

УДК 004.852:621.643
DOI <https://doi.org/10.32689/maup.it.2025.4.31>

Лідія ШТАЄР

кандидат технічних наук, доцент,
доцент кафедри інформаційно-телекомунікаційних технологій та систем,
Івано-Франківський національний технічний університет нафти і газу, lidia.shtaier@nung.edu.ua
ORCID: 0000-0003-1013-9869

Наталія ВОЗНА

доктор технічних наук, професор,
професор кафедри спеціалізованих комп'ютерних систем,
Західноукраїнський національний університет, n.vozna@wupn.edu.ua
ORCID: 0000-0002-8856-1720

Володимир ГРИГА

кандидат технічних наук, доцент,
доцент кафедри комп'ютерної інженерії та електроніки,
Карпатський національний університет імені Василя Стефаника, gr.volodymyr2018@gmail.com
ORCID: 0000-0001-5458-525X

Леся МИЧУДА

доктор технічних наук, професор,
професор кафедри безпеки інформаційних технологій,
Національний університет «Львівська політехніка», lesia.z.mychuda@lpnu.ua
ORCID: 0000-0001-8266-1782

Оксана БЕЛЕЙ

кандидат технічних наук, доцент,
доцент кафедри інформаційно-телекомунікаційних технологій та систем,
Івано-Франківський національний технічний університет нафти і газу, oksana.belei@nung.edu.ua
ORCID: 0000-0002-2386-4106

ПІДВИЩЕННЯ СТІЙКОСТІ АКУСТИЧНОГО МЕТОДУ КОНТРОЛЮ ВИТОКІВ ШЛЯХОМ ЗАСТОСУВАННЯ ЕНТРОПІЙНИХ ТА РЕЛЯЦІЙНИХ ОЗНАК

Анотація. Вирішення питання надійного транспортування енергоносіїв з використанням трубопровідного транспорту є першочерговим завданням для забезпечення екологічної та економічної стійкості держави. Достовірність відомих методів контролю стану трубопроводів, зокрема, методів акустичного контролю витоків, суттєво залежить від зміни технологічних умов транспортування. Ця проблема потребує розроблення вдосконалених підходів до формування ознак стану, стійких до експлуатаційної невизначеності.

Мета роботи полягає у вдосконаленні алгоритму опрацювання діагностичних сигналів в акустичному методі контролю витоків з трубопроводів шляхом розроблення методу класифікації станів, стійкого до зміни параметрів експлуатації. Завдання включають: обґрунтування вибору інформативних ознак акустичного сигналу, які є стійкими до зміни робочого тиску та характеристик тестового акустичного сигналу; вибір оптимальної архітектури класифікатора, який здатний до ефективного узагальнення на нових даних, шляхом порівняння лінійних, нелінійних та ансамблевих методів машинного навчання.

Методологія дослідження ґрунтується на опрацюванні зареєстрованих відбитих від неоднорідностей стінок трубопроводу акустичних сигналів, які записані в середовищі заповнення трубопроводу як результат поширення згенерованих тестових сигналів (тестовий метод діагностування). При проведенні експериментальних досліджень сигнали реєструвались для випадків моделювання п'яти різних станів трубопроводу (однорідні стінки та наявні витокі діаметрами 1, 3, 5 та 10 мм) при різних значеннях тиску в трубопроводі та різних характеристиках тестових сигналів. На відміну від традиційних підходів до формування вектора ознак з використанням абсолютних значень спектрально-енергетичних параметрів сигналу в роботі запропоновано розширити вектор ознак ентропійними складовими та реляційними, визначеними зі співвідношення між оцінками, одержаними для кожного каналу реєстрування. Такий підхід дозволив компенсувати синфазні завади, зумовлені зміною робочих умов. Проведено порівняння лінійної (перцепторон), нелінійної (MLP) та ансамблевої (Random Forest) архітектур класифікатора. Якість навчених моделей оцінено з використанням стратифікованої перехресної валідації на навчальному наборі та тестуванні на новій вибірці з невідомими моделям варіаціями тиску та параметрами тестового сигналу.

© Л. Штаєр, Н. Возна, В. Грига, Л. Мичуда, О. Белей, 2025

Стаття поширюється на умовах ліцензії CC BY 4.0

Наукова новизна дослідження полягає у: формуванні комбінованого вектора ознак з використанням ентропійних та реляційних характеристик акустичного сигналу, що забезпечило стійкість класифікатора до змін робочого тиску та несучої частоти тестового сигналу; експериментальному підтвердженні нелінійності задачі класифікації акустичних сигналів витоків у пропонованому просторі ознак; встановленні вищої здатності до узагальнення ансамблевого класифікатора порівняно з нейромережевими, що забезпечив точність класифікації на незалежному наборі даних на рівні 91%).

