UDC 681.5.004

DOI: 10.15587/1729-4061.2024.317000

DESIGN OF AN INTELLIGENT MODULE FOR DETECTING SIGNS OF INFORMATION SECURITY THREATS AND THE EMERGENCE OF UNRELIABLE DATA

Irina Cherepanska
Corresponding author
Doctor of Technical Sciences, Professor*
E-mail: cherepanskairina@gmail.com

Artem Sazonov
PhD, Associate Professor
Department of Automation Hardware and Software**

Yuriy Kyrychuk
Doctor of Technical Sciences, Associate Professor*

Petro Melnychuk
Doctor of Technical Sciences, Professor
Department of Manufacturing Engineering***

Dmytro Melnychuk
Doctor of Economic Sciences, Professor
Department of Psychology and Social Welfare***

Nataliia Nazarenko
PhD, Senior Lecturer*

Volodymyr Pryadko
Department of Electrification, Production Automation
and Engineering Ecology
Polissia National University
Stary blvd., 7, Zhytomyr, Ukraine, 10008

Serhii Bakhman
PhD Student
Department of Manufacturing Engineering***

Davyd Khraban
PhD Student
Department of Manufacturing Engineering***

*Department of Automation and Non-Destructive Testing Systems**
**National Technical University of Ukraine
“Igor Sikorsky Kyiv Polytechnic Institute”
Beresteysky ave., 37, Kyiv, Ukraine, 03056
***Zhytomyr Polytechnic State University
Chudnivska str., 103, Zhytomyr, Ukraine, 10005

At the stage of production preparation, there is an urgent need for an automated system that would timely detect signs of threats to information security and the emergence of unreliable data. To solve this problem, an intelligent module capable of detecting such threats and unreliable and/or anomalous data has been designed. The proposed intelligent module is the state-of-art, original, and effective toolkit. It can be recommended for practical use as part of the well-known information and computer system for automated modeling of the system of automatic orientation of production objects at the stage of technological preparation of machine and instrument-building production. Its application makes it possible to increase information security and reliability of important production data at the stage of technological preparation of production, in particular, when modeling systems for automatic orientation of production objects. In addition, the use of the proposed intelligent module makes it possible to obtain a number of important social and economic effects. Some of these effects are manifested in the prevention or reduction of material, intellectual and time costs for saving and restoring information, etc.

Automated analysis of important production data regarding their reliability and abnormality is carried out by machine learning methods using a specially designed advanced variational autoencoder based on classification algorithms and using wavelet transformation.

The designed intelligent module for detecting signs of a threat to information security and the emergence of unreliable and/or anomalous data works in real time with a high accuracy of 97.53 %. It meets the requirements of modern production

Keywords: information-computer system, automated modeling, control, artificial intelligence, flexible production system

Received 20.09.2024
Received in revised form 18.11.2024
Accepted 30.11.2024
Published 25.12.2024

***How to Cite**: Cherepanska, I., Sazonov, A., Kyrychuk, Y., Melnychuk, P., Melnychuk, D., Nazarenko, N., Pryadko, V., Bakhman, S., Khraban, D. (2024). Design of an intelligent module for detecting signs of information security threats and the emergence of unreliable data. Eastern-European Journal of Enterprise Technologies, 6 (2 (132)), 49–63. https://doi.org/10.15587/1729-4061.2024.317000*

1. Introduction

At the current stage of development of the machine and instrument industry, the processing of large volumes of production data is becoming an increasingly important strategic task facing specialists when solving various problems of technological preparation of production. This also applies to the problem of modeling systems for automatic orientation of production objects (SAOPOs). After all, SAOPO is one of the important components of flexible production systems (FPSs) in the machine and instrument building industry. That is why the reliability of data on structural features and physical and mechanical properties of production objects (POs), technical and economic characteristics and functional capabilities of technological equipment (TE), parameters of technological processes (TPs) plays a key role. Given the constant growth of production data volumes generated by numerous sensors, mechanisms, specialized information and computer systems (ICSs) and software, it is necessary to detect signs of threats to information security and unreliable data. It is evident that automating the detection of signs of a threat to information security and the emergence of unreliable or so-called anomalous data is the key to the successful operation of SAOPO, as well as FPS, and the enterprise as a whole.

It is obvious that the effective solution of this problem when modeling SAOPO requires complex and integrated application of effective and high-performance methods and tools. Artificial intelligence (AI) technologies, deep machine learning, cluster and regression analysis play a decisive role here. The appropriate toolkit based on the complex and integrated application of these technologies will allow efficient processing of data under an automated mode, detecting unreliable and/or anomalous data. An example can be an intelligent module for detecting signs of a threat to information security and the emergence of unreliable and/or anomalous data integrated into the well-known ICS for automated simulation of SAOPO [1].

ICS for automated simulation of SAOPO is designed to solve the important task of technological preparation of production of a scientific and applied nature. The purpose of its operation is to provide specialists with the necessary data to quickly make an informed, correct, and effective decision regarding the SAOPO composition and organization technique. This decision is based on the use of reliable information about the structural features and physical and mechanical properties of POs, technical and economic characteristics, and functional capabilities of a number of orientation devices (ODs) and other TE, as well as TP parameters [1]. Taking into account the production needs of ICS for automated modeling, SAOPO provides the possibility of collective access of the company's specialists to its individual functional modules, industrial database management systems (DB), etc. [1]. That is why the protection of production information, the preservation of commercial secrets, the immediate detection of unreliable and/or anomalous data indicating potential problems, the application of countermeasures to prevent future problems are extremely important.

2. Literature review and problem statement

It is known that the task of identifying signs of a threat to information security and the emergence of unreliable and/or anomalous data is one of the most difficult and important tasks facing modern specialists. This problem is the subject of a number of relevant studies of a scientific and applied nature, which are reflected in related works.

Thus, paper [2] provides an approach to detecting anomalies in the process of deep learning with representative hidden features for low-discrimination and insufficiently anomalous data. The method of detecting anomalies using artificial intelligence is used to monitor technological processes. The feature fusion strategy is designed to study hidden representational features. Minor deviations of operating parameters are reflected by statistical indicators. Characteristics related to the quality of the final product are selected taking into account unbalanced data. The uncontrolled anomaly detection module solves the problem of unbalanced data of technological processes. The given material is certainly valuable basic research, but it needs to be refined in terms of information protection at the early stages of production, in particular, during the technological preparation of production and design of appropriate tools.

The Isolation Forest model is presented in [3]. This model is an effective tool for detecting anomalies in measurement data when determining the mass and volume flow of liquid flowing through a pipeline with a Coriolis flowmeter. The data collected during metrological studies are processed by the anomaly detection model. This model analyzes the data and detects anomalous or outlier values that may indicate systematic or random measurement errors. At the same time, the authors do not consider other cases of anomalies, in particular in production data, at the stage of technological preparation of production, etc. However, the studies, after their appropriate additional detailed refinement, can be valuable for solving the problem of detecting anomalies and assessing the reliability of other types of data.

Papers [4, 5] propose data-driven AI architectures that are trained exclusively on so-called healthy signals combining LSTM regressors and OC-SVM classifiers. Algorithms for automated detection of any anomalous mechanical behavior recorded during vibration measurements during bearing operation are presented. However, the given information is somewhat declarative in nature and requires additional refinement and adaptation to solve the problem of detecting anomalies in production data when modeling SAOPO at the stage of technological preparation of production.

Work [6] reports the architecture of an artificial NN for data management. It is trained on signals combining LSTM regressors and OC-SVM classifiers. An algorithm for automated detection of abnormal vibration of rolling bearings during their operation is also given. Undoubtedly, this NN is a useful toolkit, but it has significant limitations in its application because the results of work [6] are limited to operational studies of rolling bearings and do not extend to other types of production data. Thus, the lack of an effective toolkit for the automated detection of threats and anomalies at the stage of production preparation, in particular during the simulation of SAOPO, is a serious obstacle to the practical application of the results from [6].

Paper [7] presents an approach to machine learning for detecting anomalies in industrial control systems based on measurement data. This approach is based on measurement data in the supervisory control and data acquisition (SCADA) system. A measuring intrusion detection system (MIDS) has been designed. This system makes it possible to detect any anomalous activity in SCADA, even if an attacker tries to hide it only at the SCADA control level. However, the information in the work is somewhat declarative and is presented in fragments. It is obvious that the material presented in the work needs additional detailed refinement and adaptation.

Work [8] reports an approach to detecting anomalies to ensure the industrial quality of finished products of a machine-building enterprise. To this end, models based on deep neural networks were used, which are not limited to a fixed set of categories and can generally assess the overall quality of PO. A quality control case from a European car manufacturer is used and the detection performance of three unsupervised models (e.g., Skip-GANomaly, PaDiM, PatchCore) is evaluated. Based on an in-depth evaluation study, it is demonstrated that reliable results can be achieved with fully unsupervised approaches that are even competitive with their supervised counterparts. The approach presented in the work is a valuable basic result but it needs to be refined and adapted to the problem because it does not take into account the specificity of tasks related to technological preparation of production.

Work [9] describes a study on the implementation of security solutions at an innovative manufacturing enterprise using the Internet of Things and machine learning. The study was based on the collection of archival data from telemetry sensors, IoT cameras, and control devices at an intelligent manufacturing facility. This data became the basis for training machine learning models used to detect anomalies over time. The study in [9], although it demonstrated the potential of machine learning, has a limited scope of application, and does not cover the problem of automated detection of threats and data anomalies in the simulation of SAAOV.

Work [10] reports the results of a study on the detection of anomalies in quantitative time series of data on the state of industry based on correlation and long-term short-term memory. A model for detecting anomalies in time series data of the number of states is presented, which is built using correlation supported by long-term short-term memory. The operability and efficiency of the model was checked on the data of real physical production processes. Although the study was successful in detecting anomalies, its application to the modeling of SAOPO is limited by the complexity of the data on its components.

Work [11] presents an anomaly detection system based on quantile regression forests. This system detects any abnormal deviations from the normal behavior of an individual device. Device behavior is defined as the number of network traffic events observed during a predetermined time period. The behavior of each device in a normal state is modeled depending on its historical behavior in defined time intervals. Based on the time intervals, the anomaly detection system characterizes as anomalous any behavior observed outside these intervals. Although the system has shown promising results in anomaly detection, it requires significant refinement and adaptation to input data on the constituents of SAOPO, which is time-consuming and labor-intensive. This limits the possibilities of its application.

Paper [12] proposes a combination of Zero Trust architecture with a three-layer security system to provide access control and authentication of users and devices in Cyber-Physical Systems (CPS) environments. CPS are integrated systems, including production systems, combining software and physical components. Although this combination of Zero Trust architecture with a three-level protection system has advanced the protection of information in cyber-physical systems environments to a new level, there are still unsolved issues related to the automated detection of signs of threats to information security and the emergence of unreliable and/or anomalous data right at the stage of technological preparation production, in particular, in the simulation of SAOPO.

