УДК 004.056.55:004.032.6

МЕТОД ВИЯВЛЕННЯ ЦИФРОВИХ ВОДЯНИХ ЗНАКІВ У МУЛЬТИМЕДІЙНОМУ КОНТЕНТІ

Шевченко С.Р., Хвалюк В.В., Бологова Н.М.
stanislav.shevchenko1@nure.ua, volodymyr.khvaliuk@nure.ua
Науковий керівник – доц., к.т.н., Бологова Н.М.
Харківський національний університет радіоелектроніки, каф. ЕОМ
м. Харків, Україна

This paper focuses on the development of a method for detecting digital watermarks in multimedia content using modern machine learning techniques. With the rapid growth of digital media and the increasing risk of unauthorized distribution and manipulation, reliable watermark detection has become essential for copyright protection and content authentication. The proposed approach is based on the analysis of visual and statistical features of multimedia data, enabling the identification of embedded watermarks even without access to the original content. Particular attention is given to the use of deep learning models, such as convolutional neural networks, which provide high accuracy and robustness against various types of distortions and attacks. The advantages of the method, including adaptability and effectiveness in real-world conditions, are discussed along with challenges related to computational complexity and data requirements. The results demonstrate that deep learning-based approaches significantly improve the performance of digital watermark detection systems.

Сучасний розвиток цифрових технологій супроводжується стрімким зростанням обсягів мультимедійного контенту та підвищенням ризиків його несанкціонованого використання, модифікації та розповсюдження. У таких умовах важливим інструментом захисту авторських прав і забезпечення автентичності даних є цифрові водяні знаки. Водночас зростає потреба у ефективних методах їх виявлення, особливо в умовах застосування різноманітних атак, стиснення та обробки контенту. Сучасні дослідження демонструють, що традиційні методи поступово поступаються місцем підходам на основі глибинного навчання, які забезпечують вищу точність і стійкість до спотворень [1].

Метою роботи є дослідження методу виявлення цифрових водяних знаків у мультимедійному контенті з використанням сучасних алгоритмів машинного навчання. Запропонований підхід базується на аналізі візуальних та статистичних характеристик зображень або відео, що дозволяє визначити наявність прихованої інформації навіть за відсутності знань про алгоритм її вбудовування. У сучасних роботах розглядаються так звані «сліпі» (blind) методи виявлення, які не потребують оригінального сигналу та демонструють високу ефективність у практичних застосуваннях [2].

Основу сучасних методів виявлення водяних знаків складають нейронні мережі, зокрема згорткові (CNN), автоенкодери та глибокі

генеративні моделі. Такі підходи дозволяють автоматично виділяти складні ознаки, що характеризують наявність водяного знака, та ефективно працювати з різними типами мультимедійного контенту. Дослідження показують, що використання глибинного навчання значно покращує процес як вбудовування, так і виявлення водяних знаків, забезпечуючи високу стійкість до атак і збереження якості контенту [3].

Важливим напрямом є також виявлення водяних знаків у контенті, створеному за допомогою штучного інтелекту. У сучасних умовах генеративні моделі здатні створювати реалістичні зображення, відео та аудіо, що ускладнює визначення їх походження. Для вирішення цієї проблеми активно розробляються методи цифрового маркування та подальшого виявлення таких міток з метою забезпечення достовірності контенту та протидії дезінформації [4].

Окрему увагу приділено задачі локалізації змін у мультимедійному контенті. Сучасні підходи дозволяють не лише виявити наявність водяного знака, але й визначити області його порушення або видалення. Наприклад, новітні моделі глибинного навчання здатні виконувати одночасно детекцію та локалізацію втручань у зображеннях і відео, що є важливим для задач цифрової криміналістики та захисту інформації [5].

Попри значні досягнення, існують певні обмеження таких систем. До них належать висока обчислювальна складність, потреба у великих обсягах навчальних даних, а також проблема узагальнення моделей на нові типи атак. Крім того, складність сучасних моделей ускладнює інтерпретацію отриманих результатів, що є важливим аспектом для практичного використання.

Отже, методи виявлення цифрових водяних знаків у мультимедійному контенті на основі глибинного навчання є перспективним напрямом розвитку інформаційної безпеки. Вони забезпечують високу точність, адаптивність і стійкість до різних типів атак, що робить їх ефективним інструментом захисту цифрового контенту. Подальші дослідження спрямовані на підвищення ефективності алгоритмів, зменшення їх ресурсомісткості та розширення можливостей застосування в реальних системах.

Список використаних джерел:

1. Bistroń M. et al. Deep Learning for Image Watermarking: A Comprehensive Survey // Sensors. 2026.

2. Pan M., Wang Z., Dong X. et al. A Black-Box Approach to Invisible Watermark Detection // arXiv. 2024.

3. Singh H. K., Singh A. K. Digital Image Watermarking Using Deep Learning // Multimedia Tools and Applications. 2023.

4. Watermarking for AI Content Detection: A Review // arXiv. 2025.

5. Abrar I., Sheikh J. A. DeepMark: Deep Learning-Based Watermarking for Tamper Detection // Information Processing & Management. 2026.