Висновки. Доведено ефективність застосування запропонованого вектора ознак та ансамблевої моделі Random Forest в алгоритмі виявлення стану трубопроводу при застосуванні акустичного методу контролю, який є стійким до змін робочого тиску та параметрів тестового сигналу. Результати дослідження можуть бути використані при розробленні алгоритмічного забезпечення систем моніторингу стану трубопроводів, стійких до зміни технологічних умов транспортування.

Ключові слова: контроль витоків, опрацювання сигналів, машинне навчання, інженерія ознак, ентропія, перцептрон, імовірнісні оцінки.

Lidiia SHTAIER, Natalia VOZNA, Volodymyr HRYHA, Lesya MYCHUDA, Oksana BELEI. ENHANCING THE ROBUSTNESS OF ACOUSTIC LEAK DETECTION METHOD BY APPLYING ENTROPY AND RELATIONAL FEATURES

Abstract. Addressing the challenge of reliable transportation of energy carriers via pipelines is a priority for ensuring the state's environmental and economic sustainability. The reliability of existing pipeline condition monitoring methods, particularly acoustic leak detection techniques, is significantly affected by variations in operational conditions. This issue necessitates the development of advanced approaches for generating condition features that are robust to operational uncertainty.

Objective. The objective of this work is to improve the diagnostic signal processing algorithm within the acoustic pipeline leak detection method by developing a state classification method robust to changes in operational parameters. The tasks include justifying the selection of informative acoustic signal features that are robust to variations in operating pressure and characteristics of the test acoustic signal, selecting the optimal classifier architecture capable of effective generalization to new data by comparing linear, non-linear, and ensemble machine learning methods.

Methodology. The research methodology is based on processing acoustic signals reflected from pipeline wall inhomogeneities. These signals were recorded within the pipeline medium because of the propagation of generated test signals (active diagnostic method). Experimental data were acquired for five simulated pipeline states (closed pipeline and pipeline with leaks with diameters of 1, 3, 5, and 10 mm) under varying pressure conditions and using different test signal characteristics. In contrast to traditional approaches that rely on absolute spectral-energy parameters for feature vector formation, this study proposes augmenting the feature vector with entropy components and relational features derived from the ratios between estimates obtained for each recording channel. This approach enabled the compensation of common-mode noise caused by variations in operating conditions. A comparison was conducted between linear (Perceptron), non-linear (MLP), and ensemble (Random Forest) classifier architectures. Model performance was evaluated using stratified cross-validation on the training set and by testing on a new dataset containing pressure variations and test signal parameters unseen by the models.

Scientific Novelty. The scientific novelty of the research lies in: developing a combined feature vector using entropy and relational characteristics of the acoustic signal, which ensured classifier robustness to variations in operating pressure and test signal carrier frequency; experimentally confirming the non-linearity of the acoustic leak signal classification task within the proposed feature space; establishing the superior generalization capability of the ensemble classifier compared to neural network-based approaches, which achieved a classification accuracy of 91% on an independent dataset.

Conclusion. The effectiveness of applying the proposed feature vector and the Random Forest ensemble model within the acoustic pipeline state detection algorithm has been proven. The method demonstrated robustness to variations in operating pressure and test signal parameters. The research results can be applied in the development of algorithmic support for pipeline condition monitoring systems that are resilient to changes in technological transportation conditions.

Key words: leak detection, signal processing, machine learning, feature engineering, entropy, perceptron, probabilistic estimation.

Вступ. Трубопровідний транспорт складає вагомую частку у переміщенні енергоносіїв на значні відстані. При цьому забезпечення надійного транспортування є першочерговим завданням для підтримання екологічної та економічної стійкості держави. Природа появи витоків з трубопроводів може бути різною – від корозії до механічного пошкодження стінки трубопроводу при проведенні земляних робіт або несанкціонованого доступу до енергоносія. Наслідки від появи витоків оцінюються не тільки економічними та екологічними збитками, а й часто супроводжуються людськими втратами. Зважаючи на це, розроблення ефективних методів контролю стану трубопроводу є актуальною науковою задачею.

Перспективним напрямом побудови систем моніторингу стану трубопроводів є застосування акустичних (тестових та пасивних) методів контролю витоків, які характеризуються можливістю дистанційного виявлення витoku. Основною проблемою впровадження таких систем є суттєве погіршення здатності до виявлення витоків в залежності від зміни технологічних умов транспортування та високий рівень завад в поєднанні з малою амплітудою корисного сигналу від витoku. Ця проблема потребує розроблення вдосконалених підходів до формування ознак стану трубопроводу, стійких до експлуатаційної невизначеності.