Paper [13] reports a method of CPS protection using a software-defined network (SDN) to create deterministic data flows between CPS components. At the same time, any deviation is classified as an anomaly. This method is appropriate for distributed CPS environments that use networks such as the Internet.

However, this method is not effective for such physically isolated environments as FPS, including SAOPO, which use traditional models of network security. In addition, issues related to the automated detection of signs of information security threats and the emergence of unreliable and/or anomalous data at the stage of technological preparation of production, in particular during the simulation of SAOPO, remained unresolved. Also, the problem of designing the appropriate toolkit has not been solved.

Work [14] presents the technology of network anomaly detection and security protection based on machine learning. This technology is aimed at improving network anomaly detection and security using machine learning. Issues of scalability and interpretability of machine learning models and resistance to attacks are considered. Although this technology has advanced anomaly detection and information protection to a new level, it relies heavily on neural network architectures. Limitations inherent in neural network architectures in detecting global dependences can affect the technology's ability to generalize different groups of input information about the components of SAOPO.

In [15], it is shown that the state of CPS depends on the measurements of sensors for monitoring and controlling the plant. Moreover, any change in the behavior of a physical process due to a cyber attack can also be detected from sensor readings. Under various circumstances, these sensor measurements follow typical patterns. A Boolean-based supervised classification method known as Logical Data Analysis (LAD) can extract patterns (or rules) from archived sensor readings, and these rules can determine plant status. In the work, based on these rules, the anomaly detection system (ADS) of the plant's behavior as a whole functions. Despite the fact that ADS is able to record changes in the behavior of a physical process due to a cyber attack, it may not effectively record the local content of the specific content of the components of SAOPO. This is a limitation of ADS, which makes it difficult to effectively apply it to data on SAOPO constituents, especially under complex production conditions.

Thus, our review of the literature demonstrates that available studies do not offer comprehensive solutions and appropriate tools for the automated detection of signs of threats to information security and the emergence of unreliable and/or anomalous data when modeling SAOPO. Thus, the need for research into the problem of timely identification of signs of threats to information security and the emergence of unreliable data is justified. It includes preserving commercial secrets, identifying unreliable and/or anomalous important production data when modeling SAOPO, and proactively preventing potential problems.

3. The aim and objectives of the study

The purpose of our research is to design an intelligent module for detecting signs of a threat to information security and the emergence of unreliable and/or anomalous data when modeling SAOPO at the stage of technological preparation of production. This would make it possible to automatically detect signs of unauthorized access and anomalies in data critical for modeling SAOPO in a timely manner, which could increase the stability of ICS and the quality of decision-making at the stage of technological preparation of production.

To achieve the goal, the following tasks must be solved:

– to build a PDCA model and a structural diagram of an intelligent module for detecting signs of a threat to information security and the emergence of unreliable and/or anomalous data, which is integrated into ICS for the automated modeling of SAOPO;

– to define methods and design means for automated detection of signs of threats to information security and the emergence of unreliable and/or anomalous data;

– to conduct training and experimental studies of the Conditional Variational Autoencoder using an example of the Conditional Variational Autoencoder for automatic analysis of data anomalies on production objects;

– to develop an algorithm for the functioning of the intelligent module to detect signs of a threat to information security and the emergence of unreliable and/or anomalous data.

4. The study materials and methods

The object of our research is the process of identifying signs of threats to information security and the emergence of unreliable and/or anomalous data.

The subject of the study is an intelligent module for detecting signs of a threat to information security and the emergence of unreliable and/or anomalous data.

The research hypothesis is based on the assumption that the implementation of the tasks of designing an intelligent module to detect signs of a threat to information security and the emergence of unreliable and/or anomalous data would improve:

– the level of protection of production data at the stage of technological preparation of production;

– speed of detection of signs of unauthorized access to ICS for automated simulation of SAOPO and timeliness of application of appropriate countermeasures;

– speed and accuracy of detection of unreliable and/or anomalous data and their removal.

In addition, the social and economic effect could be provided, which would involve the following:

– reduction of intellectual and time costs for data reliability analysis;

– prevention of leakage of important production data and innovative proposals;

– prevention of conflicts and claims as a result of non-fulfillment of obligations to partners and counterparties due to damaged and inaccurate information;

– increasing the quality of production and competitiveness of the enterprise as a whole.

Research methods are based on the application of tools, technologies, and methods of artificial intelligence, machine learning, classification or clustering algorithms, wavelet transformations, regression and comparative analysis.

The results of experimental studies, confirming the efficiency of the resulting advancements, were obtained during computer simulation. They are based on the application of computer modeling methods and special software, such as neurostimulators. A modern basic personal computer with an Intel(R) Core(TM) i3-7020U CPU @ 2.30 GHz and 4.00 GB of RAM was used. The methods of graphical representation and statistical processing were used to treat the experimental data. The selection of these methods and software was based on their availability and the possibility of practical implementation.

5. Results of research on the intelligent module for detecting threats and anomalous data

5. 1. PDCA model and structural diagram of an intelligent module for detection of threats and abnormal data

The well-known ICS for the automated simulation of SAOPO meets the basic requirements of information security control in accordance with international standards ISO 27005 and ISO 27001 [1]. From these positions, four levels of access to the processes of data accumulation, storage, processing, and transmission are established in this ICS [1]. The general scheme of the information network integration of ICS for the automated modeling of SAOPO in the general industrial information network of the machine- and instrument-making industries GKIS is described in work [1] and shown in Fig. 1.

However, in ICS for automated modeling of SAOPO, automated data analysis is not performed to establish their unreliability and/or abnormality. From these positions, the structure of ICS for the automated simulation of SAOPO, which is described in detail in paper [1] and shown in Fig. 2, was additionally supplemented with an intelligent module for detecting signs of a threat to information security and the emergence of unreliable and/or anomalous data.

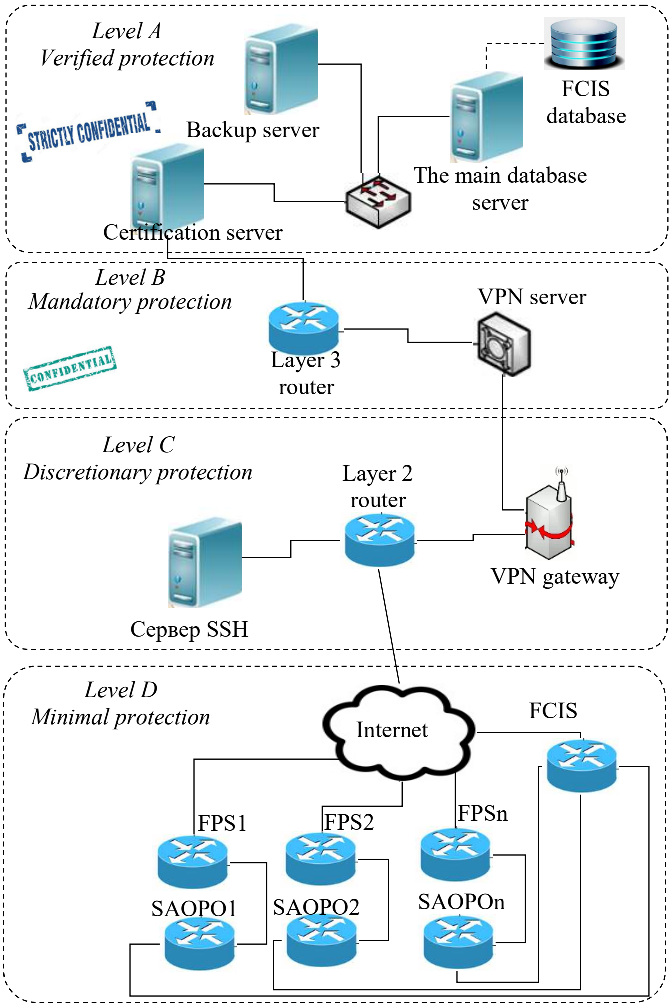

Fig. 1. General scheme of the information network integration of the information-computer system for the automated modeling of systems of automatic orientation of production objects into a general industrial information network with four levels of protection in accordance with TCSEC criteria [1]

