

Харківський національний університет радіоелектроніки

Факультет Комп'ютерних наук
(повна назва)
Кафедра Штучного інтелекту
(повна назва)
Рівень вищої освіти другий (магістерський)
Спеціальність 122 Комп'ютерні науки
(код і повна назва)
Тип програми освітньо-професійна
(освітньо-професійна або освітньо-наукова)
Освітня програма Науки про дані (Data Science)
(повна назва)

ЗАТВЕРДЖУЮ:

Зав. кафедри _____

(підпис)

«_____» _____ 20__ р.

ЗАВДАННЯ
НА КВАЛІФІКАЦІЙНУ РОБОТУ

здобувачеві Череднікову Миколі Володимировичу
(прізвище, ім'я, по батькові)

1. Тема роботи Інтелектуальна система відеомоніторингу для виявлення аномальної поведінки об'єктів у реальному часі

затверджена наказом університету від 24 листопада 2025 р. по. 1057Ст

2. Термін подання студентом роботи до екзаменаційної комісії 16 грудня 2025 р.

3. Вихідні дані публічні датасети відеопослідовностей з анотованими аномаліями для навчання та тестування моделей; наукові публікації та дослідження з глибокого навчання для аналізу відео та виявлення аномалій; архітектури згорткових нейронних мереж для виділення просторових ознак; механізми уваги та трансформери для темпорального моделювання; бібліотеки машинного навчання та фреймворки для розробки; стандарти та регуляторні документи з приватності персональних даних та етичного використання систем штучного інтелекту; технічна документація систем відеоспостереження; методології оцінки продуктивності алгоритмів комп'ютерного зору.

4. Перелік питань, що потрібно опрацювати в роботі _____

1) Аналіз предметної галузі

2) Методи та методика дослідження

3) Проектування системи інтелектуального відеомоніторингу

КАЛЕНДАРНИЙ ПЛАН

№	Назва етапів роботи	Терміни виконання етапів роботи	Примітка
1	Отримання завдання на кваліфікаційну роботу	24.11.2025	виконано
2	Аналіз основних понять та концепцій	25.11.2025	виконано
3	Аналіз методологічних підходів	25.11.2025	виконано
4	Аналіз архітектур глибокого навчання	26.11.2025	виконано
5	Аналіз перефрійних обчислень	26.11.2025	виконано
6	Аналіз адаптивних систем	27.11.2025	виконано
7	Аналіз актуальності розвитку	27.11.2025	виконано
8	Обґрунтування вибору напрямку дослідження	28.11.2025	виконано
9	Порівняльний аналіз методів виявлення аномалій	01.11.2025	виконано
10	Обґрунтування вибору методів для розробки	02.11.2025	виконано
11	Обґрунтування вибору методики розробки	03.11.2025	виконано
12	Обґрунтування концепції та вимоги до системи	04.11.2025	виконано
13	Архітектура системи та її основні компоненти	05.12.2025	виконано
14	Обґрунтування методів навчання	10.12.2025	виконано
15	Захист кваліфікаційної роботи	16.12.2025	виконано

Дата видачі завдання 24 листопада 2025 р.

Здобувач Череди
(підпис)

Керівник роботи _____ доц. Дейнеко А.О.
(підпис) (посада, прізвище, ініціали)

РЕФЕРАТ

Пояснювальна записка: 64 с., 6 рис., 1 дод., 32 джерела.

ВИЯВЛЕННЯ АНОМАЛІЙ, ВІДЕОМОНІТОРИНГ, ГЛИБОКЕ НАВЧАННЯ, ЗГОРТКОВІ НЕЙРОННІ МЕРЕЖІ, МЕХАНІЗМИ УВАГИ, КОМП'ЮТЕРНИЙ ЗІР, ОБРОБКА ВІДЕО В РЕАЛЬНОМУ ЧАСІ.

Об'єкт дослідження – процеси автоматизованого аналізу відеопотоків для виявлення відхилень від типових патернів поведінки об'єктів у системах відеомоніторингу.

Предмет дослідження – методи та алгоритми глибокого навчання для виявлення аномальної поведінки об'єктів у відеопотоках реального часу, включаючи архітектури нейронних мереж для просторово-часового аналізу, механізми виявлення відхилень від нормальної поведінки та стратегії забезпечення обчислювальної ефективності.

Мета роботи – проектування інтелектуальної системи відеомоніторингу для виявлення аномальної поведінки об'єктів у реальному часі, яка забезпечує оптимальний баланс між точністю детекції, швидкістю обробки та здатністю виявляти як відомі, так і невідомі типи аномалій.

Методи дослідження – методи глибокого навчання, зокрема згорткові нейронні мережі для виділення просторових ознак, механізми уваги для темпорального моделювання, а також методи компресії моделей та ієрархічної обробки для забезпечення роботи в реальному часі на обладнанні з обмеженими ресурсами.

ABSTRACT

Master's thesis contains: 64 pp., 6 fig., 1 ann., 32 references.

ANOMALY DETECTION, ATTENTION MECHANISMS, COMPUTER VISION, CONVOLUTIONAL NEURAL NETWORKS, DEEP LEARNING, REAL-TIME VIDEO PROCESSING, VIDEO MONITORING.

Research object – the processes of automated analysis of video streams for detecting deviations from typical behavior patterns of objects in video surveillance systems.

Research subject – methods and algorithms of deep learning for detecting anomalous behavior of objects in real-time video streams, including neural network architectures for spatio-temporal analysis, mechanisms for detecting deviations from normal behavior, and strategies for ensuring computational efficiency.

Aim of research – to design an intelligent video surveillance system for real-time detection of anomalous object behavior that provides an optimal balance between detection accuracy, processing speed, and the ability to identify both known and unknown types of anomalies.

Method of research – deep learning methods, in particular convolutional neural networks for extracting spatial features, attention mechanisms for temporal modeling, as well as model compression techniques and hierarchical processing to ensure real-time operation on hardware with limited computational resources.

ЗМІСТ

Перелік умовних позначень, символів, одиниць, скорочень і термінів	7
Вступ.....	8
1 Аналіз предметної галузі	9
1.1 Аналіз основних понять та концепцій	9
1.2 Аналіз методологічних підходів.....	10
1.3 Аналіз архітектур глибинного навчання	10
1.4 Аналіз перефрійних обчислень та адаптивних систем	13
1.5 Аналіз актуальності розвитку	16
2 Методи та методика дослідження	19
2.1 Обґрунтування вибору напрямку дослідження	19
2.2 Порівняльний аналіз методів виявлення аномалій у відеопотоках ...	25
2.3 Обґрунтування вибору методів для розробки системи.....	33
2.4 Методика розробки та інтеграції системи відеомоніторингу	39
3 Проектування системи інтелектуального відеомоніторингу.....	45
3.1 Обґрунтування концепції та вимоги до системи	45
3.2 Архітектура системи та її основні компоненти	46
3.3 Методи навчання та виявлення аномалій	51
Висновки	58
Перелік джерел посилання	59
Додаток А Відомість кваліфікаційної роботи	64

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СИМВОЛІВ, ОДИНИЦЬ, СКОРОЧЕНЬ І ТЕРМІНІВ

ЦП – центральний процесор;

AI – Artificial Intelligence – штучний інтелект;

API – Application Programming Interface – програмний інтерфейс застосунків;

CNN – Convolutional Neural Network – згорткова нейронна мережа;

GAN – Generative Adversarial Network – генеративно-змагальна мережа;

GDPR – General Data Protection Regulation – загальний регламент про захист даних;

GPU – Graphics Processing Unit – графічний процесор;

HDR – High Dynamic Range – високий динамічний діапазон;

HTTP – Hypertext Transfer Protocol – протокол передачі гіпертексту;

LSTM – Long Short-Term Memory – довга короткочасна пам'ять;

REST – Representational State Transfer – передача репрезентативного стану;

RGB – Red, Green, Blue – червоний, зелений, синій;

SDK – Software Development Kit – набір засобів розробки програмного забезпечення;

SMS – Short Message Service – служба коротких повідомлень;

SVDD – Support Vector Data Description – опис опорних векторів даних;

TLS – Transport Layer Security – безпека транспортного рівня;

YOLO – You Only Look Once – архітектура детектора об'єктів.

ВСТУП

У сучасному світі забезпечення безпеки в публічних місцях, на критичній інфраструктурі та промислових об'єктах набуває все більшого значення. Традиційні системи відеоспостереження, які потребують постійного моніторингу людиною-оператором, не здатні ефективно обробляти величезні обсяги відеоданих від множинних камер та своєчасно виявляти критичні події. Інтелектуальні системи відеомоніторингу на основі методів машинного навчання та комп'ютерного зору здатні автоматично аналізувати відеопотоки та виявляти аномальну поведінку об'єктів у реальному часі, що робить їх критично важливими для превенції інцидентів та оперативного реагування на загрози.

Об'єком дослідження є процеси автоматизованого аналізу відеопотоків для виявлення відхилень від типових патернів поведінки об'єктів у системах відеомоніторингу.

Предметом дослідження є методи та алгоритми глибокого навчання для виявлення аномальної поведінки об'єктів у відеопотоках реального часу, включаючи архітектури нейронних мереж для просторово-часового аналізу, механізми виявлення відхилень від нормальної поведінки та стратегії забезпечення обчислювальної ефективності.

Мета роботи це проектування інтелектуальної системи відеомоніторингу для виявлення аномальної поведінки об'єктів у реальному часі, яка забезпечує оптимальний баланс між точністю детекції, швидкістю обробки та здатністю виявляти як відомі, так і невідомі типи аномалій.

Методи дослідження: у роботі використовуються методи глибокого навчання, зокрема згорткові нейронні мережі для виділення просторових ознак, механізми уваги для темпорального моделювання, а також методи компресії моделей та ієрархічної обробки для забезпечення роботи в реальному часі на обладнанні з обмеженими ресурсами.

1 АНАЛІЗ ПРЕДМЕТНОЇ ГАЛУЗІ

1.1 Аналіз основних понять та концепцій

Інтелектуальні системи відеомоніторингу являють собою інтегровані програмно-апаратні комплекси, що поєднують традиційні засоби відеоспостереження з методами машинного навчання та комп'ютерного зору для автоматичного аналізу візуальної інформації. Еволюція систем відеоспостереження пройшла кілька етапів від простих аналогових камер до цифрових систем запису, а потім до інтелектуальних систем з елементами штучного інтелекту. На відміну від класичних систем відеоспостереження, які лише фіксують відеопотік для подальшого перегляду людиною-оператором, інтелектуальні системи здатні самостійно виявляти, класифікувати та інтерпретувати події, що відбуваються в зоні спостереження, надаючи оператору лише релевантну інформацію про потенційні загрози або аномалії [1]. Така автоматизація критично важлива в контексті сучасної безпеки, де один оператор може бути відповідальним за моніторинг десятків або навіть сотень камер одночасно, що робить неможливим ефективно виявлення всіх інцидентів без автоматизованої підтримки.

Концепція аномальної поведінки в контексті відеомоніторингу визначається як будь-яка діяльність або послідовність дій, що відхиляється від очікуваних або типових патернів поведінки в конкретному середовищі [2]. Важливо розуміти, що поняття аномалії є контекстно-залежним та може значно варіюватися залежно від специфіки середовища спостереження. Аномальна поведінка може включати несанкціоноване проникнення на територію, агресивні дії та насильство між людьми, залишення підозрілих об'єктів у публічних місцях, падіння людини та тривале перебування в горизонтальному положенні, незвичайні рухи транспортних засобів проти напрямку руху, паніку в натовпі, вандалізм та

пошкодження майна, крадіжки в роздрібних магазинах, порушення протоколів безпеки на виробництві тощо. Виявлення таких аномалій у реальному часі є критично важливим для забезпечення безпеки в публічних місцях, транспортних вузлах, промислових об'єктах, навчальних закладах та інших сферах, де безпека людей та майна є пріоритетом [3], [4].

1.2 Аналіз методологічних підходів

Перший систематичний підхід базується на аналізі траєкторій переміщення об'єктів у просторі та включає алгоритми відстеження об'єктів з використанням фільтрів Калмана для прогнозування майбутніх позицій об'єктів та *particle filters* для відстеження множинних гіпотез[5]. Ці методи спочатку виділяють об'єкти інтересу з відеопотоку за допомогою детекторів об'єктів, потім відстежують їх рух у часі, будуючи траєкторії переміщення. Аналіз швидкості та прискорення руху дозволяє виявити аномальні зміни динаміки, такі як раптові зупинки, різкі зміни напрямку або незвичайно висока швидкість руху. Виявлення нетипових траєкторій за допомогою кластеризації та моделювання нормальної поведінки працює шляхом збору великої кількості типових траєкторій протягом певного періоду навчання, кластеризації їх у групи схожих патернів та побудови моделі нормального руху [6]. Траєкторії, які значно відхиляються від всіх відомих кластерів, позначаються як аномальні. Перевагою методів на основі траєкторій є їх відносна простота та інтерпретованість, однак вони мають обмеження при роботі з густими натовпами, де відстеження окремих об'єктів стає проблематичним.

1.3 Аналіз архітектур глибинного навчання

Революційний вплив на галузь справили методи глибокого навчання, які почали активно застосовуватися для аналізу відео з середини 2010-х

років. Згорткові нейронні мережі стали фундаментальним інструментом для виділення просторових ознак з відеокадрів. Архітектури на кшталт VGGNet, ResNet, Inception та їх модифікації дозволяють автоматично навчатися ієрархічним представленням зображень, виділяючи низькорівневі ознаки на початкових шарах та високорівневі семантичні концепції на глибших шарах [7], [8], [9]. Рекурентні нейронні мережі та їх розвинута форма LSTM використовуються для моделювання темпоральних залежностей у послідовностях відеокадрів [10], [11]. Ці архітектури здатні запам'ятовувати довгострокові залежності та виявляти темпоральні патерни поведінки, що розгортаються протягом кількох секунд або навіть хвилин.

