УДК 004.056.5:517.9
DOI https://doi.org/10.32689/maup.it.2025.3.14

Олена НЄМКОВА
доктор технічних наук, професор, професор кафедри безпеки інформаційних технологій,
Національний університет «Львівська політехніка»,
olena.a.niemkova@lpnu.ua
***ORCID:** 0000-0003-0690-2657*

Артем АХЕКЯН
кандидат фізико-математичних наук, академік МКА, заступник директора
Львівського інституту ПрАТ «ВНЗ» Міжрегіональна Академія управління персоналом»,
arachekyan@gmail.com
***ORCID:** 0000-0002-7826-8256*

Мирослава СКОЛОЗДРА
кандидат технічних наук, доцент, доцент кафедри безпеки інформаційних технологій,
Національний університет «Львівська політехніка»,
myroslava.m.skolozdra@lpnu.ua
***ORCID:** 0009-0004-4559-0101*

МАТЕМАТИЧНИЙ МЕТОД ІДЕНТИФІКАЦІЇ ШІ-ГЕНЕРОВАНИХ ЗОБРАЖЕНЬ НА ОСНОВІ SVD ТА ЛІНІЙНОЇ РЕГРЕСІЇ

Анотація. *Стрімкий розвиток технологій штучного інтелекту, зокрема генеративних моделей, таких як Stable Diffusion, спричинив зростання кількості ШІ-генерованих зображень, що створює значні виклики для протидії дезінформації та забезпечення цілісності цифрового контенту в соціальних мережах, журналістиці та юридичних контекстах. Запропонований математичний метод вирішує цю проблему, забезпечуючи автоматизований і ефективний підхід до ідентифікації синтетичних патернів у зображеннях, що має практичну цінність для етичного нагляду за ШІ та судово-медичних застосувань. Дослідження є особливо актуальним з огляду на зростаючу потребу в надійних інструментах для виявлення маніпуляцій із зображеннями, таких як deepfakes та копіювання-переміщення, в епоху швидкого розвитку ШІ-технологій.*

Мета роботи *полягає у розробці та апробації математичного методу виявлення фальсифікації цифрових зображень, який базується на аналізі сингулярного розкладання (SVD) та лінійної регресії з використанням тангенса кута нахилу (slope) як ключового критерію для розрізнення реальних зображень і тих, що створені штучним інтелектом (ШІ). Запропонований підхід спрямований на визначення відмінностей у розподілі енергії зображень, що дозволяє ідентифікувати синтетичні патерни, характерні для AI-генерації, та оцінити ефективність методу на практичних прикладах.*

Методологія дослідження *включає перетворення цифрового зображення в матрицю пікселів, застосування сингулярного розкладання для отримання сингулярних значень, їх логарифмічної апроксимації та побудови лінійної регресії. Тангенс кута нахилу обчислюється як коефіцієнт регресії, що відображає швидкість розпаду енергії. Для підвищення точності аналізу використовуються блочні методи, де зображення розбивається на підматриці розміром 16x16 пікселів, а отримані значення slope порівнюються з емпіричним порогом, наприклад, <-0,8 для автентичних зображень. Експерименти проводилися на наборі даних, що включає реальні фотографії та зображення, створені моделями типу Stable Diffusion, з подальшою статистичною оцінкою результатів.*

Наукова новизна *полягає в інтеграції SVD із лінійною регресією для моделювання розпаду логарифмів сингулярних значень із акцентом на тангенс нахилу як диференціальну ознаку. На відміну від традиційних методів, що спираються на частотний аналіз або ключові точки, запропонований підхід забезпечує автоматизовану класифікацію без потреби в ручному налаштуванні параметрів. Це дозволяє ефективно розпізнавати маніпуляції, включаючи copy-move forgery та deepfakes, що є актуальним у контексті стрімкого розвитку ШІ-технологій.*

Висновки *роботи підтверджують високу ефективність методу для розрізнення реальних і ШІ-згенерованих зображень, де середнє значення slope для автентичних зображень становить -1,4026, а для ШІ-зображень відповідно -0,5829. Метод демонструє точність 87,76% на тестовому наборі з 98 зображень, а також Recall 93,55% і Specificity 85,07%, хоча виявлено обмеження при аналізі зображень із однорідною текстурою та наявність 7 хибнопозитивів. Результати підкреслюють практичне значення підходу для захисту від дезінформації, підтримки юриспруденції та етичного контролю ШІ, з перспективою подальшого вдосконалення через комбінацію з такими техніками, як SIFT (Scale-Invariant Feature Transform, Трансформація ознак, інваріантна до масштабу) чи CNN (Convolutional Neural Network, Згортальна нейронна мережа).*

Ключові слова: *сингулярне розкладання (SVD), ідентифікація зображень, лінійна регресія, ШІ-зображення, аналіз нахилу, комп'ютерний зір, виявлення фальсифікації зображень.*

Olena NYEMKOVA, Artem AKHEKYAN, Myroslava SKOLOZDRA. MATHEMATICAL METHOD FOR IDENTIFYING AI-GENERATED IMAGES BASED ON SVD AND LINEAR REGRESSION

