CYBER-PHYSICAL SECURITY THROUGH FACIAL RECOGNITION AND SENSOR DATA ANALYSIS

Ivaylo Atanasov, Dimitar Pilev

University of Chemical Technology and Metallurgy
8 Kliment Ohridski Blvd., Sofia 1797, Bulgaria
E-mail: ivaylo@uctm.edu

Received 11 July 2023
Accepted 19 January 2024

DOI: 10.59957/jctm.v59.i2.2024.27

ABSTRACT

The digital age has brought tremendous opportunities for innovation and efficiency. However, it has also exposed businesses, governments, and individuals to a range of cyber threats, such as data breaches, network attacks, ransomware, malicious insiders, and identity theft. This requires the implementation of robust cybersecurity measures to safeguard sensitive information and ensure the uninterrupted operation of all critical IT systems. This paper aims to provide a facial recognition security system for cyber-physical security that incorporates a neural network and intelligent algorithms to assess the severity level of security breaches. The system also includes alarms with severity levels ranging from 1 (low severity) to 4 (critical), based on facial recognition and data from carbon dioxide and temperature sensors. In the event of a security breach, an incident response plan is presented. The proposed system is applicable to offices, workspaces, server rooms, data centers and other areas where information is stored, to enhance physical security and protect against cybersecurity threats.

Keywords: cybersecurity, facial recognition, convolutional neural network, artificial intelligence.

INTRODUCTION

Cybersecurity addresses the protection of digital assets, systems, and networks from unauthorized access, malicious activities, and potential threats [1, 2]. With the fast evolution of technology and the increasing reliance on digital infrastructure, cybersecurity has become a critical concern for individuals, organizations, and governments worldwide [3].

Cybersecurity defense mechanisms encompass a range of technical and non-technical measures aimed at preventing, detecting, and mitigating cyber threats. Technical solutions include firewalls, intrusion detection (IDS) and intrusion prevention systems (IPS), strong encryption, authentication protocols and other measures. Non-technical measures involve access control policies, security awareness training and incident response planning [3, 4].

Cybersecurity faces several major challenges, including the proliferation of sophisticated attack techniques, shortage of skilled professionals, complexity of interconnected systems, and the rapid expansion of the Internet of Things (IoT) [5]. Additionally, emerging trends such as artificial intelligence (AI)-powered attacks [6], quantum computing, and the integration of cybersecurity in critical infrastructure pose new challenges that demand proactive and adaptive defense strategies. Effective cybersecurity requires a proactive and multi-layered approach.

The protection of physical infrastructure from cyber threats is a crucial aspect of cyber-physical security [7]. This includes securing data centers, workplaces and other critical physical assets from cyber-attacks that aim to disrupt or damage them, by limiting access to spaces where data is stored.

The use of facial recognition technology in cybersecurity has gained significant attention in recent years due to its potential to improve the security of sensitive information and protect against unauthorized access [8, 9]. Social engineering poses a significant cybersecurity threat that can impact both digital and physical security and can result in devastating breaches of sensitive information [10]. Hackers can use social engineering tactics to exploit human vulnerabilities and gain unauthorized access to sensitive areas, steal confidential information, and cause physical damage to assets.

Another possible cybersecurity risk is a breach of confidentiality by company employees. Facial recognition can determine whether a particular employee is authorized to reside in an area where information is stored.

Physical security of sensitive information

We propose using a complex system for cyber-physical security containing video surveillance with an integrated convolutional neural network (CNN) for facial recognition, temperature and carbon dioxide sensors, MongoDB database and pre-trained neural network model including training, data representation and decision-making processes [11].

A possible breach of confidentiality situation is when an employee breaks into a room with sensitive information to which they do not have access by default. (e.g. the finance department). Facial recognition will immediately detect a security breach and trigger the appropriate action.

In the case of social engineering, facial recognition can recognize employees from unrecognized persons and raise the corresponding alarm, but also reading information from sensors can give additional information when the camera has some limitations in different situations. A sudden rise or fall in temperature and carbon dioxide can indicate a potential problem. Changes in these indicators may indicate a fire, an open window or something else. Combining more parameters as an input layer gives more accuracy to the model. All information is recorded and stored in a MongoDB database [12].

