Evaluation the Efficiency of Information Technology of Big Data Intelligence Analysis and Processing

Myroslav Komar [1], Oleg Savenko [2], Anatoliy Sachenko [1,3], Taras Lendiuk [1], Khrystyna Lipianina-Honcharenko [1], Grygoriy Hladiy [1], Nadiia Vasylkiv [1]

[1] *West Ukrainian National University, Lvivska str., 11, Ternopil, 46000, Ukraine*
[2] *Khmelnytskyi National University, Instytutska str., 11, Khmelnytskyi, 29016, Ukraine*
[3] *Kazimierz Pulaski University of Technology and Humanities in Radom, Department of Informatics, Jacek Malczewski str., 29, Radom, 26 600, Poland*

Abstract

The paper is devoted to the solution of an actual scientific and applied problem of efficiency estimation of information technology development at intelligent analysis and big data processing. A study of the problem of big data analysis and processing is conducted. The concept of missing data recovery based on the integration of the big data model, the method of missing data recovery based on functional dependencies and associative rules and the complexity assessment of the missing data recovery method have been developed. Methods for classifying network packets, recognizing objects in satellite imagery, recognizing objects in images of text documents based on deep neural networks, creating and operating deep neural networks based on an evolutionary approach, increasing the speed of analysis and processing of big data have been developed. The methodology and information technology of intelligent analysis and big data processing have been developed by authors. The evaluation of the developed information technology efficiency is carried out.

Keywords

information technology, intelligent analysis and big data processing, missing data recovery, functional dependencies, associative rules and assessment of missing data recovery method complexity, network packet classification methods, object recognition in text document images based on deep neural networks.

1. Introduction

Today, big data [5] has become a defining and growing feature of the modern world economy, a vector of global change in almost all its sectors. But complex, unstructured data are difficult to assess using traditional analytical methods. This is where deep neural networks come to the rescue [14], [21], which have high reliability of nonlinear data conversion and presentation compared to traditional neural networks and allow processing and analyzing large amounts of data in various fields, including speech recognition [8], [13], computer vision [7], [39] and others [30], [41].

Currently, in the theory and practice of analysis and processing of big data, the contradiction between: the growing number of different sources that generate large amounts of data, which are not always high quality; availability of information technologies that are widely used for analysis and processing of big data, but do not provide sufficient reliability and efficiency in the processing of semi-structured and unstructured data; the growing demands of users for real-time big data analysis for operational decision-making, on the one hand, and the lack of concept of big data analysis and

COLINS-2022: 6th International Conference on Computational Linguistics and Intelligent Systems, May 12–13, 2022, Gliwice, Poland
EMAIL: mko@wunu.edu.ua (M. Komar); savenko_oleg_st@ukr.net (O. Savenko); as@wunu.edu.ua (A. Sachenko); tl@wunu.edu.ua (T. Lendiuk); xrustya.com@gmail.com (K. Lipianina-Honcharenko); hladiy@yahoo.com (G. Hladiy); nvs@wunu.edu.ua (N. Vasylkiv)
ORCID: 0000-0001-6541-0359 (M. Komar); 0000-0002-4104-745X (O. Savenko); 0000-0002-0907-3682 (A. Sachenko); 0000-0001-9484-8333 (T. Lendiuk); 0000-0002-2441-6292 (K. Lipianina-Honcharenko); 0000-0002-5585-8472 (G. Hladiy); 0000-0002-4247-7523 (N. Vasylkiv)

processing in their absence, incompleteness, vagueness and uncertainty, as well as limited computing resources – on the other hand.

The imperfection and limitations of existing approaches and methods do not allow to ensure sufficient efficiency of analysis and processing of big data. Therefore, the development of methods for assessing the effectiveness of mining and processing of big data in the conditions of their partial absence, incompleteness, vagueness and uncertainty, is certainly relevant.

2. Related Work

Because in-depth learning is based on hierarchical learning and obtaining different levels of complex data abstractions, it is suitable for simplifying big data analysis, semantic indexing, data labeling, information retrieval and image recognition [1], including image classification and network packet parameters to detect invasion.

In [9], [16], [28], [31], [38] methods of intrusion detection based on Autoencoder (AE) are used. The AE architecture can be considered as a data compression algorithm that is, in fact, able to first compress the input data and then recover it. Cybersecurity researchers also use AE to detect intrusions, in particular in [31] offer asymmetric deep AE, which successfully reduces the computational cost of data analysis. In [38] offer AE for various cybersecurity programs, consisting of two stages of training, i.e. pre-training and fine-tuning. Since the collected network raw data may have an unbalanced distribution, in [9] deep AE is used to create classification models to detect abnormal behavior. To create a flexible intrusion detection system in [16] use a sparse AE with a softmax layer. [16] presents a more powerful approach to creating an intrusion detection system with scalable, self-adaptive and autonomous characteristics.