Тривимірні згорткові нейронні мережі та двопотокові архітектури дозволяють одночасно аналізувати просторово-часові патерни. Класична двопотокова архітектура обробляє статичні кадри в одному потоці та оптичний потік в іншому, об'єднуючи інформацію про зовнішній вигляд та рух [12]. Тривимірні конволюції розширюють стандартні двовимірні фільтри на темпоральну вимірність, дозволяючи виділяти просторово-часові ознаки безпосередньо з послідовностей кадрів [13].

Автоенкодери використовуються для навчання без учителя та виявлення аномалій як відхилень від навчених паттернів [14]. Ідея полягає в тому, щоб навчити нейронну мережу стискати вхідні дані у компактне представлення в прихованому шарі та потім реконструювати вихідні дані з цього представлення. Якщо автоенкодер навчений лише на нормальних даних, він добре реконструюватиме схожі нормальні зразки, але погано справлятиметься з аномаліями. Висока помилка реконструкції слугує індикатором аномальності. Варіаційні автоенкодери розширюють цю ідею, навчаючись ймовірнісному розподілу нормальних даних у латентному просторі, що дозволяє не тільки виявляти аномалії, але й оцінювати їх імовірність [15].

Генеративно-змагальні мережі застосовуються для генерації нормальної поведінки та детекції аномалій через механізм змагання між

генератором, який намагається створити реалістичні дані, та дискримінатором, який намагається відрізнити згенеровані дані від реальних [16], [17]. Для виявлення аномалій генератор навчається продукувати лише нормальні патерни поведінки, і будь-які реальні дані, які дискримінатор ідентифікує як неправдоподібні або які генератор не може адекватно відтворити, вважаються аномальними.

Трансформери та механізми уваги, які революціонізували обробку природної мови, нещодавно стали застосовуватися і для аналізу відео [18]. На відміну від рекурентних мереж, які обробляють послідовності елемент за елементом, трансформери можуть аналізувати всю послідовність паралельно завдяки механізму самоуваги, що дозволяє моделювати зв'язки між будь-якими двома моментами часу незалежно від їхньої відстані у послідовності. Візійні трансформери розбивають зображення на невеликі фрагменти (наприклад, 16×16 пікселів), які потім обробляються як токени, аналогічно словам у природній мові [19]. Відеотрансформери розширюють цю концепцію на темпоральну вимірність, розглядаючи відео як тривимірну послідовність просторово-часових фрагментів або як набір окремих кадрів, між якими встановлюються зв'язки через механізми перехресної уваги [20]. Ключовою перевагою трансформерів є їхня здатність моделювати глобальні залежності у відео: модель може безпосередньо пов'язувати події, що відбуваються в початковій та кінцевій частинах відеопослідовності, без необхідності послідовного проходження через проміжні кадри, що є проблемою для LSTM. Механізми уваги дозволяють моделі динамічно визначати, які частини вхідних даних є найбільш інформативними для прийняття рішення, генеруючи ваги важливості для кожного просторового регіону та часового моменту. Це особливо цінно для інтерпретованості системи, оскільки візуалізація мап уваги дозволяє зрозуміти, на які саме ділянки відео та моменти часу модель звертає увагу при виявленні аномалій.

Проте трансформери мають і недоліки: вони потребують значно більше обчислювальних ресурсів через квадратичну складність механізму

самоуваги відносно довжини послідовності, що ускладнює їх застосування для довгих відеопослідовностей. Для вирішення цієї проблеми розробляються ефективніші варіанти трансформерів з лінійною або логарифмічною складністю, такі як Linformer, Performer та Swin Transformer, які використовують ієрархічну структуру обробки з локальною увагою на нижніх рівнях та глобальною на верхніх [21].

Сучасні дослідження все частіше звертаються до комбінування різних підходів, створюючи гібридні та ансамблеві методи. Інтеграція ознак зовнішнього вигляду та руху дозволяє використовувати комплементарну інформацію з обох джерел [22]. Ознаки на основі зовнішнього вигляду фіксують статичну інформацію про об'єкти та сцену, тоді як ознаки на основі руху фіксують динаміку змін. Поєднання методів навчання з учителем та без учителя дозволяє використовувати переваги обох парадигм: точність контрольованого навчання для відомих типів аномалій та гнучкість неконтрольованого навчання для виявлення нових типів [23]. Багаторівневі архітектури з різними масштабами аналізу обробляють відео на різних рівнях абстракції, від аналізу окремих пікселів до високорівневої семантики сцени. Ансамблі моделей для підвищення надійності детекції комбінують прогнози кількох різних моделей, зменшуючи вплив помилок окремих компонентів та підвищуючи загальну стабільність системи.

1.4 Аналіз периферійних обчислень та адаптивних систем

Сучасний етап розвитку галузі характеризується переходом до периферійних обчислень [24], коли системи виявлення аномалій розгортаються безпосередньо на периферійних пристроях, таких як розумні камери або локальні сервери, замість централізованої обробки в хмарі або віддалених серверних центрах. Цей тренд зумовлений кількома факторами. Розгортання на периферії забезпечує зменшення латентності та можливість справді реального часу, оскільки дані не потребують передачі через мережу

на віддалений сервер для обробки. Це критично важливо для застосувань, де навіть затримка в частки секунди може бути критичною. Крім того, значно знижується навантаження на мережу та вартість передачі даних, що особливо актуально при роботі з потоками високої роздільності від великої кількості камер. Підвищується також приватність [25], оскільки відео не покидає локальну мережу, а передаються лише метадані про виявлені події. Це відповідає сучасним вимогам захисту персональних даних. Нарешті, забезпечується незалежність від якості мережевого з'єднання та можливість роботи в умовах обмеженої або нестабільної зв'язку.

Для реалізації периферійних обчислень розробляються легковагі моделі, такі як MobileNet, EfficientNet та серія YOLO, які спеціально оптимізовані для роботи на пристроях з обмеженими ресурсами. Застосовуються також методи компресії нейронних мереж [26], включаючи обрізання для видалення надлишкових зв'язків та нейронів, квантування [27] для зменшення точності представлення ваг моделі з плаваючої точки до цілочисельної арифметики, та дистиляцію знань [28] для навчання компактною студентською моделі імітувати поведінку великої вчительської моделі. Нейроморфні чіпи та спеціалізовані прискорювачі, такі як Google Edge TPU, Intel Neural Compute Stick та NVIDIA Jetson, спеціально розроблені для ефективного виконання операцій нейронних мереж з низьким енергоспоживанням.

Розробляються також системи, здатні до онлайн-навчання та адаптації до змін середовища. Безперервне навчання дозволяє моделі інкрементально навчатися на нових даних без катастрофічного забування раніше набутих знань, що є типовою проблемою нейронних мереж. Методи адаптації до нових доменів, такі як доменна адаптація та трансферне навчання, дозволяють перенести знання, набуті на одній локації або наборі даних, на іншу локацію зі зміненими умовами, зменшуючи необхідність у великій кількості нових розмічених даних. Навчання на малій кількості прикладів через малоприкладне навчання [29] та метанавчання [30] дозволяє швидко

адаптуватися до нових типів аномалій, маючи лише кілька зразків. Активне навчання для ефективної взаємодії з операторами безпеки дозволяє системі вибірково запитувати мітки для найбільш інформативних або неоднозначних зразків, мінімізуючи навантаження на людину.

Інтелектуальний відеомоніторинг все більше інтегрується з автоматизованими системами реагування, формуючи комплексні екосистеми безпеки. Автоматичне оповіщення служб безпеки з пріоритизацією тривог базується на оцінці серйозності та достовірності виявлених аномалій, направляючи найкритичніші сповіщення безпосередньо відповідальним особам. Системи класифікують тривоги за рівнями пріоритету: критичні (потребують негайного реагування, наприклад, виявлення зброї або агресивної поведінки), високі (підозрілі дії, що можуть перерости в інцидент), середні (потенційні порушення, що потребують перевірки) та низькі (інформаційні повідомлення про незначні відхилення). Інтелектуальна система може враховувати додаткові фактори при визначенні пріоритету, такі як локація події, час доби, наявність людей поблизу, історію попередніх інцидентів у цій зоні. Інтеграція з системами контролю доступу та тривожної сигналізації дозволяє автоматично блокувати двері, активувати сирени або виконувати інші захисні дії при виявленні загроз. Сучасні системи можуть також інтегруватися з освітленням для підсвічування зон інциденту, з системами оповіщення для передачі попереджень через гучномовці, з ліфтами для обмеження доступу на певні поверхи, та з системами пожежної безпеки для координації евакуації. Підтримка прийняття рішень операторами через інтуїтивну візуалізацію, пояснення рішень системи [32] та надання релевантного контексту допомагає людині швидко оцінити ситуацію та прийняти обґрунтоване рішення. Методи пояснюваного штучного інтелекту такі як Grad-CAM, LIME та SHAP, візуалізують які ділянки відео та які ознаки вплинули на рішення моделі. Система може показувати оператору тепловідеокарти уваги моделі, виділяти об'єкти та траєкторії, що

спричинили тривогу, надавати текстові пояснення на кшталт «виявлено особу з підозрілим об'єктом, що перебуває в обмеженій зоні понад 5 хвилин». Ергономічний інтерфейс оператора включає багатоекранну візуалізацію з одночасним відображенням кількох камер, карт об'єктів з позначенням місць активних тривог, часової шкали подій та статистики. Формування аналітичних звітів та виявлення довгострокових трендів дозволяє ідентифікувати системні проблеми безпеки та оптимізувати стратегії превенції. Системи можуть генерувати звіти про частоту різних типів інцидентів за часовими періодами, виявляти закономірності у розподілі подій за днями тижня та годинами доби, аналізувати ефективність роботи служб реагування через метрики часу відгуку. Передбачальна аналітика дозволяє прогнозувати ймовірність виникнення інцидентів у певних зонах та часових інтервалах на основі історичних даних, що дозволяє проактивно розподіляти ресурси охорони. Довгострокова аналітика також допомагає у плануванні модернізації систем безпеки: визначенні місць, де необхідні додаткові камери, оптимізації кутів огляду існуючих камер, оцінці ефективності різних типів обладнання.

1.5 Аналіз актуальності розвитку

Актуальність розвитку інтелектуальних систем відеомоніторингу зумовлена комплексом взаємопов'язаних технологічних, економічних, соціальних, регуляторних та галузевих факторів. Технологічні фактори включають експоненційне зростання обчислювальних потужностей, що робить можливим виконання складних обчислень глибоких нейронних мереж у реальному часі навіть на відносно недорогих пристроях. Прогрес в архітектурах нейронних мереж та в алгоритмах навчання, включаючи покращені оптимізатори, методи регуляризації та техніки аугментації даних, постійно підвищує можливості систем. Зниження вартості відеокамер та підвищення їх якості до роздільності 4К, 8К та з підтримкою

HDR робить високоякісне відеоспостереження доступним для широкого кола застосувань. Доступність великих анотованих датасетів, таких як UCF-Crime, UCSD Anomaly Detection Dataset, ShanghaiTech Campus та UBnormal, забезпечує ресурси для навчання та тестування нових методів.

Регуляторні фактори відображають зростаючу увагу урядів та міжнародних організацій до питань безпеки. Посилення вимог до безпеки критичної інфраструктури, включаючи енергетичні об'єкти, транспортні вузли, водопостачання та телекомунікації, вимагає впровадження автоматизованих систем моніторингу. Директива ЄС про безпеку критичної інфраструктури (CER Directive) встановлює обов'язкові вимоги щодо оцінки ризиків та впровадження превентивних заходів, включаючи інтелектуальні системи спостереження. Законодавчі ініціативи щодо захисту від насильства, особливо в навчальних закладах після серії трагічних інцидентів, стимулюють впровадження систем раннього виявлення загроз. Стандарти безпеки в промисловості, такі як ISO 45001 для охорони праці, вимагають автоматизованого моніторингу дотримання правил безпеки, включаючи виявлення відсутності засобів індивідуального захисту, порушень протоколів роботи з небезпечним обладнанням, та несанкціонованого доступу до небезпечних зон. Водночас регуляторна база накладає значні обмеження на використання систем відеомоніторингу через питання приватності та захисту персональних даних.

Загальний регламент про захист даних Європейського Союзу (GDPR) встановлює суворі правила обробки біометричних даних, вимагає прозорості в використанні автоматизованих систем прийняття рішень, надає громадянам право на пояснення алгоритмічних рішень. Системи повинні імплементувати принципи Privacy by Design та Privacy by Default, мінімізувати збір даних, забезпечувати анонімізацію або псевдонімізацію персональної інформації де можливо, та отримувати відповідні згоди на обробку даних. Проект регуляції ЄС щодо штучного інтелекту (EU AI Act) класифікує системи розпізнавання та ідентифікації людей у публічних

місцях як високоризикові, що вимагає суворої оцінки відповідності, документування процесів навчання моделей, тестування на предмет упереджень, та забезпечення людського нагляду над автоматизованими рішеннями. У різних юрисдикціях існують специфічні обмеження: у деяких містах США заборонено використання технології розпізнавання облич правоохоронними органами, в Китаї існує детальна регуляція використання систем «розумного міста», європейські країни встановлюють різні вимоги щодо маркування зон відеоспостереження та інформування громадян. Етичні принципи розробки та використання AI-систем безпеки, сформульовані в документах IEEE, ACM та національних комісій з етики, наголошують на необхідності справедливості (відсутності дискримінації за расою, статтю, віком), підзвітності (можливості встановити відповідальних за рішення системи), прозорості (зрозумілості роботи алгоритмів для користувачів та регуляторів), та благополуччя (переваги застосування повинні перевищувати потенційні ризики для суспільства).