If facial recognition fails to detect a human presence and classify them as known or unknown, the change in carbon dioxide levels within a specific area can serve as an indicator for making such an inference, that there are more people than the system detects.

The ability of the proposed security system to detect previously unknown attacks is of great importance. This could be solved with behavior analysis methods and unsupervised or semi-supervised machine learning techniques. The efficiency of the attack detection system strongly depends on the datasets used to train the machine learning models [12].

In Fig. 1, it is shown that there are four recognized faces, and one person is identified as unrecognized. Nevertheless, there is also an individual not detected by the facial recognition component of the system. In such cases, the deviation in carbon dioxide levels within the specific room will indicate a departure from the expected value based on the number of people present.

Cyber-physical security system architecture

As discussed in the previous point, to achieve better efficiency, cyber-physical security systems must consider parameters of a different nature [13]. Fig. 2 shows the architecture of a cyber-physical security system, reporting indicators from video surveillance, temperature, and CO_2 sensors. The system uses an artificial neural network and machine learning models, and based on the information received in real time, determines the level (degree) of threat for a monitored room - low, medium, high, critical.

The proposed system is built from three main modules: Face recognition, Category converter and Predictions. The system considers nine parameters,

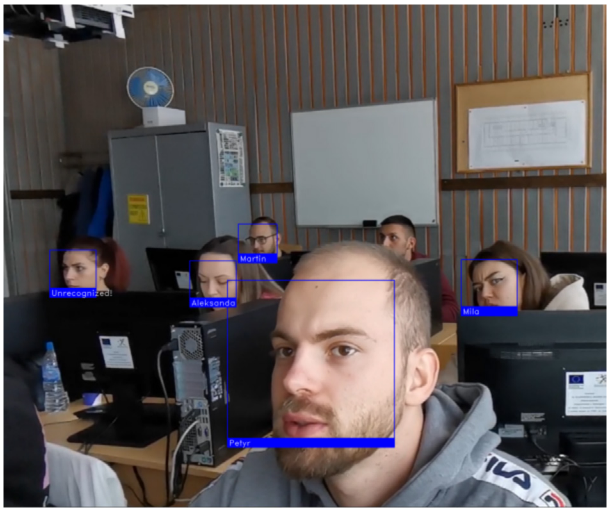


Fig. 1. Results of facial recognition.

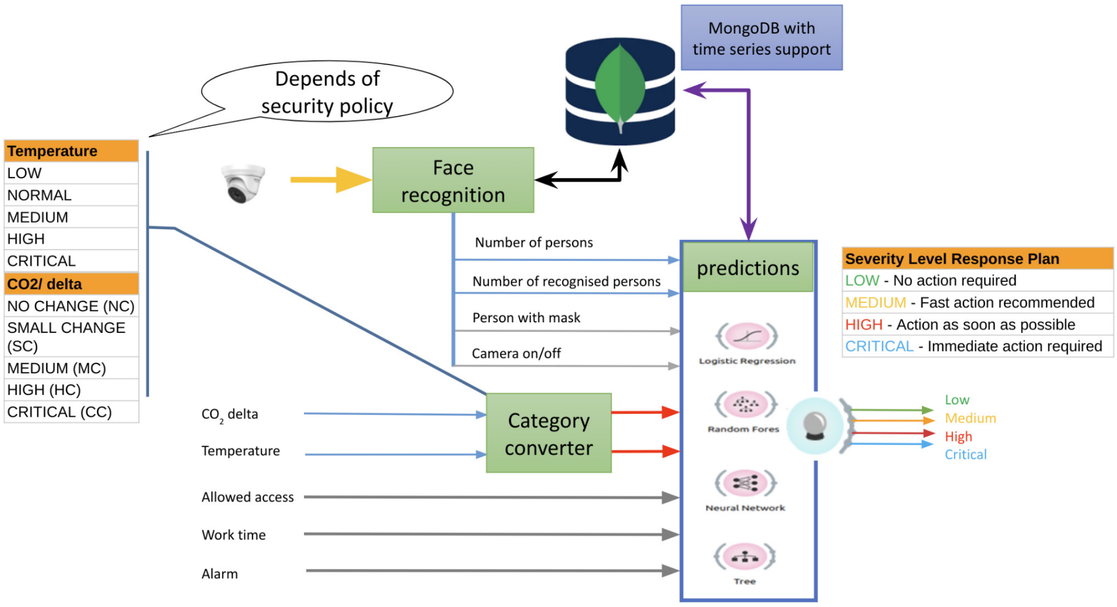


Fig. 2. Architecture of the system.

conventionally divided into three categories: video surveillance, sensors, and access level.

