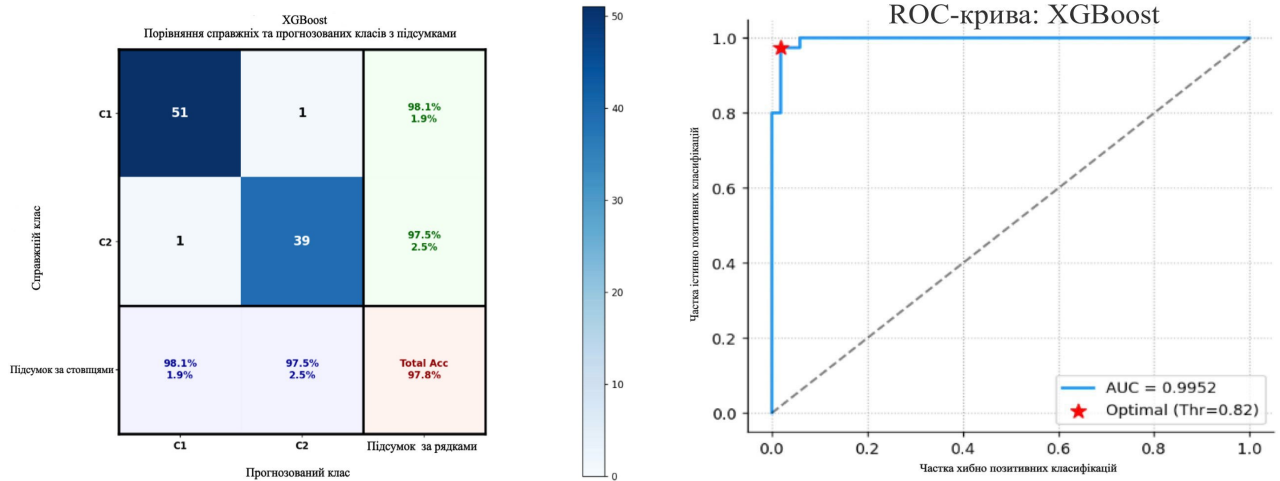


а)

б)

Рисунок 3.27 – Модель MLP Neural Net:

а) матриця плутанини; б) ROC-крива

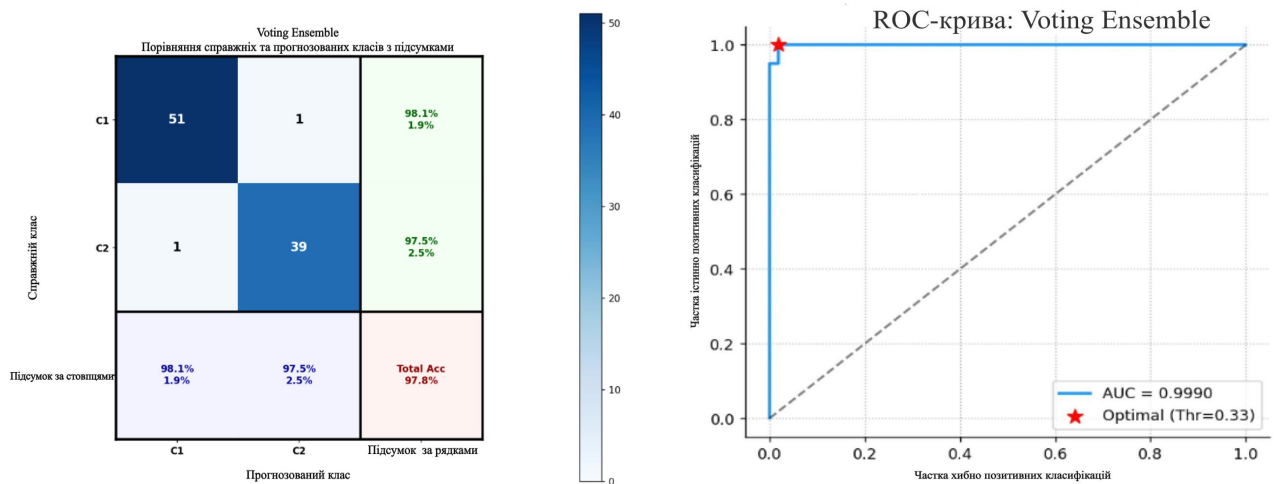


а)

б)

Рисунок 3.28 – Модель XGBoost:

а) матриця плутанини; б) ROC-крива

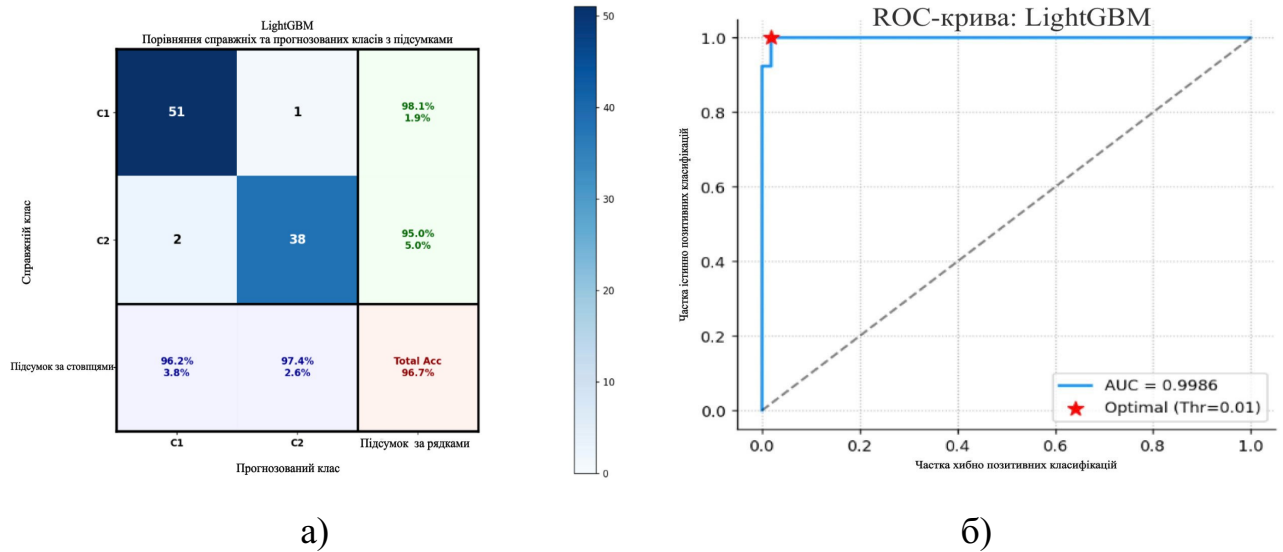


а)

б)

Рисунок 3.29 – Модель Voting Ensemble:

а) матриця плутанини; б) ROC-крива



а)

б)

Рисунок 3.30 – Модель LightGBM:

а) матриця плутанини; б) ROC-крива

Отримані матриці плутанини та ROC-криві підтверджують високу якість класифікації для більшості досліджених моделей. ROC-криві та значення AUC (Area Under the Curve) для кожної моделі наведено нижче.

Значення AUC перевищує 0.99 для дванадцяти з чотирнадцяти методів машинного навчання, що свідчить про ефективність сформованого ознакового простору на основі апарату ПКВП. Нижчі результати Decision Tree та Gradient Boosting пояснюються відомими обмеженнями цих методів без додаткового налаштування гіперпараметрів.

Таблиця 3.3

Значення AUC для розглянутих методів машинного навчання

Модель МН	AUC
Naive Bayes	1,000000
KNN (k=5)	0,999760
Extra Trees	0,999760
Random Forest	0,999519

## Продовження таблиці 3.3

Модель МН	AUC
SVM (RBF)	0,999519
AdaBoost	0,999038
Voting Ensemble	0,999038
MLP Neural Net	0,999038
LightGBM	0,998558
LDA	0,997115
Logistic Regression	0,996154
XGBoost	0,995192
Decision Tree	0,952885
Gradient Boosting	0,952644

Для аналізу дискримінантної здатності побудовано ROC-криві для всіх 14 досліджуваних класифікаторів (рис. 3.31). Переважна більшість моделей демонструє значення AUC в діапазоні 0,995–0,999, що свідчить про їх високу здатність розрізнити збалансований та незбалансований режими електроспоживання незалежно від обраного порогу класифікації. Нижчі значення AUC зафіксовано для Decision Tree (0,9528) та Gradient Boosting (0,9526), що пояснюється схильністю цих моделей до перенавчання на окремих підвибірках. Logistic Regression (AUC = 0,9961) та LDA (AUC = 0,9971), незважаючи на нижчі значення F1-score, демонструють прийнятну роздільну здатність у просторі ймовірностей. Значення AUC = 1,0 для методу Naive Bayes зумовлено особливістю його апарату оцінювання апостеріорних ймовірностей: базове припущення про умовну незалежність ознак у поєднанні з наявними кореляційними зв'язками між компонентними ознаками призводить до

граничного розподілу ймовірностей у просторі класів, що формально відображається як ідеальна роздільна здатність. Проте нижчий рівень F1-score та відсутність стійкості крос-валідаційної оцінки не підтверджують практичної переваги цієї моделі над іншими класифікаторами.

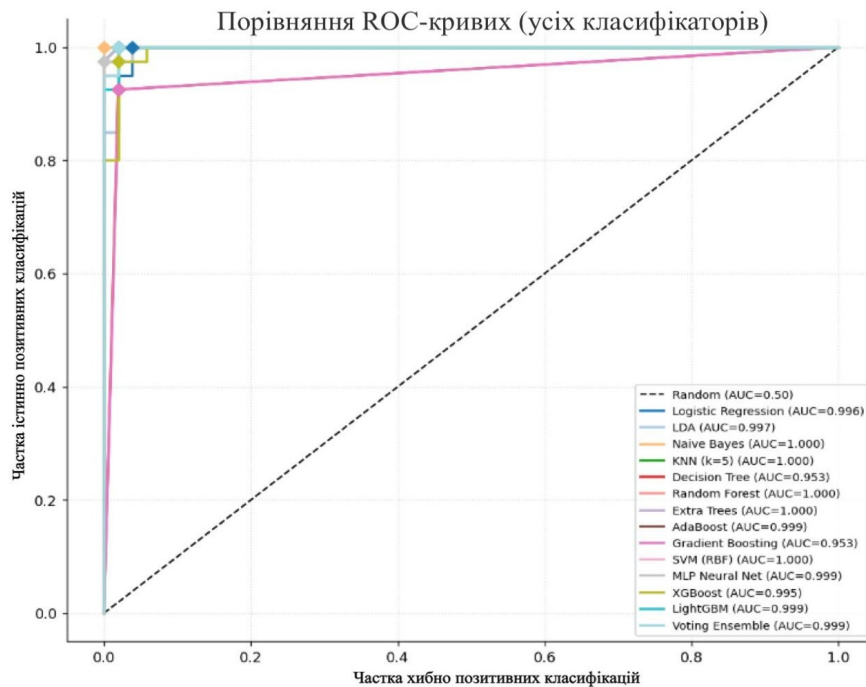


Рисунок 3.31 – Зведена ROC–крива для всіх моделей машинного навчання

Високі значення AUC ( $>0.99$ ) підтверджують здатність моделей ефективно розрізняти режими електроспоживання незалежно від порогу класифікації. Для визначення внеску кожної ознаки у процес класифікації проведено аналіз важливості на основі двох методів: MDI (Mean Decrease in Impurity) та Permutation Importance (рис. 3.32). Результати аналізу важливостей ознак підтверджують, що як глобальні, так і локальні характеристики процесу відіграють важливу роль у завданнях класифікації.

### Аналіз важливості ознак

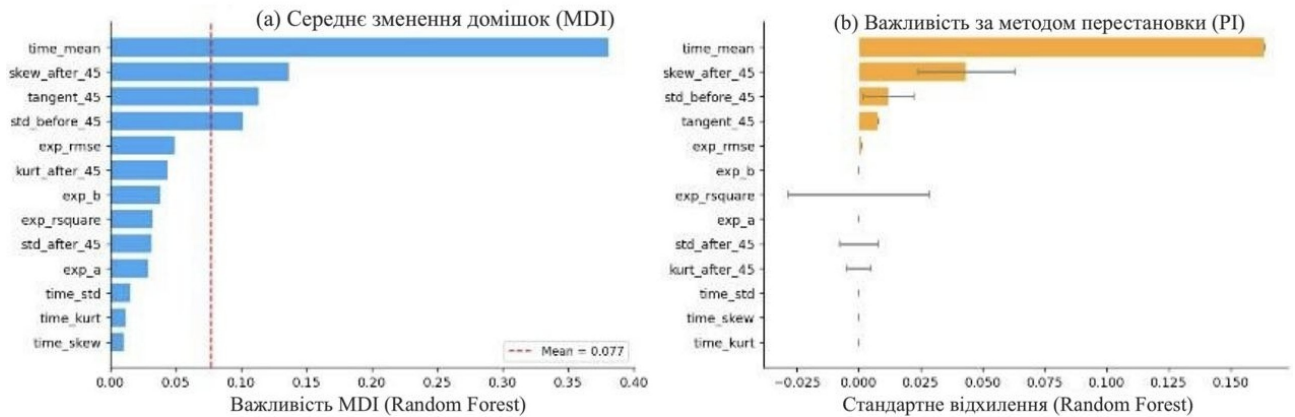


Рисунок 3.32 – Аналіз важливості ознак: а) середнє зменшення домішок (MDI, Random Forest); б) важливість за методом перестановки (Permutation Importance)

Наведені графіки (рис. 3.32) відображають важливість ознак у моделі Random Forest, але з двох принципово різних математичних підходів до її оцінки. Їх слід інтерпретувати не ізольовано, а у взаємозв'язку, оскільки кожен з них підкреслює різні аспекти впливу ознак на класифікацію.

У лівій частині наведено результати методу середнього зменшення нечистоти (MDI), який оцінює внесок кожної ознаки на основі зменшення неоднорідності вузлів дерева рішень у процесі навчання. Формально, у кожному вузлі дерева обчислюється зменшення критерію неоднорідності (наприклад, індексу Джині), і для кожної ознаки агрегується сумарний внесок у це зменшення по всіх деревах ансамблю. Відповідно, якщо ознака часто використовується для розбиття і суттєво покращує однорідність підвибірок, її важливість зростає.

На графіку чітко видно, що ознака `time_mean` має домінуючий внесок, значно перевищуючи середній рівень (позначений червоною пунктирною лінією), що означає її ключову роль у розділенні класів. Ознаки `skew_after_45`, `tangent_45` та `std_before_45` також мають значний вклад, тобто модель активно використовує як глобальні статистики, так і локальні характеристики після точки експоненційного перелому. Натомість ознаки на кшталт `time_skew` чи `time_kurt`

мають дуже малий внесок, що означає їхню низьку дискримінативну здатність у межах дерева.

Права частина відображає Permutation Importance, яка є більш строгою з точки зору причинно-наслідкової інтерпретації. Ідея полягає у тому, що значення однієї ознаки випадково перемішуються, руйнуючи її зв'язок із цільовою змінною, після чого вимірюється зниження точності моделі:

$$\Delta \text{Accuracy} = \frac{TP_{orig} + TN_{orig}}{TP_{orig} + TN_{orig} + FP_{orig} + FN_{orig}} - \frac{TP_{shuf} + TN_{shuf}}{TP_{shuf} + TN_{shuf} + FP_{shuf} + FN_{shuf}} \quad (3.15)$$

Чим більше падає точність, тим важливішою є ознака. Тут знову `time_mean` демонструє найбільший вплив, що підтверджує її фундаментальну роль. Проте на відміну від MDI, деякі ознаки змінюють свій ранг: наприклад, `std_before_45` та `tangent_45` залишаються важливими, але частина експоненційних параметрів (`exp_a`, `exp_b`, `exp_rsquare`) має майже нульовий або навіть нестабільний внесок (що видно по великих довірчих інтервалах). Це означає, що модель може використовувати їх у структурі дерев, але їх фактичний внесок у узагальнюючу здатність є обмеженим.

Особливо важливою є наявність похибок у permutation importance. Вони відображають стандартне відхилення ефекту при багаторазовому перемішуванні, і якщо вони великі або перекривають нуль, це означає, що важливість ознаки є статистично нестабільною. Наприклад, для `exp_rsquare` видно значну невизначеність, що вказує на її слабку та непослідовну роль у класифікації.

Узагальнюючи, лівий графік (MDI) показує, як модель використовує ознаки під час побудови дерев, тоді як правий (Permutation Importance) показує, наскільки модель залежить від цих ознак для досягнення точності. У даному випадку обидва підходи узгоджено вказують на те, що `time_mean` є домінуючою ознакою, тоді як більшість експоненційних параметрів мають другорядне або

допоміжне значення. Це свідчить про те, що глобальний рівень сигналу є ключовим для розділення класів «Збалансований» і «Незбалансований», тоді як складніші характеристики форми сигналу відіграють уточнюючу роль. Фінальний результат формується шляхом голосування: кожна модель виносить свій вердикт (збалансована / незбалансована), а підсумковий результат визначається за принципом більшості. Такий підхід забезпечує робастність класифікації та зменшує вплив хибних спрацювань окремих моделей.

Отже, результати дослідження підтверджують, що інформативність ознак, сформованих на основі компонентного методу ПКВП, є достатньою для надійної класифікації режимів електроспоживання незалежно від складності обраного алгоритму машинного навчання.

### **3.4. Висновки до розділу 3**

У даному розділі представлено результати експериментального аналізу даних електроспоживання приватного підприємства. На основі компонентного методу періодично корельованих випадкових процесів проведено структурний аналіз сигналу, сформовано ознаковий простір та виконано порівняльне дослідження 14 алгоритмів машинного навчання для класифікації збалансованого та незбалансованого режимів електроспоживання вузла розподіленої електроенергетичної системи.

1. На основі компонентного методу аналізу ПКВП сформовано набір із 13 інформативних ознак (9 параметрів при усередненні за компонентами та 4 при усередненні за часом), який забезпечує перехід від часових реалізацій електроспоживання до параметричного представлення, що відображає як часову динаміку добового профілю, так і кореляційну структуру між гармонічними складовими процесу.

2. Статистичний аналіз виділених ознак (KDE–оцінки розподілів, діаграми розмаху, матриця кореляцій Пірсона) підтвердив їх високу

дискримінантну здатність. Встановлено, що найбільш виражене розділення класів забезпечують ознаки `time_mean` (середній рівень енергії), `skew_after_45` (інтенсивність флуктуацій після критичної точки), `tangent_45` (часова локалізація точки перелому), `std_before_45` (стандартне відхилення сигналу до точки перелому), для яких розподіли мають мінімальне перекриття між збалансованим та незбалансованим режимами.

3. Проведено класифікацію режимів електроспоживання із застосуванням 14 моделей машинного навчання різних парадигм: лінійних (Logistic Regression, LDA), ймовірнісних (Naive Bayes), метричних (KNN), деревоподібних (Decision Tree, Random Forest, Extra Trees), бустингових (AdaBoost, Gradient Boosting, XGBoost, LightGBM), нейромережових (MLP) та ансамблевих (Voting Ensemble). Продемонстровано, що усі моделі забезпечують високу якість класифікації ( $F1 > 0,93$ ,  $MCC > 0,88$ ), при цьому найкращі результати ( $F1=0,9877$ ) досягнуто для моделей KNN ( $k=5$ ), Extra Trees та SVM (RBF). Extra Trees демонструє найвищу крос-валідаційну стійкість ( $CV-F1=0,9826$ ), що обґрунтовує його рекомендацію для практичного застосування.

4. Проведено аналіз важливості ознак методами MDI та Permutation Importance та показано, що домінуючу роль у розділенні класів відіграє глобальний рівень сигналу (`time_mean`), тоді як характеристики в зоні точки експоненціального перелому `std_before_45` (до точки), `tangent_45` та `skew_after_45` (після точки) виконують уточнюючу функцію, що підтверджує фізичну обґрунтованість запропонованого ознакового простору.

5. Висока якість класифікації незалежно від типу алгоритму свідчить про те, що дискримінантна здатність забезпечується інформативністю ознак, сформованих на основі компонентного методу ПКВП, а не складністю конкретної моделі машинного навчання, що дозволило підтвердити доцільність практичного застосування запропонованого підходу в завданнях класифікації режимів електроспоживання в розподілених електроенергетичних системах.

## РОЗДІЛ 4. РОЗРОБКА ТА РЕАЛІЗАЦІЯ ПРОГРАМНОГО МОДУЛЯ АНАЛІЗУ ДАНИХ ЕЛЕКТРОСПОЖИВАННЯ

### 4.1. Обґрунтування архітектури та вибору програмних засобів

У даному розділі наведено опис розробленої програмної реалізації запропонованого модуля, який дозволяє провести аналіз даних електроспоживання з використанням моделі ПКВП для формування ознакового простору, класифікації режимів електроспоживання, а також процес вибору протоколів передавання даних на основі моніторингу мережевих метрик. [128,129].

Запропонований підхід базується на таких основних принципах:

1. Використання моделі у вигляді періодично корельованих випадкових процесів для опису циклічної та стохастичної природи електроспоживання і формування інформативних ознак [43, 49, 50].

2. Застосування методів машинного навчання для класифікації режимів електроспоживання на основі сформованого ознакового простору [130,131,132,133].

3. Організація послідовного опрацювання даних, що забезпечує перехід від вхідних даних вимірювань до структурованого ознакового представлення [129,130].

Загальна логіка роботи системи включає такі етапи: отримання даних електроспоживання, вибір протоколу передавання даних з урахуванням характеристик мережі (затримки, втрати, пропускну здатність) для забезпечення якості вхідної інформації, застосування методів ПКВП, формування інформативних ознак, класифікацію режимів електроспоживання за допомогою методів машинного навчання, формування результатів та їх відображення [134,135].

Основні етапи опрацювання та аналізу даних електроспоживання наведені на рисунку 4.1, поєднують методи ПКВП та машинного навчання і реалізовані в програмному модулі.

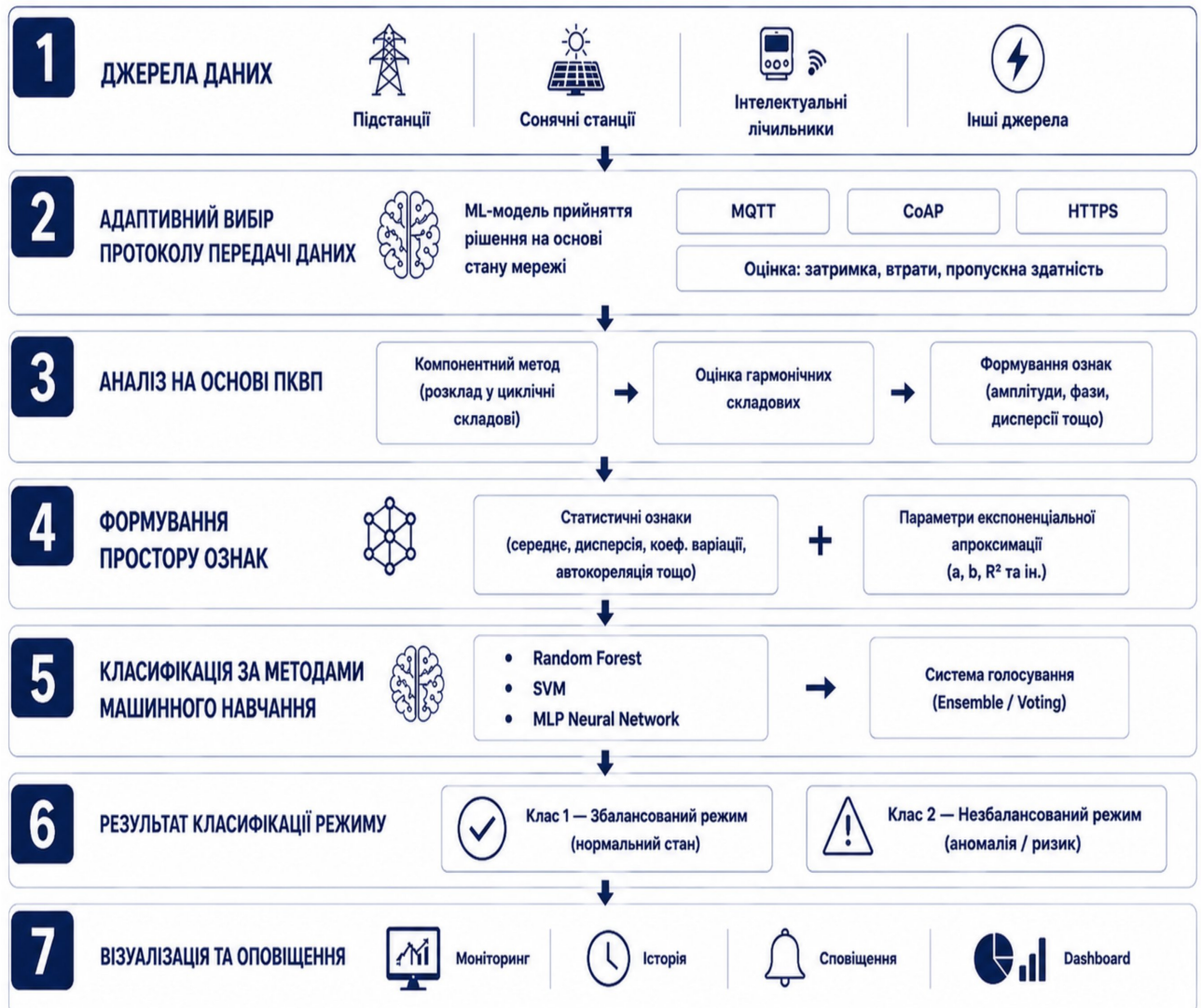


Рисунок 4.1 – Етапи опрацювання та аналізу даних електроспоживання

На початковому етапі використовуються “сирі” дані електроспоживання, які надходять від джерел вимірювання та подаються на подальше опрацювання. Перед аналітичним етапом здійснюється вибір протоколу передавання даних з урахуванням характеристик мережі, зокрема затримок, втрат та пропускну здатності [136]. Це дозволяє забезпечити якість вхідних даних, що є важливим для подальшого аналізу. Подальше опрацювання даних здійснюється на основі

моделі у вигляді періодично корельованих випадкових процесів з метою формування набору інформативних ознак.