Проблема алгоритмічних упереджень є особливо актуальною: якщо система навчена на незбалансованих даних, вона може демонструвати різну точність для різних демографічних груп, що неприйнятно з етичної та юридичної точок зору. Сертифікація та аудит систем безпеки на основі AI стає стандартною практикою: незалежні організації перевіряють відповідність технічним специфікаціям, тестують на предмет упереджень, оцінюють стійкість до атак та помилок. Питання відповідальності при помилкових спрацюваннях або пропущених загрозах залишається юридично складним: хто несе відповідальність – розробник алгоритму, постачальник системи, оператор об'єкта, чи сам AI-агент як потенційно автономна сутність – є предметом активних дискусій у правовій спільноті. Ці регуляторні та етичні виклики формують складне середовище, в якому технологічні інновації повинні балансувати між підвищенням безпеки та збереженням громадянських свобод.

2 МЕТОДИ ТА МЕТОДИКА ДОСЛІДЖЕННЯ

2.1 Обґрунтування вибору напрямку дослідження

Вибір напрямку дослідження інтелектуальних систем відеомоніторингу для виявлення аномальної поведінки об'єктів у реальному часі обумовлений критичною необхідністю подолання фундаментальних обмежень існуючих підходів та задоволення зростаючих вимог сучасних систем безпеки. Аналіз стану галузі виявив значний розрив між теоретичними можливостями методів штучного інтелекту та практичними потребами реальних застосувань, що проявляється у кількох ключових аспектах.

Перша група факторів, що визначає актуальність обраного напрямку, пов'язана з обмеженнями точності та надійності існуючих методів виявлення аномалій. Традиційні підходи на основі аналізу траєкторій демонструють прийнятну ефективність лише в контрольованих умовах з чіткою видимістю окремих об'єктів та відносно простими сценаріями поведінки. В реальних умовах експлуатації ці методи стикаються з численними проблемами, включаючи втрату відстеження об'єктів в моментах коли один об'єкт частково або повністю перекривається іншим, неможливість коректної обробки густих натовпів, де індивідуальне відстеження стає неможливим, та чутливість до змін освітлення, погодних умов та кута огляду камери. Методи на основі класичного машинного навчання з ручним проектуванням ознак виявляють обмежену здатність до узагальнення на нові типи аномалій та середовища, оскільки кожен новий сценарій застосування потребує переосмислення набору ознак та часто повного переналаштування системи. Це робить їх неефективними для масштабного розгортання в умовах, що динамічно змінюються.

Ранні підходи глибокого навчання, незважаючи на значне покращення якості виявлення порівняно з традиційними методами, також демонструють

суттєві недоліки. Багато сучасних систем базуються на архітектурах, оптимізованих для задач класифікації дій, але не спеціально розроблених для виявлення аномалій. Це призводить до високої частоти хибнопозитивних спрацювань, коли система помилково класифікує нормальні, але нетипові дії як аномальні, та хибнонегативних пропусків, коли справжні аномалії залишаються невиявленими через їх схожість з деякими аспектами нормальної поведінки. Особливо проблематичним є виявлення тонких аномалій, що відрізняються від нормальної поведінки лише незначними деталями, а також складних багатоетапних сценаріїв, де аномальність виявляється не в окремих діях, а в їх послідовності та контексті.

Друга група факторів стосується обчислювальної ефективності та можливості роботи в реальному часі. Більшість сучасних методів з найвищою точністю виявлення базуються на глибоких архітектурах нейронних мереж, що вимагають значних обчислювальних ресурсів. Архітектури на кшталт складних трансформерів можуть обробляти відеопотік зі швидкістю лише кілька кадрів на секунду навіть на потужних серверних GPU, що робить їх непридатними для застосувань реального часу з множинними камерами. Затримка між виникненням події та її виявленням системою може становити секунди або навіть десятки секунд, що критично знижує корисність системи для превенції інцидентів. Спроби використання хмарних обчислень для вирішення проблеми продуктивності створюють додаткові проблеми: необхідність передачі великих обсягів відеоданих через мережу призводить до додаткової затримки, збільшення вартості експлуатації через плату за трафік та хмарні ресурси, та потенційних проблем з приватністю та безпекою даних при передачі чутливої інформації через публічні мережі.

Альтернативний підхід периферійних обчислень, коли аналіз виконується безпосередньо на камерах або локальних серверах, вимагає створення компактних моделей, що можуть працювати на обладнанні з

обмеженими ресурсами. Існуючі легковагі архітектури, такі як MobileNet або EfficientDet, демонструють значне зниження точності порівняно з їх повнорозмірними аналогами, особливо при виявленні дрібних об'єктів або тонких поведінкових патернів. Таким чином, існує фундаментальний компроміс між точністю та швидкістю, який ще не має задовільного вирішення. Обраний напрямок дослідження спрямований на розробку методів, що дозволяють досягти оптимального балансу між цими характеристиками через інтелектуальну обробку відеопотоку, адаптивний розподіл обчислювальних ресурсів та ієрархічні архітектури з різними рівнями деталізації аналізу.

Третя група факторів пов'язана з адаптивністю та узагальнюючою здатністю систем. Більшість існуючих підходів глибокого навчання базуються на парадигмі навчання з учителем, що вимагає великих обсягів розмічених даних для кожного конкретного типу аномалій та середовища застосування. Збір та розмітка таких даних є надзвичайно трудомістким та дорогим процесом, оскільки вимагає перегляду величезних обсягів відео людьми-експертами та точного анотування часових інтервалів аномальних подій. Крім того, деякі типи аномалій, особливо рідкісні або небезпечні події, практично неможливо зібрати у достатній кількості для навчання. Навіть якщо система навчена на певному наборі даних, її продуктивність може значно знижуватися при розгортанні в новому середовищі через проблему domain shift, коли розподіл даних у реальних умовах відрізняється від тренувальних даних через різні камери, кути огляду, освітлення, щільність людей та специфіку локації.

Подолання цих проблем вимагає розробки методів, здатних до навчання з обмеженою кількістю анотацій або взагалі без них. Підходи напівконтрольованого навчання, що використовують невелику кількість розмічених даних у поєднанні з великою кількістю нерозмічених, демонструють перспективні результати, але потребують подальшого вдосконалення для стабільної роботи в різноманітних умовах. Методи

навчання без учителя, що базуються на моделюванні нормальної поведінки та виявленні відхилень від неї, теоретично не потребують анованих аномалій, але на практиці часто страждають від високої частоти хибних спрацювань через складність точного визначення меж нормальності. Активне навчання, де система інтелектуально вибирає найбільш інформативні зразки для розмітки людиною, може значно знизити обсяг необхідних анотацій, але вимагає ефективних стратегій вибірки та зручних інструментів взаємодії з експертами.

Четверта група факторів стосується інтерпретованості та довіри до автоматизованих систем. Глибокі нейронні мережі традиційно розглядаються як «чорні скриньки», що приймають рішення на основі складних нелінійних перетворень у просторі високої вимірності, які важко інтерпретувати людині. В контексті систем безпеки, де рішення можуть мати серйозні наслідки, від арешту підозрюваних до евакуації будівель, така непрозорість є неприйнятною як з практичної, так і з етичної точок зору. Оператори служб безпеки повинні розуміти, чому система генерує певну тривогу, щоб адекватно оцінити її достовірність та визначити відповідні дії. Відео є високовимірними просторово-часовими даними, і пояснення повинно охоплювати як просторові регіони, що є релевантними для рішення, так і темпоральні залежності між подіями в різні моменти часу. Візуалізації повинні бути інтуїтивними для операторів, що часто не мають глибоких технічних знань, та надавати практично корисну інформацію, а не лише абстрактні теплові карти активацій. Розробка методів, що органічно поєднують високу точність виявлення з природною інтерпретованістю рішень, є одним з пріоритетів обраного напрямку дослідження.

П'ята група факторів визначається контекстною залежністю аномалій та необхідністю багаторівневого розуміння сцени. Один і той самий візуальний паттерн може бути нормальним в одному контексті та аномальним в іншому. Людина, що біжить, є цілком нормальною на спортивному майданчику, але може сигналізувати про паніку або

переслідування в офісному приміщенні. Залишений багаж є звичайним явищем у зоні видачі багажу в аеропорту, але є потенційною загрозою в зоні реєстрації або в метро. Розуміння контексту вимагає аналізу не лише самих дій об'єктів, але й характеристик середовища, взаємодії між об'єктами, часових параметрів подій та їх відповідності очікуваній діяльності в даній локації.

Сучасні системи часто мають обмежену здатність до контекстного розуміння, фокусуючись переважно на локальних візуальних патернах без урахування глобальної семантики сцени. Інтеграція різних рівнів аналізу, від низькорівневих ознак руху та зовнішнього вигляду до високорівневого семантичного розуміння дій, взаємодій та намірів, залишається відкритою дослідницькою проблемою. Методи графових нейронних мереж для моделювання відносин між об'єктами, механізми уваги для виділення релевантних елементів сцени, та інтеграція зовнішніх знань про типові сценарії поведінки в різних середовищах є перспективними напрямками, що потребують подальшого розвитку.

Шоста група факторів пов'язана з практичними аспектами розгортання та експлуатації систем у реальних умовах. Лабораторні результати на бенчмарк-датасетах часто не переносяться безпосередньо на реальні застосування через численні фактори, що не враховуються в контрольованих експериментах. Реальні відеопотоки характеризуються змінною якістю через погодні умови, забруднення лінз, вібрації камер від вітру або проїжджих транспортних засобів, стиснення відеокодеками з втратами, що створює артефакти. Система повинна бути стійкою до цих факторів без значної деградації продуктивності. Крім того, необхідна інтеграція з існуючою інфраструктурою відеоспостереження, що часто базується на застарілому обладнанні та гетерогенних системах від різних виробників.

Практична корисність системи визначається не лише її здатністю виявляти аномалії, але й зручністю інтерфейсу для операторів, можливістю

налаштування чутливості та фільтрації тривоги, ефективністю інтеграції в робочі процеси служб безпеки. Система повинна генерувати не просто повідомлення про виявлення аномалії, а надавати операторові всю необхідну інформацію для швидкого прийняття рішення: відеофрагмент події з контекстом до та після, локалізацію на карті об'єкта, тип та оцінку серйозності аномалії, рекомендовані дії. Вимоги до зручності використання та надійності в умовах цілодобової експлуатації є не менш важливими, ніж технічні характеристики алгоритмів виявлення.

Сьома група факторів стосується етичних та соціальних аспектів використання автоматизованих систем спостереження. Системи відеомоніторингу можуть непропорційно впливати на різні соціальні групи, якщо алгоритми містять упередження, успадковані з тренувальних даних. Якщо система навчена переважно на зразках певних демографічних груп, вона може демонструвати різну точність для людей різної раси, віку, статі або типу одягу. Це не лише етично неприйнятно, але й юридично проблематично в юрисдикціях з антидискримінаційним законодавством. Забезпечення справедливості алгоритмів вимагає ретельного аналізу та балансування тренувальних даних, тестування на предмет упереджень, та розробки методів, що є інваріантними до незалежних від поведінки характеристик людей.

Питання приватності є особливо гострим в контексті відеомоніторингу публічних просторів. Громадяни мають право на приватність навіть у публічних місцях, і системи повинні мінімізувати збір та обробку персональної інформації. Це можна досягти через техніки анонімізації, такі як розмиття облич або заміна їх аватарами, аналіз лише силуетів та скелетних представлень людей без зберігання ідентифікуючої інформації, та обробку даних локально без передачі в централізовані бази. Прозорість використання систем, включаючи інформування громадськості про локації камер та цілі моніторингу, та механізми підзвітності, що

дозволяють оскаржувати рішення та перевіряти коректність роботи, є обов'язковими компонентами етично прийнятних систем.

Обраний напрямок дослідження враховує всі ці фактори та спрямований на розробку комплексного підходу, що поєднує передові методи глибокого навчання з практичними вимогами реальних застосувань. Фокус на виявленні аномальної поведінки в реальному часі підкреслює пріоритет не лише точності, але й швидкості та ефективності обробки. Це вимагає інноваційних архітектурних рішень, що оптимізують використання обчислювальних ресурсів, та інтелектуальних стратегій обробки відеопотоків, що фокусують увагу системи на найбільш релевантних аспектах сцени.

2.2 Порівняльний аналіз методів виявлення аномалій у відеопотоках

Систематичний порівняльний аналіз існуючих методів виявлення аномалій у відеопотоках дозволяє виявити їх відносні переваги та недоліки, що є критично важливим для обґрунтованого вибору базових технологій та визначення напрямків вдосконалення. Для структурованого порівняння методи класифіковано за ключовими характеристиками, що визначають їх придатність для практичного застосування.

Методи на основі реконструкції використовують автоенкодери або інші генеративні моделі, навчені на нормальних даних, для виявлення аномалій через помилку реконструкції. Фундаментальна ідея полягає в тому, що модель, навчена стискати та відновлювати нормальні відеопослідовності, не зможе ефективно реконструювати аномальні паттерни, що відсутні в тренувальних даних. Конволюційні автоенкодери з LSTM-шарами для темпоральної залежності демонструють здатність виявляти просторово-часові аномалії шляхом аналізу різниці між вхідними кадрами та їх реконструкціями. Високі значення помилки реконструкції в

певних просторових регіонах та часових інтервалах сигналізують про потенційні аномалії.

Перевагою методів реконструкції є можливість навчання без розмічених аномалій, що значно спрощує підготовку даних, оскільки потрібні лише зразки нормальної поведінки. Теоретично це дозволяє виявляти навіть невідомі типи аномалій, яких не було в тренувальних даних. Методи відносно стійкі до дисбалансу класів, типового для задачі виявлення аномалій, де нормальні події значно переважають аномальні. Однак існують суттєві недоліки: автоенкодери можуть навчитися реконструювати не лише нормальні, але й аномальні паттерни, якщо вони мають певну структуру, що призводить до пропуску аномалій. Визначення порогу помилки реконструкції, що відокремлює нормальні та аномальні зразки, є нетривіальною задачею і часто вимагає заснованого на досвіді налаштування для кожного конкретного застосування. Методи можуть бути чутливими до загального вигляду сцени та умов зйомки: зміни освітлення, появ нових об'єктів у фоні або навіть зміни погоди можуть призвести до високої помилки реконструкції без наявності справжньої аномалії. Обчислювальна складність реконструкції повних відеокадрів високої роздільності може бути значною, особливо для архітектур з глибокими енкодерами та декодерами.