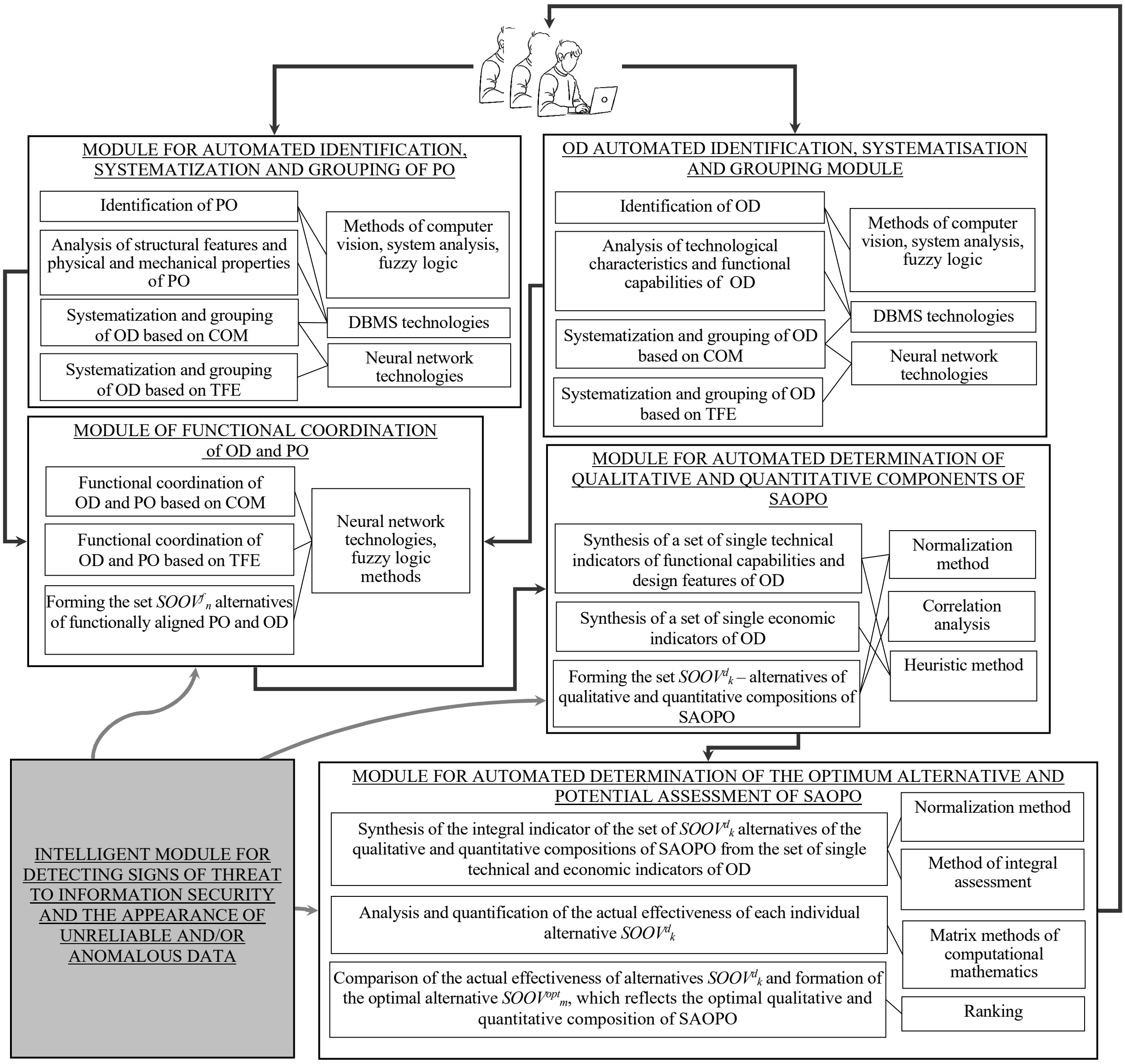
MODULE FOR AUTOMATED IDENTIFICATION, SYSTEMATIZATION AND GROUPING OF PO
Identification of PO
Analysis of structural features and physical and mechanical properties of PO
Systematization and grouping of OD based on COM
Systematization and grouping of OD based on TFE
Methods of computer vision, system analysis, fuzzy logic
DBMS technologies
Neural network technologies
OD AUTOMATED IDENTIFICATION, SYSTEMATISATION AND GROUPING MODULE
Identification of OD
Analysis of technological characteristics and functional capabilities of OD
Systematization and grouping of OD based on COM
Systematization and grouping of OD based on TFE
Methods of computer vision, system analysis, fuzzy logic
DBMS technologies
Neural network technologies
MODULE OF FUNCTIONAL COORDINATION of OD and PO
Functional coordination of OD and PO based on COM
Functional coordination of OD and PO based on TFE
Forming the set $SOOV^f_n$ alternatives of functionally aligned PO and OD
Neural network technologies, fuzzy logic methods
MODULE FOR AUTOMATED DETERMINATION OF QUALITATIVE AND QUANTITATIVE COMPONENTS OF SAOPO
Synthesis of a set of single technical indicators of functional capabilities and design features of OD
Synthesis of a set of single economic indicators of OD
Forming the set $SOOV^d_k$ – alternatives of qualitative and quantitative compositions of SAOPO
Normalization method
Correlation analysis
Heuristic method
INTELLIGENT MODULE FOR DETECTING SIGNS OF THREAT TO INFORMATION SECURITY AND THE APPEARANCE OF UNRELIABLE AND/OR ANOMALOUS DATA
MODULE FOR AUTOMATED DETERMINATION OF THE OPTIMUM ALTERNATIVE AND POTENTIAL ASSESSMENT OF SAOPO
Synthesis of the integral indicator of the set of $SOOV^d_k$ alternatives of the qualitative and quantitative compositions of SAOPO from the set of single technical and economic indicators of OD
Analysis and quantification of the actual effectiveness of each individual alternative $SOOV^d_k$
Comparison of the actual effectiveness of alternatives $SOOV^d_k$ and formation of the optimal alternative $SOOV^{opt}_m$, which reflects the optimal qualitative and quantitative composition of SAOPO
Normalization method
Method of integral assessment
Matrix methods of computational mathematics
Ranking

Fig. 2. Structure of the information and computer system for automated modeling of the system of automatic orientation of production objects with an integrated intelligent module for detecting signs of a threat to information security and the emergence of unreliable and/or anomalous data

The basis of the work of the intelligent module for detecting signs of threats to information security and the emergence of unreliable and/or anomalous data is the PDCA model. The PDCA model reflects the life cycle structure of all processes of an arbitrary system in quality management [16–18], including information quality management. Information quality management involves analyzing data to determine their unreliability and/or abnormality and to assess the level of threats. In accordance with the PDCA model and the requirements of the ISO/IEC 27001 international standard for information security management and regulation, the functioning of the intelligent module for detecting signs of a threat to information security and the emergence of unreliable and/or anomalous data occurs in stages (Fig. 3).

At the first stage, the monitoring of the content of ICS is carried out for the automated simulation of SAOPO. In particular, the company's specialists form a so-called semantic core of symbols and words that define the subject matter and key tasks of the tasks of technological preparation of production, in particular, in the case of modeling SAOPO. For example, a set of symbols, words, and phrases that describe the structural features and physical and mechanical properties of PO, technological characteristics and functional capabilities of PE, features of technological processes, indicate the cost price of products, potential partners, competitors, and counterparties, etc. After the formation of the semantic core, the indexing of the content of ICS is performed for the automated modeling of SAOPO.

At the second stage, the signs of threats to information security, unreliability and/or data anomalies are evaluated.

In particular, to assess the signs of threats to information security, such indicators of the signs of threats are calculated as the risk level for one factor $Th_{i,j,k}$, the risk level for three possible factors $Th_{i,j,k\Sigma}$, the risk level for resource Th_R and the risk for resource R.

Based on the calculation results, a conclusion is drawn about the level of detected threats.

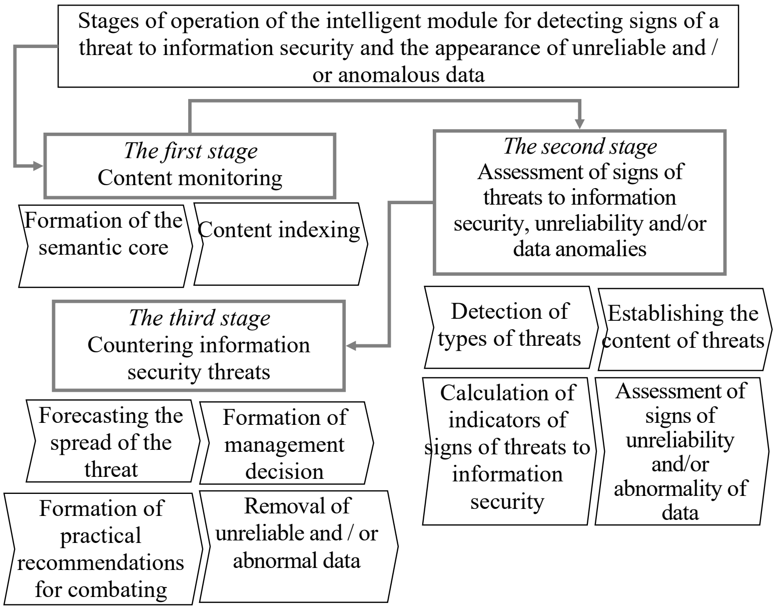
Stages of operation of the intelligent module for detecting signs of a threat to information security and the appearance of unreliable and / or anomalous data
The first stage
Content monitoring
Formation of the semantic core
Content indexing
The second stage
Assessment of signs of threats to information security, unreliability and/or data anomalies
Detection of types of threats
Establishing the content of threats
Calculation of indicators of signs of threats to information security
Assessment of signs of unreliability and/or abnormality of data
The third stage
Countering information security threats
Forecasting the spread of the threat
Formation of management decision
Formation of practical recommendations for combating
Removal of unreliable and / or abnormal data

Fig. 3. PDCA model of the functioning of the intelligent module for detecting signs of a threat to information security and the emergence of unreliable and/or anomalous data

Evaluation of signs of unreliability and/or data anomalies is carried out by machine learning and AI methods using an extended database (DB) with clear criteria of values and formalized descriptions of structural features and physical and mechanical properties of PO, technological characteristics and functional capabilities of software [19], features of TP. Given that unreliable data can be considered as deviations from expected values, that is, as anomalies, an advanced database is required for their effective detection and training of AI systems. At the same time, it is necessary that this database contains clearly defined criteria for normal indicators. This will allow new data to be checked for compliance with these criteria. For example, a formalized description of structural features and physical-mechanical properties of PO according to expression (1) and a formalized description of technological characteristics and functional capabilities of software according to expression (2) from [19]:

$$\mathrm{PO}\rightarrow\left\{\begin{array}{l}\left[\Phi_{\mathrm{C}}\left(o,p_i,\tau\right)\right];\\ \left[\Phi_{Frm}\left(b,h,l,\alpha,d_{\max},d_{\min}\right)\right];\\ \left[\Phi_{DF}\left(r,m,\xi,z\right)\right];\\ \left[\Phi_{CM}\left(\begin{array}{l}M_t\begin{pmatrix}M_{St},M_{Al},M_{CI},M_{Brz},\\ M_{Br},M_{Tr},M_{Plc},M_{Gls}\end{pmatrix},\\ F_e\left(F_f,F_p,F_d\right),E_l\left(E_p,E_{\bar{p}},E_{\bar{\bar{p}}}\right),\\ W_s\left(W_z,W_u,W_i\right),P_P\left(P_P,P_{\bar{P}}\right),\\ G_g\left(G_g,G_{\bar{g}}\right),Pg\left(Pg,P\bar{g}\right)\end{array}\right)\right]\end{array}\right\};\ (1)$$

where Ù is the logical conjunction function; Ñ – logical function of exclusive disjunction; Φ_C – a formalized description of the signs of symmetry of PO; o is the axis of rotation, $o\in(x, y)$; p_i is the plane of symmetry, $p_i\in(p_1, p_2, p_3)$, which is respectively perpendicular to the axis of rotation; parallel to the axis of rotation, i.e. passing through it, perpendicular to another plane if there is no axis of rotation in PO; τ – mutual parallel and perpendicular arrangement of axes and planes of symmetry, respectively, $\tau\in(\|, \perp)$; Φ_{Frm} – formalized description of PO shape; b – height; h – width; l – length; α – angle at the base, which can be determined by the slope of the generatrices of PO; $d_{\max}$, $d_{\min}$ – the smallest and largest diameters, respectively, of PO, which has a round shape; Φ_{DF} – a formalized description of the design features of PO; r – the ratio between the overall dimensions of PO, $r\in(r_1, r_2, r_3)$, where r_1 – all dimensions are the same, r_2 – some dimensions are almost the same, r_3 – a significant difference in dimensions; m is the value of the largest size of PO, $m\in(m_1, m_2, m_3)$, where m_1, m_2, m_3 – the largest size of PO up to 100 mm, PO size from 100 to 250 mm, PO size from 250 mm and more; ξ is the symmetry of PO ends, $\xi\in\{\xi, \bar{\xi}\}$, where $\xi, \bar{\xi}$ are symmetric and asymmetric ends, respectively; z is the presence of structural elements of PO, $z\in(z_{ext}, z_{int})$, where z_{ext}, z_{int} – the presence of external and internal structural elements, respectively; Φ_{CM} – a formalized description of the physical and mechanical properties of structural materials of PO; M_t is a parameter indicating the structural material of PO, $t\in${St, Al, CI, Brz, Br, Tr, Plc, Gls}, where St, Al, CI, Brz, Br, Tr, Plc, Gls is an abbreviated conditional designation of structural materials: steel, aluminum, cast iron, bronze, brass, wood, plastic, glass, respectively; F_e – ferromagnetic properties of materials, $e\in\{f, p, d\}$, where f, p, d are conventional designations of ferromagnet, paramagnet, and diamagnet, respectively; E_l is the electrodynamic properties of the material, $l\in\{p, \bar{p}, \bar{\bar{p}}\}$, where p, $\bar{p}$, $\bar{\bar{p}}$ is the conventional designation of the conductor, dielectric, and semiconductor properties of the material, respectively; W_s is a parameter that characterizes the reflective properties of the material, sÎ{z, u, i}, where z, u, are conventional designations of the properties of the material to reflect sound and ultraviolet waves, infrared radiation, respectively; P_P is a parameter characterizing the pneumatic properties of the material; $P\in\{P, \bar{P}\}$, where P, $\bar{P}$ is a conventional designation of expressed and unexpressed pneumatic properties of the material, respectively; G_g is a parameter characterizing the hydraulic properties of the material; $g\in\{\mathrm{g}, \bar{g}\}$, where g, $\bar{g}$ is a conventional designation of expressed and unexpressed hydraulic properties of the material, respectively; Pg is a parameter characterizing the pneumohydraulic properties of the material; $g\in\{\mathrm{g}, \bar{g}\}$, where g, $\bar{g}$is a conventional designation of expressed and unexpressed pneumohydraulic properties of the material, respectively [19]:

$$OD\rightarrow\left\{\begin{array}{l}\Phi_F;\left[M_{or_i},S_k,Z_{FOP_i},\varpi,k_{extr}\right];\left[\Phi_{FD}\left(D_{\Omega\Psi}\right)\right];\\ \left[\Phi_{P\ln}\left(\Pi\nabla H\nabla N_S\nabla U_I\nabla P_AV_K\nabla G\nabla A\right)\nabla N\nabla L\nabla K_{KK}\right];\\ \left[\Phi_{StF}\left(BLD\nabla Tr\nabla TS\nabla S-OG\nabla DSO\nabla DDOP\nabla DMO\right)\right];\\ \left[\Phi_{SB}\left(SB_i\,|\,i=\overline{1;19}\right)\right];\left[\Phi_U\left(D_U\nabla D_{\bar{U}}\right)\right];\left[\Phi_{AOM}\left(R_\varepsilon\nabla R_\eta\nabla R_\infty\right)\right]\end{array}\right\},\ (2)$$

where Ñ – logical function of exclusive disjunction; Φ_F – a formalized description of PE functionality, i.e. the PE implementation of the corresponding composition of orienting movements, a detailed description of which is given in work [5]; M_{ori} is an orientation method, i∈{Act, Pas}, where Act, Pas are conventional designations of the active M_{orAct} and passive M_{orPas} orientation methods, respectively; S is a parameter indicating the coordinate system relative to which PO is oriented, $k\in\{OD, PO, Abs\}$, where OD, PO, Abs are conventional designations of the PO coordinate system, the coordinate system of another PO, and the absolute coordinate system, respectively; Z_{FOPi} is a parameter that indicates the sign of the final oriented position (FOP), which is achieved by PO during the implementation of the orientation subfunction, a detailed description of which is given in [5]; Φ_{FD} – a formalized description of PO functional differences, detailed information on which is given in work [5]; Φ_{PIn} – a formalized description of the types of force effects of PO on PE during their automatic orientation, detailed information on which is given in work [1], by which the type of orientation system can be determined; *BLD* – conditional designation of bunker loading devices; *Tr* – conditional designation of trays that can perform orientation; *TS* – conditional designation of transport systems with orientation functions; *S-OG* – conditional designation of self-orienting grips of industrial robots; *DSO* – conditional designation of devices for spatial orientation; *DDOP* – conditional designation of devices for determining the occupied position; *DMO* – conditional designation of devices for mutual orientation; Φ_{SB} – conditional formalized description of the features of the constructions of the target bodies of PO; SB_i is a parameter that determines the design features of the target organs of PO, here i is a digital index corresponding to the value of the classification feature, $i\in\left\{1,\ 20\right\}$, where 1–20 – tubular, tray, frictional, vibrating, vibrating tray with shaped cutouts, hook, pocket, pin, with a bow, finger, sector, knife, disc, screw, slot, drum, star, slotted, non-contact, combined type of the target organ of PO, respectively; Φ_U – conditional formalized description of the universality of PO, i.e. the ability to change TE; D_U – universal PO, which has the possibility of reconfiguration for orientation of PO of a different size or shape; D_s – special PO that does not have the possibility of reconfiguration; Φ_{AOM} is a formalized description of the process mode of automatic orientation of OM, in particular R_ε – mode of individual orientation of OM, R_η – mode of orientation of OM in batches, R_Γ – mode of orientation of PO by continuous flow [19].

Data analysis and identification of those most associated with anomalies is carried out by deep machine learning methods, for example, using a special neural network or a so-called autoencoder based on classification or clustering algorithms.

At the third stage, mechanisms for countering threats to information security are launched, and unreliable and/or anomalous data are removed. The specific mechanisms for combating threats are determined depending on the level of the threat that was detected at the second stage of the operation of the intelligent module for detecting signs of a threat to information security and the emergence of unreliable and/or anomalous ICS data for the automated simulation of SAOPO. In particular, if the threat level is low, then it is advisable to forecast the distribution of content and requests for it in order to promptly adjust the control influences. In the case of a medium or high threat level, it is necessary to apply regulatory information protection tools. At the end of the stage of combating threats, practical recommendations are formed for the relevant units of the enterprise, depending on the level of the threat and the domain of activity it affects.

The structural diagram of the intelligent module for detecting signs of a threat to information security and the emergence of unreliable and/or anomalous data is shown in Fig. 4.

The structural diagram has been developed on the basis of the PDCA model of the functioning of this module.

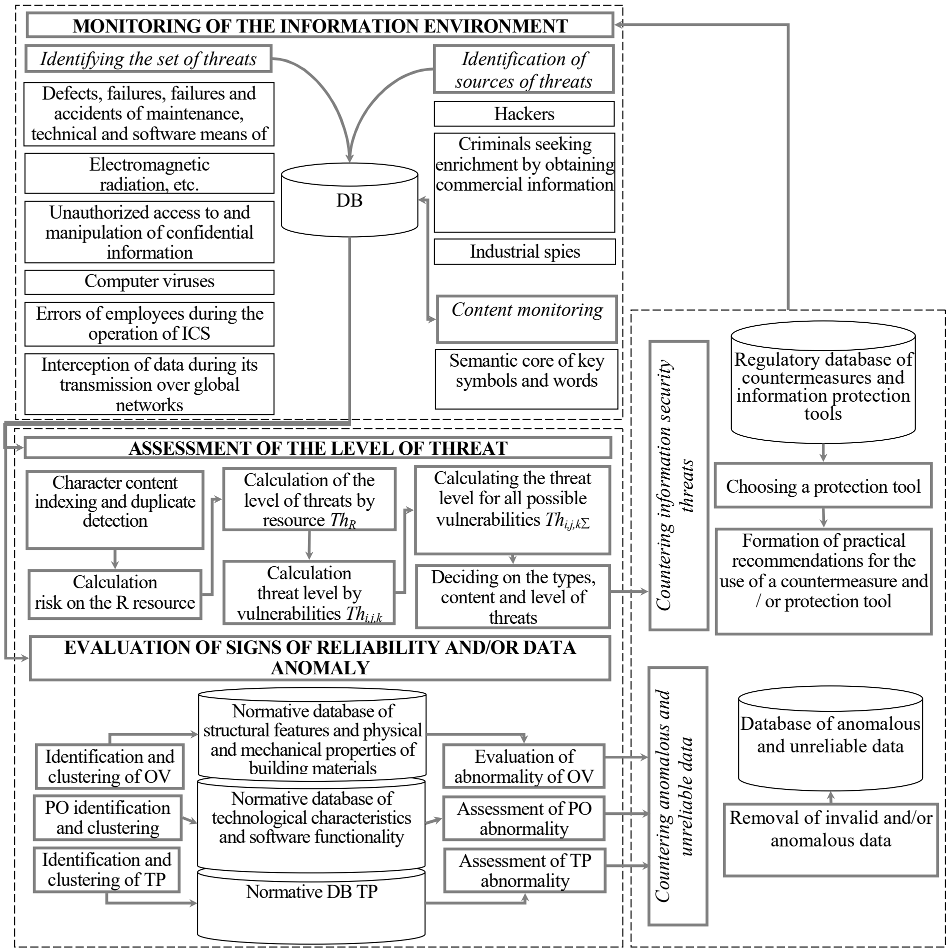
MONITORING OF THE INFORMATION ENVIRONMENT
Identifying the set of threats
Identification of sources of threats
Defects, failures, failures and accidents of maintenance, technical and software means of
Electromagnetic radiation, etc.
Unauthorized access to and manipulation of confidential information
Computer viruses
Errors of employees during the operation of ICS
Interception of data during its transmission over global networks
DB
Hackers
Criminals seeking enrichment by obtaining commercial information
Industrial spies
Content monitoring
Semantic core of key symbols and words
ASSESSMENT OF THE LEVEL OF THREAT
Character content indexing and duplicate detection
Calculation risk on the R resource
Calculation of the level of threats by resource Th_R
Calculation threat level by vulnerabilities $Th_{i,j,k}$
Calculating the threat level for all possible vulnerabilities $Th_{i,j,k\Sigma}$
Deciding on the types, content and level of threats
Countering information security threats
Regulatory database of countermeasures and information protection tools
Choosing a protection tool
Formation of practical recommendations for the use of a countermeasure and / or protection tool
EVALUATION OF SIGNS OF RELIABILITY AND/OR DATA ANOMALY
Identification and clustering of OV
PO identification and clustering
Identification and clustering of TP
Normative database of structural features and physical and mechanical properties of building materials
Normative database of technological characteristics and software functionality
Normative DB TP
Evaluation of abnormality of OV
Assessment of PO abnormality
Assessment of TP abnormality
Countering anomalous and unreliable data
Database of anomalous and unreliable data
Removal of invalid and/or anomalous data

Fig. 4. Structural diagram of the intelligent module for detecting signs of a threat to information security and the emergence of unreliable and/or anomalous data

5. 2. Methods and means for automated detection of signs of a threat to information security and the emergence of unreliable and/or anomalous data

To identify signs of threats to information security and anomalous data, it is suggested to use a comprehensive approach. In particular, the use of neural networks or autoencoders based on them is proposed, in combination with traditional methods such as regression analysis and wavelet transformation. The selection of methods and tools for the automated detection of signs of a threat to information security and the emergence of unreliable and/or anomalous data was made in view of their availability and practical implementation.

