Vol. 6, No. 1, 2024

DOI: https://doi.org/10.23939/csn2024.01.124

УДК 004: 054

КОНЦЕПТУАЛЬНИЙ ПІДХІД ДО ВИЯВЛЕННЯ DEEPFAKE-МОДИФІКАЦІЙ БІОМЕТРИЧНОГО ЗОБРАЖЕННЯ ЗАСОБАМИ НЕЙРОННИХ МЕРЕЖ

Г. В. Микитин, Х. С. Руда

Національний університет "Львівська політехніка",
кафедра захисту інформації
E-mail: halyna.v.mykytyn@lpnu.ua, khrystyna.s.ruda@lpnu.ua

Національний кластер кібербезпеки України функціонально орієнтований на побудову систем захисту різних платформ інформаційної інфраструктури, зокрема створення безпечних технологій виявлення deepfake-модифікацій біометричного зображення на основі нейронних мереж у кіберпросторі.

У цьому просторі запропоновано концептуальний підхід до виявлення deepfake-модифікацій, який розгорнуто на основі функціонування згорткової нейронної мережі та алгоритму роботи класифікатора біометричних зображень за структурою "чутливість – показник Юдена – оптимальний поріг – специфічність".

Представлено аналітичну структуру безпеки нейромережевих інформаційних технологій (ІТ) на основі багаторівневої моделі "ресурси – системи – процеси – мережі – управління" відповідно до концепції "об'єкт – загроза – захист". Ядром безпекової структури ІТ є цілісність функціонування нейромережевої системи виявлення deepfake-модифікацій біометричного зображення обличчя людини і системи аналізу даних, що реалізують інформаційний процес "розділення відеофайлу на кадри – детекція, опрацювання ознак – оцінювання точності класифікатора зображень".

Розроблено конструктивний алгоритм виявлення deepfake-модифікацій біометричних зображень: розбиття відеофайлу біометричного зображення на кадри – розпізнавання детектором – відтворення нормалізованих зображень облич – обробка засобами нейронної мережі – обчислення матриці ознак – побудова класифікатора зображень.

Ключові слова: біометричне зображення, deepfake-модифікації, інформаційна нейромережева технологія, згорткова нейронна мережа, класифікація, система підтримки прийняття рішення, концептуальний підхід, аналітична структура безпеки.

Вступ

У сучасних умовах гібридної війни проблема виявлення deepfake-модифікацій біометричних зображень набуває все більшої актуальності не лише в Україні, а й у всьому світі. Сучасні системи підміни особи – діпфейки [1], застосовуються з метою пропаганди і введення населення в оману, генеруючи правдоподібні звернення офіційних осіб, що несуть потрібні супротивнику наративи. Проте сфера використання цих систем не обмежується пропагандою [2], вони також можуть бути застосованими для встановлення несанкціонованого доступу до конфіденційної інформації. Це становить сьогодні значну небезпеку не лише для особи, чию конфіденційність було порушено, а й для держави загалом, що і відображено в Стратегії кібербезпеки України [3]. Виявлення такого роду загроз є значною проблемою, оскільки регулярно з'являються все більш реалістичні технології створення підробок, для яких потрібні більш новітні інформаційні технології виявлення deepfake-модифікацій біометричного зображення.

1. Огляд літературних джерел

У просторі задач цифрової трансформації суспільства за стратегією інтелектуалізації є актуальною задача забезпечення конфіденційності біометричного зображення в різних предметних сферах інфраструктури суспільства [4]. В цьому напрямку вчені розглядають застосування технологій виявлення deepfake-модифікацій біометричного зображення. Наприклад, в [5] автори визначають оригінальність цифрових зображень, використовуючи фрактальний характер цифрових сигналів. У цьому випадку перевірка оригінальності цифрових біометричних зображень відбувається у взаємозв'язку до ідентифікаційних функціональних характеристик цифрової апаратури їх відеозапису. Водночас відбувається порівняння записів, збережених на носіях або в пам'яті цифрової апаратури відеозапису в електронному вигляді. Якщо ідентифікаційні ознаки експериментального і досліджуваного записів збігаються, то останній є оригінальним.