Abstract. *The rapid advancement of artificial intelligence technologies, particularly generative models such as Stable Diffusion, has led to an increase in AI-generated images, creating significant challenges for countering disinformation and ensuring the integrity of digital content in social media, journalism, and legal contexts. The proposed mathematical method addresses this issue by providing an automated and effective approach to identifying synthetic patterns in images, offering practical value for ethical AI oversight and forensic applications. This research is particularly timely given the growing need for robust tools to detect image manipulations, such as deepfakes and copy-move forgeries, in an era of rapidly evolving AI capabilities.*

The purpose *of this study is to develop and test a mathematical method for detecting the authenticity of digital images, utilizing singular value decomposition (SVD) and linear regression, with the tangent of the slope (slope) as the key criterion for distinguishing real images from those generated by artificial intelligence (AI). The proposed approach aims to identify differences in the energy distribution of images, enabling the detection of synthetic patterns characteristic of AI-generated content, and to evaluate the method's effectiveness through practical examples.*

The methodology *involves transforming a digital image into a pixel matrix, applying singular value decomposition to obtain singular values, performing their logarithmic approximation, and constructing linear regression. The tangent of the slope is calculated as the regression coefficient, reflecting the rate of energy decay. To enhance accuracy, a block-based method is employed, dividing the image into 16x16 pixel submatrices, with the resulting slope values compared against an empirical threshold (e.g., <-0,8 for authentic images). Experiments were conducted on a dataset comprising real photographs and images generated by models such as Stable Diffusion, followed by statistical evaluation of the results.*

The scientific novelty *lies in the integration of SVD with linear regression to model the decay of logarithms of singular values, emphasizing the slope as a differential feature. Unlike traditional methods relying on frequency analysis or keypoints, this approach enables automated classification without the need for manual parameter tuning. It effectively detects manipulations, including copy-move forgery and deepfakes, addressing the rapid advancement of AI technologies.*

Conclusions. *The findings of the study confirm the high effectiveness of the method for distinguishing between real and AI-generated images, where the average slope value for authentic images is -1,4026, and for synthetic images, it is -0.5829. The method demonstrates an accuracy of 87,76% on a test set of 98 images, along with a Recall of 93,55% and Specificity of 85,07%, though limitations were identified in analyzing images with uniform textures and the presence of 7 false positives. The results underscore the practical significance of the approach for protecting against disinformation, supporting jurisprudence, and ethical AI control, with prospects for further improvement through integration with techniques such as SIFT (Scale-Invariant Feature Transform) or CNN (Convolutional Neural Network).*

Key words: *Singular Value Decomposition (SVD), Identification of Images, Linear Regression, AI-Generated Images, Slope Analysis, Computer Vision, Deepfake Detection.*

Постановка проблеми. Маніпуляція цифровими зображеннями, що впливає на їхню автентичність, є об'єктом аналізу, який може бути виявлений за допомогою математичних методів, зокрема сингулярного розкладання та лінійної регресії з аналізом тангенса нахилу. Такі маніпуляції, зокрема фальсифікація зображень, що передбачає процес зміни їхнього змісту з метою введення в оману, мають різне сприйняття залежно від контексту застосування. Термін «фальсифікація» несе дещо негативний відтінок, хоча його оцінка значною мірою залежить від сфери використання. У галузі розваг ця техніка застосовується позитивно, сприяючи створенню креативного та захопливого контенту. Натомість у контексті юриспруденції чи журналістики вона набуває негативного значення, часто асоціюючись із обманом чи поширенням дезінформації. У зв'язку з цим розробка та застосування методів виявлення таких маніпуляцій, включаючи фальсифікацію, є надзвичайно важливими в цих областях.

Існують різні методи маніпуляції зображень, які класифікуються залежно від технік і цілей. Відомі методи та їхнє застосування представлено у (табл. 1).

Методи маніпуляції варіюються від простого редагування (ретушування) до складних ШІ-генерацій. Їхнє застосування залежить від цілей – від нешкідливої творчості до серйозних злочинів. Оскільки різні методи маніпуляції використовують різні техніки, то їх виявлення вимагає специфічних підходів. Розрізняють таки основні підходи: блочні методи (розділення зображення на блоки, наприклад, 16x16 і порівняння їхніх ознак), ключово-точкові методи (використання SIFT, SURF для виявлення дублікатів чи аномалій), та SVD-аналіз (оцінка розпаду сингулярних значень для виявлення CMF чи ШІ-генерації).

Мета даної роботи полягає у розробці та апробації методу виявлення автентичності цифрових зображень на основі аналізу розпаду сингулярних значень із застосуванням сингулярного розкладання (SVD) та лінійної регресії. Основною ознакою для класифікації зображень як реальних чи створених штучним інтелектом є тангенс кута нахилу, отриманий шляхом апроксимації логарифмічної залежності сингулярних значень $\log(\sigma_i)$. Запропонований підхід спрямований на визначення характерних відмінностей у розподілі енергії зображень, що дозволяє диференціювати природні візуальні структури від синтетичних патернів, характерних для зображень, сформованих ШІ, та оцінити ефективність методу на практичних прикладах.

