



DOI 10.28925/2663-4023.2025.28.844

УДК 004.75

Ніщененко Дмитро Олександрович

аспірант

Державний університет інформаційно-комунікаційних технологій, Київ, Україна

ORCID ID: 0009-0006-1781-4109

dima.nishchemenko@gmail.com**Волошук Олена Борисівна**

кандидат технічних наук, доцент

Харківський національний університет радіоелектроніки, Харків, Україна

ORCID ID: 0000-0002-5912-4126

Olena.voloshchuk@nure.ua

АДАПТИВНЕ ОЧИЩЕННЯ РІЗНОРІДНИХ СЕНСОРНИХ ДАНИХ У СИСТЕМАХ РОЗУМНОГО БУДИНКУ НА ОСНОВІ КЛАСИФІКАЦІЇ ШУМІВ

Анотація. Сучасні системи розумного будинку генерують значні обсяги даних, якість котрих є критично важливою для ефективного управління, аналізу та прогнозування. Однак необроблені потоки даних часто містять складні комбінації шумів та аномалій, включаючи викиди, дрейф показників та періоди незмінних значень. Такі артефакти суттєво знижують надійність даних, що може призводити до некоректної роботи інтелектуальних систем та прийняття помилкових рішень. Існуючі методи очищення даних часто демонструють обмежену ефективність при роботі з гетерогенними типами шумів у реальному часі, особливо коли йдеться про обробку даних від сенсорів з різними фізичними характеристиками. У статті представлено новий метод ACRA, розроблений для адаптивного очищення різномірних часових рядів сенсорних даних у контексті розумного будинку. Метод ACRA включає класифікатор на основі ансамблю дерев рішень (Random Forest) для ідентифікації специфічних типів шумів у ковзному вікні оброблюваних даних. Результати класифікації доповнюються евристичним правилом, яке використовує аналіз дисперсії вікна для точнішого виявлення періодів з низьким рівнем шуму. На основі комбінації висновку класифікатора та правила, активується модуль адаптивної стратегії, який динамічно обирає найбільш доцільний оператор для корекції поточного значення даних. Для оцінки ефективності запропонованого методу було проведено експериментальне дослідження на реальних часових рядах температури, вологості та загального енергоспоживання, зібраних у житловому приміщенні, до яких було синтетично додано контрольовані типи шумів. Результати експериментів продемонстрували, що метод ACRA забезпечує значно вищу якість очищення порівняно з поширеними базовими методами, такими як ковзна медіана та фільтр Калмана, для всіх трьох досліджуваних типів сенсорних даних. Запропонований метод ACRA є надійним та гнучким інструментом для підвищення якості сенсорних даних у системах розумного будинку, що створює основу для розробки більш точних та ефективних інтелектуальних додатків.

Ключові слова: очищення даних; розумний будинок; Інтернет речей; адаптивне усунення аномалій; класифікація шумів; машинне навчання.

ВСТУП

Сучасні системи «розумного будинку» все глибше інтегруються у повсякденне життя, пропонуючи користувачам підвищений рівень комфорту, безпеки та оптимізації енергоспоживання. Основою функціонування таких систем є безперервні потоки даних, що надходять від численних сенсорів, таких як датчики температури, вологості,



освітленості, руху та лічильники споживання енергоресурсів. Якість цих даних має вирішальне значення для коректної роботи систем автоматизованого управління, інтелектуальної аналітики та розробки прогнозних моделей. Однак, необроблені сенсорні дані, що збираються в реальному часі, часто містять значну кількість різноманітних шумів та аномалій. Ці артефакти можуть бути спричинені несправностями самих сенсорів, електромагнітними перешкодами під час передачі даних, непередбачуваним впливом зовнішнього середовища або іншими стохастичними факторами [1], [2]. Типові проблеми якості даних включають випадкові короточасні викиди, поступовий дрейф показників сенсорів, періоди незмінних значень, а також наявність загального фоновому шуму. Присутність таких пошкоджень у даних може призвести до невірних інтерпретацій стану контрольованого середовища, неефективного управління наявними ресурсами та, як наслідок, до суттєвого зниження загальної ефективності та надійності функціонування систем розумного будинку [3].

Постановка проблеми. У цій частині статті описується проблема, розгляду якої присвячене дослідження, у загальному вигляді та її зв'язок із важливими науковими чи практичними завданнями.