У методі, що описується в [6], deepfake-модифікації виявляються за допомогою локалізації обличчя людини і правильного виявлення зони навколо нього для аналізу. З'ясовано, що під час заміни обличчя його біометричне зображення буде складатись з трьох зон: фонового оригінального зображення, заміненого обличчя, зони переходу, яка буде згладжувати межу між першими двома. Система, побудована на основі цього методу, реалізує пошук артефактів, які з'являються в зоні накладання облич і на основі їх виявлення або невиявлення робить висновок про те, оригінальне чи модифіковане зображення їй представили.

Інший метод, розглянутий в [7], ґрунтується на тому, що поточний deepfake-алгоритм не може генерувати зображення з роздільною здатністю, вищою за певне прийняте, згідно з відповідними критеріями, порогове значення. Отже, ці зображення обмеженої роздільної здатності згодом потрібно додатково трансформувати за допомогою афінних перетворень для відповідності обличчям, що будуть замінені у вихідному відео. Такі перетворення залишають певні характерні артефакти в отриманих модифікованих deepfake-відеофайлах, які можна ефективно зареєструвати моделлю нейронної мережі, побудованою за відповідними параметрами. В праці [8] аналізується ступінь відмінності невідфільтрованих послідовних кадрів між собою за допомогою дисперсії і відбувається класифікація не біометричних зображень самих по собі, а класифікація показників дисперсії. Автори статті також обґрунтовують неефективність застосування згорткової нейронної мережі тим, що така мережа, маючи в своїй будові згортковий фільтр, ігнорує вплив розмиття, яскравості, контрасту та шуму і відповідно втрачає частину даних, потрібних для навчання класифікатора. Розглянуті методи стають основою для розвитку підходів до виявлення deepfake-модифікацій біометричного зображення на основі згорткової нейронної мережі.

2. Постановка завдання

Виходячи з проведеного аналізу, було встановлено завдання цієї роботи: 1) запропонувати ідеологію виявлення deepfake-модифікацій біометричного зображення основі нейромережевих ІТ на рівні концептуального підходу; 2) створити модель безпеки виявлення deepfake-модифікацій на рівні аналітичної структури багаторівневої нейромережевої ІТ; 3) розробити методологію виявлення deepfake-модифікацій біометричного зображення засобами нейронних мереж на рівні конструктивного алгоритму. Мета статті – розроблення концептуального підходу, який забезпечуватиме точність класифікації біометричних зображень на основі нейромережевих ІТ у просторі виявлення deepfake-модифікацій та конфіденційність даних у сегменті розв'язання основних задач кібербезпеки.

3. Концептуальний підхід до виявлення deepfake-модифікацій біометричного зображення на основі нейромережевих інформаційних технологій

З метою цілісного представлення проблеми виявлення deepfake-модифікацій для забезпечення конфіденційності біометричних зображень та розв'язання цієї задачі з використанням інформаційних нейромережевих технологій запропоновано концептуальний підхід до виявлення deepfake-модифікацій на основі згорткових нейронних мереж (рис. 1).

