# Data Processing Centre's Cyberattack Protection Directions on the Base of Neural Network Algorithms

Yanina Shestak [1], Serhii Toliupa [1], Anatolii Shevchenko [1], Anna Torchylo [1] and Ogbu James Onyigwang [2]

[1] *Taras Shevchenko National University of Kyiv, 24 B. Havrylyshyna Str., Kyiv, 04116, Ukraine*
[2] *University of Ibadan, Ibadan 200284 Oyo State, Nigeria*

**Abstract**

This paper describes the methods of organization of the data center protection strategy, presented as a network distributed infrastructure, against potential external threats. This work indicates the advantages of using neural network algorithms and deep learning neural network architecture in the specified field. In accordance with the set of quantitative target indicators, mathematical modeling of the evaluation of the effectiveness of the selection of cyber attack software code was carried out. Based on the proposed mathematical apparatus, an evaluation of the protection of the infrastructure of the data center against cyber attacks was carried out. In particular, this article analyses using a neural network architecture such as an autoencoder, a multi-layer autoencoder, a deep belief network, a convolutional neural network, a recurrent neural network, a recursive neural network with the inclusion of algorithms based on a restricted Boltzmann machine and a long-chain scheme of short-term memory. According to a set of factors that correspond to the effectiveness of the application of neural network algorithms in solving the task of organizing a data center infrastructure protection strategy, objective functions were proposed. Besides, the determination of global extrema of these functions provides an opportunity to solve the problem of optimizing the machine code analysis system for the presence of a cyber attack.

**Keywords** [1]

Data center, cyber attack, multi-layer autoencoder, deep belief network, convolutional neural network, recurrent neural network, recursive neural network.



## 1. Introduction

The organization of data processing centers based on a Distributed Information System (DIS) provides an opportunity to significantly expand the functionality of the specified network services and increase the flexibility of the corresponding architecture depending on the requirements for optimization, reorganization and scaling of the general infrastructure, which determines the prevalence of the specified approach today. However, it should be noted that at the same time, the toolkit of a potential attacker is also expanding, which can be used in the implementation of unauthorized access to the service of the data processing center, with the subsequent task of significant material and reputational damage to the owners of the service [1 5]. This indicates the high urgency of solving the task of developing a holistic methodology for protecting network services from external threats, in accordance with the concept of a Security Information and Event Management System (SIEM), the generalized scheme of which is presented in fig. 1. This technology supports threat detection, compliance and security incident management through the collection and analysis (both near real time and historical) of security events, as well as a wide variety of other event and contextual data sources. The development and optimization of SIEM system architecture is a complex task, the solution of which

*Information Technology and Implementation (IT&I-2022), November 30 - December 02, 2022, Kyiv, Ukraine*
EMAIL: yaninashestak@gmail.com (A. 1); tolupa@i.ua (A. 2); atorouss@gmail.com (A. 3); jamesisaac2000@hotmail.com (A. 4)
ORCID: 0000-0002-1703-0316 (A.1); 0000-0002-1919-9174 (A. 2); 0000-0002-9100-6939 (A. 3)

includes the definition of the following key components, according to which the following groups of target indicators can be obtained:

1. Peculiarities of identifying signs of a cyber attack by the SIEM system: typical samples from the library of the training set or high-level signs.
2. The object of the cyber attack: the hardware platform of the SIEM complex, network protocols of the service, the operating system, software applications and blocks of customer data stored on the service platform.
3. The purpose of the cyber attack: unauthorized access, illegal copying of data or making changes, disruption of the stable operation of the information network, external control by an unauthorized user.
4. The method of monitoring the actions of a potential attacker using the SIEM system: machine analysis of software code samples, the order of execution of procedures, the life cycle of a cyber attack.