Deep Belief Neural Network (DBN) is widely used to detect intrusions [3], [4], [10], [40]. In a study [10], the authors focused on working with big raw data and used DBN to build an intrusion detection system, adjusting parameters such as the number of hidden layers and the number of neurons to find better settings than other machine learning methods on the KDDCup 99 dataset.

To explore DBN's ability to detect attacks, [3] proposed a system that first uses digital coding and a standardized method to select functions, and then uses DBN to classify network intrusions by assigning a class label to each function vector. To solve the problem of redundant information in [40] it is proposed to detect intrusions using DBN and probabilistic neural network (PNN). To solve the problem of real-time attack detection, [4] proposed an anomaly detection method based on DBN, which consists of only one hidden RBM layer and a fine-tuning layer.

Malicious attacks are constantly changing and occur on very large amounts of data, which requires scalable solutions. To solve this problem, [35] proposed a Deep Neural Network (DNN) structure that scales and can monitor network traffic and host-level events in real time, actively preventing possible network attacks. In [29], a deep learning-based network intrusion detection method is used, which uses a deep neural network to obtain characteristics of network monitoring data, and a BP neural network is used to classify intrusion types. Also, to solve the problem of network security, [33] presents the DNN model for detecting anomalies based on data flow.

Methods of intrusion detection based on recurrent neural networks (RNN) are presented in [17], [19], [20], [32], [37]]. Thus, in [32] it is proposed to take into account the characteristics of the time series of known malicious behavior and network traffic, which can increase the accuracy of attack detection algorithms. In [32], RNN is used to solve the problem of classification of attacks, where the proposed classification model is based on self-study.

Similarly, [37] investigated the use of RNN to detect intrusions using forms of binary classification and multiclass classification.

Since the LSTM (Long short-term memory) neural network, which is a type of RNN, overcomes the disappearance of the vanishing gradient during training, [17] uses the LSTM architecture to detect intrusions. Compared to [17], the constructed LSTM model has a higher level of error detection in training using the KDDCup 99 data set. Following the trend of using LSTM to detect attacks, in [20] also built a classifier LSTM to detect intrusions.

Combined deep learning methods are also used to detect attacks [22], [24]. In particular, [22] uses a hybrid method of deep learning based on AE and DNN to detect malicious code. In [22] an ensemble

network is used to classify different types of attacks. To distinguish between normal and abnormal behavior, the proposed method combines AE, BNN, DNN.

[25] describes a method based on the use of SVM for automatic detection of solar panels using high-resolution satellite images. This approach first uses a pre-screening operation that identifies regions, which are then processed to identify features. The source information obtained with this model is a list of regions and confidence values that show how likely the presence of a solar panel in a given area.

The task of direct detection of solar panels is insufficiently presented in scientific publications. [26] uses an approach based on the use of decision trees. Achieved panel localization rate was 90% in the case of using certain parameters of the algorithm, and the method itself consists of four stages. In [6], the Aerial Imagery Dataset is used as a training sample, which includes images with a resolution of up to 5000 × 5000 pixels.

3. Proposed Approach

3.1. Methodology of intelligent analysis and big data processing

The methodology of intelligent analysis and processing of big data in the absence, incompleteness, vagueness and uncertainty is based on the use of the following principles, which are the basis of the developed models and methods:

1. Basic principles of working with big data, which can be formulated on the basis of definitions of big data [2], [15], [23]:

- the principle of horizontal scalability;
- the principle of resistance to failures;
- the principle of data locality.

2. Principles that increase the efficiency and reliability of analysis and processing of big data:

- the principle of ensuring resistance to data errors;
- the principle of ensuring the ability to learn;
- the principle of ensuring the ability to evolve and adapt;
- the principle of ensuring sufficient speed and data security.

3. Principles that increase the efficiency and reliability of analysis and of big data processing:

- the principle of ensuring resistance to data errors;
- the principle of ensuring the ability to learn;
- the principle of ensuring the ability to evolve and adapt;
- the principle of ensuring sufficient speed and data security.

The basic principle of big data processing is mainly considered to be horizontal scalability, which provides data processing distributed over hundreds and thousands of computing nodes, without performance degrading [27].