ISSN 2786-5460 (Print), ISSN 2786-5479 (Online)

Таблиця 1

Класифікація методів маніпуляції зображень

№ з/п	Назва	Опис методу	Застосування	Небезпека
1	Ретушування (Image Retouching)	Метод передбачає коригування або покращення певних частин зображення, таких як видалення дефектів, зміни кольору, корекція освітлення чи текстури. Використовуються інструменти типу Photoshop для точкового редагування.	У рекламі чи соціальних мережах для «покращення» зовнішності моделей (згладжування шкіри, зміна форм). У реставрації старих фото для видалення подряпин чи плям.	Оскільки зміни не завжди приховують критичну інформацію, цей метод вважається відносно нешкідливим.
2	Сплайсинг (Image Splicing)	Об'єднання двох або більше зображень у одне, щоб приховати чи додати елементи. Зазвичай використовуються шари та маски для безшовного злиття.	У маніпуляції новинами шляхом створення фальшивих сцен, наприклад, додавання людини до події, де її не було. У художніх проєктах для створення сюрреалістичних зображень.	Приховування інформації у судових або кримінальних справах для фальсифікації доказів.
3	Копіювання та переміщення (Copy-Move Forgery, CMF)	Найпоширеніший метод, який включає копіювання частини зображення та вставлення її в іншу область того ж зображення. Часто супроводжується редагуванням (масштабування, обертання, додавання шуму) для маскування.	У фальсифікації доказів шляхом Додавання об'єктів (наприклад, зброї) до фото злочинних сцен. У соціальних мережах для створення ілюзій багатства чи присутності, наприклад, дублювання предметів.	Складність виявлення: завдяки редагуванню та шумам цей метод важко розпізнати без спеціальних алгоритмів, таких як SVD або SIFT.
4	Генерація зображень за допомогою ШІ (AI-Generated Forgery)	Використання генеративних моделей, наприклад, GANs - Generative Adversarial Networks, для створення реалістичних зображень із нуля або модифікації існуючих. Приклади: DALL·E, Stable Diffusion.	З метою дезінформації створюються фальшиві фото осіб чи подій (deepfakes). Для розваг та у мистецтві застосовується генерація унікальних зображень для творчих проєктів.	Складність виявлення: ШІ-зображення мають синтетичні патерни, які важко відрізнити від реальних без аналізу розпаду сингулярних значень чи інших методів.
5	Підміна облич (Face Swapping)	Заміна обличчя однієї людини на обличчя іншої з використанням технологій розпізнавання та синтезу, часто на основі deep learning.	Для розваг: популярно в додатках типу FaceApp або Zao. Зі злочинними намірами створюються фальшиві відео чи фото для шахрайства чи шантажу.	Для виявлення вимагає аналізу мікровиразів або аномалій у текстурах.
6	Додавання шуму чи артефактів (Noise Addition/ Artifact Insertion)	Додавання штучного шуму (наприклад, гаусового) або компресійних артефактів для маскування маніпуляцій.	У маскуванні редагувань, наприклад, у CMF для приховання швів. Для створення імітації старіння, наприклад, створення ефекту старого фото.	Для виявлення вимагає аналізу спектрального розподілу або SVD (виявлення невідповідностей).

Практичне значення даної роботи полягає в розробці та впровадженні методу виявлення автентичності цифрових зображень, який базується на аналізі сингулярного розкладання (SVD) та лінійної регресії з використанням тангенса кута нахилу (slope) як ключового критерію. Це має низку важливих прикладних аспектів, особливо в сучасному цифровому середовищі, де фальсифікація зображень стала поширеним явищем.

Аналіз останніх досліджень і публікацій. Метод сингулярного розкладання знайшов застосування серед кількох українських дослідників, особливо в галузях обробки зображень, аналізу даних і машинного навчання. Серед відомих українських дослідників SVD застосовували Алла Кобозєва, Олександр Потьомкін, Ігор Сердюк, Наталія Бондаренко та Володимир Лук'янчук, переважно в обробці

зображень, стеганографії та аналізі даних. У дослідженнях Кобозєвої SVD застосовувався для розкладання матриць зображень на компоненти, що дозволяло аналізувати локальні особливості, наприклад, максимальні сингулярні значення, і виявляти аномалії, пов'язані з маніпуляціями чи стеганографією [7].

Методи виявлення маніпуляції зображень за допомогою сингулярного розкладання (SVD), або близькі до нього підходи, активно досліджуються науковцями, інтереси яких лежать у галузі обробки зображень, комп'ютерного зору та цифрової криміналістики. SVD є узагальненням спектрального розкладання (eigenvalue decomposition) для несиметричних матриць і використовується у задачах зменшення розмірності, стиснення даних, видалення шуму та аналізу зображень. У контексті виявлення фальсифікації зображень SVD застосовується для аналізу розпаду сингулярних значень: натуральні зображення мають швидкий експоненціальний розпад log(σi) приблизно лінійний з від'ємним нахилом, тоді як фальсифіковані – повільніший або нелінійний через порушення природної структури пікселів [5]. Це робить SVD ефективним для виявлення copy-move forgery або ШІ-генерації, оскільки маніпуляції змінюють статистичні властивості матриці зображення [12]. Перевагою SVD над спектральним аналізом є можливість аналізувати несиметричні структури, як зображення, без перетворення у симетричну форму.