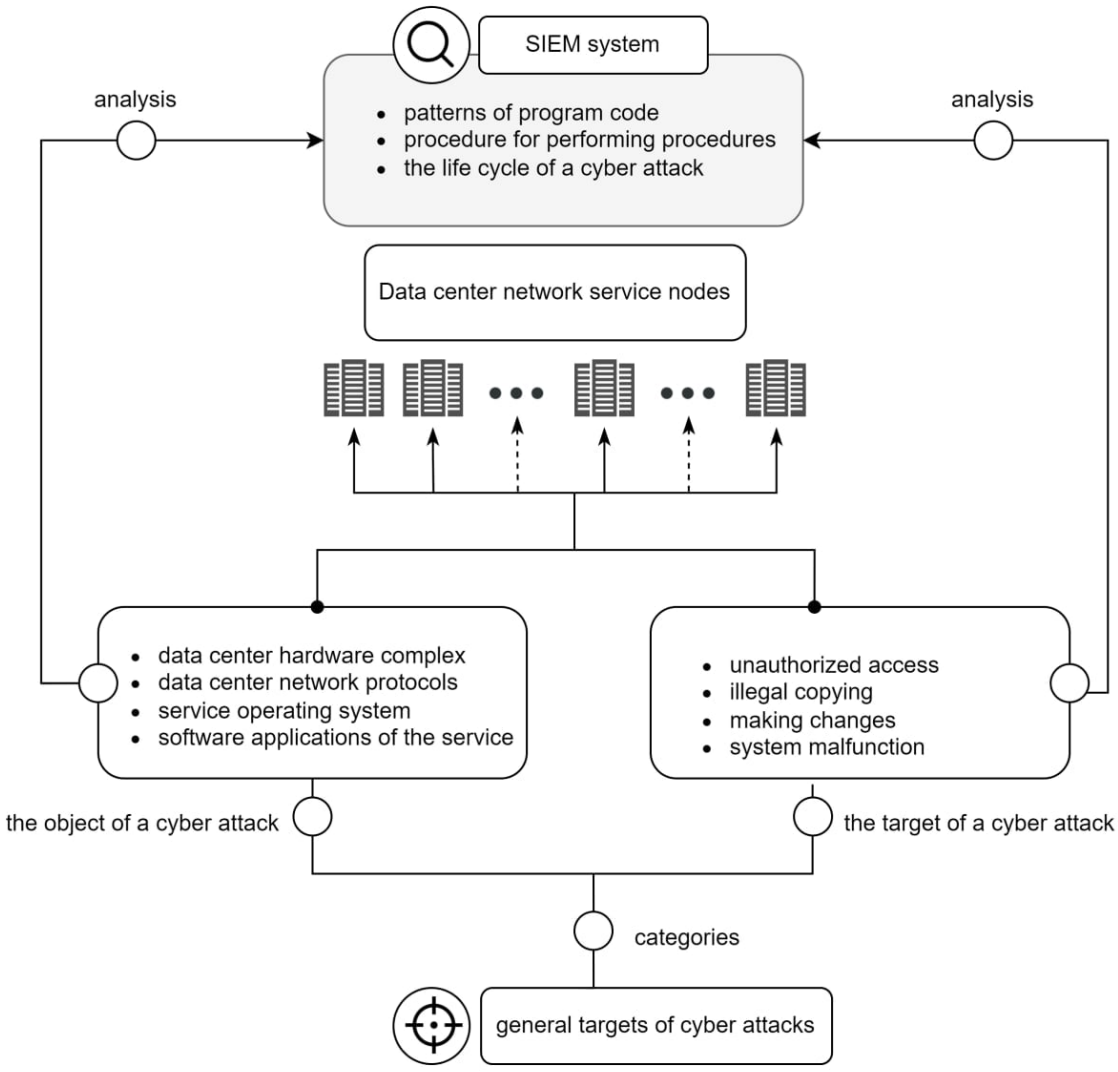


**Figure 1**: Scheme of detection of external threats at the level of SIEM system

The identification of features of the organization of software code samples, behavior and the life cycle of a cyber attack, both at the level of typical components and at the level of high-level features, is most effectively implemented through the use of machine analysis based on neural network algorithms [2-11], and deep learning neural network algorithms (DL-ANN: Deep Learning Artificial Neural Networks) in particular.

Within the framework of this study, an analysis of relevant scientific publications was conducted and it was noted that neural network algorithms that can be used in the construction of algorithms for machine analysis of software code samples in order to identify signs of cyber attacks should be divided into the following groups according to the organization of the architecture:

- a neural network of the autoencoder type, which can be extended to a multilevel autoencoder type architecture for the selection of high-level features [2, 3] which can be extended to a neural network architecture of deep learning such as a multilayer autoencoder;
- deep belief neural networks (DBN), considered as generative graph models [4, 5];
- recurrent neural networks (RNN), on the basis of which machine analysis with sets of event sequences is effectively carried out [6, 7];
- convolutional neural networks (CNN), used to highlight typical code patterns; within the framework of this task, it should be noted that the choice of CNN provides an opportunity to reduce the load on the computing resource of the general complex of machine analysis [8, 9];
- recursive neural networks (RvNN), based on the recursive application of one set of weights to a structured data set [10, 11].

At the same time, it is necessary to build appropriate mathematical models, propose quantitative indicators of the effectiveness of neural network algorithms, determine the extrema of the objective functions, and correlate the obtained results with statistical data in order to assess the performance of neural network algorithms accurately. This is considered as an unresolved part of the general research.

Thus, the aim of this work is to develop a methodology for optimizing SIEM system neural network algorithms, which can be effectively used within the framework of the organization of the distributed network infrastructure scheme in data processing center.

## 2. Principles of adaptation of deep learning neural network architecture for cyber attack detection

The research is based on the construction of an adequate mathematical model of the machine analysis procedure for the purpose of detecting a cyber attack on the DIS infrastructure. Thus, includes the need to calculate such target indicators as the accuracy of the classification of program code patterns and the order of execution of procedures, the total load on the components that determine the computing resource, RAM and information storage of the hardware platform of the corresponding service, as well as time processing of the flow of input data in accordance with the actual task of working in real time (Fig. 2). In order to justify the costs of the event collection and correlation system, it is necessary that the data not only was entered into the consolidated storage for their further analysis by the fact of the incident, but also processed.

It is obvious that the tools of the given system will significantly speed up the incident analysis process. However, the main task of cybersecurity system is timely detection, prompt response and prevention of threats. For this, it is necessary to draw up the rules of correlation by drawing the risks relevant for the company, as well as constant updating of the rules by specialists.