- CCTV - includes four parameters: persons detected in the room, number of persons detected, presence of masked people and whether the CCTV module is working.
- Sensors - temperature and CO_2 change in the room are recorded. The reading of CO_2 change, in combination with detected people in the room, allows early determination of the threat level, which is extremely important in this type of system. Separately, if the system detects high CO_2 change and at the same time a small number of persons in the room, or vice versa, this is a sign of a security-related irregularity.
- Access - contains information regarding the access allowed to the room by employees, and whether the access is during working hours.

The proposed system is built from three main modules: face recognition, category converter and predictions.

Face recognition

The face recognition module uses a CNN based on the VGG Face architecture [14]. Fig. 3 shows its architecture built from:

Module for processing the incoming frame:

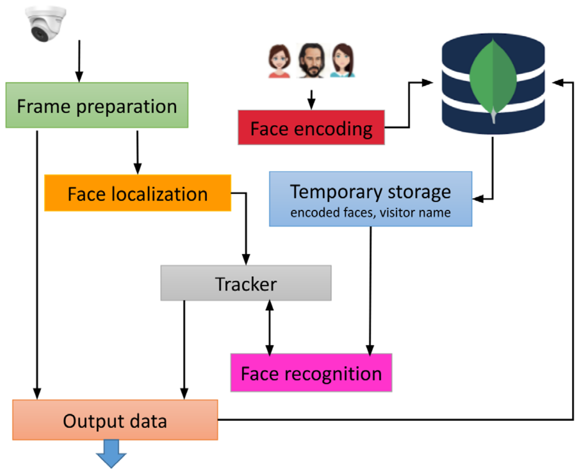


Fig. 3. Architecture of the facial recognition module.

- Face localization module;
- Visitor recognition module;
- Visitor tracking module between frames;
- Output module;
- Face image encoding module;
- Module for temporary storage of processed data.

Incoming Frame Processing Module - aims to convert the incoming frame to the appropriate format used by the other modules of the system. The main tasks of the module are:

- Reduce the image size to ¼ to save computation in subsequent operations;
- Transforming the image into black and white format;
- Implement a mechanism whereby incoming images are processed by the system through a frame.

Face localization module - determines and returns the coordinates of all recognized faces. The coordinates are set as vertices of a rectangle for each of the *x* and *y* axes. ***Visitor recognition module*** - returns the name of the visitor in case the face is recognized (it is in the database). ***Face image encoding module*** - the purpose of this module is to encode the face images of the so-called known visitors that the system needs to keep track of. The data of these visitors is submitted as an image (*jpg* or *png* format) containing a profile of the visitor's face. The name of the visitor is extracted from the name of the file containing his image. ***Temporary storage of processed data modules*** - the purpose of the module is to extract the data related to the names and face encodings of the known visitors from the database and store them in the global variables for the module. The idea is not to send constant queries to the database, thus increasing the system's speed. ***Inter-frame visitor tracking module*** (Tracker) – it uses a centroid tracking algorithm, designed to track visitors detected in the room in order to avoid subsequent calculations related to determining the face code of the persons registered by the system. ***Output module*** - passes the information related to the persons detected, how many of them are recognized, and whether there is a person wearing a mask, to the prediction module, stores the available data in MongoDB. ***Category converter*** - **c**ategorizes the sensor data. The values for the individual categories depend on the security policy of the organization. ***MongoDB*** - real-time information coming into the system is stored in MongoDB with support for time series data [12]. Using this data we can make a forecast, several days ahead, about the expected room temperature, number of visitors, etc. For example, if the prediction of the expected number of people in the room differs drastically from what is currently reported, this would be a sign of an irregularity that would be reflected by an increase in the threat level (low → medium, high → critical) [15].

