Informatyka techniczna i sztuczna inteligencja
Technical informatics and Artificial Intelligence
Engineer of XXI Century'2024

MONOGRAFIA

INFORMATYKA TECHNICZNA I SZTUCZNA INTELIGENCJA

2024

The paradigm of building secure information technologies for detecting deepfake modifications of biometric images based on neural networks

Halyna Mykytyn [1], Khrystyna Ruda [2*], Mariia Shved [3]

[1] DSc, Lviv Polytechnic National University, Ukraine, Professor of Information Protection Department halyna.v.mykytyn@lpnu.ua
[2] MSc, Lviv Polytechnic National University, Ukraine, Postgraduate of Information Protection Department, khrystyna.s.ruda@lpnu.ua
[3] PhD, Lviv Polytechnic National University, Ukraine, Senior Lecturer of Information Protection Department mariia.y.shved@lpnu.ua
** Corresponding author: khrystyna.s.ruda@lpnu.ua*

Abstract: The work proposes a paradigm of safe information technologies (IT) for detecting deepfake modifications of a biometric image based on convolutional neural networks, which is developed by: a concept based on a deepfake detection system and a decision support system; the methodology of creating neural network IT as a constructive algorithm for the functioning of information resources (IR) in the information system (IS) based on information processes (IP) and information networks (IN) and elements of security management (M); complex security system (CSS) of neural network IT at the level of "deepfake modification - detection". A multi-level complex security system is deployed with protective belts and security technologies per IT components and subsystems of information protection at the detection, blocking, and neutralization of probable random and targeted threats.

An algorithm for the detection of deepfake modifications of biometric images has been developed: splitting the video file of the biometric image into frames - recognition by the detector - reproduction of normalized images of faces - processing through a neural network - calculation of the feature matrix - construction of an image classifier. The created paradigm is transformed into various subject areas of intellectualization of infrastructure objects to ensure appropriate security profiles.

Keywords: Intellectualization; cyber security; biometric image; deepfake; information technology; neural networks; paradigm; detection system; algorithm; comprehensive security system;

Paradygmat budowania bezpiecznych technologii informacyjnych do wykrywania modyfikacji deepfake obrazów biometrycznych opartych na sieciach neuronowych

Halyna Mykytyn [1], Khrystyna Ruda [2*], Mariia Shved [3]

[1] DSc, Lviv Polytechnic National University, Ukraine, Professor of Information Protection Department halyna.v.mykytyn@lpnu.ua
[2] MSc, Lviv Polytechnic National University, Ukraine, Postgraduate of Information Protection Department, khrystyna.s.ruda@lpnu.ua
[3] PhD, Lviv Polytechnic National University, Ukraine, Senior Lecturer of Information Protection Department mariia.y.shved@lpnu.ua
** Corresponding author: khrystyna.s.ruda@lpnu.ua*

Streszczenie: Praca proponuje paradygmat bezpiecznych technologii informacyjnych (IT) do wykrywania modyfikacji deepfake obrazu biometrycznego oparty na konwolucyjnych sieciach neuronowych, który został opracowany poprzez: koncepcję opartą na systemie wykrywania deepfake oraz systemie wspomagania decyzji; metodologię tworzenia sieci neuronowych IT jako konstruktywnego algorytmu funkcjonowania zasobów informacyjnych (IR) w systemie informacyjnym (IS) na podstawie procesów informacyjnych (IP), sieci informacyjnych (IN) oraz elementów zarządzania bezpieczeństwem (M); kompleksowy system bezpieczeństwa (CSS) sieci neuronowych IT na poziomie „modyfikacja deepfake – wykrywanie". Wielopoziomowy kompleksowy

DOI: https://doi.org/10.53052/9788367652292.04

system bezpieczeństwa jest wdrażany z pasami ochronnymi i technologiami zabezpieczeń dla komponentów IT oraz podsystemów ochrony informacji na etapie wykrywania, blokowania i neutralizacji prawdopodobnych losowych i celowych zagrożeń.

