

УДК 004.8

Муляр Е.Р., Багрій Р.О., Пасічник О.А., Манзюк Е.А.

*Хмельницький національний університет*

## **МЕТОД ВИЯВЛЕННЯ ОЗНАК НАСИЛЬСТВА У ВІДЕОМАТЕРІАЛАХ НЕЙРОМЕРЕЖЕВИМИ ЗАСОБАМИ**

*Проблема виявлення насильства у відеопотоці є актуальною в сучасному світі, де зростає кількість відеоматеріалів з насильницькими сценами. Виявлення та реагування на такі сцени мають велике значення для забезпечення безпеки громадських просторів та захисту прав людини. Запропоновано метод виявлення ознак насильства у відеоматеріалах нейромережевими засобами. Метод відкриває шлях до розв'язання завдань виявлення ознак насильства на відеоматеріалах у реальному часі, а також фільтрації контенту на потокових мультимедійних платформах.*

*The problem of detecting violence in a video stream is relevant in the modern world, where the number of videos with violent scenes is growing. Detecting and responding to such scenes is important for ensuring the safety of public spaces and protecting human rights. The paper proposes a method for detecting signs of violence in video materials using neural network tools. The method opens the way to solving the problems of detecting signs of violence in real-time video materials, as well as filtering content on streaming multimedia platforms.*

### **Вступ**

Сьогодні для протидії такій суспільній проблемі як насильство розпочали активно використовувати системи відеоспостережень. Такі країни, як Китай та Південна Корея, є лідерами у встановленні камер відеоспостереження, і результати використання цих систем є вражаючими. У Китаї рівень насильства у громадських місцях знизився на 60%, а в Південній Кореї - на приблизно 50% [1]. Однак, ці системи мають деякі недоліки. Основною проблемою є людський фактор, зокрема неуважність та недбалість операторів-спостерігачів. Зазвичай оператор може ефективно контролювати лише обмежену кількість камер відеоспостереження. Однак, коли кількість камер перевищує їх межі, оператори можуть допускати помилки або пропускати випадки насильства. Для вирішення цієї проблеми сьогодні активно використовуються передові інформаційні технології, зокрема штучні нейронні мережі. Завдяки цим технологіям, системи відеоспостереження можуть автоматично аналізувати великий обсяг відеоданих і виявляти потенційні випадки насильства. Штучний інтелект допомагає зменшити навантаження на операторів і забезпечує більш точне та ефективне виявлення подій [2].

### Аналіз існуючих публікацій

Для реалізації інтелектуального відеоспостереження сьогодні активно використовують підходи модифікації та поєднання різних методів та архітектур нейромереж [3].

Прикладом такого підходу є модель нейронних мереж BiConvLSTM (Bidirectional Convolutional LSTM). Ця модель поєднує в собі два потужних компоненти: двосторонню згорткову мережу (BiConv) та рекурентну нейронну мережу LSTM (Long Short-Term Memory). BiConvLSTM використовується для аналізу послідовних даних, таких як зображення або відео, з метою виявлення шаблонів та залежностей у цих даних [4].

Іншим прикладом використання нейронних мереж для задачі виявлення насильства у відеопотоці є 3D CNN (3D Convolutional Neural Network). Дана модель являється згортковою нейронною мережею, що використовується для обробки тривимірних даних, таких як відео, медичні зображення або тривимірні моделі [5].

Приклад використання моделі BiConvLSTM наведено у роботі [6], де розглянуто підхід до виявлення насильства у відео за допомогою методу «просторово-часовий кодер». «Просторово-часовий кодер» представляє собою метод який побудований на основі декількох архітектур: BiConvLSTM, VGG13 [7] для кодування кожного відеокадру як набір карт функцій. Ці карти функцій потім передаються до BiConvLSTM для подальшого кодування у часовому напрямку відео. Виконується поелементна максимізація кодувань для створення представлення відео, яке передається класифікатору для виявлення насильства. Щодо результатів роботи даного підходу, то для набору даних «Hockey fights» точність склала 96.54%, для набору даних «Violent flows» - 92.18%.

