УДК 004.932+004.85 DOI: 10.31673/2412-4338.2025.038710

Бондарчук Андрій Петрович
Київський столичний університет імені Бориса Грінченка, м. Київ, Україна
ORCID 0000-0001-5124-5102

Мельник Ірина Юріївна
Київський столичний університет імені Бориса Грінченка, м. Київ, Україна
ORCID : 0000-0001-6041-6145

Суханевич Євгеній Іванович
Державний університет інформаційно-комунікаційних технологій, Київ
ORCID 0009-0003-6631-829X

Абрамов Вадим Олексійович
Київський столичний університет імені Бориса Грінченка
ORCID 0000-0002-8026-1475

ПІДВИЩЕННЯ ЕФЕКТИВНОСТІ СИСТЕМ ВІДЕОСПОСТЕРЕЖЕННЯ ЗА РАХУНОК ГІБРИДНОГО МЕТОДУ ВІДБОРУ КЛЮЧОВИХ КАДРІВ І ІНТЕРПРЕТАЦІЇ РІШЕНЬ

Анотація. *Сучасні системи відеоспостереження генерують значні обсяги даних, що ускладнює їх ефективний аналіз у реальному часі. Існуючі методи автоматичного виявлення аномалій часто є обчислювально-вимогливими та працюють за принципом "чорного ящика", що обмежує їх практичне застосування у критично важливих сферах, таких як охорона громадського порядку та публічна безпека. У цій статті запропоновано новий комбінований підхід, що вирішує дві ключові проблеми: неефективну обробку надлишкових даних та недостатню прозорість роботи алгоритмів штучного інтелекту. Метидика базується на поєднанні інноваційного методу відбору інформативних кадрів з подальшою їх обробкою інтерпретованою моделлю детекції. Перший етап передбачає оптимізацію вибору ознак за допомогою гібридного алгоритму, що поєднує згорткову нейромережу InceptionV3 з генетичним алгоритмом, що дозволяє зменшити обсяг даних на 70-85% зі збереженням показника повноти на рівні 98%. Другий етап забезпечує не лише класифікацію аномалій, але й генерацію зрозумілих для людини пояснень шляхом інтеграції методів пояснюваного ШІ (XAI), зокрема Grad-CAM та керованого зворотного поширення. Експериментальна перевірка на стандартних наборах даних демонструє переваги запропонованого підходу порівняно з сучасними аналогами. Отримані результати свідчать про можливість покращення точності класифікації на 3-5% при одночасному зменшенні обчислювального навантаження. Крім того, система надає візуальні пояснення прийнятих рішень у вигляді теплових карт, що підвищує довіру до її роботи. Запропонований підхід відкриває перспективи для впровадження ефективних систем відеомоніторингу у реальному часі з функцією обґрунтування прийнятих рішень.*

Ключові слова: *відеоспостереження, інформаційні системи, генетичний алгоритм, штучний інтелект, моделювання, згорткові нейронні мережі, обробка відеоданих, комп'ютерний зір.*

Bondarchuk Andrii
Borys Grinchenko Kyiv Metropolitan University, Kyiv, Ukraine
ORCID: 0000-0001-5124-5102

Melnyk Iryna
Borys Grinchenko Kyiv Metropolitan University, Kyiv, Ukraine
ORCID: 0000-0001-6041-6145

Sukhanevich Evhenii
State University of Information and Communication Technologies, Kyiv, Ukraine
ORCID: 0009-0003-6631-829X

Abramov Vadym
Borys Grinchenko Kyiv Metropolitan University
ORCID 0000-0002-8026-1475

ENHANCING THE EFFICIENCY OF VIDEO SURVEILLANCE SYSTEMS THROUGH A HYBRID METHOD OF KEY FRAME SELECTION AND DECISION INTERPRETATION

Abstract. *Modern video surveillance systems generate significant volumes of data, complicating their effective real-time analysis. Existing automated anomaly detection methods are often computationally intensive and operate as "black boxes," limiting their practical application in critical areas such as public order maintenance and public safety. This article proposes a novel combined approach that addresses two key problems: inefficient processing of redundant data and insufficient transparency of artificial intelligence algorithms. The methodology is based on combining an innovative method for selecting informative frames with their subsequent processing by an interpretable detection model. The first stage involves optimizing feature selection using a hybrid algorithm that combines the InceptionV3 convolutional neural network with a genetic algorithm, enabling a 70-85% reduction in data volume while maintaining a 98% recall rate. The second stage ensures not only anomaly classification but also the generation of human-understandable explanations through the integration of explainable AI (XAI) methods, particularly Grad-CAM and guided backpropagation. Experimental validation on standard datasets demonstrates the advantages of the proposed approach compared to contemporary alternatives. The obtained results indicate a 3-5% improvement in classification accuracy while simultaneously reducing computational load. Furthermore, the system provides visual explanations of decisions in the form of heatmaps, enhancing trust in its operation. The proposed approach opens prospects for implementing efficient real-time video monitoring systems with decision justification capabilities.*

Keywords: *video surveillance, information systems, genetic algorithm, artificial intelligence, modeling, convolutional neural networks, video data processing, computer vision.*

1. Вступ.