ШІ-генеровані зображення, наприклад, за допомогою GANs або diffusion models, можуть частково протидіяти SVD-детекції, але не повністю, через фундаментальні статистичні відмінності. Моделі ШІ можуть бути навчені імітувати статистичні властивості реальних зображень, включаючи розпад сингулярних значень. Наприклад, є відомості, що генератори можуть оптимізуватися для створення зображень із подібним SVD-розпадом, щоб обійти детектори [1]. Дослідження показують, що постобробка, наприклад, додавання шуму, дозволяє уникнути виявлення, оскільки SVD чутливий до таких маніпуляцій [15]. У 2025 році гібридні моделі, таки як Stable Diffusion, можуть генерувати зображення з реалістичним розпадом енергії, що робить SVD менш ефективним для нових генераторів [14]. Тим не менш, ШІ не може повністю протидіяти статистичним відмінностям; ШІ-зображення часто мають неприродний розподіл шуму або текстур, що порушує експоненціальний розпад сингулярних значень. Детектори на основі SVD, наприклад, блочні методи, виявляють локальні аномалії, які важко ідеально імітувати [8]. Дослідження показують, що навіть удосконалені GANs не можуть повністю відтворити SVD-розпад реальних зображень через обмеження тренувальних даних [3]. SVD комбінується з ML-моделями (наприклад, CNN), що робить імітацію складнішою. ШІ, щоб протидіяти, повинен тренуватися проти конкретних детекторів, але загальні SVD-методи стійкі [11]. Отже, повна імітація вимагає обчислювальних ресурсів і може призвести до артефактів в інших доменах, наприклад, частотному, що виявляються іншими методами. У змаганнях з SVD-детекцією, ШІ може перемагати в обмежених випадках, але не завжди, оскільки базові математичні властивості важко ідеально відтворити. Для повного виявлення імітацій потрібно комбінувати SVD з іншими техніками [2].

Отже, метод SVD є ефективним у виявленні діпфейків. Наведемо декілька прикладів досліджень на цю тему за останні роки. У статті [6] описано метод виявлення копіювання-вставки (copy-move forgery) на основі SVD. Зображення розбивається на блоки, для кожного з яких обчислюється SVD, а максимальне значення діагональної матриці (норма) використовується для групування схожих блоків. Ключовим для виявлення аномалій є аналіз сингулярних значень та їх розподілу, але лінійна регресія прямо не згадується. Стаття [10] пропонує метод виявлення фальсифікації типу «зшивання» (image splicing) з використанням SVD у комбінації з дискретним косинусним перетворенням (DCT). Зображення розбивається на блоки, для кожного обчислюються DCT-коефіцієнти, а потім застосовується SVD для вилучення особливостей. Автори статті [9] використовують метод SVD для вилучення особливостей з RGB-зображень (зокрема, з червоної матриці), після чого сингулярні значення та вектори передаються в одновимірний клітинний автомат для створення ключа автентифікації. У статті [4] описується метод виявлення маніпуляцій із зображеннями на основі SVD, де порушення лінійних залежностей у рядках або стовпцях зображення використовується для ідентифікації фальсифікацій. Метод фокусується на аналізі сингулярних значень що може бути сумісним з регресійним підходом для оцінки розпаду. Зауважимо, що у згаданих статтях, в яких статистичні методи застосовуються для оцінки аномалій, не описано чіткої комбінації SVD і лінійної регресії для моделювання розпаду логарифмів сингулярних значень $\log(\sigma_i)$.

Постановка завдання. Оскільки метою даної роботи є розробка математичного методу для визначення автентичності цифрових зображень шляхом розрізнення реальних зображень та тих, що створені штучним інтелектом (ШІ), з використанням сингулярного розкладання (SVD) та лінійної регресії, було сформульовано наступне **завдання**, яке передбачає наступне:

1. Аналіз структури цифрових зображень шляхом перетворення їх у матриці пікселів і обчислення сингулярних значень за допомогою SVD.

ISSN 2786-5460 (Print), ISSN 2786-5479 (Online)

2. Побудова логарифмічної апроксимації сингулярних значень та застосування лінійної регресії для визначення тангенса кута нахилу (slope) як кількісного показника розпаду енергії зображення.

3. Встановлення емпіричних порогів для класифікації зображень на основі значення slope, де автентичні зображення характеризуються різким експоненціальним розпадом (slope < -0,8), а ШІ-згенеровані – повільнішим (slope > -0,8).

4. Проведення експериментальної перевірки методу на тестовому наборі даних, що включає реальні фотографії та зображення, сформовані ШІ-моделями (наприклад, Stable Diffusion), з оцінкою точності та ідентифікації обмежень.