The principle of fault tolerance is derived from the principle of horizontal scalability. As there may be many computing nodes in the cluster (sometimes dozens of thousands) and their number it is possible to increase the probability of machine failure increases. Methods of big data processing should take into account the likelihood of such situations and provide preventive measures in the process.

The principle of data locality. Because data is distributed across a large number of computing nodes, if it is physically located on one server and processed on another, the cost of data transfer may be unreasonably high. Therefore, it is desirable to process data on the same machine on which they are stored.

The principle of ensuring resistance to data errors. When using information technology to obtain reliable and high-quality results, it is important not only the methods, ways and means of obtaining them, but also the initial data quality. Small variations or specifics of the initial data, especially incomplete data, can lead to inaccurate results, and in the case of the neural network technologies usage – completely unbalance the settings of such models. Qualitative data are characterized by different parameters.

Data completeness is an indicator of the amount of available data relative to the desired amount. It is used to confirm how data deficiencies will affect their usefulness.

Accuracy of big data can be defined as the degree to which data accurately describes the object or real world being considered. To measure data set or data element accuracy, the data is compared with standards or data that are commonly used or accepted.

The timeliness of big data is one of the critical parameters for assessing the quality of big data, because they can change very quickly and if the importance of this parameter is neglected, the relevance of the result is lost. The timeliness of big data is measured by the degree of data that represents reality at the right time.

The uniqueness of big data is defined as the measurement of a data item relative to itself or its counterpart in another data set or database. The validity of big data indicates the syntax, i.e. the correctness of the data, their format, type and range.

To measure reliability, compare the data with the actual rules defined for them.

Consistency of big data refers to the extent to which the logical relationship between correlated data is correct and complete, i.e. it determines the absence of difference when comparing two or more data from an event or object. To measure the data consistency parameter, the data item is measured against an object or event and its counterpart in another data set.

Reliability of the big data system is defined as the ability of the network to provide reliable data transmission in a state of constant change in the structure of the network or the ability of the device to provide reliable data output.

The need for big data is responsible for determining the usefulness of data and the satisfaction of user needs. Timeliness, accuracy and completeness are calculated to measure the parameter, as the value of this parameter determines the possibility of using the data.

The principle of resilience to data errors based on data quality assessment involves recovering missing data by creating additional data values using the base domain and functional dependencies and adding these values to existing training data to ensure their completeness.

The development of intelligent systems must include the ability to learn in order to memorize the associated data, to summarize similar data, referring them to one class. This principle is provided by the use of neural network technologies for analysis and processing of big data.

The principle of providing the ability to evolve and adapt provides the ability to develop the system over time in order to acquire better properties, such as finding the optimal parameters of the neural network to work in conditions of limited computing resources. In the case of analyzing the parameters of network traffic to detect intrusions – the ability to take into account the parameters of previous attacks in order to create better classifiers.

The principle of ensuring sufficient speed and data security involves the possibility of efficient hardware implementation (based on random access memory and programmable logic matrices) and real-time decision-making based on fuzzy logic. The implementation of this principle will make it possible to process and analyze data in real time, or close to it, which is extremely important when building modern data analysis systems for critical infrastructure. An additional argument in favor of hardware implementation of the decision-making subsystem based on fuzzy logic is that the operation of hardware implementation cannot be blocked by targeted exposure to a computer virus. At software implementation, it is possible to completely block this subsystem (which leads to server downtime, and therefore relatively easy to detect), and purposeful choice of such a method of protection against computer viruses, which, to increase the effectiveness of the attack, provides minimal and consistent protection (which does not lead to server downtime, and therefore such an action is difficult to detect). At the same time, the computer virus can neither disable nor modify the hardware implementation of the decision-making subsystem.

The proposed principles are the basis for the development of models and methods of data mining and data processing:

1. Big data quality models that take into account seven quality parameters (p_1 corresponds to data completeness, p_2 – accuracy, p_3 – timeliness, p_4 – uniqueness, p_5 – reality, p_6 – sequence, p_7 – reliability), which allowed to determine the characteristics of these parameters due to lack correlation between them and assess the quality of big data, in particular to obtain information about the presence of missing values.
2. Big data models for missing data recovery, which is based on a hierarchy of objects that allows processing structured and semi-structured data from sources with different data structures [36].