Аналіз останніх досліджень і публікацій. Акустичні методи діагностування стану трубопроводів використовуються для трубопроводів різного призначення та матеріалу виготовлення. Зокрема, методи кореляційного аналізу [1, 7], які вважаються класичним методом локалізації витоків, ґрунтуються на розрахунку взаємно кореляційної функції (ВКФ) сигналів від давачів (двох або більше) та встановленні координати витoku за часом затримки поширення шуму від витoku. В роботі [7] запропоновано вдосконалену реалізацію для сигналів акустичної емісії (АЕ) шляхом розрахунку вторинної ВКФ з фазовим перетворенням та додатковим застосуванням інтерполяції. Робота [1] зорієнтована на поширенні тестових акустичних сигналів в середовищі заповнення труби та опрацюванні відбитих від неоднорідностей стінок сигналів. Підвищення завадостійкості корисного сигналу тут реалізовано з використанням тестового сигналу, згенерованого за кодом Баркера, реєструванні сигналів двома давачами (канали реєстрування), обчислення ВКФ (та її обвідної) каналів для уточнення швидкості поширення акустичних коливань в середовищі, формуванні імпульсної характеристики трубопроводу при відсутності витoku та різницевої – за наявності витoku. Проте, за наявності значного рівня завад, зміні технологічних умов транспортування (тиск, температура) та зміні характеру сигналу надійність встановлення факту появи витoku суттєво зменшується. Це веде до ускладнення застосування методів, що базуються на обчисленні ВКФ, для автоматичної класифікації стану контрольованої ділянки. Можливість врахування впливу появи витoku з використанням математичного моделювання показано в [8]. Особливістю такого підходу є можливість адаптації моделі до різних типів речовини та режимів течії, проте успішність підходу суттєво залежить від коректності заданих характеристик процесу, стійкості чисельної схеми та фізичної картини витoku.

Сьогодні активно розвиваються методи, що поєднують опрацювання акустичних сигналів та методи машинного навчання (ML). Спільним в таких роботах [3-4, 9] є розгляд побудови моделей на основі вектора ознак, що включає енергетичні, спектральні та часові характеристики. Автори [3] відзначають важливість якості та кількості даних для побудови якісних моделей ML, а також необхідності додаткових досліджень ефективності моделей при перенесенні їх на інші середовища транспортування за рахунок різниці у фізиці процесу поширення хвиль. Робота [9], яка присвячена опрацюванню сигналів АЕ, демонструє залежність точності класифікації від стратегії формування вектора ознак (зокрема, включення статистичних ознак) та обраного алгоритму класифікації. Зазначено, що найефективнішим класифікатором в лабораторних умовах виявився ансамблевий метод ML Random Forest. Недолік методу полягає у швидкості загасання сигналів АЕ, що обмежує відстань виявлення витoku, та відсутність незалежної вибірки для тестування моделі в промислових умовах. На противагу [9], автори [4] розглядають питання про високий рівень хибного спрацювання для мереж водопостачання на основі даних з автономних витокошукачів. В роботі проведено порівняльний аналіз ефективності різних алгоритмів ML для класифікації зареєстрованих сигналів, запропоновано рейтинг алгоритмів за критеріями точності та чутливості, проте робота зосереджена виключно на даних з автономних витокошукачів, якість зареєстрованих сигналів яких не дозволяє побудувати складні ознаки для навчання моделей та необхідно враховувати значний вплив матеріалу трубопроводу.

Основним викликом для моделей ML є суттєва залежність від якості навчальних даних та експлуатаційна невизначеність при виявленні витоків. В роботі [5] запропоновано методику формування тренувальних наборів. Стійкість до експлуатаційних умов вирішується шляхом застосування глибокого навчання та адаптації домену [11], де пропонується використання глибокої згорткової нейромережі (DACNN – Domain Adaptive Convolutional Neural Networks) з спеціальною функцією втрат. Переваги підходу полягають у стійкості до експлуатаційних умов, автоматичному виділенні ознак та перенесенні знань, здобутих на лабораторних даних, на промислові умови. Недоліки полягають у значній обчислювальній складності, великих обсягах даних та складності архітектури мережі.

Перспективним компромісом між складністю глибокого навчання та слабкою інформативністю статистичних ознак авторам дослідження видається підхід [6, 10] з додаванням ентропійних ознак при навчанні моделі. Обрана нейронна мережа в [6] продемонструвала високу швидкість навчання та ефективність на малих вибірках, але вона суттєво залежить від способу визначення ентропійної ознаки. Розвитком такого підходу є запропонований в [10] спосіб формування вектора ознак, який включає зменшений в розмірності набір ознак, що сформовані як лінійне зважування елементів вектора ентропій різного типу. Таке рішення дозволяє врахувати вплив різного типу ентропій і одночасно прискорити навчання моделі, уникнувши при цьому надлишковості та корельованої інформації при формуванні вектора ознак. З іншого боку, складність обчислень різних типів ентропії для кожного вікна сигналу та згортання їх в узагальнену ознаку вимагає відповідних потужностей.

Проведений аналіз показав обмежену здатність до виявлення витоків в умовах високого рівня шуму в класичних методах опрацювання акустичних сигналів [1, 7], методи глибокого навчання

володіють значною обчислювальною складністю [11], статистичні методи [9] характеризуються високим рівнем помилок [4]. Використання ентропії в процесі опрацювання сигналів дозволяє перейти до інформаційного аналізу сигналів [6, 10], але здатність до виявлення витоків в таких методах залежить від наявності корельованих завод. Це визначає потребу в розробленні вдосконалених підходів до формування простору ознак стану, які поєднують інформативність ентропійних показників з стійкістю реляційних (відносних) характеристик.