5. Оцінка практичної придатності розробленого підходу для захисту від дезінформації, підтримки юриспруденції та етичного контролю ШІ-технологій.

Постановка завдання зумовлена необхідністю протидії стрімкому поширенню фальсифікованих зображень у цифровому середовищі, що вимагає ефективних і автоматизованих методів аналізу, а також актуальністю інтеграції математичних інструментів у задачі комп'ютерного зору.

Виклад основного матеріалу дослідження. Метод сингулярного розкладання (Singular Value Decomposition, SVD) – це фундаментальна техніка лінійної алгебри, яка застосовується для розкладання матриці на три компоненти. Формально, будь-яку матрицю A розміром $m \times n$ можна представити у вигляді SVD розкладання:

$$A = U\Sigma V^T$$

де U – ортогональна матриця розміром $m \times m$, стовпці якої є лівими сингулярними векторами,

Σ – діагональна матриця розміром $m \times n$ з невід'ємними сингулярними значеннями $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_k \geq 0$ (де $k = \min(m,n)$), які відображають «енергію» або важливість компонент матриці,

V^T – транспонована ортогональна матриця розміром $n \times n$, стовпці якої є правими сингулярними векторами.

Після обчислення сингулярних значень матриці зображення необхідно виконати розрахунок їхніх десяткових логарифмів і побудувати лінійну регресію залежно від порядкового номера сингулярного значення. На наступному етапі визначається тангенс кута нахилу отриманої прямої, який слугує ключовим параметром для класифікації зображень на реальні та фальсифіковані.

Був розроблений наступний алгоритм (який було реалізовано мовою програмування Python), основні його кроки наведено нижче:

1. Підготовка вхідних даних (завантажити цифрове зображення у форматі JPG і перетворити його на двовимірну матрицю пікселів A розміром $m \times n$, m – висота, n – ширина зображення; розбити матрицю A на блоки 16×16 для локального аналізу, отримавши множину підматриць A_i, i = 1, 2, … k, де k – кількість блоків).

2. Обчислення сингулярного розкладання (SVD) (для кожної підматриці A_i виконати сингулярне розкладання; витягти вектори сингулярних значень σ_i для подальшого аналізу).

3. Обчислення десяткових логарифмів (для кожного вектора σ_i обчислити десяткові логарифми сингулярних значень; сформувати залежність $y_j = \log_{10}(\sigma_{ij})$ від індексу j).

4. Побудова лінійної регресії (використовуючи метод найменших квадратів, апроксимувати отримані дані (j, y_j) лінійною функцією $y_j = aj + b$, де a – тангенс кута нахилу (slope); обчислити параметр a).

5. Визначення тангенсів нахилу блоків (взяти отримане значення a як тангенс кута нахилу для блоку A_i; повторити кроки 2 – 4 для всіх блоків з утворенням множини значень a_i).

6. Класифікація зображень (обчислити середнє значення тангенсу нахилу для всього зображення slope як середнє аріфметичне; порівняти його з емпіричним порогом, наприклад, якщо slope менше за -0,8, зображення класифікується як реальне, якщо slope не менше за -0,8, то фальсифіковане (ШІ-генероване)).

Для тестування алгоритму було використано 97 зображень, з яких 31 зображення (клас 1) отримане за допомогою ШІ-генератора [13] і 67 зображень (клас 0) взято з колекції реальних фотографій, отриманих з мобільного телефону. Всі зображення мали обсяг 100÷200 kB і розширення JPG. Роздільна здатність камери мобільного телефону 45 Мрх, час генерування одного ШІ-зображення 10,8 сек. Тематично реальні та генеровані зображення були різноманітними: одна людина, група людей, архітектура, природа, тварини.

Було отримано наступні результати. Для ШІ-генерованих зображень отримана наступна множина тангенсів кутів нахилу (slope), відсортована від мінімального до максимального: {-1,2855; -1,1473; -0,7889; -0,5886; -0,5701; -0,5679; -0,5674; -0,5647; -0,5641; -0,5519; -0,5492; -0,5455; -0,5402; -0,5343; -0,5278; -0,5273; -0,5267; -0,5256; -0,5247; -0,5175; -0,5144; -0,5134; -0,5102; -0,5081; -0,5041; -0,5000; -0,4934; -0,4933}; множина slope для реальних зображень відповідно така: {-7,1863; -5,5992; -4,4227;

-2,2959; -2,2162;, -1,7976; -1,7859;, -1,7567; -1,7381; -1,6716; -1,6427; -1,6412; -1,5873; -1,5612; -1,5589; -1,5526; -1,5303; -1,5043; -1,4736; -1,4441; -1,4348; -1,4233; -1,4229; -1,3669; -1,3588; -1,3501; -1,3109; -1,2676; -1,2661; -1,2149; -1,2094; -1,1981; -1,1911; -1,1552; -1,1455; -1,1422; -1,1389; -1,1346; -1,1266; -1,1223; -1,0716; -1,0706; -1,0637; -1,0594; -1,0361; -1,0119; -1,0051; -0,9976; -0,9797; -0,9629; -0,9588; -0,9233; -0,9229; -0,8921; -0,8869; -0,8359; -0,8147; -0,7753; -0,7384; -0,6939; -0,6525; -0,6518; -0,632; -0,6305; -0,6147; -0,6075; -0,5357}.