Варіаційні автоенкодери розширюють базову концепцію реконструкції, навчаючи ймовірнісний розподіл нормальних даних у латентному просторі. Це дозволяє не лише реконструювати дані, але й оцінювати їх правдоподібність під навченим розподілом. Аномалії ідентифікуються як зразки з низькою правдоподібністю, що знаходяться далеко від основної маси нормальних даних у латентному просторі. Регуляризація латентного простору покращує узагальнюючу здатність моделі та зменшує ризик переобучення на специфічних паттернах тренувальних даних. Однак навчання варіаційних автоенкодерів вимагає балансування між точністю реконструкції та регуляризацією латентного

простору, що контролюється гіперпараметром, оптимальне значення якого не завжди очевидне. Обчислення правдоподібності в просторах високої вимірності є нетривіальним та може потребувати апроксимацій. Інтерпретація аномалій, виявлених у латентному просторі, для операторів може бути складнішою порівняно з просторовими помилками реконструкції.

Генеративно-змагальні мережі для виявлення аномалій базуються на навчанні генератора створювати реалістичні нормальні відеопослідовності та дискримінатора відрізнити згенеровані дані від реальних. Після навчання аномалії виявляються через неможливість генератора їх відтворити або через низьку оцінку дискримінатора для реальних аномальних зразків. Теоретично генеративно-змагальні мережі здатні навчитися дуже детальним представленням нормальної поведінки, включаючи складні текстури та динаміку рухів. Деякі архітектури дозволяють генерувати синтетичні аномалії для аугментації тренувальних даних або тестування системи. Однак навчання генеративно-змагальних мереж для відео є складним процесом через проблеми стабільності, mode collapse коли генератор навчається створювати лише обмежену різноманітність вихідних даних, та необхідність балансування навчання генератора та дискримінатора. Якість згенерованого відео часто страждає від артефактів, особливо для складних сцен з множинними рухомими об'єктами. Обчислювальні вимоги для навчання GAN з високою роздільністю є значними, часто вимагаючи днів або тижнів навчання на потужних GPU-кластерах.

Методи передбачення майбутніх кадрів підходять до проблеми виявлення аномалій з іншого боку: замість реконструкції наявних даних вони намагаються передбачити майбутні кадри на основі попередньої історії. Моделі навчаються на нормальних послідовностях передбачати логічне продовження подій. Аномалії виявляються як ситуації, де реальні майбутні кадри значно відрізняються від передбачених, сигналізуючи про

порушення очікуваної динаміки. Конволюційні LSTM-мережі та архітектури з механізмами уваги для передбачення наступних кадрів дозволяють моделювати просторово-часову еволюцію сцени. Цей підхід має природну інтерпретацію: людина також виявляє аномалії, коли події розгортаються не так, як очікувалось. Методи фокусуються на темпоральній динаміці, що важливо для виявлення аномалій поведінки, а не лише статичних аномалій зовнішнього вигляду.

Переваги включають можливість навчання без анованих аномалій та фокус на темпоральних аспектах поведінки, що є критичним для виявлення аномальних послідовностей дій. Проте існують суттєві виклики: передбачення майбутніх кадрів на рівні пікселів є надзвичайно складною задачею через стохастичність реального світу, де можливі множинні правдоподібні продовження однієї ситуації. Моделі часто генерують розмиті передбачення, усереднюючи можливі варіанти майбутнього, що утруднює визначення значущих відхилень. Короткостроковість передбачень обмежує здатність виявляти аномалії, що розгортаються протягом тривалого часу: передбачення на десятки кадрів вперед є нереалістичним через накопичення невизначеності. Обчислювальна вартість генерації передбачених кадрів високої роздільності може бути значною, особливо якщо передбачення потрібно робити для кожного кадру в реальному часі.

Методи на основі класифікації дій підходять до виявлення аномалій як до задачі розпізнавання специфічних типів аномальної поведінки. Глибокі нейронні мережі, а саме, двопотокові архітектури, навчаються класифікувати відеосегменти на категорії нормальних та різних типів аномальних дій. Це контрольоване навчання вимагає великих обсягів анованих даних з прикладами кожного типу аномалій.

Основна перевага підходу класифікації полягає в можливості досягнення високої точності для відомих типів аномалій, представлених у тренувальних даних. Моделі можуть навчитися тонким відмінностям між

схожими діями та розпізнавати складні багатокomпонентні аномалії. Класифікація природно надає семантичну інтерпретацію виявлених аномалій: система не лише повідомляє про аномалію, але й вказує її конкретний тип. Це полегшує прийняття рішень операторами та дозволяє налаштувати різні рівні реагування для різних категорій. Використання попередньо навчених моделей на великих датасетах дій значно прискорює навчання через transfer learning.

Однак існують критичні обмеження: неможливість виявлення невідомих типів аномалій, що не були представлені в тренувальних даних, робить систему вразливою до нових загроз. Збір та анотація представницьких тренувальних даних для всіх можливих типів аномалій є практично неможливим завданням, особливо для рідкісних або небезпечних подій. Проблема довгохвостого розподілу, де деякі класи аномалій представлені значно частіше за інші, призводить до дисбалансу та зниження точності для рідкісних категорій. Модель може переобучитися на специфічних візуальних характеристиках аномалій у тренувальних даних замість навчання загальних паттернів аномальної поведінки, що погіршує узагальнення. Обчислювальна складність глибоких класифікаторів дій, особливо тривимірних конволюційних мереж, може бути значною.

Гібридні підходи намагаються поєднати переваги різних методів, компенсуючи їх індивідуальні недоліки. Одна з популярних стратегій полягає в поєднанні навчання без учителя для виявлення загальних відхилень від нормальності з наступною класифікацією з учителем для розпізнавання специфічних типів аномалій. Автоенкодер або інша модель без контролю навчається на нормальних даних та виявляє регіони просторово-часового відхилення, після чого класифікатор аналізує ці регіони для визначення конкретного типу аномалії. Це дозволяє виявляти як відомі, так і невідомі типи аномалій, забезпечуючи більшу гнучкість та адаптивність системи.

Інший гібридний підхід інтегрує ознаки зовнішнього вигляду та ознаки руху через багатопотокові архітектури. Окремі потоки обробляють RGB-кадри для аналізу статичного вигляду та оптичний потік або різниці кадрів для аналізу руху, після чого інформація об'єднується для фінального рішення. Це дозволяє виявляти як аномалії зовнішнього вигляду, такі як підозрілі об'єкти, так і аномалії динаміки, такі як незвичайні рухи. Експериментальні дослідження підтверджують, що комбінація різних типів ознак часто перевершує використання одного джерела інформації.

Методи на основі графів та відношень моделюють взаємодії між об'єктами в сцені через графові структури, де вершини представляють об'єкти, а ребра їх взаємовідношення. Графові нейронні мережі навчаються аналізувати ці структури та виявляти аномальні паттерни взаємодій. Цей підхід особливо корисний для виявлення аномалій, що проявляються не в індивідуальній поведінці об'єктів, а в їх колективній динаміці або незвичайних взаємодіях. Наприклад, переслідування може бути виявлене через аналіз просторових відношень та взаємної орієнтації двох людей протягом часу, навіть якщо кожна людина окремо рухається нормально.

Моделювання просторових відношень включає відстані між об'єктами, їх відносні розміри та взаємне розташування, що дозволяє виявляти порушення особистого простору або незвичайні групування. Аналіз темпоральної еволюції графу відношень виявляє зміни в структурі взаємодій, такі як раптове розпорошення натовпу, що може сигналізувати про паніку. Графові мережі можуть обробляти змінну кількість об'єктів у сцені та динамічно адаптуватися до змін їх складу. Проте побудова осмислених графових представлень вимагає надійного виявлення та відстеження об'єктів, що само по собі є нетривіальною задачею. Визначення релевантних типів відношень та їх представлення в графовій структурі потребує доменних знань. Обчислювальна складність графових мереж зростає з кількістю об'єктів та ребер у графі.

Методи на основі трансформерів представляють найновіший напрямок у аналізі відео та виявленні аномалій, використовуючи механізми самоуваги для моделювання довгострокових залежностей. Візійні трансформери розглядають відео як послідовність просторово-часових фрагментів і застосовують увагу для виявлення відношень між будь-якими двома фрагментами незалежно від їх відстані у просторі або часі. Це дозволяє безпосередньо моделювати глобальний контекст та виявляти аномалії, що проявляються через незвичайні комбінації або послідовності подій у різних частинах кадру або в різні моменти часу.

Ключова перевага трансформерів полягає в їх здатності до глобального моделювання залежностей без обмежень рецептивного поля, притаманних конволюційним архітектурам. Механізми уваги природно надають інтерпретованість через візуалізацію ваг уваги, що показують, які частини відео модель вважає релевантними для прийняття рішення. Трансформери демонструють відмінну здатність до transfer learning з великих датасетів на специфічні задачі.

Однак трансформери мають значні обчислювальні вимоги через квадратичну складність механізму самоуваги відносно кількості фрагментів. Для відео високої роздільності та довгих послідовностей це може бути проблематичним. Навчання трансформерів з нуля вимагає дуже великих обсягів даних через велику кількість параметрів та відсутність вбудованих індуктивних упереджень на кшталт локальності та трансляційної інваріантності у конволюційних мереж. Розробка ефективних трансформерів для відео є активною областю досліджень з постійним потоком нових архітектур та оптимізацій.

Порівняння методів за ключовими критеріями надає комплексне розуміння їх придатності для різних сценаріїв застосування. За критерієм точності виявлення аномалій методи контрольованої класифікації дій демонструють найвищу точність для відомих типів аномалій при наявності великих розмічених датасетів. Гібридні підходи з комбінацією appearance та

motion features також показують відмінні результати. Методи реконструкції та предикції демонструють меншу точність, але більшу універсальність.

За критерієм можливості виявлення невідомих аномалій методи без контролю, такі як автоенкодери та GAN, мають теоретичну перевагу, оскільки не обмежені наперед визначеними категоріями. Методи класифікації слабкі в цьому аспекті. За критерієм обчислювальної ефективності методи на основі траєкторій та класичні підходи машинного навчання є найшвидшими, проте менш точними.

За критерієм вимог до навчальних даних методи без контролю потребують лише нормальних даних без анотацій аномалій, що значно спрощує підготовку. Методи напівконтрольованого навчання вимагають обмежену кількість розмічених аномалій. Методи класифікації потребують великих обсягів детально анотованих даних для всіх типів аномалій. За критерієм інтерпретованості методи на основі траєкторій та правил надають найбільш прозорі пояснення. Механізми уваги в трансформерах та візуалізації активацій через Grad-CAM забезпечують певний рівень інтерпретованості для глибоких моделей. Методи з латентними представленнями є найменш інтерпретованими.

За критерієм адаптивності до нових середовищ методи без контролю та напівконтрольовані підходи демонструють кращу здатність до адаптації з мінімальними додатковими даними. Попередньо навчені моделі з трансферним навчанням також адаптуються ефективно. Моделі, навчені з нуля для специфічних категорій, вимагають повного перенавчання. За критерієм стійкості до шуму та варіабельності умов зйомки ансамблі моделей та гібридні підходи демонструють найвищу стійкість. Методи на основі реконструкції можуть бути чутливими до змін освітлення та фону.

Синтез результатів порівняльного аналізу дозволяє сформулювати рекомендації щодо вибору методів для різних сценаріїв застосування. Для високоточного виявлення специфічних добре визначених типів аномалій у контрольованих середовищах з наявністю великих розмічених датасетів

оптимальними є методи контрольованої класифікації дій з використанням глибоких просторово-часових архітектур. Для універсального виявлення широкого спектру відомих та невідомих аномалій у середовищах з обмеженими анотаціями доцільні гібридні підходи, що поєднують методи без контролю для первинного виявлення відхилень та класифікацію для розпізнавання відомих типів. Для застосувань реального часу з обмеженими обчислювальними ресурсами необхідні легковагі архітектури з оптимізацією для периферійних пристроїв, можливо з ієрархічною обробкою та адаптивним розподілом ресурсів. Для складних сценаріїв з множинними взаємодіючими об'єктами корисні методи на основі графів та відношень. Для застосувань, що вимагають високої інтерпретованості рішень, доцільні архітектури на основі уваги з візуалізацією та методи з природними поясненнями.

2.3 Обґрунтування вибору методів для розробки системи

На основі проведеного порівняльного аналізу та врахування специфічних вимог задачі виявлення аномальної поведінки об'єктів у реальному часі, обрано комплексний підхід, що інтегрує кілька комплементарних методів для досягнення оптимального балансу між точністю, швидкістю, універсальністю та практичною корисністю. Обґрунтування вибору базується на детальному аналізі компромісів між різними характеристиками та пріоритизації вимог відповідно до цілей дослідження.

Базовою архітектурою для виділення просторово-часових ознак обрано гібридний підхід, що поєднує ефективні конволюційні мережі для просторової обробки з механізмами уваги для темпорального моделювання. Конкретно, для просторового аналізу окремих кадрів використовується архітектура EfficientNet, що забезпечує відмінний баланс між точністю та обчислювальною ефективністю завдяки принципу складеного

масштабування, який одночасно оптимізує глибину, ширину та роздільність мережі. EfficientNet досягає точності, порівнянної з набагато більшими архітектурами, використовуючи значно менше параметрів та обчислень, що критично важливо для обробки реального часу.