At the same time, deep learning neural network architecture require to learn the attack model from historical threat data and use the trained models to detect intrusions for unknown cyber threats. Thus, machine learning-driven solutions used to detect rare or anomalous patterns can improve detection of new cyber threats and zero-day vulnerabilities.

DLNN solutions collect and correlate alerts, allowing analysts to gain more insight into a security incident or attack and free up more time for more important investigations. Accordingly, these systems analyze large volumes of data coming from multiple sources, monitor suspicious behavior, automatically respond to potential attacks, and eliminate them, detect threats, notify about them, then investigate and remediate them. Robust analytics are critical to understanding threats and make it easy for experts to find threats that might otherwise go unnoticed, and provide visibility into their timeline.

At the same time, the researchers indicate the advantages of using the DL-ANN architecture within the specified task, the adaptation features of which can be defined to the following categories (Fig. 3):

1. Productivity of machine analysis, which includes increasing the accuracy of pattern selection, tools for selecting high-level features, as well as effective work with large data sets, the relevance of which increases with the exponential growth of the bandwidth of information channels and the volume of information storage.
2. Increasing the load on the computing resource, as well as the RAM resource and information storage of the machine analysis complex, which may be unacceptable according to the limitations of the hardware complex.

3. An increase in the time of processing input data, which may be unacceptable in accordance with peak loads when processing input data under real-time operating conditions, as well as the time of learning neural network algorithms on the training sample.

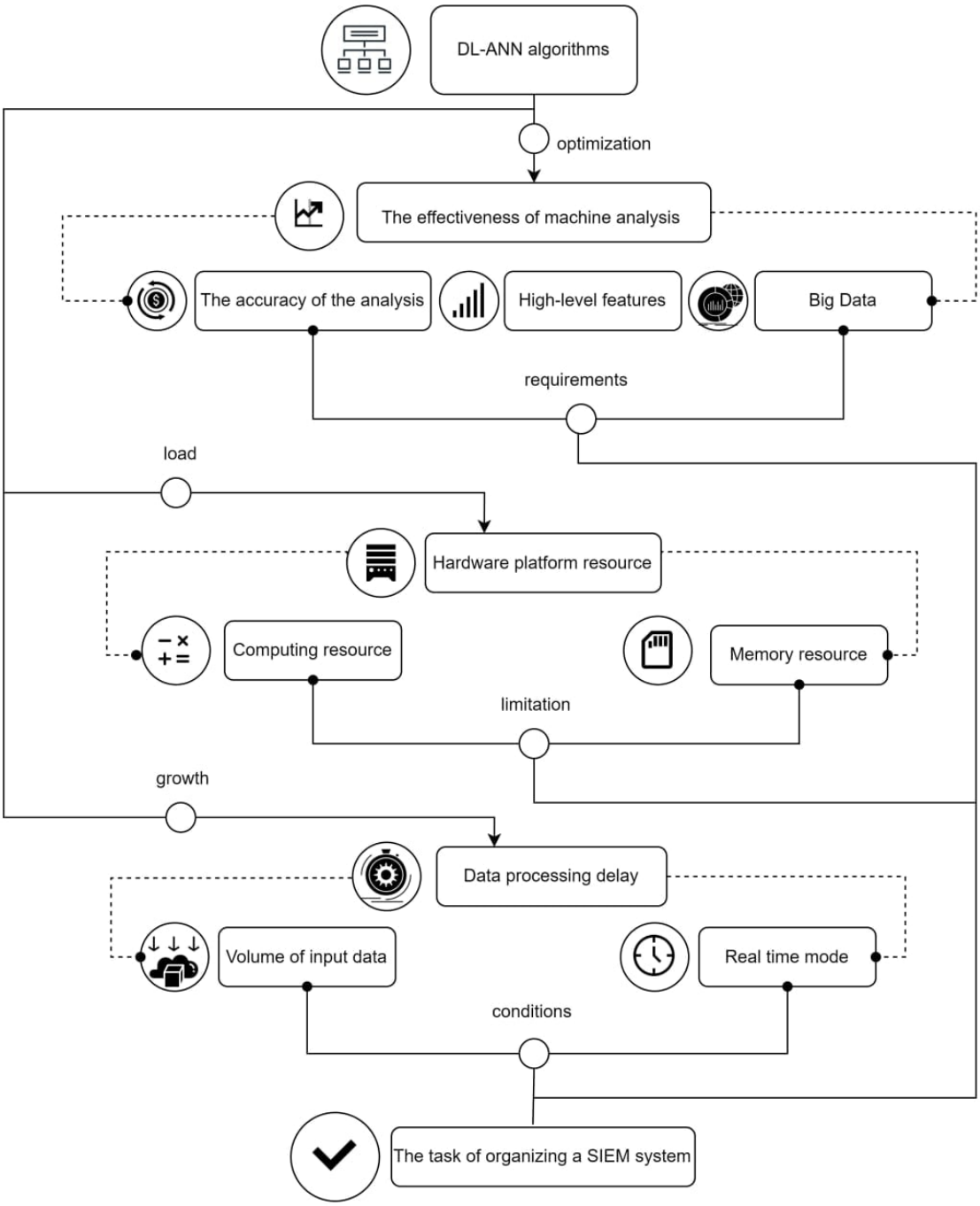


**Figure 2**: Scheme of adaptation of neural network algorithms of deep learning in the organization of the SIEM system of the data center