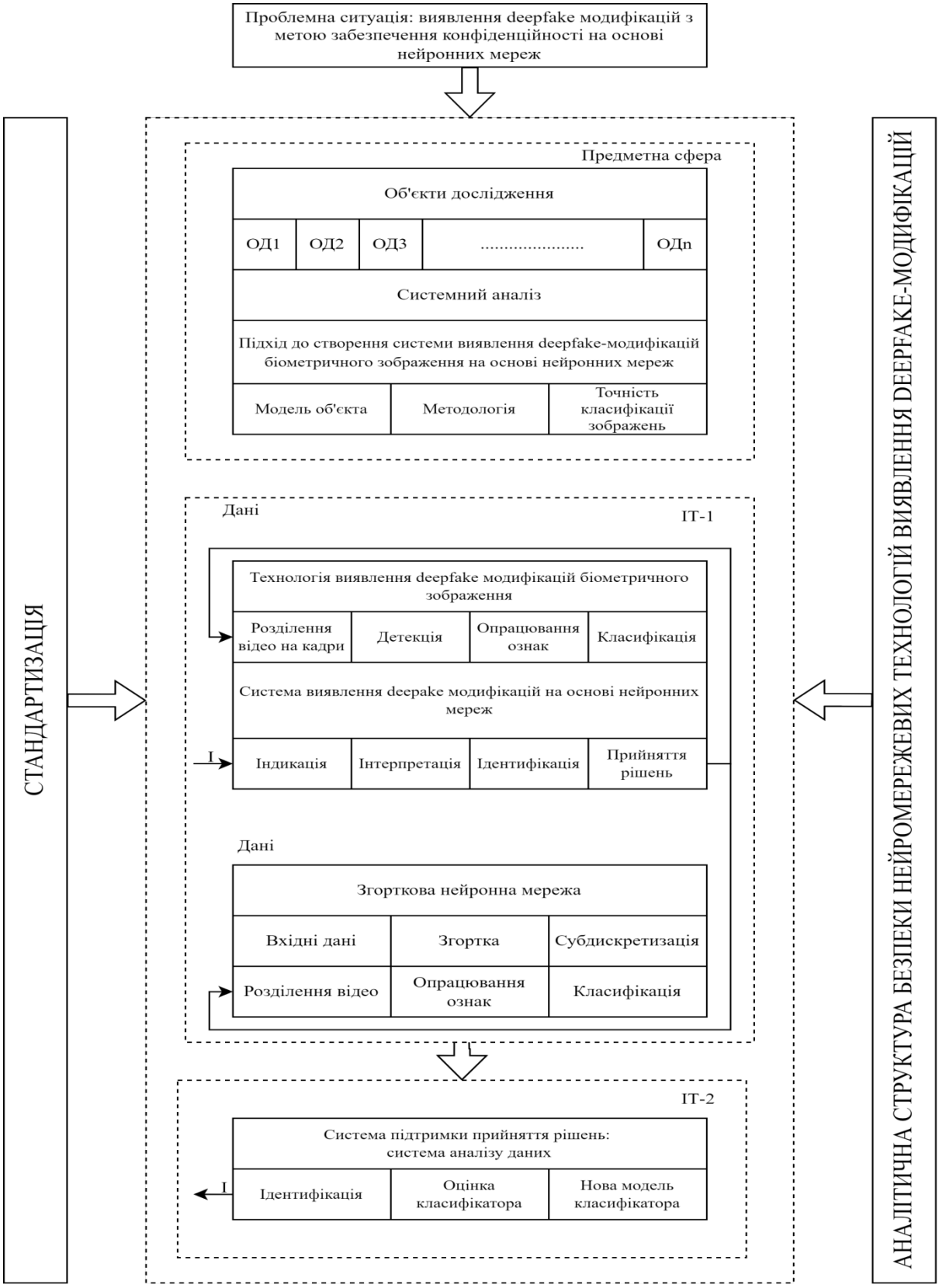

Рис. 1. Концептуальний підхід до виявлення deepfake-модифікацій на основі нейромережевих технологій

Розроблення інформаційних нейромережевих технологій виявлення deepfake-модифікацій біометричного зображення залежить від рівня структурованості завдань і потребує відповідних підходів у контексті гарантування безпеки об'єктів дослідження як біометричних зображень за профілем конфіденційності. Відповідно до проблемної ситуації в контексті виявлення deepfake-модифікацій біометричних зображень, концептуальний підхід до виявлення deepfake-модифікацій біометричного зображення має таку структуру: 1) предметна сфера інфраструктури суспільства – об'єкти дослідження (ОД1, ОД2, .., ОДn); системний аналіз (принципи цілісності, структурованості та ієрархічності); підхід до створення системи виявлення deepfake-модифікацій на основі згорткової нейронної мережі; 2) інформаційна нейромережева технологія (ІТ1) згідно із структурою "модель об'єкта – методологія – точність класифікації зображень"; 3) система підтримки прийняття рішень (ІТ2). Конструктивний алгоритм ІТ1 "розділення відео на кадри – детекція – опрацювання ознак – класифікація" реалізується системою на основі нейронних мереж за структурою "індикація – інтерпретація – ідентифікація – прийняття рішення", зокрема функціональністю згорткової нейронної мережі "вхідні дані – згортка – субдискретизація". Конструктивний алгоритм ІТ2 реалізується за допомогою системи аналізу даних для ідентифікації оцінки класифікатора і прийняття управлінського рішення на встановлення відповідності щодо виявлення deepfake-модифікації. Система підтримки прийняття рішення ідентифікує ознаки досліджуваного біометричного зображення, встановлюючи відповідність вибраній моделі класифікатора.