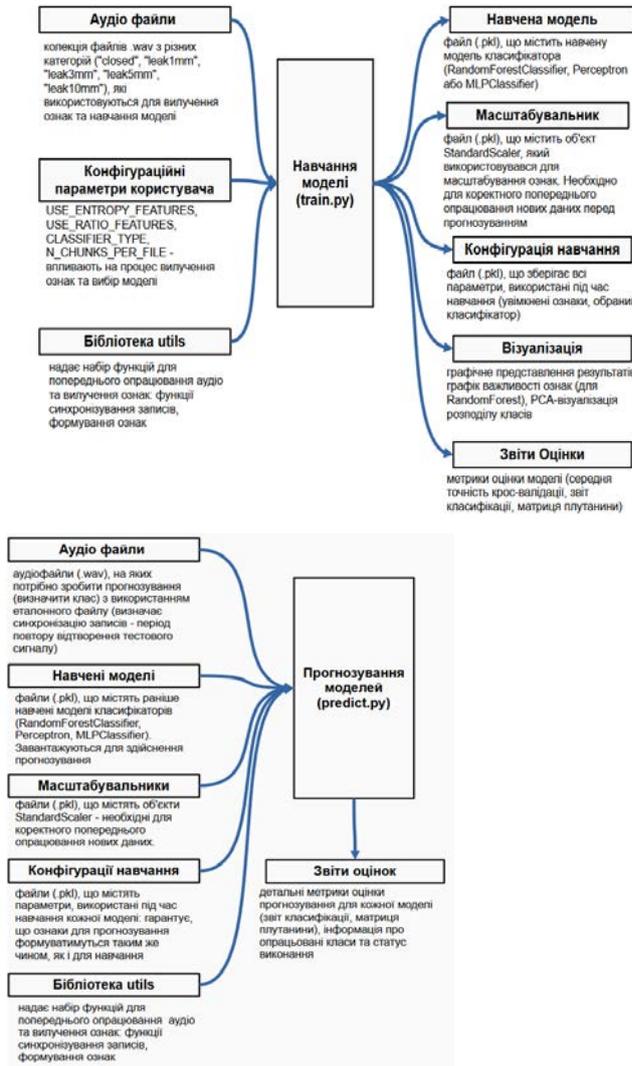
Мета статті полягає у вдосконаленні алгоритму опрацювання діагностичних сигналів в акустичному методі контролю витоків з трубопроводів шляхом розроблення методу класифікації станів, стійкого до зміни параметрів експлуатації. Для досягнення поставленої мети необхідно вирішити такі завдання: обґрунтувати вибір інформативних ознак акустичного сигналу, які є стійкими до зміни робочого тиску та характеристик тестового акустичного сигналу; здійснити вибір оптимальної архітектури класифікатора, який здатний до ефективного узагальнення на нових даних, шляхом порівняння лінійних, нелінійних та ансамблевих методів машинного навчання.

Виклад основного матеріалу дослідження. Методика дослідження. Основою акустичного тестового методу контролю є генерування у середовищі заповнення труби тестових сигналів та фіксування відбитих від неоднорідностей акустичних сигналів [1]. Експериментальні дані, які використано для вдосконалення алгоритму опрацювання сигналів, одержано з використанням змодельованої ділянки трубопроводу на основі розгорнутого на території розбірного металевого трубопроводу діаметром 150 мм загальною довжиною 174 м. При проведенні експериментальних досліджень тиск у системі нагнітався за допомогою пересувної компресорної станції та контролювався манометром. У трубопроводі були інтегровані два ключові модулі: модуль випромінювання тестових коливань (генератор імпульсів, змонтований у патрубку довжиною 1 м, діаметр 150 мм) та модуль імітації витоку (ділянка труби довжиною 6 м та діаметром 150 мм) з краном та змінними насадками (1, 3, 5, 10 мм) для моделювання витоку різного розміру. Схема та опис методики проведення досліджень наведено в [1]. Для реєстрації акустичного відгуку системи використовувалась пара акустичних датчиків (мікрофонів МКЭ-3), розташованих в середовищі заповнення трубопроводу на відстані 0,442 м один від одного. Аналогові сигнали з мікрофонів оцифровані за допомогою звукової карти ESI Juli@ з частотою дискретизації 192 кГц (24-біт).

