https://doi.org/10.31891/2307-5732-2025-359-54
УДК 004.056.52

ПОДДУБНИЙ ВАДИМ
Харківський національний університет радіоелектроніки
https://orcid.org/0000-0002-4380-491X
e-mail: vadym.poddubnyi@nure.ua
СЄВЄРІНОВ ОЛЕКСАНДР
Харківський національний університет радіоелектроніки
https://orcid.org/0000-0002-6327-6405
e-mail: oleksandr.sievierinov@nure.ua

СТІЙКІСТЬ СХЕМИ АВТЕНТИФІКАЦІЇ ЗАСНОВАНОЇ НА НУЛЬОВОМУ ВОДЯНОМУ ЗНАКУ

У статті представлений аналіз стійкості схеми автентифікації заснованої на нульовому водяному знаку. В роботі розглядається двохфакторна схема автентифікації що в якості фактору використовує "знання чогось" (пароль) та "володіння чимсь" (цифрове RGB-зображення).

У якості алгоритму нульового знаку обраний метод заснований на перетвореннях DWT та K-means з додатковим використанням Swish-функції. Дослідження проводилися шляхом розгляду теоретичної складності алгоритму при знанні зловмисником його параметрів, таких як: пароль, геш-значення паролю, зображення, еталонний водяний знак, результат перетворення та інших параметрів.

Попередні дослідження показали високу теоретичну стійкість даної схеми, яка базується на складності пароля та на розмірності зображень. При великій розмірності (для зображень розміром 512x512 та більше пікселів) забезпечується досить висока криптографічна стійкість. Проте теоретична стійкість не доведена, і значення можуть бути набагато нижчими, оскільки специфіка зображень та перетворень може створювати додаткові слабкі місця. Алгоритм має досить велику колізію, пов'язану з перетвореннями цифрових зображень та множенні матриць, що зменшує його стійкість.

Схеми автентифікації та алгоритми нульового водяного знаку потребують подальшого дослідження, доведення криптографічних параметрів та методів інтеграції в системи контролю доступом, оскільки вони можуть забезпечувати досить високу стійкість автентифікації в системах з високим рівнем шумів, а висока зручність і дешевизна даних схем надає перевагу перед іншими методами автентифікації. В роботі надаються рекомендації, щодо покращення можливих характеристик алгоритму.

***Ключові слова**: RGB, нульовий водяний знак, автентифікація, пароль, DWT, K-means.*

PODDUBNYI VADYM
SIEVIERINOV OLEKSANDR
Kharkiv National University of Radio Electronic

ROBUSTNESS OF AN AUTHENTICATION SCHEME BASED ON ZERO-WATERMARKING

The article presents an analysis of the robustness of an authentication scheme based on zero watermarking. The study examines a two-factor authentication scheme that uses "knowledge of something" (a password) and "possession of something" (a digital RGB image) as its factors.

The zero watermarking algorithm chosen is based on DWT and K-means transformations, with additional use of the Swish function. The analysis is conducted by considering the theoretical complexity of the algorithm assuming the adversary knows its parameters, such as the password, the hash of the password, the image, the reference watermark, the transformation result, and other parameters.

Previous studies have shown a high theoretical robustness of the scheme, which relies on the complexity of the password and the dimensionality of the image. For large image sizes (512×512 pixels and above), a relatively high level of cryptographic resistance is achieved. However, this robustness is not formally proven, and the actual strength may be significantly lower due to the specifics of the images and transformations, which can introduce additional vulnerabilities. The algorithm is subject to a relatively high rate of collision, associated with digital image transformations and matrix multiplications, which weakens its resistance.

Authentication schemes and zero watermarking algorithms require further research, formal proof of cryptographic properties, and methods for integration into access control systems, as they can provide a high level of authentication robustness in systems with high noise levels. Additionally, the convenience and low cost of such schemes give them an advantage over other authentication methods. The study provides recommendations for improving the potential characteristics of the algorithm.

