

УДК 004.08:519.2

<https://doi.org/10.31713/vt4202514>

Волощук А. В. [1; ORCID ID: 0009-0007-1478-1601],
аспірант,
Осухівська Г. М. [1; ORCID ID: 0000-0003-0132-1378],
к.т.н., доцент,
Паляниця Ю. Б. [1; ORCID ID:],
к.т.н., доцент,
Луцків А. М. [1; ORCID ID:],
к.т.н., доцент,
Валяшек В. Б. [1; ORCID ID:],
к.т.н., доцент

¹Тернопільський національний технічний університет імені І. Пулюя, м. Тернопіль

ВИЗНАЧЕННЯ ЗБАЛАНСОВАНOSTІ ЕЛЕКТРОЕНЕРГЕТИЧНИХ СИСТЕМ МЕТОДАМИ МАШИННОГО НАВЧАННЯ НА ОСНОВІ ІНФОРМАТИВНИХ ОЗНАК ОТРИМАНИХ КОМПОНЕНТНИМ МЕТОДОМ

У статті запропоновано підхід до визначення збалансованості електроенергетичної системи на основі поєднання компонентного методу аналізу періодично корельованих випадкових процесів та методів машинного навчання. Процедура формування інформативного ознакового простору базується на експоненціальній апроксимації компонентного усереднення та статистичних характеристиках сигналу.

Сформовано датасет із 365 добових реалізацій, на якому проведено порівняльне дослідження 14 класифікаторів. Результати статистичного аналізу підтвердили значущість ключових ознак, а найкращі моделі (KNN, Extra Trees, SVM) досягли точності 98,9% та F1-міри 0,9877.

Показано, що перехід до ознакового представлення сигналу забезпечує високу точність класифікації, інтерпретованість результатів у термінах фізичних характеристик електроспоживання та зменшує розмірність вихідних даних, що знижує обчислювальні витрати й підвищує ефективність алгоритмів машинного навчання. Запропонований підхід може бути використаний для визначення стану збалансованості розподілених електроенергетичних систем та ідентифікації незбалансованих режимів їх функціонування.

Ключові слова: електроспоживання; електроенергетична система; випадковий процес; математична модель; ПКВП; компонентний метод; машинне навчання.

Вступ. Сучасна енергетична інфраструктура еволюціонувала у складні багаторівневі системи з високим рівнем інтеграції



розподілених джерел генерації та активних споживачів, що суттєво ускладнює процеси управління та моніторингу режимів функціонування електроенергосистем [1].

Розподілені електроенергетичні системи характеризуються децентралізованою структурою генерації та споживання, значною варіативністю навантаження та наявністю стохастичних збурень, що призводить до підвищеної нестабільності режимів і ускладнює забезпечення їх збалансованості. У таких умовах особливої актуальності набуває задача своєчасного виявлення незбалансованих станів, оскільки навіть короточасні відхилення режимів можуть призводити до перевантаження елементів системи, зниження якості електроенергії та порушення нормального функціонування електроенергетичної мережі [2].

Тому актуальним завданням є виокремлення таких інформативних ознак сигналу електроспоживання, які б дозволили аналізувати збалансованість електроенергетичних систем. Підходи до аналізу та моделювання електроспоживання в задачах забезпечення збалансованості електроенергетичних систем охоплюють широкий спектр методів. Проте складність аналізу профілів електроспоживання зумовлена впливом різних факторів, які спричиняють випадкові флуктуації, але поряд з цим відмічається і повторюваність структури сигналу [3; 4].

Математичний апарат періодично корельованих випадкових процесів (ПКВП), зокрема компонентний метод аналізу дозволяє здійснити формалізацію структури сигналів шляхом урахування періодичної нестационарності його ймовірнісних характеристик [5; 6]. Отримання інформативних ознак із сформованих компонентних матриць для подальшого їх використання методами машинного навчання дає змогу визначити стан збалансованості системи.

Метою роботи є визначення стану збалансованості розподілених електроенергетичних систем на основі виділених інформативних ознак про електроспоживання компонентним методом аналізу періодично корельованих випадкових процесів, використовуючи методи машинного навчання.

Аналіз останніх досліджень і публікацій. Сучасні дослідження у сфері електроенергетичних систем демонструють ефективність застосування методів машинного навчання для прогнозування стабільності режимів функціонування [6]. У таких роботах задачі аналізу стану системи формулюються як задачі класифікації на основі параметрів навантаження та інших характеристик [7].

Альтернативним підходом є використання методів кластеризації, що дозволяють виділяти типові режими функціонування без попереднього задання класів [8]. Подальший розвиток цих підходів пов'язаний із застосуванням ансамблевих і гібридних моделей, здатних враховувати складну нелінійну структуру даних та підвищувати точність прогнозування [9; 10].