Został opracowany algorytm wykrywania modyfikacji deepfake obrazów biometrycznych: podział pliku wideo obrazu biometrycznego na klatki – rozpoznawanie przez detektor – odtwarzanie znormalizowanych obrazów twarzy – przetwarzanie przez sieć neuronową – obliczanie macierzy cech – konstrukcja klasyfikatora obrazów. Stworzony paradygmat jest transformowany do różnych dziedzin inteligentnej infrastruktury, aby zapewnić odpowiednie profile bezpieczeństwa.

Słowa kluczowe: Intelektualizacja; Cyberbezpieczeństwo; Obraz biometryczny; Deepfake; Technologia informacyjna; Sieci neuronowe; Paradygmat; System wykrywania; Algorytm; Kompleksowy system bezpieczeństwa;

1. Introduction

The influence of technocracy on the development of modern civilization significantly determines the widespread application of various cutting-edge technologies to the essential processes of vital activity within the key subject areas of society. This influence is particularly evident in the integration and application of artificial intelligence (AI) and neural network technologies, especially in the processes of intellectualization. One of the most urgent and pressing problems in this field is the phenomenon of deepfake technology, which represents a mechanism controlled by artificial intelligence to replace a person's face in video files with the face of another individual through complex machine learning algorithms. This process modifies biometric images and poses a serious challenge, requiring advanced neural network-based technology to detect and counter the threat.

The foundation for addressing and solving this problem lies in the structure of the "intellectualization - cyber security" paradigm, which is based on several core principles: Industry 4.0, the Cyber Security Strategies of Ukraine, and the EU4Digital Programs: Cyber Security - East [1, 2, 3]. The confidentiality and security of a person's biometric image, particularly of their face, are determined by the successful implementation of secure recognition models using sophisticated information technologies based on neural networks. These models are designed to ensure that facial biometric data remains protected against unauthorized alterations and deepfake manipulations. Analysis of recent achievements and publications: With the growing concern and increasing debate about the misuse and abuse of deepfake technology, researchers in the field are continuously working to enhance and improve methods for detecting the manipulation of biometric images. For example, the work described in [4] proposes an innovative method for detecting deepfakes by analyzing convolutional traces that are generated when this type of image is created. This particular technique is built upon a detailed analysis of the relationship between each pixel and its neighboring pixels, aiming to find the relationships of adjacent pixels by calculating the maximum value of mathematical expectation. Once these relationships are identified, the technique uses machine learning algorithms to classify the images based on the detected patterns.

Another unique approach to deepfake detection, discussed in [5], is based on a model that includes three distinct convolutional neural network (CNN) streams, each receiving different types of input data. The first stream is provided with an input image and is responsible for learning general facial characteristics of the person, including features such as the shape of the head and the color of the hair. The second stream is given a blurred version of the image and is tasked with determining the skin tone or color of the individual. The third CNN stream works with a sharper, more detailed version of the image and is focused on analyzing local features of the face. These three CNN streams are then combined at the feature level, and a decision about the classification is made at the original classification stage. In [6], the research focuses on the dispersion indicators of consecutive video frames, which are used to train a classifier for biometric image recognition. The authors explain their decision to reject the use of convolutional neural networks, citing the presence of convolutional filters in the network structure, which they claim can cause the loss of important data and thus affect the dispersion indicators of biometric images.

The article in [7] proposes a novel method for detecting deepfake modifications by implementing face localization techniques and searching for artifacts that arise during the modification process. When a face is replaced in a video or image, the modified image typically consists of three distinct areas: the background, which remains unmodified, the replaced face, and a transition area that smooths the border between the two. The proposed system searches for inconsistencies in the areas where the face overlaps with the background, and based on the detection or non-detection of such inconsistencies, the system makes a conclusion regarding the authenticity of the biometric image.

DOI: https://doi.org/10.53052/9788367652292.04

Another noteworthy approach is presented in [8], where the authors utilize a neural network that is capable of isolating the noisy background in an image. They argue that the output file will display different noise patterns in the areas of the original image compared to those altered by deepfake algorithms. The neural network, trained on a large dataset of images captured by various cameras, examines the noise levels in both the face and background areas of the image. If a discrepancy in noise patterns is detected, the system concludes that the image has been modified.