***Keywords:** RGB, zero watermarking, authentication, password, DWT, K-means.*

Стаття надійшла до редакції / Received 03.10.2025
Прийнята до друку / Accepted 15.11.2025

Постановка проблеми

Автентифікація – процедура встановлення належності користувачеві інформації в системі пред'явленого ним ідентифікатора. Автентифікація користувачів в інформаційно-комунікаційній системі є обов'язковим методом забезпечення рольової моделі користувачів. Існує безліч типів автентифікації: автентифікація процесів, користувачів, пристроїв ітд. Одним з найрозповсюдженішим типів є автентифікація користувачів, це може бути як і цифрова автентифікація (наприклад введення паролю до свого персонального комп'ютеру чи використання електронного підпису на сайті держпослуг тощо), так і автентифікація користувача для фізичного доступу (використання ключів, карток, біометрії в системах контролю доступом). Під час автентифікації користувача, останній пред'являє системі ідентифікатори засновані на одному з трьох факторів: знання чогось, володіння чимсь, або перевірка характеристики об'єкту. Поєднання декількох факторів при проходженні процедури автентифікації створює багатофакторну автентифікацію (MFA).

Багатофакторна автентифікація є необхідним методом автентифікації користувачів, так багато рекомендацій в сфері інформаційної безпеки встановлюють вимогу наявності мінімум двохфактроної автентифікації[1]. Однією із схем багатофакторної автентифікації є схема з використанням нульових водяних знаків. Такі схеми зазвичай базуються на декількох факторах, зазвичай це пароль в поєднанні з медіа або текстовими даними [2][3]. Проте все ще актуальна проблема атак на схеми автентифікації з використанням нульового водяного знаку та їх криптографічної стійкості. Дана стійкість повинна бути достатньою, оскільки від коректної автентифікації залежать права доступу користувачів в системі. Ситуація стає складнішою у зв'язку з природою водяного знаку, яка дозволяє виникнення шумів в медіаданих.

Метою роботи є аналіз стійкості схем багатофакторної автентифікації заснованих на алгоритмах нульового водяного знаку.

Вибір схеми та алгоритмів водяного знаку для аналізу

Для аналізу була обрана схема автентифікації з використанням зображення в якості контейнеру та пароль в якості ключа для водяного знаку. Був обраний алгоритм нульового знаку заснований на перетвореннях DWT та K-means. Даний алгоритм був розроблений з метою застосування в схемах автентифікації тому підходить для дослідження.

Схема автентифікації заборонована в роботі [4], алгоритм нульового водяного знаку запропонований в [5], в даних роботах було доведено потенційну придатність алгоритму для автентифікації користувачів, та доведено його стійкість до шумів та пошкоджень. На рис. 1 зображена загальна схема, на рис. 2 зображена схема роботи самого алгоритму.