У разі невідповідності система аналізу даних приймає рішення для користувача про побудову нової моделі класифікатора. З метою забезпечення уніфікації методів створення системи виявлення deepfake-модифікацій на основі нейронних мереж у просторі забезпечення конфіденційності біометричних зображень структура концептуального підходу передбачає елементи стандартизації в галузі нейромережевих технологій, біометричних зображень, кібербезпеки за: ISO/IEC 30107[9], ISO/IEC TR 24029[10], ДСТУ ISO/IEC 15408[11], C2PA Specification [12]. Інструментарієм виявлення deepfake-модифікацій біометричного зображення є інформаційні нейромережеві технології, ядром яких є система. Основними етапами нейромережевої технології виявлення deepfake-модифікацій обличчя людини є:

- Розділення відео на кадри. Оскільки об'єктом дослідження є окремі біометричні зображення, а не відео загалом, для подальшого опрацювання потрібне виділення зображень з потоку кадрів.
- Детекція. На цьому етапі згорткова нейронна мережа виявляє обличчя людей на кадрі і виділяє їх для наступного етапу. Оскільки на наступному етапі відбувається дослідження зображення на основі біометричних ознак, доцільним є виокремлення біометричних зображень облич людей із загального фону.
- Опрацювання ознак. На цьому етапі спеціально спроєктована нейронна згорткова нейронна мережа генерує матрицю ознак кожного біометричного зображення на основі обчислених ваг, що були сформовані в процесі її навчання.
- Класифікація. Ваги, що були згенеровані на попередньому етапі, разом з класами, визначеними для кожного біометричного зображення, використовуються для навчання класифікатора.

За проведеним аналізом відомих підходів безпечного виявлення deepfake-модифікацій біометричних зображень запропоновано створення аналітичної структури безпеки нейромережевих інформаційних технологій виявлення deepfake-модифікацій біометричних зображень обличчя людини у просторі безпечної інтелектуалізації об'єктів інфраструктури суспільства. Ядром аналітичної структури безпечної нейромережевої інформаційної технології (рис. 2) є система виявлення deepfake-модифікацій біометричних зображень на основі нейронних мереж та система аналізу даних, які програмно орієнтовані на цілісну реалізацію інформаційного процесу "розділення відео на кадри – виявлення deepfake – обробка ознак – оцінювання класифікації зображень" і на цій основі прийняття рішення про достатню точність класифікатора deepfake-модифікацій відповідно до вибраної моделі з можливістю її оновлення.

4. Конструктивний алгоритм виявлення deepfake-модифікації біометричних зображень засобами згорткової нейронної мережі

Згорткова нейронна мережа. Беручи до уваги те, що об’єктом дослідження є біометричне зображення, для його опрацювання було вибрано саме згорткову архітектуру нейронної мережі, оскільки ця архітектура спроєктована для роботи із зображеннями [13].