Thus, the task of optimizing algorithms based on the neural network architecture of deep learning for the detection of cyber attack patterns, within the framework of the study, is solved by determining the global maxima of the objective functions of the accuracy of machine analysis, the global minima of the load on the computing resource and the resource of the hardware platform, as well as the global minimum of the input flow processing time data when working in real-time mode under the conditions of hardware platform resource limitation. In order to determine the quantitative indicators of the accuracy of the identification and classification of cyberattack patterns, according to the statistical

results of the study in relation to the total number of objects of analysis $N_\Sigma$, the following designations are introduced:

- $N_{TP}$ as the number of true positives (TP) results of machine analysis;
- $N_{TN}$ as the number of true negatives (TN) results of machine analysis;
- $N_{FP}$ as the number of false positives (FP) results of machine analysis;
- $N_{FN}$ as the number of false negatives (FN) of machine analysis results.

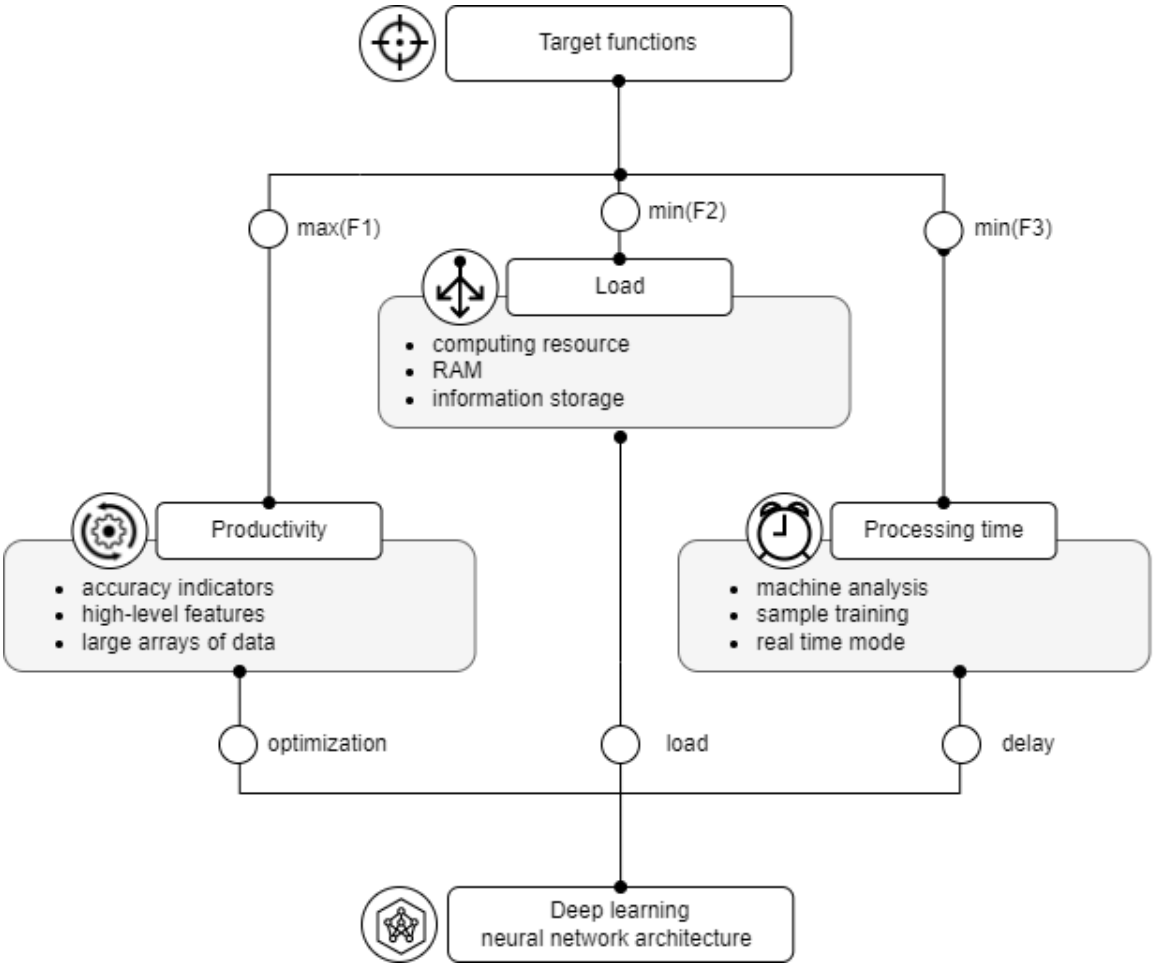


**Figure 3**: Features of adaptation of deep learning neural network architecture for detection of cyber attack patterns

In addition, for the convenience of building a mathematical apparatus, additional statistical indicators can be introduced:

- $N_{\Sigma T}$ and $N_{\Sigma F}$ as the total number of true and false classification results;
- $N_{\Sigma N}$ and $N_{\Sigma P}$ as the total number of negative and positive classification results.