3. The method of recovering missing data, which allows analyzing the hidden dependencies in the data set and take into account the nature of the data set and predict the lack of data for each data source separately based on the specifics and nature of these sources. The method creates additional data values based on functional dependencies and association rules and adds these values to the existing training data, which allowed to increase the efficiency and reliability of further data analysis [36].
4. The method of classification of network packets based on deep neural networks, which is characterized by reducing the dimensionality of the analyzed information in the middle of the network and minimizing the standard error of recovery of analyzed information and provides classification of the main components of texture features. This, in turn, made it possible to analyze unstructured big data, increased the speed and reliability of their processing and created the possibility of application in intelligent intrusion detection systems in real time [18].
5. A method of recognizing objects in the images of satellite images based on a deep convolutional neural network, which allows to increase the reliability of images classification with poor quality and low resolution [12].
6. A method of recognizing objects in images of text documents, based on image pre-processing, which simplifies the localization of individual parts and subsequent recognition of localized blocks using a deep convolutional neural network, which increases the reliability of classification of localized parts of the document [11].
7. A method of creating and operating deep neural networks based on an evolutionary approach, which allows in parallel with the algorithm to find the optimal neural network parameters to analyze big data based on the neural network with minimal learning error at each step of the genetic algorithm. conditions of limited computing resources, as well as the ability to model the neural network depending on the required performance and reliability.
8. Method of increasing the speed of analysis and processing of big data, which is based on pre-processing options for possible outputs of the deep neural network or possible solutions that meet all combinations of required and current parameters of fuzzy inference, which provided efficient hardware implementation (based on random access memory) and programmable logic matrices) and real-time decision making based on fuzzy logic.

3.2. Information technology of intelligent analysis and big data processing

For information support of the offered models and methods the information technology of intelligent analysis and processing of big data in the conditions of their absence, incompleteness, vagueness and uncertainty is developed.

The methodology of data mining and big data processing developed above provides basic support for the proposed information technology, which, in turn, determines the information flows and functions for their processing, the relationships between them, control information and tools.

The input data of the proposed information technology of intelligent analysis and big data processing are:

- parameters of network packets;
- satellite images with the image of solar panels;
- images of text documents.

The implementation of information technology involves the principle of modularity, which allows, if necessary, the analysis of other data to implement the corresponding function without restructuring the entire system. It is enough to implement a neural network to analyze new data. As a tool for building models of deep learning, it is advisable to use the Tensorflow framework [34].

The source information of information technology of intelligent analysis and big data processing are:

- the result of the classification of network packets;
- the result of the recognition of solar panels;
- the result of recognition of text documents.

Under certain conditions, a method of increasing the speed of analysis and processing of big data can be used, which involves an approach to reducing the hardware complexity of the neural network.

The decision-making unit analyzes the result obtained and decides whether to return to the original conditions for data re-analysis or use of the obtained results.

The developed information technology of intelligent analysis and big data processing is presented in the form of a structural model (Fig. 1).