Сформований набір ознак подається на вхід моделей машинного навчання, які виконують класифікацію режимів електроспоживання. Результатом є визначення стану системи [132, 133], зокрема збалансованого або незбалансованого режиму. Результатом класифікації є віднесення стану системи до одного з двох класів:

Клас 1 – Збалансований режим, що відповідає нормальному функціонуванню системи зі стабільною динамікою зміни та низькою варіативністю.

Клас 2 – Незбалансований режим, що характеризується наявністю аномалій, підвищеною стохастичною мінливістю та потенційною нестабільністю [131].

Стан системи відображається на інформаційній панелі оператора для подальшого прийняття рішень. Для апробації та верифікації запропонованих рішень розроблено програмний модуль, який реалізує всі описані функціональні рівні у єдиному інтегрованому середовищі.

Вибір технологічного апарату та засобів дослідження базується на необхідності вирішення завдань аналізу нестационарних сигналів електроспоживання з подальшою класифікацією режимів електроспоживання електроенергетичної системи. З огляду на це, у якості базового інструменту реалізації обрано мову програмування Python. Її використання дозволяє суттєво скоротити час розробки за рахунок застосування вже апробованих алгоритмічних рішень, забезпечуючи при цьому гнучкість, читабельність коду та можливість інтеграції з низькорівневими мовами для оптимізації обчислень. Для реалізації числових обчислень та опрацювання даних електроспоживання використано бібліотеки NumPy та Pandas, які забезпечують ефективну роботу з багатовимірними масивами часових рядів вимірювань [130].

Для формування інформативного ознакового простору реалізовано процедуру експоненційної апроксимації матриці компонентного усереднення, параметри якої включаються до вектора ознак разом із статистичними характеристиками сигналу [129].

Модулі машинного навчання побудовано засобами бібліотеки `scikit-learn`, яка забезпечує уніфіковану реалізацію різних парадигм класифікації: лінійних (Logistic Regression, LDA), метричних (`k-Nearest Neighbors`), деревоподібних (Decision Tree, Random Forest, Extra Trees), бустингових (AdaBoost, Gradient Boosting), методу опорних векторів (SVM) та ймовірнісних (Naive Bayes) моделей. Вибір бібліотеки `scikit-learn` обумовлений її уніфікованим API для навчання, крос-валідації та інференсу, а також достатньою продуктивністю при розв'язанні завдань класифікації режимів електроспоживання.

Додатково використано бібліотеки `XGBoost` та `LightGBM` як оптимізовані реалізації градієнтного бустингу, що забезпечують високу продуктивність при роботі з табличними ознаковими просторами. Для реалізації багат шарового перцептрона застосовано модуль `MLPClassifier` бібліотеки `scikit-learn` з адаптивним оптимізатором `Adam`, що забезпечує ефективну оптимізацію параметрів моделі в процесі навчання для завдань класифікації режимів електроспоживання.

Архітектурно систему побудовано на основі фреймворку `Django` (Python 3.13) з модульною організацією коду, що дозволяє забезпечити баланс між простотою реалізації та масштабованістю програмної системи. Вибір `Django` обумовлений наявністю вбудованого ORM для роботи з базою даних, модуля `Django REST Framework` для реалізації REST API у форматі JSON, а також підтримкою модульної організації коду, що дозволяє незалежно розробляти та тестувати аналітичні, комунікаційні та інтерфейсні компоненти системи. Основний аналітичний модуль реалізовано як окремий компонент у складі проєкту, що дозволяє незалежно розвивати та тестувати кожен функціональний

блок. Взаємодія між модулями здійснюється через внутрішній API на базі Django REST Framework у форматі JSON.

Для зберігання проміжних та результуючих даних використовується вбудована база даних SQLite, що повністю відповідає потребам дослідницького прототипу з локальним розгортанням. Графічний інтерфейс користувача забезпечує інтерактивну візуалізацію результатів аналізу та доступний через локальний браузер без необхідності встановлення додаткового клієнтського програмного забезпечення. Контейнеризацію середовища забезпечено засобами Docker, що гарантує відтворюваність обчислювального оточення та спрощує розгортання застосунку на різних платформах.

З огляду на зазначені вимоги до точності, інтерпретованості та обчислювальної ефективності алгоритмів класифікації, обрана сукупність програмних засобів та архітектурних рішень забезпечує ефективну реалізацію інформаційної технології аналізу режимів електроспоживання. Запропонована комбінація алгоритмів дозволяє враховувати як лінійні, так і нелінійні закономірності у сформованому ознаковому просторі, що забезпечує високу якість класифікації збалансованих та незбалансованих режимів роботи електроенергетичної системи.

Слід зазначити, що обрана архітектура є масштабованою: при переході від експериментального до промислового впровадження SQLite може бути замінено на PostgreSQL, локальний сервер на хмарне або Edge–Cloud розгортання, а REST API забезпечує безшовну інтеграцію з зовнішніми системами диспетчерського керування (SCADA) та платформами Smart Grid без модифікації ядра аналітичних алгоритмів. Це створює передумови для подальшого розвитку системи моніторингу електроспоживання та інтеграції нових алгоритмічних рішень без необхідності суттєвої модифікації ядра програмної системи.

#### **4.2. Адаптивний вибір оптимального протоколу передавання даних на основі методів машинного навчання**

Ефективність класифікації режимів електроспоживання, описаної у попередньому підрозділі, безпосередньо залежить від якості та повноти вхідних даних, що надходять від інтелектуальних лічильників, підстанцій тощо [130]. З огляду на архітектурні рішення, описані у підрозділі 4.1, критичним елементом є забезпечення якості вхідних даних через адаптивний вибір протоколу передавання.

В умовах реальної інфраструктури електроенергетичних мереж якість передавання даних визначається обраним комунікаційним протоколом [137,138]. Використання неоптимального протоколу може призводити до втрати пакетів, збільшення затримок та зниження частоти оновлення даних, що у свою чергу погіршує точність формування ознак ПКВП та результати класифікації [135, 137].

Впровадження єдиної системи стандартів та адаптивних протоколів дозволяє мінімізувати накладні витрати на трафік, забезпечуючи масштабованість системи та цілісність даних [135]. Це створює передумови для формування єдиного інформаційного середовища, у якому дані від кінцевих споживачів до центрів прийняття рішень передаються без втрати семантичного змісту. Такий підхід дозволяє нівелювати ризики інформаційної асиметрії на ринках електроенергії, запобігаючи економічним збиткам від недостовірного визначення режиму електроспоживання системи та гарантуючи стабільність функціонування електроенергосистеми. Адже наявні інфраструктури мають різні типи систем і пристроїв електроенергетичних мереж вибір комунікаційного протоколу є критичним фактором [136, 137], що впливає на якість, повноту та своєчасність отримання даних електроспоживання.

На рис. 4.2 представлено схему передавання даних електроспоживання, яка відображає взаємодію між джерелами вимірювання, мережевою інфраструктурою, яка відповідає за збір даних і подальше їх використання.

Збір даних здійснюється на рівні підстанцій та енергетичних об'єктів (зокрема сонячної станції) за допомогою інтелектуальних лічильників, які передають інформацію через L2-комутатори до локальних шлюзів. Для сегментації трафіку використовується технологія VLAN, що забезпечує логічне розділення даних між різними об'єктами.



Рисунок 4.2 – Запропонована схема передавання даних електроспоживання

Передавання даних електроспоживання здійснюється через модеми/RTU із використанням захищених каналів зв'язку VPN (IPsec/SSL), що забезпечують надійність та безпеку обміну. Отримані дані передаються до системи, де реалізовано їх зберігання, опрацювання та подальший аналіз для підтримки

прийняття рішень. Отримані результати відображаються на робочому місці оператора, що забезпечує моніторинг стану енергосистеми та прийняття оператором рішень щодо її функціонування.

Запропонована архітектура забезпечує масштабованість, захищеність та надійність передавання даних у розподілених енергетичних системах [135]. Необхідність адаптивного вибору протоколу передавання даних електроспоживання зумовлена також сучасними вимогами до інформаційно–комунікаційної взаємодії в рамках функціонування ОЕС України в синхронному режимі з ENTSO–E, коли надійність передавання телеметричної інформації є надзвичайно важливою [14].

У роботі розглянуто протоколи MQTT, CoAP та HTTPS як найпоширеніші для передавання даних у системах моніторингу [134, 135]. MQTT є протоколом типу publish/subscribe, що забезпечує баланс між швидкістю та надійністю передавання [137]. CoAP орієнтований на мінімальні затримки та низькі накладні витрати; у стандартному режимі використовує UDP без гарантії доставки, однак підтримує режим Confirmable для надійної передачі з підтвердженням. HTTPS – забезпечує високий рівень надійності та безпеки, проте має більші затримки.

Вибір між протоколами визначається умовами каналу зв'язку, вимогами до надійності та допустимими затримками у конкретній конфігурації системи моніторингу.

Кожен з розглянутих протоколів має визначені специфікацією робочі характеристики, при дотриманні яких забезпечується його ефективне функціонування. Вибір цих протоколів зумовлений тим, що вони працюють на прикладному рівні та безпосередньо використовуються для передавання даних у системах на базі IP. Типові характеристики протоколів, що використовуються як базові, наведено у таблиці 4.1 [137, 138].

Адаптивний вибір протоколу базується на неперервному моніторингу мережевих метрик та порівнянні їх з типовими значеннями [139, 140]. Датасет мережевих метрик сформовано на основі даних, зареєстрованих від IoT-

пристроїв приватного підприємства, що використовують протоколи MQTT, CoAP та HTTPS для передавання даних електроспоживання в умовах реальної інфраструктури.

Таблиця 4.1

Типові характеристики протоколів передавання даних.

Характеристика	MQTT	CoAP	HTTPS
Затримка, мс	10–30	5–25	50–150
Пропускна здатність, Мбіт/с	1–5	0.5–4	5–15
Втрати пакетів, %	< 0.3	< 0.2	< 1.5
Розмір корисного навантаження, байт	20–300	10–200	500–3000
Джитер (Jitter)	0.2–0.5	0.05–0.25	0.3–0.7

Для кожного інтервалу агрегації аналізується набір параметрів, який включає середню затримку (latency), пропускну здатність (throughput), рівень втрат пакетів (packet loss), середній розмір корисного навантаження (payload size) та джитер (jitter, як показник варіативності затримки).

Ці параметри порівнюються з типовими діапазонами поточного протоколу. Якщо всі значення знаходяться в межах норми, тоді система продовжує роботу без змін. У разі виявлення відхилень від нормативних діапазонів класифікатор формує рекомендацію щодо перемикавання на більш відповідний протокол передавання даних.

Для оцінки ефективності запропонованого підходу було досліджено чотири моделі машинного навчання, що відрізняються характеристиками навчання та узагальнювальної здатності.

Навчання моделей виконувалося на датасеті, сформованому на основі даних електроспоживання підприємства. Результати порівняння моделей наведено у таблиці 4.2.

Таблиця 4.2

Результати порівняння моделей класифікації для вибору протоколу  
передавання даних

Модель	Accuracy	F1-score	Precision	Recall
Logistic Regression	0,7277	0,7256	0,7277	0,7260
Gradient Boosting	0,8207	0,8197	0,8207	0,8201
Random Forest	0,8717	0,8711	0,8717	0,8710
MLP Neural Network	0,9023	0,9020	0,9023	0,9021

Аналіз результатів засвідчує, що модель MLP Neural Network (NN) демонструє найвищу точність класифікації (90,2%), що обумовлено її здатністю моделювати нелінійні залежності між ознаками через приховані шари мережі (архітектура 64–32 нейрони). Модель Random Forest (RF) показує другий за точністю результат (87,2%), що пояснюється ефективністю ансамблевого підходу у роботі з перекриттям класів.

Logistic Regression (LR) демонструє найнижчу точність (72,8%) серед досліджених моделей, що є очікуваним результатом: лінійна модель не здатна адекватно розділити класи MQTT та CoAP, які мають значне перекриття у просторі ознак. Отримані результати підтверджують доцільність застосування ансамблевих методів машинного навчання для завдань адаптивного вибору протоколу передавання даних електроспоживання.

Для детальнішого аналізу якості класифікації побудовано матриці плутанини та ROC-криві для кожної з досліджених моделей. Матриці наведено на рис. 4.3–4.6.

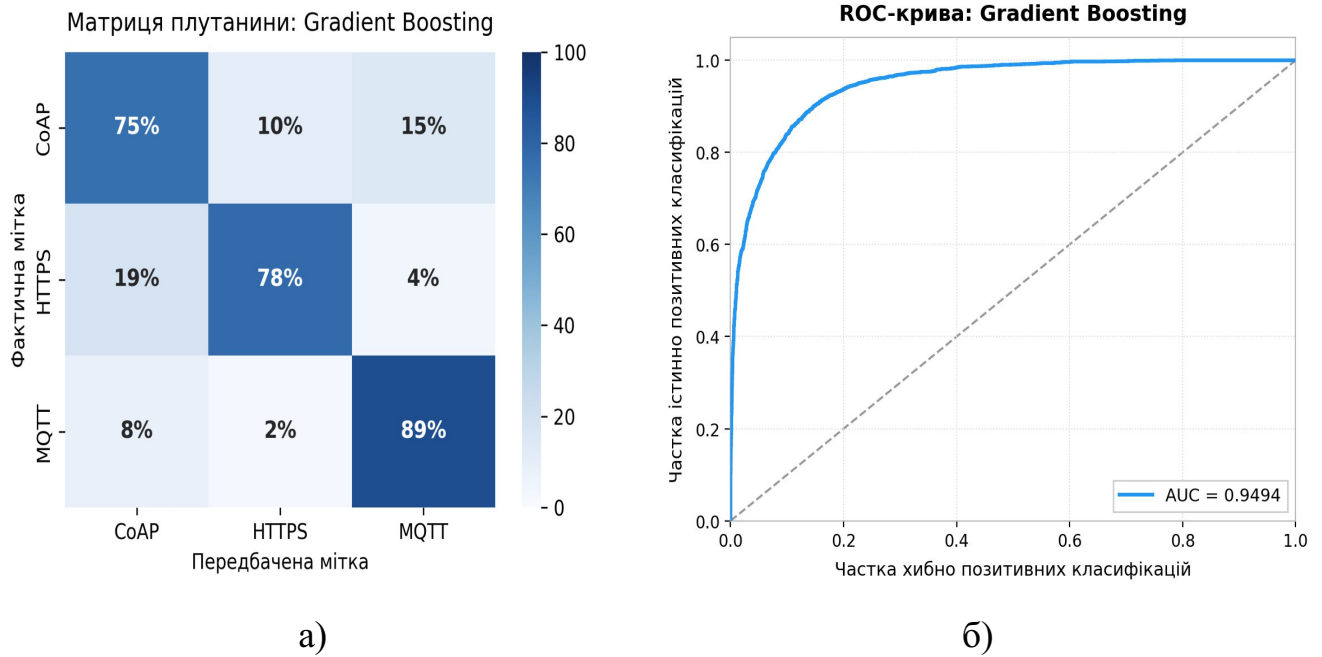


Рисунок 4.3 – Модель Gradient Boosting:

а) матриця плутанини; б) ROC-крива

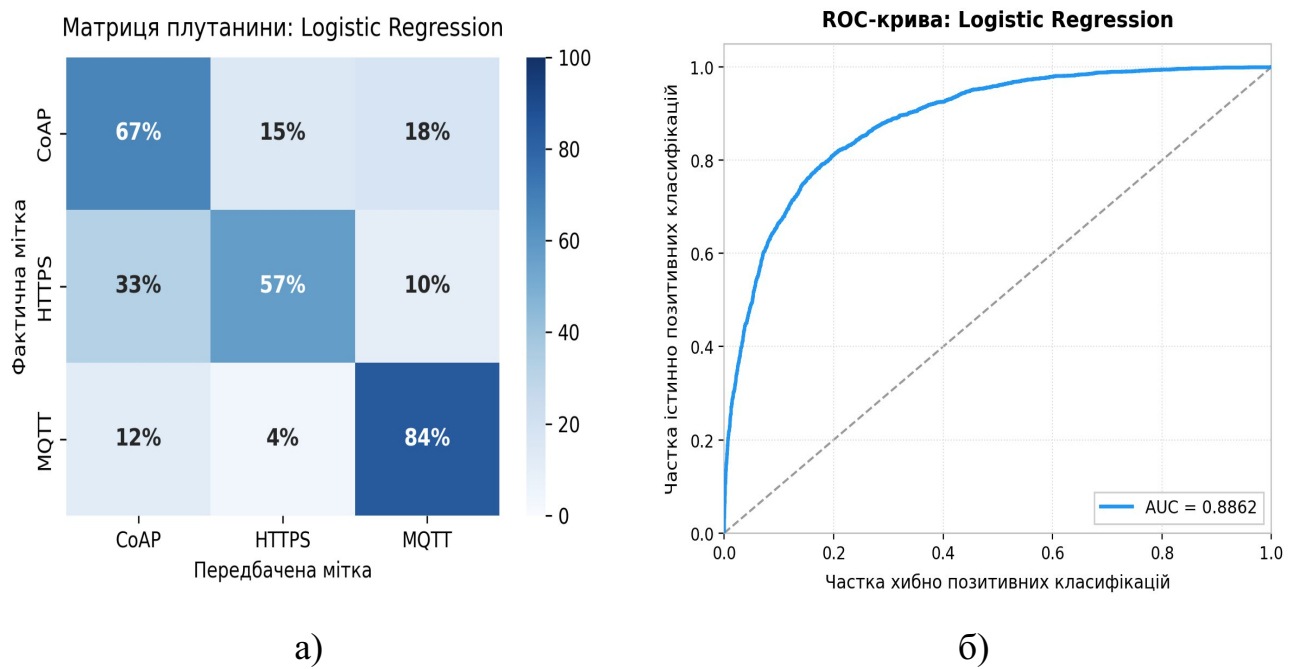


Рисунок 4.4 – Модель Logistic Regression:

а) матриця плутанини; б) ROC-крива

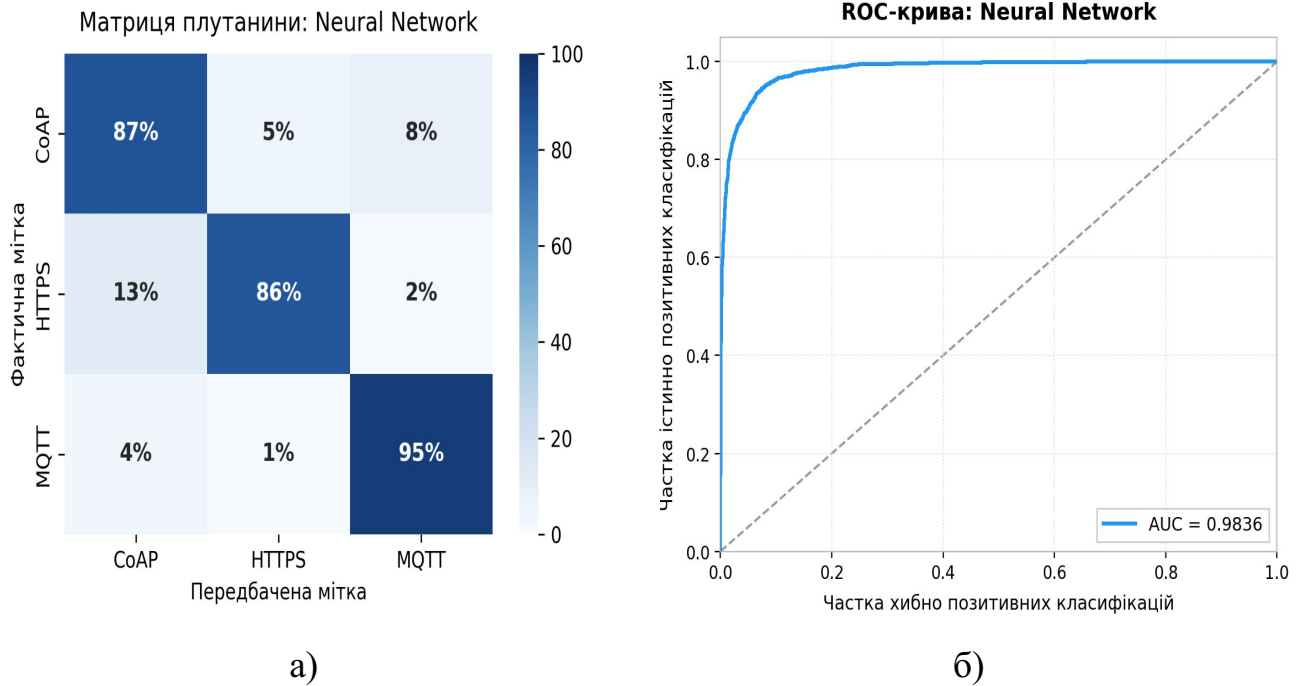


Рисунок 4.5 – Модель Neural Network:

а) матриця плутанини; б) ROC-крива

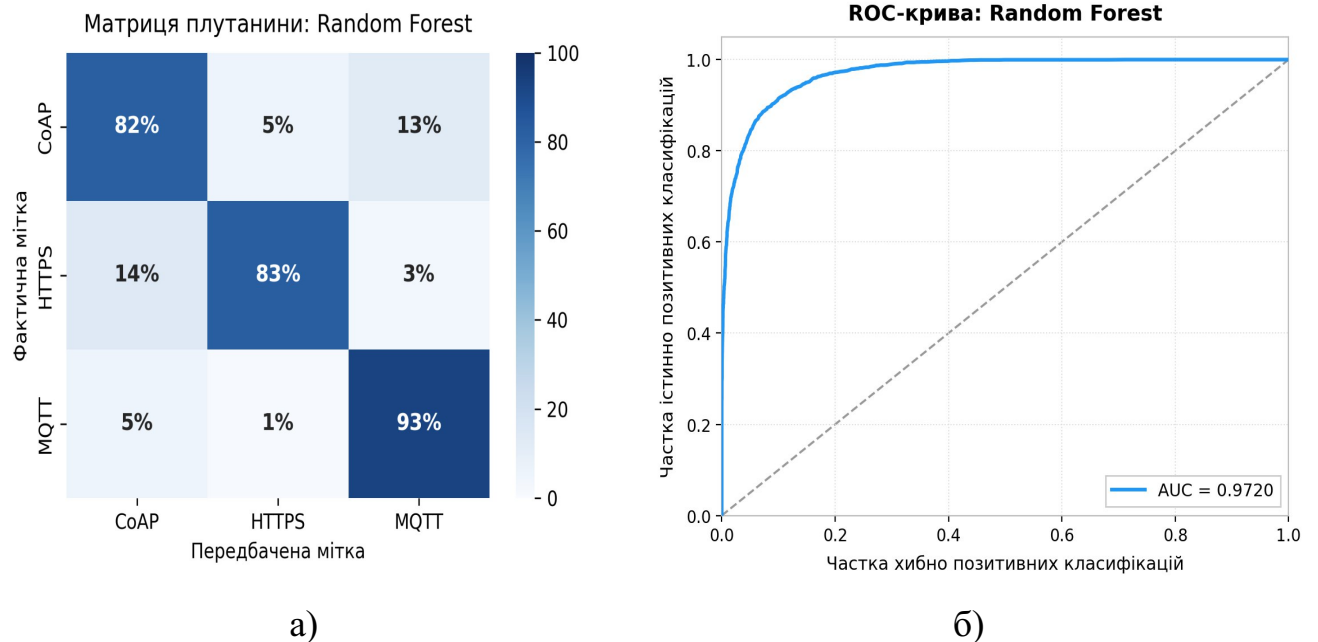


Рисунок 4.6 – Модель Random Forest:

а) матриця плутанини; б) ROC-крива

Аналіз матриць плутанини (рис. 4.3–4.6) виявив характерну закономірність у розподілі помилок класифікації: основна частина хибних спрацювань

зосереджена між протоколами CoAP та MQTT, що зумовлено схожістю їх поведінки в умовах обмеженої пропускної здатності каналу. Протокол HTTPS стабільно відокремлюється від CoAP завдяки принципово різній архітектурі та накладним витратам, однак у режимах інтенсивного навантаження його латентні характеристики можуть зближуватися з профілем MQTT, спричиняючи локальні помилки класифікації.

Для інтерпретації результатів класифікації та обґрунтування вибору ознакового простору проведено аналіз важливості ознак з використанням двох методів: середнього зменшення домішок (MDI) та важливості за методом перестановки (PI). Для цього аналізу використано модель Random Forest як інтерпретовану ансамблеву модель, що дозволяє безпосередньо оцінити внесок кожної ознаки незалежно від фінальної класифікуючої моделі. Результати аналізу наведено на рисунку 4.7.