Based on the specified static indicators, the objective functions of the accuracy of program code pattern classification can be calculated as $F_{AL}$(AL: Accuracy Level) and $F_{PL}$(PL: Precision Level):

$$\begin{bmatrix} F_{AL} = \dfrac{N_{\Sigma T}}{N_\Sigma} \\ F_{PL} = \dfrac{N_{TP}}{N_{\Sigma P}} \end{bmatrix}, \text{де } \begin{bmatrix} N_{\Sigma T} = N_{TP} + N_{TN} \\ N_{\Sigma P} = N_{TP} + N_{FP} \end{bmatrix} \quad (1)$$

In accordance with the specified objective functions of the accuracy of classification, an assessment of the performance of the application of neural network algorithms can be carried out while limiting the allowable processing time of the input stream of a fixed data volume for identical computing resources and memory resources of the hardware platform.

## 3. Evaluation of the performance of neural network algorithms according to the target indicators of accuracy of machine analysis

In order to evaluate the performance of the application of neural network algorithms in accordance with the target indicators of the accuracy of machine analysis, it is proposed to conduct research for neural network architectures, which are considered relevant in solving the problem of identifying and classifying cyber attack patterns nowadays. At the same time, it is proposed to compare classical neural network architectures, which are characterized by a minimal load on the hardware resource of the machine analysis system, with neural network architectures of deep learning [4 -7, 10-13] .

At the first stage, based on a set of statistical analysis indicators $\{N_{TP}, N_{TN}, N_{FP}, N_{FN}\}$ presented in studies [2, 3], it is proposed to determine the ranges of values for the above objective functions $\{F_{AL}\}$and $\{F_{PL}\}$for the architecture of the autoencoder and multi-level autoencoder according to equation (1).

The calculation results are shown in fig. 3. It demonstrates that the specified neural network architecture shows mediocre performance values.

When moving from a standard autoencoder (values of the objective functions $F_{AL}^{0}$and $F_{PL}^{0}$) to a deep learning neural network of the multi-layered autoencoder (values of the objective functions $F_{AL}^{+}$and $F_{PL}^{+}$), the accuracy of the analysis does not increase, but the spread of values significantly decreases, which makes it possible to fix the target indicators at the maximum possible level of this model of values. The reliability of the application of the deep learning architecture is based on the comparison of the values obtained at the output of each layer, which corresponds to a separate autoencoder, with the next layer.

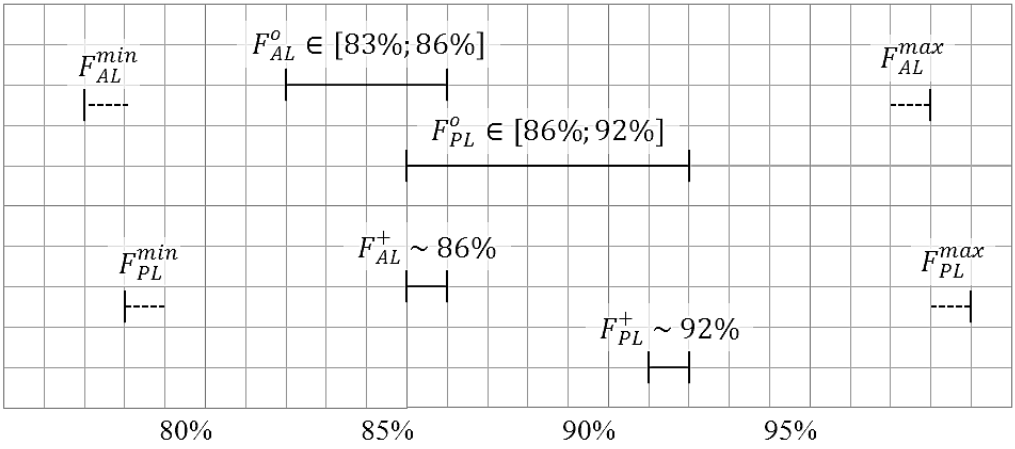


**Figure 4**: Ranges of accuracy for detecting patterns of cyberattacks when using autoencoder and multi-level autoencoder neural networks

In turn, neural network algorithms based on DBN architecture are based on the composition of basic neural networks and classification layers. Within the framework of the SIEM system , the specified approach is used to highlight high-level features of the software code at the level of deobfuscation.

In addition to the basic architecture (values of the objective functions $F_{AL}^{D}$and $F_{PL}^{D}$), the research paper presents modeling results for DBN with a linear regression function (LR: Linear Regression). It is the value of the objective functions $F_{AL}^{LR}$ and $F_{PL}^{LR}$ and the elements of the architecture of the probabilistic neural network (PNN) are the values of the objective functions $F_{AL}^{P}$ and $F_{PL}^{P}$, respectively [12, 13]. As the simulation results show, neural network algorithms based on the DBN architecture with LR layers provide maximum accuracy in both indicators ( $F_{AL}^{LR} \in [97\%, 98\%]$ and $F_{AL}^{LR} \sim 98\%$, respectively) with minimal spread ( $\Delta F_{AL}^{LR} \sim 2\%$and $\Delta F_{PL}^{LR} \sim 1\%$, respectively).