Водночас більшість існуючих підходів орієнтована на аналіз параметрів електроенергетичної системи в цілому та не враховує внутрішню структуру процесу електроспоживання як окремого об'єкта дослідження. Це обумовлює доцільність розробки підходів, спрямованих на аналіз структури електроспоживання та формування інформативних ознак на її основі.

При цьому ключовим фактором, що визначає стан системи є характер зміни електроспоживання, зокрема його варіативність та наявність відхилень від типових профілів. У зв'язку з цим аналіз структури електроспоживання та виділення інформативних ознак на його основі є релевантним підходом до ідентифікації збалансованих і незбалансованих режимів.

Дослідження когерентної оцінки коваріації ПКВП підтвердили обчислювальну ефективність для аналізу сигналів із періодичними структурами [11]. Впровадження ПКВП для моделювання добових патернів трафіку комп'ютерних мереж [12] та електроспоживання організацій [4; 6] продемонструвало практичну універсальність методології та її придатність для визначення збалансованості електроенергетичних систем.

Виклад основної частини. Для визначення стану збалансованості розподілених електроенергетичних систем здійснюється класифікація методами машинного навчання з використанням інформативних ознак, виділених компонентним методом ПКВП із даних електроспоживання. Тобто для формування датасету ознак використано результати компонентного методу аналізу сигналів електроспоживання в рамках моделі ПКВП.

Такий підхід забезпечує перехід від сирих часових реалізацій до компактного стохастичного представлення процесу. Для кожної добової реалізації електроспоживання формуються дві ключові структури: матриця компонентного усереднення $\underline{V}(t)$ та вектор часового усереднення \underline{V}_k , які використовуються як база для подальшого виділення інформативних ознак. Кожна доба розглядається як окреме спостереження, загальна тривалість



спостереження 365 діб, 206 відносяться до збалансованого режиму (Class 1) та 159 до незбалансованого (Class 2).

Збалансований режим електроспоживання (Class 1) характеризується узгодженою зміною навантаження споживача, за якої сумарне електроспоживання формується плавно та передбачувано у часі без різких пікових відхилень. Такий режим відповідає типовому добовому профілю, коли внесок окремих електроприймачів є узгодженим, а їх робота не призводить до суттєвих стрибків потужності, забезпечуючи рівномірний розподіл навантаження.

Незбалансований режим електроспоживання (Class 2), навпаки, пов'язаний з неузгодженою динамікою роботи електроприймачів і проявляється у вигляді різких змін навантаження та відхилень від типового профілю. Це зумовлює підвищену варіативність процесу споживання, формування пікових навантажень і нерівномірний розподіл навантаження в часі, що може призводити до відхилення параметрів режиму від допустимих значень.

Розподіл класів у вибірці є відносно збалансованим (206 спостережень для Class 1 та 159 для Class 2), що дозволяє уникнути суттєвого впливу дисбалансу на результати класифікації. На рис. 1 показано відмінності у профілях компонентного усереднення для збалансованого (C1) та незбалансованого (C2) режимів.

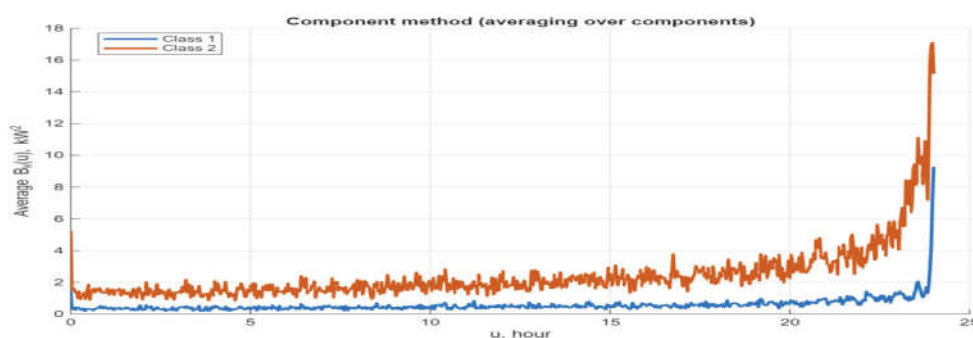


Рис. 1. Профілі компонентного усереднення для збалансованого (C1) та незбалансованого (C2) режимів

Як бачимо з рис. 1, незбалансований режим (C2) характеризується систематично вищими значеннями компонент сигналу та більш вираженим зростанням у кінці добового інтервалу. Натомість збалансований режим (C1) демонструє відносно стабільну динаміку з меншою амплітудою коливань.