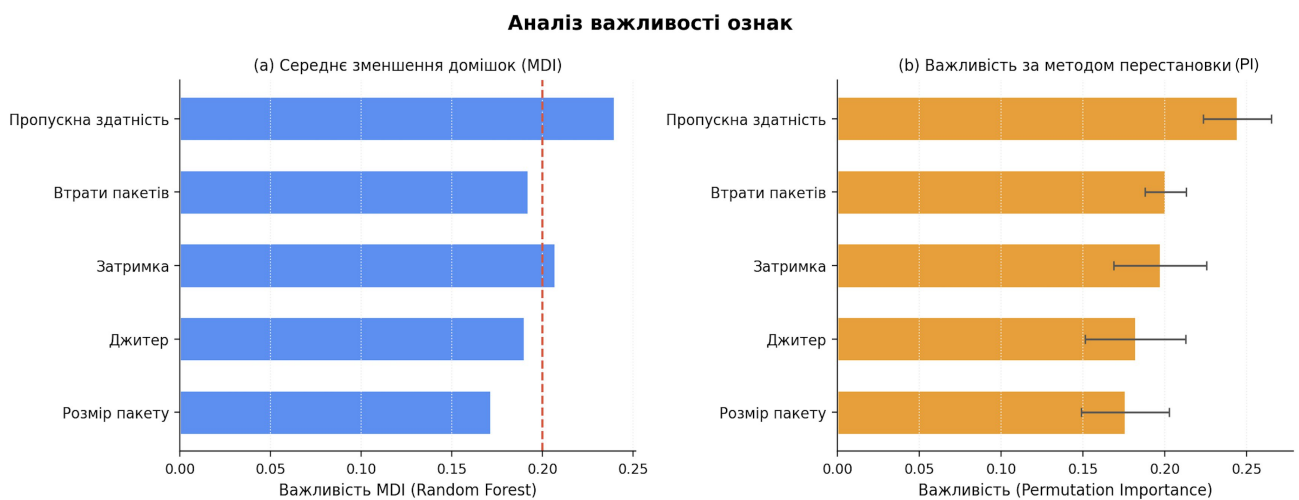


Рисунок 4.7 – Важливість ознак для вибору оптимального протоколу

За обома методами найвищу важливість демонструють пропускна здатність, за нею втрати пакетів та затримка, тоді як джитер і розмір пакета мають нижчі значення. Результати MDI та PI демонструють узгоджену ієрархію важливості ознак, що підтверджує стійкість отриманих оцінок. Рівномірний розподіл важливості між усіма п'ятьма ознаками свідчить про те, що випадіння

однієї з них не призводить до суттєвого зниження точності класифікації. Підвищений розкид значень  $PI$  для ознаки розміру пакета вказує на її нестабільну інформативність залежно від умов вибірки.

Порівняння моделей свідчить про суттєву перевагу ансамблевих підходів: Neural Network та Random Forest демонструють значно рівномірніше заповнення діагональних елементів матриці при мінімальних позадіагональних значеннях порівняно з Gradient Boosting та Logistic Regression. Відповідні ROC-криві підтверджують цю тенденцію вищі значення AUC для Neural Network та Random Forest вказують на кращу роздільну здатність цих моделей у просторі ймовірностей незалежно від обраного порогу класифікації.

Отримані результати аналізу та класифікації потребують зручного представлення для подальшої інтерпретації та оцінки стану системи. З цією метою у межах роботи реалізовано інтерфейс програмного модуля адаптивного вибору протоколу (модуль АВП), який забезпечує візуалізацію основних характеристик та результатів опрацювання даних.

На вході модуль отримує п'ять характеристик мережевого каналу, які подаються на ML-класифікатор, обраний за критерієм зваженого F1-score. На виході формується рекомендація щодо оптимального протоколу передавання даних, яка відображається на панелі оператора разом з оцінками важливості ознак та моніторингом стану каналів зв'язку.

На рис. 4.8 наведено загальну схему роботи модуля адаптивного вибору протоколу передавання даних. Запропонований підхід дозволяє враховувати поточний стан мережевого каналу та на основі сукупності параметрів визначати доцільний протокол передавання даних. Це забезпечує більш стабільне передавання даних та зменшує вплив мережевих обмежень на якість подальшого аналізу електроспоживання.

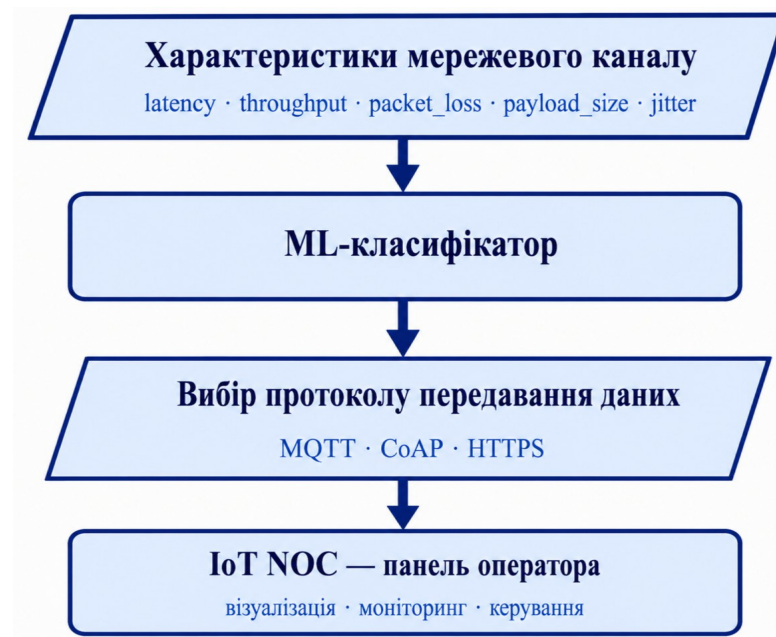


Рисунок 4.8 – Функціональна схема модуля адаптивного вибору протоколу передавання даних

Використання моделей машинного навчання у завданнях вибору протоколу дозволяє адаптувати передавання даних до змінних умов функціонування системи. З огляду на це, у розробленому програмному модулі адаптивного вибору протоколу передавання даних, завданням є забезпечення максимальної повноти та своєчасності надходження даних електроспоживання для подальшого аналізу. Модуль АВП здійснює моніторинг актуальних характеристик мережевого каналу та на основі ML-аналізу визначає, який з доступних протоколів є найвідповіднішим за поточних умов. Інтерфейс програмного модуля АВП візуалізує результати вибору протоколів передавання даних у зручному для користувача вигляді. Інтерфейс модуля класифікації наведено на рис. 4.9.

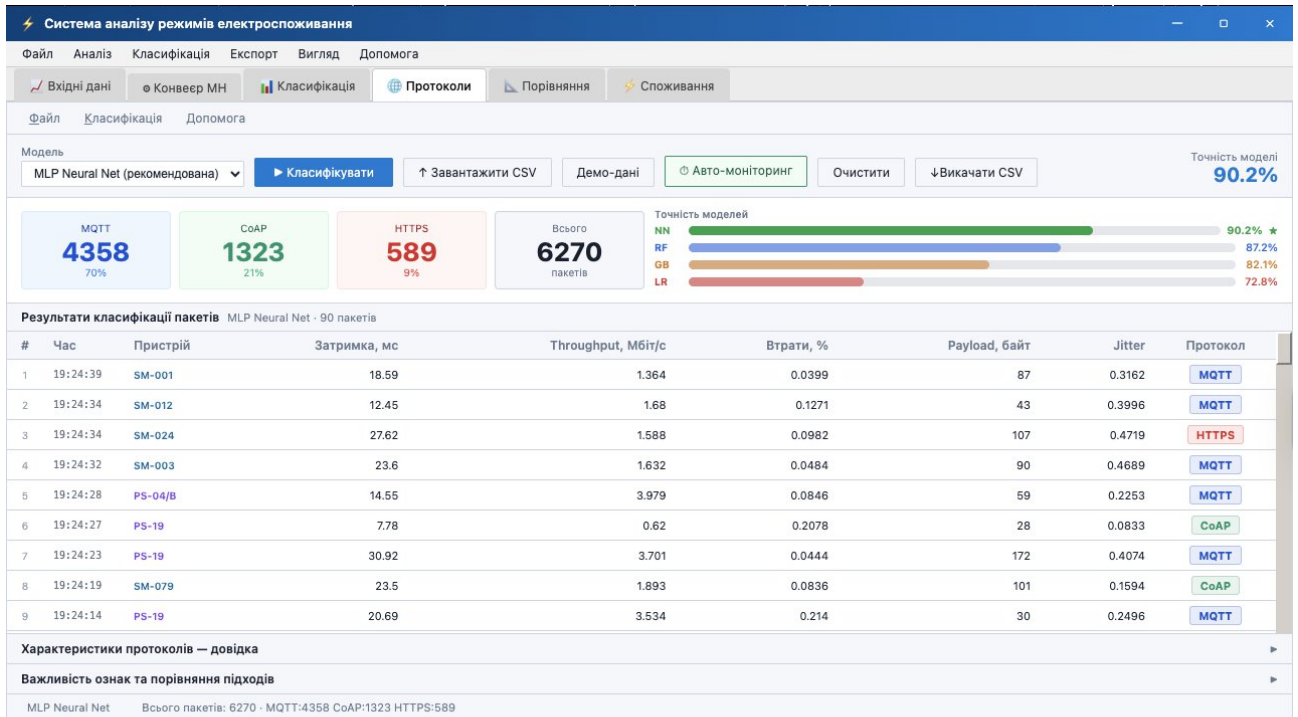


Рисунок 4.9 – Інтерфейс модуля класифікації з деталізацією даних по протоколах

На рис. 4.9 представлено інтерфейс модуля АВП, що відображає розподіл класифікованих пакетів за протоколами, загальну точність обраної моделі (90,2%), а також деталізовану таблицю із зазначенням мережевих метрик визначеного протоколу.

На рис. 4.10 представлено інтерфейс модуля АВП, що відображає узагальнені показники стану мережі, графічне представлення мережевих характеристик протоколів, розподіл трафіку та журнал подій для оперативного моніторингу.

Центральна частина інтерфейсу містить графічне представлення ключових мережевих характеристик, зокрема затримки передавання даних та пропускну здатності для різних протоколів, що дозволяє наочно оцінити їх ефективність. Додатково наведено розподіл всього трафіку між протоколами, а також блок із зазначенням точності використаних методів машинного навчання. У нижній

частині інтерфейсу представлено детальну інформацію щодо параметрів кожного протоколу, включаючи значення затримки, обсяг трафіку та втрати пакетів, а також панель керування.

Праворуч розташовано журнал подій, який відображає зміни стану вузлів мережі та результати виконання команд, що забезпечує можливість оперативного моніторингу та аналізу роботи системи.

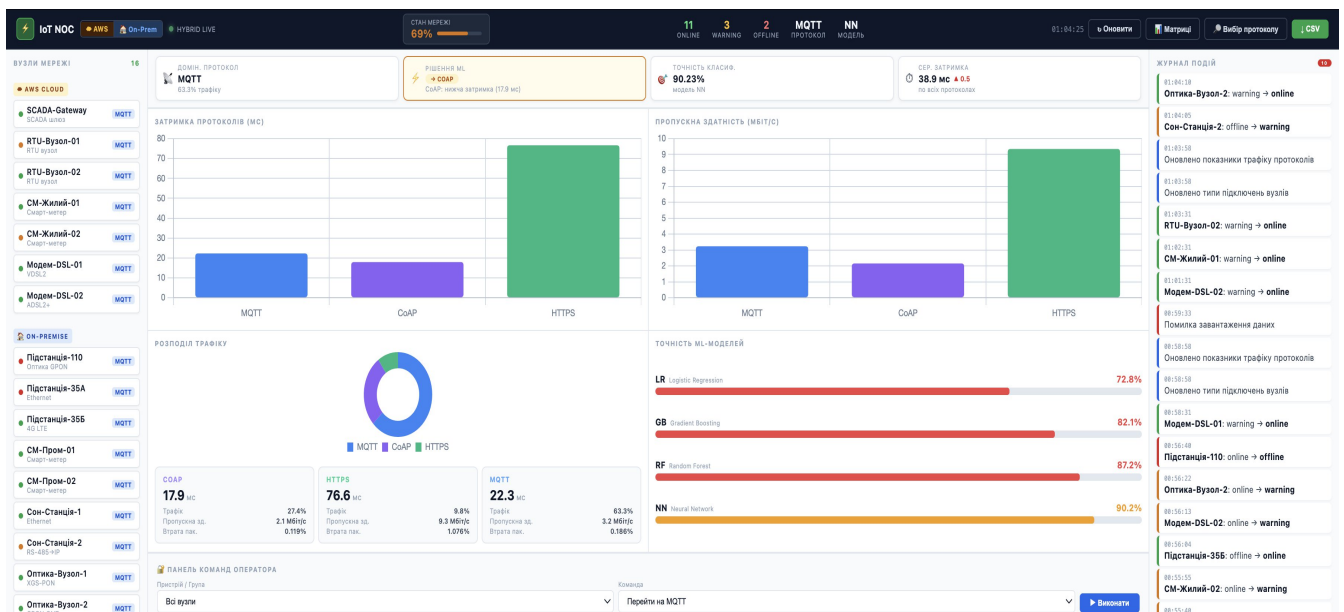


Рисунок 4.10 – Вікно візуалізації стану мережі

Таким чином, інтерфейс модуля АВП забезпечує комплексне відображення як мережевих параметрів, так і результатів вибору протоколу передавання даних відповідно до поточного стану мережі.

### 4.3. Програмний модуль аналізу даних електроспоживання та класифікації режимів електроспоживання

Програмний модуль реалізує опрацювання даних електроспоживання від моменту надходження даних електроспоживання з інтелектуальних лічильників через канали зв'язку до формування відповідних результатів щодо стану

електроенергетичної системи та параметрів комунікаційної інфраструктури [128,129,135].

Концепція базується на інтеграції трьох функціональних рівнів, збору та передавання даних, отримання інформативних ознак із застосуванням ПКВП, прийняття рішень про стан збалансованості на основі методів машинного навчання [130,131,133].

З урахуванням зазначених принципів та етапів на рис. 4.11 представлено загальну концепцію функціонування запропонованого підходу. Представлена схема відображає послідовний процес опрацювання даних у системі.

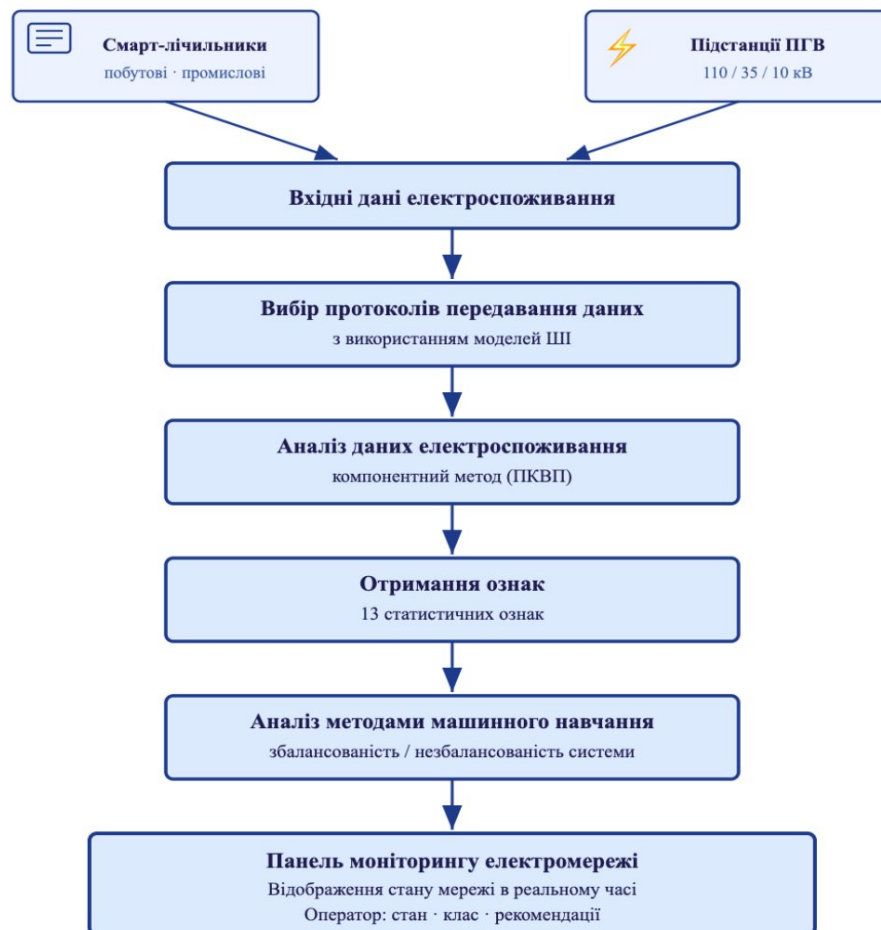


Рисунок 4.11 – Загальна концепція запропонованого аналізу електроспоживання в розподілених електроенергосистемах

На першому етапі збираються метрики з інтелектуальних лічильників, які опрацьовуються моделлю ПКВП для отримання компонентів (інформативних ознак), які є вхідними даними для моделей машинного навчання, що класифікують стан електроенергетичної системи (збалансований або незбалансований).

На верхньому рівні розташовані інтелектуальні лічильники, підстанції, з яких отримуються дані електроспоживання. Дані для аналізу надходять через канали зв'язку одним із адаптивно вибраних протоколів [135, 136, 137].

На аналітичному рівні виконується формування інформативних ознак на основі компонентного методу ПКВП. Отримані 13 ознак (9 від усереднення за компонентами та 4 від усереднення за часом) подаються на вхід моделей машинного навчання, які здійснюють класифікацію режиму електроспоживання електроенергосистеми [131,132,133]. Збалансований режим характеризується стабільною поведінкою електроспоживання без суттєвих відхилень від типових значень, тоді як незбалансований режим відображає наявність аномальних змін та нестабільності у процесі електроспоживання.

Модуль забезпечує опрацювання даних від їх завантаження в систему визначення режиму електроспоживання вузла. Інтерфейс користувача модуля побудований на основі вкладкової навігації, що забезпечує послідовний доступ до всіх етапів аналізу.

В інтерфейсі системи доступні такі вкладки: «Вхідні дані» для завантаження даних, «ПКВП» для налаштування аналізу, «Синфазний» і «Компонентний» для відображення результатів відповідних методів та «ЦВП» для результатів аналізу циклічних процесів.

Вкладка «Вхідні дані» надає оператору можливість завантажити файл із даними електроспоживання (формат TXT/CSV із стовпцями День, Година, кВт), обрати роздільник, задати кількість рядків для пропуску або згенерувати демо-матрицю для тестування. Представлені дані та графічні залежності (рис. 4.12) використано як тестовий приклад, програмний модуль підтримує завантаження

довільних часових рядів без обмежень на обсяг вибірки. Після завантаження система автоматично відображає таблицю даних та графік електроспоживання із зазначенням середнього значення.

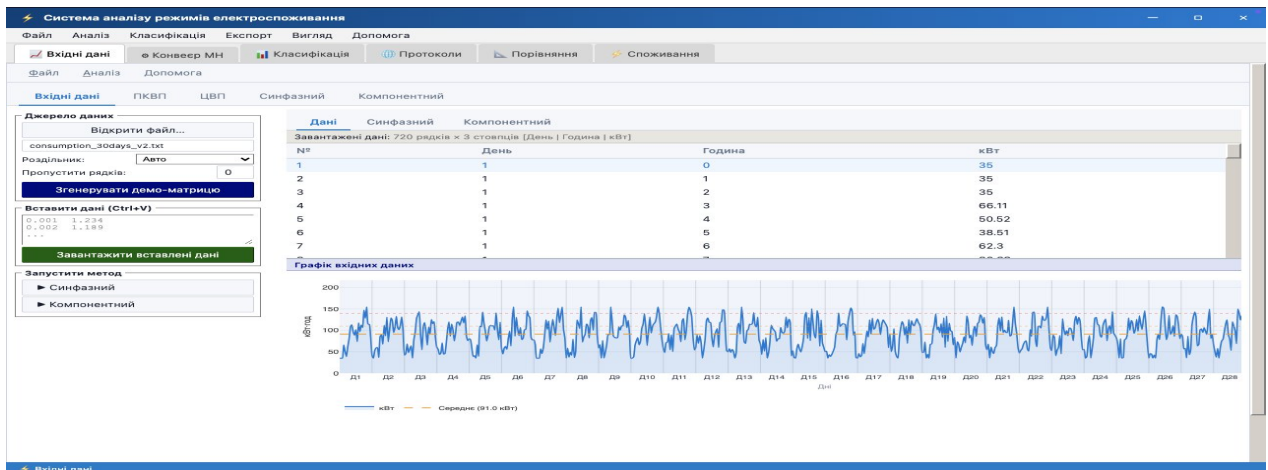


Рисунок 4.12 – Інтерфейсне вікно із завантаженими даними електроспоживання

Розроблений програмний модуль також забезпечує можливість комплексного опрацювання даних електроспоживання із застосуванням різних методів аналізу та візуалізацію результатів. На рис. 4.13–4.14 наведено приклади візуалізації результатів проведеного аналізу.

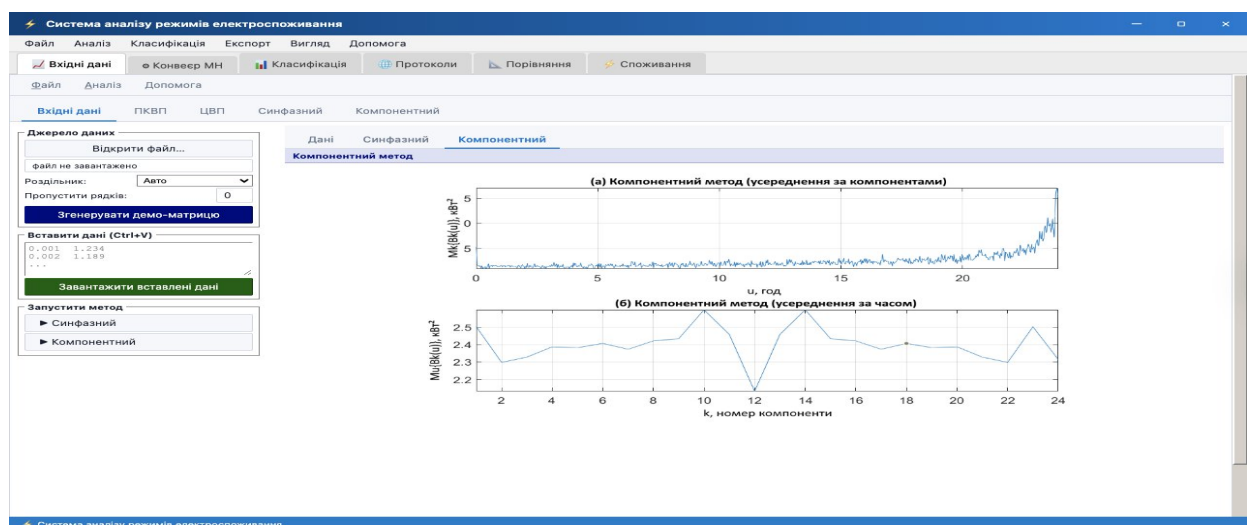


Рисунок 4.13 – Результати компонентного методу аналізу електроспоживання (усереднення за часом та компонентами)

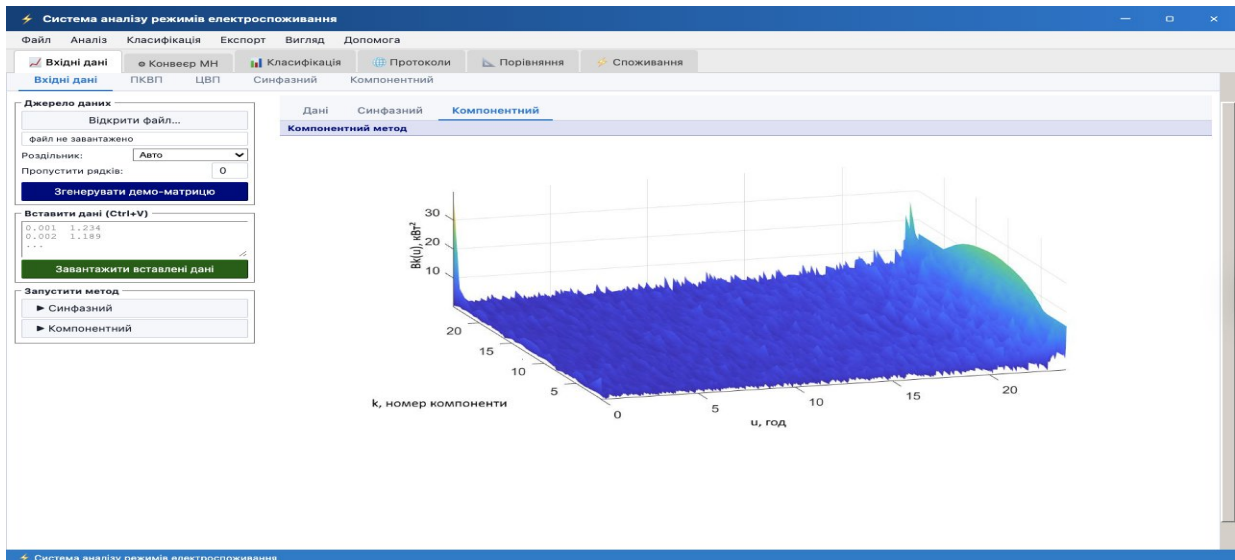


Рисунок 4.14 – Тривимірна візуалізація результатів компонентного методу аналізу електроспоживання

Сформований набір ознак подається на вхід моделей машинного навчання для класифікації режимів електроспоживання. Результати порівняння наведено на рис. 4.15.

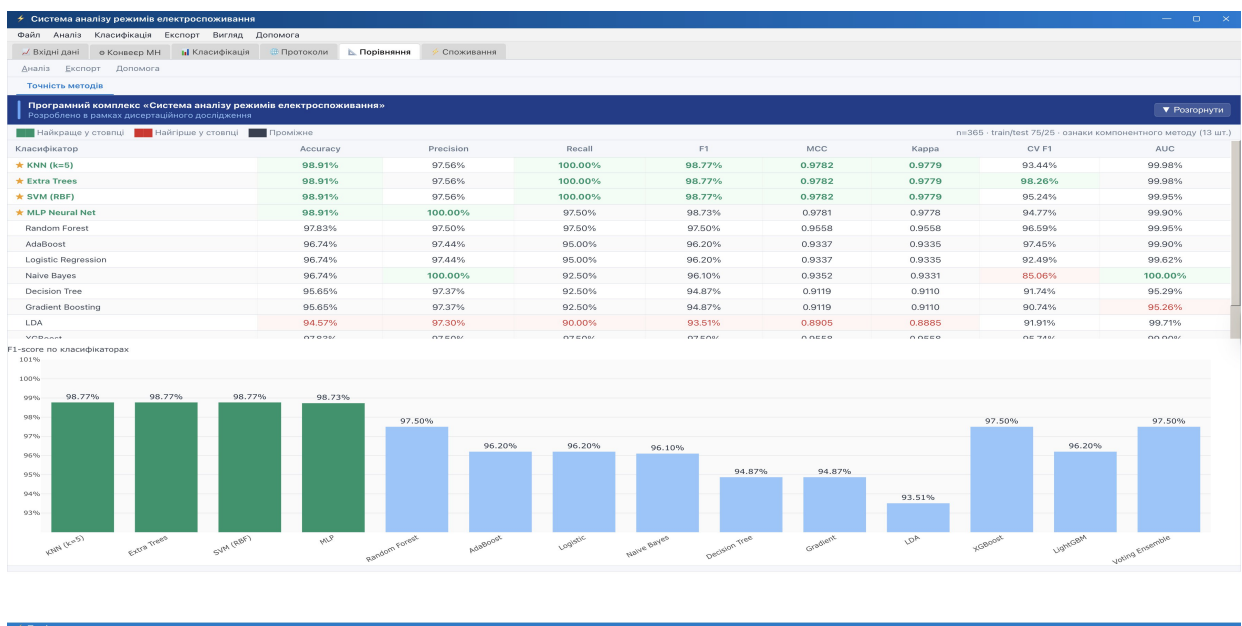


Рисунок 4.15 – Результати порівняння моделей машинного навчання для класифікації режимів електроспоживання

На рис. 4.16 представлено інтерфейс модуля визначення режиму електроспоживання для конкретного об'єкта. Ліва панель відображає значення всіх 13 статистичних ознак, центральна колонка відображає результати кожної з 14 моделей машинного навчання, а права – фінальне рішення. У випадку рівного розподілу голосів між класами остаточне рішення приймається на основі результату моделі з найвищим значенням F1-score.