Ідея підвищення стійкості алгоритму контролю стану трубопроводу полягає у використанні методів машинного навчання для класифікації можливих станів за невизначених експлуатаційних умов. Для проведення порівняння ефективності використання класифікаторів було обрано приклад лінійної (перцепторон), нелінійної (багатошаровий перцептрон – MLP) та ансамблевої (Random Forest) архітектур. Для навчання та тестування було зібрано дані, що відповідають п'яти різним станам трубопроводу: closed – герметичний стан (відсутність витоку); leak1mm – витік через отвір діаметром 1 мм; leak3mm, leak5mm, leak10mm – витіки через отвори діаметрами 3, 5 та 10 мм відповідно. Для забезпечення неоднорідності (гетерогенності) даних та перевірки стійкості моделей записи проводились за змінних умов експлуатації. Зміни у тестовому сигналі, згенерованому за принципом заводостійкого кодування (5ти розрядний код Баркера) було реалізовано через: обвідну тестового сигналу (прямокутна форма, Гаусова, синусоїдна), частоту несучої (480 та 960 Гц). Врахування зміни робочого тиску проведено шляхом реєстрування окремих серій записів сигналів для усіх модельованих станів трубопроводу при різному значенні тиску (відсутній тиск; 1; 1,4; 1,6 атмосфер).

Для коректної оцінки здатності моделей до узагальнення, увесь масив даних було розділено на два незалежні набори: навчальний та тестовий. Навчальний набір використовувався для навчання та 5-ти кратної стратифікованої перехресної валідації. Цей набір містив 11 файлів (загалом 216 зразків реєстрування діагностичного сигналу для кожного каналу), що включали приклади усіх класів, типів тестових сигналів та записів з різним значенням тиску. Тестовий набір використовувався тільки для фінальної оцінки навчених моделей і складався з 3х файлів (135 зразків), які ніколи не демонструвалися моделям під час навчання та валідації. Ці дані містили записи сигналів при різному значенні тиску в трубопроводі та при поширенні згенерованих тестових сигналів з різними характеристиками.

Програмна реалізація дослідження можливості використання методів ML для виявлення стану трубопроводу проведена мовою Python (версія 3.10) та побудована за модульним принципом: спільний модуль опрацювання сигналів (utils.py), процедура навчання (train.py) та процедура прогнозування та оцінювання моделей (predict.py). В проєкті застосовано бібліотеки sklearn (побудова та валідація моделей), librosa (аналіз сигналів), scipy (математичний аналіз), joblib (збереження моделей, об'єктів масштабування та файлів конфігурування), matplotlib та seaborn (візуалізація результатів). Функціональне навантаження складових процедури навчання та прогнозування із можливістю порівняння результатів трьох моделей наведено на контекстній діаграмі (рис. 1).



а

б

Рис. 1. Контекстна діаграма процедури: а) навчання моделей; б) тестування

Джерело: складено авторами.

Вибір інформативних ознак акустичного сигналу, які є стійкими до зміни робочого тиску та характеристик тестового акустичного сигналу проводився в два етапи. На першому етапі було сформовано базовий вектор ознак, які традиційно використовуються в задачах акустичного діагностування. Цей набір включав: енергетичні ознаки (енергія для кожного з двох каналів energy_ch1, energy_ch2 та їх співвідношення energy_ratio); статистичні (середньоквадратичне відхилення rms1, rms2 та пікове значення амплітуди peak1, peak2 для кожного каналу); спектральні (середньозважене значення частот spec_centroid1, spec_centroid2 та спектральна ширина смуги spec_bw1, spec_bw2 для кожного каналу); кореляційні (максимальне значення нормалізованої ВКФ між каналами norm_corr). Такий набір ознак продемонстрував високу чутливість до зміни несучої частоти тестового сигналу, що приводило до помилок класифікації.

На другому етапі було доповнено вектор ознак значенням ентропії, яка розраховувалась для кожного каналу (entropy1, entropy2) шляхом усереднення ентропії Шенона, обчисленої у ковзних вікнах по всьому зразку. Розмір вікна становить 2048 відліків, що зумовлено структурою тестового сигналу у вигляді 5ти розрядного коду Баркера (тривалість одного розряду коду – 0,02 с, що відповідає 3840 відлікам). Обраний розмір вікна дозволяє аналізувати локальні властивості сигналу в межах одного кодового імпульсу без впливу переходів між розрядами. Для зменшення чутливості моделей до змін тиску та виділення диференційного ефекту витоку остаточний вектор ознак доповнено реляційними ознаками (відношеннями) rms_ratio та entropy_ratio. Така комбінація базових, ентропійних та

реляційних ознак дозволила побудувати стійку модель, здатну ефективно опрацьовувати дані, отримані за різних експлуатаційних умов.

