Detecting Deepfake Modifications of Biometric Images using Neural Networks

Valeriy Dudykevych[1], Serhii Yevseiev[2], Halyna Mykytyn[1], Khrystyna Ruda[1], and Hennadii Hulak[3]

[1] *Lviv Polytechnic National University, Lviv, 79013, Ukraine*
[2] *National Technical University "Kharkiv Polytechnic Institute," Kharkiv, 61000, Ukraine*
[3] *Borys Grinchenko Kyiv Metropolitan University, 18/2 Bulvarno-Kudriavska str., Kyiv, 04053, Ukraine*

Abstract

The National Cybersecurity Cluster of Ukraine is functionally oriented towards building systems for the protection of various platforms within the information infrastructure, including the development of secure technologies for detecting deepfake modifications of biometric images based on neural networks in cyberspace. The paper introduces an instrumental platform for detecting deepfake modifications of biometric images and an analytical security structure of neural network Information Technologies (IT) based on a multi-level model of "resources—systems—processes—networks—management" according to the "object—threat—protection" concept. The instrumental platform integrates information neural network technology and decision support information technology, employing a modular architecture of the neural network detection system for deepfake modifications in the "preprocessing data—feature processing—classifier training" space. The core of the IT security structure is the integrity of the functioning of the neural network system for detecting deepfake modifications of human facial biometric images and data analysis systems that implement the information process of "splitting a video file into frames—detection, feature processing—classifier accuracy assessment". The security of the multi-level model of neural network IT is based on systemic and synergistic approaches, enabling the construction of a comprehensive IT security system, considering the emergent property in the presence of potential targeted threats and the application of advanced technologies at the hardware and software levels. The proposed comprehensive security system for the information process of detecting deepfake modifications of biometric images covers hardware and software means by segments: automated classifier accuracy assessment; real-time detection of deepfake modifications; sequential image processing; accuracy evaluation of classification using cloud computing.

Keywords

Intellectualization, cybersecurity, biometric image, deepfake, information technology, neural networks, detection system, instrumental platform, analytical security structure, comprehensive security system.

1. Introduction

The problem statement. The security of critical state infrastructure objects in both physical and cyberspace is currently a pressing issue within the realm of intellectualization across various societal domains. In the context of Industry 4.0 tasks, the Cybersecurity Strategy of Ukraine, and the National Cybersecurity Cluster, one of the paramount tools for addressing the challenge of safely intellectualizing critical infrastructure objects is the utilization of neural network information

CPITS-2024: Cybersecurity Providing in Information and Telecommunication Systems, February 28, 2024, Kyiv, Ukraine
EMAIL: vdudykev@gmail.com (V. Dudykevych); serhii.yevseiev@gmail.com (S. Yevseiev); cosmos-zirka@ukr.net (H. Mykytyn); khrystyna.s.ruda@lpnu.ua (K. Ruda); h.hulak@kubg.edu.ua (H. Hulak)
ORCID: 0000-0001-8827-9920 (V. Dudykevych); 0000-0003-1647-6444 (S. Yevseiev); 0000-0003-4275-8285 (H. Mykytyn); 0000-0001-8644-411X (K. Ruda); 0000-0001-9131-9233 (H. Hulak)

technologies for detecting deepfake modifications in the biometric images of individuals' faces [1–3]. The accuracy criterion for classifying biometric images through neural networks hinges on the safety of detecting deepfake modifications, a determination guided by the comprehensive security system of a multi-level information technology framework [4, 5].

Analysis of recent achievements and publications. The ongoing development of methodological principles for establishing cybersecurity systems in information technologies that support the functioning of critical infrastructure objects remains pertinent [6, 7]. Currently, security processes are being implemented in tasks related to the detection of deepfake modifications in biometric facial images using neural networks. Investigations into security issues within the realm of machine learning, particularly dealing with complex threat models and corresponding protective measures, are actively underway [8, 9]. The study [10] delves into the efficiency assessment of contemporary algorithms designed to detect fake content, shedding light on their performance within the context of information warfare scenarios. This comparative analysis contributes valuable insights into the ongoing efforts to bolster defenses against deceptive information dissemination. In [11], the security model and data privacy in deep learning, as part of machine learning, are examined under the influence of relevant attacks. This includes poisoning attacks and evasion attacks, both of which impact decision-making processes in deep learning. Countermeasures against such attacks involve the recognition and removal of malicious data, training models to be insensitive to such data, and concealing the model's structure and parameters. The confidentiality of data during deep learning is also jeopardized by specific attacks, such as the inversion of the security model. Effective tools to counter privacy threats include cryptographic methods, notably homomorphic encryption [12, 13].