Статистичні показники множин для ШІ-генерованих та реальних зображень наведено у (табл. 2).

Таблиця 2

Описова статистика для набору даних slope

Показник	Реальні зображення	ШІ-зображення
Середнє значення	-1,4026	-0,5829
Медіана	-1,1552	-0,5278
Дисперсія вибірки	1,0933*	0,0318
Стандартне відхилення	1,0456**	0,1782
Інтервал	6,6506	0,7962
Мінімум	-7,1863	-1,2855
Максимум	-0,5357	-0,4893

*Примітка: *,** – такі великі значення дисперсії та стандартного відхилення пояснюються наявністю викидів {-7,1863; -5,5992; -4,4227}, що не є характерними для набору даних slope реальних зображень*

Виконано бінарну класифікацію отриманих результатів. Для порогу slope = -0,8 з 31 ШІ-зображень 2 помилково класифікуються як реальні (False Negatives, FN = 2), решта 29 класифікуються як фейкові (True Positives, TP = 29). З 67 справжніх зображень 10 класифікуються як фейкові (False Positives, FP = 10) і 57 як справжні (True Negatives, TN = 57).

Для оцінки якості класифікації застосуємо наступні параметри: точність (*Accuracy*), точність прогнозу позитивного класу (*Precision*), чутливість (*Recall*), специфічність (*Specificity*), площу під ROC-кривою.

Точність (*Accuracy*) – це метрика, яка вимірює частку правильних прогнозів моделі відносно загальної кількості прогнозів. Точність обчислюють, щоб оцінити, наскільки добре запропонований метод розрізняє реальні зображення та зображення, створені штучним інтелектом:

$$Accuracy = \frac{TN + TP}{TN + TP + FN + FP} * 100\% = 87{,}6\%$$

Точність 87.76% висока, але менша 95%, що може вказувати на жорсткість порогу -0,8. Оптимізація порогу може підвищити точність.

Точність прогнозу позитивного класу (*Precision*) вимірює частку правильних позитивних прогнозів серед усіх позитивних прогнозів, що означатиме частку правильно класифікованих зображень, створених штучним інтелектом, серед усіх зображень, позначених як ШІ:

$$Precision = \frac{TP}{TP + FP} * 100\% = 74{,}36\%$$

Precision показує, що серед усіх зображень, класифікованих як ШІ, 74,36% дійсно є ШІ. Решта 25,64% (FP = 10) – це реальні зображення, помилково віднесені до ШІ. Порівняно з *Accuracy* (87,76%) *Precision* нижча, оскільки враховує лише позитивний клас (ШІ) і чутлива до FP. Це вказує на те, що поріг -0,8 допускає певну кількість помилок, особливо серед реальних зображень із slope близьким до -0,8.

Чутливість (*Recall*, або *True Positive Rate*) вимірює частку правильно виявлених позитивних випадків серед усіх реальних позитивних випадків. *Recall* допомагає оцінити, наскільки добре метод виявляє ШІ-зображення, використовуючи поріг -0,8:

$$Recall = \frac{TP}{TP + FN} * 100\% = 93{,}55\%$$

Ця метрика показала, що серед усіх ШІ-зображень метод правильно виявляє 93,55% як фальсифіковані. Решта 6.45% (FN = 2) – це ШІ-зображення, помилково класифіковані як реальні, що може бути пов'язано з їхньою високою реалістичністю. Загалом, *Recall* вищий за *Precision* (74,36%), що вказує на те, що запропонований метод для обраного порогу краще виявляє ШІ-зображення, але з ризиком

ISSN 2786-5460 (Print), ISSN 2786-5479 (Online)

помилково класифікувати реальні як ШІ. Високий *Recall* робить метод корисним для сфер, де пропуск ШІ-зображення критичний (наприклад, дезінформація чи юриспруденція), але низький *Precision* вимагає балансу, наприклад, оптимізації порогу до -0,9 для зменшення FP. Загалом, метод краще «ловить» ШІ (93,55%), ніж виключає помилки в реальних (TN/(TN+FP)*100% = 85,07%).

Специфічність (*Specificity*, або *True Negative Rate*) – це метрика, яка вимірює частку правильно виявлених негативних випадків (реальних зображень) серед усіх реальних негативних випадків.

$$Specificity = \frac{TN}{TN + FP} * 100\% = 85{,}07\%$$

Метрика показала, що серед усіх реальних зображень метод правильно класифікує 85,07% як реальні. Решта 14,93% – це реальні зображення, помилково класифіковані як ШІ, що може бути пов'язано з шумом або текстурами, де slope ≥ -0,8. Отже, є ризик помилково «позначати» реальні зображення як ШІ. *Specificity* = 85,07% свідчить про добру здатність методу виключати фальсифікації для реальних зображень, що є сильною стороною для захисту від дезінформації. Різниця з ШІ (де *Recall* = 93,55%) вказує, що оптимізація порогу (наприклад, до -0,9) може підвищити *Specificity*.