Статистичні ознаки – ПС-110/35 «Промислова»		Результати класифікаторів		Фінальне рішення
Ознака	Значення	Модель	Рішення	
exp_a – коеф. амплітуди (y=ae*bx)	0.0478	Logistic Regression	✓ ЗБАЛАНСОВАНА	<div style="text-align: center;"> <p>Голоси моделей</p> <p>✓ 12 збалансована    ✗ 2 незбалансована</p> <p><b>ЗБАЛАНСОВАНА</b></p> <p>Більшість: 12 з 14 моделей</p> </div>
exp_b – швидкість exp зростання/спадання	0.1121	LDA	✓ ЗБАЛАНСОВАНА	
exp_rsquate – якість апроксимації (R <sup>2</sup> )	0.0427	Naive Bayes	✓ ЗБАЛАНСОВАНА	
exp_rmse – середньоквадр. похибка	2.0945	KNN (k=5)	✗ НЕЗБАЛАНСОВАНА	
tangent_45 – індекс точки нахилу 45°	428.9	Decision Tree	✓ ЗБАЛАНСОВАНА	
std_before_45 – std до точки x <sub>н</sub>	1.9993	Random Forest	✓ ЗБАЛАНСОВАНА	
std_after_45 – std після точки x <sub>н</sub>	2.0866	Extra Trees	✓ ЗБАЛАНСОВАНА	
skew_after_45 – асиметрія залишків	-0.3976	AdaBoost	✗ НЕЗБАЛАНСОВАНА	
kurt_after_45 – ексцес залишків	12.5616	Gradient Boosting	✓ ЗБАЛАНСОВАНА	
time_mean – середнє компонент	0.601	SVM (RBF)	✓ ЗБАЛАНСОВАНА	
time_std – розкид (варіативність)	2.7	MLP Neural Net	✓ ЗБАЛАНСОВАНА	
time_skew – асиметрія розподілу	0.0562	XGBoost	✓ ЗБАЛАНСОВАНА	
time_kurt – ексцес розподілу	3.8755	LightGBM	✓ ЗБАЛАНСОВАНА	
		Voting Ensemble	✓ ЗБАЛАНСОВАНА	

Рисунок 4.16 – Інтерфейс модуля класифікації збалансованості електроенергосистеми з результатами голосування моделей

Наявність двох відмінних рішень (KNN та AdaBoost класифікували стан як незбалансований) не впливає на підсумковий результат, проте фіксується у системі для подальшого аналізу оператором. Це дозволяє своєчасно виявляти ознаки переходу системи до незбалансованого режиму електроспоживання.

#### 4.4. Висновки до розділу 4

У розділі продемонстровано програмний модуль аналізу даних електроспоживання, що забезпечує аналіз на основі методів ПКВП,

класифікацію режимів електроспоживання розподілених електроенергетичних системах та візуалізацію результатів. Обґрунтовано вибір програмних засобів, визначено архітектуру системи та підтверджено ефективність запропонованого підходу за результатами експериментального дослідження.

1. Обґрунтовано архітектуру програмного модуля для аналізу даних електроспоживання, яка забезпечує інтеграцію методів математичного аналізу на основі моделі періодично корельованих випадкових процесів та алгоритмів машинного навчання з метою реалізації повного циклу опрацювання даних від отримання неопрацьованих даних до класифікації режимів електроспоживання.

2. Запропоновано метод адаптивного вибору комунікаційного протоколу передавання даних електроспоживання на основі аналізу п'яти мережевих параметрів із застосуванням методів машинного навчання. За результатами порівняльного аналізу чотирьох класифікаторів встановлено, що модель MLP Neural Network забезпечила найвищу точність вибору протоколу ( $F1\text{-score}=0,9020$ ), а застосування адаптивного механізму вибору оптимального протоколу залежно від поточного стану мережі забезпечує надійне та своєчасне надходження даних електроспоживання для подальшого аналізу та класифікації.

3. Реалізовано програмний модуль аналізу даних електроспоживання, який забезпечує формування кореляційних компонент сигналу на основі ПКВП, побудову 13-вимірної ознакової простору та автоматизовану класифікацію збалансованого і незбалансованого режимів електроспоживання з підтримкою інтерактивної візуалізації результатів для оперативного прийняття рішень щодо стану електроенергетичної системи.

## ВИСНОВКИ

У дисертаційній роботі розв'язано актуальне науково-прикладне завдання аналізу даних електроспоживання на основі моделі у вигляді періодично корельованих випадкових процесів у поєднанні із методами машинного навчання для класифікації режимів електроспоживання в розподілених електроенергетичних системах. При цьому отримано такі наукові та практичні результати.

1. *Проаналізовано* відомі математичні моделі та методи аналізу даних електроспоживання, включаючи статистичні, спектральні, нейромережеві підходи, методи машинного навчання, та охарактеризовано області їх практичного застосування. На основі проведеного аналізу обґрунтовано актуальність теми дисертаційного дослідження та сформульовано завдання для її реалізації.

2. *На основі аналізу експериментальних даних встановлено*, що процес електроспоживання є нестационарним, проте характеризується подвійною природою – повторюваною за структурою та стохастичною за характером реалізації кожного окремого циклу, що обґрунтовує доцільність застосування математичних моделей, які одночасно враховують повторюваність та стохастичність електроспоживання, зокрема моделі циклічного випадкового процесу та моделі у вигляді періодично корельованого випадкового процесу.

3. *Обґрунтовано математичну модель* електроспоживання у вигляді періодично корельованого випадкового процесу, для визначення режиму електроспоживання (збалансований/незбалансований), яка одночасно враховує стохастичну природу, повторювану структуру сигналів електроспоживання, імовірнісні характеристики якої безпосередньо інтегровані у кореляційну структуру моделі через умову періодичності статистичних характеристик, що дозволяє формалізувати нестационарність процесу без надлишкової параметризації та мінімізує кількість обчислювальних ресурсів.

4. *Обґрунтовано застосування компонентного методу аналізу* періодично корельованого випадкового процесу для формування інформативних ознак, що дозволило розкласти кореляційну структуру сигналу електроспоживання на систему кореляційних компонент та використати їх як інформативні ознаки для класифікації режимів електроспоживання.

5. *На основі компонентного методу ПКВП сформовано 13* інформативних ознак (9 компонентних та 4 часових), дискримінантну здатність яких підтверджено аналізом розподілів, побудовою діаграм розмаху та кореляційним аналізом. За результатами аналізу важливості ознак ідентифіковано чотири найважливіші характеристики: середній рівень енергії (*time\_mean*), інтенсивність флуктуацій після критичної точки (*skew\_after\_45*), часова локалізація точки перелому (*tangent\_45*) та стандартне відхилення сигналу до точки перелому (*std\_before\_45*), як основні індикатори режиму електроспоживання.

6. *За результатами порівняльного аналізу ефективності 14* моделей машинного навчання із застосуванням матриць плутанини, ROC-кривих та операційних характеристик встановлено, що найвищий F1-score (0,9877) досягнуто для Extra Trees, KNN (k=5) та SVM (RBF), за критерієм крос-валідаційної стабільності перевагу має Extra Trees (CV-F1=0,9826). Значення AUC перевищує 0,999 для дванадцяти з чотирнадцяти моделей, а MCC та Карра вище 0,88 для всіх моделей підтверджують надійну узгодженість класифікації. Висока якість класифікації незалежно від типу алгоритму свідчить про достатню дискримінантну здатність сформованого ознакового простору на основі компонентного методу ПКВП.

7. *Запропоновано метод* адаптивного вибору оптимального протоколу передавання даних на основі аналізу п'яти мережевих параметрів: затримки, пропускної здатності, втрат пакетів, розміру корисного навантаження та джитер ( *jitter*), із застосуванням методів машинного навчання. Порівняльний аналіз чотирьох моделей (Logistic Regression, Gradient Boosting, Random Forest, MLP

Neural Network) показав, що модель MLP Neural Network забезпечила найкращі результати класифікації, досягнувши точність (Accuracy) та F1-score на рівні 0,9020.

8. *Розроблено* програмні модулі класифікації режимів електроспоживання та адаптивного вибору протоколу передавання даних, які забезпечують визначення поточного режиму роботи вузлів розподіленої електроенергетичної системи, адаптивне перемикання протоколів передавання даних відповідно до поточних характеристик каналу зв'язку, а також візуалізацію результатів для підтримки оперативного прийняття рішень.

**СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ**

1. Khalid M. Energy 4.0: AI-enabled digital transformation for sustainable power networks. *Computers & Industrial Engineering*. 2024. Art. 110253. DOI: <https://doi.org/10.1016/j.cie.2024.110253>.
2. Fouad M., Mali R., Lmouatassime A., Bousmah M. Machine Learning and IoT for Smart Grid. *ISPRS – International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*. 2020. Vol. XLIV-4/W3-2020. P. 233–240. DOI: <https://doi.org/10.5194/isprs-archives-XLIV-4-W3-2020-233-2020>.
3. Tuballa M. L., Abundo M. L. A review of the development of Smart Grid technologies. *Renewable and Sustainable Energy Reviews*. 2016. Vol. 59. P. 710–725. DOI: <https://doi.org/10.1016/j.rser.2016.01.011>.
4. Data Analytics for Smart Grids Applications: A Key to Smart City Development / ed. by D. Kumar Sharma et al. Cham : *Springer Nature Switzerland*, 2023. DOI: <https://doi.org/10.1007/978-3-031-46092-0>.
5. Shi W., Cao J., Zhang Q., Li Y., Xu L. Edge Computing: Vision and Challenges. *IEEE Internet of Things Journal*. 2016. Vol. 3, no. 5. P. 637–646. DOI: <https://doi.org/10.1109/jiot.2016.2579198>.
6. Number of connected IoT devices growing 14% to 21.1 billion. IoT Analytics Your Global IoT Market Research Partner. URL: <https://iot-analytics.com/number-connected-iot-devices/> (дата звернення: 27.02.2026).
7. Smart Grid Technologies: Communication Technologies and Standards / V. C. Gungor et al. *IEEE Transactions on Industrial Informatics*. 2011. Vol. 7, no. 4. P. 529–539. DOI: <https://doi.org/10.1109/tii.2011.2166794>.
8. Enhancing resilience in complex energy systems through real-time anomaly detection: a systematic literature review / A. Aghazadeh Ardebili et al. *Energy Informatics*. 2024. Vol. 7, no. 1. Art. 96. DOI: <https://doi.org/10.1186/s42162-024-00401-8>.
9. *Інтелектуальні електричні мережі: елементи та режими* : монографія / за заг. ред. О. В. Кириленка. Київ : Ін-т електродинаміки НАН України, 2016. 400 с.

10. Electric storage systems based on the use of powerful bi-directional converters / A. Zharkin et al. *Bulletin of the National Technical University «KhPI» Series: New solutions in modern technologies*. 2018. Vol. 1, no. 26(1302). P. 25–33. DOI: <https://doi.org/10.20998/2413-4295.2018.26.04>.
11. Лежнюк П. Д., Рубаненко О. Є., Гунько І. О. *Оптимізація режимів електричних мереж з відновлюваними джерелами електроенергії* : монографія. Вінниця : ВНТУ, 2017. 164 с.
12. Лежнюк П. Д., Комар В. О., Добровольська Л. Н., Повстянко К. О. Відносне оцінювання засобів балансування режимів електроенергетичних систем з відновлюваними джерелами енергії. *Вісник Вінницького політехнічного інституту*. 2022. № 3. С. 24–30. DOI: <https://doi.org/10.31649/1997-9266-2022-162-3-24-30>.
13. Komenda N., Volynets V., Hrytsiuk I., Bandura I., Romaniuk M. Estimation of the total influence of methods for increasing the line natural load. *Polityka Energetyczna – Energy Policy Journal*. 2024. Vol. 27, no. 1. P. 27–48. DOI: <https://doi.org/10.33223/epj/175239>.
14. Hrytsiuk I., Volynets V., Hrytsiuk Yu., Bandura I., Komenda N. Prospects for the integration of distributed energy sources into the Ukrainian power grid. *Machinery & Energetics*. 2025. Vol. 16, no. 1. P. 130–145. DOI: <https://doi.org/10.31548/machinery/1.2025.130>.
15. Davydenko L. V., Davydenko V. A. Aspects of integration into the electric network of decentralized generation based on renewable energy sources: an overview of issues and tasks for the sustainable development of the energy system. *Journal of Electrical and Power Engineering*. 2024. Vol. 30, no. 1. P. 22–29. DOI: <https://doi.org/10.31474/2074-2630-2024-1-22-29>.
16. Khomiak A., Rozen V. Analysis of methods of changing volumes of electricity generation at solar power plants. *Power Engineering: Economics, Technique, Ecology*. 2023. no. 3. DOI: <https://doi.org/10.20535/1813-5420.3.2023.289652>.

17. Про затвердження Національного плану дій з відновлюваної енергетики на період до 2030 року та плану заходів з його виконання. Офіційний вебпортал парламенту України. URL: <https://zakon.rada.gov.ua/laws/show/761-2024-p#Text> (дата звернення: 27.02.2026).
18. Power system stability in the Era of energy Transition: Importance, Opportunities, Challenges, and future directions / A. M. Saleh et al. *Energy Conversion and Management*: X. 2024. Art. 100820. DOI: <https://doi.org/10.1016/j.ecmx.2024.100820>.
19. Hahn H., Meyer–Nieberg S., Pickl S. Electric load forecasting methods: Tools for decision making. *European Journal of Operational Research*. 2009. Vol. 199, no. 3. P. 902–907. DOI: <https://doi.org/10.1016/j.ejor.2009.01.062>.
20. Predicting peak day and peak hour of electricity demand with ensemble machine learning / T. Fu et al. *Frontiers in Energy Research*. 2022. Vol. 10. DOI: <https://doi.org/10.3389/fenrg.2022.944804>.
21. Artificial intelligence based anomaly detection of energy consumption in buildings: A review, current trends and new perspectives / Y. Himeur et al. *Applied Energy*. 2021. Vol. 287. Art. 116601. DOI: <https://doi.org/10.1016/j.apenergy.2021.116601>.
22. AI-Empowered Methods for Smart Energy Consumption: A Review of Load Forecasting, Anomaly Detection and Demand Response / X. Wang et al. *International Journal of Precision Engineering and Manufacturing-Green Technology*. 2023. DOI: <https://doi.org/10.1007/s40684-023-00537-0>.
23. A data-driven methodology for deriving electricity consumption typologies from smart meters / C. Quesada et al. *Energy Reports*. 2025. Vol. 14. P. 2420–2434. DOI: <https://doi.org/10.1016/j.egyr.2025.09.002>.
24. Comparative Analysis of ARIMA and SARIMA Models in Electrical Load Forecasting: Insights for Long and Short–Term Projections / M. A. Momani et al. *International Journal of Electrical Engineering and Computer Science*. 2025. Vol. 7. P. 83–89. DOI: <https://doi.org/10.37394/232027.2025.7.8>.

25. Peak Electrical Energy Consumption Prediction by ARIMA, LSTM, GRU, ARIMA–LSTM and ARIMA–GRU Approaches / A. A. Pierre et al. *Energies*. 2023. Vol. 16, no. 12. P. 4739. DOI: <https://doi.org/10.3390/en16124739>.
26. Electricity load forecasting using clustering and ARIMA model for energy management in buildings / B. Nepal et al. *Japan Architectural Review*. 2019. Vol. 3, no. 1. P. 62–76. DOI: <https://doi.org/10.1002/2475–8876.12135>.
27. Papaioannou G. P., Dikaiakos C., Dramountanis A., Papaioannou P. G. Analysis and modeling for short- to medium-term load forecasting using a hybrid manifold learning principal component model and comparison with classical statistical models (SARIMAX, exponential smoothing) and artificial intelligence models (ANN, SVM): the case of Greek electricity market. *Energies*. 2016. Vol. 9, no. 8. P. 635. DOI: <https://doi.org/10.3390/en9080635>.
28. Fryz M., Scherbak L. Properties of Discrete–Time Conditional Linear Cyclostationary Random Processes in the Problems Of Energy Informatics. *System Research in Energy*. 2023. Vol. 2023, no. 1. P. 72–79. DOI: <https://doi.org/10.15407/srenergy2023.01.072>.
29. Data-driven load profiles and the dynamics of residential electricity consumption / M. Anvari et al. *Nature Communications*. 2022. Vol. 13, no. 1. DOI: <https://doi.org/10.1038/s41467-022-31942-9>.
30. Терещенко Т. О., Ямненко Ю. С., Клепач Л. Є., Лайкова Л. Г., Палій Д. М. Огляд основних моделей прогнозування енергоспоживання в MicroGrid. *Вчені записки Таврійського національного університету імені В.І. Вернадського. Серія: Технічні науки*. 2018. Т. 29 (68), № 4 (1). С. 36–41.
31. Марченко Б. Г., Приймак М. В. Побудова моделі та аналіз стохастично–періодичних навантажень енергосистем. *Праці Інституту електродинаміки НАН України*. Київ : ІЕД НАН України, 1999. Вип. 1. С. 129–153.
32. Приймак М. В., Мацюк О. В., Маєвський О. В., Прошин С. Ю. Моделі та методи дослідження систем масового обслуговування марківського типу в умовах

стохастичної періодичності та їхнє застосування в енергетиці. *Технічна електродинаміка*. 2014. № 2. С. 11–16.

33. Марценко С. В. Математичне моделювання та статистичні методи обробки даних вимірювань в задачах моніторингу електронавантаження : автореф. дис. ... канд. техн. наук : 01.05.02. Тернопіль : ТНТУ ім. І. Пулюя, 2011. 22 с.
34. Scherbak L. M., Fryz M. Y., Hotovych V. A. Electricity consumption simulation using random coefficient periodic autoregressive model. *IOP Conference Series: Earth and Environmental Science*. 2023. Vol. 1254, no. 1. Art. 012027. DOI: <https://doi.org/10.1088/1755-1315/1254/1/012027>.
35. Gotovych V., Nazarevych O., Shcherbak L. Mathematical modeling of the regular-mode electric power supply and electric power consumption processes of the organization. *Scientific journal of the Ternopil national technical university*. 2018. Vol. 91, no. 3. P. 134–142. DOI: [https://doi.org/10.33108/visnyk\\_tntu2018.03.134](https://doi.org/10.33108/visnyk_tntu2018.03.134).
36. Fryz M., Scherbak L. Statistical Analysis of Random Coefficient Periodic Autoregression and Its Application for Short-Term Electricity Consumption Forecasting. *Tekhnichna Elektrodynamika*. 2019. Vol. 2019, no. 2. P. 38–47. DOI: <https://doi.org/10.15407/techned2019.02.038>.
37. Fryz M., Mlynko B. Discrete-time conditional linear random processes and their properties. *Herald of Khmelnytskyi National University. Technical Sciences*. 2022. Vol. 309, no. 3. P. 7–12. DOI: <https://doi.org/10.31891/2307-5732-2022-309-3-7-12>.
38. Application of hilbert transform for power quality indicators monitoring in general purpose grids / S. Kovtun et al. *System Research in Energy*. 2024. Vol. 2024, no. 2. P. 71–83. DOI: <https://doi.org/10.15407/srenergy2024.02.071>.
39. Nowotarski J., Weron R. Recent advances in electricity price forecasting: A review of probabilistic forecasting. *Renewable and Sustainable Energy Reviews*. 2018. Vol. 81. P. 1548–1568. DOI: <https://doi.org/10.1016/j.rser.2017.05.234>.
40. Hong T., Fan S. Probabilistic electric load forecasting: a tutorial review. *International Journal of Forecasting*. 2016. Vol. 32, no. 3. P. 914–938. DOI: <https://doi.org/10.1016/j.ijforecast.2015.11.011>.

41. Jaworski I., Derecka M., Bielinski W. Ocena możliwości zastosowania teorii procesów losowych okresowo niestacjonarnych do badania dobowej zmienności mocy pobieranej przez odbiorców. VIII konferencja naukowo-techniczna «*Rynek energii elektrycznej: liberalizacja – szanse i zagrożenia*». Lublin : Politech. Lubelska, 2001. S. 45–52. (польськ.).
42. Bielinski W., Derecka M. Komputerowo wspomagane wyznaczanie charakterystyk stochastycznych przebiegów obciążeń elektroenergetycznych. VII Konferencja naukowo-techniczna «*Rynek energii elektrycznej: liberalizacja – szanse i zagrożenia*» (ZKwE 2002). Poznań, Poland, 2002. S. 227–232. (польськ.).
43. Яворський І. М. Математичні моделі та аналіз стохастичних коливань. Львів : ФМІ НАНУ, 2013. 804 с.
44. Приймак М. В. Побудова прогнозних графіків енергонавантажень на основі періодичного лінійного випадкового процесу. *Технічна електродинаміка*. 2000. № 3. С. 50–52.
45. Мацюк О. В., Приймак М. В. Періодично корельовані випадкові процеси із модульованими через період значеннями ймовірнісних характеристик. *Вісник Тернопільського державного технічного університету*. 2006. Т. 11, № 2. С. 142–149.
46. Voloshchuk A., Osukhivska H., Khvostivskyi M., Sverstiuk A. Application of periodically correlated stochastic processes for forecasting electricity consumption. *Measuring and Computing Devices in Technological Processes*. 2025. no. 3. P. 393–403. DOI: <https://doi.org/10.31891/2219-9365-2025-83-48>.
47. Voloshchuk A., Osukhivska H., Khvostivskyi M., Sverstiuk A. Analysis of electricity consumption using the component method of periodically correlated random processes. *Computer Systems and Information Technologies*. 2025. no. 3. P. 74–82. DOI: <https://doi.org/10.31891/csit-2025-3-8>.
48. Voloshchuk A., Osukhivska H. Application of a cyclic stochastic process model for analyzing electricity consumption. *Computer-Integrated Technologies: Education,*

*Science, Production*. 2025. no. 61. P. 35–42. DOI: <https://doi.org/10.36910/6775-2524-0560-2025-61-05>.

49. Драган Я. Енергетична теорія лінійних моделей стохастичних сигналів. Львів : Фіз.-мех. ін-т ім. Г. В. Карпенка НАН України, 1997. 361 с. (Теорія сигналів і систем: моделі, алгоритми, структури ; т. 1).
50. Драган Я., Сікора Л., Яворський Б. Основи сучасної теорії стохастичних сигналів і енергетична концепція: математичний апарат, фізичне тлумачення. Львів : Центр стратегічних досліджень ЕБТЕС, 1999. 133 с.
51. Драган Я., Євтух П., Сікора Л., Яворський Б. Періодично корельовані випадкові процеси як адекватні моделі сигналів кратної ритміки природних явищ і технологічних процесів. *Комп'ютерні технології друкарства*. 2000. № 4. С. 269–290.
52. Alnuaimi A. F. A. H., Albaldawi T. H. K. An overview of machine learning classification techniques. *BIO Web of Conferences*. 2024. Vol. 97. P. 00133. DOI: <https://doi.org/10.1051/bioconf/20249700133>.
53. A machine learning ensemble framework based on a clustering algorithm for improving electric power consumption performance / T. Sim et al. *Scientific Reports*. 2025. Vol. 15, no. 1. DOI: <https://doi.org/10.1038/s41598-025-23978-w>.
54. Advancing ensemble learning techniques for residential building electricity consumption forecasting: Insight from explainable artificial intelligence / J. Moon et al. *PLOS ONE*. 2024. Vol. 19, no. 11. Art. e0307654. DOI: <https://doi.org/10.1371/journal.pone.0307654>.
55. Guato Burgos M. F., Morato J., Vizcaino Imacaña F. P. A Review of Smart Grid Anomaly Detection Approaches Pertaining to Artificial Intelligence. *Applied Sciences*. 2024. Vol. 14, no. 3. Art. 1194. DOI: <https://doi.org/10.3390/app14031194>.
56. Swarnkar V., Ralhan S., Singh M., Parashar D., Singh M. Correlation based feature importance analysis for improving machine learning stability predictions in hybrid PV systems. *Scientific Reports*. 2026. Vol. 16. Art. 10041. DOI: <https://doi.org/10.1038/s41598-026-37270-y>.