Сучасні системи відеоспостереження генерують колосальні обсяги даних, аналіз яких вручну є практично неможливим. Автоматизоване виявлення аномальних подій, зокрема актів насильства, за допомогою методів штучного інтелекту стикається з двома ключовими проблемами: неефективністю обробки непотрібних даних та "чарівністю" у прийнятті рішень глибокими нейромережами, коли людина-оператор не може зрозуміти логіку машини. Це обмежує довіру до системи та ускладнює її впровадження у критично важливих сферах, таких як охорона громадського порядку та безпека на об'єктах.

2. Аналіз останніх досліджень і публікацій.

Аналіз сучасних досліджень з підвищення ефективності систем відеоспостереження показує значний прогрес у напрямку підвищення точності детекції, однак залишаються невирішеними ключові проблеми обчислювальної ефективності та інтерпретованості результатів. Великий крок у вирішенні даної проблеми зроблено у роботі Salman, et al. [1]. Дослідження останніх років можна умовно поділити на три основні категорії. Перша категорія це методи на основі глибокого навчання. Роботи [2-4] демонструють високу ефективність архітектур на основі 3D-згорток та трансформерів у завданнях виявлення аномалій. Однак ці підходи вимагають обробки повного відеопотоку, що призводить до надмірних обчислювальних витрат. Моделі часто працюють як "чорні ящики", що ускладнює їх застосування в критично важливих системах, де необхідна зрозумілість прийняття рішень. Друга категорія це підходи до відбору ключових кадрів. В роботах [5-6] пропонують методи обробки та стиснення відеоданих зокрема на основі аналізу руху. Незважаючи на зменшення обсягу даних, ці методи часто втрачають важливі кадри з раптовими аномаліями, що не пов'язані з рухом. В інших роботах пропонується впроваджують методи машинного навчання для відбору кадрів, але вони не враховують оптимізацію ознак на рівні окремих кадрів, що обмежує їх ефективність. Третя категорія обробка масивних потоків даних, забезпечення реального часу та відповідності вимогам прозорості рішень. Це питання розглянуто у роботах [6-7] Однак ці рішення часто не враховують оптимізацію на рівні окремих вузлів, що призводить до неефективного використання ресурсів.

3. Мета і задачі дослідження.

Метою є розробка комплексного підходу для автоматизованого виявлення аномалій у відеозаписах шляхом поєднання ефективного методу відбору ключових кадрів з інтерпретованою моделлю глибокого навчання (XAI-Inv3), спрямованого на подолання обмежень сучасних систем відеоаналітики щодо обчислювальної неефективності та недостатньої прозорості прийняття рішень.

Завдання дослідження:

Провести огляд сучасних підходів до виявлення аномалій у відео, виявити їхні недоліки, зокрема високі обчислювальні витрати на обробку надлишкових даних.

Визначити та реалізувати функцію, яка забезпечує баланс між точністю класифікації та кількістю обраних ознак.

Запропонувати механізм порівняння кадрів на основі розрахунку евклідової відстані між оптимізованими векторами ознак та адаптивного визначення порогу для ідентифікації ключових кадрів.

4. Результати дослідження.

Для подолання цих обмежень пропонується методика, яка опирається на використання інтелектуального фільтра, відібраючи з відеопотоку лише найважливіші кадри, що містять потенційні аномалії та інструменти аналізу відібраних кадрів як для класифікації подій, так і для надання зрозумілих для людини пояснень щодо прийнятого рішення. Такий підхід дозволяє значно зменшити обчислювальне навантаження та підвищити прозорість роботи системи.

Дана методика поєднує потужність згорткової нейромережі InceptionV3 для витягування високорівневих ознак із кожного кадру та генетичного алгоритму для оптимізації вибору найбільш релевантних ознак. Генетичний алгоритм ітеративно еволюціонує набір ознак, максимізуючи точність при мінімізації їх кількості. Фінальним етапом є розрахунок евклідової відстані між послідовними кадрами на основі вибраних ознак та відбір тих, чия відстань перевищує динамічний поріг, що вказує на значну зміну у сцені.