The purpose of the work is to create a comprehensive paradigm for the construction of safe and secure information technologies aimed at the detection of deepfake modifications of biometric images through the use of advanced neural networks. The broader goal is the development of a fully integrated IT security system and, on this basis, the creation of an algorithmic implementation of a deepfake detection system. This system would operate within the context of addressing the challenges posed by the safe intellectualization of society's subject areas, ultimately contributing to the enhancement of cybersecurity and the protection of biometric data from manipulation.

2. The paradigm for building secure information technologies for detecting deepfake modifications of biometric images based on neural networks

As a result of the analysis of the existing technologies for detecting deepfake modifications, it is proposed to create a new paradigm of cyber security of technologies for detecting deepfake modifications of biometric images under the tasks of intellectualization and the tasks of cyber security of the subject spheres of society in the space of Industry 4.0, which includes the following tasks: 1) development of the concept of detecting deepfake modifications of biometric images based on neural networks; 2) construction of a methodology for creating information technology for detecting deepfake modifications based on neural networks ((IT-1) and information technology for decision support (IT-2); 3) development of a comprehensive security system of information neural network technology.

The vertical decomposition of the construction of secure IT for the detection of deepfake modifications based on neural networks in order to ensure the confidentiality of biometric images following the regulatory framework (Fig. 1.) provides: problematic oriented situation "intellectualization - information neural network technology for detecting deepfake modifications of biometric images - cyber security"; 2) the concept of detecting deepfake modifications of biometric images; 3) the methodology of creating an information neural network technology for detecting deepfake modifications; 4) a complex IT security system for detecting deepfake modifications based on a convolutional neural network.

The core of the concept of detecting deepfake modifications of biometric images is the constructive algorithms of IT1 and IT2 functioning. The system for detecting deepfake modifications based on neural networks, which has the structure "indication - interpretation - identification - decision-making", in particular, the functionality of the convolutional neural network "input data - convolution - subsampling" implements the IT1 algorithm "video division into frames - deepfake detection - processing features - classification of images".

The data analysis system for evaluating the biometric image classifier implements the constructive IT2 algorithm "identification - evaluation of the classifier - new model". It allows deciding whether the accuracy of the classifier is sufficient for detecting deepfake modifications of biometric images. The user can create a new biometric image classifier model in case of inconsistency according to the analysis system.

The basis of the methodology for creating IT for the detection of deepfake modifications of biometric images is its presentation as a constructive algorithm for the functioning of information resources in information systems (systems for detecting deepfake modifications using convolutional neural networks, data analysis systems) based on information processes (phases, operations, processing), information networks (computer, wireless), information security management based on the "plan - do - check - act" model (DSTU ISO 9001:2015). Multi-level CSS IT detection of deepfake in accordance with its components and probable threats are represented by - methodological, technical, software, communication, and management security subsystems and deployed by security belts with information protection technologies at the level of countermeasures - detection, blocking, and neutralization. The structure of the paradigm for the construction of safe information neural network technologies for the detection of deepfake modifications of biometric images enables its application in subject areas in accordance with problematic tasks and security profiles.

DOI: https://doi.org/10.53052/9788367652292.04

2.1. The algorithm of the detection system of deepfake modifications of biometric images

On the basis of the proposed paradigm for constructing safe IT for the detection of deepfake modifications of biometric images (Fig. 1), we will consider the algorithm of the system for step-by-step detection of deepfake modifications using convolutional neural networks.