57. Misiurek K., Olkuski T., Zyśk J. Review of Methods and Models for Forecasting Electricity Consumption. *Energies*. 2025. Vol. 18, no. 15. Art. 4032. DOI: <https://doi.org/10.3390/en18154032>.
58. Modelling the properties of shape memory alloys using machine learning methods / O. Yasniy et al. *Procedia Structural Integrity*. 2025. Vol. 68. P. 132–138. DOI: <https://doi.org/10.1016/j.prostr.2025.06.033>.
59. Романюк М. В., Коменда Н. В., Волинець В. І., Гадай А. В., Бандура І. О. Енергетичне моделювання Smart Grid: морфометричний підхід до аналізу розподілених енергетичних систем. *Наукові праці ДонНТУ. Серія: Електротехніка і енергетика*. 2025. Т. 33, вип. 2. С. 37–42. DOI: <https://doi.org/10.31474/2074-2630-2025-2-37-42>.
60. Electricity Consumption Classification using Various Machine Learning Models / B. K. Paikaray et al. *EAI Endorsed Transactions on Energy Web*. 2024. Vol. 11. DOI: <https://doi.org/10.4108/ew.6274>.
61. Zhang W., Quan H., Srinivasan D. Parallel and reliable probabilistic load forecasting via quantile regression forest and quantile determination. *Energy*. 2018. Vol. 160. P. 810–819. DOI: <https://doi.org/10.1016/j.energy.2018.07.019>.
62. Kim T.-Y., Cho S.-B. Predicting residential energy consumption using CNN-LSTM neural networks. *Energy*. 2019. Vol. 182. P. 72–81. DOI: <https://doi.org/10.1016/j.energy.2019.05.230>.
63. Li N., Li B., Gao L. Transient Stability Assessment of Power System Based on XGBoost and Factorization Machine. *IEEE Access*. 2020. Vol. 8. P. 28403–28414. DOI: <https://doi.org/10.1109/access.2020.2969446>.
64. A Comprehensive Survey on Graph Neural Networks / Z. Wu et al. *IEEE Transactions on Neural Networks and Learning Systems*. 2021. Vol. 32, no. 1. P. 4–24. DOI: <https://doi.org/10.1109/tnnls.2020.2978386>.
65. Kea K., Han Y., Kim T.-K. Enhancing anomaly detection in distributed power systems using autoencoder-based federated learning. *PLOS ONE*. 2023. Vol. 18, no. 8. Art. e0290337. DOI: <https://doi.org/10.1371/journal.pone.0290337>.

66. Imani M., Beikmohammadi A., Arabnia H. R. Comprehensive Analysis of Random Forest and XGBoost Performance with SMOTE, ADASYN, and GNUS Under Varying Imbalance Levels. *Technologies*. 2025. Vol. 13, no. 3. P. 88. DOI: <https://doi.org/10.3390/technologies13030088>.
67. Kumar Panda D., Das S., Townley S. Hyperparameter Optimized Classification Pipeline for Handling Unbalanced Urban and Rural Energy Consumption Patterns. *Expert Systems with Applications*. 2023. Vol. 214. Art. 119127. DOI: <https://doi.org/10.1016/j.eswa.2022.119127>.
68. Artificial Intelligence Techniques for Power System Transient Stability Assessment / P. Sarajcev et al. *Energies*. 2022. Vol. 15, no. 2. Art. 507. DOI: <https://doi.org/10.3390/en15020507>.
69. Duan J. Deep learning anomaly detection in AI-powered intelligent power distribution systems. *Frontiers in Energy Research*. 2024. Vol. 12. Art. 1364456. DOI: <https://doi.org/10.3389/fenrg.2024.1364456>.
70. Deep learning for intelligent demand response and smart grids: A comprehensive survey / P. Boopathy et al. *Computer Science Review*. 2024. Vol. 51. Art. 100617. DOI: <https://doi.org/10.1016/j.cosrev.2024.100617>.
71. Aguiar-Pérez J. M., Pérez-Juárez M. Á. An Insight of Deep Learning Based Demand Forecasting in Smart Grids. *Sensors*. 2023. Vol. 23, no. 3. Art. 1467. DOI: <https://doi.org/10.3390/s23031467>.
72. A meta-learning based distribution system load forecasting model selection framework / Y. Li et al. *Applied Energy*. 2021. Vol. 294. Art. 116991. DOI: <https://doi.org/10.1016/j.apenergy.2021.116991>.
73. Informer: Beyond Efficient Transformer for Long Sequence Time-Series Forecasting / H. Zhou et al. *Proceedings of the AAAI Conference on Artificial Intelligence*. 2021. Vol. 35, no. 12. P. 11106–11115. DOI: <https://doi.org/10.1609/aaai.v35i12.17325>.
74. Dudek G. Pattern-based local linear regression models for short-term load forecasting. *Electric Power Systems Research*. 2016. Vol. 130. P. 139–147. DOI: <https://doi.org/10.1016/j.epsr.2015.09.001>.

75. Dudek G. A Comprehensive Study of Random Forest for Short-Term Load Forecasting. *Energies*. 2022. Vol. 15, no. 20. Art. 7547. DOI: <https://doi.org/10.3390/en15207547>.
76. Load Forecasting Techniques and Their Applications in Smart Grids / H. Habbak et al. *Energies*. 2023. Vol. 16, no. 3. Art. 1480. DOI: <https://doi.org/10.3390/en16031480>.
77. A novel multi-time-scale modeling for electric power demand forecasting: From short-term to medium-term horizon / K. G. Boroojeni et al. *Electric Power Systems Research*. 2017. Vol. 142. P. 58–73. DOI: <https://doi.org/10.1016/j.epsr.2016.08.031>.
78. Bahman S., Zareipour H. Long-Term Multi-Resolution Probabilistic Load Forecasting Using Temporal Hierarchies. *Energies*. 2025. Vol. 18, no. 11. Art. 2908. DOI: <https://doi.org/10.3390/en18112908>.
79. Napolitano A. Cyclostationarity: New trends and applications. *Signal Processing*. 2016. Vol. 120. P. 385–408. DOI: <https://doi.org/10.1016/j.sigpro.2015.09.011>.
80. Лупенко С. А., Литвиненко Я. В. Методи статистичного опрацювання циклічних випадкових процесів у задачах моделювання сигналів. *Вісник ТНТУ*. 2015. № 2 (78). С. 177–188.
81. Periodically correlated random processes: Application in early diagnostics of mechanical systems / I. Javorskyj et al. *Mechanical Systems and Signal Processing*. 2017. Vol. 83. P. 406–438. DOI: <https://doi.org/10.1016/j.ymssp.2016.06.022>.
82. Properties of characteristics estimators of periodically correlated random processes in preliminary determination of the period of correlation / I. N. Yavorskyj et al. *Radioelectronics and Communications Systems*. 2012. Vol. 55, no. 8. P. 335–348. DOI: <https://doi.org/10.3103/s0735272712080018>.
83. Драган Я. П. Періодично корельовані та споріднені з ними випадкові процеси – моделі сигналів у коливних системах. *Імовірнісні моделі та обробка випадкових сигналів і полів*. Харків : Ін-т радіоелектр., 1992. Т. 1. С. 26–41.
84. Lytvynenko I. V., Maruschak P. O., Lupenko S. A. Processing and modeling of ordered relief at the surface of heat-resistant steels after laser irradiation as a cyclic

- random process. *Automatic Control and Computer Sciences*. 2014. Vol. 48, no. 1. P. 1–9. DOI: <https://doi.org/10.3103/s0146411614010040>.
85. Lupenko S., Butsiy R. Isomorphic Multidimensional Structures of the Cyclic Random Process in Problems of Modeling Cyclic Signals with Regular and Irregular Rhythms. *Fractal and Fractional*. 2024. Vol. 8, no. 4. P. 203. DOI: <https://doi.org/10.3390/fractalfract8040203>.
86. Лупенко С. А., Литвиненко Я. В., Сверстюк А. С. Статистичний сумісний аналіз кардіосигналів на основі вектора циклічних ритмічно пов'язаних випадкових процесів. *Електроніка та системи керування*. Київ : Національний авіаційний університет, 2008. С. 22–29.
87. Lupenko S. Abstract Cyclic Functional Relation and Taxonomies of Cyclic Signals Mathematical Models: Construction, Definitions and Properties. *Mathematics*. 2024. Vol. 12, no. 19. Art. 3084. DOI: <https://doi.org/10.3390/math12193084>.
88. Литвиненко Я. В., Лупенко С. А., Студена Ю. Методи статистичної обробки сигналів серця на базі їх моделі у вигляді циклічного випадкового процесу із зонною часовою структурою. *Вісник Тернопільського державного технічного університету*. 2006. Т. 11, № 4. С. 189–200.
89. Сверстюк А. С. Математичне моделювання та методи обробки синхронно зареєстрованих сигналів серця з використанням циклічних ритмічно пов'язаних випадкових процесів : дис. ... канд. техн. наук : 01.05.02. Київ, 2010. 196 с.
90. Mathematical model of gas consumption process in the form of cyclic random process / I. Lytvynenko et al. 2021 *IEEE 16th International Conference on Computer Sciences and Information Technologies (CSIT)*, LVIV, Ukraine, 22–25 September 2021. 2021. DOI: <https://doi.org/10.1109/csit52700.2021.9648621>.
91. Горкуненко А. Б., Лупенко С. А. Порівняльний аналіз математичних моделей циклічних економічних процесів в інформаційних системах підтримки прийняття економічних рішень. *Науковий вісник НЛТУ України*. 2012. Вип. 22.5. С. 345–351.

92. Lupenko S. The Mathematical Model of Cyclic Signals in Dynamic Systems as a Cyclically Correlated Random Process. *Mathematics*. 2022. Vol. 10, no. 18. Art. 3406. DOI: <https://doi.org/10.3390/math10183406>.
93. Lupenko S. Rhythm–Adaptive Statistical Estimation Methods of Probabilistic Characteristics of Cyclic Random Processes. *Digital Signal Processing*. 2024. Vol. 151. Art. 104563. DOI: <https://doi.org/10.1016/j.dsp.2024.104563>.
94. Detection of distributed and localized faults in rotating machines using periodically non–stationary covariance analysis of vibrations / I. Yavorsky et al. *Measurement Science and Technology*. 2023. DOI: <https://doi.org/10.1088/1361-6501/acbc93>.
95. Methods of Hidden Periodicity Discovering for Gearbox Fault Detection / I. Javorskyj et al. *Sensors*. 2021. Vol. 21, no. 18. Art. 6138. DOI: <https://doi.org/10.3390/s21186138>.
96. Investigation of the correlation structure of the vibration signal of the decanter bearing assembly / I. M. Javorskyj et al. *Tehnička diagnostika i nerazrušajúšij kontrol'*. 2024. Vol. 2024, no. 2. P. 3–10. DOI: <https://doi.org/10.37434/tdnk2024.02.0>.
97. Дедів Л. Є., Сверстюк А. С., Дедів І. Ю., Хвостівський М. О., Дозорський В. Г., Яворська Є. Б. Математичне та комп'ютерне моделювання електрокардіосигналів у системах голтерівського моніторингу : монографія. Львів : Магнолія 2006, 2021. 120 с.
98. Хвостівська Л.В., Осухівська Г.М., Хвостівський М.О., Шадріна Г.М., Дедів І.Ю. Розвиток методів та алгоритмів обчислення періоду стохастичних біомедичних сигналів для медичних комп'ютерно–діагностичних систем. *Вісник НТУУ КПІ. Серія Радіотехніка, Радіоапаратобудування*, (79). С. 78–84. DOI: <https://doi.org/10.20535/radap.2019.79.78-84>.
99. Дедів Л.Є., Хвостівський М.О., Дунець В.Л., Шадріна Г.М. Імітаційна модель електрокардіосигналу на основі періодично корельованого випадкового процесу. *Вісник Тернопільського державного технічного університету*. Тернопіль: ТДТУ ім. І.Пулюя. 2008. №3. С. 201–205.

100. Дунець В.Л., Драган Я.П. Опрацювання електрокардіосигналу синфазним методом для виявлення патології. *Комп'ютерні технології друкарства*. Львів: Українська академія друкарства, 2007. № 18. С. 51–59.
101. Khvostivskyi M., Osukhivska H., Khvostivska L., Lobur T., Velychko D. Mathematical modelling of daily computer network traffic. ITTAP'2021: *1st International Workshop on Information Technologies: Theoretical and Applied Problems*. Ternopil, Ukraine, 2021. Vol. 3039. URL: <https://ceur-ws.org/Vol-3039/short24.pdf>.
102. Хвостівський М. О., Хвостівська Л. В., Дедів І. Ю., Дедів Л. Є. Інтелектуальна система прогнозування трафіку комп'ютерних мереж на основі синфазної обробки даних. *Вісник Херсонського національного технічного університету*. 2025. Т. 2, № 3(94). С. 497–503. DOI: <https://doi.org/10.35546/kntu2078-4481.2025.3.2.63>.
103. Javors'kyj I., Isayev I., Zakrzewski Z., Brooks S. P. Coherent covariance analysis of periodically correlated random processes. *Signal Processing*. 2007. Vol. 87, no. 1. P. 13–32. DOI: <https://doi.org/10.1016/j.sigpro.2006.04.002>.
104. Linear filtration methods for statistical analysis of periodically correlated random processes—Part II: Harmonic series representation / I. Javorskyj et al. *Signal Processing*. 2011. Vol. 91, no. 11. P. 2506–2519. DOI: <https://doi.org/10.1016/j.sigpro.2011.04.031>.
105. Linear filtration methods for statistical analysis of periodically correlated random processes—Part I: Coherent and component methods and their generalization / I. Javorskyj et al. *Signal Processing*. 2012. Vol. 92, no. 7. P. 1559–1566. DOI: <https://doi.org/10.1016/j.sigpro.2011.09.030>.
106. Javorskyj I., Isayev I., Majewski J., Yuzefovych R. Component covariance analysis for periodically correlated random processes. *Signal Processing*. 2010. Vol. 90, no. 4. P. 1083–1102. DOI: <https://doi.org/10.1016/j.sigpro.2009.07.031>.

107. Javorskyj I., Matsko I., Yuzefovych R., Zakrzewski Z. Discrete estimators of characteristics for periodically correlated time series. *Digital Signal Processing*. 2016. Vol. 53. P. 25–40. DOI: <https://doi.org/10.1016/j.dsp.2016.03.003>.
108. Yavorskyj I. N., Yuzefovych R., Matsko I. Y., Shevchik V. Component correlation analysis of vectorial periodically non-stationary random processes. *Radioelectronics and Communications Systems*. 2014. Vol. 57, no. 9. P. 403–417. DOI: <https://doi.org/10.3103/S0735272714090039>.
109. Javorskyj I., Yuzefovych R., Matsko I., Zakrzewski Z. The least square estimation of the basic frequency for periodically non-stationary random signals. *Digital Signal Processing*. 2022. Vol. 122. Art. 103333. DOI: <https://doi.org/10.1016/j.dsp.2021.103333>.
110. Tymoshchuk D., Voloshchuk A., Sverstiuk A., Osukhivska H., Bahrii-Zaiats O. Machine learning-based information technology for analyzing energy peaks in power grid balancing. *Proceedings of the 2nd International Workshop on Advanced Applied Information Technologies (AdvAIT 2025)*. Khmelnytskyi, Ukraine – Žilina, Slovakia, 2025. Vol. 4163. P. 204–220. URL: <https://ceur-ws.org/Vol-4163/paper18.pdf>.
111. Марченко Б. Г., Приймак М. В. Математичні основи теорії лінійних періодичних процесів : монографія. Київ : *Наукова думка*, 1996. 232 с.
112. Tymoshchuk D., Yasniy O., Didych I., Stanko A., Medvid V. Prediction of SMA hysteresis behavior by ensemble stacking machine learning model with SHAP-based explanation. *Progress in Engineering Science*. 2026. Art. 100237. DOI: <https://doi.org/10.1016/j.pes.2026.100237>.
113. Huang Z., Yi Y. Short-Term Load Forecasting for Regional Smart Energy Systems Based on Two-Stage Feature Extraction and Hybrid Inverted Transformer. *Sustainability*. 2024. Vol. 16, no. 17. Art. 7613. DOI: <https://doi.org/10.3390/su16177613>.
114. Бунько В., Кондрат О., Бунько Н. Аналіз використання та впровадження системи «розумного» моніторингу електричної енергії навчального корпусу. *Вісник Кременчуцького національного університету імені Михайла*

*Остроградського*. 2024. № 4. С. 169–176. DOI: <https://doi.org/10.32782/1995-0519.2024.4.22>.

115. Ahmad T., Chen H. Utility companies strategy for short-term energy demand forecasting using machine learning based models. *Sustainable Cities and Society*. 2018. Vol. 39. P. 401–417. DOI: <https://doi.org/10.1016/j.scs.2018.03.002>.
116. Ofori-Ntow Jnr E., Ziggah Y. Y. Electricity demand forecasting based on feature extraction and optimized backpropagation neural network. *e-Prime – Advances in Electrical Engineering, Electronics and Energy*. 2023. Vol. 6. Art. 100293. DOI: <https://doi.org/10.1016/j.prime.2023.100293>.
117. Tymoshchuk D., Didych I., Maruschak P., Yasniy O., Mykytyshyn A., Mytnyk M. Machine Learning Approaches for Classification of Composite Materials. *Modelling*. 2025. Vol. 6, no. 4. 118. DOI: <https://doi.org/10.3390/modelling6040118>.
118. Węglarczyk S. Kernel density estimation and its application. *ITM Web of Conferences*. 2018. Vol. 23. Art. 00037. DOI: <https://doi.org/10.1051/itmconf/20182300037>.
119. Williamson D. F., Parker R. A., Kendrick J. S. The Box Plot: A Simple Visual Method to Interpret Data. *Annals of Internal Medicine*. 1989. Vol. 110, no. 11. P. 916–921. DOI: <https://doi.org/10.7326/0003-4819-110-11-916>.
120. Fu H., Tang S., Zhao X. Limitations of Correlation Coefficients in Research on Functional Connectomes and Psychological Processes. *Human Brain Mapping*. 2025. Vol. 46, no. 10. Art. e70287. DOI: <https://doi.org/10.1002/hbm.70287>.
121. Hastie T., Tibshirani R., Friedman J. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. 2nd ed. New York : Springer, 2009. 745 c. DOI: <https://doi.org/10.1007/978-0-387-84858-7>.
122. Breiman L. Random forests. *Machine Learning*. 2001. Vol. 45, no. 1. P. 5–32. DOI: <https://doi.org/10.1023/a:1010933404324>.
123. Geurts P., Ernst D., Wehenkel L. Extremely randomized trees. *Machine Learning*. 2006. Vol. 63, no. 1. P. 3–42. DOI: <https://doi.org/10.1007/s10994-006-6226-1>.

124. Cortes C., Vapnik V. Support–vector networks. *Machine Learning*. 1995. Vol. 20, no. 3. P. 273–297. DOI: <https://doi.org/10.1007/bf00994018>.
125. Friedman J. H. Greedy function approximation: A gradient boosting machine. *Annals of Statistics*. 2001. Vol. 29, no. 5. P. 1189–1232. DOI: <https://doi.org/10.1214/aos/1013203451>.
126. Fawcett T. An introduction to ROC analysis. *Pattern Recognition Letters*. 2006. Vol. 27, no. 8. P. 861–874. DOI: <https://doi.org/10.1016/j.patrec.2005.10.010>.
127. Chicco D., Jurman G. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics*. 2020. Vol. 21, no. 1. Art. 6. DOI: <https://doi.org/10.1186/s12864-019-6413-7>.
128. Mitra S., Chakraborty B., Mitra P. Smart meter data analytics applications for secure, reliable and robust grid system: Survey and future directions. *Energy*. 2023. Art. 129920. DOI: <https://doi.org/10.1016/j.energy.2023.129920>.
129. Madran U., Qaisar S. M., Soyoglu D. Efficient Time–Series Feature Extraction and Ensemble Learning for Appliance Categorization Using Smart Meter Data. *Computer Modeling in Engineering & Sciences*. 2025. P. 1–10. DOI: <https://doi.org/10.32604/cmescs.2025.072024>.
130. Akinrogunde O. O., Adelakun A., Theophilus E. E., Thomas T. G. A systematic review of machine learning and deep learning approaches for load and energy consumption prediction in contemporary power systems. *Journal of Innovation Materials, Energy, and Sustainable Engineering*. 2025. Vol. 3, no. 1. P. 1–20. DOI: <https://doi.org/10.61511/jimese.v3i1.2025.1949>.
131. Ali A. M., Dawoud O. K., Ghoneim O. A., Abdel-Basset M. Machine learning models for smart grid stability prediction: a comparative analysis. *Scientific Reports*. 2026. Vol. 16, no. 1. DOI: <https://doi.org/10.1038/s41598-026-47385-x>.
132. Machine learning for power system stability and control / R. Islam et al. *Results in Engineering*. 2025. Vol. 26. Art. 105355. DOI: <https://doi.org/10.1016/j.rineng.2025.105355>.

133. Integrating Machine Learning for Anomaly Detection and Pattern Recognition in Smart Grid Power Data / C. Xu, P. Zhang, N. Luo, F. Zheng, W. He. *Distributed Generation & Alternative Energy Journal*. 2025. Vol. 40, no. 3. P. 595–614. DOI: <https://doi.org/10.13052/dgaej2156-3306.4037>.
134. Kashyap M., Sharma V. A comparative analysis and implementation of CoAP and MQTT protocol for IoT communication. *Life Cycle Reliability and Safety Engineering*. 2025. Vol. 14, no. 4. P. 715–730. DOI: <https://doi.org/10.1007/s41872-025-00324-7>.
135. IoT-enabled interoperability for energy-mobility integration: A systematic review of protocols, platforms, and standards / X. Liu et al. *Applied Energy*. 2026. Vol. 411. Art. 127611. DOI: <https://doi.org/10.1016/j.apenergy.2026.127611>.
136. Voloshchuk A., Osukhivska H. Adaptive multi-protocol communication for energy systems. *Scientific journal of the Ternopil national technical university*. 2025. Vol. 119, no. 3. P. 97–106. DOI: [https://doi.org/10.33108/visnyk\\_tntu2025.03.097](https://doi.org/10.33108/visnyk_tntu2025.03.097).
137. Voloshchuk A., Velychko D., Osukhivska H., Palamar A. Computer system for energy distribution in conditions of electricity shortage using artificial intelligence. CEUR Workshop Proceedings: *2nd International Workshop on Computer Information Technologies in Industry 4.0 (CITI 2024)*. Ternopil, Ukraine, 2024. Vol. 3742. P. 66–75. URL: <https://ceur-ws.org/Vol-3742/paper5.pdf>.
138. Kovtun V., Kovtun O., Grochla K., Yasniy O. The quality of service assessment of eMBB and mMTC traffic in a clustered 5G ecosystem of a smart factory. *Egyptian Informatics Journal*. 2025. Vol. 29. Art. 100598. DOI: <https://doi.org/10.1016/j.eij.2024.100598>.
139. Žatuchin D., Azarskov M. An Adaptive Protocol Selection Framework for Energy-Efficient IoT Communication: Dynamic Optimization Through Context-Aware Decision Making. *Informatics*. 2025. Vol. 12, no. 4. P. 125. DOI: <https://doi.org/10.3390/informatics12040125>.
140. Kasaraneni P. P., Kumar Y. V. P., Moganti G. L. K., Kannan R. Machine learning-based ensemble classifiers for anomaly handling in smart home energy

consumption data. *Sensors*. 2022. Vol. 22, no. 23. Art. 9323. DOI:  
<https://doi.org/10.3390/s22239323>.

## ДОДАТКИ

## ДОДАТОК А

## Список публікацій здобувача за темою дисертації

1. Voloshchuk A.<sup>1</sup>, Osukhivska H.<sup>2</sup>, Khvostivskyi M.<sup>3</sup>, Sverstiuk A.<sup>4</sup>. Application of periodically correlated stochastic processes for forecasting electricity consumption. *Measuring and Computing Devices in Technological Processes*. 2025. no. 3. P. 393–403. DOI: <https://doi.org/10.31891/2219-9365-2025-83-48>. (<sup>1</sup> – обґрунтував застосування синфазного методу ПКВП для аналізу сигналів електроспоживання, провів обчислювальні експерименти та підготував рукопис; <sup>2</sup> – здійснювала наукове керівництво, формулювала завдання дослідження та брала участь у написанні статті; <sup>3</sup> – брав участь в обговоренні результатів та оформленні публікації; <sup>4</sup> – здійснював наукове керівництво та брав участь у плануванні експериментів).
2. Voloshchuk A.<sup>1</sup>, Osukhivska H.<sup>2</sup>, Khvostivskyi M.<sup>3</sup>, Sverstiuk A.<sup>4</sup>. Analysis of electricity consumption using the component method of periodically correlated random processes. *Computer Systems and Information Technologies*. 2025. no. 3. P. 74–82. DOI: <https://doi.org/10.31891/csit-2025-3-8>. (<sup>1</sup> – запропонував процедуру виділення інформативних ознак із результатів компонентного методу ПКВП, провів експериментальне дослідження та підготував рукопис; <sup>2</sup> – здійснювала наукове керівництво та формулювала постановку завдання; <sup>3</sup> – брав участь в обговоренні методології та оформленні публікації; <sup>4</sup> – здійснював консультації щодо інтерпретації результатів класифікації, брав участь в обговоренні результатів та перевірці рукопису).
3. Voloshchuk A.<sup>1</sup>, Osukhivska H.<sup>2</sup>. Adaptive multi-protocol communication for energy systems. *Scientific journal of the Ternopil national technical university*. 2025. Vol. 119, no. 3. P. 97–106. DOI: [https://doi.org/10.33108/visnyk\\_tntu2025.03.097](https://doi.org/10.33108/visnyk_tntu2025.03.097). (<sup>1</sup> – розробив метод

адаптивного вибору протоколу передавання даних на основі ансамблю алгоритмів машинного навчання, реалізував систему динамічного перемикавання протоколів та підготував рукопис; <sup>2</sup> – здійснювала наукове керівництво, брала участь у постановці завдання та підготовці статті до публікації.