Це підтверджує наявність структурних відмінностей між класами та обґрунтовує доцільність подальшого виділення інформативних ознак.

Для подальшого формування інформативних ознак виконується апроксимація матриці компонентного усереднення $\underline{B}(t)$ експоненціальною функцією вигляду:

$$y(t) = a \cdot e^{bt}, \quad (1)$$

де a – коефіцієнт початкового рівня, b – параметр швидкості зростання.

Таким чином, для кожної доби сформовано вектор із 12 інформативних ознак, зокрема: коефіцієнт амплітуди експоненціальної моделі (exp_a), швидкість експоненціального зростання або спадання (exp_b), коефіцієнт детермінації R^2 ($exp_rsquare$), середньоквадратична похибка (exp_rmse), індекс точки з похідною 1 ($tangent_45$), стандартне відхилення до точки (std_before_45), стандартне відхилення після точки (std_after_45), коефіцієнт асиметрії після точки ($skew_after_45$), коефіцієнт ексцесу після точки ($kurt_after_45$), середнє значення сигналу ($time_mean$), стандартне відхилення сигналу ($time_std$), коефіцієнт асиметрії сигналу ($time_skew$). З яких 9 ознак для «усереднення по компонентах» та ще 3 для «усереднення по часу».

Для кількісного аналізу динаміки електроспоживання було введено параметр x_{45} – точку переходу до агресивного зростання. Математично вона визначається як момент, у якому швидкість зміни сигналу стає рівною швидкості плину часу $f'(x)=1$, геометрично – це відповідає нахилу дотичної до тренду під кутом 45 градусів. Вибір цієї точки дозволяє об'єктивно розділити добовий цикл на два статистично різні періоди: зона стабільності (до x_{45}) – приріст енергії менший за одиницю, що відповідає штатному режиму роботи та зона ризику (після x_{45}) – швидкість зростання перевищує одиницю, що вказує на формування енергетичного зростання.

Перед опрацюванням даних класифікаторами проведено статистичний аналіз сформованих ознак з метою оцінки їх дискримінативної здатності. Попередній візуальний аналіз розподілів ознак для двох класів (рис. 2) свідчить про наявність суттєвих відмінностей для окремих дескрипторів.

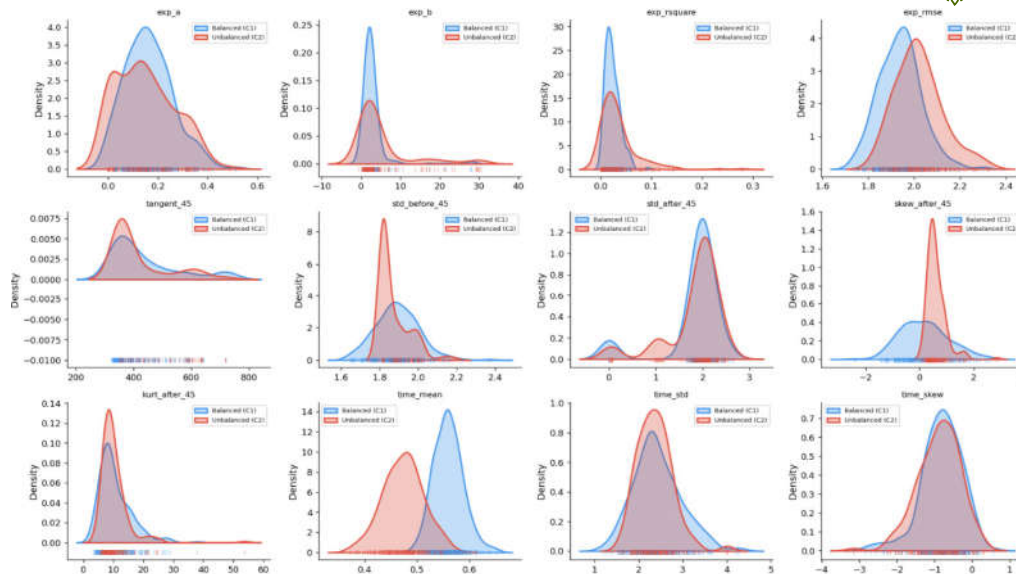


Рис. 2. Розподіли ознак за класами

Як видно з рис. 2, ознаки `exp_rmse`, `skew_after_45` та `time_mean` демонструють чітке розділення між класами, тоді як для інших ознак спостерігається значне перекриття розподілів. Це вказує на різний рівень інформативності окремих дескрипторів.

Для кількісної оцінки статистичної значущості відмінностей між класами застосовано непараметричний критерій Манна–Уїтні, що не потребує припущення про нормальність розподілу ознак.