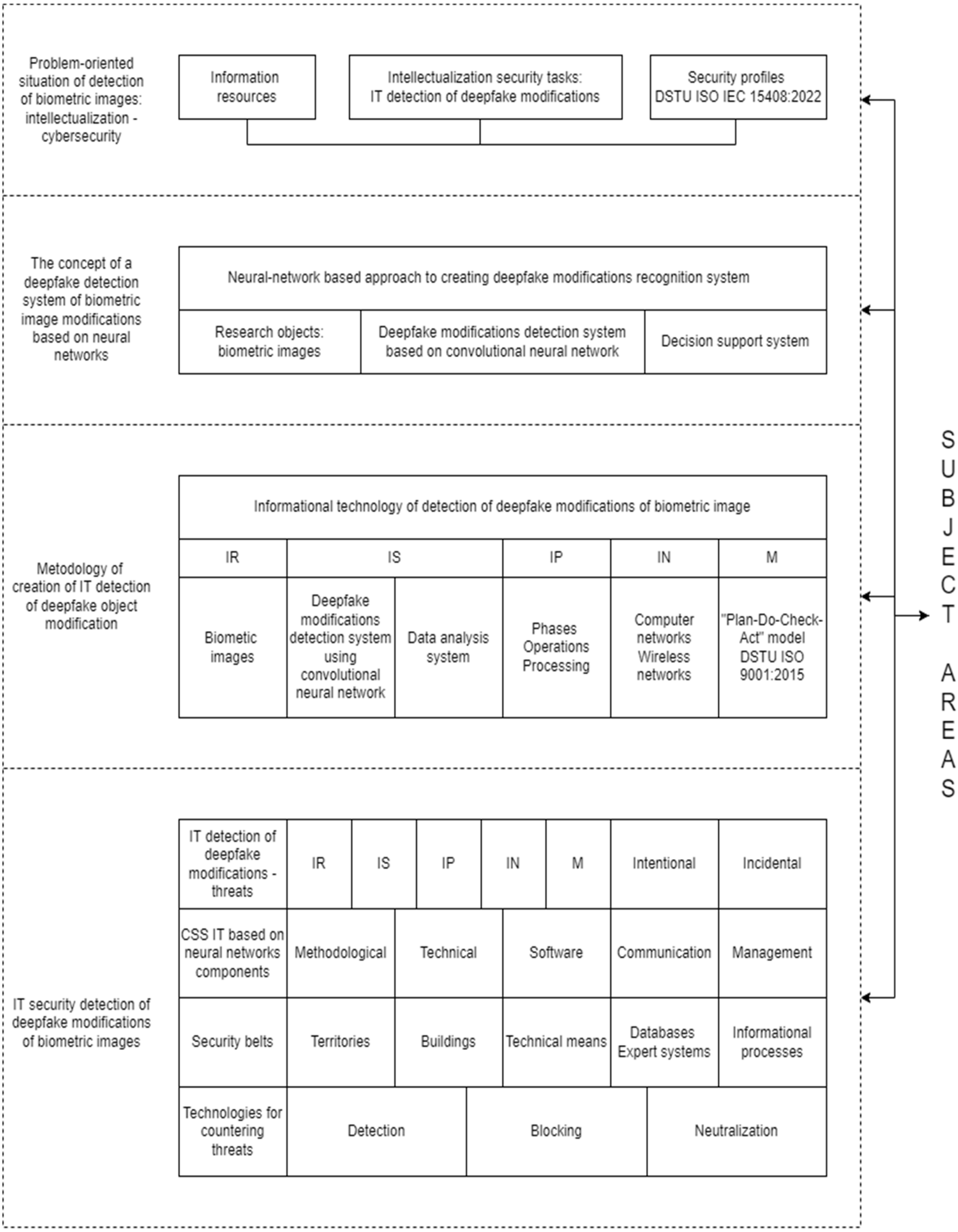
Problem-oriented situation of detection of biometric images: intellectualization - cybersecurity
Information resources
Intellectualization security tasks: IT detection of deepfake modifications
Security profiles DSTU ISO IEC 15408:2022
The concept of a deepfake detection system of biometric image modifications based on neural networks
Neural-network based approach to creating deepfake modifications recognition system
Research objects: biometric images
Deepfake modifications detection system based on convolutional neural network
Decision support system
Metodology of creation of IT detection of deepfake object modification
Informational technology of detection of deepfake modifications of biometric image
IR
IS
IP
IN
M
Biometic images
Deepfake modifications detection system using convolutional neural network
Data analysis system
Phases Operations Processing
Computer networks Wireless networks
"Plan-Do-Check-Act" model DSTU ISO 9001:2015
IT security detection of deepfake modifications of biometric images
IT detection of deepfake modifications - threats
IR
IS
IP
IN
M
Intentional
Incidental
CSS IT based on neural networks components
Methodological
Technical
Software
Communication
Management
Security belts
Territories
Buildings
Technical means
Databases Expert systems
Informational processes
Technologies for countering threats
Detection
Blocking
Neutralization
SUBJECT AREAS