Аналіз останніх досліджень і публікацій. Необхідність забезпечення високої якості сенсорних даних у системах розумного будинку є актуальною науково-практичною задачею. Різноманітність типів сенсорів та відмінності у фізичній природі вимірюваних ними величин, призводять до того, що потоки даних характеризуються різними динамічними властивостями та амплітудними діапазонами. Більшість існуючих методів очищення даних або є занадто загальними, не враховуючи специфіку конкретного типу шуму, або вузькоспеціалізованими. Крім того, багато методів не призначені для роботи в режимі реального часу. Таким чином, існує потреба у розробці універсальних, але водночас адаптивних методів очищення, здатних працювати з різнорідними потоками сенсорних даних в реальному часі, ідентифікувати та усувати різні типи шумів, зберігаючи при цьому корисну інформацію, що міститься в сигналі.

Проблематиці очищення даних часових рядів, зокрема в контексті IoT та сенсорних мереж, присвячено значну кількість наукових праць. Традиційні підходи включають методи на основі фільтрації, такі як ковзне середнє, експоненційне згладжування та медіанна фільтрація [4]. Ці методи прості в реалізації, однак можуть призводити до згладжування важливих деталей сигналу та є малоефективними проти структурованих аномалій. Більш складні модельні підходи, наприклад, фільтр Калмана та його модифікації, демонструють кращі результати за умов відповідності статистичних характеристик сигналу та шуму прийнятим модельним припущенням [5].

Однак, у реальних умовах експлуатації сенсорів ці припущення часто порушуються через наявність змішаних та неГаусових типів шумів. Активно розвиваються методи очищення даних на основі машинного навчання. До них належать підходи, що використовують кластеризацію для виявлення викидів, регресійні моделі для прогнозування та заміни аномальних значень, а також нейронні мережі, зокрема автокодуювальники та рекурентні нейронні мережі, для вивчення нормальної поведінки часового ряду та ідентифікації відхилень [6] – [9].

Деякі дослідження зосереджені на класифікації типів аномалій перед їх усуненням, що дозволяє застосовувати більш специфічні стратегії корекції [10]. Наприклад, у роботі запропоновано систему для промислових часових рядів, що інтегрує декілька стратегій очищення [11]. Проте, більшість таких систем або орієнтовані на пакетну обробку, або не забезпечують достатньої адаптивності до одночасної присутності різних типів шумів у різнорідних потоках даних, характерних для систем розумного будинку.

Невирішеною частиною загальної проблеми залишається розробка комплексного методу, який би поєднував точність машинного навчання в ідентифікації типів шумів з гнучкістю адаптивного вибору стратегій очищення в режимі реального часу для сенсорних даних з різними характеристиками.

Мета статті. Метою даного дослідження є розробка та експериментальна оцінка ефективності методу адаптивного очищення різнорідних сенсорних даних ACRA (Adaptive Classification-based Real-time Anomaly cleaning) для підвищення якості та надійності інформації в системах розумного будинку.

РЕЗУЛЬТАТИ ДОСЛІДЖЕННЯ

Запропоновано метод ACRA призначений для адаптивного очищення одновимірних часових рядів сенсорних даних у режимі реального часу. Основою методу є інтелектуальна ідентифікація типу присутнього шуму з подальшим вибором відповідної стратегії його усунення. ACRA складається з ключових компонентів, що взаємодіють послідовно при надходженні кожної нової точки даних (рис. 1).