Таблиця 1

Результати тесту Манна–Уїтні для інформативних ознак

Ознака	U-стат.	p-значення	Значущість
<code>exp_a</code>	18056,0	$9,31 \times 10^{-2}$	ні
<code>exp_b</code>	14941,0	$1,51 \times 10^{-1}$	ні
<code>exp_rsquare</code>	14780,0	$1,10 \times 10^{-1}$	ні
<code>exp_rmse</code>	9001,0	$1,59 \times 10^{-13}$	ТАК
<code>tangent_45</code>	17234,5	$3,86 \times 10^{-1}$	ні
<code>std_before_45</code>	19165,0	$5,29 \times 10^{-3}$	ТАК
<code>std_after_45</code>	15488,0	$3,74 \times 10^{-1}$	ні
<code>skew_after_45</code>	10189,0	$5,96 \times 10^{-10}$	ТАК
<code>kurt_after_45</code>	16294,0	$9,34 \times 10^{-1}$	ні
<code>time_mean</code>	31639,0	$1,23 \times 10^{-52}$	ТАК
<code>time_std</code>	17244,0	$3,86 \times 10^{-1}$	ні
<code>time_skew</code>	17765,0	$1,65 \times 10^{-1}$	ні

Найкращі результати показали, що найбільш статистично значущими є ознаки $time_mean$ ($p < 10^{-52}$), exp_rmse ($p < 10^{-13}$), $skew_after_45$ ($p < 10^{-9}$) та std_before_45 ($p < 5.3 \times 10^{-3}$). Ці ознаки забезпечують основний внесок у розділення класів.

Інші ознаки не досягають рівня статистичної значущості $\alpha = 0,05$, однак не виключаються з подальшого аналізу, оскільки ансамблеві методи машинного навчання здатні ефективно використовувати їх у комбінації з іншими предикторами.

Діаграми розмаху (рис. 3) дозволяють оцінити медіани, міжквартильні інтервали та наявність викидів для кожної ознаки.

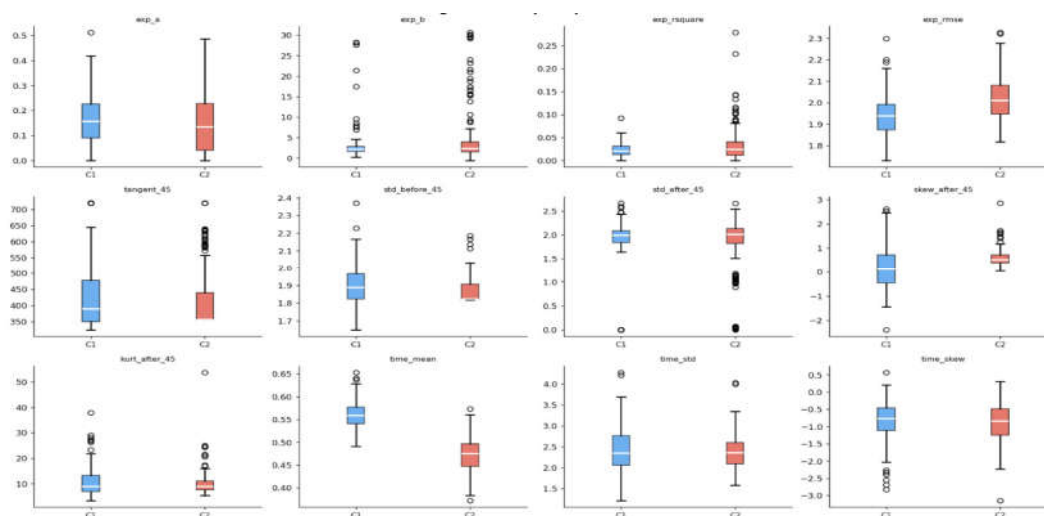


Рис. 3. Діаграми розмаху ознак для двох класів

Для класифікації стану електроенергетичних систем на («збалансований» чи «незбалансований») було протестовано 14 алгоритмів машинного навчання, які представляють різні підходи до розв'язання задачі.

Датасет було розділено на навчальну (273 зразки, 75%) та тестову (92 зразки, 25%) вибірки зі стратифікацією за класами для збереження пропорційного представництва обох режимів. Перед подачею на моделі, чутливі до масштабу даних (SVM, KNN, MLP), усі ознаки було стандартизовано за допомогою z-нормалізації. Для надійної оцінки узагальнюючої здатності моделей застосовано 5-кратну стратифіковану крос-валідацію на навчальній вибірці.

До базових моделей належать лінійні та ймовірнісні алгоритми: Logistic Regression, Linear Discriminant Analysis (LDA) та Naive Bayes.