With the help of regression and comparative analysis, it is possible to assess the risks and potential losses from the use of software and ICS, in particular, in the modeling of SAOPO. The simplest way to measure the risks (probability) of the occurrence of adverse events or threats is a comparative analysis, which allows one, two, and three factors to be evaluated. Thus, the risk level for one factor $Th_{i,j,k}$ is calculated according to known formula (3) and is determined in the interval of values $Th_{i,j,k} \in [0; 1]$:

$$Th_{i,j,k} = \frac{ER_{i,j,k}}{100} \cdot \frac{P(V)_{i,j,k}}{100}, \tag{3}$$

where $ER_{i,j,k}$ – criticality of threat implementation, %; $P(V)_{i,j,k}$ – probability of threat realization due to the i-th, j-th, k-th factor during the given time interval, %.

The level of risk (probability) of the occurrence of adverse events or threats for three factors $Th_{i,j,k\Sigma}$ is calculated as the sum of products of individual risk levels $Th_{i,j,k}$ from formula (4) and is determined in the range of values $Th_{i,j,k} \in [0; 1]$:

$$Th_{ijky} = Th_i + Th_j + Th_k = \left(1 - \prod_{i=1}^{I}(1 - Th_i)\right) + \\ + \left(1 - \prod_{j=1}^{J}(1 - Th_j)\right) + \left(1 - \prod_{k=1}^{K}(1 - Th_k)\right). \tag{4}$$

The risk level for resource Th_R is calculated from formula (5):

$$Th_R = Th_{Ri} + Th_{Rj} + Th_{Rk} = \left(1 - \prod_{i=1}^{I}(1 - Th_{iy})\right) + \\ + \left(1 - \prod_{j=1}^{J}(1 - Th_{jy})\right) + \left(1 - \prod_{k=1}^{K}(1 - Th_{ky})\right). \tag{5}$$

The risk for resource R is calculated from formula (6):

$$R = Th_R \cdot D, \tag{6}$$

where D is the criticality level of the resource.

The amount of costs (losses) Z from the occurrence of an adverse event is estimated according to formula (7):

$$Z = Th_{ijk\acute{O}} C_{\Sigma} = Th_i \mathrm{C}_i + Th_j C_j + Th_k C_k, \tag{7}$$

where $Th_{i,j,k\Sigma}$ – threat level by three factors; C – the amount of loss from the occurrence of the threat due to the i-th, j-th, k-th factor.

Determining the integral indicator of the level of threats Th, from a mathematical point of view, is a multi-criteria problem that requires an assessment of the level of threats with different weighting factors. This problem can be solved according to the non-linear trade-off scheme given in [21], according to which a set of predefined threats is represented in vector form according to expression (8):

$$Th = \arg\min_{Th \in M} \sum_{j=1}^{J} \omega_j \left[1 - Th_j\right]^{-1}, \tag{8}$$

where ω_j is the weight coefficient of the j-th threat.

The weighting coefficients of threats are determined from formula (9) given in the literature [20]:

$$\omega_{\mathrm{j}} = \frac{r_j}{\sum_{j=1}^{n} r_j}, \tag{9}$$

where ω_j is the weight coefficient of the j-th threat; r_j – coefficient and correlation index of the j-th threat, which is established on the basis of subjective assessments of experts in view of a specific situation; n is the number of threats.

An integral estimate of the level of threats is obtained in the range of values [0; 1]. Based on the received assessment of the level of threats, a general conclusion is formed and an appropriate tool and/or countermeasure for information protection is determined (Fig. 4). For example, reducing the level of threat can be carried out both by the separate application of appropriate technical, management, and administrative means, and by their combination. Examples of possible values of the integrated assessment of the probability of occurrence of adverse events (threats) and the corresponding levels of threats are given in Table 1.

Table 1

Integral assessments of the probability of occurrence of adverse events (threats) and corresponding levels of threats

Quantitative value of the integral assessment of the level of threats	Threat level	Comment
0.00–0.20	Missing	The threat can be ignored
0.21–0.40	Low	The threat is insignificant, the consequences are easily eliminated, the losses (costs) are small, the impact on ICS is insignificant
0.41–0.50	Average	A threat with moderate negative results, elimination of the consequences does not bear large costs, damages from the occurrence of the event are not large, the impact on ICS is not critical
0.51–0.70	Above average	A threat with serious negative results, liquidation of the consequences involves large costs, losses from the occurrence of the event are significant, the impact on ICS is significant
0.71–1.00	High	A threat with extremely negative results, no possibility of ICS functioning, damages from the occurrence of the event are critical

Automated data analysis is carried out using machine learning methods. Thus, multilayer neural networks or an extended variational autoencoder (CVAE) based on them are used to detect anomalies. Additionally, wavelet transformation and classification algorithms are used. The latter, based on multi-scale analysis, help assign data to the appropriate categories. Multiscale data analysis is performed using a wavelet transform. It implies that the digital signal, which displays data about PO, PE, and technological processes (TP), is decomposed based on basis. It is formed by shifts and large-scale copies of functions describing OV, PO, and TP. Fig. 5 demonstrates multiscale image analysis. The object of analysis is an abstract PO of the conical roller type made of St20kp steel, which meets the standards of DSTU 7809 and DSTU 7808. PO has a specific geometry: different ends and a blind hole. The mass of PO is 5.5 kg. This example is described in detail in the literature [1].

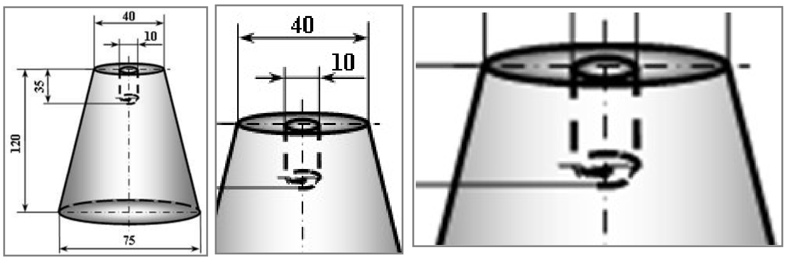

Fig. 5. An example of multi-scale image analysis of an abstract production object of the conical roller type

Consistent refinement of the signal using multi-scale analysis makes it possible to reveal its local features, in particular, to determine the dynamics of changes in the signal and their intensity. The subsequent clustering of the data, based on multi-scale analysis, makes it possible to detect anomalies because anomalies are usually characterized by sharp jumps or deviations of values at the micro or macro levels.

The data is then automatically analyzed by a specially designed CVAE, which detects those containing anomalies. To this end, the designed CVAE is pre-trained on reliable data without anomalies. These data are contained in a previously built expanded database, which contains clear criteria for the values of structural features and physical and mechanical properties of PO, technological characteristics and functional capabilities of PE and TE under normal conditions. In automatic data analysis, the designed CVAE takes a training sample of valid data without anomalies and reconstructs it.

The basis of the designed CVAE consists of several serially connected multilayer NNs, one of which is an encoder, the other a decoder (Fig. 6). The encoder is a constructed NN with a multilayer perceptron architecture, which performs the function of encoding the *Input* signal. The input signal *Input* is represented in the vector form $X=\{x_1, ..., x_n\}$. The corresponding output code (*Code* signal) is formed at the output of the encoder, which also has the vector form $Y=\{y_1, ..., y_m\}$ and is fed to the input of the decoder (Fig. 6). The decoder is also a built NN with a multilayer perceptron architecture. The decoder performs the function of decoding the vector Y of the *code* output signal from the encoder into the *Output* signal, which has the vector form $X=\{x_1, ..., x_n\}$ corresponding to the input signal *Input* (Fig. 6). The number of input and output neurons of the proposed CVAE is the same.

The weights w of artificial neurons of the proposed CVAE are adjusted in the range of values $w\in[0; 1]$ so that the data recovery error L_{Eam} is minimal $L_{Eam}\rightarrow\min$. L_{Eam} is determined on the basis of the mean squared error E_{am} and is calculated according to formula (10) when training NNs, which form the basis of the proposed CVAE, using the Back Propagation algorithm:

$$L_{E_{am}} = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - X_i'\right)^2, \tag{10}$$

where n is the size of the analyzed data set, X_i – vector representation of the true value of the i-th parameter, X_i' – representation of the restored CVAE value of the i-th parameter.

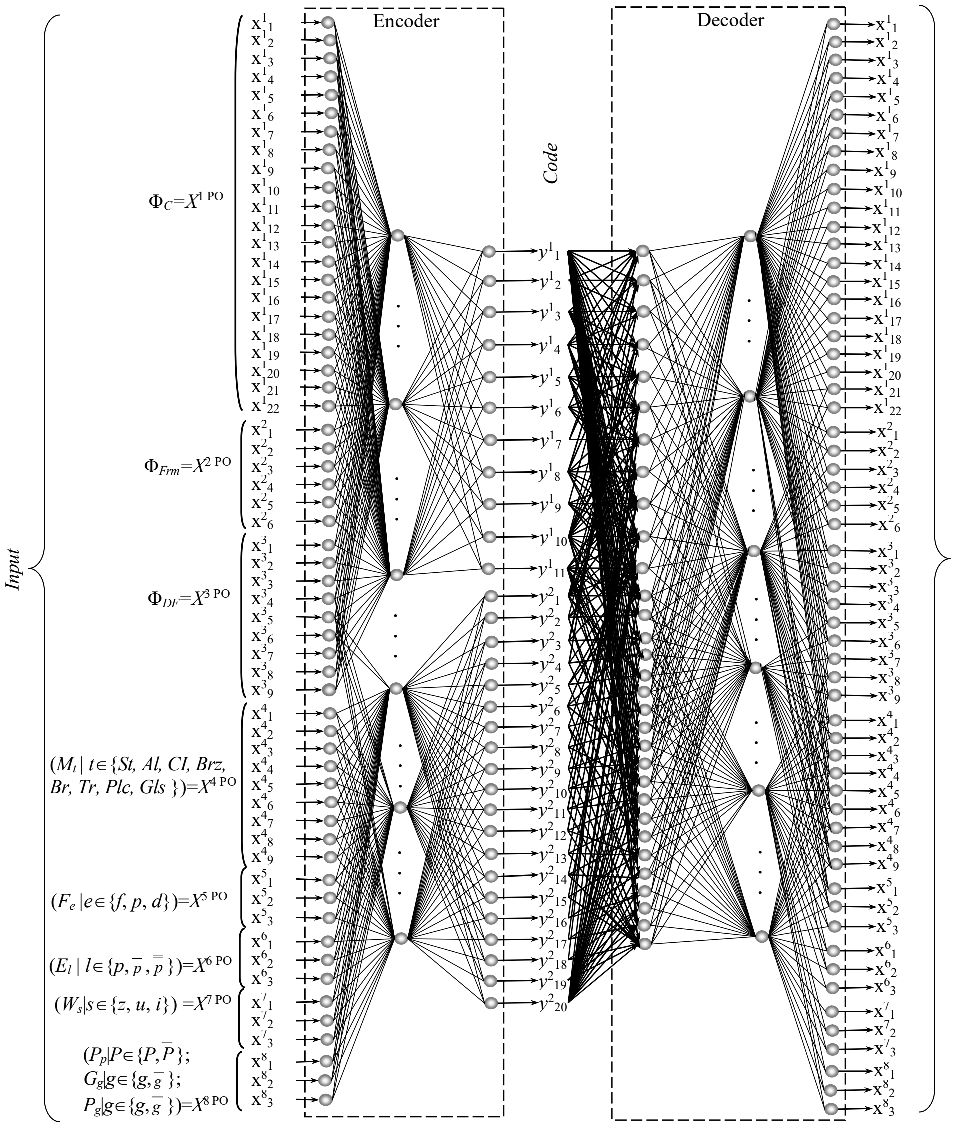
Input
$\Phi_C=X^{1\ PO}$
x^1_1 … x^1_{22}
$\Phi_{Frm}=X^{2\ PO}$
x^2_1 … x^2_6
$\Phi_{DF}=X^{3\ PO}$
x^3_1 … x^3_9
$(M_t \mid t\in\{St, Al, CI, Brz, Br, Tr, Plc, Gls\})=X^{4\ PO}$
x^4_1 … x^4_9
$(F_e \mid e\in\{f, p, d\})=X^{5\ PO}$
x^5_1 … x^5_3
$(E_l \mid l\in\{p, \overline{p}, \overline{\overline{p}}\})=X^{6\ PO}$
x^6_1 … x^6_3
$(W_s \mid s\in\{z, u, i\})=X^{7\ PO}$
x^7_1 … x^7_3
$(P_p \mid P\in\{P, \overline{P}\}; G_g \mid g\in\{g, \overline{g}\}; P_g \mid g\in\{g, \overline{g}\})=X^{8\ PO}$
x^8_1 … x^8_3
Encoder
Code
y^1_1 … y^1_{11}
y^2_1 … y^2_{20}
Decoder
x^1_1 … x^8_3
Output