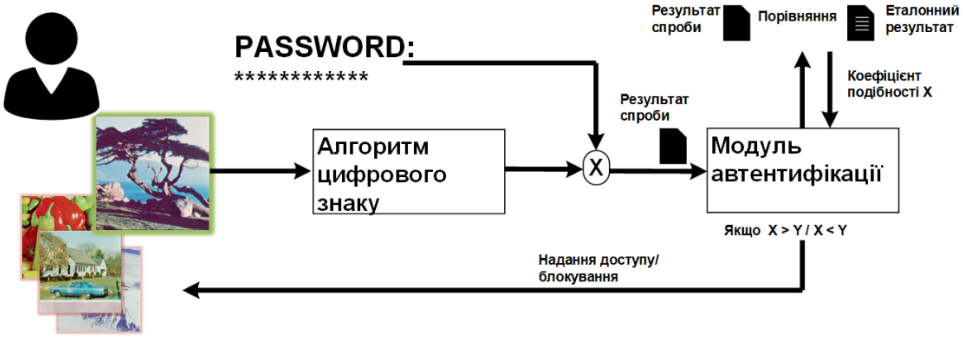

Рис. 1. Загальна схема автентифікації заснована на водяному знакові

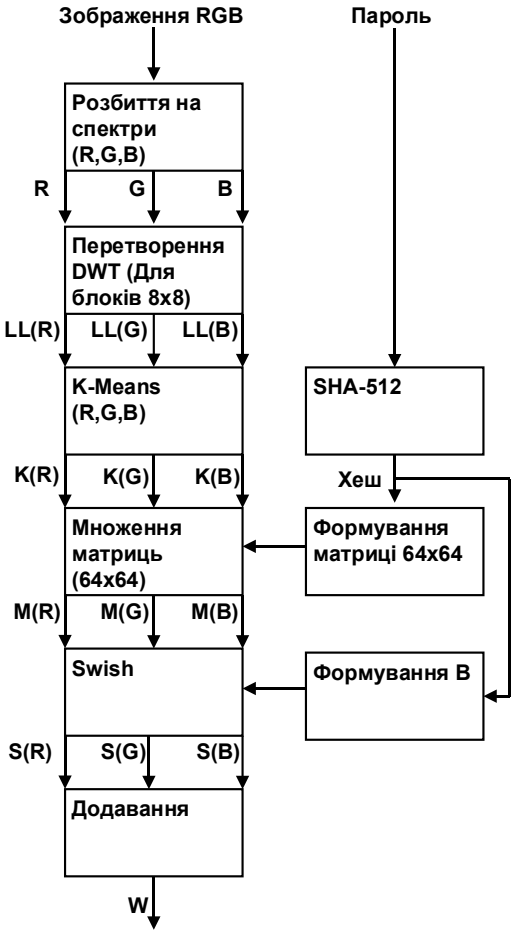

Рис. 2. Схема роботи алгоритму нульового водяного знаку

Метою роботи є розгляд ймовірних атак та підрахунок стійкості обраного алгоритму нульового знаку в схемах автентифікації користувачів. Додатковою метою є надання рекомендацій щодо використання даного алгоритму (та аналогічних йому за структурою) в схемах автентифікації.

Результати аналізу

В обраній схемі з запропонованим алгоритмом існують наступні дані:

– зображення користувача (ключові дані що використовуються для автентифікації);
– пароль (ключові дані що використовуються для автентифікації);
– геш від пароля (використовується для пришвидшення перевірки);
– еталонний водяний знак (водяний знак що зберігається в модулі автентифікації як еталона версія);
– коефіцієнт подібності (використовується виключно в модулі автентифікації при прийнятті рішень, для користувача доступна тільки значення 0 – не пройшов автентифікацію та 1 – автентифікацію пройдено).

Деякі параметри даних наведені в таблиці 1.

Таблиця 1

Параметри даних алгоритму

Параметр	**Діапазон даних (1 символ)**	**Розмір набору даних**
Зображення	0-255	X*Y*3, де X- ширина зображення, Y довжина
Пароль	1 байт	64 байти
Водяний знак	0-12 484 800	(X*Y)/2*3 де X- ширина зображення, Y довжина
Геш від пароля	0-255	64

Розглянемо декілька варіантів в залежності від обізнаності зловмисника.

Чорна коробка. Зловмисник не знає принципів роботи алгоритму водяного знаку (чорна коробка) з нульовим значенням ключових та внутрішніх даних. Це найбільш реалістичний сценарій ґрунтується на тому, що зловмисник не знає принципів роботи модуля автентифікації, має доступ лише до засобів вводу даних автентифікації та отримує результат лише у вигляді "успіху" або "невдачі".

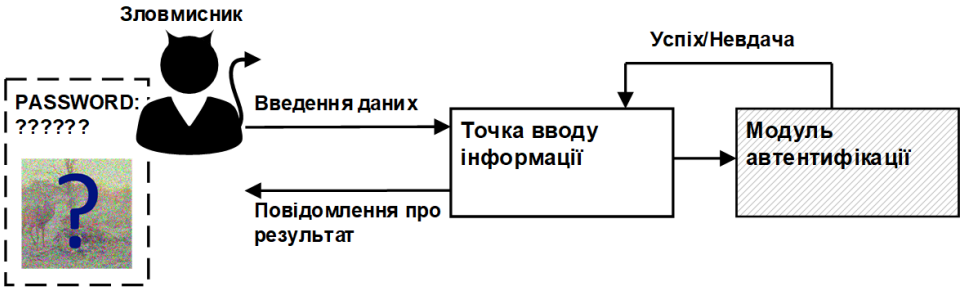

Рис. 3. Загальна схема "Чорна коробка"

При такому варіанті зловмисник може здійснювати атаки лише перебором, крипостійкість такої схеми базується на:

– складності пароля;
– обраному зображенні.