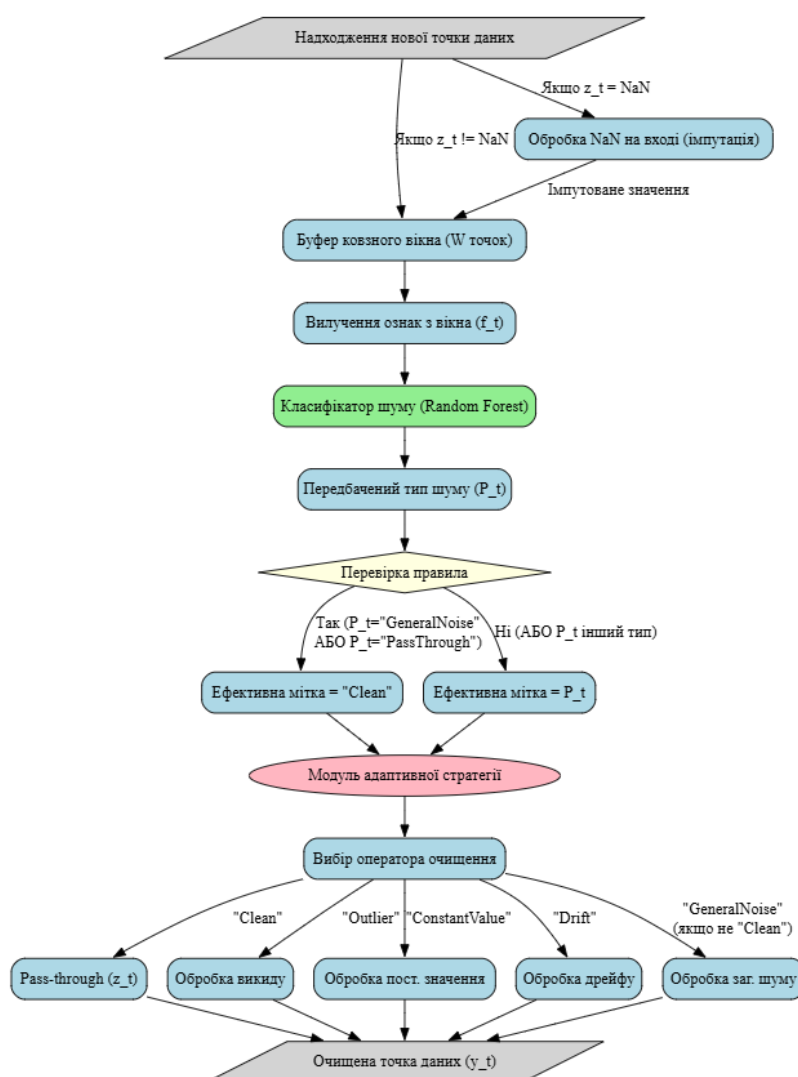


Рис. 1. Блок-схема роботи методу ACRA



Метод обробляє дані послідовно, приймаючи рішення на основі інформації, доступної в межах ковзного вікна останніх точок даних. Архітектура складається з наступних ключових компонентів, що працюють ітераційно по мірі надходження нових точок даних.

Буфер ковзного вікна (Sliding Window Buffer): Для аналізу даних у контексті їх часової послідовності використовується буфер ковзного вікна фіксованого розміру. У даному дослідженні розмір вікна було встановлено на 10 останніх точок даних. При надходженні нової точки даних вона додається до буфера, а найстаріша точка, якщо буфер повний, видаляється. Буфер забезпечує необхідну історію для вилучення ознак та прийняття рішень щодо очищення. Якщо вхідне значення є пропущеним (NaN), воно попередньо імпутується медіаною останніх кількох вже очищених значень перед додаванням до буфера.

Модуль вилучення ознак. Для кожної повної конфігурації ковзного вікна розраховується набір статистичних ознак. Ці ознаки призначені для кількісного опису характеристик даних у поточному вікні, що можуть вказувати на наявність та тип шуму. До набору ознак входять: статистичні моменти (середнє, дисперсія, стандартне відхилення), порядкові статистики (мінімум, максимум, розмах), робастні статистики (інтерквартильний розмах, медіанне абсолютне відхилення), коефіцієнт варіації, середнє абсолютне відхилення між послідовними точками та індикатори пропущених даних.

Класифікатор типів шуму. Основою методу ACRA є класифікатор на основі машинного навчання, реалізований за допомогою алгоритму Random Forest. Цей класифікатор навчається на попередньо підготовленому наборі даних з розміченими типами шумів і призначений для ідентифікації домінуючого типу шуму у поточному вікні на основі вилучених ознак. У даному дослідженні класифікатор навчався розрізняти чотири основні категорії шуму: «Outlier», «ConstantValue», «Drift» та «GeneralNoise».

Правило для визначення «чистого» стану. Оскільки класифікатор не видає мітку «Clean», для ідентифікації періодів з низьким рівнем шуму використовується додаткове правило. Якщо класифікатор визначає поточний стан як «GeneralNoise» (або «PassThrough» у випадку невпевненого передбачення), але при цьому розрахована дисперсія даних у вікні є нижчою за попередньо визначений поріг, то ефективна мітка для поточної точки встановлюється як «Clean».

Модуль адаптивної стратегії та оператори очищення. На основі ефективної мітки модуль адаптивної стратегії обирає та застосовує відповідний оператор для корекції поточного значення, отримуючи на виході очищене значення. Стратегії включають пропуск даних («Clean»), заміну на медіану вікна («Outlier»), заміну на медіану/середнє останніх кількох точок («ConstantValue», «Drift», «GeneralNoise»).

Таким чином, інтеграція IoT-технологій у розумні будинки дозволяє автоматизувати управління ресурсами та підвищити ефективність їх використання, що особливо важливо для забезпечення енергозбереження та підвищення комфортності проживання.

Основою для експериментів слугував реальний набір даних, зібраний за допомогою сенсорів, встановлених у житловому приміщенні на північному сході Мексики [12]. Базовий набір містить 605 260 зразків, зібраних протягом 14 місяців (з 5 листопада 2022 року по 5 січня 2024 року) з хвилинною частотою та охоплює 19 змінних, пов'язаних зі споживанням енергії та погодними умовами.