Fig. 6. Example of the designed CVAE for automatic anomaly analysis of production object data

The principle of anomalous data detection is that when the proposed CVAE encounters anomalous data characterized by sharp jumps and deviations in signal values, an error L_{Eam} will occur, which will significantly exceed some threshold value ε: $L_{mse}>>\varepsilon$.

5. 3. Training and experimental studies of Conditional Variational Autoencoder

A training sample $OV_i|i=\overline{1;60}$ of 60 examples was built for CVAE training. Examples for the formation of the training sample were chosen arbitrarily from the known set of mechanical and instrument engineering POs with various structural features, from St20kp steel according to DSTU 7809 and DSTU 7808. Data on the physical and mechanical properties and design features of PO in the training sample $OV_i|i=\overline{1;60}$ were formalized according to expression (1), reduced to the same size, vector form, and converted into digital signals. Also, information on the physical and mechanical properties and structural features of POs were pre-processed using wavelet transformation. As a result, a set of training digital signals *Input* without anomalies is obtained.

In addition, digital signals about the physical and mechanical properties and structural features of each PO were arranged in rows of 58 elements. Each element of the row corresponds to a certain physical and mechanical property and constructive feature of the formalized description of a certain PO from the training sample $OV_i|i=\overline{1;60}$. As a result, the size of the training sample is 60×58=3,480 digits in the range of values [0; 1]. Fig. 7 shows an example of a fragment of a digital signal *Input* of a training sample without anomalies. This signal is a curve that reflects the level of the signal about the presence of certain physical and mechanical properties and design features of PO1 and PO2, which were arbitrarily selected from the set of POs that formed the training sample.

To test the CVAE encoder, a test sample containing anomalous and non-anomalous *Input* signals about the physical and mechanical properties and design features of a set of POs was previously built. An example of *Input* signals with anomalies for PO1 and PO2, which were arbitrarily selected from the training set of POs, is shown in Fig. 8.

At the output of the CVAE encoder, a set of *Code* signals was received, which are a set of curves that reflect the dynamics and intensity of changes in the *Input* signals depending on the presence or absence of anomalies (Fig. 9, 10). Thus, in the case of anomalies in the *Input* signal (Fig. 8), the *Code* curve is characterized by sharp jumps and deviations of values (Fig. 10) compared to the main data distribution.

Based on the statistical set of CVAE training results, as well as the value of data recovery error L_{Eam}, a threshold value of ε=0.15 of data recovery error L_{Eam} was derived. Signals for which the data recovery error L_{Eam} significantly exceeds the threshold value $L_{Eam}>>\varepsilon=0.15$ are considered to contain anomalies.

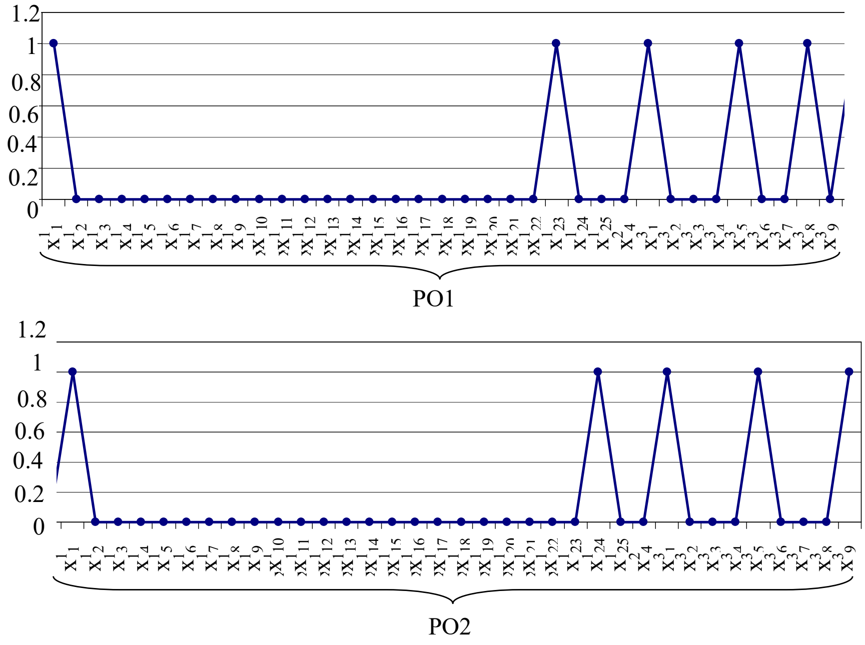

Fig. 7. Example of *Input* signals without anomalies for PO1 and PO2 arbitrarily selected from the training sample

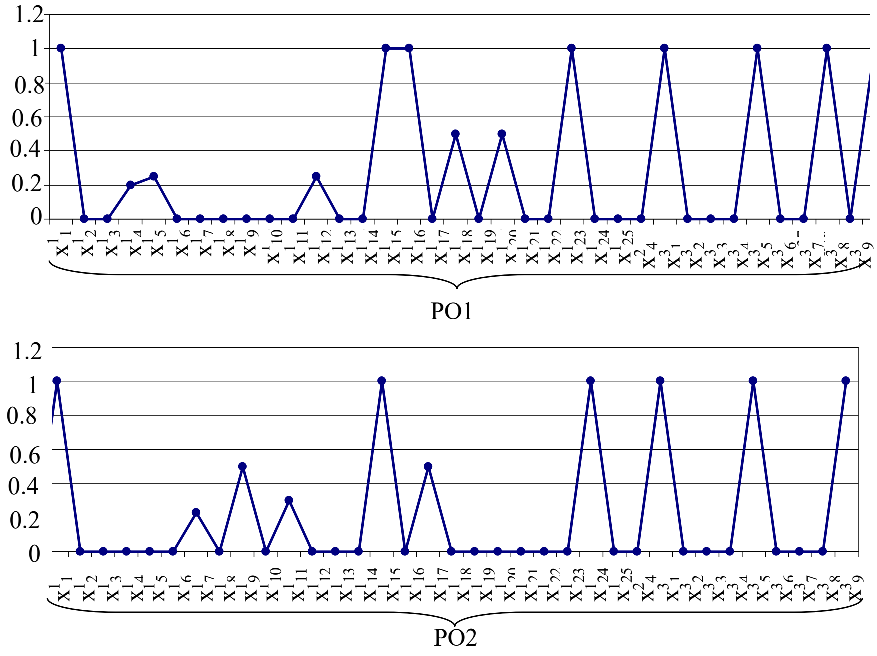

Fig. 8. Example of *Input* signals with anomalies for PO1 and PO2 arbitrarily selected from the training sample

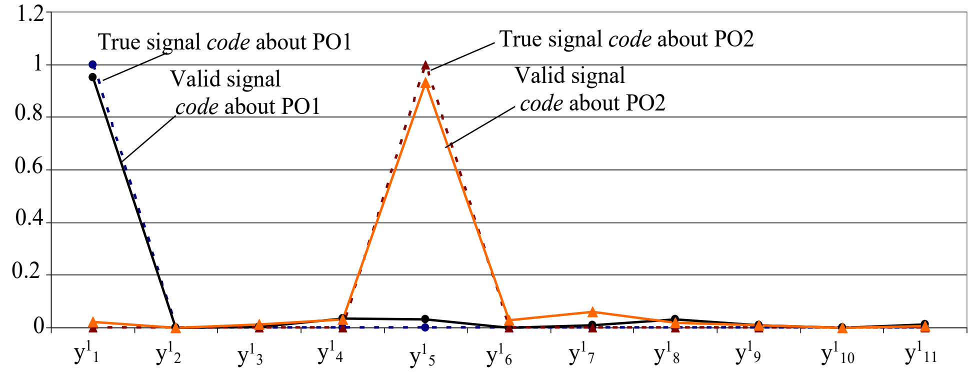

Fig. 9. Example of *Code* signals obtained on the basis of *Input* signals without anomalies based on Fig. 7

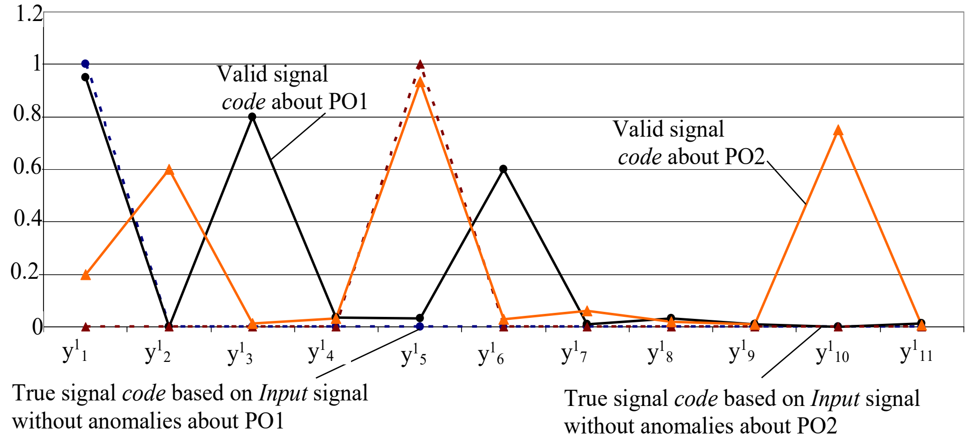

Fig. 10. Example of *Code* signals obtained on the basis of *Input* signals with anomalies based on Fig. 8

Plots in Fig. 9, 10 indicate the effectiveness of CVAE training. When anomalous PO data was fed to the CVAE input, the encoder was able to detect five areas of anomaly and identify the PO data as anomalous. Therefore, the use of the designed CVAE makes it possible to accurately identify data that have deviations from data under normal conditions. This is the basis for establishing the fact of cyber-attacks or technical failures.