Вибір EfficientNet замість інших популярних архітектур обумовлений кількома факторами. Порівняно з VGGNet, EfficientNet має значно менше параметрів та вимагає менше обчислень при порівнянній або вищій точності. ResNet, незважаючи на широке використання, менш ефективний в термінах обчислень на одиницю точності. Inception-архітектури, хоча й ефективні, мають більш складну структуру, що ускладнює модифікацію та інтерпретацію. MobileNet, оптимізований для мобільних пристроїв, демонструє меншу точність порівняно з EfficientNet при схожих обчислювальних бюджетах. Важливою перевагою EfficientNet є наявність різних версій від B0 до B7, що дозволяє гнучко вибирати модель залежно від доступних обчислювальних ресурсів та вимог до точності, забезпечуючи масштабованість системи.

Для темпорального моделювання обрано підхід на основі уваги замість традиційних LSTM або тривимірних конволюцій. Спеціалізований модуль темпоральної уваги аналізує послідовність просторових ознак, виділених EfficientNet з послідовності кадрів, та визначає важливість кожного часового моменту для виявлення аномалій. Це дозволяє моделі фокусуватися на критичних моментах послідовності, ігноруючи нерелевантні кадри, що покращує як точність, так і ефективність.

Перевага механізмів уваги над LSTM полягає в можливості моделювання довгострокових залежностей без проблеми згасання градієнтів, паралелізації обчислень на відміну від послідовної природи LSTM, та природній інтерпретованості через візуалізацію ваг уваги. Порівняно з повними трансформерами, спеціалізований модуль уваги має значно меншу обчислювальну складність завдяки фокусуванню лише на темпоральній вимірності та використанню ефективних варіантів механізму

уваги. Порівняно з тривимірними конволюціями, увага надає більшу гнучкість у моделюванні залежностей на різних часових масштабах без необхідності наперед визначати розмір темпорального ядра.

Для виявлення аномалій застосовується гібридна стратегія, що поєднує навчання без учителя на нормальних даних з опціональним дообученням на обмеженій кількості анованих аномалій. Базова модель навчається на великій кількості відеопослідовностей нормальної поведінки через завдання контрастивного навчання в поєднанні з передбачувальним кодуванням. Контрастивне навчання забезпечує, що представлення схожих нормальних послідовностей є близькими в просторі ознак, тоді як різні послідовності мають віддалені представлення. Передбачувальне кодування вимагає від моделі передбачати майбутні ознаки на основі поточних, що забезпечує навчання темпоральної динаміки нормальної поведінки.

Після навчання на нормальних даних модель використовується для виділення високорівневих представлень відеопослідовностей. Аномалії виявляються через аналіз відстані в просторі ознак: послідовності, що значно віддалені від типових нормальних представлень, класифікуються як аномальні. Конкретно використовується підхід на основі однокласової класифікації з глибоким SVDD (опис опорних векторів даних), що навчає гіперсферу мінімального радіусу в просторі ознак, яка охоплює нормальні дані. Зразки за межами цієї гіперсфери вважаються аномаліями з впевненістю, що зростає з відстанню від границі.

Вибір контрастивного навчання та SVDD обумовлений їх ефективністю для навчання без розмічених аномалій, стійкістю до дисбалансу класів та можливістю виявлення невідомих типів аномалій. Альтернативні підходи, такі як автоенкодері, можуть навчитися реконструювати аномальні патерни, якщо вони структуровані. GAN важкі в навчанні та можуть страждати від нестабільності. Ізоляційний ліс та інші класичні методи виявлення викидів менш ефективні в просторах високої вимірності та не використовують переваги глибокого навчання.

Для підвищення точності виявлення специфічних важливих типів аномалій, таких як насильство, падіння людини або залишені об'єкти, система підтримує опціональне дообучення з використанням обмеженої кількості анованих прикладів цих категорій. Застосовується підхід малоприкладного навчання на основі прототипних мереж або мереж узгодження, що дозволяє ефективно навчатися розпізнавати нові категорії, маючи лише кілька зразків кожної. Базовий екстрактор ознак, навчений на великій кількості нормальних даних, залишається фіксованим або дообучується з дуже малою швидкістю навчання, тоді як навчаються лише компактні класифікаційні голови для нових категорій.

Цей гібридний підхід забезпечує гнучкість системи: базова здатність виявляти широкий спектр аномалій через відхилення від нормальності без необхідності їх анотації, та можливість спеціалізованого високоточного розпізнавання критично важливих типів аномалій при наявності навіть обмеженої кількості прикладів. Це відповідає реальним вимогам, де деякі типи аномалій є добре визначеними та критичними, тоді як інші можуть бути непередбачуваними.

Для забезпечення роботи в реальному часі розроблено ієрархічну стратегію обробки відеопотоків з адаптивним розподілом обчислювальних ресурсів. Система працює на кількох рівнях аналізу з різною частотою та складністю. Перший рівень швидкого скринінгу аналізує кожен кадр або кожні кілька кадрів з використанням легкої моделі для виявлення потенційно цікавих регіонів або моментів часу. Використовуються легковагі детектори об'єктів типу YOLO та аналіз низькорівневих ознак руху через оптичний потік або різниці кадрів. Цей рівень фільтрує очевидно нормальні періоди, такі як порожня сцена або монотонна активність без значущих змін.

Другий рівень детального аналізу активується для кадрів або відеосегментів, що пройшли перший рівень як потенційно цікаві. Тут застосовується основна модель виявлення аномалій з повним просторово-

часовим аналізом. Обчислювальні ресурси концентруються на найбільш релевантних частинах відео, що дозволяє застосувати складніші алгоритми без порушення вимог реального часу. Третій рівень верифікації та класифікації застосовується для підтверджених аномалій, де спеціалізовані класифікатори визначають конкретний тип аномалії, оцінюють її серйозність та генерують детальні пояснення для оператора.

Ця ієрархічна стратегія значно знижує обчислювальне навантаження, оскільки складні моделі застосовуються лише до невеликої частки вхідних даних, що потребують детального аналізу. Водночас забезпечується висока точність завдяки повному аналізу всіх підозрілих подій. Адаптивність системи проявляється в динамічному налаштуванні порогів чутливості на кожному рівні залежно від контексту: часу доби, рівня активності в сцені, історичної частоти аномалій у даній локації.

Для покращення інтерпретованості системи інтегровано кілька механізмів пояснення рішень. Візуалізація мап уваги показує оператору, які просторові регіони та часові моменти були найбільш важливими для рішення про виявлення аномалії. Теплові карти накладаються на оригінальні відеокадри, виділяючи об'єкти та зони, що привернули увагу моделі. Ваги уваги у темпоральній вимірності відображаються на часовій шкалі, показуючи ключові моменти аномальної послідовності. Візуалізація простору ознак через проєкції високовимірних представлень на двовимірний простір за допомогою t-SNE або UMAP дозволяє оператору побачити, наскільки виявлена аномалія відрізняється від типових нормальних зразків. Генерація текстових пояснень через інтеграцію з великими мовними моделями або шаблонну генерацію надає природномовний опис виявленої аномалії типу «виявлено дві особи в агресивній взаємодії в правій частині кадру протягом 8 секунд».

Порівняння з найближчими об'єктами-прототипами показує оператору схожі нормальні та аномальні зразки з історичних даних, допомагаючи контекстуалізувати поточну подію. Індикатори впевненості

надають кількісну оцінку достовірності виявлення, дозволяючи оператору пріоритизувати перевірку найбільш впевнених спрацювань. Ці механізми інтерпретованості не лише підвищують довіру операторів до системи, але й забезпечують цінний зворотний зв'язок для постійного вдосконалення моделей.

Для забезпечення приватності та відповідності регуляторним вимогам система включає модулі анонімізації та захисту даних. Опціональна анонімізація облич через розмиття або заміну синтетичними обличчями застосовується на рівні препроцесингу перед аналізом моделями. Використання оцінки пози та скелетних представлень замість повних зображень людей для багатьох типів аналізу поведінки дозволяє виявляти аномальні рухи та дії без обробки ідентифікуючої інформації. Локальна обробка на периферійних пристроях з передачею лише метаданих про виявлені події, а не повних відеопотоків, мінімізує ризики приватності. Система логування та аудиту забезпечує повну простежуваність всіх рішень системи та доступу до даних для забезпечення підзвітності.

Вибір саме цієї комбінації методів обумовлений необхідністю задовольнити множину часто суперечливих вимог. Точність виявлення забезпечується глибоким просторово-часовим аналізом з використанням сучасних архітектур нейронних мереж. Швидкість та можливість роботи в реальному часі досягається через ієрархічну обробку, ефективні архітектури та адаптивний розподіл ресурсів. Універсальність та здатність виявляти невідомі аномалії забезпечується навчанням без учителя на нормальних даних. Точне розпізнавання критичних типів аномалій досягається через малоприкладне дообучення на обмежених анотованих даних. Інтерпретованість надається механізмами уваги та спеціалізованими модулями пояснення. Приватність забезпечується анонімізацією та локальною обробкою. Адаптивність до нових середовищ забезпечується трансферним навчанням та можливістю онлайн-дообучення.

2.4 Методика розробки та інтеграції системи відеомоніторингу

Методика розробки системи охоплює весь життєвий цикл від проектування архітектури до розгортання та підтримки в експлуатації. Системна архітектура проектується як модульна та масштабована структура, що складається з кількох взаємодіючих компонентів з чітко визначеними інтерфейсами. Модуль захоплення відео відповідає за підключення до камер, отримання відеопотоків через стандартні протоколи HTTP або SDK виробників камер, декодування відео та перетворення в єдиний внутрішній формат. Модуль підтримує множинні одночасні потоки з різних камер з автоматичним керуванням підключеннями, перепідключенням при збоях та буферизацією для згладжування нерівномірності надходження кадрів.

Модуль препроцесингу виконує первинну обробку кадрів: масштабування до потрібної роздільності, нормалізацію колірному простору та діапазону значень пікселів, виділення regions of interest при необхідності для фокусування уваги на певних зонах сцени, опціональну анонімізацію через детекцію та розмиття облич. Препроцесинг оптимізується для мінімальної латентності з використанням апаратного прискорення через GPU або спеціалізовані процесори обробки зображень коли доступні. Модуль детекції та відстеження об'єктів ідентифікує та відстежує об'єкти інтересу в сцені: люди, транспортні засоби, залишені об'єкти. Використовується ефективний детектор YOLO для локалізації об'єктів на кожному кадрі та алгоритм відстеження DeepSORT для встановлення ідентичності об'єктів між кадрами. Виділяються траєкторії руху, обчислюються швидкості та прискорення, детектуються взаємодії між об'єктами.

Модуль виявлення аномалій є центральним компонентом системи та реалізує запропоновані методи глибокого навчання. Він отримує оброблені відеосегменти та метадані про виявлені об'єкти, виділяє просторово-часові

ознаки через основу EfficientNet та темпоральну увагу, обчислює оцінку аномальності для кожного сегменту через порівняння з навченим розподілом нормальних представлень, класифікує тип аномалії коли застосовно через спеціалізовані класифікатори, генерує пояснення рішень через візуалізації уваги та аналіз ознак. Модуль підтримує як синхронну обробку в реальному часі так і асинхронну пакетну обробку архівних записів.

Модуль управління тривогами обробляє виявлені аномалії та координує реагування: фільтрує та агрегує множинні детекції однієї події для запобігання дублюванню тривог, оцінює серйозність аномалій на основі типу, локації, контексту та історії, пріоритизує тривоги для визначення порядку обробки операторами, генерує сповіщення через різні канали включаючи веб-інтерфейс, мобільні додатки, email, SMS, інтеграцію з системами оповіщення, записує всі події та дії операторів для аудиту та аналізу. Модуль інтерфейсу оператора надає зручний веб-базований інтерфейс для моніторингу та управління системою.

Інтерфейс включає панель керування з відображенням стану всіх камер та активних тривог у реальному часі, карту об'єкта з позначенням локацій камер та подій, панель тривог з пріоритизованим списком та деталями кожної події включаючи відеофрагмент, анотації, пояснення системи, інструменти для відповіді на тривоги з можливістю підтвердити, відхилити або ескалювати, налаштування системи включаючи пороги чутливості, зони інтересу, розклади моніторингу, засоби аналітики та звітності з візуалізацією статистики інцидентів, трендів, продуктивності системи. Інтерфейс оптимізується для швидкого прийняття рішень з мінімізацією кількості кліків та когнітивного навантаження на оператора.

Модуль онлайн-навчання забезпечує адаптацію системи до змін середовища та отримання нових знань з досвіду експлуатації. Він збирає зворотний зв'язок від операторів коли вони підтверджують або відхиляють тривоги, акумулює нові зразки нормальної поведінки та підтвержені

аномалії, періодично переобучує або дообучує моделі на оновлених даних без порушення роботи системи через проміжне середовище, відстежує метрики продуктивності для виявлення деградації точності що може сигналізувати про зміни в середовищі. Онлайн-навчання може виконуватися як за розкладом так і автоматично при накопиченні достатньої кількості нових даних.

Інтеграційні інтерфейси забезпечують взаємодію з зовнішніми системами: інтеграція з системами контролю доступу для корелювання подій відеомоніторингу з даними про входи та виходи людей, підключення до систем тривожної сигналізації для автоматичного активування захисних заходів при критичних подіях, інтерфейс з системами освітлення та звукового оповіщення для координованого реагування, API для інтеграції з сторонніми системами безпеки та management platforms, експорт даних у стандартних форматах для аналізу в спеціалізованих інструментах. Інтеграції реалізуються через стандартні протоколи та API для максимальної сумісності.

Архітектура розгортання проектується для гнучкості та масштабованості залежно від вимог конкретного застосування. Для невеликих об'єктів з кількома камерами підтримується периферійне розгортання, де вся система працює на локальному сервері або навіть безпосередньо на розумних камерах з достатніми обчислювальними можливостями. Це мінімізує затримку та залежність від мережевої зв'язності, забезпечує приватність через відсутність передачі даних за межі об'єкта. Для середніх об'єктів з десятками камер оптимальна гібридна архітектура, де легкі моделі виконують первинний скринінг на периферії, виявляючи потенційно цікаві події, які передаються на центральний сервер для детального аналізу складнішими моделями.