Для цього дослідження було обрано три репрезентативні одновимірні часові ряди: внутрішню температуру (temp), внутрішню вологість (humidity) та загальне активне енергоспоживання будинку (active_power). З них було виділено календарний місяць



(січень 2023 року) з частотою в одну хвилину. В результаті було отримано 20 160 точок даних для кожного ряду. Крім того, для побудови «еталонного» сегмента було відібрано безперервний тижневий фрагмент внутрішньої температури (10 080 точок) з 1 по 7 січня 2023 року, який демонстрував типові добові коливання.

Усі обрані сегменти використовувалися як чистий сигнал, до якого додатково вводилися контрольовані синтетичні аномалії: викиди, періоди сталих значень, дрейф, адитивний гаусівський шум та пропущені значення (NaN). Параметри кожного типу шуму підбиралися індивідуально для кожного ряду, з урахуванням характерних особливостей відповідного сенсорного сигналу, з метою забезпечення реалістичності моделювання.

Експериментальне дослідження включало наступні етапи для кожного з трьох типів сенсорних даних:

1. Розділення зашумлених даних та відповідних міток типів шуму на навчальну (70%) та тестову (30%) вибірки.
2. Навчання окремої моделі класифікатора ACRA (Random Forest) для кожного типу сенсорних даних на відповідній навчальній вибірці.
3. Розрахунок індивідуальних порогів для правила "Clean" та параметрів R і Q для фільтра Калмана на основі характеристик навчальних даних.
4. Ініціалізація методів ACRA, Rolling Median і Kalman Filter.
5. Послідовна обробка (симуляція в реальному часі) тестових зашумлених даних кожним з методів очищення.
6. Оцінка якості очищення.
7. Оцінка ефективності внутрішньої класифікації ACRA.

Ефективність усіх методів очищення оцінювалася за допомогою наступних показників на тестовому наборі.

Середньоквадратична похибка (RMSE) вимірює квадратний корінь із середньоквадратичної різниці між очищеними/оціненими значеннями та еталонними значеннями. Середня абсолютна похибка (MAE) вимірює середню абсолютну різницю між очищеними/оціненими значеннями та еталонними значеннями.

Відсоткове зниження похибок, досягнуте методом ACRA порівняно з базовими методами, було розраховано шляхом обчислення різниці між значенням метрики базового методу та значенням метрики запропонованого методу, ділення цієї різниці на значення метрики базового методу, а потім множення результату на 100.

Щоб зрозуміти поведінку внутрішнього механізму класифікації, ми порівняли ефективну мітку, призначену методом ACRA на кожному кроці, з істинним базовим типом шуму, введеним у тестові дані. Метрики включали: загальну точність, прецизійність, повноту, F1-оцінку (для кожного класу, зважену та макроусереднену), матрицю плутанини.

Середній час обробки однієї точки даних (у мілісекундах) вимірювався для кожного методу під час моделювання для порівняння їх обчислювальної ефективності.

Підсумкові значення метрик та середній час обробки однієї точки даних для всіх досліджуваних методів на тестових вибірках наведено в табл. 1–3 та рис. 2–4.

Таблиця 1

Результати ефективності очищення для даних температури

Метод	RMSE	MAE	AvgTimeMs (мс)
ACRA	0.3671	0.1520	33.46
RollingMedian	0.5270	0.2157	0.07
KalmanFilter	1.0695	0.7829	0.01
Noisy	4.1988	0.6362	0.00



Порівняння ефективності очищення для: temp

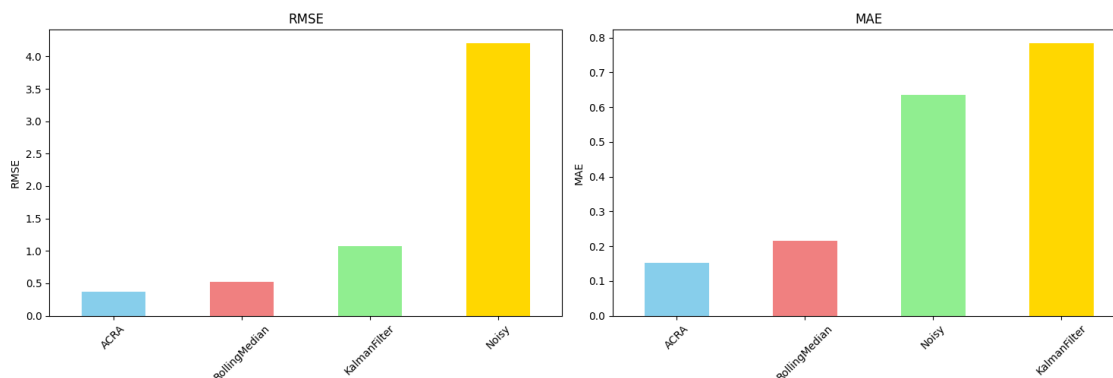


Рис. 2. Порівняння ефективності очищення за для температури

Таблиця 2

Результати ефективності очищення для даних вологості

Метод	RMSE	MAE	AvgTimeMs (мс)
ACRA	2.2730	0.6205	33.32
RollingMedian	3.2679	1.0532	0.07
KalmanFilter	6.0318	4.3208	0.01
Noisy	11.6855	1.6722	0.00