Figure 1. The paradigm of building secure information technologies for detecting deepfake modifications of biometric images based on neural networks.

DOI: https://doi.org/10.53052/9788367652292.04

This approach is aimed at ensuring the information confidentiality profile (DSTU ISO/IEC 15408-1:2017) (Fig. 2).

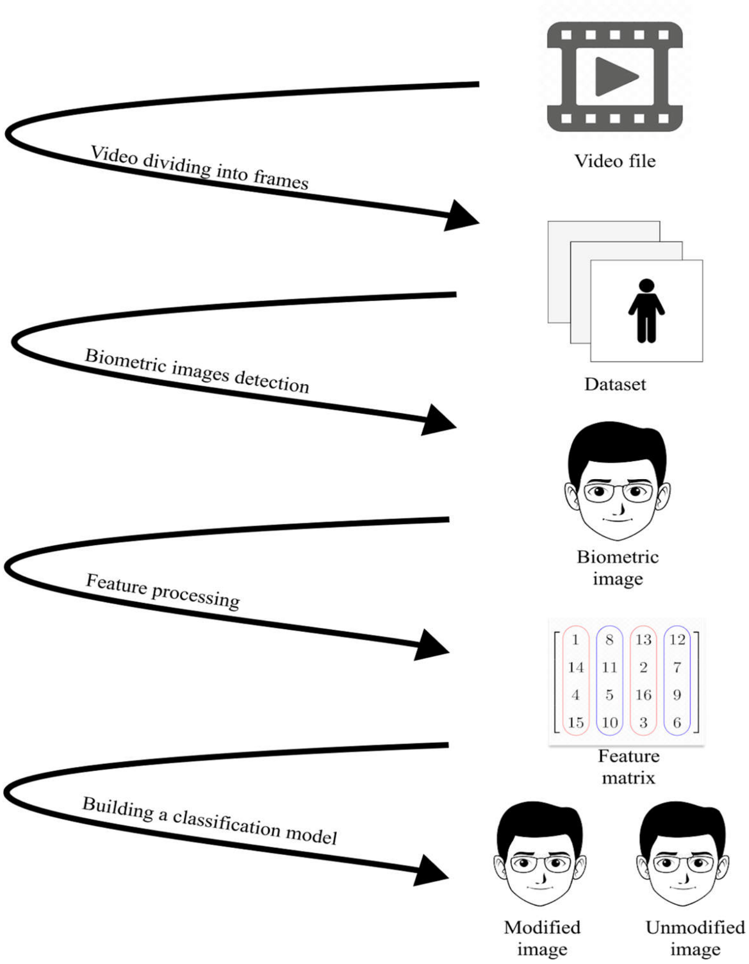

Figure 2. The algorithm of the detection system of deepfake modifications of biometric images.

At the first stage of functioning of the detection of deepfake modifications, the algorithmic structure of recognition is displayed: dividing the video file into separate frames - recognition of biometric images by the detector - creation of new standardized images.

The second stage of detection of deepfake modifications is characterized by the algorithmic structure of processing: reproduction of normalized biometric images of faces - processing of images with a neural network - calculation of feature matrices - construction of an image classifier

The third stage of detecting deepfake modifications is characterized by the algorithmic structure of classifier accuracy assessment: sensitivity and specificity of the classifier – Youden index – optimal threshold value of image classification – calculation of classifier accuracy.