Для обчислення ROC-AUC (Receiver Operating Characteristic – Area Under Curve) на основі наданих даних нахилу slope для реальних зображень і ШІ-зображень було використано Python із бібліотеками numpy, sklearn та matplotlib для візуалізації. ROC-AUC вимірює якість бінарної класифікації, порівнюючи справжні позитивні та хибні позитивні ставки при різних порогах (рис.1).

Значення AUC (Area Under Curve) = 0,95 свідчить про високу ефективність запропонованого методу, заснованого на аналізі тангенса кута нахилу за допомогою сингулярного розкладання та лінійної регресії, для розрізнення реальних зображень і тих, що створені штучним інтелектом. AUC = 0,95 наближається до ідеального значення 1,0, що вказує на хорошу здатність методу класифікувати зображення з мінімальною кількістю помилок.

Висновки та перспективи подальших розвідок. Проведене дослідження підтвердило ефективність розробленого математичного методу для виявлення автентичності цифрових зображень на основі сингулярного розкладання (SVD) та лінійної регресії з аналізом тангенса кута нахилу (slope). Аналіз розподілу енергії зображень через логарифмічну апроксимацію сингулярних значень із порогом -0,8 дозволив чітко розмежувати реальні зображення (середнє slope = -1,4026; 85,1% значень < -0,8) та зображення, створені штучним інтелектом (середнє slope = -0,5829; 93,5% значень > -0,8). Експериментальна оцінка на тестовому наборі з 67 реальних фото з мобільного телефону та 31 зображення, згенерованого моделями типу Stable Diffusion, показала високу точність класифікації (Accuracy = 87,76%) та AUC = 0,95, що свідчить про добру здатність методу розпізнавати синтетичні патерни, включаючи deepfakes та copy-move forgery. Експериментальні результати: 57 справжніх

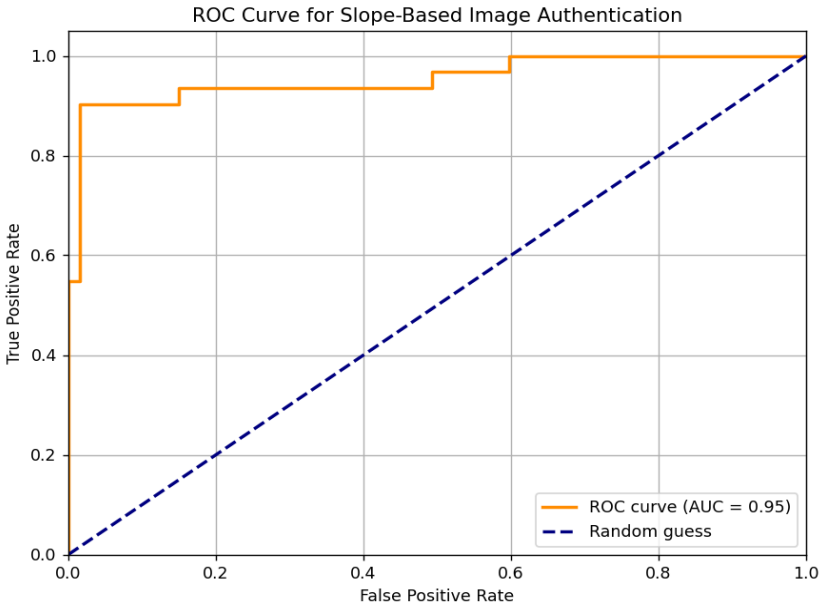

Рис. 1. Графік ROC-кривої (суцільна лінія) показує ефективність класифікації. Додано лінію випадкового вибору (пунктир, FPR = TPR) для порівняння

негативних, 29 справжніх позитивних, 10 хибнопозитивних і 2 хибнонегативних підкреслюють збалансованість класифікації з Precision = 74,36%, Recall = 93,55% та Specificity = 85,07%.

Практична придатність методу підтверджена його потенціалом для захисту від дезінформації в соціальних мережах, підтримки юриспруденції через верифікацію фото як доказів та етичного контролю ШІ-технологій. Високе значення Recall (93,55%) забезпечує ефективне виявлення ШІ-зображень, що критично для боротьби з фальсифікаціями, тоді як Specificity (85,07%) гарантує надійність виключення реальних зображень. Отримані результати підкреслюють наукову новизну підходу, що полягає в застосуванні тангенса нахилу як диференціальної ознаки.

Обмеження методу пов'язані з обробкою зображень із однорідною текстурою, де slope наближається до порогу -0,8, а також із викидами, наприклад, slope < -4,0 у реальних зображеннях, що потребує фільтрації. Перспективи подальших досліджень включають оптимізацію порогу через ROC-аналіз для підвищення Precision до 85% і вище, розширення тестового набору даних із різноманітними ШІ-моделями (DALL·E, Midjourney) та інтеграцію з іншими техніками для підвищення стійкості до аномалій. Для покращення результатів планується впровадження адаптивної фільтрації малих сингулярних значень, комбінацію з глибоким навчанням для автоматичного визначення порогу та застосування методу до відеоаналізу для виявлення deepfakes у динамічних сценах. Отримані результати підкреслюють практичне значення роботи для сучасних викликів у комп'ютерному зорі й безпеці даних, з потенціалом для автоматизованого впровадження в реальному часі.