Furthermore, the study of hardware security for deep neural networks within the "threat—protection" space is discussed in [14]. Modern methods ensuring the detection of deepfake modifications in biometric facial images with an accuracy ranging from 0.94 to 0.99 are known [15].

The aim of the study. The primary objective of this study is to formulate an analytical security structure for information technology designed to detect deepfake modifications in biometric images. This structure aligns with the instrumental platform and a multi-level model of neural network IT, encompassing Information Resources (IR), Information Systems (IS), Information Processes (IP), Information Networks (IN), and Information Security Management (ISM). The constructed algorithm within this structure is aimed at facilitating the secure operation of neural network IT.

2. Instrumental Platform for Detecting Deepfake Modifications of Biometric Images

The creation of an analytical security structure for detecting deepfake modifications of biometric images is based on the following prerequisites: an instrumental platform (Fig. 1)—information neural network technology (IT1); decision support information technology (IT2). The development of information technologies for detecting deepfake modifications of biometric images relies on: the use of a staged approach for detecting modified biometric images using convolutional neural networks [16]; the application of a neural network system for detecting deepfake modifications based on its architecture and decision support systems for assessing the classifier's performance according to the evaluation methodology [17]. The information neural network technology is based on the following components: the object model, methodology for detecting deepfake modifications, accuracy of biometric image classification, and an evaluation methodology for assessing the classifier's performance. The constructive algorithm of IT1: "video segmentation—detection—feature processing —classification" is implemented through the architecture of the neural network system using a modular approach, incorporating individual functional modules to enhance the efficiency and adaptability of the deepfake modification detection algorithm as shown on Fig. 2.

The modular architecture of the neural network system for detecting deepfake modifications implements an interconnected algorithm comprising "preprocessing data—feature processing—classifier training" flow. This algorithm is functionally deployed with a convolutional neural network in the space of "input data—convolution—subsampling" and ensures "indication—interpretation—identification—decision-making" [18].

The data preprocessing module of the deepfake modification detection system functionally executes an algorithm that involves:

1. Splitting the video file into individual frames utilizing Python libraries.
2. Face detection using neural network-based tools.
3. Processing detected biometric images (cropping, adjusting height and width, reformatting) to create new standardized samples.

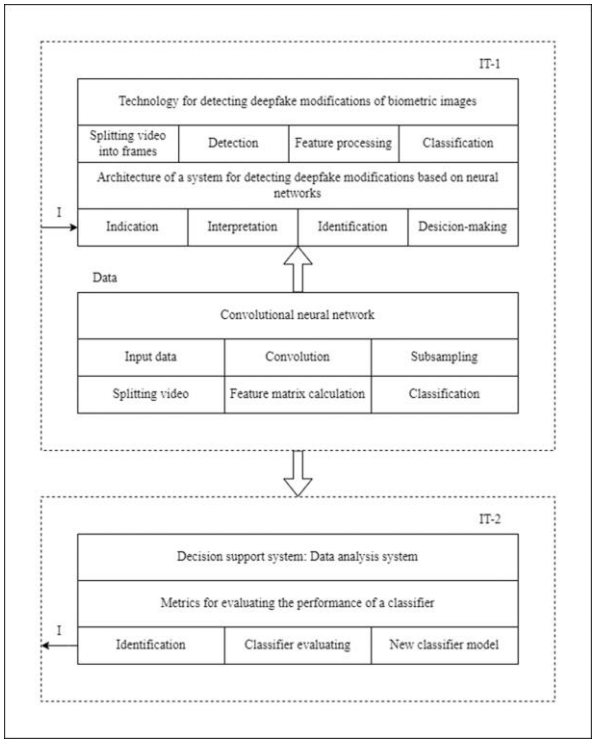

Figure 1: Instrumental platform for detecting deepfake modifications of biometric images

The feature processing module of the deepfake modification detection system is characterized by an algorithmic structure that includes:

1. The utilization of normalized facial biometric images.
2. The extraction of feature matrices using neural network tools.
3. The saving of these features in formatted arrays to be processed as input data for classifier training.

The classifier training module of the deepfake modification detection system implements a functional algorithm that includes:

1. Classifier training.
2. Evaluation of the classifier based on selected metrics.
3. Decision on classifier admission—modified image; unmodified image.