The last group of neural network architectures, which was considered in this paper, includes such architectures as CNNs using a long chain of short-term memory elements (LSTM - Long Short-Term Memory) which includes the next values:

- the value of the objective functions $F_{AL}^{C}$and $F_{PL}^{C}$, RNN , where the problem of short-term memory is also solved through the use of the LSTM scheme ,

- the value of the objective functions $F_{AL}^{C}$and $F_{PL}^{C}$, RvNN with internal memory, which makes it possible to perform machine analysis of code sequences of arbitrary length, which increases performance systems when working with large arrays,
- the values of the objective functions $F_{AL}^{C}$and $F_{PL}^{C}$, as well as neural network algorithms based on the restricted Boltzmann machine ( Restricted Boltzmann Machines , RBM ) are the values of the objective functions $F_{AL}^{RB}$and $F_{PL}^{RB}$.

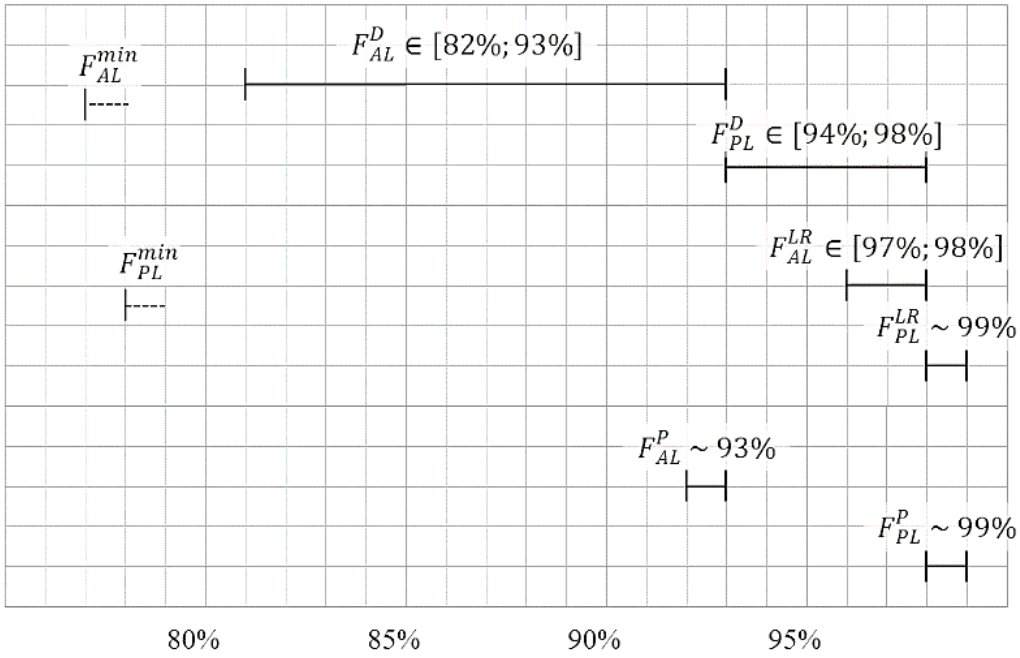


**Figure 5**: Ranges of accuracy for detecting patterns of cyberattacks when applying deep neural networks of beliefs

As the simulation results demonstrates, neural network algorithms based on the RNN architecture with the LSTM scheme provide maximum accuracy in both indicators ( $F_{AL}^{R}$~96%and $F_{AL}^{R}$~96%, respectively) with minimal spread ( $\Delta F_{AL}^{R}$~1%and $\Delta F_{PL}^{R}$~1%, respectively).

The main advantage of the particular scheme is the flexibility of convolutional operators in reducing the number of parameters. As a result, the CNN-LSTM network is becoming deeper. Such networks provide superior performance by simulating signals in temporal information and provide highly efficient threat level detection of suspicious events using a long chain of short-term memory cells. The final classification layer of the CNN-LSTM architecture is a fully connected layer that provides a final decision on the threat level within a certain period of time for each new SIEM instance.

In this way, the proposed approach can be adapted for a wide range of tasks in the field of organization, configuration and optimization of the SIEM scheme through the assessment of the accuracy of machine analysis in the selection and classification of cyber attack patterns according to a specific task (threat level, volume of incoming data flow and available computing resource and the memory resource of the hardware platform of the service).

## 4. A comprehensive technique for evaluating the performance of neural network algorithms

In order to build a universal methodology for evaluating the performance of machine analysis with the aim of further adapting it according to specific tasks, the above target functions of the accuracy of determining and classifying cyber attack patterns should be supplemented with such categories as the ratio of false results and the ratio of correct detection results ( $\kappa_F$and $\kappa_T$, respectively) and the completeness indicator $F_{Rec}$:

$$\begin{cases} \begin{cases} \kappa_F = \dfrac{N_{\Sigma F}}{N_\Sigma} \\ \kappa_T = \dfrac{N_{\Sigma T}}{N_\Sigma} \end{cases} \\ F_{Rec} = \dfrac{N_{TP}}{N_{TP} + N_{FN}} \end{cases} \quad (2)$$

and on the basis of the objective functions $F_{Rec}$, and $F_{PL}$ in turn, the F1 -classification accuracy indicator can be determined as:

$$F_{F1} = \frac{2F_{Rec} \cdot F_{PL}}{F_{Rec} + F_{PL}} \quad (3)$$

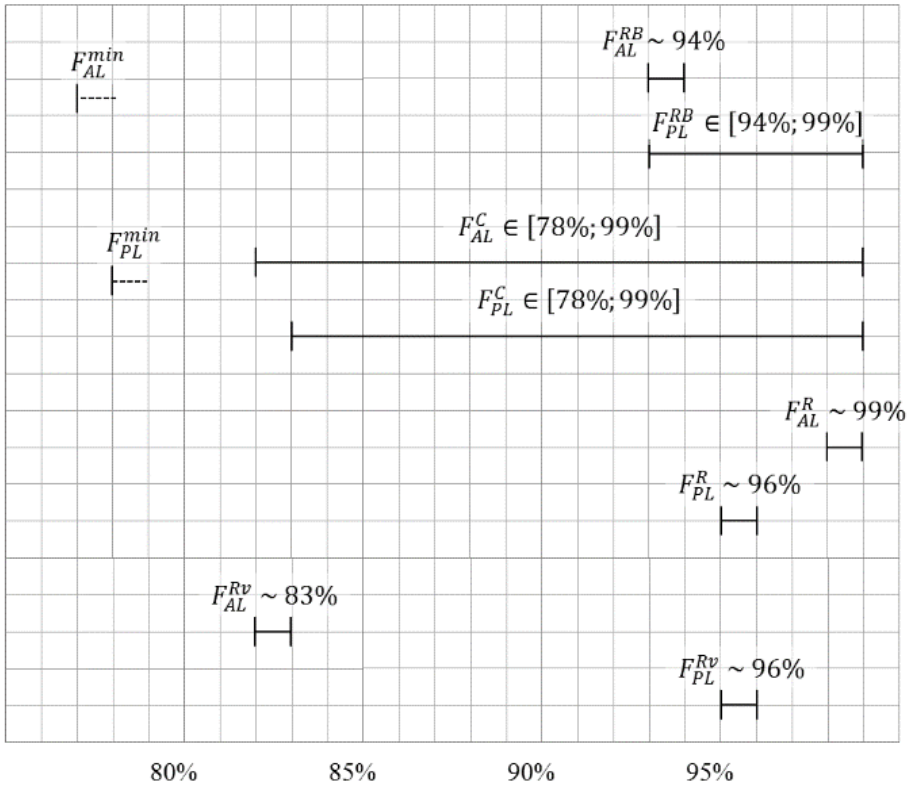


**Figure 6**: Ranges of accuracy for detecting patterns of cyberattacks when using algorithms based on RBM , CNN , RNN and RvNN

Thus, the task of optimizing the complex of machine analysis based on neural network algorithms for the detection and classification of cyber attack patterns can be reduced to the task of finding the global extremum of the objective function, as suggested above. At the same time, in accordance with the specific task of the organization of the SIEM system, one of the accuracy functions of the set $F_{AL}$, $F_{PL}$, $F_{Rec}$ and $F_{F1}$ is considered as a target function, and for the others, together with the indicator of the number of false classification results $\kappa_F$ or the indicator of the number of true classification results $\kappa_T$, permissible limits are introduced:

- permissible limits of the number of false or true pattern classification results as $\kappa_F \in [0\%; \kappa_F^{max}]$ or $\kappa_T \in [\kappa_T^{min}; 100\%]$, respectively, where the values of $\kappa_F^{max}$ and $\kappa_T^{min}$ are chosen in accordance with the requirements determined at the level of the SIEM -system organization depending on the task;
- permissible limits of the classification accuracy function $F_{AL} \in [F_{AL}^{min}; 100\%]$, where the value $F_{AL}^{min}$ is chosen in accordance with the requirements determined at the level of the SIEM -system organization, depending on the task;

- permissible limits of the classification accuracy function $F_{PL} \in [F_{PL}^{min}; 100\%]$, where the value $F_{PL}^{min}$is chosen in accordance with the requirements determined at the level of the SIEM -system organization, depending on the task;
- permissible limits of the classification accuracy function $F_{Rec} \in [F_{Rec}^{min}; 100\%]$, where the value $F_{Rec}^{min}$is chosen in accordance with the requirements determined at the level of the SIEM -system organization, depending on the task;
- permissible limits of the F 1 classification accuracy indicator $F_{\mathrm{F1}} \in [F_{\mathrm{F1}}^{min}; 100\%]$, where the value $F_{\mathrm{F1}}^{min}$is chosen in accordance with the requirements determined at the level of the SIEM -system organization, depending on the task.

According to the proposed approach, the arguments of the objective function and the functions that determine the permissible limits set by the researcher will be the following categories:

- a set of parameters defining the neural network architecture;
- a set of parameters that determine the selection of the activation function;
- set parameters that determine the peculiarities of neural network training and preparation of the training sample.

The specified technique allows to generalize the currently relevant approaches to optimizing machine analysis in order to detect cyberattacks on the components of the infrastructure of the data processing center, represented as DIS , and can be used in the future to solve a wide range of tasks related to the organization, configuration, reorganization, scaling and optimization of the SIEM -system.

## 5. Conclusion

In conclusion, there was presented a strategy based on neural network algorithms to raise the data center protection against distributed cyber attacks in this paper. As a result of the work carried out, the peculiarities of building mathematical models, which are used for the evaluation and optimization of neural network algorithms for the selection and classification of cyber attack patterns, in particular algorithms based on the neural network architecture of deep learning, were analyzed.