4. Voloshchuk A.<sup>1</sup>, Osukhivska H.<sup>2</sup>. Application of a Cyclic Stochastic Process Model for Analyzing Electricity Consumption. *Computer–Integrated Technologies: Education, Science, Production*. 2025. no. 61. P. 35–42. DOI: <https://doi.org/10.36910/6775–2524–0560–2025–61–05>. (<sup>1</sup> – провів верифікацію моделі електроспоживання у вигляді циклічного випадкового процесу на реальних даних погодинного споживання, та підготував рукопис; <sup>2</sup> – здійснювала наукове керівництво, брала участь у формулюванні завдань та написанні статті).
5. Волощук, А. В.<sup>1</sup>, Осухівська, Г. М.<sup>2</sup>, Паляниця, Ю. Б.<sup>3</sup>, Луцків, А. М.<sup>4</sup>, Валяшек, В. Б.<sup>5</sup>. Визначення збалансованості електроенергетичних систем методами машинного навчання на основі інформативних ознак отриманих компонентним методом. *Вісник Національного університету водного господарства та природокористування*. 2025. Т. 4, № 112. С. 148–161. DOI: <https://doi.org/10.31713/vt4202514>. (<sup>1</sup> – запропонував і реалізував метод виділення інформативних ознак на основі компонентного методу ПКВП, провів порівняльний аналіз алгоритмів машинного навчання для класифікації режимів електроспоживання та підготував рукопис; <sup>2</sup> – здійснювала наукове керівництво та брала участь у написанні статті; <sup>3,4,5</sup> – брали участь в обговоренні методології, результатів та оформленні публікації).

*Наукові праці, які засвідчують апробацію матеріалів дисертації:*

1. Tymoshchuk D., Voloshchuk A., Sverstiuk A., Osukhivska H., Bahrii-Zaiats O. Machine learning-based information technology for analyzing energy peaks in power grid balancing. *Proceedings of the 2nd International Workshop on Advanced Applied Information Technologies (AdvAIT 2025)*. Khmelnytskyi, Ukraine – Žilina, Slovakia, 2025. Vol. 4163. P. 204–220. URL: <https://ceur-ws.org/Vol-4163/paper18.pdf>.
2. Voloshchuk A., Osukhivska H., Khvostivskyi M., Sverstiuk A., Khvostivska L. Component method for analyzing the energy consumption signal as a periodically correlated random process. *Proceedings of the International Workshop on Information Technologies: Theoretical and Applied Problems (ITTAP 2025)*. Ternopil, Ukraine – Opole, Poland, 2025. Vol. 4146. P. 128–137. URL: <https://ceur-ws.org/Vol-4146/paper13.pdf>.
3. Voloshchuk A., Velychko D., Osukhivska H., Palamar A. Computer system for energy distribution in conditions of electricity shortage using artificial intelligence. *Proceedings of the 2nd International Workshop on Computer Information Technologies in Industry 4.0 (CITI 2024)*. Ternopil, Ukraine, 2024. Vol. 3742. P. 66–75. URL: <https://ceur-ws.org/Vol-3742/paper5.pdf>.
4. Voloshchuk A., Osukhivska H. Artificial intelligence technologies in power grids. *Modern Trends of Social Transformations of Society in Conditions of Sustainable Development : abstracts of VIII Intern. Sci. and Pract. Conf., Stockholm, Sweden, February 24–26, 2025*. Stockholm, 2025. P. 210.
5. Волощук А., Осухівська Г. Архітектура системи енергетичного підприємства для отримання даних про споживання електроенергії. *Матеріали XI наук.-техн. конф. «Інформаційні моделі, системи та технології»*, м. Тернопіль, 13–14 груд. 2023 р. Тернопіль : ТНТУ ім. Івана Пулюя, 2023. С. 140.
6. Осухівська Г., Волощук А. Технології передавання та опрацювання даних в комп'ютеризованих системах обліку електроенергії. *Матеріали X наук.-*

*техн. конф. «Інформаційні моделі, системи та технології»*, м. Тернопіль,  
7–8 груд. 2022 р. Тернопіль : ТНТУ ім. Івана Пулюя, 2022. С. 136.

## ДОДАТОК Б

## Довідки/акти впровадженнь результатів дисертаційних досліджень



МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ  
 ТЕРНОПІЛЬСЬКИЙ НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ  
 імені ІВАНА ПУЛЮЯ

вул. Руська, 56, м. Тернопіль, 46001. Тел. (0352)52-41-81. Факс (0352)25-49-83  
<http://www.tntu.edu.ua>, E-mail: [univ@tntu.edu.ua](mailto:univ@tntu.edu.ua), Код ЄДРПОУ 05408102

12.05.2026 № 2/28-689 На № \_\_\_\_\_ від \_\_\_\_\_



«ЗАТВЕРДЖУЮ»

Перший проректор

Тернопільського національного  
 технічного університету ім. І. Пулюя

Олег ЛЯШУК

05 2026 р.

## АКТ

про впровадження результатів дисертаційної роботи

*Волощука Андрія Володимировича*

на тему «Моделі та методи аналізу даних електроспоживання в розподілених електроенергетичних системах», поданого на здобуття ступеня доктора філософії за спеціальністю 122 «Комп'ютерні науки» в освітній процес Тернопільського національного технічного університету імені Івана Пулюя

Запропоновані у дисертаційній роботі Волощука А.В. математичні моделі, методи аналізу та комп'ютерного моделювання електроспоживання на основі періодично корельованих випадкових процесів, а також підходи до формування інформативного ознакового простору та застосування методів машинного навчання використано у навчальному процесі кафедри комп'ютерних наук при проведенні лекційних занять з дисципліни «Моделювання систем» для здобувачів освітнього рівня «Бакалавр» спеціальності 122 «Комп'ютерні науки». Це дало змогу підвищити рівень компетентності здобувачів у питаннях аналізу та моделювання процесів електроспоживання, методів виявлення особливостей динаміки навантаження, а також практичного застосування алгоритмів машинного навчання для задач класифікації станів електроенергетичних систем.

Гарант освітньо-професійної програми

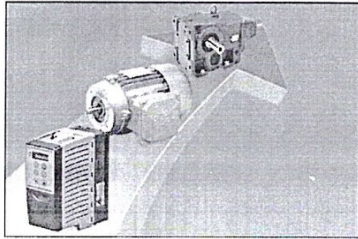
Леся ДМИТРОЦА

Завідувач кафедри комп'ютерних наук

Ігор БОДНАРЧУК

Начальник навчального відділу

Ігор ТКАЧЕНКО



**Науково-виробниче об'єднання  
"Енергоощадні технології"**

Адреса 46011 м.Тернопіль, вул. Глиняна,27,  
ЄДРПОУ 35308407

тел./факс.: (0352) 533909

Р/р UA393510050000026000144615800 в АТ  
"УКРСИББАНК",ХАРКІВ МФО 351005

Є платником податку на прибуток на загальних підставах

## АКТ ВПРОВАДЖЕННЯ

результатів дисертаційного дослідження

Волощука Андрія Володимировича

*«Моделі та методи аналізу даних електроспоживання в розподілених  
електроенергетичних системах»*

у ПП «НВО «ЕНЕРГООЩАДНІ ТЕХНОЛОГІЇ»

Цим актом підтверджується, що результати дисертаційного дослідження здобувача ступеня доктора філософії (PhD) за спеціальністю 122 «Комп'ютерні науки» Тернопільського національного технічного університету імені Івана Пулюя Волощука Андрія Володимировича впроваджено у діяльність ПП «НВО «ЕНЕРГООЩАДНІ ТЕХНОЛОГІЇ».

У діяльність підприємства впроваджено такі результати дисертаційного дослідження:

- використання математичної моделі електроспоживання у вигляді періодично корельованого випадкового процесу, що враховує циклічну нестационарність сигналів навантаження;
- модуль автоматичного визначення станів збалансованості електроенергетичної системи, інтегрований у роботу підприємства для виявлення стану збалансованості системи;
- модуль адаптивного вибору комунікаційного протоколу (MQTT, CoAP, HTTPS) для передавання даних електроспоживання в електроенергетичній системі підприємства залежно від поточного стану мережевих параметрів;
- програмні засоби візуалізації процесу передавання даних електроспоживання з відображенням обраного комунікаційного

протоколу в умовах обмеженої пропускнуї здатності та нестабільних каналів зв'язку.

Впровадження результатів дисертаційного дослідження сприяло розширенню функціональних можливостей програмних продуктів підприємства у сфері моніторингу електроспоживання та підвищенню надійності їх роботи в реальних умовах експлуатації.

*Даний акт не є підставою для проведення фінансових розрахунків.*

Директор  
ПІНВО «ЕНЕРГООЩАДНІ  
ТЕХНОЛОГІЇ»



\_\_\_\_\_ 2026 р.

/ Іскерський І.С./  
(ПІБ)



**Національна академія аграрних наук України  
Інститут сільського господарства Карпатського регіону  
Тернопільська державна сільськогосподарська дослідна станція  
Державне підприємство "Дослідне господарство "Подільське"**

48240, м.Хоростків, вул. Незалежності 19, Тернопільської обл.

тел.: (035-57) 5-11-47, факс 5-11-31.

№ 45 від 15.12.2025р.

код : 36979291

На № \_\_\_\_\_ від \_\_\_\_\_

**АКТ ВПРОВАДЖЕННЯ**  
результатів дисертаційного дослідження  
Волощука Андрія Володимировича  
«Моделі та методи аналізу даних електроспоживання в  
розподілених електроенергетичних системах»  
у ДП "ДГ "ПОДІЛЬСЬКЕ"

Цим актом підтверджується впровадження результатів дисертаційного дослідження Волощука Андрія Володимировича у діяльність ДП «ДГ «ПОДІЛЬСЬКЕ» ТДСГДС ІСГ Карпатського регіону НААН.

У процесі діяльності підприємства використано розроблені в рамках дослідження програмні рішення, спрямовані на аналіз електроспоживання та оптимізацію процесів передавання даних. Зокрема, застосовано модуль визначення стану збалансованості електроенергетичної системи, модуль адаптивного вибору протоколу передавання даних, а також засоби візуалізації відповідних процесів.

Застосування зазначених рішень дало змогу підвищити ефективність функціонування електроенергетичної системи підприємства, забезпечити більш оперативне виявлення відхилень у режимах роботи обладнання, знизити ризик перевантаження та покращити організацію передавання даних в умовах нестабільної якості каналів зв'язку.

*Цей акт не є підставою для проведення фінансових розрахунків.*

Директор  
ДП "ДГ "ПОДІЛЬСЬКЕ"  
« 15 » грудня 2025 р.



  
/ Сергій РЕВЕГА

«ЗАТВЕРДЖУЮ»

Проректор закладу вищої освіти  
з наукової роботи

Тернопільського національного  
медичного університету  
імені І.Я. Горбачевського



проф. Олещук О.М.  
2026 р.

### АКТ

про впровадження результатів дисертаційної роботи здобувача ступеня доктора філософії PhD за спеціальністю 122 «Комп'ютерні науки» кафедри комп'ютерних наук Тернопільського національного технічного університету імені Івана Пулюя *Волощука Андрія Володимировича* на тему: «Моделі та методи аналізу даних електроспоживання в розподілених електроенергетичних системах» при виконанні науково-дослідної роботи «Інтелектуальні моделі в кібер-фізичних системах медико-біологічних процесів», № державної реєстрації 0125U000105 Тернопільського національного медичного університету імені І.Я. Горбачевського

Цим актом підтверджується, що результати дисертаційної роботи здобувача ступеня доктора філософії *Волощук А.В.* на тему: «Моделі та методи аналізу даних електроспоживання в розподілених електроенергетичних системах» використано під час виконання науково-дослідної роботи «Інтелектуальні моделі в кібер-фізичних системах медико-біологічних процесів», № державної реєстрації 0125U000105, Тернопільський національний медичний університет імені І.Я. Горбачевського Міністерства охорони здоров'я України.

*Волощук А.В.* здійснив аналіз сигналів електроспоживання на основі компонентного методу періодично корельованих випадкових процесів, зокрема виконав оцінювання кореляційних компонент автоковаріаційної функції, сформував матрицю компонентного усереднення та провів статистичну обробку отриманих результатів для ідентифікації режимів роботи електроенергетичної системи.

Відповідальний виконавець:

кандидат технічних наук, доцент закладу вищої освіти  
кафедри медичної фізики діагностичного  
та лікувального обладнання Тернопільського національного  
медичного університету імені І.Я. Горбачевського

Міністерства охорони здоров'я України

Багрій-Заяць О.А.

## ДОДАТОК В

### Результати аналізу даних електроспоживання районів синфазним та компонентним методами ПКВП

З метою додаткової верифікації отриманих результатів проведено аналіз даних електроспоживання Чортківського та Збарзького районів із застосуванням синфазного та компонентного методів ПКВП. Результати аналізу даних електроспоживання Чортківського району наведено на рисунках В.1–В.6.

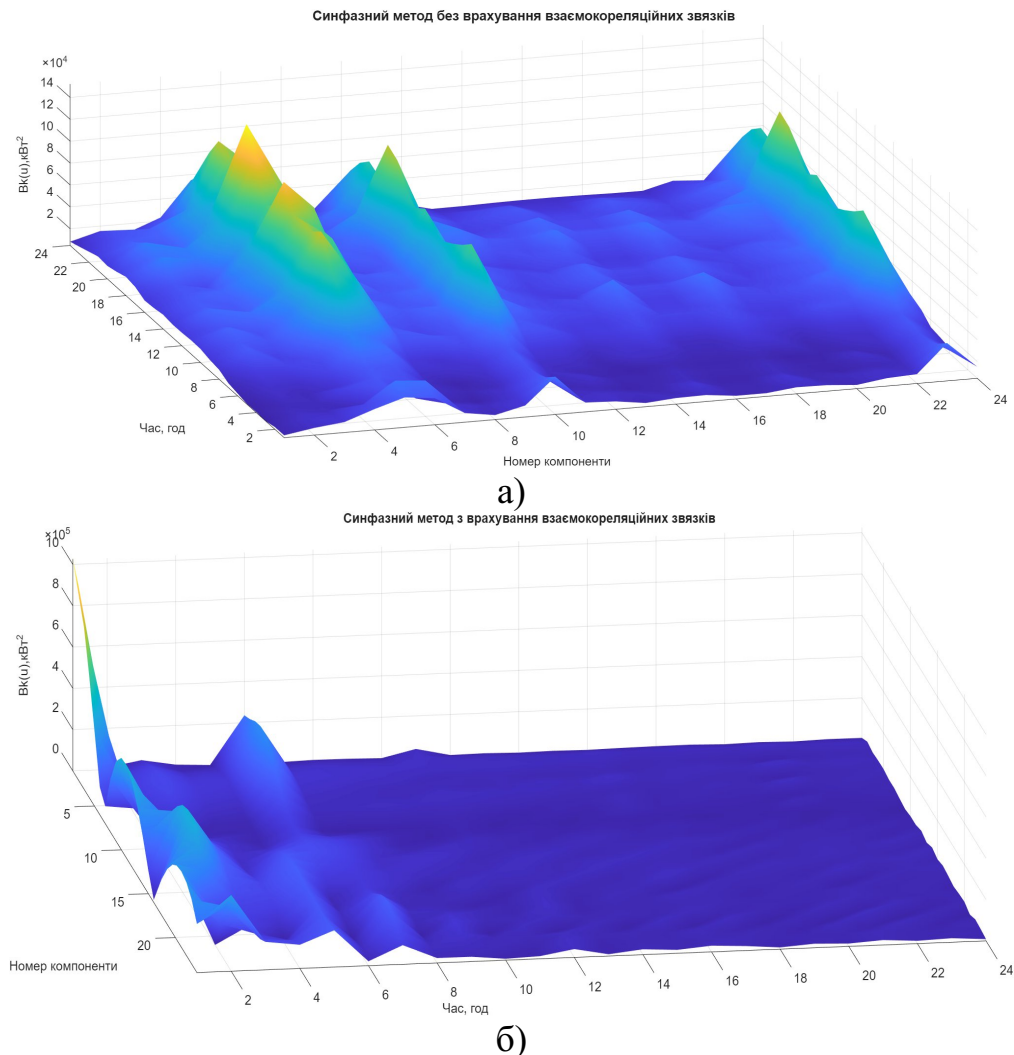


Рисунок В.1 – Тривимірне представлення синфазного методу аналізу ПКВП для Чортківського району:

- а) без врахування взаємкореляційних зв'язків;
- б) з врахуванням взаємкореляційних зв'язків

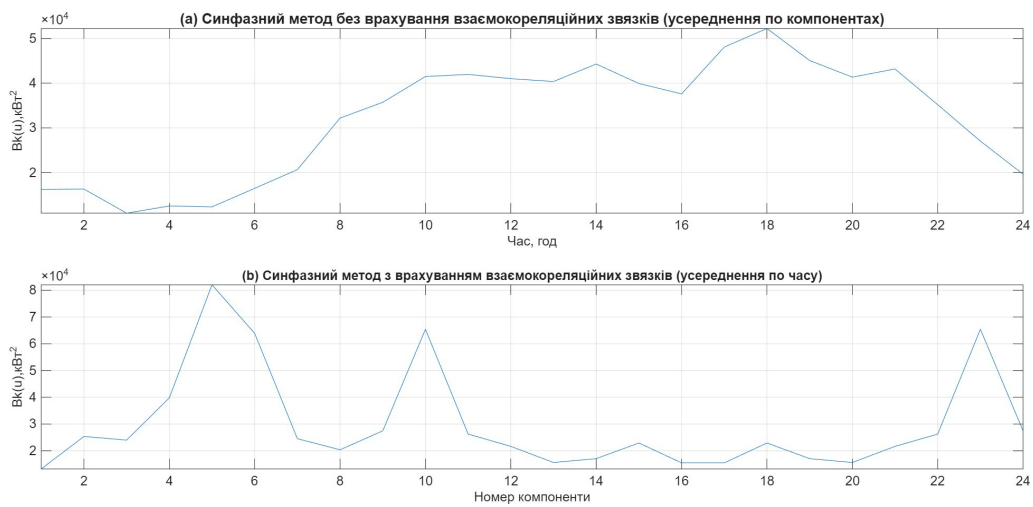


Рисунок В.2 – Усереднені кореляційні компоненти синфазного методу без урахування взаємкореляційних зв'язків для Чортківського району:

- а) усереднення за компонентами;
- б) усереднення за часовим зсувом

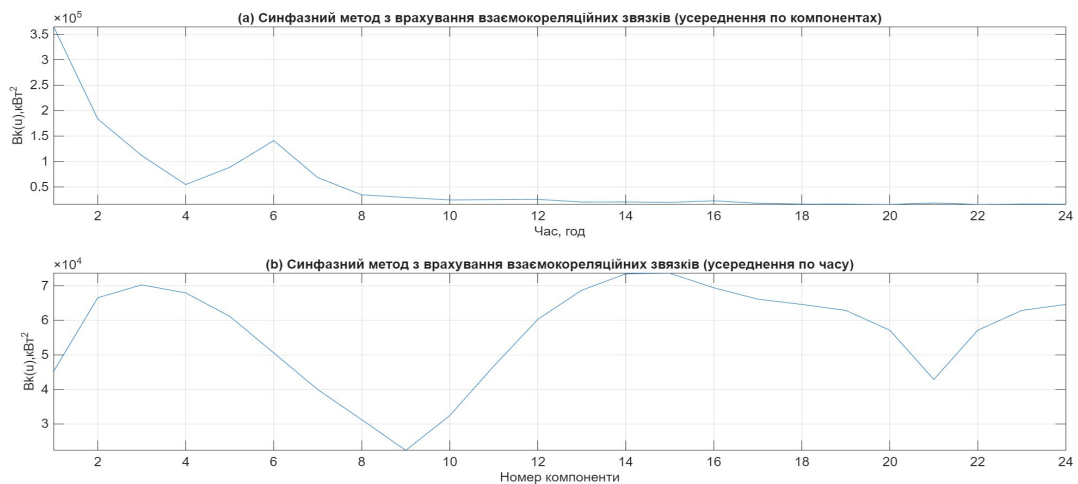


Рисунок В.3 – Усереднені кореляційні компоненти синфазного методу з урахуванням взаємкореляційних зв'язків для Чортківського району:

- а) усереднення за компонентами;
- б) усереднення за часовим зсувом

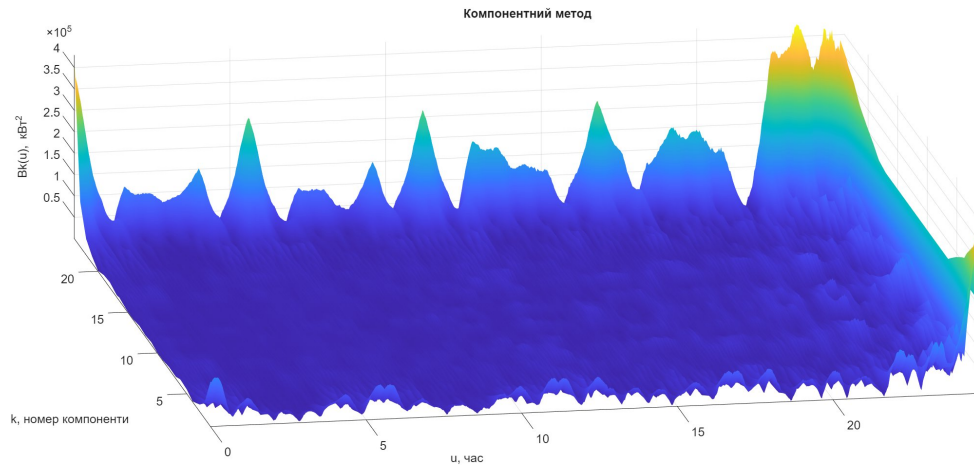


Рисунок В.4 – Тривимірне представлення матриці кореляційних компонент компонентного методу аналізу ПКВП для Чортківського району

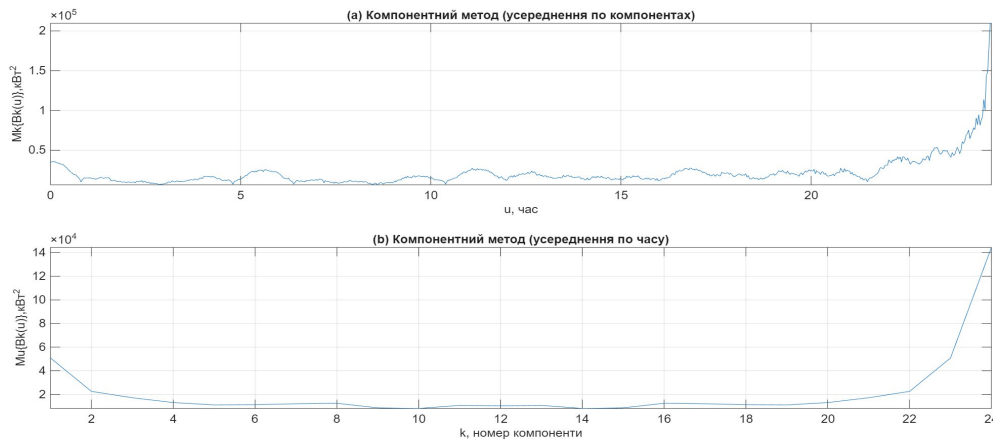


Рисунок В.5 – Результати компонентного методу аналізу ПКВП для Чортківського району:

- а) усереднення за компонентами;
- б) усереднення за часовим зсувом

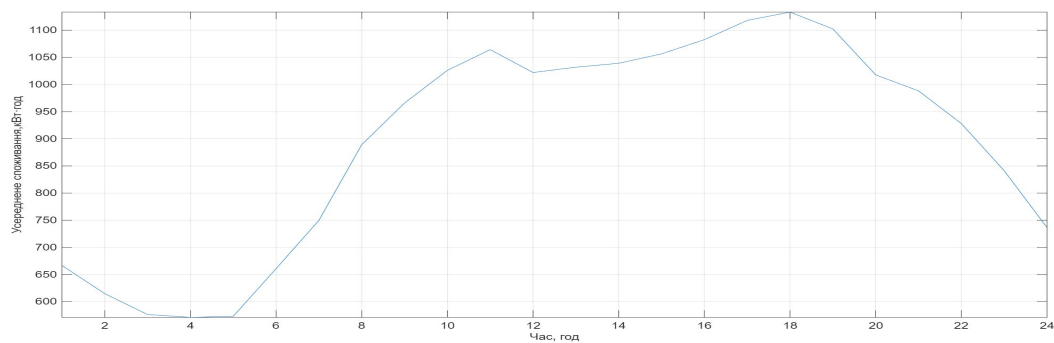
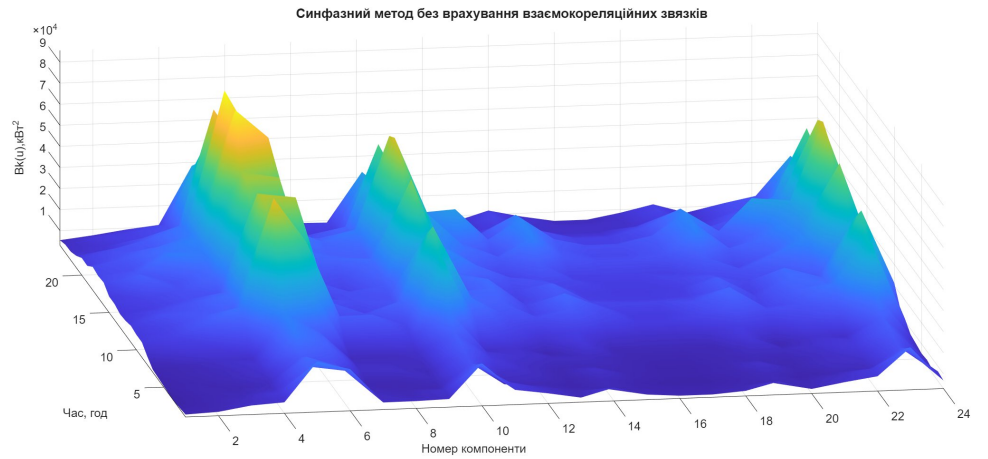
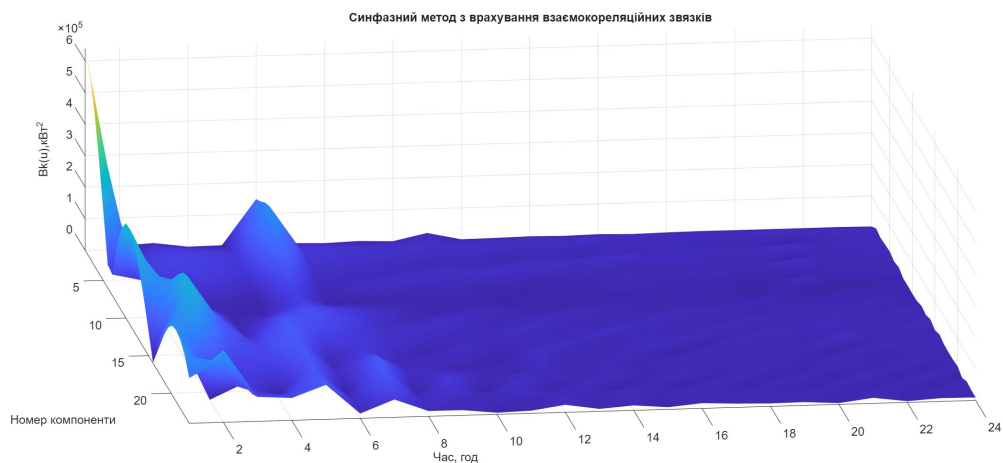


Рисунок В.6 – Усереднений добовий профіль електроспоживання для Чортківського району

Для даних електроспоживання Збарзького району проведено аналогічний аналіз із використанням синфазного та компонентного методів ПКВП. Отримані результати наведено на рисунках В.7–В.12.