The analysis of the stored data will, in turn, allow security policies to be updated according to detected gaps, threats and irregularities.

Severity alarm generation

Once the severity level has been assessed, the system should generate a severity alarm with a level from 1 (low severity) to 4 (critical). The severity alarm should trigger appropriate action based on the severity level. We propose the following severity levels to assess the impact of security breaches:

Severity Level 1: Low severity - No action required

This level represents a security incident that has minimal or no impact on the organization's operations or security.

Severity Level 2: Moderate severity - Fast action recommended

In such situations, the system will notify the system administrator or security personnel for immediate investigation. To prevent further unauthorized access or damage, it is advisable to temporarily disable access to the affected system or area. Maintaining a record of the incident and monitoring it is crucial to prevent its escalation over time.

Severity Level 3: High severity - Action as soon as possible

High severity incidents may include unauthorized access to sensitive data, physical breaches of critical IT infrastructure, malware infections or denial of service attacks. Access to the affected system/area should be promptly disabled, and an incident response plan should be initiated to mitigate the damage, based on the incident's nature and the organization's policies.

Severity Level 4: Critical severity - Immediate action required

This level represents the most severe security incident, demanding immediate action from the system administrator or security personnel. In such cases, the system will notify the system administrator or security personnel for immediate investigation. Access to the affected system or area should be disabled as soon as possible and without delay. Incident response plan must be initiated to contain the attack and minimize the extent of damage.

Incident response plan for cyber-physical security

This plan should be prepared in accordance with the organization's security policies. Once the severity level has been assessed, the incident response plan can provide a structured approach to responding to the security breach. The incident response plan should include procedures for investigating and responding to security breaches based on the severity level. This plan should include procedures for notifying appropriate persons, conducting investigations, and restoring the system to normal operation. The incident response plan should be a well-documented and well-communicated set of procedures. For example, the NIST framework for incident response includes four stages: preparation and prevention; detection and analysis; containment, eradication, and recovery; and post-incident activity [16].

Predictions

A neural network as well as machine learning algorithms are used to predict the threat level. The implemented neural network (Fig. 4) consists of 9 inputs, 4 outputs and two hidden layers.

The network is trained with a dataset of 480 records (Fig. 5), referring to a development office (several rooms), which houses 2 servers, the developers' computers, work documentation, etc. The alarm is activated when a fire sensor, a panic button and/or a door or window is opened. Individual rooms have different access levels, and the number of employees is up to 30.

The accuracy achieved by the neural network is 86.6 %. Since no similar datasets currently exist, as the data accumulates the training of the neural network will be refined, which will affect the accuracy of the model. Separately, as can be seen from Fig. 6, a large percentage of the errors are in the transitions between low → medium and medium → high. This is because the set boundary between the respective levels is fuzzy, which is normal for this type of system. The problem would be solved by optimizing the boundary levels depending on the security policies of the organization.

Similar results have been achieved using machine learning models such as Logistic Regression, Random Forest and Decision Tree (Fig. 7).

RESULTS AND DISCUSSION

The proposed architecture allows for flexibility regarding the parameters used in the process of determining the threat level associated with the cyber-physical security of a site. The criteria regarding the threat level can be easily and quickly adapted when security policies change, and the adaptation of the CNN