At the same time, within the framework of this study:

- a generalized scheme for detecting external threats at the level of the information security event management system was developed;
- approaches for adapting the neural network architecture of deep learning to identify cyber attack patterns are proposed;
- the ranges of accuracy of detection of cyber attack patterns when using autoencoder neural networks, multi-level autoencoder, deep belief neural networks, convolutional neural networks, recurrent neural networks, recursive neural networks and restricted Boltzmann machine are determined;
- the toolkit for evaluating the performance of the application of neural network algorithms has been expanded through the introduction of a set of machine analysis accuracy indicators, which act as target functions and permissible limits defined at the quantitative level.

It is shown that the presented mathematical model is generalized and when the mathematical apparatus is expanded, it provides an opportunity to organize, configure, reorganize, scale and optimize the SIEM system at an automatic level for a wide range of tasks that arise in the organization of data centers.

## 6. References


[1] Aiyetoro, G., & Owolawi, P. (2019). Spectrum management schemes for internet of remote things (IORT) devices in 5G networks via Geo Satellite. *Future Internet, 11* (12), 257. https://doi.org/10.3390/fi11120257.

[2] Yu, Y., Long, J., & Cai, Z. (2017). Network intrusion detection through stacking dilated convolutional autoencoders. *Security and Communication Networks, 2017*, 1–10. https://doi.org/10.1155/2017/4184196.

[3] Song, Y., Hyun, S., & Cheong, Y.-G. (2021). Analysis of autoencoders for network intrusion detection. *Sensors, 21* (13), 4294. https://doi.org/10.3390/s21134294.
[4] He, J., Tan, Y., Guo, W., & Xian, M. (2020). A small sample DDOS attack detection method based on Deep Transfer Learning. *2020 International Conference on Computer Communication and Network Security (CCNS)*. https://doi.org/10.1109/ccns50731. 2020.00019.
[5] Sarker, IH (2021). Deep cybersecurity: A comprehensive overview from neural network and Deep Learning Perspective. *SN Computer Science, 2* (3). https://doi.org/10.1007/ s42979-021-00535-6.
[6] Ma, X., Zhang, X., Dong, C., & Chen, X. (2021). A survey on Secure Outsourced Deep Learning. *Cyber Security Meets Machine Learning*, 129–163. https://doi.org/10.1007/978-981-33-6726-5_6.
[7] Toliupa, S., Buchyk, S., Nakonechnyi, V., ...Parkhomenko, I., Lukova-Chuiko, N. Building an Intrusion Detection System in Critically Important Information Networks with Application of Data Mining Methods. Proceedings - 16th International Conference on Advanced Trends in Radioelectronics, Telecommunications and Computer Engineering, TCSET 2022, 2022, стр. 128–133.
[8] Shtanenko, S., Samokhvalov, Y., Toliupa, S., Silko, O. Increasing survivability of technological systems based on the technology of programmable logic device. *CEUR Workshop Proceedings*, 2022, 3132, стр. 237–245.
[9] Abu Al-Haija, Q., & Al-Dala'ien, M. (2022). Elba-IoT: An ensemble learning model for botnet attack detection in IoT networks. *Journal of Sensor and Actuator Networks, 11* (1), 18. https://doi.org/10.3390/jsan11010018.
[10] Jiang, P., Wu, H., & Xin, C. (2021). DeepPOSE: Detecting GPS spoofing attack via deep recurrent neural network. *Digital Communications and Networks*. https://doi.org/10.1016/j.dcan.2021.09.006
[11] Mao, X., & Li, Q. (2020). Generative Adversarial Networks (GANS). Generative adversarial networks for image generation, 1–7. https://doi.org/10.1007/978-981-33-6048-8_1.
[12] Saisindhutheja, R., & Shyam, GK (2021). A deep belief network based attack detection using a secure SAAS framework. *2021 International Conference on Innovative Practices in Technology and Management (ICIPTM)*. https://doi.org/10.1109/ iciptm52218.2021. 9388329.
[13] Ma, X., Zhang, X., Dong, C., & Chen, X. (2021). A survey on Secure Outsourced Deep Learning. *Cyber Security Meets Machine Learning*, 129–163. https://doi.org/10.1007/978-981-33-6726-5_6.
[14] Hnatiienko, H., Kiktev, N., Babenko, N., Desiatko, A., Myrutenko, L. Prioritizing Cybersecurity Measures with Decision Support Methods Using Incomplete Data *// Selected Papers of the XXI International Scientific and Practical Conference "Information Technologies and Security*", Kyiv, Ukraine, December 9, 2021 / CEUR Workshop Proceedings, 2021, 3241, pp. 169–180.
[15] Song, H., Zhuqing, J., Men, A., Yang, B. A hybridsemi-supervised anomaly detection model for high-dimensional data. *Computational intelligence and neuroscience*, vol. 2017, 2017
[16] Manikopoulos, C., Papavassiliou, S. Network in-trusion and fault detection: a statistical anomaly approach. *IEEE Communications Magazine*, vol. 40, no. 10, pp. 76–82, 2002.