У роботі [8] запропоновано метод до виявлення насильства у відео за допомогою 3D-CNN. 3D-CNN спочатку обробляє кожний кадр відео, використовуючи набори фільтрів для виявлення важливих ознак, таких як рух, форма та колір. Потім 3D-CNN обробляє послідовність кадрів, використовуючи 3D-фільтри. Класифікація здійснюється шляхом застосування логістичної регресії до вихідного тензора 3D-CNN. На рахунок результатів роботи даного методу, то для набору даних «Hockey fights» точність склала 98.3%, для набору даних «Violent flows» - 97.17%.

В наведених публікаціях не розглянуто підходи до взаємодії запропонованих методів з відеоматеріалами у реальному часі, тому виникають питання в ефективності даних підходів при роботі з відеопотоком даних в реальному часі.

**Метою роботи** є розробка методу для виявлення ознак насильства у відеоматеріалах нейромережевими засобами. Метод повинен працювати як зі статичним відеоматеріалом (відеоролик) так і з динамічним (відеопотік в реальному часі).

## Метод виявлення ознак насильства у відеоматеріалах неймережевими засобами

Робота запропонованого методу полягає в отриманні ознак насильства з кадрів вхідного відео за допомогою згорткової нейронної мережі і визначення ступеню насильства у відсотковому відношенні у відеопотоці за допомогою SVM (методу опорних векторів). На рисунку 1 представлено архітектуру даного методу.

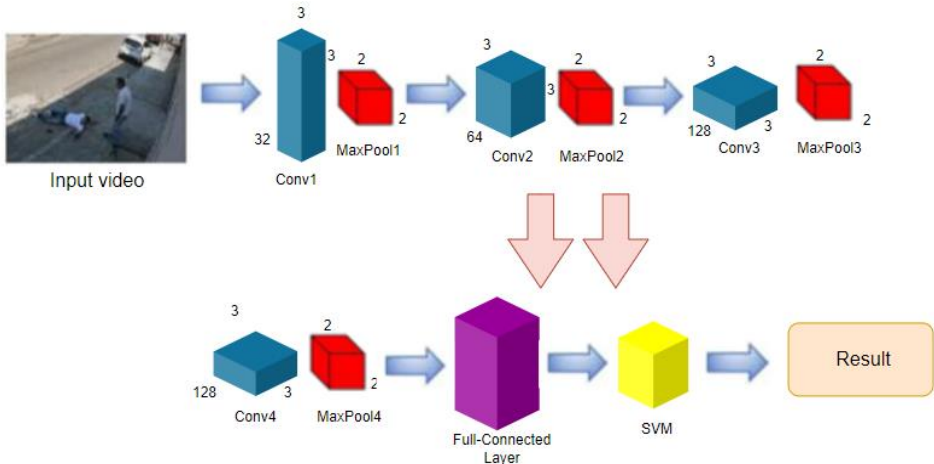


Рисунок 1 – Архітектура запропонованого методу

### Етап 1 – Отримання кадрів із вхідного відеоматеріалу

Необхідно розбити вхідний відеоматеріал на послідовність кадрів та перетворити кожен кадр у карту зображень.

### Етап 2 – Операція згортки

На даному етапі необхідно виконати операцію згортки на вхідне зображення для того, щоб отримати карту ознак. Для виконання даної операції використовуються фільтри (матриця параметрів). Для даної нейронної мережі було обрано 4 рівня фільтрів по 32, 64, 128, 128 фільтрів на кожному рівні відповідно. Формування карти ознак можна здійснити за допомогою наступної математичної формули:

$$M(i, j) = (K * X)(i, j) = \sum_m \sum_n K(m, n) X(i - m, j - n), \quad (1)$$

де  $M$  – елемент карти ознак з координатами  $i$  та  $j$ ,  $X$  – вхідне зображення,  $K$  – детектор ознак,  $(m, n)$  – розмірності детектора ознак.