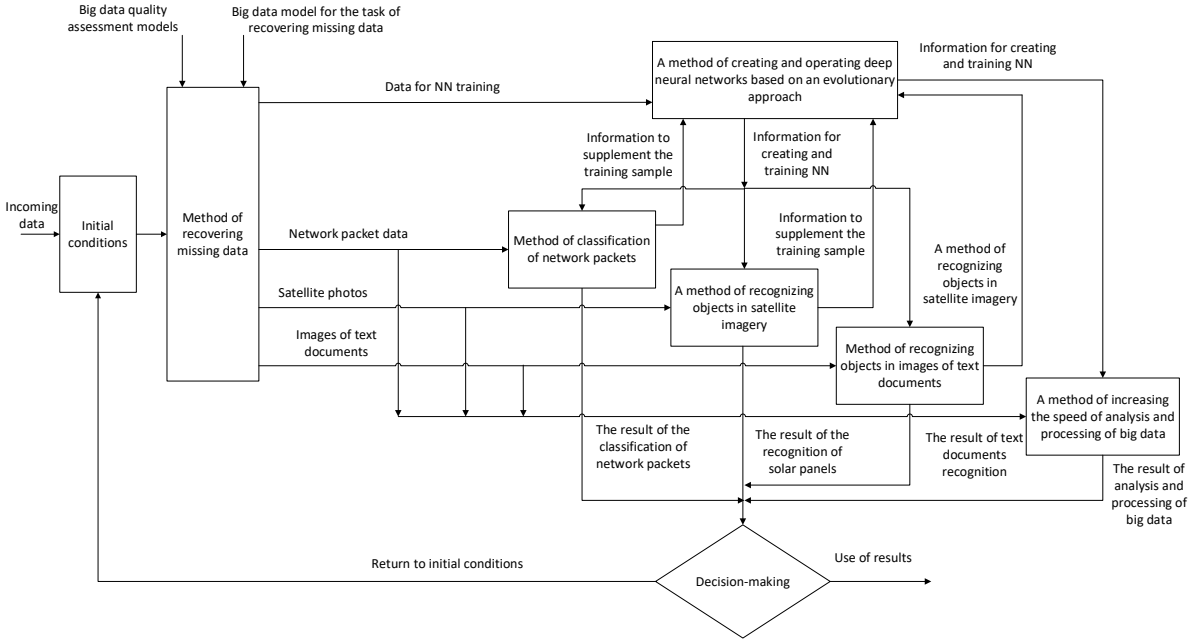

Figure 1: Structural model of information technology for intelligent analysis and big data processing

3.3. Evaluation of the information technology efficiency for intelligent analysis and big data processing.

Let's evaluate the effectiveness of the developed information technology for intelligent analysis and big data processing.

Let t_0 – be the time spent on data recovery, and t_0^j – be the time spent on data recovery for the i-th iteration or the j-th processed task.

Then, let the data recovery be performed to process the next $(q+1)$ big data. The time for processing is denoted by t_0^{q+1} and defined as follows:

$$t_{0,Oc}^{q+1} = \frac{\sum_{i=1}^{q} t_0^i}{q * \max\limits_{1 \le i \le q} t_0^i}. \quad (1)$$

Thus, $t_{0,Oc}^{q+1}$ will be evaluated and normalized to take it into account in the following stages of calculating the effectiveness of the developed methods. The value of $t_{0,Oc}^{q+1}$ is obtained as the average of all previous processing time for data recovery divided by the maximum among the previous values.

Therefore, then $t_{0,Oc}^{q+1} \le 1$.

In the future, the next value of time t_0^{q+2} can be estimated from its previous values

$$t_{0,Oc}^{q+2} = \frac{\sum_{i=1}^{q+1} t_0^i}{(q+1) * \max\limits_{1 \le i \le (q+1)} t_0^i}. \quad (2)$$

Let's introduce an iterative formula to optimize calculations

$$t_{0,Oc}^{q+2} = \frac{\left(\frac{\sum_{i=1}^{q} t_0^i}{q} + t_0^{q+1}\right)}{2 * max(\max\limits_{1 \le i \le q}(t_0^i); t_0^{q+1})}. \quad (3)$$

The use of three different methods separately depending on the type of input big data is taken into account by introducing the coefficients $\alpha_j, j = \overline{1,3}$, and:

$\alpha_j = 1$, if the method is applicable to the input data;

$\alpha_j = 0$, if the input method is not applicable.

The method processing time will be determined analogously to formula (2) and the following notation will be introduced:

$t_1^{q_1}$ – time spent on processing input data by the first method in q_1 step;

$t_2^{q_2}$ – time spent on processing input data by the second method in q_2 step;

$t_3^{q_3}$ – is the time spent processing the input data by the third method in step q_3.

We introduce the time $t_4^{q,p}$, which indicates the additional costs associated with making decisions about returning to the original conditions. It will include the time spent on data recovery and processing one of the methods.

Then

$$t_4^{q,p} = t_0^{q,p} + (\alpha_1 t_1^{q,p} + \alpha_2 t_2^{q,p} + \alpha_3 t_3^{q,p}) \quad (4)$$

The number of such returns p when processing one series of input data is set by the decision-making system and depends on the recovery step. If the input data is not restored, then the system captures and issues a calculation in the previous step of the value with the appropriate message.

The efficiency of the system as a whole is defined as a value that takes into account the time for data recovery, time for data processing and time spent on reprocessing:

$$K_e = \frac{t_0^q + (\alpha_1 t_1^q + \alpha_2 t_2^q + \alpha_3 t_3^q) + t_4^{q,p}}{2}. \quad (5)$$

The minimization function f_{min} of the efficiency factor will reflect the best efficiency, in particular we will present it as follows:

$$f_{min} : K_e \rightarrow min. \quad (6)$$

If $p = 0$, then

$$K_e = \frac{t_0^q + (\alpha_1 t_1^q + \alpha_2 t_2^q + \alpha_3 t_3^q)}{2}, \quad (7)$$

while $K_e \le 1$.