Результати порівняння обраних класифікаторів (перцепторон, MLP та Random Forest) дозволяють встановити характер задачі (лінійна/нелінійна) та обґрунтувати необхідність у використанні складніших структур. Для моделювання нелінійних залежностей було обрано багатосаровий перцепторон (MLP) з одним прихованим шаром (`hidden_layer_sizes=(50,)`) та функцією активації ReLU. Вибір ансамблевого методу Random Forest зумовлений його специфікою подолання ефекту перенавчання та здатністю до узагальнення. Для запобігання зміщенню прогнозів в бік домінуючих класів, для RandomForest та Perceptron під час навчання було застосовано гіперпараметр `class_weight='balanced'`. Кількісний аналіз продуктивності моделей оцінено з використанням метрик, обчислених на основі матриці плутанини [2]:

- продуктивність (*accuracy*) – загальна частка правильних прогнозів;
- точність (*precision*) – відображає частку зразків, які класифіковані моделлю як певний клас, дійсно належить до цього класу (надійність позитивного прогнозу);
- повнота (*recall*) – показує, яку частку зразків певного класу модель змогла коректно знайти з-поміж усіх наявних зразків цього класу (здатність виявляти цільовий клас);
- оцінка загальної ефективності (*F1-score*) – гармонійне середнє між точністю та повнотою:

$$F1\text{-score} = 2 * \frac{\text{precision} * \text{recall}}{\text{precision} + \text{recall}}.$$

Результати та обговорення. Приклад вигляду згенерованого тестового сигналу в формі коду Баркера, ВКФ між зразками одного запису тривалості 2,1 с (для визначення швидкості поширення сигналу), реєстрування відбитих сигналів та ВКФ з тестовим сигналом для станів закрито та наявності витоку розміром 3 мм показано на (рис. 2).

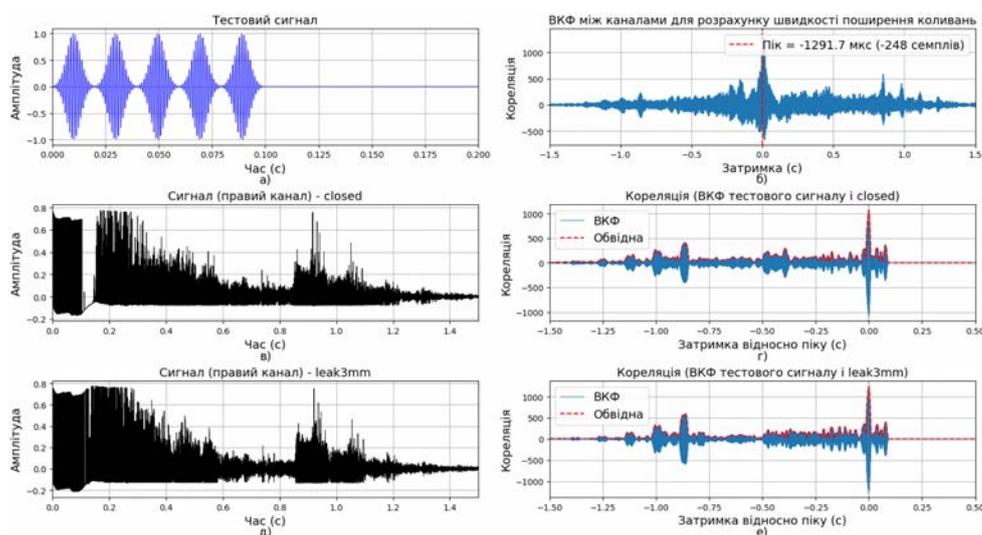


Рис. 2. Часовий та кореляційний аналіз діагностичних сигналів: а) тестовий сигнал; б) ВКФ каналів; в) сигнал правого каналу за відсутності витоку; г) ВКФ для стану закрито; д) сигнал правого каналу за наявності витоку 3 мм; е) ВКФ для стану з витоком 3 мм

Джерело: складено авторами.

Результати дослідження впливу архітектури класифікатора та складу вектора ознак на ефективність виявлення стану трубопроводу наведено в табл. 1. Результати проаналізовано щодо впливу формування складу вектора ознак та щодо здатності моделі до узагальнення на незалежних даних (тестова вибірка).

У випадку використання базового вектору ознак усі моделі продемонстрували найгірші показники (табл. 1), зокрема, лінійна модель показала 25 помилок зі значенням середньоквадратичного відхилення (нестабільність) $\pm 0,06$ при проведенні 5ти кратної стратифікованої перехресної валідації. Нелінійні моделі також продемонстрували 6–8 помилок, при цьому плутали стани. Додавання тільки реляційних ознак не значно покращило ситуацію, а отже, компенсація впливу тиску без врахування хаотичної природи сигналу є недостатньою для чіткого розділення класів. Додавання ентропійних ознак разом з реляційними дозволило досягти найкращих результатів в оцінці загальної ефективності нелінійних моделей (MLP та RandomForest), яка сягнула значення 98,6-99,1% з кількістю помилок

Таблиця 1

Результати тестування моделей на навчальній та тестовій вибірках

Класифікатор	Тип	Ентропійні ознаки	Реляційні ознаки	F1-score (перехресна валідація)	F1-score (незалежні дані)	Розрив
RandomForest	Нелінійний (ансамбль)	-	-	0,9725±0,0334, к-сть помилок: 6	0,9431	-4,8%
		-	+	0,9770±0,0206, к-сть помилок: 5		
		+	+	0,9908±0,0113, к-сть помилок: 2		
MLP (неймережа)	Нелінійний	-	-	0,9630±0,0236, к-сть помилок: 8	0,8206	-16,6%
		-	+	0,9630±0,0315, к-сть помилок: 8		
		+	+	0,9862±0,0113, к-сть помилок: 3		
Perceptron	Лінійний	-	-	0,8846±0,0615, к-сть помилок: 25	0,7670	-15,5%
		-	+	0,8521±0,0363, к-сть помилок: 32		
		+	+	0,9216±0,0303, к-сть помилок: 17		

Джерело: складено авторами.