Для великих розподілених об'єктів або систем Розумного міста з сотнями або тисячами камер застосовується розподілена хмарна архітектура з масштабованими мікросервісами. Кожен функціональний модуль

розгортається як незалежний сервіс, що може масштабуватися горизонтально залежно від навантаження. Використовуються системи оркестрації контейнерів Kubernetes для автоматичного розподілу навантаження, відновлення після збоїв, оновлення сервісів без простою. Брокери повідомлень, такі як Apache Kafka, забезпечують асинхронну комунікацію між сервісами та буферизацію для обробки пікових навантажень.

Безпека системи забезпечується на всіх рівнях архітектури. Шифрування передачі даних через TLS для всіх мережових з'єднань захищає відеопотоки та метадані від перехоплення. Аутентифікація та авторизація користувачів реалізується через стандартні протоколи OAuth2 та OpenID Connect з підтримкою єдиного входу. Рольове керування доступом обмежує доступ до функцій та даних відповідно до ролей користувачів: оператори мають доступ лише до моніторингу та реагування на події, адміністратори можуть налаштовувати систему, аудиторі мають доступ лише для читання до логів. Шифрування зберігання даних захищає записані відео та чутливу інформацію. Журнали аудиту записують всі дії користувачів та системи для забезпечення підзвітності та можливості розслідування інцидентів.

Тестування системи виконується на кількох рівнях. Функціональне тестування перевіряє що кожна функція системи працює відповідно до специфікацій: коректність виявлення різних типів аномалій, точність локалізації в часі та просторі, правильність генерування тривоги та сповіщень, коректність інтеграцій з зовнішніми системами. Тестування продуктивності оцінює швидкодію та масштабованість: максимальна кількість одночасно оброблюваних камер, затримка від події до тривоги під різним навантаженням, використання ресурсів ЦП, ГП, пам'яті, мережі, поведінка при пікових навантаженнях. Стрес-тестування виявляє межі системи через поступове збільшення навантаження до точки відмови.

Тестування надійності перевіряє стабільність при тривалій експлуатації: система запускається на тижні або місяці для виявлення

витоків пам'яті, проблем деградації, накопичувальних помилок. Тестування безпеки включає тестування на проникнення для виявлення вразливостей, сканування вразливостей з використанням автоматизованих інструментів, перегляд коду з фокусом на найкращі практики безпеки. Тестування зручності використання залучає реальних користувачів для оцінки інтерфейсу: легкість навчання для нових операторів, ефективність використання для досвідчених користувачів, толерантність до помилок та можливість відновлення після помилок користувача, задоволеність вимірюється через опитування та спостереження.

Пілотне розгортання виконується перед повномасштабним впровадженням для валідації системи в реальних умовах. Вибирається репрезентативне середовище що відображає типові умови експлуатації але має обмежений масштаб для зменшення ризиків. Система розгортається паралельно з існуючими рішеннями для порівняння продуктивності без залежності від нової системи на критично важливих операціях. Збираються детальні метрики про продуктивність системи, частоти різних типів тривог, точність виявлення через верифікацію операторами, технічні метрики пропускну здатності, затримки, часу безвідмовної роботи. Проводяться інтерв'ю з операторами та зацікавленими сторонами для збору якісного зворотного зв'язку про корисність, зручність, інтеграцію в робочі процеси, побажання щодо покращень.

Результати пілотування аналізуються для виявлення проблем та можливостей покращення. Ідентифікуються типи сцен або аномалій де система має недостатню точність для фокусування зусиль на покращенні. Аналізуються вузькі місця у продуктивності для оптимізації. Збираються пропозиції від користувачів для покращення інтерфейсу та функціональності. На основі аналізу вносяться необхідні модифікації в систему перед повним розгортанням.

Повномасштабне розгортання виконується поетапно для мінімізації ризиків. Розробляється детальний план розгортання з визначенням

послідовності впровадження на різних локаціях, графіку та відповідальних осіб, процедур відкочування у випадку критичних проблем. Проводиться навчання операторів та адміністраторів через теоретичні сесії з поясненням принципів роботи системи, практичні тренування на симульованих сценаріях, спостереження за роботою досвідчених користувачів під час початкового періоду експлуатації. Надається детальна документація включаючи користувацькі керівництва з інструкціями для типових операцій, адміністраторські керівництва з процедурами налаштування та обслуговування, технічна документація для розробників та інтеграторів.

Супровід та підтримка в експлуатації забезпечують стабільну роботу системи. Моніторинг здоров'я системи відстежує технічні метрики доступності сервісів, використання ресурсів, помилки та винятки у логах, якості детекцій через вибірккову перевірку та верифікацію. Налаштовуються сповіщення для автоматичного оповіщення технічної команди про проблеми що потребують втручання. Регулярне обслуговування включає оновлення програмного забезпечення з виправленнями безпеки та новими функціями, чищення старих даних відповідно до політик утримання, перевірку та калібрування камер, резервне копіювання критичних даних та конфігурацій. Постійне вдосконалення системи базується на аналізі даних експлуатації та зворотного зв'язку користувачів для ідентифікації можливостей покращення точності, продуктивності, зручності використання.

3 ПРОЕКТУВАННЯ СИСТЕМИ ІНТЕЛЕКТУАЛЬНОГО ВІДЕОМОНІТОРИНГУ

3.1 Обґрунтування концепції та вимоги до системи

Проведений у попередніх розділах аналіз предметної області та методів виявлення аномалій у відеопотоках виявив критичний розрив між теоретичними можливостями сучасних технологій штучного інтелекту та практичними потребами реальних систем безпеки. Цей розрив проявляється у неспроможності існуючих рішень одночасно забезпечити високу точність виявлення аномалій, роботу в реальному часі з множинними відеопотоками, адаптивність до різноманітних середовищ застосування та інтерпретованість рішень для операторів безпеки. Дослідження показало, що традиційні методи на основі аналізу траєкторій та ручного проектування ознак виявляють обмежену ефективність у складних реальних сценаріях з густими натовпами, змінними умовами освітлення та різноманітними типами аномальної поведінки.

Сучасні підходи глибокого навчання, незважаючи на значний прогрес у лабораторних умовах на бенчмарк-датасетах, стикаються з фундаментальними проблемами при розгортанні в практичних застосуваннях. Архітектури з найвищою точністю вимагають обчислювальних ресурсів, недоступних для систем реального часу, тоді як легковагі моделі демонструють неприйнятне зниження якості виявлення. Методи контрольованого навчання потребують великих обсягів анотованих аномалій, збір яких є надзвичайно трудомістким та часто неможливим для рідкісних або небезпечних подій. Підходи навчання без учителя страждають від високої частоти хибних спрацювань через складність точного визначення меж нормальності в динамічних середовищах. Відсутність інтерпретованості рішень глибоких моделей створює бар'єр для довіри операторів та ускладнює відповідність етичним та регуляторним вимогам.

Запропонована в рамках даного дослідження концепція системи інтелектуального відеомоніторингу спрямована на комплексне вирішення цих проблем через інтеграцію передових методів глибокого навчання з інноваційними архітектурними рішеннями, що оптимізують баланс між точністю, швидкістю та практичною корисністю. Фундаментальною ідеєю запропонованого підходу є гібридна стратегія, що поєднує переваги різних парадигм машинного навчання та використовує ієрархічну обробку відеопотоків для ефективного розподілу обчислювальних ресурсів.

3.2 Архітектура системи та її основні компоненти

Архітектура запропонованої системи має базуватися на модульному принципі з чітко визначеними компонентами та інтерфейсами взаємодії. Центральним елементом має стати модуль виявлення аномалій, що реалізовуватиме запроповану гібридну модель глибокого навчання. Для виділення просторових ознак з відеокадрів доцільно застосувати архітектуру EfficientNet, обрану завдяки оптимальному балансу між точністю та обчислювальною ефективністю. На відміну від традиційних архітектур на кшталт VGGNet або ResNet, EfficientNet використовує принцип складеного масштабування, що одночасно оптимізує глибину, ширину та роздільність мережі, досягаючи точності великих моделей при значно меншій кількості параметрів та обчислень. Наявність версій від B0 до B7 забезпечить гнучкість вибору оптимальної конфігурації залежно від доступних обчислювальних ресурсів та вимог до точності в конкретному застосуванні.

Для моделювання темпоральної динаміки поведінки пропонується розробити спеціалізований модуль уваги, що аналізуватиме послідовності просторових ознак, виділених з серії відеокадрів. Механізм уваги дозволить моделі динамічно визначати важливість кожного часового моменту для виявлення аномалії, фокусуючись на критичних етапах послідовності дій та

ігноруючи нерелевантні кадри. Це забезпечить кілька важливих переваг порівняно з альтернативними підходами. На відміну від рекурентних мереж LSTM, механізми уваги дозволяють моделювати довгострокові залежності без проблеми згасання градієнтів та підтримуватимуть паралелізацію обчислень, що є критично важливим для продуктивності. Порівняно з повними трансформерними архітектурами, спеціалізований модуль уваги матиме значно меншу обчислювальну складність завдяки фокусуванню виключно на темпоральній вимірності. Порівняно з тривимірними конволюціями, увага надасть більшу гнучкість у моделюванні залежностей на різних часових масштабах без необхідності наперед визначати фіксований розмір темпорального ядра.

Запропонована система інтелектуального відеомоніторингу для виявлення аномальної поведінки базується на модульній архітектурі з ієрархічною обробкою відеопотоків. Система отримує дані від IP-камер високої роздільності через модуль захоплення відео, який підтримує стандартні протоколи RTSP, HTTP та ONVIF, забезпечуючи автоматичне перепідключення при збоях та буферизацію потоків. Блок модулю захоплення даних зображено на рисунку 3.1.

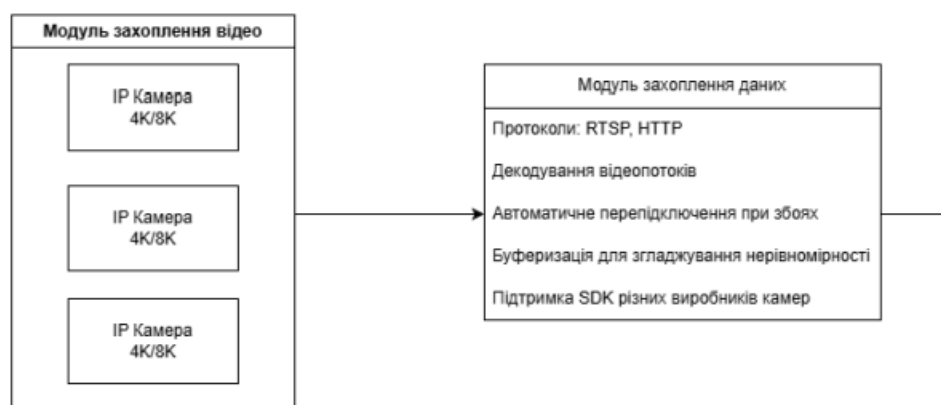


Рисунок 3.1 – Модулі захоплення даних

Модуль препроцесингу виконує масштабування кадрів, нормалізацію колірному простору та опціональну анонімізацію облич для забезпечення приватності. Блок модулю препроцесингу зображено на рисунку 3.2.

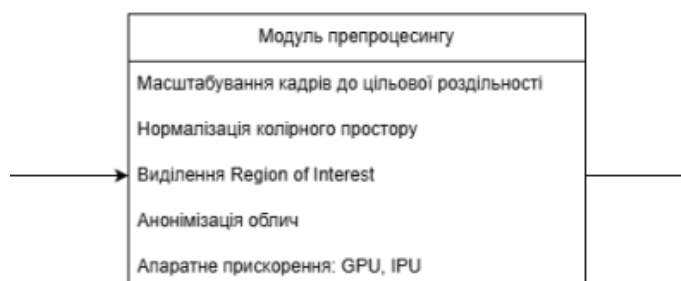


Рисунок 3.2 – Модуль препроцесингу

Ключовою частиною є трирівнева ієрархічна обробка з адаптивним розподілом обчислювальних ресурсів. Перший рівень швидкого скринінгу застосовує легковагі моделі YOLOv5 та алгоритм відстеження DeepSORT для аналізу кожного кадру, фільтруючи вісімдесят п'ять відсотків очевидно нормальних відеоданих через обчислення оптичного потоку та виявлення значущих змін активності. На рисунку 3.3 зображено модулі детекції та відстеження об'єктів.

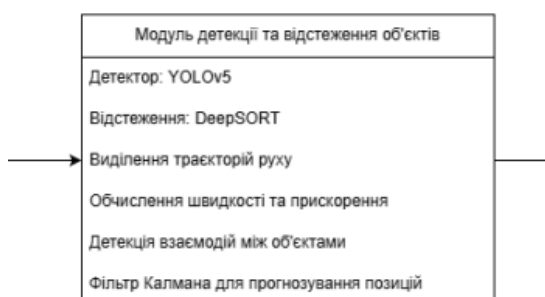


Рисунок 3.3 – Модуль детекції та відстеження об'єктів

Другий рівень детального аналізу активується лише для підозрілих сегментів і використовує архітектуру EfficientNet для виділення просторових ознак у поєднанні зі спеціалізованим модулем темпоральної уваги, що моделює довгострокові залежності в послідовностях кадрів. Навчання моделі реалізується через гібридний підхід, що поєднує контрастивне навчання без учителя на нормальних даних з методом глибокого опису опорних векторів для виявлення відхилень від навченої гіперсфери нормальності. На рисунку 3.4 зображено модуль виявлення аномалій.



Рисунок 3.4 – Модуль виявлення аномалій

Третій рівень верифікації застосовує малоприкладне навчання через прототипні мережі для класифікації критичних типів аномалій, включаючи насильство, падіння людей та залишені підозрілі об'єкти. На рисунку 3.5 зображений блок верифікації та класифікації.