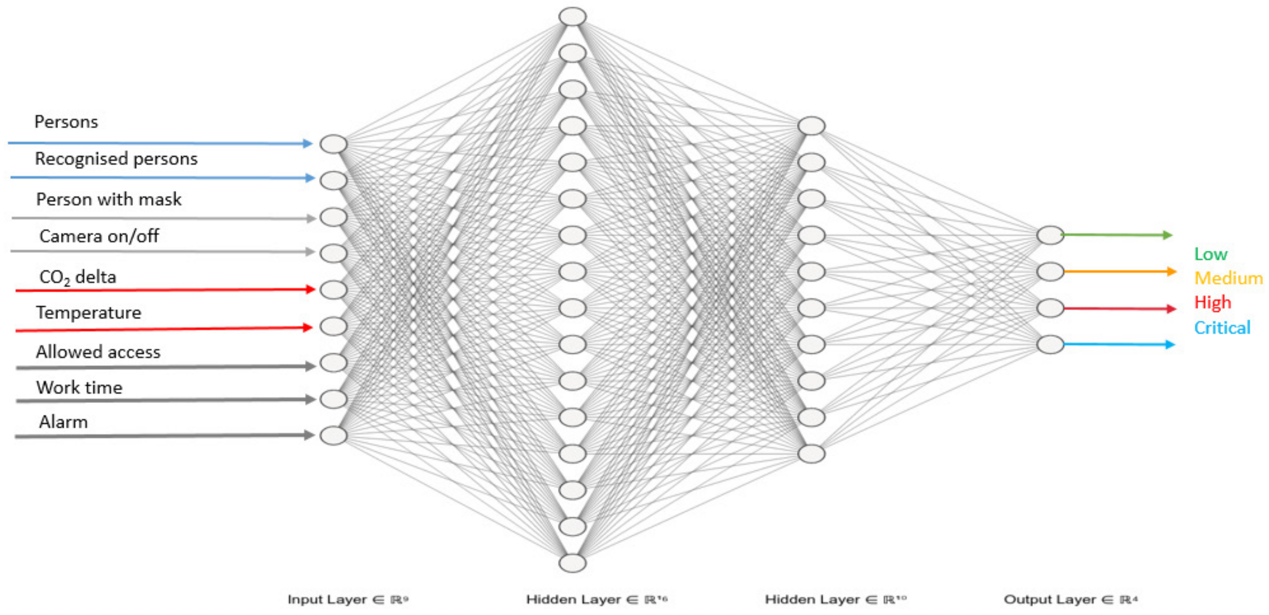
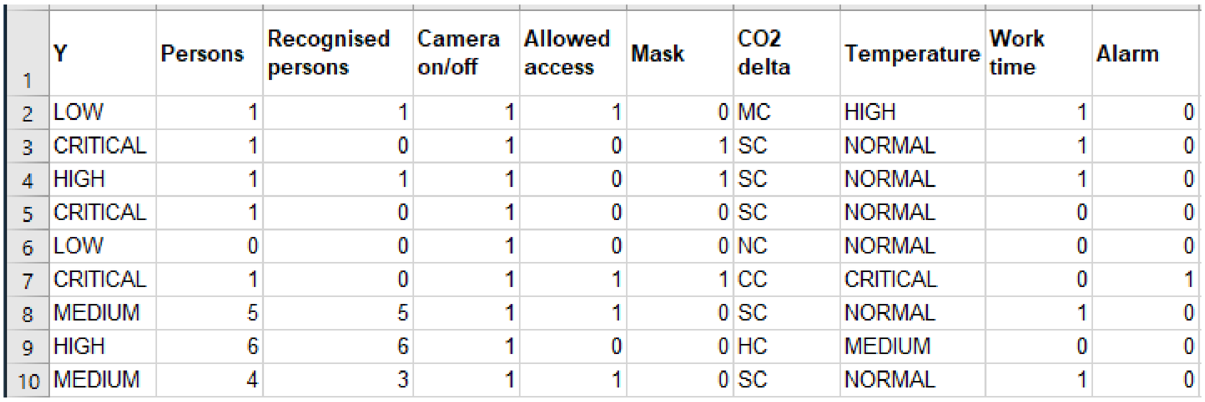
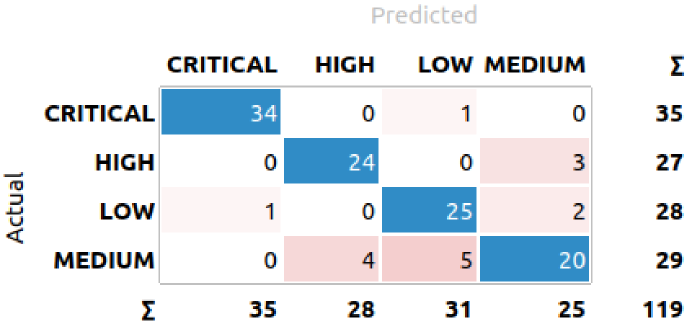

Persons
Recognised persons
Person with mask
Camera on/off
CO_2 delta
Temperature
Allowed access
Work time
Alarm
Low
Medium
High
Critical
Input Layer $\in \mathbb{R}^9$
Hidden Layer $\in \mathbb{R}^{16}$
Hidden Layer $\in \mathbb{R}^{10}$
Output Layer $\in \mathbb{R}^4$

Fig. 4. Neural network model.