Метричний підхід представлений алгоритмом k-Nearest Neighbors (KNN, $k=5$). Ієрархічні правила реалізовані за допомогою Decision Tree з обмеженою глибиною (максимум 6 рівнів).



Також були протестовані різні ансамблеві методи. Методи беггінгу включають Random Forest та Extra Trees (по 300 дерев у кожній моделі). Методи бустингу представлені алгоритмами AdaBoost (200 класифікаторів), Gradient Boosting, XGBoost та LightGBM.

Для моделювання складних нелінійних залежностей використані Support Vector Machine з RBF-ядром ($C=10$) та багат шаровий перцептрон (MLP) з двома прихованими шарами (128 та 64 нейрони).

Результати оцінки ефективності класифікаторів на тестовій вибірці наведено у табл. 2, відсортованій за значенням F1-score. Для всебічної оцінки використано комплекс метрик: Accuracy, Precision, Recall, F1-score (гармонійне середнє Precision та Recall), MCC, Карра, CV-F1 та AUC.

Таблиця 2

Порівняльна оцінка ефективності класифікаторів

Класифікатор	Acc	Prec	Rec	F1	MCC	Карра	CV-F1	AUC
KNN (k=5)	0,9891	0,9756	1,0000	0,9877	0,9782	0,9779	0,9344	0,9998
Extra Trees	0,9891	0,9756	1,0000	0,9877	0,9782	0,9779	0,9826	0,9998
SVM (RBF)	0,9891	0,9756	1,0000	0,9877	0,9782	0,9779	0,9524	0,9995
MLP	0,9891	1,0000	0,9750	0,9873	0,9781	0,9778	0,9477	0,9990
XGBoost	0,9783	0,9750	0,9750	0,9750	0,9558	0,9558	0,9574	0,9952
Random Forest	0,9783	0,9750	0,9750	0,9750	0,9558	0,9558	0,9659	0,9995
Voting Ens.	0,9783	0,9750	0,9750	0,9750	0,9558	0,9558	0,9626	0,9990
AdaBoost	0,9674	0,9744	0,9500	0,9620	0,9337	0,9335	0,9745	0,9990
LightGBM	0,9674	0,9744	0,9500	0,9620	0,9337	0,9335	0,9743	0,9986
Logistic Reg.	0,9674	0,9744	0,9500	0,9620	0,9337	0,9335	0,9249	0,9962
Naive Bayes	0,9674	1,0000	0,9250	0,9610	0,9352	0,9331	0,8506	1,0000
Decision Tree	0,9565	0,9737	0,9250	0,9487	0,9119	0,9110	0,9174	0,9529
Grad. Boost.	0,9565	0,9737	0,9250	0,9487	0,9119	0,9110	0,9074	0,9526
LDA	0,9457	0,9730	0,9000	0,9351	0,8905	0,8885	0,9191	0,9971

Найкращі результати показали KNN (k=5), Extra Trees та SVM (RBF), досягнувши точності 98,91% і F1-score 0,9877. Високі значення AUC (більшість моделей перевищують 0,99) підтверджують відмінну

здатність моделей розрізнати два стани збалансованості електроенергетичної системи.

Для оцінки внеску кожної ознаки в класифікацію було застосовано два методи в рамках моделі Random Forest: Mean Decrease Impurity (MDI) та Permutation Importance.

Метод MDI показує, наскільки сильно ознака зменшує неоднорідність у вузлах дерев, тоді як Permutation Importance вимірює падіння точності після випадкового перемішування значень ознаки. Другий підхід вважається більш надійним, оскільки менш залежить від кореляцій між ознаками.

Обидва методи чітко виділяють `time_mean` як домінуючу ознаку її важливість значно перевищує середній рівень (рис. 4). Це узгоджується з результатами тесту Манна-Уїтні і має просте фізичне пояснення: незбалансований стан електроенергетичної системи характеризується систематично вищим середнім рівнем електроспоживання.

Наступні за значимістю ознаки `skew_after_45`, `tangent_45` та `std_before_45` відображають як загальну динаміку сигналу, так і його поведінку після критичної точки зростання.

Натомість ознаки `time_skew`, `time_kurt` та `kurt_after_45` демонструють мінімальний внесок і можуть бути виключені при спрощенні моделі. Великі довірчі інтервали у Permutation Importance для `exp_rsquare` та `exp_a` вказують на їхню нестабільну і другорядну роль.