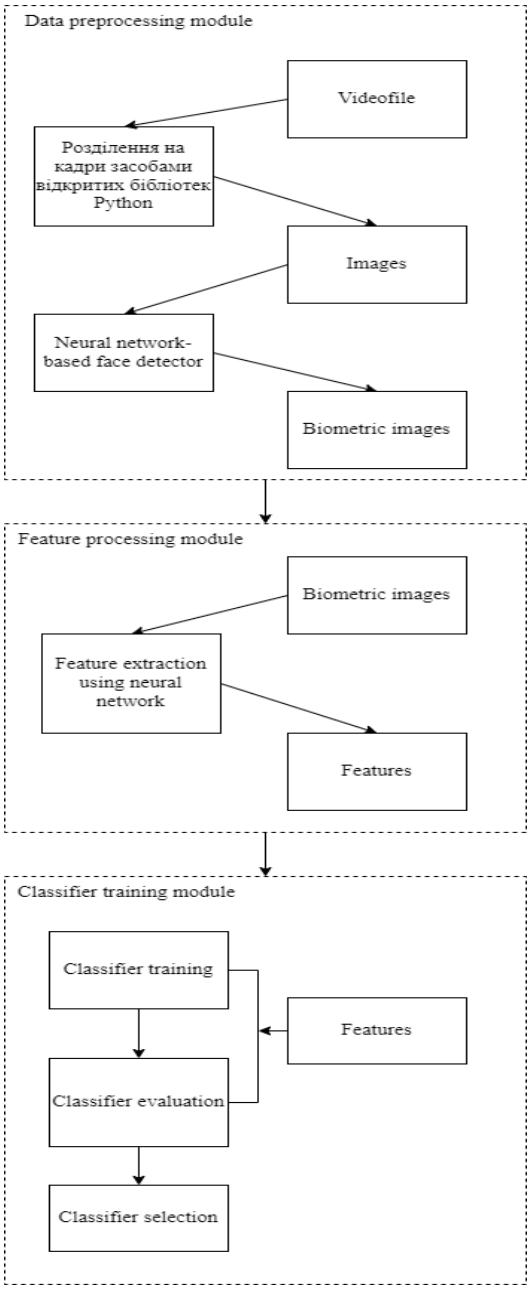

Figure 2: Architecture of a system for detecting deepfake modifications based on neural networks

The evaluation of the classifier in the system for detecting deepfake modifications of biometric images takes into account:

1. Sensitivity and specificity of the classifier.
2. Youden's index, determining the optimal threshold value for the classification of biometric images.
3. Informatively classified biometric images.

The constructive algorithm of IT2, involving "identification—classifier evaluation—new classifier model," is implemented by the decision support system in the data analysis space, considering evaluation metrics such as:

1. Classifier accuracy.
2. The area under the curve.
3. Logarithmic loss function, which positions the difference between the predicted probability of an element belonging to a certain class and the actual probability of belonging from the classifier [19].

3. Security Structure for Detecting Deepfake Modifications based on a Multi-Level Model of Neural Network IT

After analyzing existing approaches to secure detection of deepfake modifications in biometric images, the following proposals are made:

1. the creation of an analytical security structure for neural network information technologies designed for the detection of deepfake modifications in human facial biometric images within the space of secure object intelligence for critical infrastructure [20].
2. the development of a comprehensive security system for the information process "phase—operation—processing" based on levels such as "splitting video files into frames—detection, feature processing—evaluation of image classifier accuracy".

The analytical security structure of neural network IT for detecting deepfake modifications, aiming to ensure the confidentiality and integrity of human facial biometric images (Fig. 3), incorporates a systemic and synergistic approach. The systemic approach adheres to principles of hierarchy, structuring, and integrity, providing grounds for the creation of a comprehensive IT security system within the space of optimal integration of methodological, technical (hardware), software, and normative support for secure functioning throughout the information life cycle in the system, and the algorithm of the information process at the "phase—operation—processing" level. The synergistic approach, exhibiting the emergent property, presents one facet of the integrity of information protection in IT, assuming the presence of properties specific to a comprehensive IT security system as a whole but not specific to its elements—complex security systems of information resources, systems, processes, networks, and management.