Іншим чином зловмиснику необхідно обрати таке зображення *I* та пароль *P*, щоб одночасно в комбінації вони утворювали необхідний водяний знак.

Щодо паролю, зловмисник не знає довжини паролю, його складності, типів символів що використовуються та інших параметрів. Математична складність перебору пароля:

$$S(pass) = y^c, \quad (1)$$

де *c*- кількість символів в паролі; *y* – розмір алфавіт (діапазон символів які може набувати пароль).

Розглянемо систему з нульовим рівнем шуму (в якій коефіцієнт достовірності дорівнює 1).

Характеристики зображення дорівнюють:

$$S(image) = 256^{M*M*3}=2^{24M^2}, \quad (2)$$

де 256 – значення яскравості кожного біту; *M* – розмір зображення; 3 – кількість спектрів (R,G,B).

Але оскільки під час перетворень виникають колізії то зловмиснику не необхідно перебирати всі зображення.

Після перетворення Хаара[7]отримується значення діапазону:

$$S\big(HAAR(image)\big) = 256^{\frac{M}{2}*\frac{M}{2}*3} = 2^{6M^2}, \quad (3)$$

Отже це найменш необхідний діапазон для перебору це кількість зображень яка дорівнює LL, оскільки одне LL може проектуватися у декілька зображень. Проте зловмисник не знає чи обране зображення генерує нове значення з пулу 2^{6M^2}, чи те яке вже проходило перевірку, отже це мінімальний необхідний пул перебору за умови що всі зображення будуть генерувати унікальні значення LL.

Після цього використовується перетворення K-Means[8], яке також зменшує та групує дані, проте у K-Means кожен піксель групується в залежності від кластеру для якого він потрапив, в алгоритмі використовуються кількість кластерів яка дорівнює 5, тобто кількість згрупувань в загальному випадку наведена в формулі нижче, оскільки значення однакове для всіх пікселів спектру.

Тобто в загальному мінімальна необхідність перебору K-Means

$$S(K\_means) = K^{M^2}, \quad (4)$$

Формула для перетворення K-Means після перетворення Хаара виглядає наступним чином:

$$S(K\_means(HAAR(image))) = K^{M^2/4}, \quad (5)$$

Отже чим більша кластеризація тим більше складність підбору, проте вона ніколи не більше 2^{6M^2}, оскільки це максимальний пул зображень після перетворення Хаара. Проте зловмисник не знає чи обране зображення генерує нове значення з пулу $K^{M^2/4}$, чи те яке вже проходило перевірку, отже це мінімальний необхідний пул перебору за умови що всі зображення будуть генерувати унікальні значення K-Means.

Під час наступного множення матриці зображення на матрицю пароля також втрачається набір даних.

Кількість можливих проекцій матриці A при множенні на вироджену матрицю B в загальному випадку:

$$\mathrm{A}' = \mathrm{c}^{M*r}, \quad (6)$$

де *c*- кількість значень елементів матриці; *M* - кількість рядків матриці; *r* – ранг матриці B.

Тобто для матриці B в яких повторюються рядки ранг матриці *r=1*, та зображення A:

$$\mathrm{A}' = 256^M, \quad (7)$$

Як вже було з'ясовано, розмірність зменшується в два рази під час перетворення Хаара, тому кінцева стійкість:

$$S` = 2^{4M}, \quad (8)$$

Проте оскільки на відміну від K-Means використовуються 3 матриці окремо які сумуються в кінці, то загальна кількість комбінацій матриць:

$$S` = 3 * 2^{4M}, \quad (9)$$

Проте значення не може бути більшим за $K^{M^2/4}$.