1	Y	Persons	Recognised persons	Camera on/off	Allowed access	Mask	CO2 delta	Temperature	Work time	Alarm
2	LOW	1	1	1	1	0	MC	HIGH	1	0
3	CRITICAL	1	0	1	0	1	SC	NORMAL	1	0
4	HIGH	1	1	1	0	1	SC	NORMAL	1	0
5	CRITICAL	1	0	1	0	0	SC	NORMAL	0	0
6	LOW	0	0	1	0	0	NC	NORMAL	0	0
7	CRITICAL	1	0	1	1	1	CC	CRITICAL	0	1
8	MEDIUM	5	5	1	1	0	SC	NORMAL	1	0
9	HIGH	6	6	1	0	0	HC	MEDIUM	0	0
10	MEDIUM	4	3	1	1	0	SC	NORMAL	1	0

Fig. 5. Part of used dataset.



Predicted

Actual

	CRITICAL	HIGH	LOW	MEDIUM	Σ
CRITICAL	34	0	1	0	35
HIGH	0	24	0	3	27
LOW	1	0	25	2	28
MEDIUM	0	4	5	20	29
Σ	35	28	31	25	119

Fig. 6. Neural network Confusion matrix.

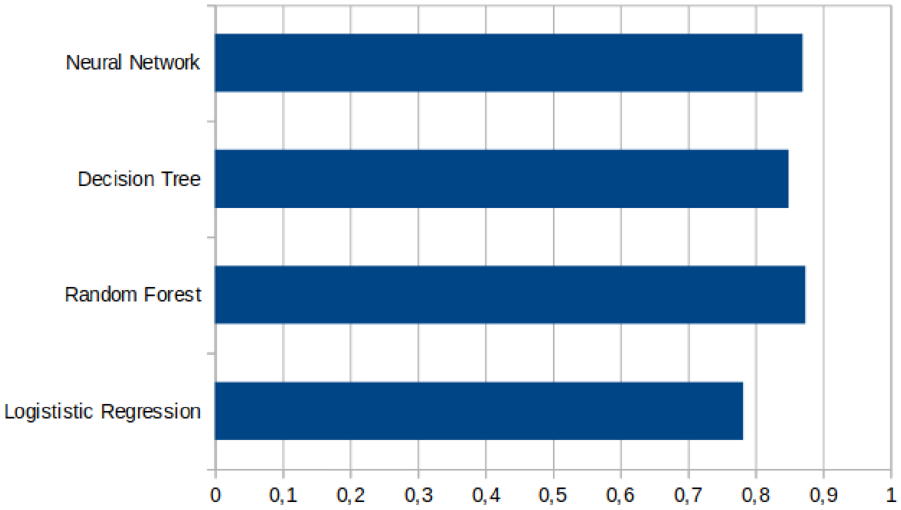


Fig. 7. Accuracy of the models.

associated with the facial recognition of visitors provides new perspectives for the development of cyber-physical security assurance systems. For example, timely tagging of unauthorized individuals increases security regarding copying, modification, or destruction of data and/or property. Separately, in work premises such as a bank or development office, even the emotional state of the persons present (visitors/workers) is relevant for security. There are developments using CNN that can determine the emotions of the detected visitors, and reporting this type of data would increase the security of both the workers and the data and equipment available [17, 18].