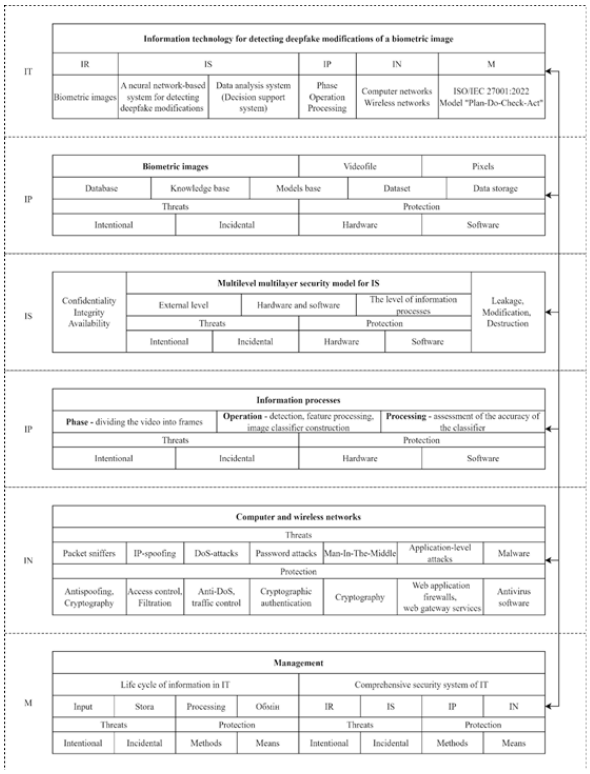

Figure 3: Analytical structure of the security of neural network-based informational technology

The core of the analytical structure of secure neural network information technology is the system for detecting deepfake modifications in biometric images based on neural networks and the data analysis system, programmatically oriented towards the comprehensive implementation of the information process "splitting the video into frames—deepfake

detection—feature processing—evaluation of image classification". On this basis, decisions are made regarding the sufficient accuracy of the deepfake modification classifier according to the chosen model, with the possibility of updating it. Table 1 presents a comprehensive security system for the information process of detecting deepfake modifications at the processing level of biometric images according to the "object—threat—protection" concept.

Table 1
The comprehensive security system of the deepfake detection process of biometric image modification at the processing level

Object: informational process	Threats		Protection	
Processing	**Intentional**	**Incidental**	**Hardware**	**Software**
The automated classifier accuracy assessment	Leakage and/or violation of confidentiality, integrity of data and models Unauthorized access Malicious software Distributed Denial of Service Attacks (DDoS)	Failures and/or instability of technical devices Operator errors Unpatched software vulnerabilities	Luna SA HSM Luna SP Luna XML	Encrypt Easy Suricata Webroot DNS Protection 1Password BitLocker Bitdefender Antivirus
The deepfake detection in real-time	Data manipulation Model inversion Data poisoning Adversarial examples Denial of service	Technical malfunctions of the network and components	nShield Connect HSM Gryada-301 Baryer-301 Canal-301	ManageEngine Log360 BitLocker
The sequential image processing	Data poisoning Adversarial examples Model manipulation Leakage and/or violation of confidentiality, integrity of data and models Cracking of cryptographic protection algorithms	Network failures Physical damage to equipment Poor data management practices	Luna SA4 HSM Luna PCM	Cisco UVPN-ZAS BitLocker
The classification accuracy assessment using cloud computing	Leakage and/or violation of confidentiality, integrity of data and models Malicious software Distributed denial of service attacks (DDoS) Phishing and social engineering	Data corruption Network failures Unpatched software vulnerabilities DoS on the side of the service provider Operator errors	Cisco Firepower Palo Alto Networks PA-7000 Series	Webroot DNS Protection AlienVault USM

Regulatory support for the analytical structure of neural network IT security is grounded in several international standards in the field of cybersecurity, including ISO/IEC 27034:2017, IEC 61508-3:2010, and ISO/IEC 13335-1:2004. The C2PA Specification 1.0, a pioneering functional standard by the Content Provenance and Authenticity Coalition, establishes scenarios, workflows, and requirements for validating and ensuring the digital provenance of content. These methods validate information about the creation and modification of media files, empowering content editors to create tamper-proof media by documenting who created or modified digital content, the specifics of modifications made, implementing robust security measures, and fostering transparency in the content creation process. [17].

4. Conclusions

In the paper, we introduce a security methodology for IT detection of deepfake modifications in biometric images using neural networks. The methodology is based on:

1. an instrumental platform.
2. an analytical security structure of neural network information technologies according to a multi-level model.
3. a comprehensive security system for the information process of detecting deepfake modifications at the processing level, following the concept of "object—threat—protection".

This serves as the foundation for the development of systematic approaches to secure deepfake detection within the security profiles of critical infrastructure.

References

[1] H. Kagermann, W. Wahlster, J. Helbig, Securing the Future of German Manufacturing Industry: Recommendations for Implementing the Strategic Initiative Industrie 4.0. Final Report of the Industrie 4.0 Working Group, Acatech, National Academy of Science and Engineering (2013).