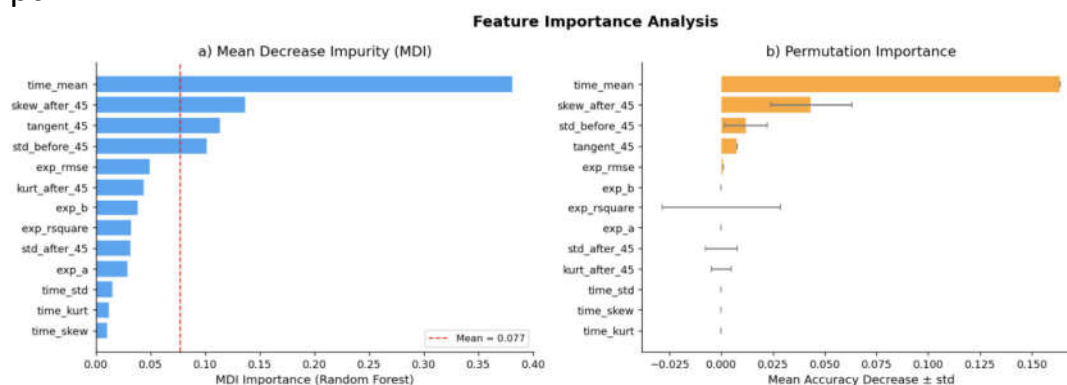


Рис. 4. Аналіз важливості ознак у моделі Random Forest:
(a) Mean Decrease Impurity (MDI);
(b) Permutation Importance

Результати дослідження свідчать про високу ефективність підходу, що поєднує компонентний метод аналізу періодично корельованих випадкових процесів із подальшим виділенням ознак та застосуванням алгоритмів машинного навчання.



Порівняльний аналіз моделей показав, що найкращі результати забезпечують ансамблеві методи (Random Forest, Extra Trees) та метод опорних векторів (SVM). Їхня ефективність пояснюється здатністю враховувати складні нелінійні взаємозв'язки між ознаками. Водночас бустинг-алгоритми не продемонстрували суттєвої переваги над методами беггінгу, що, ймовірно, пов'язано з відносно невеликим обсягом вибірки та високою розділеністю класів.

Навіть прості лінійні моделі (Logistic Regression та Naive Bayes) показали достатньо високий рівень точності (до 96,7%), що свідчить про спрощення структури задачі після застосування компонентного методу ПКВП. У результаті задача набуває властивостей, близьких до лінійно роздільної, що має практичне значення з точки зору обчислювальної ефективності та інтерпретованості моделей.

Серед обмежень підходу слід відзначити залежність якості ознак від параметрів експоненціальної апроксимації, наявність незначного дисбалансу класів, а також необхідність додаткової валідації на інших об'єктах і більших вибірках. Перспективним напрямком подальших досліджень є застосування методів відбору та зменшення розмірності ознак для підвищення ефективності та інтерпретованості моделей.

Висновки. У роботі запропоновано підхід до класифікації режимів електроспоживання, що поєднує компонентний метод аналізу періодично корельованих випадкових процесів із методами машинного навчання. На основі компонентного представлення сигналу сформовано ознаковий простір, який дозволяє ефективно описувати як глобальну динаміку, так і стохастичні особливості електроспоживання.

Однією з ключових переваг запропонованого підходу є фізична інтерпретованість сформованого ознакового простору. На відміну від моделей, що працюють безпосередньо з вихідними часовими рядами, кожна із 12 ознак має чітке математичне підґрунтя та відповідає конкретним характеристикам процесу електроспоживання. Це дозволяє не лише виконувати точну класифікацію стану збалансованості електроенергетичних систем, а й інтерпретувати результати в термінах, зрозумілих фахівцям енергетичної галузі.

Проведений статистичний аналіз підтвердив значимість ключових ознак, зокрема домінуючу роль показника середнього рівня сигналу (time_mean), що узгоджується з результатами аналізу

важливості ознак. Експериментальні результати показали, що запропонований підхід забезпечує високу точність класифікації.

Отримані результати свідчать про те, що використання компонентного методу ПКВП суттєво спрощує структуру задачі класифікації та дозволяє застосовувати як складні, так і відносно прості моделі без істотної втрати точності. Таким чином, зменшення розмірності вихідних даних за рахунок переходу до ознакового простору дозволяє скоротити обчислювальні витрати та підвищити ефективність застосування алгоритмів машинного навчання. Практична значущість роботи полягає у можливості побудови автоматизованих систем моніторингу режимів електроспоживання, орієнтованих на застосування в інтелектуальних енергетичних мережах.

Перспективи подальших досліджень пов'язані з розширенням обсягів даних, валідацією підходу на різних типах об'єктів та застосуванням методів зменшення розмірності ознак для підвищення ефективності моделей.