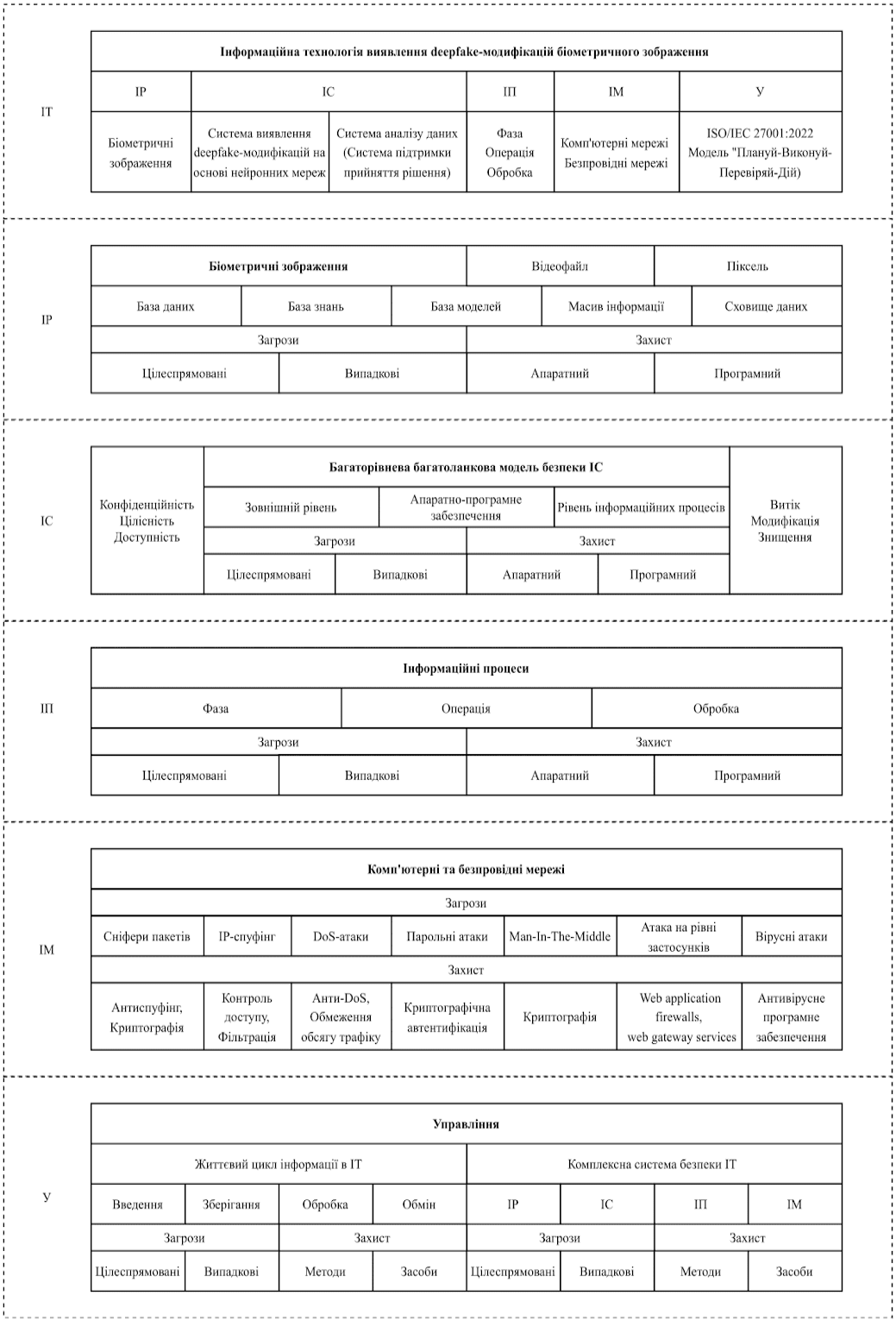

Рис. 2. Аналітична структура безпеки нейромережевих технологій виявлення deepfake-модифікацій

У пропонованому рішенні згорткові нейронні мережі застосовуються для детекції біометричних зображень, а також для обчислення значущих ознак біометричного зображення. Брак вихідного шару в мережі, що відповідає за опрацювання ознак, компенсується наявністю відокремленого класифікатора, який використовує обчислені мережею матриці ознак біометричних зображень. Вхідними даними для нейронної мережі виступають біометричні зображення, виділені з попередньо розділеного на кадри відеофайлу. На наступному етапі біометричні зображення опрацьовуються згортковою нейронною мережею, результатом чого є матриця ознак зображення.