5. 4. Algorithm of the intelligent module for detecting anomalies and threats to information security

ICS for automated simulation of SAOPO is a space- and time-distributed dynamic system that functions as part of space- and time-distributed dynamic macro- and hyper-systems – FPS and GKIS. Under such conditions, the process of identifying signs of threats to information security and the emergence of unreliable and/or anomalous data should be continuous and cyclical. From these positions, the cyclical nature of the intelligent module's functioning algorithm for detecting signs of a threat to information security and the emergence of unreliable and/or anomalous data is obvious. The generalized block diagram of the algorithm that reproduces the functioning of the intelligent module for detecting signs of threats to information security and the emergence of unreliable and/or anomalous data is shown in Fig. 11. Its implementation involves 23 steps in the corresponding blocks. The beginning of the algorithm – block 1. In blocks 2–4, monitoring of the information environment is performed. In particular, block 2 identifies possible threats that may arise due to:

– defects, failures, accidents of TE, technical and software means of information processing;

– electromagnetic radiation, etc.;

– unauthorized access to confidential information for the purpose of manipulation;

– computer viruses;

– errors of employees when working with ICS for automated simulation of SAOPO;

– interception of data during their transmission over information networks.

Block 3 identifies the sources of potential threats to information security, which can be various types of hackers, criminals seeking enrichment by obtaining and selling trade secrets, and industrial spies.

In block 4, content monitoring is performed with the active use of the semantic core formed from symbols, words, and phrases that determine the subject and key tasks of SAOPO modeling tasks.

In blocks 5–15, the level of threats is evaluated. In particular, block 5 performs indexing of symbolic content and detection of duplicates. Also, block 5 calculates risk levels (probabilities) of adverse events or threats. Risk levels are calculated by one factor $Th_{i,j,k}$, by three possible factors $Th_{i,j,k\Sigma}$, for resource Th_R, risk for resource R from expressions (1) to (4), respectively. In addition, the amount of costs (losses) Z from the occurrence of an adverse event is estimated from expression (5) and the value of the integral indicator of the level of threats Th (6) is calculated.

In blocks 6–15, based on the received integral indicator of the level of threats Th, a corresponding conclusion is formed according to Table 1.

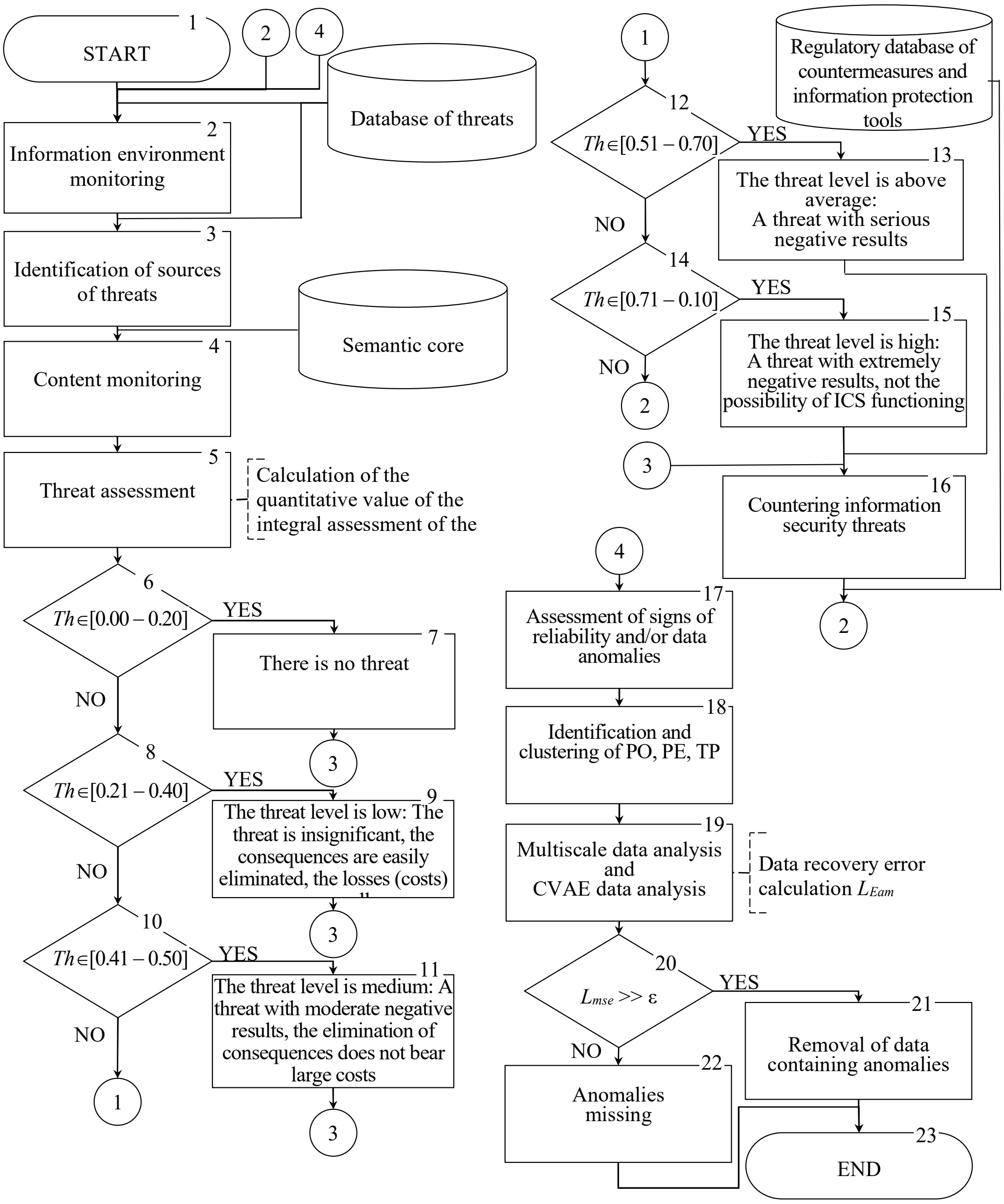
1
START
2
4
Database of threats
2
Information environment monitoring
3
Identification of sources of threats
Semantic core
4
Content monitoring
5
Threat assessment
Calculation of the quantitative value of the integral assessment of the
6
$Th \in [0.00 - 0.20]$
YES
7
There is no threat
NO
3
8
$Th \in [0.21 - 0.40]$
YES
9
The threat level is low: The threat is insignificant, the consequences are easily eliminated, the losses (costs)
NO
3
10
$Th \in [0.41 - 0.50]$
YES
11
The threat level is medium: A threat with moderate negative results, the elimination of consequences does not bear large costs
NO
1
3
1
Regulatory database of countermeasures and information protection tools
12
$Th \in [0.51 - 0.70]$
YES
13
The threat level is above average: A threat with serious negative results
NO
14
$Th \in [0.71 - 0.10]$
YES
15
The threat level is high: A threat with extremely negative results, not the possibility of ICS functioning
NO
2
3
16
Countering information security threats
4
2
17
Assessment of signs of reliability and/or data anomalies
18
Identification and clustering of PO, PE, TP
19
Multiscale data analysis and CVAE data analysis
Data recovery error calculation L_{Eam}
20
$L_{mse} >> \varepsilon$
YES
21
Removal of data containing anomalies
NO
22
Anomalies missing
23
END

Fig. 11. Generalized block diagram of the functioning algorithm of the intelligent module for detecting signs of a threat to information security and the emergence of unreliable and/or anomalous data

Block 16 defines the appropriate tool and/or countermeasure to be applied to protect information.

In blocks 17–23, data is automatically evaluated using machine learning methods. In particular, a specially designed CVAE is used. CVAE makes it possible to identify and cluster PO, PE, and TP (block 17). Multiscale data analysis is performed in block 18, and anomaly detection in block 19 and block 20. Blocks 21 and 22 decide what to do with the data. Thus, on the basis of the received data, a decision is made to remove data containing anomalies (block 21) or to leave them if there are no anomalies (block 22). Block 23 is the end of the algorithm.

6. Discussion of results based on the proposed intellectual module

The proposed intelligent module for detecting signs of threats to information security and the emergence of unreliable and/or anomalous data is the state-of-the-art, original, and effective tool. It, in contrast to known solutions [2–15], is used at the stage of technological preparation of production during the automated simulation of SAOPO as part of known ICS. In addition, unlike solutions in [2–6], the application of the proposed intelligent module makes it possible to obtain social and economic effects. These effects are manifested in increasing the level of protection of important production data, preventing losses from their damage or loss. The material, intellectual, and time costs of restoring information are reduced. Risks associated with non-fulfillment of obligations and lawsuits are reduced. The competitiveness of the enterprise is increasing.

Unlike solutions in [2–4], the proposed intelligent module contains CVAE, which is automatically trained on a previously formed set of examples. At the same time, considering the production necessity, CVAE is relatively easy to train and retrain to solve new tasks without changing the structure and without involving new equipment. After all, the problem of detecting unreliable and/or anomalous data from a mathematical point of view was stated as a classic problem of classification. At the same time, the machine learning used to solve it provided high accuracy. Thus, five anomalies on the test sample were detected with an accuracy of 97.53 %.

Unlike solutions from [2–15], the proposed intelligent module uses a combined approach based on machine learning, classic algorithms of regression and comparative analysis. This obviously makes it possible to expand the functionality of the proposed module and makes it possible to solve two tasks. The first task is to identify signs of threats to the information security of ICS of the automated simulation of SAOPO and apply appropriate countermeasures in a timely manner. The second is to identify unreliable and/or anomalous data on the components of SAOPO, in particular PO, PE, TE, and TP, and remove them at the stage of technological preparation of production.

Our results for each task are explained as follows.

The designed PDCA model of the functioning of an intelligent module for detecting signs of a threat to information security and the emergence of unreliable and/or anomalous data (Fig. 3) clearly reflects the life cycle and the relationship of all information security management processes. Thanks to this, the essence of the processes of assessing the level of threats and analyzing data to determine their unreliability and/or abnormality, as well as the application of appropriate countermeasures against threats and the extraction of anomalous data, has been established. In general, the constructed PDCA model made it possible to form the stages of information security management and determine their content.

The developed structural diagram (Fig. 4) of the newly designed intelligent module offers a complex sequence of actions aimed at identifying signs of a threat to information security and the emergence of unreliable and/or anomalous data. This is realized by the combined application of regression and comparative analysis methods and machine learning. These methods are used at the stages of content monitoring, assessment of signs of threats to information security, unreliability, and/or data anomalies.