Модель системи побудована на архітектурі InceptionV3, але ключовою є її модифікація для забезпечення інтерпретованості (eXplainable AI, XAI). Вона інтегрує методи на основі градієнтів, такі як **Grad-CAM** (Gradient-weighted Class Activation Mapping) та **кероване зворотне поширення**. Ці методи дозволяють візуалізувати у вигляді "теплових карт" ті області зображення, які найбільше вплинули на прогноз моделі (наприклад, кулак, зброя, хаотичний рух), надаючи оператору наочне і зрозуміле пояснення, чому подія була класифікована як аномальна.

Генетичний алгоритм у даній методиці починає роботу зі створення початкової популяції. Кожен індивідуум у цій популяції, який називається "хромосомою", є двійковим вектором. Довжина вектора відповідає повній кількості ознак, витягнутих згортковою нейромережею InceptionV3 з кадру. Кожен біт у хромосомі (ген) вказує на включення (1) або виключення (0) конкретної ознаки з подальшого аналізу. Таким чином, хромосома представляє собою компактно закодований підмножина ознак-кандидатів.

Кожна хромосома оцінюється за допомогою функції придатності. Ця функція є найважливішим компонентом, оскільки керує напрямком еволюції. Вона не просто максимізує точність класифікації, але й містить штрафний коефіцієнт (λ) за надто велику кількість обраних ознак. Математично це виражається як

$$F\ (c) = A\ (c) - \lambda * D\ (c)$$

де $A\ (c)$ - це точність класифікації, яку демонструє модель (наприклад, XAI-Inv3), коли в якості вхідних даних використовується не повний набір ознак, а лише підмножина, закодована в хромосомі, вимірює, наскільки добре модель може виявляти аномалії, використовуючи лише ті ознаки, для яких у хромосомі c встановлено значення 1; $D\ (c)$ - це загальна кількість ознак, які вибрані (активовані) в хромосомі c. Це кількість одиниць (1) у двійковому векторі хромосоми. Наприклад, якщо хромосома має вигляд [1, 0, 0, 1, 1, 0], то = 3.

Це забезпечує пошук компромісу між високою точністю та ефективністю. На основі отриманих оцінок придатності відбувається селекція: хромосоми з вищим значенням мають більшу ймовірність бути обраними для "розмноження" за допомогою механізму "рулетки" або турнірного відбору.

Обрані "батьківські" хромосоми піддаються генетичним операторам. Оператор схрещування (кросовер) обмінюється частинами між двома хромосомами, створюючи нове "потомство" з комбінацією ознак батьків. Оператор мутації випадковим чином змінює окремі біти в хромосомах (з малою ймовірністю), запроваджуючи в популяцію нову генетичну інформацію та запобігаючи застряганню в локальних оптимумах. Цей цикл (оцінка, селекція, схрещування, мутація) повторюється протягом заданої кількості поколінь, поки не буде знайдено субоптимальний набір ознак, що максимізує значення функції придатності.

Після того як генетичний алгоритм для кожного кадру визначив оптимальний підмножина ознак, наступним кроком є кількісне вимірювання подібності між послідовними кадрами. Для цього використовується евклідова відстань. Нехай два вектори оптимізованих ознак, визначених генетичним алгоритмом для кадрів i та j, позначаються як:

$$F_i = \left(f_i^{(1)}, f_i^{(2)}, f_i^{(d)}\right), F_j = \left(f_j^{(1)}, f_j^{(2)}, f_j^{(d)}\right),$$

де d — кількість ознак, обраних у хромосомі.

Евклідова відстань D_{ij} між цими векторами обчислюється за формулою:

$$D_{ij} = \sqrt{\sum_{k=1}^{d} (f_i^{(k)} - f_j^{(k)})^2}.$$

де $f_i^{(k)}$ та $f_j^{(k)}$— значення k -ї ознаки для кадрів i та j відповідно.

Для кожної ознаки k обчислюється різниця$(f_i^{(k)} - f_j^{(k)})$, яка потім підноситься до квадрату, щоб усунути негативні значення та підсилити великі розбіжності.