Навчання класифікатора. Підхід до виявлення модифікованих зображень засобами згорткової нейронної мережі реалізується в кілька етапів. Перший етап полягає у розбитті кожного з відеофайлів на окремі кадри, які надалі будуть опрацьовані за допомогою детектора біометричних зображень. Детектор розпізнає об'єкт, після чого з кожного розпізнаного обличчя людини у кадрі відеозапису отримуємо біометричне зображення, нормалізоване до визначених розмірів з маркуванням, до якого класу (реального чи модифікованого) належить досліджуване зображення. Другий етап використовує нормалізовані зображення облич, які підлягають обробці згортковою нейронною мережею для обчислення векторів ознак, що використовуються для навчання класифікатора. Вони розподіляються на тренувальну і тестову частину для, власне, навчання і перевірки результатів.

В алгоритмі роботи класифікатора в просторі виявлення модифікованих біометричних зображень задіяні такі взаємопов'язані характеристики: точність, чутливість, специфічність класифікатора, оптимальне порогове значення класифікатора, показник Юдена. Показник Юдена є у взаємозв'язку з чутливістю та специфічністю класифікатора.

Результатом роботи класифікатора біометричних зображень є матриця невідповідностей у просторі: правильно визначені модифіковані зображення, правильно визначені немодифіковані зображення, неправильно визначені модифіковані зображення і неправильно визначені немодифіковані зображення.

Чутливість роботи класифікатора характеризує частку істинно позитивних біометричних зображень, які визначені правильно [14]:

$$TPR = \frac{TP}{TP + FN}, \quad (1)$$

де TPR – чутливість класифікатора, TP – кількість правильно класифікованих позитивних досліджуваних зразків, FN – кількість хибно класифікованих негативних досліджуваних зразків.

Досліджуване біометричне зображення незалежно визначається імовірністю належності до відповідного класу. Показник Юдена [15] використовується для визначення оптимального порогового значення класифікації зображень:

$$J = max\,(TPR(t) + TNR(t) - 1), \quad (2)$$

де J – показник Юдена, t – порогове значення правильного вибору біометричного зображення, TNR – специфічність класифікатора, що визначається як

$$TNR = \frac{TN}{TN + FP}, \quad (3)$$

де TN – кількість правильно класифікованих негативних досліджуваних зразків, FP – хибно класифікованих позитивних досліджуваних зразків.

На рис. 3 представлено порогове значення відсікання інформовано класифікованих зображень від неінформовано класифікованих у просторі взаємозв'язку індексу Юдена з чутливістю класифікатора (TPR) та його специфічністю TNR. Його максимальне значення зумовлює оптимальне порогове значення класифікатора, що забезпечує критерій збалансованості характеристик чутливості класифікатора і його специфічності.

Точність класифікації біометричних зображень у просторі виявлення deepfake-модифікацій зумовлюється кількістю досліджуваних тренувальних зображень, що уможливлює досягнення значення чутливості класифікатора більшим, ніж 0,85.