Thus, machine learning methods, in particular specially designed CVAE (Fig. 6), provide an opportunity to automatically analyze the reliability of information about the components of SAOPO with high accuracy in real time. Experimental studies have shown that the accuracy of detecting anomalies on the test sample is 97.53 %, which meets the requirements of modern machine and instrument manufacturing.

The methods of regression and comparative analysis make it possible to quantitatively estimate the probability of occurrence of adverse events, that is, threats and the amount of possible costs according to expressions (3) to (8). Thanks to this, even at the stage of technological preparation of production, the values of possible costs according to expression (7) and Table 1 from the occurrence of threats during the operation of ICS for automated simulation of SAOPO are estimated. Also, owing to this, appropriate countermeasures are applied in time and losses are prevented.

Training and experimental studies of the designed CVAE were carried out on the example of CVAE for automatic analysis of anomalies of data on PO. Training of the proposed CVAE was conducted on an arbitrarily formed training sample. The examples for the training sample are formalized descriptions according to expression (1) of a set of machine and instrument engineering POs with different structural features, made of St20kp steel according to DSTU 7809 and DSTU 7808. In addition, the examples of the training sample are reduced to the same size, vector form, and converted into digital signals (Fig. 7, 8). The results (Fig. 9, 10) made it possible to verify not only the functionality of the proposed CVAE but also to determine high accuracy. Thus, five anomalies on the test sample were detected by the proposed CVAE with an accuracy of 97.53 %, which meets the accuracy requirements of modern machine and instrument manufacturing.

The developed algorithm for the functioning of the intelligent module reproduces an ordered sequence of actions to detect signs of a threat to information security and the emergence of unreliable and/or anomalous data. Decomposition of this process was carried out and local tasks were determined, which are the main tasks of the newly designed intelligent module. Owing to the representation of the algorithm in the form of a block diagram (Fig. 11), the steps and the necessary sequence of actions for assessing information security are clearly displayed. This makes it possible to improve the perception of the process of detecting signs of threats and the detection and extraction of abnormal production data, to apply appropriate countermeasures in a timely manner.

In addition, the limitations of this research when trying to apply it in practice may be an insufficient level of knowledge and skills in working with machine learning methods and artificial intelligence technologies. They can cause difficulties for developers along the way.

The disadvantage of this study is that it is focused only on the protection and analysis of production data when modeling SAOPO while it is necessary to detect signs of threats to information security and deviations from the norm of data when solving other tasks of technological preparation of production.

Further development of this research can be carried out in several directions. First of all, the need to expand the functionality of the proposed intellectual in the direction of detecting signs of threats to information security and deviation from the data norm when solving other tasks of technological preparation of production is obvious. Secondly, there is a need to develop appropriate original software with a convenient and intuitive interface for the newly designed intelligent module.

7. Conclusions

1. A PDCA model and a structural diagram of an intelligent module for detecting signs of a threat to information security and the emergence of unreliable and/or anomalous data, which is integrated into ICS for automated simulation of SAOPO, have been developed. The PDCA model clearly displays the life cycle and interrelationship of all processes for identifying signs of a threat to information security and the emergence of unreliable and/or anomalous data. The constructed structural scheme of the newly designed intelligent module reproduces a complex sequence of actions aimed at identifying signs of a threat to information security and the emergence of unreliable and/or anomalous data. In addition, the structural diagram reflects information flows,

connections between them, defined methods and means of automated data processing to determine anomalies and identify signs of a threat to information security. It is also planned to integrate the newly designed intelligent module into known ICS for automated simulation of SAOPO.

2. Methods have been defined and means have been designed for the automated detection of signs of a threat to information security and the emergence of unreliable and/ or anomalous data. Methods of regression and comparative analysis were used to identify signs of a threat to information security. These methods are the basis for automating the quantitative assessment of the probability of occurrence of adverse events or threats, the amount of possible costs even at the stage of technological preparation of production. Original CVAEs have been designed for automatic analysis of production data for anomalies.

3. Training and experimental studies of the designed CVAE were carried out on the example of CVAE automatic analysis of anomalies of data on PO. Our results confirmed the feasibility of the designed CVAE, the ability to analyze a set of production data and detect anomalies in real time with a high accuracy of 97.53 %. The results of experimental research reported in this work allow us to consider the proposed intelligent module as the newest, original, and effective tool for detecting signs of a threat to information security and the emergence of unreliable and/or anomalous data. It is obvious that the intelligent module can be recommended for practical use as part of known ICS for automated simulation of SAOPO at the stage of technological preparation of production.

4. An algorithm for the functioning of the intelligent module has been developed, which reproduces an ordered sequence of actions for solving local problems of identifying signs of a threat to information security and the emergence of unreliable and/or anomalous data. This algorithm can be recommended for practical application in the development of appropriate original software with a convenient and intuitive interface in further scientific research.

Conflicts of interest

The authors declare that they have no conflicts of interest in relation to the current study, including financial, personal, authorship, or any other, that could affect the study, as well as the results reported in this paper.

Funding

The study was conducted without financial support.

Data availability

All data are available, either in numerical or graphical form, in the main text of the manuscript.

Use of artificial intelligence

The authors confirm that they did not use artificial intelligence technologies when creating the current work.

References

1. Cherepanska, I., Sazonov, A., Melnychuk, P., Melnychuk, D., Kalchuk, S., Pryadko, V., Yanovsky, V. (2024). Design of an information-computer system for the automated modeling of systems for automatic orientation of production objects in the machine and instrument industries. Eastern-European Journal of Enterprise Technologies, 3 (2 (129)), 6–19. https://doi.org/10.15587/1729-4061.2024.306516
2. Gao, Y., Yin, X., He, Z., Wang, X. (2023). A deep learning process anomaly detection approach with representative latent features for low discriminative and insufficient abnormal data. Computers & Industrial Engineering, 176, 108936. https://doi.org/10.1016/j.cie.2022.108936
3. Aschepkov, V. (2024). The use of the Isolation Forest model for anomaly detection in measurement data. Innovative technologies and scientific solutions for industries, 1 (27), 236–245. https://doi.org/10.30837/itssi.2024.27.236
4. Vos, K., Peng, Z., Jenkins, C., Shahriar, M. R., Borghesani, P., Wang, W. (2022). Vibration-based anomaly detection using LSTM/SVM approaches. Mechanical Systems and Signal Processing, 169, 108752. https://doi.org/10.1016/j.ymssp.2021.108752
5. Huang, X., Wen, G., Dong, S., Zhou, H., Lei, Z., Zhang, Z., Chen, X. (2021). Memory Residual Regression Autoencoder for Bearing Fault Detection. IEEE Transactions on Instrumentation and Measurement, 70, 1–12. https://doi.org/10.1109/tim.2021.3072131
6. Panza, M. A., Pota, M., Esposito, M. (2023). Anomaly Detection Methods for Industrial Applications: A Comparative Study. Electronics, 12 (18), 3971. https://doi.org/10.3390/electronics12183971
7. Mokhtari, S., Abbaspour, A., Yen, K. K., Sargolzaei, A. (2021). A Machine Learning Approach for Anomaly Detection in Industrial Control Systems Based on Measurement Data. Electronics, 10 (4), 407. https://doi.org/10.3390/electronics10040407
8. Zipfel, J., Verworner, F., Fischer, M., Wieland, U., Kraus, M., Zschech, P. (2023). Anomaly detection for industrial quality assurance: A comparative evaluation of unsupervised deep learning models. Computers & Industrial Engineering, 177, 109045. https://doi.org/10.1016/j.cie.2023.109045
9. Jaramillo-Alcazar, A., Govea, J., Villegas-Ch, W. (2023). Anomaly Detection in a Smart Industrial Machinery Plant Using IoT and Machine Learning. Sensors, 23 (19), 8286. https://doi.org/10.3390/s23198286
10. Tang, M., Chen, W., Yang, W. (2022). Anomaly detection of industrial state quantity time-Series data based on correlation and long short-term memory. Connection Science, 34 (1), 2048–2065. https://doi.org/10.1080/09540091.2022.2092594
11. Evangelou, M., Adams, N. M. (2020). An anomaly detection framework for cyber-security data. Computers & Security, 97, 101941. https://doi.org/10.1016/j.cose.2020.101941

12. Ameer, S., Gupta, M., Bhatt, S., Sandhu, R. (2022). BlueSky. Proceedings of the 27th ACM on Symposium on Access Control Models and Technologies, 235–244. https://doi.org/10.1145/3532105.3535020
13. Szymanski, T. H. (2022). The “Cyber Security via Determinism” Paradigm for a Quantum Safe Zero Trust Deterministic Internet of Things (IoT). IEEE Access, 10, 45893–45930. https://doi.org/10.1109/access.2022.3169137
14. Liu, R., Shi, J., Chen, X., Lu, C. (2024). Network anomaly detection and security defense technology based on machine learning: A review. Computers and Electrical Engineering, 119, 109581. https://doi.org/10.1016/j.compeleceng.2024.109581
15. Das, T. K., Adepu, S., Zhou, J. (2020). Anomaly detection in Industrial Control Systems using Logical Analysis of Data. Computers & Security, 96, 101935. https://doi.org/10.1016/j.cose.2020.101935
16. Patel, P., Deshpande, V. (2017). Application Of Plan-Do-Check-Act Cycle For Quality And Productivity Improvement-A Review. International Journal for Research in Applied Science & Engineering Technology, 5 (1), 197–201. Available at: https://www.researchgate.net/publication/318743952_Application_Of_Plan-Do-Check-Act_Cycle_For_Quality_And_Productivity_Improvement-A_Review
17. Molodetska-Hrynchuk, K. (2017). The model of decision making support system for detection and assessment of the state information security threat of social networking services. Ukrainian Scientific Journal of Information Security, 23 (2). https://doi.org/10.18372/2225-5036.23.11803
18. Gong, X., Yu, S., Xu, J., Qiao, A., Han, H. (2023). The effect of PDCA cycle strategy on pupils’ tangible programming skills and reflective thinking. Education and Information Technologies, 29 (5), 6383–6405. https://doi.org/10.1007/s10639-023-12037-4
19. Cherepanska, I., Sazonov, A., Melnychuk, D., Melnychuk, P., Khazanovych, Y. (2023). Quaternion Model of Workpieces Orienting Movements in Manufacturing Engineering and Tool Production. Lecture Notes in Mechanical Engineering, 127–135. https://doi.org/10.1007/978-3-031-42778-7_12
20. Voronin, A. N. (2009). Nelineynaya skhema kompromissov v mnogokriterial’nyh zadachah otsenivaniya i optimizatsii. Kibernetika i sistemniy analiz, 45 (4), 106–114. Available at: http://nbuv.gov.ua/UJRN/KSA_2009_45_4_10
21. Nykolyuk, O. M., Martynchuk, V. (2018). A Methodology for Assessing Resource Potential of Innovation-Oriented Agricultural Enterprises. Problemy Ekonomiky, 1 (35), 207–213. Available at: https://www.proquest.com/openview/1716ad4663e51395c99da80118e1204e/1?pq-origsite=gscholar&cbl=2048964