If $p > 0$, then K_e may be greater than 1, which will indicate problems with the quality of the initial data, their recovery, as well as the effectiveness of the processing methods used. As a result of experimental studies obtained $K_e = 0{,}82$.

Therefore, the proposed assessment of the effectiveness of information technology allows to take into account the time to recover missing data, time to analyze and process data and, if necessary, the time spent on re-analysis and processing.

4. Conclusions

The methodology of intelligent analysis and processing of big data in the conditions of their absence, incompleteness, vagueness and uncertainty is based on the use of the following principles, which are the basis of the developed models and methods:

1. Basic principles of working with big data, which can be formulated on the basis of definitions of big data:

- the principle of horizontal scalability;
- the principle of resistance to failures;
- the principle of data locality.

2. Principles that increase the efficiency and reliability of analysis and processing of big data:

- the principle of ensuring resistance to data errors;
- the principle of ensuring the ability to learn;
- the principle of ensuring the ability to evolve and adapt;
- the principle of ensuring sufficient speed and data security.

A structural model of information technology analysis and processing of big data has been developed, which reflects the movement of information flows in the integration of big data sets with models of deep learning and provides decision-making on the need for pre-processing of input data, recovery of missing data; application of appropriate deep neural networks for big data analysis; the need to increase the speed of deep neural networks to work in real time and in conditions of limited computing resources. This has increased the efficiency and reliability of the analysis and processing of big data compared to known information technologies.

An approach to evaluate the effectiveness of mining and data processing based on the use of missing data recovery method, network packet classification method, object recognition method in satellite imagery images, object recognition method in text document image methods, method of creation and operation deep neural networks based on an evolutionary approach, as well as a method to increase the speed of analysis and processing of big data.

The results of experimental studies confirm that the proposed approach allows to ensure sufficient efficiency. In particular, it was found that the level of efficiency of the applied methods of analysis and processing of big data was obtained at the level of $K_e = 0{,}82$.

5. References

[1] A. Oussous, F.-Z. Benjelloun, A. A. Lahcen, S. Belfkih, Big data technologies: A survey, Journal of King Saud University – Computer and Information Sciences, 30 (2018) 431–448. https://doi.org/10.1016/j.jksuci.2017.06.001.

[2] H. Al-Barashdi, R. Al-Karousi, Big data in academic libraries: Literature review and future research directions, Journal of Information Studies and Technology 2018 (2019) 2-16. https://doi.org/10.5339/jist.2018.13.

[3] M. Z. Alom, V. Bontupalli, T. M. Taha, Intrusion detection using deep belief networks, in: Proceedings of the National Aerospace and Electronics Conference, Dayton, OH, USA, June 2015, pp. 339–344. https://doi.org/10.1109/NAECON.2015.7443094.

[4] K. Alrawashdeh, C. Purdy, Toward an online anomaly intrusion detection system based on deep learning, in: Proceedings of the 15th IEEE International Conference on Machine Learning and Applications, Anaheim, CA, USA, December 2016, pp. 195–200. https://doi.org/10.1109/ICMLA.2016.0040.

[5] Big Data: The Next Frontier for Innovation, Competition, and Productivity. McKinsey Global Institute. URL: http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation.

[6] K. Bradbury, R. Saboo, T. L. Johnson, J. Malof, Distributed solar photovoltaic array location and extent dataset for remote sensing object identification, Scientific Data 3 (2016). https://doi.org/10.1038/sdata.2016.106.

[7] D. Cirean, U. Meler, L. Cambardella, J. Schmidhuber, Deep, big, simple neural nets for handwritten digit recognition, Neural Computation 22 (2010) 3207-3220. https://doi.org/10.1162/NECO_a_00052.

[8] G. E. Dahl, D. Yu, L. Deng, A. Acero, Context-dependent pretrained deep neural networks for large-vocabulary speech recognition, IEEE Transactions on Audio Speech and Language Processing 20 (2012) 30-41. https://doi.org/10.1109/TASL.2011.2134090.