Порівняння ефективності очищення для: humidity

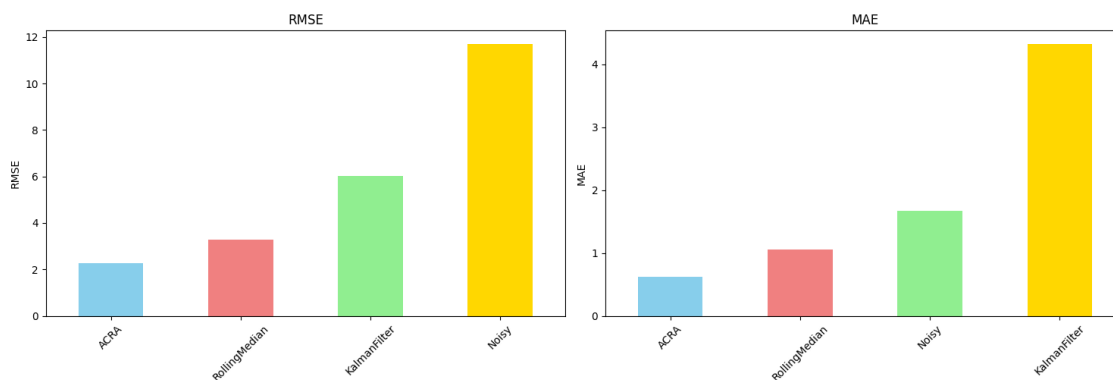


Рис. 3. Порівняння ефективності очищення за для вологості

Таблиця 3

Результати очищення для даних активного енергоспоживання

Метод	RMSE	MAE	AvgTimeMs (мс)
ACRA	104.6866	36.0271	33.30
RollingMedian	128.3362	44.0032	0.07
KalmanFilter	153.3934	93.7641	0.01
Noisy	200.3566	36.7711	0.00

Порівняння ефективності очищення для: active_power

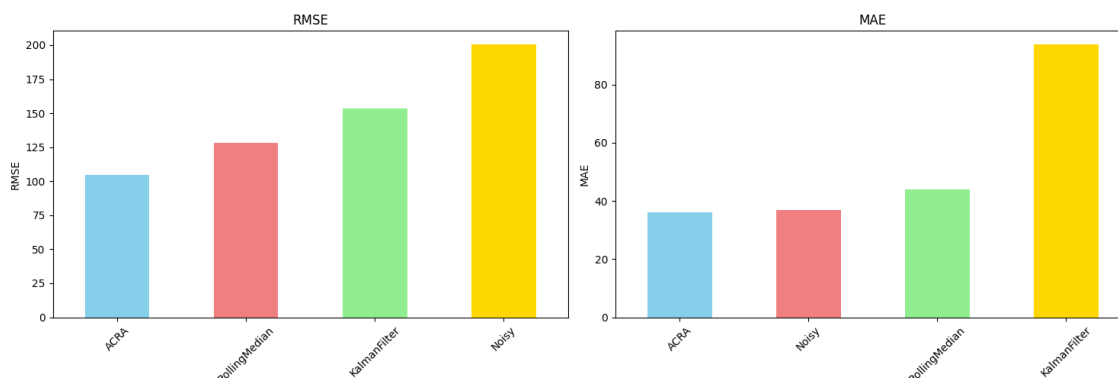


Рис. 4. Порівняння ефективності очищення за для енергоспоживання

Відповідно до результатів наведених у таблицю 1–3 та рис. 2–4, метод ACRA демонструє найкращі показники RMSE та MAE для всіх трьох типів сенсорних даних. Для даних температури ACRA забезпечує зниження RMSE на 30.3% порівняно з Rolling Median та на 65.5% порівняно з Kalman Filter. Для даних вологості відповідні показники зниження RMSE становлять 30.5% та 62.3%. Для даних активного енергоспоживання ACRA знижує RMSE на 18.4% відносно Rolling Median та на 31.9% відносно Kalman Filter. Ці результати підкреслюють ефективність адаптивного підходу ACRA.