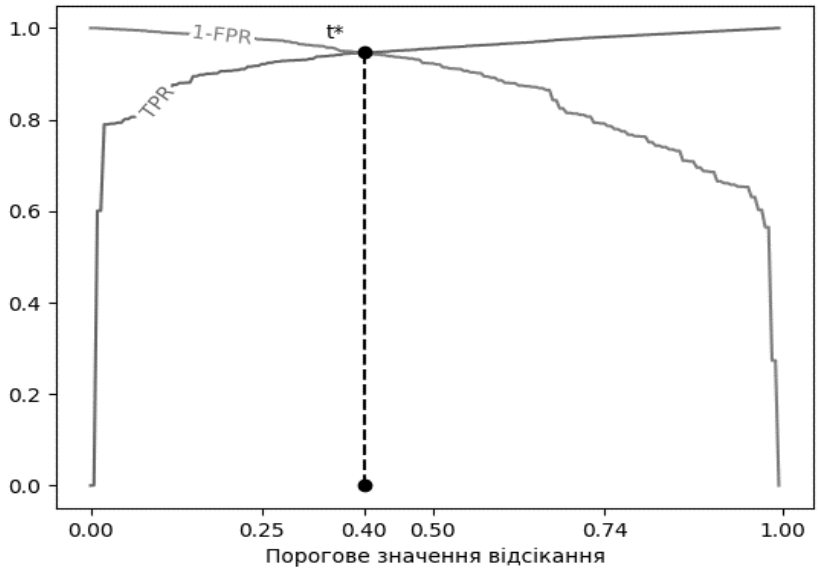

Рис. 3. Порогове значення відсікання інформовано класифікованих зображень від неінформовано класифікованих

6. Результати досліджень

Розроблено концептуальний підхід до виявлення deepfake-модифікацій біометричного зображення, що містить: модель об’єкта – методологію – точність класифікації зображень, систему виявлення deepfake-модифікацій “індикація – інтерпретація – ідентифікація – прийняття рішення” на основі згорткової нейронної мережі та алгоритмічно-програмну реалізацію обробки даних. Ці підходи відображаються на рівні аналітичної структури багаторівневої нейромережевої ІТ. Під час дослідження створено модель безпеки для виявлення deepfake-модифікацій біометричних зображень, що базується на нейромережевій аналітичній структурі. Також розроблено конструктивний алгоритм виявлення модифікованих біометричних зображень на основі згорткової нейронної мережі, який враховує параметри, такі як чутливість, показник Юдена, оптимальний поріг та специфічність.

Висновки

Запропонований концептуальний підхід до виявлення deepfake-модифікацій біометричного зображення на основі нейронних мереж дає можливість окреслити напрям досліджень – забезпечення конфіденційності інформації у просторі кібербезпеки біометричних зображень із застосуванням: 1) підходу до виявлення deepfake-модифікацій “модель об’єкта – методологія – точність класифікації зображень”; 2) системи виявлення deepfake-модифікацій “індикація – інтерпретація – ідентифікація – прийняття рішення” на основі згорткової нейронної мережі “розділення відео – опрацювання ознак – класифікація зображення”; 3) системи аналізу даних “ідентифікація – оцінка класифікатора – нова модель класифікатора”; 4) алгоритмічно-програмної реалізації обробки даних. Розвинуто модель безпеки виявлення deepfake-модифікацій біометричного зображення засобами нейронних мереж на рівні аналітичної структури багаторівневої нейромережевої ІТ. Створено конструктивний алгоритм виявлення модифікованих біометричних зображень засобами згорткової нейронної мережі на основі навчання класифікатора за структурою “чутливість – показник Юдена – оптимальний поріг – специфічність”.

Список літератури

1. Kietzmann J., Lee L. W., McCarthy I. P., and Kietzmann T. C. Deepfakes: Trick or treat? Business Horizons, vol. 63(2), pp. 135–146, 2020. DOI: 10.1016/j.bushor.2019.11.006

2. Yevseiev S., Ponomarenko V., Laptiev O., Milov O., Korol O., Milevskyi S. et. al.; Yevseiev S., Ponomarenko V., Laptiev O., Milov O. (eds.) Synergy of building cybersecurity systems. Kharkiv: PC TECHNOLOGY CENTER, p.188, 2021. DOI: https://doi.org/10.15587/978-617-7319-31-2

3. Стратегія кібербезпеки України (2021–2025). [Електронний ресурс] Available at: https://www.rnbo.gov.ua/files/2021/STRATEGIYA%20KYBERBEZPEKI/proekt%20strategii_kyberbezpeki_Ukr.pdf. (Accessed: 19 March 2024)

4. Karpinski M., Khoma V., Dudvkevych V., Khoma Y. and Sabodashko D. Autoencoder Neural Networks for Outlier Correction in ECG- Based Biometric Identification, 2018 IEEE 4th International Symposium on Wireless Systems within the International Conferences on Intelligent Data Acquisition and Advanced Computing Systems (IDAACS-SWS), pp. 210–215, 2018. DOI: 10.1109/IDAACS-SWS.2018.8525836

5. Rybalskyi O. V., Soloviev V. Y. On the development of the theory, methods and means of conducting the examination of digital photo, video and sound recording materials, methods and means of conducting the examination of digital photo, video and sound recording materials". Modern Special Technique, vol. 3 (30), pp. 119–121, 2012. [In Russian]. Available at: http://nbuv.gov.ua/UJRN/sstt_2012_3_19 (Accessed on 19 March 2024).