### Етап 3 – Операція максимального об'єднання

Максимальне об'єднання являється операцією, яка об'єднує елементи в межах фільтра на карті ознак і вибирає найбільший елемент. Тобто, після проходження через шар максимального об'єднання, отримується нова карта ознак, яка містить найбільш помітні ознаки з попередньої карти ознак. Виконати дану операцію можна за допомогою наступної формули:

$$p(i, j) = \max_{i,j}(x(i - m, j - n)), \quad (2)$$

де  $p(i, j)$  – значення елемента поточного рівня з координатами  $i$  та  $j$ ,  $x$  – вхідні дані з попередніх рівнів,  $(m, n)$  – розмірність рецептивного поля.

### Етап 4 – Повнозв'язний рівень

Повнозв'язний рівень представляє собою модель багаторівневого перцептрона, де всі нейрони з наступного шару з'єднані з нейронами попереднього шару. Цей рівень використовуються на передостанньому етапі роботи мережі для підготовки результатів на виході мережі. На даному рівні виконується обчислення скалярного добутку даних та параметрів з додаванням зсуву.

### Етап 5 – Класифікація отриманих ознак за допомогою SVM

Метод опорних векторів (SVM) використовується для знаходження параметрів гіперплощини у великому чи нескінченному вимірному просторі, яка може служити для класифікації. Головна ідея полягає в тому, щоб знайти гіперплощину, яка має найбільшу відстань до найближчих навчальних точок будь-якого класу. На виході класифікатор SVM видає оцінку, яка представляє ймовірність того, що вхідні дані належать до певного класу (насилницького або не насилницького).

Для того, щоб запропонований метод міг працювати з відеоматеріалом в реальному часі під час процесу навчання нейронної мережі використовувався метод fine-tuning. Суть даного методу полягає в тому, що нейромережу спочатку навчають на наборі даних готових відео, а потім поступово додають до набору даних відеопотоки у реальному часі. Це допомагає нейромережі навчитися розпізнавати насильство в умовах шуму, швидкої зміни та різноманітності.

### Аналіз ефективності створеного методу

Для того, щоб оцінити ефективність роботи запропонованого методу, проведено декілька експериментів щодо двох наборів даних для виявлення насильства: «Hockey fights» [9], «Livestream» [10]. Оцінка ефективності методу базується на визначенні загальної точності для кожного набору даних.

Загальна точність для набору даних «Hockey fights» склала 98.5%, представлено на рисунку 2.

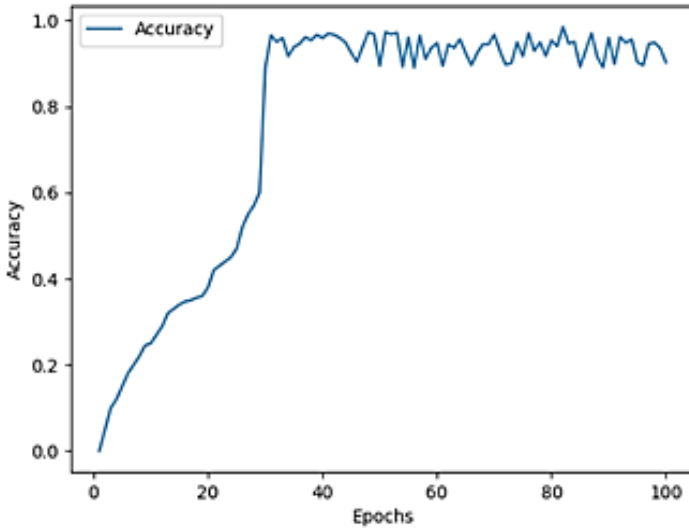


Рисунок 2 – Загальна точність набору даних «Hockey fights»

Загальна точність для набору даних «Livestream» склала 87.4%, представлено на рисунку 3.