Тобто кількість можливих матриць A' після роботи алгоритму дорівнює $3 * 2^{4M}$. Це мінімальне (проте не гарантоване) значення з набору зображень 2^{24M^2} яке необхідно перебрати зловмиснику. Стійкість падає у зв'язку з появою колізій у перетвореннях та відображенням матриці A в певний пул матриць A'. Проте зловмисник не може знати чи зображення яке він використовує зараз перетворюється у нову матрицю чи вже в ту яку він перетворював. Іншими словами $3 * 2^{4M}$ – це мінімальний перебір за умови що кожне перетворення буде призводити до унікального A'.

Додатково слід розглянути варіант того що значення зображень *I*, які пройдуть перевірку розподіляються не рівномірно по всій площині, оскільки алгоритм розрахований на відсічення шумів.

Тобто мінімально необхідно перебрати $3 * 2^{4M}$ з 2^{24M^2}, значення розподіляються не рівномірно, та можливий перебір по діапазонам даних.

Дане значення зменшується в системах з шумом, так при встановленні коефіцієнту достовірності наприклад в 0.8, стійкість зменшується до $3 * 2^{3.2M}$ мінімальних значень.

Перетворення Swish не враховується в крипостійкості, оскільки не призводить до появи колізій а слугує лише для згладжування результатів.

При врахуванні стійкості паролю, стійкість системи дорівнює;

$$S(I + pass) = y^c * 3 * 2^{4M},$$

Часткове знання. Нехай зловмисник знає результат перетворень, а саме: згенеровану матрицю A', що утворилася після послідовного перетворення зображення I за допомогою перетворення Хаара, K-Means, множення матриць. Тоді в доповнення до чорної коробки це полегшує перебір, оскільки зловмисник може відсікати зображення які вже призводили до відомого результату, уникаючи повторень. Але все ще мінімальна стійкість системи:

$$S(I + pass) = y^c * 3 * 2^{4M},$$

Нехай зловмисник знає зображення. Тоді йому доступне зображення, його перетворення Хаара та K-Means та відкритий водяний знак. Позначмо на схемі відомі та невідомі параметри, як показано на Рис 4 (червоне – те що не знає, зелене – те що знає, синє - опціонально).

Тоді складність зламу полягає в находженні пароля, тобто в переборі значень паролю.

Нехай зловмисник додатково знає додатково водяний знак W, який зберігається в базі даних КЗЗ. Як видно, зі схеми зловмиснику необхідно знайти такий геш від паролю, при якому сума матриць для кожного спектру після перетворення Switch (коефіцієнтом якого являється частина гешу) та поблокового множення на відоме зображення буде давати наближене до водяного знаку. Та навіть розв'язавши цю задачу він отримає лише геш-значення від паролю.

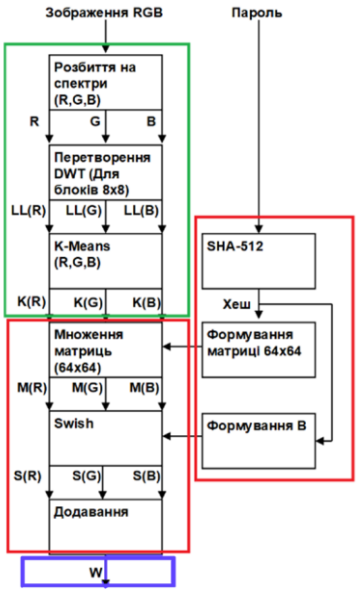

Рис. 4. Дані що знає зловмисник при знанні зображення

Нехай зловмисник знає пароль, позначено дані на Рис. 5 які він знає (червоне – те що не знає, зелене – те що знає, синє - опціонально).

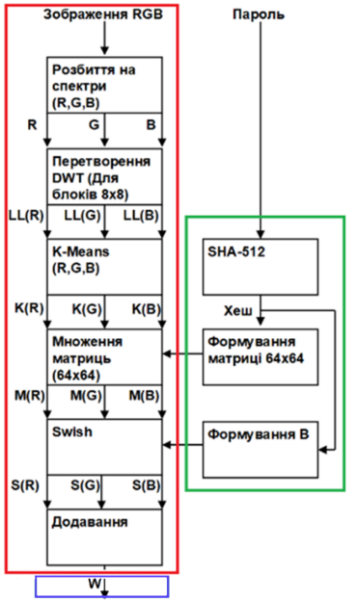

Рис. 5. Дані що знає зловмисник при знанні паролю

Тоді зловмиснику необхідно знайти таке зображення щоб після перетворення Хаара та K-Means після поблокового множення кожного спектру на відому матрицю та після функції Switch з відомим коефіцієнтом, сума даних спектрів була наближена до водяного знаку. Додатково слід визначити розмір зображення, оскільки даний параметр є невідомим.