6. Li L., Bao J., Zhang T., Yang H., Chen D., Wen F., & Guo B., Face X-ray for more general face forgery detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 5001–5010, 2020. DOI: https://doi.org/10.48550/arXiv.1912.13458

7. Li Y., Lyu S. Exposing deepfake videos by detecting face warping artifacts. In Proceedings of the IEEE 14 Conference on Computer Vision and Pattern Recognition Workshops, pp. 46–52, 2019. DOI: https://doi.org/10.48550/arXiv.1811.00656

8. Lee G., Kim M. Deepfake Detection Using the Rate of Change between Frames Based on Computer Vision. Sensors vol. 21:7367, 2021. DOI: https://doi.org/10.3390%2Fs21217367

9. (2016) ISO/IEC 30107-1: Information technology – Biometric presentation attack detection – Part 1: Framework.

10. SC42 WG3: Assessment of the robustness of neural networks – part 1: Overview.Tech. Rep. CD TR 24029-1, ISO/IEC JTC 1/SC 42 Artificial Intelligence (2019).

11. (2017) DSTU ISO/IEC 15408-1: Information security, cybersecurity and privacy protection. Evaluation criteria for IT security. Part 1: Introduction and general model.

12. C2PA. 2020. Coalition for Content Provenance and Authenticity. [Електронний ресурс] Available at: https://c2pa.org/ (Accessed: 19 March 2024).

13. Albawi S., Mohammed T. A. and Al-Zawi S. Understanding of a convolutional neural network, International Conference on Engineering and Technology (ICET), pp. 1–6, 2017. DOI: https://doi.org/10.1109/ICEngTechnol. 2017.8308186

14. Yerushalmy S. Statistical problems in assessing methods of medical diagnosis, with special reference to x-ray techniques. Public Health Rep, 1947. DOI: https://doi.org/10.2307/4586294

15. Schisterman E. F., Perkins N. J., Liu A., Bondell H. Optimal cut-point and its corresponding Youden Index to discriminate individuals using pooled blood samples. Epidemiology. vol. 16 (1), pp. 73–81, 2005. DOI: https://doi.org/10.1097/01.ede.0000147512.81966.ba

CONCEPTUAL APPROACH TO DETECTING DEEPFAKE MODIFICATIONS OF BIOMETRIC IMAGES USING NEURAL NETWORKS

H. Mykytyn, K. Ruda

Lviv Polytechnic National University,
Department of Cybersecurity

The National Cybersecurity Cluster of Ukraine is functionally oriented towards building systems to protect various platforms of information infrastructure, including the creation of secure technologies for detecting deepfake modifications of biometric images based on neural networks in cyberspace.

This space proposes a conceptual approach to detecting deepfake modifications, which is deployed based on the functioning of a convolutional neural network and the classifier algorithm for biometric images structured as "sensitivity-Yuden index-optimal threshold-specificity".

An analytical security structure for neural network information technologies is presented based on a multi-level model of "resources-systems-processes-networks-management" according to the concept of "object-threat-defense". The core of the IT security structure is the integrity of the neural network system for detecting deepfake modifications of biometric face images as well as data analysis systems implementing the information process of "video file segmentation into frames-feature detection, processing – classifier image accuracy assessment".

A constructive algorithm for detecting deepfake modifications of biometric images has been developed: splitting the video file of biometric images into frames – recognition by the detector – reproduction of normalized facial images – processing by neural network tools – feature matrix computation – image classifier construction.

Keywords: biometric image, deepfake modifications, neural network technology, convolutional neural network, classification, decision support system, conceptual approach, analytical security structure.