Щодо обчислювальної вартості, метод ACRA потребує близько 33 мс на обробку однієї точки даних, що є прийнятним для реалізації в режимі реального часу для даних з хвилинною дискретизацією.

Оцінка ефективності внутрішнього механізму прийняття рішень методом ACRA показала загальну точність зіставлених рішень на рівні 0.4883 для температури, 0.4762 для вологості та 0.3709 для енергоспоживання. Аналіз матриць помилок та звітів класифікації показує, що структуровані аномалії та викиди ідентифікуються з високою повнотою, хоча точність для викидів може бути нижчою через їх часткову схожість з інтенсивним загальним шумом.

Приклад роботи методів на сегменті даних температури наведено на рис. 5.

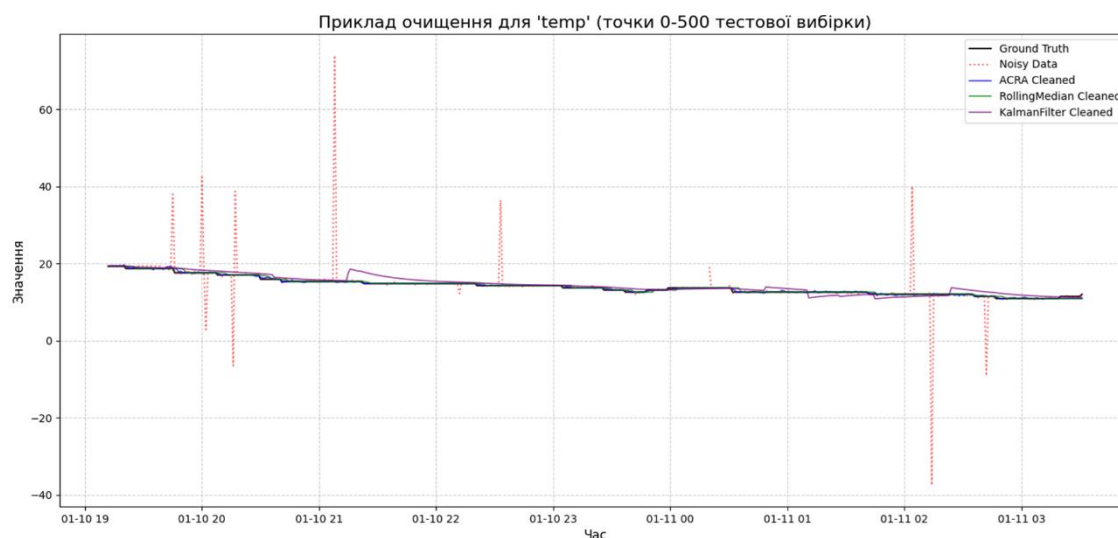


Рис. 5. Приклад очищення часового ряду температури різними методами



Незважаючи на те, що точність внутрішньої класифікації не є ідеальною, загальна ефективність очищення даних методом ACRA залишається високою. Це свідчить про те, що навіть якщо тип шуму ідентифіковано не абсолютно точно, адаптивний вибір стратегії очищення дозволяє суттєво покращити якість даних порівняно з вихідним сигналом та менш гнучкими методами.

ВИСНОВКИ ТА ПЕРСПЕКТИВИ ПОДАЛЬШИХ ДОСЛІДЖЕНЬ

У ході проведеного дослідження було розроблено та експериментально оцінено метод адаптивного очищення різнорідних сенсорних даних ACRA, призначений для підвищення якості інформації в системах розумного будинку. Запропонований метод, що базується на класифікації типів шумів за допомогою моделі Random Forest та адаптивному виборі стратегії корекції, продемонстрував високу ефективність.

Експерименти на даних температури, вологості та енергоспоживання, зашумлених синтетично, показали, що ACRA стабільно перевершує традиційні підходи, такі як ковзна медіана та фільтр Калмана, за показниками RMSE та MAE для всіх досліджуваних типів даних. Це підтверджує здатність методу ефективно усувати різноманітні типи аномалій, зберігаючи при цьому корисні характеристики сигналу, та свідчить про його гнучкість при роботі з даними, що мають різні фізичні властивості та характеристики шумів.

Таким чином, ACRA є перспективним інструментом для попередньої обробки даних у реальних системах розумного будинку, забезпечуючи надійнішу інформаційну основу для подальшого аналізу та прийняття рішень. Хоча внутрішня класифікація типів шумів не є ідеальною, загальна адаптивна стратегія дозволяє досягти високої якості очищення. Обчислювальна складність методу є прийнятною для багатьох практичних застосувань з хвилинною дискретизацією даних.