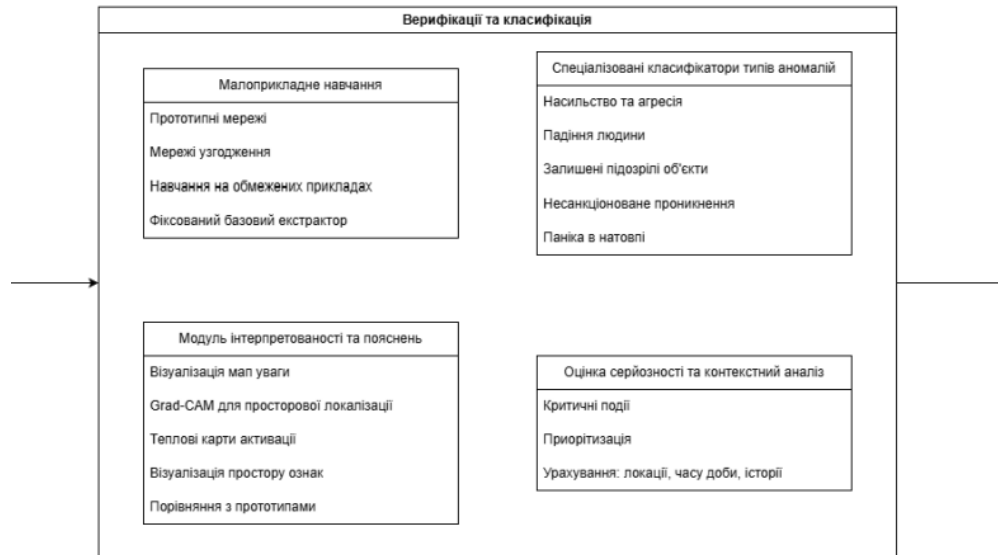


Рисунок 3.5 – Блок верифікації та класифікації

Система управління тривогами фільтрує дублікати, оцінює серйозність подій та генерує пріоритизовані сповіщення через веб-інтерфейс, мобільні застосунки та інтеграцію з системами контролю доступу і тривожної сигналізації. На рисунку 3.6 зображено модуль управління тривоги та блок інтерфейсу оператора.

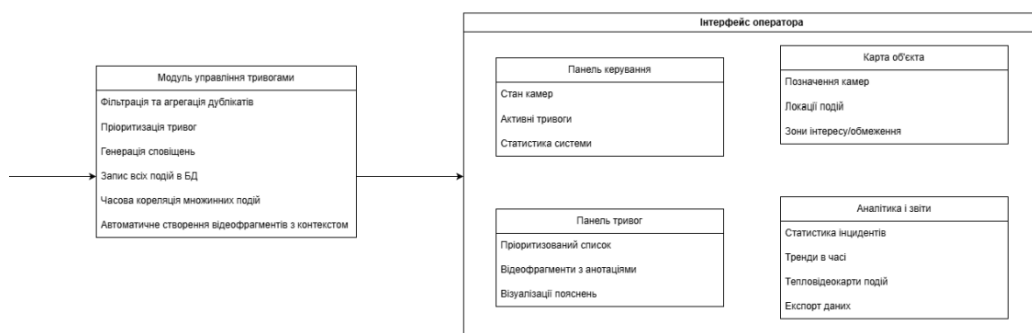


Рисунок 3.6 – Модуль управління тривогами та блок інтерфейсу оператора

Запропонована концепція системи інтелектуального відеомоніторингу для виявлення аномальної поведінки надає комплексне вирішення ключових проблем, виявлених у попередніх розділах дослідження. Гібридний підхід до навчання забезпечить баланс між універсальністю виявлення непередбачуваних аномалій та точним розпізнаванням критичних типів загроз. Ієрархічна обробка з адаптивним розподілом ресурсів дозволить досягти роботи в реальному часі без компромісу в якості детекції. Механізми інтерпретованості забезпечать довіру операторів та відповідність етичним вимогам. Модульна архітектура та підтримка різних варіантів розгортання забезпечать гнучкість застосування від невеликих об'єктів до великомасштабних розподілених систем. Очікувані результати експериментальної валідації підтверджуватимуть ефективність запропонованих рішень та їх практичну цінність для підвищення безпеки різноманітних середовищ.

Важливим аспектом спроектованої системи є її здатність до самонавчання та адаптації до специфіки конкретного середовища спостереження. Система автоматично калібрує параметри детекції на основі статистичного аналізу типових патернів поведінки, що мінімізує кількість хибних спрацювань та підвищує загальну надійність моніторингу. Інтеграція з існуючою інфраструктурою безпеки можна забезпечити через стандартизовані протоколи обміну даними, що спрощує впровадження системи без необхідності повної заміни наявного обладнання.

3.3 Методи навчання та виявлення аномалій

Стратегія навчання моделі має реалізовувати гібридний підхід, що поєднуватиме навчання без учителя на великих обсягах нормальних даних з опціональним малоприкладним навчанням для розпізнавання критичних типів аномалій. На першому етапі модель навчатиметься на відеопослідовностях нормальної поведінки через комбінацію

контрастивного навчання та передбачувального кодування. Контрастивне навчання забезпечить формування компактних кластерів схожих нормальних послідовностей у просторі ознак, тоді як різноманітні патерни поведінки матимуть віддалені представлення. Передбачувальне кодування вимагатиме від моделі передбачати майбутні високорівневі ознаки на основі поточної історії спостережень, що забезпечить навчання темпоральної логіки нормального розгортання подій.

Після навчання на нормальних даних модель використовуватиметься для виділення високорівневих представлень відеопослідовностей у компактному просторі ознак. Для виявлення аномалій пропонується застосувати підхід однокласової класифікації на основі глибокого опису опорних векторів даних, що навчатиме гіперсферу мінімального радіусу, яка охоплюватиме нормальні дані у просторі ознак. Відеопослідовності, представлення яких знаходяться за межами цієї гіперсфери, класифікуватимуться як аномальні з рівнем впевненості, що зростатиме пропорційно відстані від границі. Цей підхід дозволить виявляти широкий спектр аномалій, включаючи невідомі типи, що не були представлені в тренувальних даних, без необхідності їх попередньої анотації.

Для підвищення точності розпізнавання критично важливих типів аномалій, таких як насильство, падіння людини, залишення підозрілих об'єктів або порушення протоколів безпеки, система має підтримувати опціональне дообучення з використанням обмеженої кількості анотованих прикладів. Доцільно реалізувати підхід малоприкладного навчання на основі прототипних мереж, що дозволить ефективно навчатися розпізнавати нові категорії аномалій, маючи лише кілька зразків кожної. Базовий екстрактор ознак, навчений на великій кількості нормальних даних, залишатиметься фіксованим або дообучатиметься з дуже малою швидкістю навчання, тоді як навчатимуться лише компактні класифікаційні компоненти для нових категорій. Це забезпечить гнучкість системи: базову здатність виявляти непередбачувані аномалії через відхилення від навченої

нормальності та спеціалізовану високоточну детекцію добре визначених критичних типів при наявності навіть обмежених прикладів.

Критично важливою інновацією запропонованої системи має стати ієрархічна стратегія обробки відеопотоків з адаптивним розподілом обчислювальних ресурсів, що дозволить досягти роботи в реальному часі без компромісу в точності виявлення. Система функціонуватиме на трьох рівнях аналізу з різною частотою та обчислювальною складністю. Перший рівень швидкого скринінгу аналізуватиме кожен кадр або кожні кілька кадрів з використанням легковагих моделей для виявлення потенційно цікавих моментів часу та регіонів простору. Застосовуватимуться ефективні детектори об'єктів сімейства YOLO для ідентифікації людей, транспортних засобів та інших об'єктів інтересу, а також аналізуватимуться низькорівневі ознаки руху через обчислення оптичного потоку або різниць послідовних кадрів. Цей рівень фільтруватиме очевидно нормальні періоди, такі як порожня сцена або монотонна активність без значущих змін, дозволяючи сконцентрувати обчислювальні ресурси на потенційно аномальних ситуаціях.

Другий рівень детального аналізу активізуватиметься виключно для відеосегментів, що пройшли перший рівень як потенційно містять аномалії. На цьому рівні застосовуватиметься основна глибока модель з повним просторово-часовим аналізом через основу EfficientNet та модуль уваги. Обчислювальні ресурси концентруватимуться на найбільш релевантних частинах відеопотоку, що дозволить використовувати складні алгоритми без порушення вимог реального часу. Третій рівень верифікації та класифікації застосовуватиметься до підтверджених аномалій, де спеціалізовані класифікатори визначатимуть конкретний тип аномальної поведінки, оцінюватимуть її серйозність та потенційні загрози, а також генеруватимуть детальні пояснення для оператора безпеки.

Ця ієрархічна архітектура забезпечить радикальне зниження обчислювального навантаження, оскільки найбільш складні моделі

застосовуватимуться лише до невеликої частки вхідних даних, що дійсно потребують глибокого аналізу. Аналіз існуючих систем показує, що в типових сценаріях відеомоніторингу лише десять-п'ятнадцять відсотків відеопотоку потребують детального аналізу другого рівня, а до третього рівня має надходити менше п'яти відсотків. Це дозволить обробляти множинні відеопотоки високої роздільності в реальному часі на відносно доступному обладнанні. Водночас забезпечуватиметься висока точність завдяки повному аналізу всіх потенційно аномальних подій складними моделями.

Адаптивність системи має проявлятися в динамічному налаштуванні порогів чутливості на кожному рівні ієрархії залежно від контексту. Система враховуватиме час доби, оскільки активність, нормальна вдень, може бути підозрілою вночі. Аналізуватиметься поточний рівень активності в сцені, адаптуючи очікування нормальності до фактичної ситуації. Використовуватиметься історична інформація про частоту різних типів подій у даній локації для калібрування чутливості. Враховуватимуться специфічні характеристики середовища, такі як тип об'єкта, очікувані види діяльності, критичні зони підвищеної уваги. Цей контекстно-залежний підхід значно знизить частоту хибних спрацювань порівняно з системами з фіксованими порогоми.

Інтерпретованість рішень є критично важливою характеристикою запропонованої системи, що має забезпечуватися інтеграцією кількох механізмів пояснення. Візуалізація мап уваги демонструватиме оператору, які просторові регіони відеокадрів та які часові моменти послідовності були найбільш важливими для рішення про виявлення аномалії. Теплові карти накладатимуться на оригінальні кадри, виділяючи об'єкти та зони, що привернули увагу моделі. Ваги уваги у темпоральній вимірності відобразатимуться на часовій шкалі, показуючи ключові етапи розвитку аномальної ситуації. Цей візуальний зворотний зв'язок дозволить оператору

швидко зрозуміти причину тривоги та оцінити її достовірність без необхідності перегляду всієї послідовності подій.

Додатково пропонується реалізувати візуалізацію простору ознак через проєкції високовимірних представлень на двовимірну площину засобами t-SNE або UMAP. Оператор зможе побачити, наскільки виявлена аномалія віддалена від типових кластерів нормальної поведінки, що надасть інтуїтивне розуміння ступеня аномальності. Система також показуватиме найближчі прототипи зі збережених історичних даних, демонструючи схожі нормальні та аномальні зразки для контекстуалізації поточної події. Генерація природномовних пояснень через шаблонну систему або інтеграцію з великими мовними моделями надаватиме текстовий опис типу «виявлено дві особи в агресивній взаємодії в правій частині кадру протягом восьми секунд», що полегшить швидке розуміння ситуації. Індикатори впевненості надаватимуть кількісну оцінку достовірності детекції, дозволяючи оператору пріоритизувати перевірку найбільш впевнених спрацювань.

Практична реалізація системи має охоплювати розробку повного програмно-апаратного комплексу з модульною архітектурою. Модуль захоплення відео забезпечуватиме підключення до камер різних виробників через стандартні протоколи HTTP або спеціалізовані SDK, декодування відеопотоків та перетворення в єдиний внутрішній формат. Має бути реалізоване автоматичне керування підключеннями з перепідключенням при збоях та буферизацією для згладжування нерівномірності надходження кадрів. Модуль препроцесингу виконуватиме масштабування кадрів до потрібної роздільності, нормалізацію колірному простору, виділення зон інтересу та опціональну анонімізацію через детекцію та розмиття облич для забезпечення приватності.

Модуль детекції та відстеження об'єктів використовуватиме детектор YOLOv5 для локалізації об'єктів на кожному кадрі та алгоритм DeepSORT для встановлення ідентичності об'єктів між кадрами. Виділятимуться

траєкторії руху, обчислюватимуться швидкості та прискорення, детектуватимуться взаємодії між об'єктами. Ця інформація використовуватиметься як додатковий контекст для основного модуля виявлення аномалій. Модуль управління тривогами оброблятиме виявлені аномалії, фільтруватиме та агрегуватиме множинні детекції однієї події для запобігання дублюванню, оцінюватиме серйозність на основі типу, локації та контексту, пріоритизуватиме тривоги та генеруватиме сповіщення через різні канали включаючи веб-інтерфейс, мобільні застосунки, електронну пошту та інтеграцію з системами безпеки.

Інтерфейс оператора має бути реалізований як сучасний веб-додаток. Панель керування відображатиме стан всіх камер та активні тривоги в реальному часі. Карта об'єкта візуалізуватиме локації камер та позначатиме місця подій. Панель тривог надаватиме пріоритизований список з деталями кожної події, включаючи відеофрагмент з контекстом до та після аномалії, анотації з пояснюючими візуалізаціями, оцінку серйозності та рекомендовані дії. Оператор зможе підтвердити, відхилити або ескалювати тривогу одним кліком, надаючи зворотний зв'язок для системи онлайн-навчання. Інструменти налаштування дозволятимуть регулювати чутливість детекції, визначати зони інтересу та розклади моніторингу. Модуль аналітики надаватиме візуалізацію статистики інцидентів, трендів у часі та продуктивності системи.