а)



б)

Рисунок В.7 – Тривимірне представлення синфазного методу аналізу ПКВП для Збарзького району:

- а) без урахування взаємкореляційних зв'язків;
- б) з урахуванням взаємкореляційних зв'язків

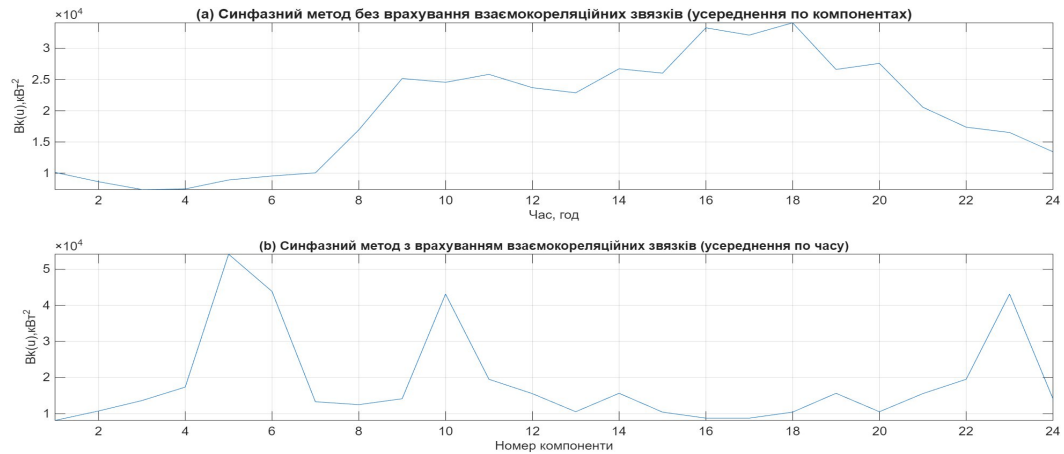


Рисунок В.8 – Усереднені кореляційні компоненти синфазного методу без урахування взаємкореляційних зв'язків для Збараського району:

- а) усереднення за компонентами;
- б) усереднення за часовим зсувом

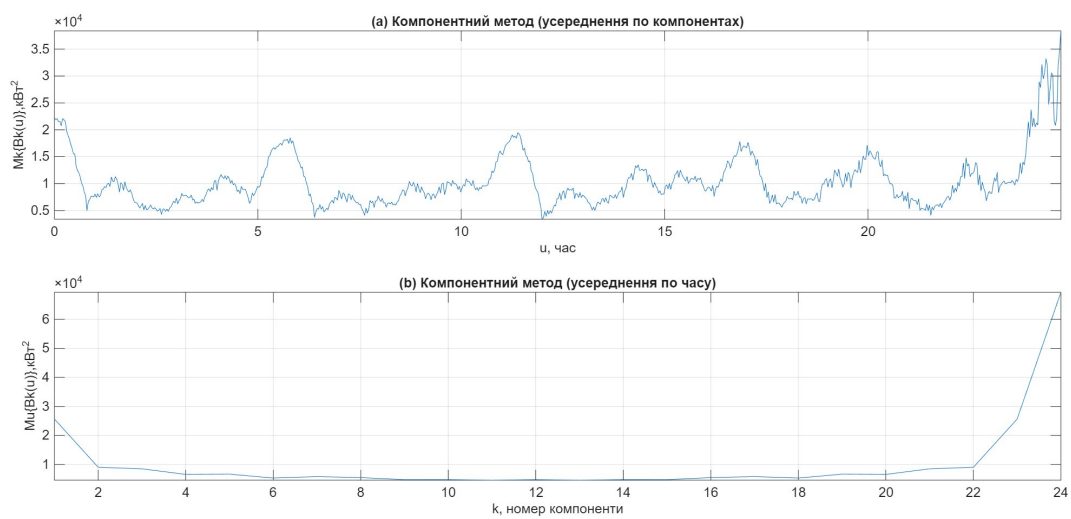


Рисунок В.9 – Усереднені кореляційні компоненти синфазного методу з урахуванням взаємкореляційних зв'язків для Збараського району:

- а) усереднення за компонентами;
- б) усереднення за часовим зсувом

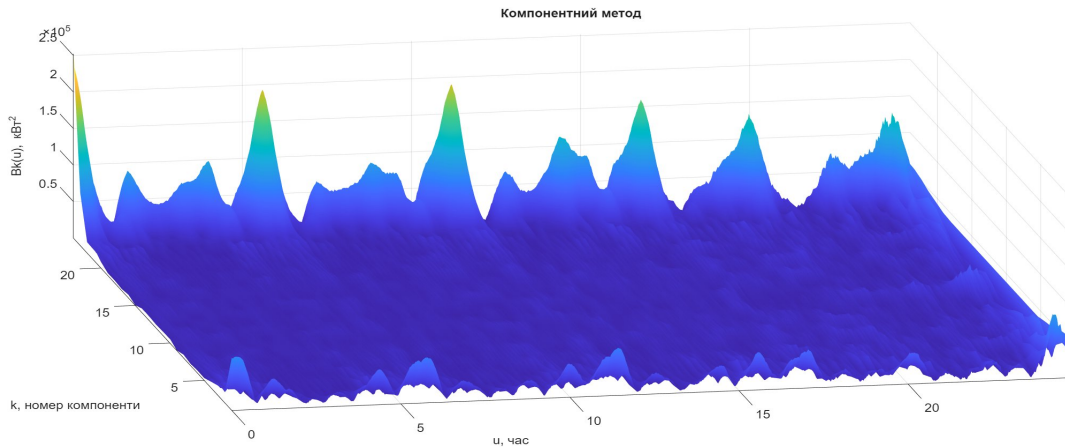


Рисунок В.10 – Тривимірне представлення матриці кореляційних компонент компонентного методу аналізу ПКВП для Збараського району

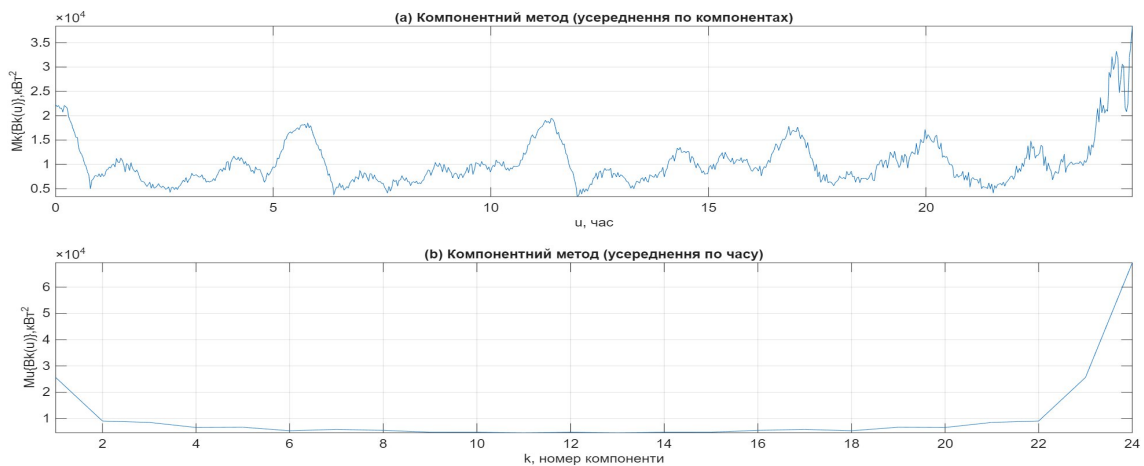


Рисунок В.11 – Результати компонентного методу аналізу ПКВП для Збараського району:

а) усереднення за компонентами;

б) усереднення за часовим зсувом

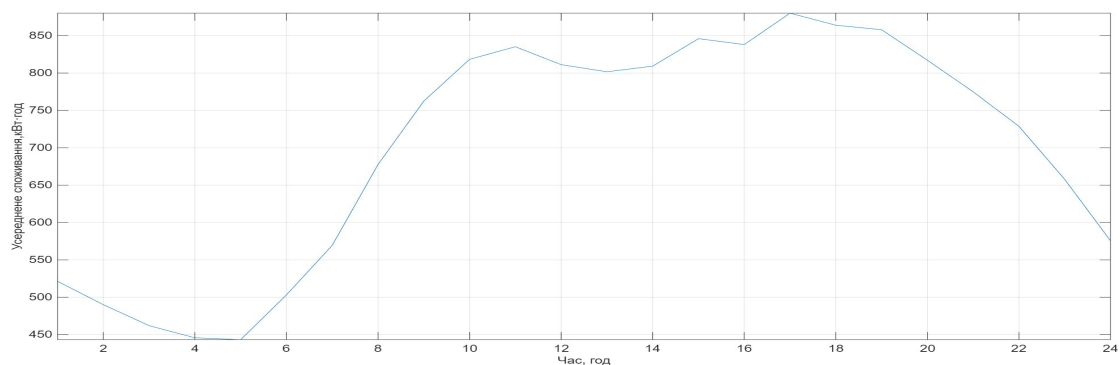


Рисунок В.12 – Усереднений добовий профіль електроспоживання для Збараського району

## ДОДАТОК Г

Фрагмент програмного коду розробленого ПЗ для вибору  
оптимального протоколу

## Фрагмент коду

```
=====
# IoT Protocol Selection
=====
```

```
import numpy as np
import pandas as pd
from sklearn.linear_model import LogisticRegression
from sklearn.ensemble import GradientBoostingClassifier, RandomForestClassifier
from sklearn.neural_network import MLPClassifier
from sklearn.preprocessing import StandardScaler, LabelBinarizer
from sklearn.model_selection import train_test_split, StratifiedKFold, cross_val_score
from sklearn.metrics import (accuracy_score, f1_score, precision_score, recall_score,
                             roc_auc_score, confusion_matrix, classification_report)
FEATURES = ['latency', 'jitter', 'packet_loss', 'throughput', 'payload_size']
TARGET = 'optimal_protocol'
PROTOCOLS = ['MQTT', 'CoAP', 'HTTPS']
MODELS = {
    'Logistic Regression': LogisticRegression(max_iter=1000, random_state=42),
    'Gradient Boosting': GradientBoostingClassifier(n_estimators=100, random_state=42),
    'Random Forest': RandomForestClassifier(n_estimators=200, random_state=42),
    'MLP Neural Network': MLPClassifier(hidden_layer_sizes=(64, 32),
                                       max_iter=500, random_state=42),
}
```

```
=====
# Section 1. Data loading
=====
```

```
def load_data(path: str):
    """
    Load the network dataset and return the feature matrix and label vector.
    Parameters
    -----
    path : str — path to network_dataset.csv.
    Returns
    -----
    X : ndarray, shape (n_samples, 5)
    y : ndarray, shape (n_samples,) — protocol labels
    """
    df = pd.read_csv(path)
    X = df[FEATURES].values
    y = df[TARGET].values
    return X, y
```

```
=====
# Section 2. Training and evaluation pipeline
```

---

```

def train_and_evaluate(X, y):
    """
    Train each classifier and compute hold-out and cross-validated metrics.
    Pipeline:
    1. Stratified 80/20 train-test split (random_state=42).
    2. Z-score standardisation — parameters fitted on training data only.
    3. For each model: fit, predict, compute accuracy / F1 / precision /
       recall / ROC-AUC (one-vs-rest, macro) on the held-out test set.
    4. 5-fold stratified cross-validation (F1-weighted) on training data.
    Returns
    -----
    trained : dict — fitted model objects keyed by name
    scaler  : StandardScaler fitted on X_train
    results : dict — evaluation metrics per model
    """
    X_train, X_test, y_train, y_test = train_test_split(
        X, y, test_size=0.2, stratify=y, random_state=42
    )
    scaler = StandardScaler()
    X_train = scaler.fit_transform(X_train)
    X_test = scaler.transform(X_test)
    cv = StratifiedKFold(n_splits=5, shuffle=True, random_state=42)
    results = {}
    trained = {}
    for name, model in MODELS.items():
        model.fit(X_train, y_train)
        y_pred = model.predict(X_test)
        # ROC-AUC: binarise labels for one-vs-rest multiclass
        if hasattr(model, 'predict_proba'):
            proba = model.predict_proba(X_test)
            auc = roc_auc_score(y_test, proba,
                                multi_class='ovr', average='macro',
                                labels=model.classes_)
        else:
            auc = float('nan')
        cv_f1 = cross_val_score(model, X_train, y_train,
                                cv=cv, scoring='f1_weighted')
        results[name] = {
            'accuracy': round(accuracy_score(y_test, y_pred), 4),
            'f1': round(f1_score(y_test, y_pred,
                                average='weighted', zero_division=0), 4),
            'precision': round(precision_score(y_test, y_pred,
                                                average='weighted', zero_division=0), 4),
            'recall': round(recall_score(y_test, y_pred,
                                         average='weighted', zero_division=0), 4),
            'auc': round(auc, 4),
            'cv_f1_mean': round(cv_f1.mean(), 4),
            'cv_f1_std': round(cv_f1.std(), 4),
            'confusion_matrix': confusion_matrix(y_test, y_pred,
                                                labels=model.classes_).tolist(),
        }

```

```

        'report': classification_report(y_test, y_pred,
                                       labels=model.classes_,
                                       output_dict=True, zero_division=0),
    }
    trained[name] = (model, y_test, y_pred)
    return trained, scaler, results
=====
# Section 3. Result reporting
=====
def print_results(results: dict) -> str:
    """
    Print a formatted comparison table and return the name of the best model.
    The best model is selected by the highest weighted F1-score on the
    held-out test partition.
    """
    sep = "-" * 80
    header = (f"\n{'Model':<22} {'Accuracy':>9} {'F1':>8} "
             f"{'Precision':>10} {'Recall':>8} {'AUC':>7} "
             f"{'CV F1':>8} {'±STD':>7}")
    print(header)
    print(sep)
    for name, m in results.items():
        auc_str = f"{m['auc']:>7.4f}" if not np.isnan(m['auc']) else f"{'—':>7}"
        print(f"{name:<22} {m['accuracy']:>9.4f} {m['f1']:>8.4f} "
              f"{m['precision']:>10.4f} {m['recall']:>8.4f} {auc_str} "
              f"{m['cv_f1_mean']:>8.4f} {m['cv_f1_std']:>7.4f}")
    best = max(results, key=lambda n: results[n]['f1'])
    print(f"\nBest model: {best}"
          f" (F1 = {results[best]['f1']:.4f}"
          f", AUC = {results[best]['auc']:.4f}"
          f", CV F1 = {results[best]['cv_f1_mean']:.4f}"
          f" ± {results[best]['cv_f1_std']:.4f})\n")
    return best
def print_confusion_matrices(trained: dict, results: dict):
    """Print confusion matrices for all classifiers side-by-side."""
    for name, (model, _, _) in trained.items():
        labels = list(model.classes_)
        cm = results[name]['confusion_matrix']
        col_w = 9
        print(f" {name}")
        print(" " + " " * 14 + "".join(f" {lbl:>{col_w}}" for lbl in labels)
              + " ← predicted")
        for lbl, row in zip(labels, cm):
            print(f" {lbl:<14}" + "".join(f" {v:>{col_w}}" for v in row))
        print()
def print_feature_importances(trained: dict):
    """Report MDI feature importances for the Random Forest classifier."""
    rf_entry = trained.get('Random Forest')
    if rf_entry is None:
        return
    rf_model = rf_entry[0]

```

```

ranked = sorted(zip(FEATURES, rf_model.feature_importances_),
                key=lambda x: x[1], reverse=True)
max_imp = ranked[0][1]
print(" Feature importances — Random Forest (MDI)")
print(f' {Feature':<16} {Importance':>11} Bar")
print(" " + "-" * 45)
for feat, imp in ranked:
    bar = " * int(imp / max_imp * 24)
    print(f' {feat:<16} {imp:>11.4f} {bar}")
print()
=====
# Section 4. Protocol inference on new observations
=====
def select_protocol(conditions: dict, model, scaler) -> dict:
    """
    Predict the optimal protocol for a single set of network conditions.
    Parameters
    conditions : dict — keys: latency, jitter, packet_loss,
                       throughput, payload_size
    model      : fitted classifier with predict / predict_proba
    scaler     : StandardScaler fitted on training data
    Returns
    -----
    dict with 'protocol' and (if available) 'probabilities'
    """
    x = np.array([[conditions[f] for f in FEATURES]])
    x = scaler.transform(x)
    protocol = model.predict(x)[0]
    proba_raw = model.predict_proba(x)[0] if hasattr(model, 'predict_proba') else None
    result = {'protocol': protocol}
    if proba_raw is not None:
        result['probabilities'] = {c: round(float(p), 4)
                                   for c, p in zip(model.classes_, proba_raw)}
    return result
=====
# Entry point
=====
if __name__ == '__main__':
    X, y = load_data('network_dataset.csv')
    print(f'Dataset: {len(y)} samples | classes: {np.unique(y).tolist()}")
    trained_models, scaler, results = train_and_evaluate(X, y)
    best_name = print_results(results)
    print("Confusion matrices (held-out test set):")
    print_confusion_matrices(trained_models, results)
    print("Feature importances:")
    print_feature_importances(trained_models)
    # Inference example: predict protocol for a new network observation
    sample = {
        'latency': 42.5, # ms
        'jitter': 8.3, # ms
        'packet_loss': 1.2, # %

```

```
'throughput': 18.7, # Mbit/s
'payload_size': 512, # bytes
}
best_model, _, _ = trained_models[best_name]
decision = select_protocol(sample, best_model, scaler)
print(f'Input conditions : {sample}')
print(f'Recommended protocol: {decision['protocol']}')
if 'probabilities' in decision:
    print("Per-class probabilities:")
    for proto, prob in sorted(decision['probabilities'].items(),
                             key=lambda x: x[1], reverse=True):
        print(f' {proto}: {prob:.4f}')
```

## ДОДАТОК Д

### Фрагмент програмного коду розробленого ПЗ для класифікації режимів електроспоживання

#### Фрагмент програми

```
# Power consumption mode classification module for smart-grid network nodes
# -----
# Fragment 1. Database models (consumption_classifier/models.py)
# -----
from django.db import models
class ClassificationRun(models.Model):
    """Single training-and-evaluation session executed on the current dataset."""
    STATUS = [('running', 'Running'), ('completed', 'Completed'), ('failed', 'Failed')]
    created_at = models.DateTimeField(auto_now_add=True)
    status = models.CharField(max_length=20, choices=STATUS, default='running')
    n_class1 = models.IntegerField(default=0) # number of Class 1 signals (balanced mode)
    n_class2 = models.IntegerField(default=0) # number of Class 2 signals (unbalanced mode)
    class Meta:
        ordering = ['-created_at']
class ClassifierResult(models.Model):
    """Per-classifier evaluation metrics recorded within a classification run."""
    run = models.ForeignKey(ClassificationRun, on_delete=models.CASCADE,
                            related_name='results')
    model_name = models.CharField(max_length=50)
    accuracy = models.FloatField()
    auc = models.FloatField()
    cv_f1_mean = models.FloatField()
    cv_f1_std = models.FloatField()
    report = models.JSONField()
    confusion_matrix = models.JSONField()
    class Meta:
        ordering = ['-accuracy']
        unique_together = ('run', 'model_name')
# -----
# Fragment 2. PKVP dataset loading and 13-feature extraction
# -----
import numpy as np
from scipy.optimize import curve_fit
from scipy.stats import skew, kurtosis
FEATURE_NAMES = [
    'exp_a', # amplitude coefficient a of the fit  $y = a \cdot \exp(b \cdot x)$ 
    'exp_b', # exponential rate b of the trend model
    'exp_rsquare', # coefficient of determination  $R^2$  for the exponential fit
    'exp_rmse', # root-mean-square error of the exponential approximation
    'tangent_45', # sample index where  $dy/dx = 1$  (45° tangent point)
    'std_before_45', # std of de-trended residuals before the tangent point
    'std_after_45', # std of de-trended residuals after the tangent point
    'skew_after_45', # skewness of residuals in the post-tangent region
    'kurt_after_45', # excess kurtosis of residuals in the post-tangent region
    'time_mean', # mean of the 24-hour averaged load profile
    'time_std', # standard deviation of the daily load profile
```

```

'time_skew', # skewness of the daily load profile
'time_kurt', # excess kurtosis of the daily load profile
]
def _exp_model(x, a, b):
    return a * np.exp(b * x)
def extract_comp_features(signal: np.ndarray) -> np.ndarray:
    """Extract 4 statistical descriptors of residuals before and after x*."""
    n = len(signal)
    x = np.linspace(0, 1, n)
    try:
        popt, _ = curve_fit(_exp_model, x, signal,
                            p0=[signal[0], 0.01], maxfev=5000)
        a, b = popt
        y_fit = _exp_model(x, a, b)
        ss_res = np.sum((signal - y_fit) ** 2)
        ss_tot = np.sum((signal - signal.mean()) ** 2)
        r2 = 1.0 - ss_res / (ss_tot + 1e-9)
        rmse = np.sqrt(ss_res / n)
    except RuntimeError:
        a, b, r2, rmse = 0.0, 0.0, 0.0, 0.0
        y_fit = np.zeros(n)
    # Tangent point: d/dx[a·exp(b·x)] = 1 → x* = ln(1/(a·b)) / b
    if a * b > 0:
        x_45 = np.log(1.0 / (a * b)) / b
        idx_45 = int(np.clip(round(x_45 * (n - 1)), 0, n - 1))
    else:
        idx_45 = n // 2 # fallback to the signal midpoint
    y_res = signal - y_fit
    before = y_res[:idx_45] if idx_45 > 0 else y_res[:1]
    after = y_res[idx_45:] if idx_45 < n else y_res[-1:]
    return np.array([
        a, b, r2, rmse, float(idx_45),
        float(np.std(before)),
        float(np.std(after)),
        float(skew(after)),
        float(kurtosis(after)), # excess kurtosis (Fisher definition)
    ])
def extract_time_features(signal: np.ndarray) -> np.ndarray:
    """Extract 4 statistical moments from a 24-sample averaged daily load profile."""
    return np.array([
        float(signal.mean()),
        float(signal.std()),
        float(skew(signal)),
        float(kurtosis(signal)),
    ])
def load_pkvp_dataset(comp1_path: str, comp2_path: str,
                     time1_path: str, time2_path: str):
    """
    Load four PKVP CSV files (produced by the MATLAB pre-processing script)
    and assemble the feature matrix X and label vector y.
    Expected file dimensions:
    data_aver_comp_class1.csv — 720 × 206 (Class 1, balanced regime)
    data_aver_comp_class2.csv — 720 × 159 (Class 2, unbalanced regime)
    data_aver_time_class1.csv — 24 × 206
    data_aver_time_class2.csv — 24 × 159
    """

```

```

Returns: X (n_samples × 13 float64),
        y (n_samples, int — 0 = Class 1, 1 = Class 2)
comp1 = np.loadtxt(comp1_path, delimiter=',') # 720 × 206
comp2 = np.loadtxt(comp2_path, delimiter=',') # 720 × 159
time1 = np.loadtxt(time1_path, delimiter=',') # 24 × 206
time2 = np.loadtxt(time2_path, delimiter=',') # 24 × 159
X, y = [], []
for i in range(comp1.shape[1]):
    feat_c = extract_comp_features(comp1[:, i])
    feat_t = extract_time_features(time1[:, i])
    X.append(np.concatenate([feat_c, feat_t]))
    y.append(0) # Class 1 — balanced consumption mode
for i in range(comp2.shape[1]):
    feat_c = extract_comp_features(comp2[:, i])
    feat_t = extract_time_features(time2[:, i])
    X.append(np.concatenate([feat_c, feat_t]))
    y.append(1) # Class 2 — unbalanced consumption mode
return np.array(X), np.array(y)
# -----
# Fragment 3. Classification pipeline
# -----
from sklearn.preprocessing import StandardScaler
from sklearn.model_selection import train_test_split, StratifiedKFold, cross_val_score
from sklearn.metrics import (accuracy_score, roc_auc_score,
                             classification_report, confusion_matrix)
from sklearn.ensemble import (RandomForestClassifier, ExtraTreesClassifier,
                              GradientBoostingClassifier, VotingClassifier)
from sklearn.svm import SVC
from sklearn.neural_network import MLPClassifier
from xgboost import XGBClassifier
RANDOM_STATE = 42
CLASSIFIERS = {
    'Extra Trees': ExtraTreesClassifier(n_estimators=300,
                                       random_state=RANDOM_STATE),
    'Random Forest': RandomForestClassifier(n_estimators=300, max_depth=8,
                                           random_state=RANDOM_STATE),
    'SVM (RBF)': SVC(kernel='rbf', C=10, gamma='scale',
                    probability=True, random_state=RANDOM_STATE),
    'MLP Neural Net': MLPClassifier(hidden_layer_sizes=(128, 64),
                                    max_iter=500, random_state=RANDOM_STATE),
    'Gradient Boosting': GradientBoostingClassifier(n_estimators=200,
                                                  learning_rate=0.05,
                                                  max_depth=5,
                                                  random_state=RANDOM_STATE),
    'XGBoost': XGBClassifier(n_estimators=200, learning_rate=0.05,
                            max_depth=5, random_state=RANDOM_STATE,
                            verbosity=0),
    'Voting Ensemble': VotingClassifier(voting='soft', estimators=[
        ('et', ExtraTreesClassifier(n_estimators=300, random_state=RANDOM_STATE)),
        ('svm', SVC(kernel='rbf', C=10, probability=True, random_state=RANDOM_STATE)),
        ('xgb', XGBClassifier(n_estimators=200, random_state=RANDOM_STATE,
                             verbosity=0)),
    ]),
}
def run_classification(data_dir: str, run) -> dict:
    import os