Список використаних джерел:

1. Ba Z., Zhang Y., Cheng P., Gong B., Zhang X., Wang Q., Ren K. Robust Watermarks Leak: Channel-Aware Feature Extraction Enables Adversarial Watermark Manipulation. arXiv:2502.06418v1 [cs.CV], 10 Feb 2025. URL: https://arxiv.org/html/2502.06418v1

2. Capasso P., Cattaneo G., de Marsico M. A Comprehensive Survey on Methods for Image Integrity. ACM Transactions on Multimedia Computing, Communications and Applications, Vol. 20, No. 11, Article No. 347, 2024, pp. 1–34. URL: https://doi.org/10.1145/3633203

3. Deb P., Deb S., Das A., Kar N. Image Forgery Detection Techniques: Latest Trends and Key Challenges. IEEE Access, Vol. PP, No. 99, January 2024, pp. 1–1. DOI: 10.1109/ACCESS.2024.3498340

4. Gul G., Avcibas I., Kurugollu F. SVD Based Image Manipulation Detection. In: 2010 IEEE International Conference on Image Processing, Hong Kong, China, September 2010. DOI: 10.1109/ICIP.2010.5652854. URL: https://ieeexplore.ieee.org/document/5652854

5. Kashyap A., Agarwal M., Gupta H. Detection of Copy-Move Image Forgery Using SVD and Cuckoo Search Algorithm. arXiv:1704.00631v1 [cs.MM], 3 Apr 2017. URL: https://arxiv.org/pdf/1704.00631. DOI: 10.14419/ijet.v7i2.13.11604

6. Khudhair Z. N., Mohamed F., Rehman A., Saba T., Bahaj S. A. Detection of Copy-Move Forgery in Digital Images Using Singular Value Decomposition. Computers, Materials & Continua, Vol. 74, No. 2, 2023, pp. 4135–4147. URL: https://doi.org/10.32604/cmc.2023.032315

7. Kobozieva A., Bobok I., Kushnirenko N. Steganalysis Method for Detecting LSB Embedding in Digital Video, Digital Image Sequence. In: 11th International Conference «Information Control Systems and Technologies» (ICST 2023), Odesa, 21–23 September 2023, pp. 78–90. [CEUR Workshop Proceedings, Vol. 3513]. URL: https://ceur-ws.org/Vol-3513/paper07.pdf

8. Lađević A. L., Kramberger T., Kramberger R., Vlahek D. Detection of AI-Generated Synthetic Images with a Lightweight CNN. Artificial Intelligence, Vol. 5, No. 3, 2024, pp. 1575–1593. URL: https://doi.org/10.3390/ai5030076

9. Malakooti M. V., Tafti A. P., Rohani F., Moghaddasifar M. A. RGB Digital Image Forgery Detection Using Singular Value Decomposition and One Dimensional Cellular Automata. In: 2012 8th International Conference on Computing Technology and Information Management (NCM and ICNIT), 2012. URL: https://ieeexplore.ieee.org/document/6268546

10. Moghaddasi Z., Jalab H. A., Noor R. M. Image Splicing Forgery Detection Based on Low-Dimensional Singular Value Decomposition of Discrete Cosine Transform Coefficients. Neural Computing and Applications, 2018. URL: https://doi.org/10.1007/s00521-018-3648-3

11. Saberi M., Sadasivan V. S., Rezaei K., Kumar A., Chegini A., Wang W., Feizi S. Robustness of AI-Image Detectors: Fundamental Limits and Practical Attacks. arXiv:2310.00076, Feb 2024. URL: https://doi.org/10.48550/arXiv.2310.00076

12. Sengupta S., Shinde P., Shah H. Image Forgery Detection Techniques for Forensic Sciences. International Journal of Software & Hardware Research in Engineering, Vol. 2, No. 8, August 2014. URL: https://ijournals.in/wp-content/uploads/2017/07/9.2814-Prajakta.pdf

13. Stable Diffusion 2.1 Demo. URL: https://huggingface.co/spaces/stabilityai/stable-diffusion

14. Vahdati D. S., Nguyen T. D., Azizpour A., Stamm M. C. Beyond Deepfake Images: Detecting AI-Generated Videos. arXiv:2404.15955v1 [cs.CV], 24 Apr 2024. URL: https://arxiv.org/html/2404.15955v1

15. Xie H., Ni J., Zhang J., Zhang W., Huang J. Evading Generated-Image Detectors: A Deep Dithering Approach. Signal Processing, Vol. 197, August 2022, 108558. URL: https://doi.org/10.1016/j.sigpro.2022.108558

Дата надходження статті: 19.09.2025
Дата прийняття статті: 20.10.2025
Опубліковано: 04.12.2025