2–3 шт (табл. 1). Ефективність прийнятого рішення щодо формування складу вектора ознак продемонстровано на рис. 3, що підтверджує значення ентропії як ключової ознаки розрізнення станів (перша позиція важливості на рис. 3, б).

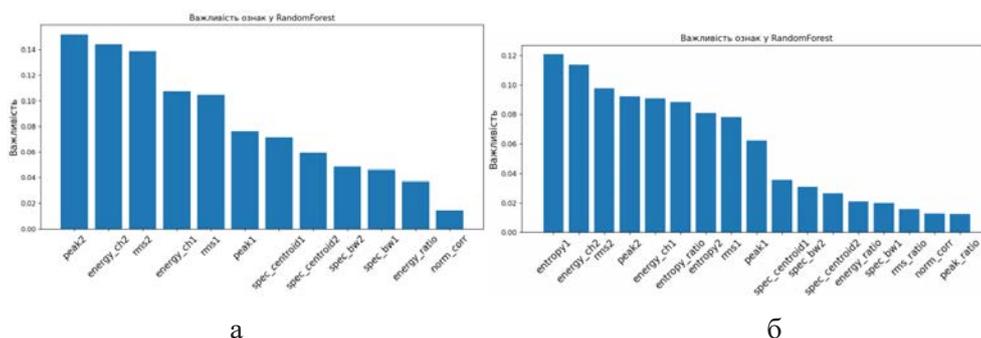


Рис. 3. Оцінка важливості ознак RandomForest для вектора: а) базового; б) повного

Джерело: складено авторами.

Аналіз ефективності застосування лінійного класифікатора (табл. 1) дозволяє стверджувати, що досліджувана задача носить нелінійний характер. Результат класифікації перцептрона за умов повного набору ознак досягає 92,2% при 17ти помилках, що вказує на неможливість лінійного розділення класів у сформованому просторі ознак, тоді як нелінійні класифікатори досягли майже рівнозначного успіху в 98,6–99,1% на етапі навчання завдяки їх можливості моделювати складні границі рішень.

Вирішальним критерієм щодо вибору моделі є результати застосування навчених моделей на тестовій вибірці, що містить дані з невідомими моделям параметрами тиску та характеристиками тестових сигналів. Тестування проведено на наборі, який містив зразки closed та leak3mm. Цей тест імітує реальну ситуацію, коли модель стикається з невизначеними умовами. Результати здатності моделей до узагальнення (табл. 1) свідчать про значне перенавчання неймережевої моделі MLP (розрив -16,6%) та лінійної моделі (розрив -15,5%). Розрив у значенні оцінки загальної ефективності для ансамблевої моделі становить -4,8 %, що підтверджує обґрунтованість її вибору для вирішення поставленої задачі. Детальніший аналіз щодо порівняння ефективності нелінійних моделей при тестуванні можна зробити на основі матриць плутанини (рис. 4).

З рис. 4, а видно, що RandomForest не сплутав стани closed і leak3mm, а отже, його точність $precision=1$. Помилки цієї моделі полягали у відкиданні зразків (4 та 10 помилок зі станами closed та leak3mm відповідно) і класифікуванні їх як інші стани (середнє значення повноти $recall=0,89$). Для моделі MLP (рис. 4, б) результати демонструють сплутування класів між собою ($precision=0,98/0,93$) та відкидання зразків (середнє значення повноти $recall=0,73$).

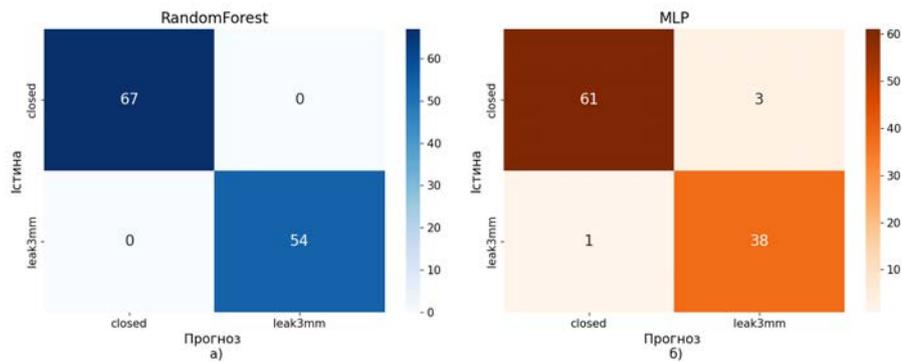


Рис. 4. Матриці плутанини на тестовому наборі для: а) RandomForest; б) MLP

Джерело: складено авторами.