1. Fang X., Misra S., Xue G., Yang D. Smart Grid – The New and Improved Power Grid: A Survey. *IEEE Communications Surveys & Tutorials*. 2012. Vol. 14, № 4. P. 944–980. doi:10.1109/SURV.2011.101911.00087.
2. Kuster C., Rezgui Y., Mourshed M. Electrical load forecasting models: A critical systematic review. *Sustainable Cities and Society*. 2017. № 35. P. 257–270.
3. Nowotarski J., Weron R. Extrapolating the long-term seasonal component of electricity prices for forecasting in the day-ahead market. *Journal of Commodity Markets*. 2025. Vol. 37. P. 100449. doi:10.1016/j.jcomm.2024.100449.
4. Voloshchuk A., Osukhivska H., Khvostivskiy M., Sverstiuk A. Application of periodically correlated stochastic processes for forecasting electricity consumption. *Measuring and Computing Devices in Technological Processes*. 2025. № 3. P. 393–403.
5. Javorskyj I., Isayev I., Majewski J., Yuzefovych R. Component covariance analysis for periodically correlated random processes. *Signal Processing*. 2010. № 90(4). P. 1083–1102.
6. Voloshchuk A., Osukhivska H., Khvostivskiy M., Sverstiuk A. Analysis of electricity consumption using the component method of periodically correlated random processes. *Computer Systems and Information Technologies*. 2025. № 3. P. 74–82.
7. Önder M., Dogan M. U., Polat K. Classification of smart grid stability prediction using cascade machine learning methods and the internet of things in smart grid. *Neural Computing and Applications*. 2023. № 35(24). P. 17851–17869.
8. Senturk U., Eyecioglu O. Clustering analysis for stability prediction in smart grid systems. *13th International Conference on Smart Grid (icSmartGrid)*. 2025. C. 743–747.
9. Satu Md. S., Riyad Md. M. H., Khan Md. I. Machine learning approaches to predict the stability of smart grid. *Franklin Open*. 2026. № 15.
10. Redwan K., Datto S., Ahmed M., Hannan N., Shufian A., Mahmood M. S. Dynamic and transformative hybrid machine learning strategies for effective stability prediction in smart energy networks. *27th International Conference on Computer and Information Technology (ICCIT)*. 2024. C. 3602–3607.
11. Javors'kyj I., Yuzefovych R., Kravets I., Matsko I. Methods of periodically correlated random processes and their generalizations. *Cyclostationarity*:



Theory and Methods. 2014. P. 73–93. **12.** Khvostivska L., Khvostivskiy M., Martsenyuk V. Statistical analysis of network traffic in computer networks based on periodically correlated random processes model. *International Journal of Computer Science and Network Security*. 2021. Vol. 21, № 4. P. 147–154.

REFERENCES:

- 1.** Fang X., Misra S., Xue G., Yang D. Smart Grid – The New and Improved Power Grid: A Survey. *IEEE Communications Surveys & Tutorials*. 2012. Vol. 14, № 4. P. 944–980. doi:10.1109/SURV.2011.101911.00087.
 - 2.** Kuster C., Rezgui Y., Mourshed M. Electrical load forecasting models: A critical systematic review. *Sustainable Cities and Society*. 2017. № 35. P. 257–270.
 - 3.** Nowotarski J., Weron R. Extrapolating the long-term seasonal component of electricity prices for forecasting in the day-ahead market. *Journal of Commodity Markets*. 2025. Vol. 37. P. 100449. doi:10.1016/j.jcomm.2024.100449.
 - 4.** Voloshchuk A., Osukhivska H., Khvostivskiy M., Sverstiuk A. Application of periodically correlated stochastic processes for forecasting electricity consumption. *Measuring and Computing Devices in Technological Processes*. 2025. № 3. P. 393–403.
 - 5.** Javorskyj I., Isayev I., Majewski J., Yuzefovych R. Component covariance analysis for periodically correlated random processes. *Signal Processing*. 2010. № 90(4). P. 1083–1102.
 - 6.** Voloshchuk A., Osukhivska H., Khvostivskiy M., Sverstiuk A. Analysis of electricity consumption using the component method of periodically correlated random processes. *Computer Systems and Information Technologies*. 2025. № 3. P. 74–82.
 - 7.** Önder M., Dogan M. U., Polat K. Classification of smart grid stability prediction using cascade machine learning methods and the internet of things in smart grid. *Neural Computing and Applications*. 2023. № 35(24). P. 17851–17869.
 - 8.** Senturk U., Eyecioglu O. Clustering analysis for stability prediction in smart grid systems. *13th International Conference on Smart Grid (icSmartGrid)*. 2025. C. 743–747.
 - 9.** Satu Md. S., Riyad Md. M. H., Khan Md. I. Machine learning approaches to predict the stability of smart grid. *Franklin Open*. 2026. № 15.
 - 10.** Redwan K., Datto S., Ahmed M., Hannan N., Shufian A., Mahmood M. S. Dynamic and transformative hybrid machine learning strategies for effective stability prediction in smart energy networks. *27th International Conference on Computer and Information Technology (ICCIT)*. 2024. C. 3602–3607.
 - 11.** Javors'kyj I., Yuzefovych R., Kravets I., Matsko I. Methods of periodically correlated random processes and their generalizations. *Cyclostationarity: Theory and Methods*. 2014. P. 73–93.
 - 12.** Khvostivska L., Khvostivskiy M., Martsenyuk V. Statistical analysis of network traffic in computer networks based on periodically correlated random processes model. *International Journal of Computer Science and Network Security*. 2021. Vol. 21, № 4. P. 147–154.
-