[2] National Security and Defense Council of Ukraine. URL: https://www.rnbo.gov.ua/files/2021/STRATEGIYA%20KYBERBEZPEKI/proekt%20strategii_kyberbezpeki_Ukr.pdf

[3] The national cybersecurity cluster. URL: https://cybersecuritycluster.org.ua/

[4] B. Bebeshko, et al., Application of Game Theory, Fuzzy Logic and Neural Networks for Assessing Risks and Forecasting Rates of Digital Currency, J. Theor. Appl. Inf. Technol. 100(24) (2022) 7390–7404.

[5] K. Khorolska, et al., Application of a Convolutional Neural Network with a Module of Elementary Graphic Primitive Classifiers in the Problems of Recognition of Drawing Documentation and Transformation of 2D to 3D Models, J. Theor. Appl. Inf. Technol. 100(24) (2022) 7426–7437.

[6] S. Yevseiev, et al., Synergy of Building Cybersecurity Systems. PC Technology Center (2021). doi: 10.15587/978-617-7319-31-2.

[7] Y. Bobalo, V. Dudykevych, H. Mykytin, Strategic Security of the "Object—Information Technology" System, Publishing House of Lviv Polytechnic National University (2020).

[8] M. Choraś, et al., Machine Learning—The Results Are Not the only Thing that Matters! What About Security, Explainability and Fairness?, Computational Science—ICCS 2020, LNTCS 12140 (2020) 615–628. doi: 10.1007/978-3-030-50423-6_46.

[9] N. Papernot, et al., SoK: Security and Privacy in Machine Learning, IEEE European Symposium on Security and Privacy (EuroS&P) (2018). 399–414. doi: 10.1109/EuroSP.2018.00035.

[10] Y. Shtefaniuk, I. Opirskyy, Comparative Analysis of the Efficiency of Modern Fake Detection Algorithms in Scope of Information Warfare, 11th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (2021) 207–211, doi: 10.1109/IDAACS53288.2021.9660924.1.

[11] H. Bae, et al., Security and Privacy Issues in Deep Learning, ArXiv (2018). doi: 10.48550/arXiv.1807.11655.

[12] V. Grechaninov, et al., Decentralized Access Demarcation System Construction in Situational Center Network, in: Workshop on Cybersecurity Providing in Information and Telecommunication Systems II, vol. 3188, no. 2 (2022) 197–206.

[13] V Grechaninov, et al., Formation of Dependability and Cyber Protection Model in Information Systems of Situational Center, in: Workshop on Emerging Technology Trends on the Smart Industry and the Internet of Things, vol. 3149 (2022) 107–117.

[14] Q. Xu, M. Tanvir Arafin, G. Qu, Security of Neural Networks from Hardware Perspective: A Survey and Beyond, 26th Asia and South Pacific Design Automation Conference (ASP-DAC), (2021) 449–454. doi: 10.1145/3394885.3431639.

[15] X. Cao, N. Gong, Understanding the Security of Deepfake Detection, Digital Forensics and Cyber Crime, LNICST 441 (2022) 360–378. doi: 10.1007/978-3-031-06365-7_22.

[16] V. Dudykevych, H. Mykytyn, K. Ruda. Application of Deep Learning for Detecting Deepfake Modifications in Biometric Images, Mod. Spec. Technol. 1 (2022) 13–22.

[17] L. Wieclaw, et al., Biometrie identification from Raw ECG Signal Using Deep Learning Techniques, 9th IEEE International Conference on Intelligent Data Acquisition and Advanced

Computing Systems: Technology and Applications (IDAACS) (2017) 129-133. doi: 10.1109/IDAACS.2017.8095063.

[18] V. Dudykevych, H. Mykytyn, K. Ruda, The Concept of a Deepfake Detection System of Biometric Image Modifications based on Neural Networks, IEEE 3rd KhPI Week on Advanced Technology (2022). doi: 10.1109/khpiweek57572.2022.9916378.

[19] E. Altuncu, V. Franqueira, S. Li, Deepfake: Definitions, Performance Metrics and Standards, Datasets and Benchmarks, and a Meta-Review, ArXiv (2022). doi: 10.48550/arXiv.2208.10913

[20] X. Wang, T. Ahonen, J. Nurmi, Applying CDMA technique to network-on-chip, IEEE Transactions on Very Large Scale Integration (VLSI) Systems 15(10) (2007) 1091–1100. doi: 10.1109/tvlsi.2007.903914.

[21] H. Hulak, et al., Dynamic Model of Guarantee Capacity and Cyber Security Management in the Critical Automated Systems, in: 2nd International Conference on Conflict Management in Global Information Networks, vol. 3530 (2022) 102–111.