In the classification process, the probability of the biometric image belonging to each class is calculated independently. The Youden index is used to determine the optimal threshold value for images[9]:

$$J = max(TPR(t) + TNR(t) - 1), \tag{1}$$

where J is the Youden index, t is the threshold value for correct image selection, TPR is the sensitivity of the classifier, which is defined as

$$TPR = \frac{TP}{TP+FN}, \tag{2}$$

where TP is the number of correctly classified modified images, FN is the number of incorrectly classified unmodified (original) images, TNR is the specificity of the classifier, defined as

$$TNR = \frac{TN}{TN+FP}, \tag{3}$$

where *TN* is the number of correctly classified unmodified images, *FP* is the number of incorrectly classified modified images.

The optimal value of the Youden index provides balanced values of sensitivity (TPR) and specificity (TNR) of the image classifier at the same level, which determines the threshold for cutting off informedly classified images from uninformedly classified biometric images in the belonging space "high probability - low probability" (Fig. 3) .

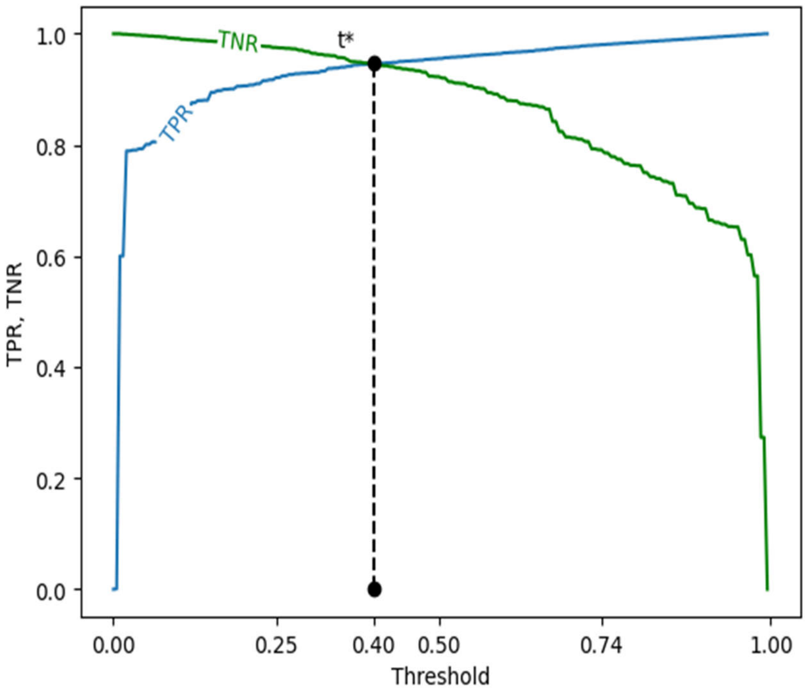

Figure 3. The experimental determination of the cut-off value of the informed classified images from uninformed classified

The accuracy of the classifier is calculated based on informed classified biometric images:

$$Acc = \frac{TP_i + TN_i}{TP_i + FP_i + TN_i + FN_i}, \tag{4}$$

where *Acc* is the classification accuracy, TP_i is the number of informed correctly classified modified images, TN_i is the number of informed correctly classified unmodified images, FP_i is the number of informed wrongly classified modified images, FN_i is the number of informed wrongly classified unmodified images.

2.2. Comprehensive IT security system for detection of deepfake modifications

The development of methodological principles for the creation of cyber security systems for the functioning of critical infrastructure objects [10, 11] is complemented by a multi-level model of a comprehensive security system (CSS) of informational neural network technology for detecting deepfake modifications of biometric images, which is presented in Fig. 4. The first level of the CSS model represents the objects of protection as components of IT - IR, IS, IP, IN, M. The second level is accidental and targeted threats to the corresponding protection objects. The third level is a complex IT security system based on neural networks, which is represented by: methodological – I; technical (hardware, physical) – II; software – III; communication - IV; management – V subsystems. Each CSS IT subsystem has regulatory support. The fourth level is a complex of protective zones that reflect the nature of information security: the outer zone, which covers the entire territory where buildings containing IT based on neural networks are located - α; the belt of buildings (premises), or IT devices - β; the belt of system components, technical means, software γ, the belt of elements of databases, expert systems – θ, the belt of information processing processes (life cycle) – λ. The fifth level is a set of methods and means of countering potential threats arising in the protective belts α, β, γ, θ. λ at the level: detection – a; blocking – b; neutralization – c.