Незважаючи на отримані позитивні результати, існують напрямки для подальшого вдосконалення методу. До них належать дослідження більш складних архітектур класифікаторів, розширення методу для обробки багатовимірних часових рядів з урахуванням кореляцій між сенсорними потоками, розробка механізмів автоматичної оптимізації параметрів ACRA, а також його оптимізація для розгортання на периферійних пристроях з обмеженими обчислювальними ресурсами. Подальша валідація методу на ширшому спектрі реальних промислових даних та оцінка його впливу на ефективність конкретних прикладних задач у системах розумного будинку також є важливими кроками для підтвердження його практичної значущості.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Gubbi, J., Buyya, R., Marusic, S., & Palaniswami, M. (2013). Internet of Things (IoT): A vision, architectural elements, and future directions. *Future Generation Computer Systems*, 29(7), 1645–1660. <https://doi.org/10.1016/j.future.2013.01.010>
2. Karkouch, A., Mousannif, H., Al Moatassime, H., & Noel, T. (2016). Data quality in internet of things: A state-of-the-art survey. *Journal of Network and Computer Applications*, 73, 57–81. <https://doi.org/10.1016/j.jnca.2016.08.002>
3. Sethi, S. (2024). Data Governance in Smart Home Systems: The S.H.I.E.L.D. Framework. *International Journal For Multidisciplinary Research*, 6(1). <https://doi.org/10.36948/ijfmr.2024.v06i01.39041>
4. Wang, X., & Wang, C. (2020). Time Series Data Cleaning: A Survey. *IEEE Access*, 8, 1866–1881. <https://doi.org/10.1109/access.2019.2962152>



5. Phan, M. Q., Vicario, F., Longman, R. W., & Betti, R. (2017). State-Space Model and Kalman Filter Gain Identification by a Kalman Filter of a Kalman Filter. *Journal of Dynamic Systems, Measurement, and Control*, 140(3). <https://doi.org/10.1115/1.4037778>
6. S., Mohanavalli, S., Sripriya, N., & Poornima, S. (2018). Outlier Detection using Clustering Techniques. *International Journal of Engineering & Technology*, 7(3.12), 813. <https://doi.org/10.14419/ijet.v7i3.12.16508>
7. Corrales, D., Corrales, J., & Ledezma, A. (2018). How to Address the Data Quality Issues in Regression Models: A Guided Process for Data Cleaning. *Symmetry*, 10(4), 99. <https://doi.org/10.3390/sym10040099>
8. Ibebuchi, C. C. (2024). Fuzzy time series clustering using autoencoders neural network. *AIMS Geosciences*, 10(3), 524–539. <https://doi.org/10.3934/geosci.2024027>
9. Kirichenko, L., Koval, Y., Yakovlev, S., & Chumachenko, D. (2024). Anomaly Detection in Fractal Time Series with LSTM Autoencoders. *Mathematics*, 12(19), 3079. <https://doi.org/10.3390/math12193079>
10. Raju, K. H. P., Sandhya, N., & Mehra, R. (2017). Supervised SVM Classification of Rainfall Datasets. *Indian Journal of Science and Technology*, 10(15), 1–6. <https://doi.org/10.17485/ijst/2017/v10i15/106115>
11. Zamani Sabzi, H., Abudu, S., Alizadeh, R., Soltanisehat, L., Dilekli, N., & King, J. P. (2018). Integration of time series forecasting in a dynamic decision support system for multiple reservoir management to conserve water sources. *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects*, 40(11), 1398–1416. <https://doi.org/10.1080/15567036.2018.1476934>
12. Aguirre-Fraire, B., Beltrán, J., & Soto-Mendoza, V. (2024). A Comprehensive Dataset Integrating Household Energy Consumption and Weather Conditions in a North-eastern Mexican Urban City. *Data in Brief*, 110452. <https://doi.org/10.1016/j.dib.2024.110452>

**Dmytro Nishchemenko**

Postgraduate Student

State university of information and communication technologies, Kyiv, Ukraine

ORCID ID: 0009-0006-1781-4109

dima.nishchemenko@gmail.com**Olena Volochchuk**

Candidate of Technical Sciences, Associate Professor

Kharkiv National University of Radio Electronics, Kharkiv, Ukraine

ORCID ID: 0000-0002-5912-4126

Olena.voloshchuk@nure.ua**ADAPTIVE REAL-TIME CLEANING OF HETEROGENEOUS SENSOR DATA
IN SMART HOME SYSTEMS BASED ON NOISE CLASSIFICATION**