Якщо зловмисник додатково знає водяний знак то ця задача спрощується, оскільки можливе обернене проходження алгоритм, проте під час знаходження водяного знаку використовується функція Switch[8], це ускладнює оберненість алгоритму, оскільки знаходження оберненого значення до Switch-функції є більш складною задачею, враховуючи що необхідно знайти всі Switch-значення матриці. В даному випадку можлива побудова таблиці значень Switch функції, за відомими геш-значеннями паролю.

Також оберненість ускладняється тим що відбувається множення матриць, що ускладнює оберненість. Далі необхідно знайти обернені функції K-Means та Хаара.

Як можна побачити стійкість заснована на тому, що:

– значення розмиваються сумуванням елементів трьох матриць;

– знаходження оберненої функції Switch складніше ніж прямої;

– додавання коефіцієнту з гешу в Switch додає невизначеність, оскільки одному значенню x може відповідати декілька y;

– знаходження матриці A, що при множенні на вирождену матрицю B дає матрицю C є складною задачею;

– алгоритм повторюється для кожного спектру зображення та для кожного блоку зображення.

Висновок

Отже запропонована схема автентифікації з використанням алгоритму нульового водяного знаку має теоретично високу крипостійкість, яка основним чином залежить від розміру зображення та складності паролю. При значеннях зображення в 512x512 пікселів, теоретична стійкість перевищує значення криптографічних алгоритмів.

Проте є певні параметри які можуть знизити реальну стійкість:

– параметри пароля не контролюються, можуть існувати слабкі паролі вигадані користувачем;

– в реальному світі зображення підпорядковуються певним законам, тому можливе усічення пулу можливих значень для перебору;

– набір зображення що проходять перевірку не розподілений рівномірно по всьому пулу зображень, що теоретично може спростити перебір;

– можливі появи "слабких" ключів при використання паролів та зображень що в результаті перетворень дають специфічні результати.

Для покращення стійкості схеми автентифікації на основі нульового водяного знаку та паролю рекомендується:

– перевіряти паролі та зображення за допомогою політики безпеки (розмір зображення, складність паролю);

– використовувати різні функції гешування в алгоритмі та для швидкої перевірки паролю, оскільки значення гешу від пароля полегшує криптоаналіз.

– не розповсюджувати еталон водяний знак, оскільки це спрощує криптоаналіз, проте схема з відкритим водяним знаком все ще має достатню криптостійкість;

– використовувати не надто великі значення (стійкість обмежена перетворенням матриці) та не надто малі оскільки зменшення кількості кластеризацій в K-Means зменшує стійкість схеми.

Запропоновані схема та алгоритми, можуть слугувати цікавим рішенням в системах що мають певні рівні шумів, проте потребують більш детального дослідження та підтвердження криптографічної стійкості.

Література

1. Адміністрація Держспецзв'язку. Про визначення Базового профілю безпеки інформації: Наказ від 24.06.2024 № 317 . – [Електронний ресурс]. – Режим доступу: https://cip.gov.ua/ua/docs/nakaz-administraciyi-derzhspeczv-yazku-vid-24-06-2024-317-pro-viznachennya-bazovogo-profilyu-bezpeki-informaciyi.;

2 Jalil Z, Mirza AM, Sabir M. Content based zero-watermarking algorithm for authentication of text documents // International Journal of Computer Science and Information Security. – 2010. – Т. 7, № 2. – С. 171–177. – DOI:10.48550/arXiv.1003.17963.

3. Bashardoost M, Rahim MSM, Hadipour N. A novel zero-watermarking scheme for text document authentication // Jurnal Teknologi (Sciences & Engineering). – 2015. – Т. 75, № 4. – С. 49–56. – DOI:10.11113/jt.v75.5066.