Критичним аспектом є визначення значення порогу, яке би розділяло кадри на ключові та надлишкові. Жорстко заданий поріг неефективний через різноманіття умов зйомки. Ми можемо використати адаптивний підхід: поріг обчислюється динамічно на основі статистики перших N кадрів відео (наприклад, 1000). Розраховується середня евклідова відстань для цієї вибірки, а потім порогове значення встановлюється як деяке кратне цій середній величині. Це дозволяє системі автоматично підлаштовуватися під конкретне відео.

Фінальний процес класифікації є порівняно простим, але ефективним. Для кожної пари послідовних кадрів обчислюється евклідова відстань між їхніми оптимізованими векторами ознак. Якщо ця відстань перевищує обчислений динамічний поріг, то другий кадр у парі позначається як ключовий. Всі кадри, відстань для яких знаходиться нижче за поріг, вважаються надлишковими і відкидаються. Це радикально зменшує обсяг даних, що надходять на вхід моделі, при збереженні інформації про всі важливі події.

Традиційні глибокі нейромережі, такі як стандартний InceptionV3, демонструють високу точність, але працюють за принципом "чорного ящика". Це означає, що ми бачимо лише вхідні дані (кадр) і результат (наприклад, "аномалія"), але внутрішній процес прийняття рішення залишається непрозорим. Для критично важливих застосувань, таких як безпека, це неприйнятно, оскільки оператор не може довіряти системі, яка не пояснює своїх висновків і може бути схильною до помилок через незбалансовані дані чи артефакти.

Завдання моделі полягає не лише у класифікації кадру, але й у генерації зрозумілих для людини пояснень: які саме області зображення, які візуальні ознаки найбільше вплинули на прийняте рішення. Це перетворює модель з "чорного ящика" на "скляний" або "прозорий ящик".

Інтерпретованість є критично важливою для побудови довіри та практичного впровадження. Коли система виявляє аномалію, оператор отримує не просто сигнал тривоги, а наочне підтвердження: теплову карту, яка виділяє область конфлікту, підозрілий предмет або незвичну дію. Це дозволяє оператору швидко перевірити обґрунтованість тривоги та прийняти рішення, уникаючи помилкових спрацьовувань. Таким чином, формується міст між високою точністю алгоритму та людським розумінням.

Grad-CAM — це один з ключових методів у XAI-Inv3. Він працює на основі аналізу градієнтів. Коли мережа робить прогноз (наприклад, "бійка"), Grad-CAM обчислює, які саме пікселі в останньому згортковому шарі найбільше "відповідальні" за цей прогноз. Технічно, метод обчислює зважену суму карт ознак останнього згорткового шару, де ваги визначаються градієнтом цільового класу щодо цих карт ознак. Результатом є теплова карта (heatmap) — напівпрозорий накладка на оригінальне зображення, де "гарячі" кольори (червоний, жовтий) вказують на найважливіші області.

Метод керованого зворотного поширення (Guided Backpropagation) доповнює Grad-CAM. Він більш детально підходить до питання інтерпретації. Цей метод також аналізує градієнти, але просувається назад по мережі аж до вхідного піксельного рівня. Його ключова особливість — фільтрація: він зберігає лише ті градієнти, які мають позитивний вплив на прогнозований клас, і відсікає негативні. В результаті генерується чітке, високодеталізоване зображення, де яскраво виділяються конкретні контури та текстури (наприклад, силует зброї, контури людей у бійці), що сприяли класифікації.

Разом ці два методи забезпечують потужний та багаторівневий інструмент інтерпретації. Grad-CAM дає загальне уявлення про локалізацію події, показуючи "де" сталася аномалія (наприклад, група людей у лівому кутку кадру). Кероване зворотне поширення деталізує це уявлення, показуючи "що саме" в цій області привернуло увагу мережі (окремі предмети). Така комбінація забезпечує максимально повне і зрозуміле пояснення рішення моделі.

XAI-Inv3 фундаментально базується на перевіреній часом архітектурі InceptionV3. Це означає, що вона зберігає всі її переваги: модулі Inception для ефективного витягування ознак на різних масштабах, техніки зменшення розмірності для боротьби з перенавчанням та допоміжні класифікатори

для поліпшення збіжності під час навчання глибокої мережі. Ця базова структура забезпечує високу продуктивність у завданні класифікації зображень, що є фундаментом для всієї системи.

5. Висновки та перспективи подальших досліджень. Можна зробити висновок, що запропонований підхід є високоефективний механізмом попередньої обробки, який усуває інформаційний шум і радикально знижує обчислювальні витрати. Це дозволяє XAI-Inv3 працювати лише з релевантними даними, що підвищує як швидкість, так і, потенційно, точність класифікації завдяки зосередженості на значущих подіях. Разом вони формують цілісний конвеєр, оптимізований як для швидкодії, так і для якості аналізу.