Висновки. В роботі запропоновано підвищення стійкості алгоритму виявлення витоків з трубопроводів з використанням акустичного методу контролю. Доведено ефективність застосування вектора ознак, який включає ентропійні та реляційні ознаки, в поєднанні з ансамблевою моделлю RandomForest при класифікації діагностичних сигналів методу.

Показано, що розширення вектора ознак ентропійними та реляційними параметрами дозволяє сформувати модель ML, яка є стійкою до зміни тиску та характеристик тестового сигналу. На основі тренувальної вибірки одержано підтвердження нелінійності задачі пошуку витоків шляхом порівняння результатів класифікації лінійної моделі Персептрон (ефективність 92.2%) та нелінійних моделей RandomForest та MLP (ефективність 99.1%). На основі тестування розглянутих моделей на незалежних даних визначено перевагу RandomForest у здатності до узагальнення. Наявність в незалежній вибірці діагностичних сигналів, зареєстрованих при інших значеннях тиску ніж в тренувальній вибірці виявило різницю між нелінійними моделями. MLP та Персептрон продемонстрували ознаки перенавчання, тоді як ансамблева модель RandomForest продемонструвала здатність до узагальнення з ефективністю 94% та точністю 1, що доводить її придатність для практичного застосування.

Результати дослідження можуть бути використані при розробленні алгоритмічного забезпечення систем акустичного моніторингу стану трубопроводів, стійких до зміни технологічних умов транспортування.

Список використаних джерел:

1. Заміховський Л. М., Штаер Л. О. Контроль витоків з магістральних та технологічних трубопроводів. Івано-Франківськ: ІФНТУНГ. 2013. 224 с.
2. Марчук Д. К., Граф М. С. Методи оцінки ефективності моделей виявлення об'єктів у комп'ютерному зорі. *Вісник Херсонського національного технічного університету*. 2023. 2(85). P. 181–186. DOI: 10.35546/kntu2078-4481.2023.2.25.
3. Chen R., Wang Q., Javanmardi A. A Review of the Application of Machine Learning for Pipeline Integrity Predictive Analysis in Water Distribution Networks. *Archives of Computational Methods in Engineering*. 2025. Vol. 32, No. 6. P. 3821–3849. DOI: 10.1007/s11831-025-10251-6.
4. El-Zahab S., Abdelkader E.M., Fares A., Zayed T. Comparative Analysis of Machine Learning Techniques in Enhancing Acoustic Noise Loggers' Leak Detection. *Water*. 2025. Vol. 17, No. 16. P. 2427. DOI: 10.3390/w17162427.
5. Hong Z., Zhao D., Dong L., Liu S., Qiu F., Jin Y. Improving Pipeline Leak Detection Accuracy by Constructing a Multistate Leakage Sounds Training Set. *Journal of Pipeline Systems Engineering and Practice*. 2025. Vol. 16, No. 4. P. 04025056. DOI: 10.1061/JPSEA2.PSENG-1864.
6. Li Y., Wang X., Wu J. Fault Diagnosis of Rolling Bearing Based on Permutation Entropy and Extreme Learning Machine. *Chinese Control and Decision Conference (CCDC) : Proceedings of the Conference, 28–30 May 2016 Yinchuan, China*. P. 2966–2971. DOI: 10.1109/CCDC.2016.7531490.
7. Liang H., Gao Y., Li H., Huang S., Chen M., Wang B. Pipeline Leakage Detection Based on Secondary Phase Transform Cross-Correlation. *Sensors*. 2023. Vol. 23, No. 3. P. 1572. DOI: 10.3390/s23031572.
8. Rybitskyi I., Voitenko S., Karpash O., Krytyn Yu. Mathematical Modelling of the Ability of Detect and Measure Small Leaks in Pipeline Systems. *European Open Science Space*. 2025. P. 132–137. DOI: 10.70286/eoss-22.09.2025.004.
9. Ullah N., Ahmed Z., Kim J.-M. Pipeline Leakage Detection Using Acoustic Emission and Machine Learning Algorithms. *Sensors*. 2023. Vol. 23, No. 6. P. 3226. DOI: 10.3390/s23063226.
10. Yang D., Lu J., Dong H., Hu Z. Pipeline signal feature extraction method based on multi-feature entropy fusion and local linear embedding. *Systems Science & Control Engineering*. 2022. Vol. 10, No. 1. P. 407–416. DOI: 10.1080/21642583.2022.2063202.
11. Zhang B., Li W., Li X.-L., Ng S.-K. Intelligent Fault Diagnosis Under Varying Working Conditions Based on Domain Adaptive Convolutional Neural Networks. *IEEE Access*. 2018. Vol. 6. P. 66367–66384. DOI: 10.1109/ACCESS.2018.2878491.

Дата надходження статті: 23.11.2025

Дата прийняття статті: 10.12.2025

Опубліковано: 30.12.2025