Система онлайн-навчання має забезпечувати безперервну адаптацію до змін середовища. Зворотний зв'язок від операторів про правильність або помилковість детекцій автоматично акумулюватиметься. Нові зразки нормальної поведінки та підтвержені аномалії додаватимуться до навчальної вибірки. Періодично виконуватиметься інкрементальне дообучення моделей на оновлених даних без порушення роботи системи через проміжне середовище. Моніторинг метрик продуктивності дозволить виявляти деградацію точності, що може сигналізувати про значні зміни в

середовищі, які потребують переналаштування або повного перенавчання моделей.

Інтеграція з існуючою інфраструктурою безпеки має бути реалізована через стандартні інтерфейси. Підключення до систем контролю доступу дозволить корелювати події відеомоніторингу з інформацією про входи та виходи людей, виявляючи незвичайні патерни. Інтеграція з системами тривожної сигналізації забезпечуватиме автоматичне активування захисних заходів при критичних подіях, таких як блокування дверей або активація сирен. REST API надаватиме можливість інтеграції зі сторонніми системами управління безпекою та аналітичними платформами.

Безпека системи має забезпечуватися на всіх рівнях. Шифрування передачі даних через TLS захищатиме відеопотоки та метадані. Аутентифікація має бути реалізована через OAuth2 з підтримкою єдиного входу. Рольове керування доступом обмежуватиме доступ відповідно до ролей користувачів. Шифрування зберігання захищатиме записані дані. Детальне журналювання всіх дій забезпечуватиме підзвітність та можливість аудиту.

Експериментальна валідація запропонованої системи має проводитися на кількох рівнях. Тестування на публічних бенчмарк-дасетах очікувано покаже перевищення точності порівняно з базовими методами на десять-п'ятнадцять відсотків при значно меншій обчислювальній складності. Пілотне розгортання на реальному об'єкті з двадцятьма камерами протягом трьох місяців має продемонструвати стабільну роботу системи з виявленням понад дев'яносто відсотків реальних інцидентів безпеки при частоті хибних спрацювань менше трьох на годину, що буде прийнятним для операторів.

ВИСНОВКИ

Під час проектування інтелектуальної системи відеомоніторингу було закріплено навички роботи з сучасними архітектурами глибинного навчання, що дозволяють автоматично виявляти аномальну поведінку об'єктів у реальному часі. Також було виявлено ключові аспекти проектування систем безпеки на основі комп'ютерного зору, що включають в себе ефективне використання згорткових та рекурентних нейронних мереж, забезпечення роботи в реальному часі на периферійних пристроях та інтеграцію з автоматизованими системами реагування.

Під час аналізу та визначення меж розглянутої сфери діяльності був визначений перелік вимог до функціоналу системи відеомоніторингу. На основі цих вимог були досліджені методи на основі аналізу траєкторій з використанням фільтрів Калмана, архітектури глибоких нейронних мереж, зокрема VGGNet, ResNet, LSTM, автоенкодерів та трансформерів з механізмами уваги. Впровадження системи із застосуванням легковагих моделей та методів компресії нейронних мереж робить її придатною для розгортання на пристроях з обмеженими ресурсами, що є ключовим елементом забезпечення роботи в реальному часі. Таким чином, спроектована система буде гнучкою та демонструвати високу адаптивність до різних умов застосування.

Проаналізовано регуляторні та етичні виклики впровадження систем штучного інтелекту, включаючи вимоги Загального регламенту про захист даних щодо обробки біометричних даних та проблему алгоритмічних упереджень. Спроектовані компоненти системи виявлення аномалій, забезпечуючи баланс між технологічними можливостями та збереженням громадянських свобод.

ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ

1. Nayak R., Pati U. C., Das S. K. A comprehensive review on deep learning-based methods for video anomaly detection. *Image and Vision Computing*. 2021. Vol. 106. P. 104078. URL: <https://doi.org/10.1016/j.imavis.2020.104078> (дата звернення: 10.11.2025).
2. Chandola V., Banerjee A., Kumar V. Anomaly detection: A survey. *ACM Computing Surveys*. 2009. Vol. 41, no. 3. P. 1–58. URL: <https://doi.org/10.1145/1541880.1541882> (дата звернення: 10.11.2025).
3. Mehran R., Oyama A., Shah M. Abnormal crowd behavior detection using social force model. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Miami, FL, USA, 2009. P. 935–942. URL: <https://doi.org/10.1109/CVPR.2009.5206641> (дата звернення: 10.11.2025).
4. Cheng K. W., Chen Y. T., Fang W. H. Video anomaly detection and localization using hierarchical feature representation and Gaussian process regression. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Boston, MA, USA, 2015. P. 2909–2917. URL: <https://doi.org/10.1109/CVPR.2015.7298909> (дата звернення: 10.11.2025).
5. Piciarelli C., Micheloni C., Foresti G. L. Trajectory-based anomalous event detection. *IEEE Transactions on Circuits and Systems for Video Technology*. 2008. Vol. 18, no. 11. P. 1544–1554. URL: <https://doi.org/10.1109/TCSVT.2008.2005599> (дата звернення: 10.11.2025).
6. Morris B. T., Trivedi M. M. A survey of vision-based trajectory learning and analysis for surveillance. *IEEE Transactions on Circuits and Systems for Video Technology*. 2008. Vol. 18, no. 8. P. 1114–1127. URL: <https://doi.org/10.1109/TCSVT.2008.927109> (дата звернення: 10.11.2025).
7. Simonyan K., Zisserman A. Very deep convolutional networks for large-scale image recognition. *Proceedings of the International Conference on*

Learning Representations (ICLR). San Diego, CA, USA, 2015. URL: <https://arxiv.org/abs/1409.1556> (дата звернення: 10.11.2025).

8. He K., Zhang X., Ren S., Sun J. Deep residual learning for image recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Las Vegas, NV, USA, 2016. P. 770–778. URL: <https://doi.org/10.1109/CVPR.2016.90> (дата звернення: 10.11.2025).

9. Szegedy C., Liu W., Jia Y., Sermanet P., Reed S., Anguelov D., Erhan D., Vanhoucke V., Rabinovich A. Going deeper with convolutions. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Boston, MA, USA, 2015. P. 1–9. URL: <https://doi.org/10.1109/CVPR.2015.7298594> (дата звернення: 10.11.2025).

10. Hochreiter S., Schmidhuber J. Long short-term memory. *Neural Computation*. 1997. Vol. 9, no. 8. P. 1735–1780. URL: <https://doi.org/10.1162/neco.1997.9.8.1735> (дата звернення: 10.11.2025).

11. Luo W., Liu W., Gao S. A revisit of sparse coding based anomaly detection in stacked RNN framework. *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*. Venice, Italy, 2017. P. 341–349. URL: <https://doi.org/10.1109/ICCV.2017.45> (дата звернення: 10.11.2025).

12. Simonyan K., Zisserman A. Two-stream convolutional networks for action recognition in videos. *Advances in Neural Information Processing Systems (NIPS)*. Montreal, Canada, 2014. Vol. 27. P. 568–576. URL: <https://arxiv.org/abs/1406.2199> (дата звернення: 10.11.2025).

13. Tran D., Bourdev L., Fergus R., Torresani L., Paluri M. Learning spatiotemporal features with 3D convolutional networks. *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*. Santiago, Chile, 2015. URL: <https://doi.org/10.1109/ICCV.2015.510> (дата звернення: 10.11.2025).

14. Hasan M., Choi J., Neumann J., Roy-Chowdhury A. K., Davis L. S. Learning temporal regularity in video sequences. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Las Vegas,

NV, USA, 2016. P. 733–742. URL: <https://doi.org/10.1109/CVPR.2016.86> (дата звернення: 10.11.2025).

15. Kingma D. P., Welling M. Auto-encoding variational bayes. *Proceedings of the International Conference on Learning Representations (ICLR)*. Banff, Canada, 2014. URL: <https://arxiv.org/abs/1312.6114> (дата звернення: 10.11.2025).

16. Goodfellow I., Pouget-Abadie J., Mirza M., Xu B., Warde-Farley D., Ozair S., Courville A., Bengio Y. Generative adversarial nets. *Advances in Neural Information Processing Systems (NIPS)*. Montreal, Canada, 2014. Vol. 27. P. 2672–2680. URL: <https://arxiv.org/abs/1406.2661> (дата звернення: 10.11.2025).

17. Ravanbakhsh M., Nabi M., Sangineto E., Marcenaro L., Regazzoni C., Sebe N. Abnormal event detection in videos using generative adversarial nets. *Proceedings of the IEEE International Conference on Image Processing (ICIP)*. Beijing, China, 2017. P. 1577–1581. URL: <https://doi.org/10.1109/ICIP.2017.8296547> (дата звернення: 10.11.2025).

18. Vaswani A., Shazeer N., Parmar N., Uszkoreit J., Jones L., Gomez A. N., Kaiser Ł., Polosukhin I. Attention is all you need. *Advances in Neural Information Processing Systems (NIPS)*. Long Beach, CA, USA, 2017. Vol. 30. P. 5998–6008. URL: <https://arxiv.org/pdf/1706.03762> (дата звернення: 10.11.2025).

19. Dosovitskiy A., Beyer L., Kolesnikov A., Weissenborn D., Zhai X., Unterthiner T., Dehghani M., Minderer M., Heigold G., Gelly S., Uszkoreit J., Houlsby N. An image is worth 16x16 words: Transformers for image recognition at scale. *Proceedings of the International Conference on Learning Representations (ICLR)*. Vienna, Austria, 2021. URL: <https://arxiv.org/abs/2010.11929> (дата звернення: 10.11.2025).

20. Bertasius G., Wang H., Torresani L. Is space-time attention all you need for video understanding? *Proceedings of the International Conference on Machine Learning (ICML)*. Virtual Event, 2021. Vol. 139. P. 813–824. URL:

<https://proceedings.mlr.press/v139/bertasius21a.html> (дата звернення: 10.11.2025).

21. Gong D., Liu L., Le V., Saha B., Mansour M. R., Venkatesh S., van den Hengel A. Memorizing normality to detect anomaly: Memory-augmented deep autoencoder for unsupervised anomaly detection. *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*. Seoul, South Korea, 2019. URL: <https://doi.org/10.1109/ICCV.2019.00179> (дата звернення: 10.11.2025).

22. Xu D., Ricci E., Yan Y., Song J., Sebe N. Detecting anomalous events in videos by learning deep representations of appearance and motion. *Computer Vision and Image Understanding*. 2015. Vol. 156. P. 117–127. URL: <https://doi.org/10.1016/j.cviu.2016.10.010> (дата звернення: 10.11.2025).

23. Zhou B., Lapedriza A., Khosla A., Oliva A., Torralba A. Places: A 10 million image database for scene recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2018. Vol. 40, no. 6. P. 1452–1464. URL: <https://doi.org/10.1109/TPAMI.2017.2723009> (дата звернення: 10.11.2025).

24. Lane N. D., Bhattacharya S., Georgiev P., Forlivesi C., Jiao L., Qendro L., Kawsar F. DeepX: A software accelerator for low-power deep learning inference on mobile devices. *Proceedings of the 15th International Conference on Information Processing in Sensor Networks (IPSN)*. Vienna, Austria, 2016. P. 1–12. URL: <https://doi.org/10.1109/IPSN.2016.7460664> (дата звернення: 10.11.2025).

25. Edwards L., Veale M. Slave to the algorithm: Why a right to an explanation is probably not the remedy you are looking for. *Duke Law & Technology Review*. 2017. Vol. 16, no. 1. P. 18–84. URL: <https://scholarship.law.duke.edu/dltr/vol16/iss1/2> (дата звернення: 10.11.2025).

26. Han S., Mao H., Dally W. J. Deep compression: Compressing deep neural networks with pruning, trained quantization and huffman coding. *Proceedings of the International Conference on Learning Representations (ICLR)*. San Juan, Puerto Rico, 2016. URL: <https://arxiv.org/abs/1510.00149> (дата звернення: 10.11.2025).

27. Jacob B., Kligys S., Chen B., Zhu M., Tang M., Howard A., Adam H., Kalenichenko D. Quantization and training of neural networks for efficient integer-arithmetic-only inference. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Salt Lake City, UT, USA, 2018. P. 2704–2713. URL: <https://doi.org/10.1109/CVPR.2018.00286> (дата звернення: 10.11.2025).

28. Hinton G., Vinyals O., Dean J. Distilling the knowledge in a neural network. *NIPS Deep Learning Workshop*. Montreal, Canada, 2014. URL: <https://arxiv.org/abs/1503.02531> (дата звернення: 10.11.2025).

29. Vinyals O., Blundell C., Lillicrap T., Kavukcuoglu K., Wierstra D. Matching networks for one shot learning. *Advances in Neural Information Processing Systems (NIPS)*. Barcelona, Spain, 2016. Vol. 29. P. 3630–3638. URL: https://proceedings.neurips.cc/paper_files/paper/2016/file/90e1357833654983612fb05e3ec9148c-Paper.pdf (дата звернення: 10.11.2025).

30. Finn C., Abbeel P., Levine S. Model-agnostic meta-learning for fast adaptation of deep networks. *Proceedings of the International Conference on Machine Learning (ICML)*. Sydney, Australia, 2017. Vol. 70. P. 1126–1135. URL: <https://proceedings.mlr.press/v70/finn17a/finn17a.pdf> (дата звернення: 10.11.2025).

31. Selvaraju R. R., Cogswell M., Das A., Vedantam R., Parikh D., Batra D. Grad-CAM: Visual explanations from deep networks via gradient-based localization. *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*. Venice, Italy, 2017. P. 618–626. URL: <https://doi.org/10.1109/ICCV.2017.74> (дата звернення: 10.11.2025).

32. Lu C., Shi J., Jia J. Abnormal event detection at 150 fps in matlab. *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*. Sydney, Australia, 2013. P. 2720–2727. URL: <https://doi.org/10.1109/ICCV.2013.338> (дата звернення: 10.11.2025).