Дане дослідження є корисним при розробці систем безпеки та автоматизованого відеомоніторингу, проектуванні розумних міст. Забезпечуючи високу точність, ефективність та, що найважливіше, зрозумілість роботи алгоритмів, запропонована методика закладає основу для більш тісної та ефективної взаємодії людини та штучного інтелекту в критично важливих сферах. Майбутні дослідження можуть бути спрямовані на адаптацію методики для роботи в реальному часі та впровадження в embedded-системи.

Список використаної літератури

1. Salman, M., Abbas, N., Rahman, S. I. U., Rehman, A., Alamri, F. S., Elyassih, A., & Saba, T. (2025). Enhancing surveillance anomaly detection with keyframes and explainable inception model. Egyptian Informatics Journal, 31, 100769.

2. Masud U, Sadiq M, Masood S, Ahmad M, Abd El-Latif AA. LW-DeepFakeNet: a lightweight time distributed CNN-LSTM network for real-time DeepFake video detection. SIViP 2023;17(8):4029–37.

3. Theobald O. Machine learning: make your recommender system; build your recommender system with machine learning insights. Packt Publishing Ltd 2024

4. Bensakhria, Ayoub. “Leveraging Real-time Edge AI-Video Analytics to Detect and Prevent Threats in Sensitive Environments.” (2023).

5. Shen G, Ouyang Y, Sanchez V. Video anomaly detection via prediction network with enhanced spatio-temporal memory exchange. In: CASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP); 2022. p. 3728–32

6. Bondarchuk, A., Dibrivniy, O., Grebenyk, V., & Onyshchenko, V. (2021, October). Motion Vector Search Algorithm for Motion Compensation in Video Encoding. In 2021 IEEE 8th International Conference on Problems of Infocommunications, Science and Technology (PIC S&T) (pp. 345-348). IEEE.

7. Shevchenko, O., Bondarchuk, A., Polonevych, O., Zhurakovskyi, B., & Korshun, N. (2021). Methods of the objects identification and recognition research in the networks with the IoT concept support. Cybersecurity Providing in Information and Telecommunication Systems 2021, 2963, 277-282.

8. Skladannyi, P., Kostiuk, Y., Rzaieva, S., Bebeshko, B., & Korshun, N. (2025). Adaptive Methods for Embedding Digital Watermarks to Protect Audio and Video Images in Information and Communication Systems.

9. Kostiuk, Y., Skladannyi, P., Sokolov, V., & Vorokhob, M. (2025). Models and technologies of cognitive agents for decision-making with integration of Artificial Intelligence. Modern Data Science Technologies Doctoral Consortium 2025, 4005, 82-96.

10. Kostiuk, Y., Skladannyi, P., Khorolska, K., Sokolov, V., & Hulak, H. (2025). Application of Statistical and Neural Network Algorithms in Steganographic Synthesis and Analysis of Hidden Information in Audio and Graphic Files.

11. Anakhov, P., Zhebka, V., Bondarchuk, A., Storchak, K., & Sablina, M. (2023). Increasing the Reliability of a Heterogeneous Network using Redundant Means and Determining the Statistical Channel Availability Factor. Cybersecurity Providing in Information and Telecommunication Systems, 231-236.

12. Костюк, Ю., Бебешко, Б., Гулак, Г., Складанний, П., Рзаєва, С., & Хорольська, К. (2024). Забезпечення кібербезпеки та швидкодії передачі даних у безпровідних мережах. Безпека інформації, 30(3), 365–375. https://doi.org/10.18372/2225-5036.30.20357

13. Крючкова, Л., Башкевич, Є., & Куцибала, М. (2024). Програмна реалізація алгоритмів виокремлення контурів зображень об'єктів в інтелектуальних системах відеоспостереження. Телекомунікаційні та інформаційні технології, 2(83), 14–29. https://doi.org/10.31673/2412-4338.2024.022439

14. V. Sokolov, P. Skladannyi, A. Platonenko, Video Channel Suppression Method of Unmanned Aerial Vehicles, in: IEEE 41st International Conference on Electronics and Nanotech-nology (2022) 473–477. doi: 10.1109/ELNANO54667.2022.9927105

15. Y. Kostiuk, et al., Integrated protection strategies and adaptive resource distribution for secure video streaming over a Bluetooth network, in: Cybersecurity Providing in Information and Telecommunication Systems II, vol. 3826 (2024) 129–138.