CONCLUSIONS

The integration of AI methods, specifically face recognition techniques combined with additional data from CO_2 and temperature sensors and neural networks in cyber-physical security, brings several benefits including improved accuracy, reduced response time and enhanced threat detection capabilities. This holds significant promise for enhancing cyber-physical security. However, it is essential to consider privacy concerns, ethical considerations, potential algorithmic biases and other limitations of AI algorithms when implementing such systems to ensure responsible and effective deployment. Neural networks have shown remarkable capabilities in analyzing complex patterns and making intelligent decisions. By leveraging pre-trained neural network models including model training, data representation and decision-making processes, physical security systems can automatically assess the severity of detected incidents or potential threats. As AI continues to advance, the application of these technologies in physical security will likely play a pivotal role in safeguarding physical assets, facilities, and individuals in an increasingly interconnected world.

REFERENCES

1. D.C. Wilson, Cybersecurity, MIT Press, 2021.
2. A. Pentland, D.L. Shrier, H.E. Shrobe, New Solutions for Cybersecurity. MIT Press, 2018.
3. I. Agrafiotis, J.R.C. Nurse, M. Goldsmith, S. Creese, D. Upton, A taxonomy of cyber-harms: Defining the impacts of cyber-attacks and understanding how they propagate, Journal of Cybersecurity, 4, 1, 2018, tyy006, https://doi.org/10.1093/cybsec/tyy006
4. M. Erbschloe, Physical security for IT, Elsevier Digital Press, 2005
5. Standards ENISA, available 11 November 2022 at https://www.enisa.europa.eu
6. R. Gürfidan, M. Ersoy, O. Kilim, AI-Powered Cyber Attacks Threats and Measures. In: Hemanth, D.J., Yigit, T., Kose, U., Guvenc, U. (eds) 4th International Conference on Artificial Intelligence and Applied Mathematics in Engineering, ICAIAME 2022, Engineering Cyber-Physical Systems and Critical Infrastructures, vol 7. Springer, Cham, 2023, https://doi.org/10.1007/978-3-031-31956-3_37
7. R.M. Clark, S. Hakim, Cyber-physical security: Protecting critical infrastructure at the state and local level, Springer, 2017.
8. U. Michelucci, Advanced Applied Deep Learning: Convolutional Neural Networks and Object Detection, 2019.
9. R.W. Gehl, S.T. Lawson, Social Engineering, MIT Press, 2022.
10. B. Mehlig, Machine learning with neural networks: An introduction for scientists and engineers, Cambridge University Press, 2022.
11. L. Alzubaidi, J. Zhang, A.J. Humaidi, A. Al-Dujaili, Y. Duan, O. Al-Shamma, J. Santamaría, M.A. Fadhel, M. Al-Amidie, L. Farhan, Review of deep learning: concepts, CNN architectures, challenges, applications, future directions, J. Big. Data, 8, 53, 2021, https://doi.org/10.1186/s40537-021-00444-8
12. Time Series Collections MongoDB Manual, available June, 2023 at https://www.mongodb.com/docs/manual/core/timeseries-collections
13. O. Tushkanova, D. Levshun, A. Branitsky, E. Fedorchenko, E. Novikova, I. Kotenko, Detection of cyberattacks and anomalies in cyber-physical systems: Approaches, data sources, evaluation. Algorithms, 16, 2, 2023, 85, https://doi.org/10.3390/a16020085
14. Q. Cao, L. Shen, W. Xie, O.M. Parkhi, A. Zisserman, VGGFace2: A dataset for recognizing faces across pose and age, In FG, 2018.
15. D. Li, D. Chen, L. Shi, B. Jin, J. Goh, S.K. Ng, MAD-GAN: Multivariate Anomaly Detection for Time Series Data with Generative Adversarial Networks. In Proceedings of the International Conference on Artificial Neural Networks, Munich, Germany, 17-

19, 2019.

16. NIST Cybersecurity Framework, available: https://www.nist.gov/cyberframework

17. A. Atanassov, D. Pilev, F. Tomova, Bimodal System for Facial and Body Gestures Emotion Recognition, J. Inform. Innov. Technol., 3, 3, 2021.

18. A. Atanassov, D. Pilev, Pre-trained Deep Learning Models for Facial Emotions Recognition, International Conference Automatics and Informatics (ICAI), 2020, DOI: 10.1109/ICAI50593.2020.9311334