```

```

X, y = load_pkvp_dataset(
    comp1_path=os.path.join(data_dir, 'data_aver_comp_class1.csv'),
    comp2_path=os.path.join(data_dir, 'data_aver_comp_class2.csv'),
    time1_path=os.path.join(data_dir, 'data_aver_time_class1.csv'),
    time2_path=os.path.join(data_dir, 'data_aver_time_class2.csv'),
)
run.n_class1 = int((y == 0).sum())
run.n_class2 = int((y == 1).sum())
run.save()
X_tr, X_te, y_tr, y_te = train_test_split(
    X, y, test_size=0.25, stratify=y, random_state=RANDOM_STATE
)
scaler = StandardScaler()
Xs_tr = scaler.fit_transform(X_tr)
Xs_te = scaler.transform(X_te)
cv = StratifiedKFold(n_splits=5, shuffle=True, random_state=RANDOM_STATE)
all_results = {}
for name, clf in CLASSIFIERS.items():
    clf.fit(Xs_tr, y_tr)
    y_pred = clf.predict(Xs_te)
    proba = (clf.predict_proba(Xs_te)[: , 1]
             if hasattr(clf, 'predict_proba')
             else clf.decision_function(Xs_te))
    acc = accuracy_score(y_te, y_pred)
    auc = roc_auc_score(y_te, proba)
    report = classification_report(
        y_te, y_pred,
        target_names=['Balanced', 'Unbalanced'],
        output_dict=True, zero_division=0)
    cm = confusion_matrix(y_te, y_pred, labels=[0, 1]).tolist()
    cv_f1 = cross_val_score(clf, Xs_tr, y_tr, cv=cv,
                           scoring='f1_weighted')
    ClassifierResult.objects.create(
        run=run, model_name=name,
        accuracy=round(acc, 6), auc=round(auc, 6),
        cv_f1_mean=round(cv_f1.mean(), 6),
        cv_f1_std=round(cv_f1.std(), 6),
        report=report, confusion_matrix=cm,
    )
    all_results[name] = {'accuracy': acc, 'auc': auc}
return all_results
# -----
# Fragment 4. REST API (consumption_classifier/views.py)
# -----
from rest_framework import status
from rest_framework.views import APIView
from rest_framework.response import Response
from django.conf import settings
import os
DATA_DIR = os.path.join(settings.BASE_DIR, 'data')
class RunClassificationView(APIView):
    """POST /api/classification/run/ — trigger the full classification pipeline."""
    def post(self, request):
        run = ClassificationRun.objects.create(status='running')
        try:
            results = run_classification(DATA_DIR, run)

```

```

run.status = 'completed'
run.save()
return Response({'run_id': run.id, 'models': results},
                status=status.HTTP_201_CREATED)
except Exception as e:
run.status = 'failed'
run.save()
return Response({'error': str(e)}, status=status.HTTP_400_BAD_REQUEST)
class ClassificationResultsView(APIView):
    """GET /api/classification/results/ — retrieve metrics for the latest completed run."""
    def get(self, request):
run = ClassificationRun.objects.filter(status='completed').first()
if not run:
return Response({'error': 'No completed run'},
                status=status.HTTP_404_NOT_FOUND)
return Response({
    'run_id': run.id,
    'n_class1': run.n_class1,
    'n_class2': run.n_class2,
    'results': [
        {
            'model': r.model_name,
            'accuracy': r.accuracy,
            'auc': r.auc,
            'cv_f1_mean': r.cv_f1_mean,
            'cv_f1_std': r.cv_f1_std,
            'confusion_matrix': r.confusion_matrix,
        }
        for r in run.results.order_by('-accuracy')
    ],
})

```

## ДОДАТОК Е

## Фрагмент програмного коду з середовища MATLAB

## Фрагмент коду

```

% Preparation of PKVP dataset: loading, averaging, signal generation
% Class 1 — balanced (n=206), Class 2 — unbalanced (n=159)
clear; close all; clc; rng('default')
%addpath(genpath('e:\matlab_libs\ThirdParty\'),'end');
addpath(genpath('e:\matlab_libs\Palaniza\'),'end'); addpath(genpath('e:\matlab_libs\m\'),'end');
addpath(genpath('d:\matlab_libs\Palaniza\'),'end');
% openvar('MuyVar')
%% =====
% LOAD .DAT DATA
%% =====
comp_b = importdata('Component method.dat'); % Class 1 "Balanced"
comp_u = importdata('Component method_peak.dat'); % Class 2 "Unbalanced"
%% =====
% TRANSPOSE TO 720x24
% Rows = 720 time samples (u, hour)
% Columns = 24 components (k)
%% =====
comp_n_csv = comp_b'; % Class 1 "Balanced"
comp_p_csv = comp_u'; % Class 2 "Unbalanced"
% writematrix(comp_n_csv, 'Component_method.csv');
% writematrix(comp_p_csv, 'Component_method_peak.csv');
%% =====
% STORE FOR LOOPING
%% =====
data_all = {comp_n_csv, comp_p_csv};
class_names = {'Class 1', 'Class 2'};
%% =====
% SURFACE PLOTS
%% =====
for i = 1:2
    data = data_all{i};
    % Dimensions
    [n_samples, n_components] = size(data); % n_samples = 720, n_components = 24
    % Axes
    u = linspace(0, 24, n_samples); % 720 time samples over 24 hours
    k = 1:n_components; % 24 components
    [K, U] = meshgrid(k, u); % columns = components, rows = time
    figure('Color','w')
    surf(K, U, data, 'EdgeColor','none')
    shading interp
    colormap(jet)
    view(-45, 30)
    xlabel('k, component number')
    ylabel('u, hour')
    zlabel('B_k(u), kW^2')
    title(['Component method - ', class_names{i}])
    colorbar
    axis tight

```

```

end
%% =====
% AVERAGING
%% =====
avg_over_components_all = cell(1,2); % average over components → vs time
avg_over_time_all      = cell(1,2); % average over time → vs component
for i = 1:2
    data = data_all{i};
    % Average over components (columns)
    avg_over_components_all{i} = mean(data, 2); % 720x1
    % Average over time (rows)
    avg_over_time_all{i} = mean(data, 1); % 1x24
end
%% =====
% PLOT: Averaging over components (vs time)
%% =====
figure('Color','w'); hold on
plot(u, avg_over_components_all{1}, 'LineWidth', 2)
plot(u, avg_over_components_all{2}, 'LineWidth', 2)
grid on
xlabel('u, hour')
ylabel('Average B_k(u), kW^2')
title('Component method (averaging over components)')
legend(class_names, 'Location', 'best')
hold off
%% =====
% PLOT: Time averaging (vs components)
%% =====
figure('Color','w'); hold on
plot(k, avg_over_time_all{1}, 'LineWidth', 2)
plot(k, avg_over_time_all{2}, 'LineWidth', 2)
grid on
xlabel('k, component number')
ylabel('Average B_k(u), kW^2')
title('Component method (time averaging)')
legend(class_names, 'Location', 'best')
hold off
%% =====
% SAVE AVERAGED CURVES TO CSV
%% =====
% Average over components → 720x1
writematrix(avg_over_components_all{1}, 'aver_comp_class1.csv')
writematrix(avg_over_components_all{2}, 'aver_comp_class2.csv')
% Average over time → 1x24 → save as 24x1
writematrix(avg_over_time_all{1}, 'aver_time_class1.csv')
writematrix(avg_over_time_all{2}, 'aver_time_class2.csv')
disp('Saved: aver_comp_class1.csv, aver_comp_class2.csv')
disp('Saved: aver_time_class1.csv, aver_time_class2.csv')
% =====
% GET DATA 'aver_comp_class1.csv'
% =====
clearvars; clc;
% =====
% LOAD BASE SIGNAL
% =====
% Averaged over time → 720x1

```

```

signal_name = 'aver_comp_class1.csv';
% signal_name = 'aver_comp_class2.csv';
% % Averaged over components → 24x1
% signal_name = 'aver_time_class1.csv';
% signal_name = 'aver_time_class2.csv';
base_signal = readmatrix(signal_name);
base_signal = base_signal(:);% ensure column vector
n_samples = length(base_signal);
% =====
% PARAMETERS
% =====
n_signals = 206; % number of signals
noise_level = 5; % max noise level (e.g. 0.10 = 10%)
% =====
% CREATE LOG TREND (0 → 1)
% =====
log_b = 0.01;
log_c = 0.1; % shift (horizontal shift, controls where log starts)
k_vec = linspace(0,1,n_samples);
trend = log(log_b + k_vec + log_c);
% Normalize to [0, 1]
% trend = (trend - min(trend)) / (max(trend) - min(trend));
trend = mapP(trend, [], [0.1 0.6]); %
% --- OPTIONAL: reverse trend ---
trend = flip(trend); %
% =====
% VISUALIZE TREND
% =====
figure('Color','w')
plot(linspace(0,24,n_samples), trend, 'LineWidth',2)
grid on
xlabel('u, hour')
ylabel('Trend (0 → 1)')
title('Logarithmic Trend (controls noise level)')
% =====
% GET SIGNALS
% =====
signal_final = zeros(n_samples, n_signals);
for i = 1:n_signals
    signal = base_signal;
    % =====
    % NOISE MODE (UNCOMMENT ONE)
    % =====
    skew_deg = 3; % >0 → stronger skew (try 1–10)
    % --- 1. Gaussian noise (symmetric) ---
    % noise = randn(n_samples,1);
    % --- 2. Right-skewed noise ---
    % Prepare skew using exponential transform
    % noise = randn(n_samples,1);
    % noise = exp(skew_deg * noise);
    % noise = noise - mean(noise); % center
    % --- 3. Left-skewed noise ---
    % noise = randn(n_samples,1);
    % noise = -exp(skew_deg * noise);
    % noise = noise - mean(noise); % center
    % 4. Alternative: Gamma-based skew (more physical)

```

```

% Right-skew:
% noise = gamrnd(skew_deg, 1, n_samples, 1);
% noise = noise - mean(noise);
% Left-skew:
noise = -gamrnd(skew_deg, 1, n_samples, 1);
noise = noise - mean(noise);
% =====
% APPLY TREND TO NOISE LEVEL
% =====
% Noise increases along signal according to trend
% Left: 0% noise → Right: noise_level %
scaled_noise = trend .* (noise_level * signal .* noise);
% Final signal
signal_final(:,i) = signal + scaled_noise;
end
% =====
% SAVE TO FILE
% =====
filename = strcat("data_", signal_name);
writematrix(signal_final, filename);
disp(strcat(filename, " | ", num2str(n_signals), " signals prepared"))
%% =====
% PLOT MEAN WITH CONFIDENCE INTERVAL
% READ SIGNALS
% =====
% Averaged over time → 720xn_signals
signal_name = 'data_aver_comp_class1.csv';
% signal_name = 'data_aver_comp_class2.csv';
% % Averaged over components → 24xn_signals
% signal_name = 'data_aver_time_class1.csv';
% signal_name = 'data_aver_time_class2.csv';
signals = readmatrix(signal_name); % size: n_samples x n_signals
[n_samples, n_signals] = size(signals);
% Time axis
u = linspace(0, 24, n_samples);% 720 samples over 24 hours
% =====
% COMPUTE MEAN AND CONFIDENCE INTERVAL
% =====
mean_signal = mean(signals, 2);% mean over signals
std_signal = std(signals, 0, 2);% standard deviation over signals
% 95% Confidence interval: mean ± 1.96*std/sqrt(N)
ci_upper = mean_signal + 1.96 * std_signal / sqrt(n_signals);
ci_lower = mean_signal - 1.96 * std_signal / sqrt(n_signals);
% =====
% PLOT MEAN WITH CONFIDENCE INTERVAL
% =====
figure('Color','w'); hold on
% Shaded confidence interval
fill([u, fliplr(u)], [ci_upper', fliplr(ci_lower')], ...
     [0.9 0.9 1], 'EdgeColor','none'); % light blue
% Plot mean
plot(u, mean_signal, 'b', 'LineWidth', 2)
grid on, %zoom xon
xlabel('u, hour')
ylabel('B_k(u), kW^2')
title('Mean Signal with 95% Confidence Interval')

```

```

legend('95% CI','Mean', 'Location','best')
hold off
% =====
% GET DATA 'aver_time_class1.csv'
% =====
clearvars; clc;
% =====
% LOAD SIGNAL
% =====
% % Averaged over components → 720x1
% signal_name = 'aver_comp_class1.csv';
% signal_name = 'aver_comp_class2.csv';
% Averaged over time → 24x1
signal_name = 'aver_time_class1.csv';
% signal_name = 'aver_time_class2.csv';
base_signal = readmatrix(signal_name);
base_signal = base_signal(:); % ensure column vector
n_samples = length(base_signal);
% =====
% PARAMETERS
% =====
n_signals = 206; % number of signal
noise_level = 3; % max noise level (e.g. 0.10 = 10%)
% =====
% CREATE LOG TREND (0 → 1)
% =====
log_b = 0.01;
log_c = 0.9; % shift (horizontal shift, controls where log starts)
k_vec = linspace(0,1,n_samples);
trend = log(log_b + k_vec + log_c);
% Normalize to [0, 1]
% trend = (trend - min(trend)) / (max(trend) - min(trend));
trend = mapP(trend, [], [0.8 0.9]); %
% --- OPTIONAL: reverse trend ---
trend = flip(trend); %
% =====
% VISUALIZE TREND
% =====
figure('Color','w')
plot(linspace(0,24,n_samples), trend, 'LineWidth',2)
grid on
xlabel('u, hour')
ylabel('Trend (0 → 1)')
title('Logarithmic Trend (controls noise level)')
% =====
% GET SIGNALS
% =====
signal_final = zeros(n_samples, n_signals);
for i = 1:n_signals
    signal = base_signal;
    % =====
    % NOISE MODE (UNCOMMENT ONE)
    % =====
    skew_deg = 3; % >0 → stronger skew (try 1–10)
    % --- 1. Gaussian noise (symmetric)
    % noise = randn(n_samples,1);

```

```

% --- 2. Right-skewed noise ---
% Prepare skew using exponential transform
% noise = randn(n_samples,1);
% noise = exp(skew_deg * noise);
% noise = noise - mean(noise); % center
% --- 3. Left-skewed noise ---
% noise = randn(n_samples,1);
% noise = -exp(skew_deg * noise);
% noise = noise - mean(noise); % center
% --- 4. Alternative: Gamma-based skew (more physical)
% Right-skew:
% noise = gamrnd(skew_deg, 1, n_samples, 1);
% noise = noise - mean(noise);
% Left-skew:
noise = -gamrnd(skew_deg, 1, n_samples, 1);
noise = noise - mean(noise);
% =====
% APPLY TREND TO NOISE LEVEL
% =====
% Noise increases along signal according to trend
% Left: 0% noise → Right: noise_level %
scaled_noise = trend .* (noise_level * signal .* noise);
% Final signal
signal_final(:,i) = signal + scaled_noise;
end
% =====
% SAVE TO FILE
% =====
filename = strcat("data_", signal_name);
writematrix(signal_final, filename);
disp(strcat(filename, " | ", num2str(n_signals), " signals prepared"))
%% =====
% PLOT MEAN WITH CONFIDENCE INTERVAL
% READ SIGNALS
% =====
% Averaged over time → 720xn_signals
% signal_name = 'data_aver_comp_class1.csv';
% signal_name = 'data_aver_comp_class2.csv';
% % Averaged over components → 24xn_signals
signal_name = 'data_aver_time_class1.csv';
% signal_name = 'data_aver_time_class2.csv';
signals = readmatrix(signal_name); % size: n_samples x n_signals
[n_samples, n_signals] = size(signals);
% Time axis
u = linspace(0, 24, n_samples); % 720 samples over 24 hours
% =====
% COMPUTE MEAN AND CONFIDENCE INTERVAL
% =====
mean_signal = mean(signals, 2); % mean over signals
std_signal = std(signals, 0, 2); % standard deviation over signals
% 95% Confidence interval: mean ± 1.96*std/sqrt(N)
ci_upper = mean_signal + 1.96 * std_signal / sqrt(n_signals);
ci_lower = mean_signal - 1.96 * std_signal / sqrt(n_signals);
% =====
% PLOT MEAN WITH CONFIDENCE INTERVAL
% =====

```

```

figure('Color','w'); hold on
% Shaded confidence interval
fill([u, flipr(u)], [ci_upper', flipr(ci_lower')], ...
     [0.9 0.9 1], 'EdgeColor','none'); % light blue
% Plot mean
plot(u, mean_signal, 'b', 'LineWidth', 2)
grid on, %zoom xon
xlabel('u, hour')
ylabel('B_k(u), kW^2')
title('Mean Signal with 95% Confidence Interval')
legend('95% CI','Mean', 'Location','best')
% =====
% GET DATA 'aver_time_class2.csv'
% =====
clearvars; clc;
% =====
% LOAD BASE SIGNAL
% =====
% Averaged over time → 720x1
% signal_name = 'aver_comp_class1.csv';
% signal_name = 'aver_comp_class2.csv';
% % Averaged over components → 24x1
% signal_name = 'aver_time_class1.csv';
signal_name = 'aver_time_class2.csv';
base_signal = readmatrix(signal_name);
base_signal = base_signal(:); % ensure column vector
n_samples = length(base_signal);
% =====
% PARAMETERS
% =====
n_signals = 159; % number of signals )noise_level = 5; % max noise level (e.g. 0.10 = 10%)
% =====
% CREATE LOG TREND (0 → 1)
% =====
log_b = 0.01;
log_c = 0.9; % shift (horizontal shift, controls where log
starts)<=====
k_vec = linspace(0,1,n_samples);
trend = log(log_b + k_vec + log_c);
% Normalize to [0, 1]
% trend = (trend - min(trend)) / (max(trend) - min(trend));
trend = mapP(trend, [ ], [0.8 0.9]);
%<=====
% --- OPTIONAL: reverse trend
% trend = flip(trend); % <=====
% =====
% VISUALIZE TREND
% =====
figure('Color','w')
plot(linspace(0,24,n_samples), trend, 'LineWidth',2)
grid on
xlabel('u, hour')
ylabel('Trend (0 → 1)')
title('Logarithmic Trend (controls noise level)')
% =====
% GET SIGNALS

```

```

% =====
signal_final = zeros(n_samples, n_signals);
for i = 1:n_signals
    signal = base_signal;
    skew_deg = 3; % >0 → stronger skew (try 1–10)
<=====
    % --- 1. Gaussian noise (symmetric) ---
    % noise = randn(n_samples,1);
    % --- 2. Right-skewed noise ---
    % Prepare skew using exponential transform
    % noise = randn(n_samples,1);
    % noise = exp(skew_deg * noise);
    % noise = noise - mean(noise); % center
    % --- 3. Left-skewed noise ---
    % noise = randn(n_samples,1);
    % noise = -exp(skew_deg * noise);
    % noise = noise - mean(noise); % center
    % --- 4. Alternative: Gamma-based skew (more physical) ---
    % Right-skew:
    % noise = gamrnd(skew_deg, 1, n_samples, 1);
    % noise = noise - mean(noise);
    % Left-skew:
    noise = -gamrnd(skew_deg, 1, n_samples, 1);
    noise = noise - mean(noise);
    % =====
    % APPLY TREND TO NOISE LEVEL
    % =====
    % Noise increases along signal according to trend
    % Left: 0% noise → Right: noise_level %
    scaled_noise = trend .* (noise_level * signal .* noise);
    % Final signal
    signal_final(:,i) = signal + scaled_noise;
end
% =====
% SAVE TO FILE
% =====
filename = strcat("data_", signal_name);
writematrix(signal_final, filename);
disp(strcat(filename, " | ", num2str(n_signals), " signals prepared"))
%% =====
% PLOT MEAN WITH CONFIDENCE INTERVAL
% READ SIGNALS
% =====
signal_name = 'data_aver_time_class2.csv';
signals = readmatrix(signal_name); % size: n_samples x n_signals
[n_samples, n_signals] = size(signals);
% Time axis
u = linspace(0, 24, n_samples); % 720 samples over 24 hours
% =====
% COMPUTE MEAN AND CONFIDENCE INTERVAL
% =====
mean_signal = mean(signals, 2); % mean over signals
std_signal = std(signals, 0, 2); % standard deviation over signals
% 95% Confidence interval: mean ± 1.96*std/sqrt(N)
ci_upper = mean_signal + 1.96 * std_signal / sqrt(n_signals);
ci_lower = mean_signal - 1.96 * std_signal / sqrt(n_signals);

```

```

% =====
% PLOT MEAN WITH CONFIDENCE INTERVAL
% =====
figure('Color','w'); hold on
% Shaded confidence interval
fill([u, fliplr(u)], [ci_upper', fliplr(ci_lower')], ...
     [0.9 0.9 1], 'EdgeColor','none'); % light blue
% Plot mean
plot(u, mean_signal, 'b', 'LineWidth', 2)
grid on, %zoom xon
xlabel('u, hour')
ylabel('B_k(u), kW^2')
title('Mean Signal with 95% Confidence Interval')
legend('95% CI','Mean', 'Location','best')
hold off
=====
% Feature Extraction from PKVP Dataset for Power Consumption Mode Classification
=====
clear; close all; clc;
COMP_HEADERS = {'exp_a','exp_b','exp_rsquare','exp_rmse','tangent_45', ...
               'std_before_45','std_after_45','skew_after_45','kurt_after_45'};
TIME_HEADERS = {'time_mean','time_std','time_skew','time_kurt'};
=====
% BLOCK 1. Component features — exponential approximation  $y = a \cdot \exp(b \cdot x)$ 
=====
% Nonlinear least squares fit options
fit_opts = fitoptions('Method', 'NonlinearLeastSquares', ...
                    'Algorithm', 'Levenberg-Marquardt', ...
                    'MaxIter', 1000, 'TolFun', 1e-8, 'ToI', 1e-8);
ft = fitype('a*exp(b*x)', 'independent', 'x', 'coefficients', {'a','b'});
for cls = 1:2
    filename = sprintf('data_aver_comp_class%d.csv', cls);
    data = readmatrix(filename); % 720 x n_signals
    [n_samples, n_signals] = size(data);
    x = linspace(0, 1, n_samples); % normalized x-axis [0,1]
    feature_val = zeros(n_signals, 9);
    for i = 1:n_signals
        y = data(:, i);
        if any(isnan(y)) || any(isinf(y))
            feature_val(i, :) = NaN;
            continue
        end
        % Fit  $y = a \cdot \exp(b \cdot x)$ 
        try
            [exp_fit, gof] = fit(x, y, ft, fit_opts);
            a = exp_fit.a;
            b = exp_fit.b;
            r2 = gof.rsquare;
            rmse = gof.rmse;
        catch
            a = 0; b = 0; r2 = 0; rmse = 0;
        end
        % Tangent-45 point:  $d/dx[a \cdot \exp(b \cdot x)] = 1 \Rightarrow x^* = \ln(1/(a \cdot b)) / b$ 
        % Splits the signal into "rising" and "flat/noisy" regions
        if a * b > 0
            x_45 = (1/b) * log(1 / (a * b));

```

```

        idx_45 = round(x_45 * (n_samples - 1)) + 1;
        idx_45 = max(1, min(n_samples, idx_45));
    else
        idx_45 = round(n_samples / 2);    % fallback to midpoint
    end
    % Detrend: subtract exponential fit, analyse residuals
    y_res = y - a * exp(b * x);
    y_before = y_res(1 : idx_45);    % residuals before tangent point
    y_after = y_res(idx_45 : end);    % residuals after tangent point
    feature_val(i, :) = [a, b, r2, rmse, idx_45, ...
                        std(y_before), std(y_after), ...
                        skewness(y_after), kurtosis(y_after) - 3]; % excess kurtosis
end
out = sprintf('features_data_aver_comp_class%d.csv', cls);
writetable(array2table(feature_val, 'VariableNames', COMP_HEADERS), out)
fprintf('Component features Class %d saved: %s\n', cls, out)
end
=====
% BLOCK 2. Time features — averaged daily load profile
=====
for cls = 1:2
    filename = sprintf('data_aver_time_class%d.csv', cls);
    data = readmatrix(filename); % 24 x n_signals
    [n, n_signals] = size(data);
    feature_val = zeros(n_signals, 4);
    for i = 1:n_signals
        y = data(:, i);
        if any(isnan(y)) || any(isinf(y))
            feature_val(i, :) = NaN;
            continue
        end
        % Statistical moments of the daily profile shape
        feature_val(i, :) = [mean(y), std(y), skewness(y), kurtosis(y) - 3];
    end
    out = sprintf('features_data_aver_time_class%d.csv', cls);
    writetable(array2table(feature_val, 'VariableNames', TIME_HEADERS), out)
    fprintf('Time features Class %d saved: %s\n', cls, out)
end
=====
% BLOCK 3. Combined dataset (13 features + class label)
=====
X = []; y_labels = [];
for cls = 1:2
    comp = readmatrix(sprintf('features_data_aver_comp_class%d.csv', cls));
    time = readmatrix(sprintf('features_data_aver_time_class%d.csv', cls));
    X = [X;    comp, time]; %#ok<AGROW>
    y_labels = [y_labels; repmat(cls - 1, size(comp, 1), 1)]; %#ok<AGROW>
end
all_headers = [COMP_HEADERS, TIME_HEADERS, {'label'}];
writetable(array2table([X, y_labels], 'VariableNames', all_headers), 'pkvp_dataset.csv')
fprintf('\nCombined dataset saved: pkvp_dataset.csv (%d samples, %d features)\n', ...
        size(X, 1), size(X, 2))

```