4. Поддубний ВО, Сєвєрінов ОВ, Гвоздьов РЮ. Використання нульових водяних знаків для підтвердження авторства зображень та багатофакторної автентифікації // Радіотехніка. – 2024. – Вип. 218. – С. 35–43. – DOI:10.30837/rt.2024.3.218.02

5. Поддубний В, Сєвєрінов О. Метод автентифікації користувачів в інформаційно-комунікаційних системах із застосуванням нульового водяного знака // Сучасний стан наукових досліджень та технологій в промисловості. – 2025. – № 2(32). – С. 69–78. – DOI:10.30837/2522-9818.2025.2.069.

6. Sethi N, Krishna R, Arora RP. Image compression using Haar wavelet transform [Електронний ресурс] // Computer Engineering and Intelligent Systems. – 2012. – Т.2, № 3. – С. 26–31. – Режим доступу: https://scispace.com/pdf/image-compression-using-haar-wavelet-transform-1q7fxnv973.pdf

7. Accord.NET Framework. KMeans Class [Електронний ресурс]. – Режим доступу: http://accord-framework.net/docs/html/T_Accord_MachineLearning_KMeans.htm (дата звернення: 08.09.2025)

8. Ramachandran P, Zoph B, Le QV. Swish: a self-gated activation function [Електронний ресурс]. – 2017. – arXiv:1710.05941. – Режим доступу: https://arxiv.org/pdf/1710.05941v1.pdf

References

1. Administratsiia Derzhavnoi sluzhby spetsial′noho zv'iazku ta zakhystu informatsii Ukrainy. Pro vyznachennia Bazovoho profiliu bezpeky informatsii: Nakaz vid 24.06.2024 № 317 [Electronic resource]. – Available at: https://cip.gov.ua/ua/docs/nakaz-administraciyi-derzhspeczv-yazku-vid-24-06-2024-317-pro-viznachennya-bazovogo-profilyu.

2. Jalil Z, Mirza AM, Sabir M. Content based zero-watermarking algorithm for authentication of text documents // International Journal of Computer Science and Information Security. – 2010. – Vol. 7, No. 2. – P. 171–177. – DOI:10.48550/arXiv.1003.17963

3. Bashardoost M, Rahim MSM, Hadipour N. A novel zero-watermarking scheme for text document authentication // Jurnal Teknologi (Sciences & Engineering). – 2015. – Vol. 75, No. 4. – P. 49–56. – DOI:10.11113/jt.v75.5066;

4. Poddubnyi VO, Sievierinov OV, Hvoz′dov RIu. Vykorystannia nul′ovykh vodianykh znakiv dlia pidtverdzhennia avtorstva zobrazhen′ ta bahatofaktornoi avtentyfikatsii // Radiotekhnika. – 2024. – Vyp. 218. – P. 35–43. – DOI:10.30837/rt.2024.3.218.02.

5. Poddubnyi V, Sievierinov O. Metod avtentyfikatsii korystuvachiv v informatsiino-komunikatsiinykh systemakh iz zastosuvanniam nul′ovoho vodianoho znaka // Suchasnyi stan naukovykh doslidzhen′ ta tekhnolohii v promyslovosti. – 2025. – № 2(32). – P. 69–78. – DOI:10.30837/2522-9818.2025.2.069

6. Sethi N, Krishna R, Arora RP. Image compression using Haar wavelet transform [Electronic resource] // Computer Engineering and Intelligent Systems. – 2012. – Vol. 2, No. 3. – P. 26–31. – Access mode: https://scispace.com/pdf/image-compression-using-haar-wavelet-transform-1q7fxnv973.pdf;

7. Accord.NET Framework. KMeans Class [Electronic resource]. – Access mode: http://accord-framework.net/docs/html/T_Accord_MachineLearning_KMeans.htm;

8. Ramachandran P, Zoph B, Le QV. Swish: a self-gated activation function [Electronic resource]. – 2017. – arXiv preprint arXiv:1710.05941. – Access mode: https://arxiv.org/pdf/1710.05941v1.pdf.