[9] F. Farahnakian, J. Heikkonen, A deep auto-encoder based approach for intrusion detection system, in: Proceedings of the 20th International Conference on Advanced Communication Technology, Chuncheon, South Korea, 11-14 February, 2018, pp. 178–183. https://doi.org/10.23919/ICACT.2018.8323687.
[10] N. Gao, L. Gao, Q. Gao, H. Wang, An intrusion detection model based on deep belief networks, in: Proceedings of the Second International Conference on Advanced Cloud and Big Data, Huangshan, China, November 2014, pp. 247–252. https://doi.org/10.1109/CBD.2014.41.
[11] V. Golovko, A. Kroshchanka, E. Mikhno, A. Sachenko, S. Bezobrazov, M. Komar, I. Shylinska, Deep convolutional neural network for recognizing the images of text documents, in: Workshop Proceedings of the 8th International Conference on "Mathematics. Information Technologies. Education", MoMLeT&DS-2019, Shatsk, Ukraine, June 2-4, 2019, CEUR-WS, vol. 2386, pp. 297-306. http://ceur-ws.org/Vol-2386/paper22.pdf.
[12] V. Golovko, A. Kroshchanka, E. Mikhno, M. Komar, A. Sachenko, Deep convolutional neural network for detection of solar panels, in: T. Radivilova, D. Ageyev, N. Kryvinska, (eds) Data-Centric Business and Applications, volume 48 of Lecture Notes on Data Engineering and Communications Technologies, Springer, Cham, 2021, pp. 371-389. https://doi.org/10.1007/978-3-030-43070-2_17.
[13] G. Hinton, L. Deng, D. Yu, G. Dahl, A. Mohamed, N. Jaitly, A. Senior, V. Vanhoucke, P. Nguyen, T. Sainath, B. Kingsbury, Deep neural network for acoustic modeling in speech recognition, IEEE Signal Processing Magazine 29 (2012) 82–97. https://doi.org/10.1109/MSP.2012.2205597.
[14] G. E. Hinton, E. S. Osindero, Y. The, A fast learning algorithm for deep belief nets, Neural Computation 18 (2006) 1527–1554. https://doi.org/10.1162/neco.2006.18.7.1527.
[15] W. H. Inmon, Big Data – getting it right: A checklist to evaluate your environment, URL: http://dssresources.com/papers/features/ inmon/inmon01162014.htm.
[16] A. Javaid, Q. Niyaz, W. Sun, M. Alam, A deep learning approach for network intrusion detection system, in: Proceedings of the 9th EAI International Conference on Bio-Inspired Information and Communications Technologies, New York, NY, USA, December 2016, pp. 21–26. https://doi.org/10.4108/eai.3-12-2015.2262516.
[17] J. Kim, J. Kim, H. Kim, Long short term memory recurrent neural network classifier for intrusion detection, in: Proceedings of the International Conference on Platform Technology and Service, Jeju, Korea, February 2016, pp. 1–5. https://doi.org/10.1109/PlatCon.2016.7456805.
[18] M. Komar, A. Sachenko, V. Golovko, V. Dorosh, Compression of network traffic parameters for detecting cyber attacks based on deep learning, in: Proceedings of the 9th IEEE International Conference on Dependable Systems, Services and Technologies, Kyiv, Ukraine, May 24-27, 2018, pp. 44-48. https://doi.org/10.1109/DESSERT.2018.8409096.
[19] R. B. Krishnan, N. Raajan, An intellectual intrusion detection system model for attacks classification using RNN, International Journal of Pharmaceutical Technology and Biotechnology 8 (2016) 23157–23164.
[20] T. Le, J. Kim, H. Kim, An effective intrusion detection classifier using long short-term memory with gradient descent optimization, in: Proceedings of the International Conference on Platform Technology and Service, Jeju, Korea, February 2017, pp. 1–6. https://doi.org/10.1109/PlatCon.2017.7883684.
[21] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, Nature 521 (2015) 436–444. https://doi.org/10.1038/nature14539.
[22] Y. Li, R. Ma, R. Jiao, A hybrid malicious code detection method based on deep learning, International Journal of Security and its Applications 9 (2015) 205–216. https://doi.org/10.14257/ijsia.2015.9.5.21.
[23] R. K. Lomotey and R. Deters, Towards knowledge discovery in big data, in: Proceedings of the 2014 IEEE 8th International Symposium on Service Oriented System Engineering, 2014, pp. 181-191. https://doi.org/10.1109/SOSE.2014.25.
[24] S. A. Ludwig, Intrusion detection of multiple attack classes using a deep neural net ensemble, in: Proceedings of the IEEE Symposium Series on Computational Intelligence, Honolulu, HI, USA, November 2017. Honolulu, 2017, pp. 1-7. https://doi.org/10.1109/SSCI.2017.8280825.
[25] J. Malof, R. Hou, L. M. Collins, K. Bradbury, R. Newell, Automatic solar photovoltaic panel detection in satellite imagery, in: Proceedings of the International Conference on Renewable