Abstract. Modern smart home systems generate substantial volumes of data, the quality of which is critically important for effective management, analysis, and forecasting. However, raw data streams often contain complex combinations of noise and anomalies, including outliers, concept drift, and periods of stagnant values. Such artifacts significantly reduce data reliability, potentially leading to incorrect operation of intelligent systems and flawed decision-making. Existing data cleaning methods often demonstrate limited effectiveness when dealing with heterogeneous types of noise in real-time, particularly when processing sensor data with varying physical characteristics. This paper presents a novel method — ACRA (Adaptive Classification-based Real-time Anomaly cleaning) — designed for adaptive cleaning of heterogeneous sensor time series data in the context of smart homes. The ACRA method incorporates a classifier based on an ensemble of decision trees (Random Forest) to identify specific types of noise within a sliding window of incoming data. The classification results are supplemented by a heuristic rule that leverages window variance analysis to more accurately detect low-noise periods. Based on the combination of the classifier's output and the rule, an adaptive strategy module is activated to dynamically select the most appropriate operator for correcting the current data value. To evaluate the effectiveness of the proposed method, an experimental study was conducted using real-world time series of temperature, humidity, and total energy consumption collected in a residential environment, with synthetically injected controlled noise types. The experimental results demonstrated that the ACRA method provides significantly better cleaning quality compared to common baseline methods such as the moving median and Kalman filter across all three types of sensor data studied. The proposed ACRA method is a robust and flexible tool for enhancing the quality of sensor data in smart home systems, laying the foundation for the development of more accurate and efficient intelligent applications.

Keywords: data cleaning; smart home; Internet of Things; adaptive anomaly removal; noise classification; machine learning.

REFERENCES (TRANSLATED AND TRANSLITERATED)

1. Gubbi, J., Buyya, R., Marusic, S., & Palaniswami, M. (2013). Internet of Things (IoT): A vision, architectural elements, and future directions. *Future Generation Computer Systems*, 29(7), 1645–1660. <https://doi.org/10.1016/j.future.2013.01.010>
2. Karkouch, A., Mousannif, H., Al Moatassime, H., & Noel, T. (2016). Data quality in internet of things: A state-of-the-art survey. *Journal of Network and Computer Applications*, 73, 57–81. <https://doi.org/10.1016/j.jnca.2016.08.002>
3. Sethi, S. (2024). Data Governance in Smart Home Systems: The S.H.I.E.L.D. Framework. *International Journal For Multidisciplinary Research*, 6(1). <https://doi.org/10.36948/ijfmr.2024.v06i01.39041>
4. Wang, X., & Wang, C. (2020). Time Series Data Cleaning: A Survey. *IEEE Access*, 8, 1866–1881. <https://doi.org/10.1109/access.2019.2962152>
5. Phan, M. Q., Vicario, F., Longman, R. W., & Betti, R. (2017). State-Space Model and Kalman Filter Gain Identification by a Kalman Filter of a Kalman Filter. *Journal of Dynamic Systems, Measurement, and Control*, 140(3). <https://doi.org/10.1115/1.4037778>



6. S., Mohanavalli, S., Sripriya, N., & Poornima, S. (2018). Outlier Detection using Clustering Techniques. *International Journal of Engineering & Technology*, 7(3.12), 813. <https://doi.org/10.14419/ijet.v7i3.12.16508>
7. Corrales, D., Corrales, J., & Ledezma, A. (2018). How to Address the Data Quality Issues in Regression Models: A Guided Process for Data Cleaning. *Symmetry*, 10(4), 99. <https://doi.org/10.3390/sym10040099>
8. Ibebuchi, C. C. (2024). Fuzzy time series clustering using autoencoders neural network. *AIMS Geosciences*, 10(3), 524–539. <https://doi.org/10.3934/geosci.2024027>
9. Kirichenko, L., Koval, Y., Yakovlev, S., & Chumachenko, D. (2024). Anomaly Detection in Fractal Time Series with LSTM Autoencoders. *Mathematics*, 12(19), 3079. <https://doi.org/10.3390/math12193079>
10. Raju, K. H. P., Sandhya, N., & Mehra, R. (2017). Supervised SVM Classification of Rainfall Datasets. *Indian Journal of Science and Technology*, 10(15), 1–6. <https://doi.org/10.17485/ijst/2017/v10i15/106115>
11. Zamani Sabzi, H., Abudu, S., Alizadeh, R., Soltanisehat, L., Dilekli, N., & King, J. P. (2018). Integration of time series forecasting in a dynamic decision support system for multiple reservoir management to conserve water sources. *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects*, 40(11), 1398–1416. <https://doi.org/10.1080/15567036.2018.1476934>
12. Aguirre-Fraire, B., Beltrán, J., & Soto-Mendoza, V. (2024). A Comprehensive Dataset Integrating Household Energy Consumption and Weather Conditions in a North-eastern Mexican Urban City. *Data in Brief*, 110452. <https://doi.org/10.1016/j.dib.2024.110452>