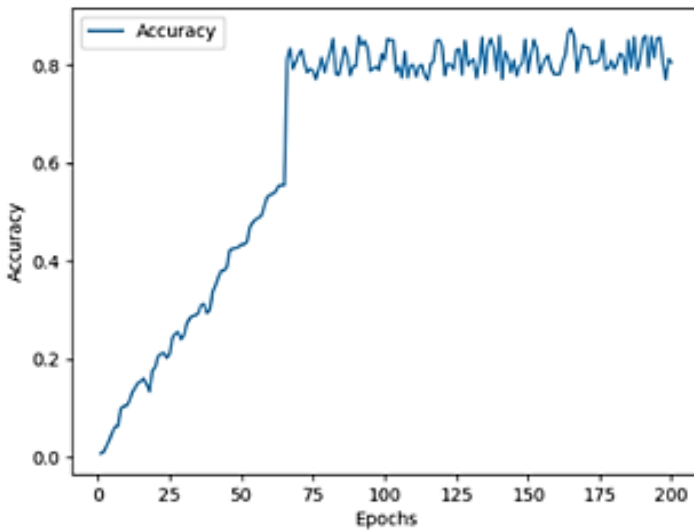


Рисунок 3 – Загальна точність набору даних «Livestream»

## Висновки

Отже, запропонований метод для виявлення ознак насильства у відеоматеріалах неймережевими засобами дозволяє визначити ступінь насильницького характеру у відсотковому відношенні, відповідно на статичних відеоматеріалах так і на динамічних (відеопотік). Подальші дослідження спрямовані на пришвидшення роботи та покращення точності запропонованого методу для набору даних, які пов'язані з насильницькими діями у реальному часі.

## Перелік посилань

1. Violence a global public health problem URL: <https://www.scielo.br/j/csc/a/3hrn64cpBqBFb9mNfP4KGXr/?lang=en>
2. Використання технологій штучного інтелекту протидії злочинності URL: [https://ivpz.kh.ua/wp-content/uploads/2020/12/Матеріали-семінару\\_Використання-техн-штучного-інтел\\_5.11.2020.pdf](https://ivpz.kh.ua/wp-content/uploads/2020/12/Матеріали-семінару_Використання-техн-штучного-інтел_5.11.2020.pdf)
3. A CNN-RNN Combined Structure for Real-World Violence Detection in Surveillance Cameras URL: <https://www.mdpi.com/2076-3417/12/3/1021>
4. NABNet: A Nested Attention-guided BiConvLSTM network URL: <https://www.sciencedirect.com/science/article/abs/pii/S1746809422007017>
5. 3D Convolutional Neural Network — A Guide for Engineers URL: <https://www.neuralconcept.com/post/3d-convolutional-neural-network-a-guide-for-engineers>
6. Convolutional LSTM for the Detection of Violence in Videos URL: [https://openaccess.thecvf.com/content\\_ECCVW\\_2018/papers/11130/Hanson\\_Bidirectional\\_Convolutional\\_LSTM\\_for\\_the\\_Detection\\_of\\_Violence\\_in\\_Videos\\_ECCVW\\_2018\\_paper.pdf](https://openaccess.thecvf.com/content_ECCVW_2018/papers/11130/Hanson_Bidirectional_Convolutional_LSTM_for_the_Detection_of_Violence_in_Videos_ECCVW_2018_paper.pdf)
7. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. In International Conference on Learning Representations (2015) URL: <http://arxiv.org/abs/1409.1556>
8. Efficient Violence Detection Using 3D Convolutional Neural Networks URL: [https://www.researchgate.net/profile/Tanfeng-Sun/publication/337537845\\_Efficient\\_Violence\\_Detection\\_Using\\_3D\\_Convolutional\\_Neural\\_Networks/links/5f75e252a6fdcc00864ccb95/Efficient-Violence-Detection-Using-3D-Convolutional-Neural-Networks.pdf](https://www.researchgate.net/profile/Tanfeng-Sun/publication/337537845_Efficient_Violence_Detection_Using_3D_Convolutional_Neural_Networks/links/5f75e252a6fdcc00864ccb95/Efficient-Violence-Detection-Using-3D-Convolutional-Neural-Networks.pdf)
9. Hockey Fight Detection Dataset URL: <https://paperswithcode.com/dataset/hockey-fight-detection-dataset>
10. Livestream URL: <https://www.twitch.tv/directory/category/just-chatting>