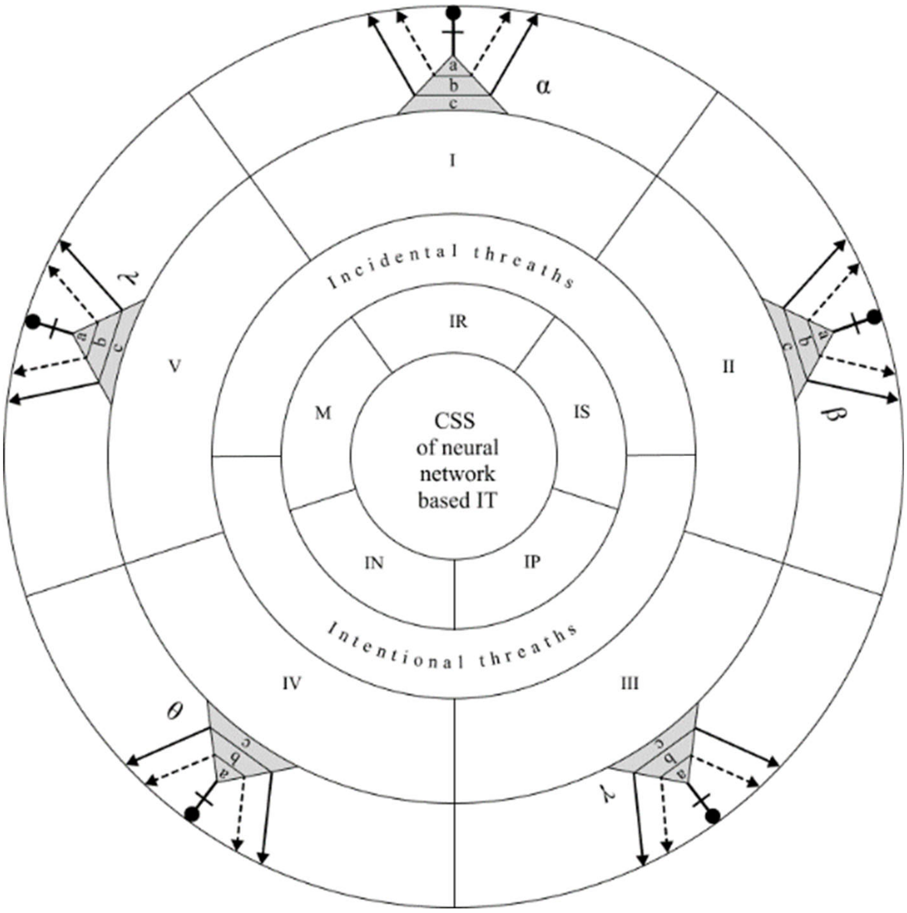

Figure 4. A multi-level model of a comprehensive IT security system for detecting deepfake modifications based on neural networks

Regulatory support of CSS IT is based on standards and regulatory documents in the field of cyber security. The standards include state standards of Ukraine's DSTU; state standards of Ukraine, unified with international DSTU ISO/IEC; international standards ISO (International Organization for Standardization), IEC (International Electrotechnical Commission), ITU (International Telecommunication Union) and others; national standards of standardization organizations of individual countries BSI, DIN, ANSI, NIST and others; regional standards of organizations that represent the interests of large regions or continents in the global standardization process, CEN (European Committee for Standardization), CENELEC (European Committee for Standardization in Electrical Engineering) and others; standards of industrial consortia and professional organizations ICC, API and others. In the context of the development of approaches to ensure IT security and the detection of deepfake modifications of biometric images, specification 1.0 of the C2PA standard [12] defines safe, tamper-proof, cross-platform standardized methods for determining the authenticity of file content based on the structure of "source and time-space coordinates of file content creation - editing technologies file in time during the life cycle", which accordingly informs the user about the change of digital content.