Voloshchuk A. V. [1; ORCID ID: 0009-0007-1478-1601],

Post-graduate Student

Osukhivska H. M. [1; ORCID ID: 0000-0003-0132-1378],

Candidate of Engineering (Ph.D.), Associate Professor,

Palianytsia Y. B. [1; ORCID ID:],

Candidate of Engineering (Ph.D.), Associate Professor,

Lutskiv A. M. [1; ORCID ID:],

Candidate of Engineering (Ph.D.), Associate Professor,

Valiashek V. B. [1; ORCID ID:],

Candidate of Engineering (Ph.D.), Associate Professor

¹*Ternopil Ivan Puluj National Technical University, Ternopil*

DETERMINATION OF ELECTRIC POWER SYSTEM BALANCE USING MACHINE LEARNING METHODS BASED ON INFORMATIVE FEATURES OBTAINED BY THE COMPONENT METHOD

This paper proposes an approach for determining the balance state of distributed electric power systems based on the combination of the component method for analyzing periodically correlated random processes (PCRP) and machine learning algorithms.

The relevance of this research is driven by the increasing complexity of modern distributed power systems, characterized by decentralized generation, high load variability, and stochastic disturbances, which necessitate timely detection of unbalanced operating states to prevent equipment overload and power quality degradation. The feature extraction procedure is based on the exponential approximation of component-averaged signals and statistical characteristics of the time series. As a result, a set of 12 informative features is formed, including parameters of the exponential model (exp_a , exp_b , exp_rsquare , exp_rmse), characteristics related to the critical growth point x_{45} (tangent_{45} , std_before_{45} , std_after_{45} , skew_after_{45} , kurt_after_{45}), and statistical descriptors of the signal (time_mean , time_std , time_skew). The x_{45} parameter is introduced as the transition point to aggressive growth, mathematically defined as the moment when the signal rate of change equals unity, which enables an objective partitioning of the daily cycle into stability and risk zones. A dataset of 365 daily realizations was constructed, comprising 206 samples of the balanced mode (Class 1) and 159 samples of the unbalanced mode (Class 2). The dataset was partitioned into training (75%) and testing (25%) subsets with class stratification.

The discriminative power of the features was evaluated using the nonparametric Mann–Whitney U test, which identified the most statistically significant features: time_mean ($p < 10^{-52}$), exp_rmse ($p < 10^{-13}$),



skew_after_45 ($p < 10^{-9}$), and std_before_45 ($p < 5.3 \times 10^{-3}$). A comparative analysis of 14 machine learning classifiers representing diverse paradigms was conducted, including linear and probabilistic methods (Logistic Regression, LDA, Naive Bayes), metric-based approaches (KNN), tree-based models (Decision Tree, Random Forest, Extra Trees), boosting algorithms (AdaBoost, Gradient Boosting, XGBoost, LightGBM), support vector machines with RBF kernel, and neural networks. The best-performing models (KNN with $k=5$, Extra Trees, and SVM with RBF kernel) achieved an accuracy of 98.91% and an F1-score of 0.9877 with perfect recall.

High AUC values (exceeding 0.99 for most models) confirm the strong ability of the models to distinguish between balanced and unbalanced operating states of the electric power system. Feature importance analysis using both Mean Decrease Impurity and Permutation Importance methods confirmed time_mean as the dominant discriminative feature. It is shown that the transition from raw time series to feature-based representation ensures high classification accuracy, interpretability of results in terms of physical characteristics of electricity consumption, and reduction of data dimensionality, which decreases computational costs and improves the efficiency of machine learning algorithms. The proposed approach can be effectively used for determining the balance state of distributed electric power systems and identifying unbalanced operating modes.

Keywords: electricity consumption; electric power system; random process; mathematical model; periodically correlated random processes; component method; machine learning.

Отримано: 08 вересня 2025 року
Прорецензовано: 10 вересня 2025 року
Прийнято до друку: 18 грудня 2025 року