Conclusions

The paper presents the quintessence of solving the problem of detecting deepfake modifications of biometric images based on: 1) the paradigm of building safe information neural network technologies for detecting deepfake modifications; 3) a comprehensive security system of IT; 2) the algorithm for the detection of deepfake modifications of biometric images of human faces; which is the basis for the development of approaches, algorithms for detecting deepfake modifications based on informational neural network technologies and the creation of their complex security system based on security profiles in various subject areas.

Reference

1. Colombo, Armando W.; Karnouskos, Stamatis; Bangemann, Thomas. Towards the Next Generation of Industrial Cyber-Physical Systems. *Industrial Cloud-Based Cyber-Physical Systems* 2014, 1, 1–22. https://doi.org/10.1007/978-3-319-05624-1_1.
2. Strategiya kiberbezpeki Ukrainy (2021–2025 roky). Available online: https://www.rnbo.gov.ua/files/2021/STRATEGIYA%20KYBERBEZPEKI/proekt%20strategii_kyberbezpeki_Ukr.p deepfake (accessed on 5 October 2024).
3. Програма EU4Digital: Кібербезпека – Схід. Available online: https://eufordigital.eu/uk/discover-eu/eu4digital-improving-cyber-resilience-in-the-eastern-partnership-countries/ (accessed on 5 October 2024).
4. Guarnera, O.; Giudice, O.; Battiato, S. Deepfake Detection by Analyzing Convolutional Traces. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2020, 666–667. https://doi.org/10.1109/CVPRW50498.2020.00100.
5. Yavuzkilic, S.; Sengur, A.; Akhtar, Z.; Siddique, K. Spotting Deepfakes and Face Manipulations by Fusing Features from Multi-Stream CNNs Models. *Symmetry* 2021, 13, 1352. https://doi.org/10.3390/sym13081352.
6. Lee, G.; Kim, M. Deepfake Detection Using the Rate of Change between Frames Based on Computer Vision. *Sensors* 2021, 21, 7367. https://doi.org/10.3390/s21217367.
7. Li, L.; Bao, J.; Zhang, T.; Yang, H.; Chen, D.; Wen, F.; Guo, B. Face X-ray for More General Face Forgery Detection. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, 5001–5010. https://doi.org/10.1109/CVPR42600.2020.00506.
8. Zhang, K.; et al. Beyond a Gaussian Denoiser: Residual Learning of Deep CNN for Image Denoising. *IEEE Transactions on Image Processing* 2017, 26, 3142–3155. https://doi.org/10.1109/TIP.2017.2699184.
9. Youden, W.J. Index for Rating Diagnostic Tests. *Cancer* 1950, 3, 32–35. https://doi.org/10.1002/1097-0142(1950)3:1<32::AID-CNCR2820030106>3.0.CO;2-3.
10. Yevseiev, S.; Ponomarenko, V.; Laptiev, O.; Milov, O.; Korol, O.; Milevskyi, S. Synergy of Building Cybersecurity Systems. *Kharkiv: PC TECHNOLOGY CENTER*, 2021, 188.
11. Bobalo, Y.; Dudykevych, V.; Mykytyn, G.; Stosyk, T. Paradigm of Safe Intelligent Ecological Monitoring of Environmental Parameters. *Proceedings of the 3rd International Conference on Information Security and Information Technologies (ISecIT 2021) Co-located with 1st International Forum "Digital Reality" (DRForum 2021)*, September 13–19, Odesa, Ukraine, 2021, 244–249.
12. Coalition for Content Provenance and Authenticity. Available online: https://c2pa.org/ (accessed on 5 October 2024).