Energy Research and Applications, Palermo, Italy, 22-25 November, 2015, pp. 1428–1431. https://doi.org/10.1109/ICRERA.2015.7418643.
[26] J. Malof, K. Bradbury, L. Collins, R. Newell, Automatic detection of solar photovoltaic arrays in high resolution aerial imagery, Applied Energy 183 (2016) 229–240. https://doi.org/10.1016/j.apenergy.2016.08.191.
[27] M. Chen, S. Mao, Y. Zhang, V. C. M. Leung, Big Data, Related Technologies, Challenges, and Future Prospects, Spinger, 2014, 100 p.
[28] D. Papamartzivanos, F. Gomez Marmol, G. Kambourakis, Introducing deep learning self-adaptive misuse network intrusion detection systems, IEEE Access 7 (2019) 13546–13560. https://doi.org/10.1109/ACCESS.2019.2893871.
[29] W. Peng, X. Kong, G. Peng, X. Li, Z. Wang, Network intrusion detection based on deep learning, in: Proceedings of the 2019 International Conference on Communications, Information System and Computer Engineering, Haikou, China, July 2019, pp. 431–435. https://doi.org/10.1109/CISCE.2019.00102.
[30] R. Salakhutdinov, A. Mnih, G. Hinton, Restricted Boltzmann machines for collaborative filtering, in: Proceedings of the 24th International Conference on Machine Learning, Corvallis, Oregon, USA, June 20-24, 2007, pp. 791–798. https://doi.org/10.1145/1273496.1273596.
[31] N. Shone, T. N. Ngoc, V. D. Phai, Q. Shi, A deep learning approach to network intrusion detection, IEEE Transactions on Emerging Topics in Computational Intelligence 2 (2018) 41–50. https://doi.org/10.1109/TETCI.2017.2772792.
[32] R. C. Staudemeyer, Applying long short-term memory recurrent neural networks to intrusion detection, South African Computer Journal 56 (2015) 136–154. https://doi.org/10.18489/sacj.v56i1.248.
[33] T. A. Tang, L. Mhamdi, D. McLernon, S. A. R. Zaidi, M. Ghogho, Deep learning approach for network intrusion detection in software defined networking, in: Proceedings of the 2016 International Conference on Wireless Networks and Mobile Communications, Reims, France, October 2016, pp. 258–263. https://doi.org/10.1109/WINCOM.2016.7777224.
[34] Tensorflow, URL: https://www.tensorflow.org.
[35] R. Vinayakumar, M. Alazab, K. P. Soman, P. Poornachandran, A. Al-Nemrat, S. Venkatraman, Deep learning approach for intelligent intrusion detection system, IEEE Access 7 (2019) 41525–41550. https://doi.org/10.1109/ACCESS.2019.2895334.
[36] C. Wang, N. Shakhovska, A. Sachenko, M. Komar, A new approach for missing data imputation in big data interface, Information Technology and Control 49 (2020) 541-555. https://doi.org/10.5755/j01.itc.49.4.27386.
[37] C. Yin, Y. Zhu, J. Fei, X. He, A deep learning approach for intrusion detection using recurrent neural networks, IEEE Access 5 (2017) 21954–21961. https://doi.org/10.1109/ACCESS.2017.2762418.
[38] M. Yousefi-Azar, V. Varadharajan, L. Hamey, U. Tupakula, Autoencoder-based feature learning for cyber security applications, in: Proceedings of the International Joint Conference on Neural Networks, San Diego, CA, USA, June 2017, pp. 3854–3861. https://doi.org/10.1109/IJCNN.2017.7966342.
[39] M. Zeiler, G. Taylor, R. Fergus, Adaptive deconvolutional networks for mid and high level feature learning, in: Proceedings of the IEEE International Conference on Computer Vision, Barcelona, Spain, November 6-13, 2011, pp. 2018–2025. https://doi.org/10.1109/ICCV.2011.6126474.
[40] G. Zhao, C. Zhang, L. Zheng, Intrusion detection using deep belief network and probabilistic neural network, in: Proceedings of the IEEE International Conference on Computational Science and Engineering and IEEE International Conference on Embedded and Ubiquitous Computing, Taipei, Taiwan, December 2017, vol, pp. 639–642. https://doi.org/10.1109/CSE-EUC.2017.119.
[41] Z.-N. Hu, Y. V. Bodyanskiy, N. Ye. Kulishova, O. K. Tyshchenko, A multidimensional extended neo-fuzzy neuron for facial expression recognition, International Journal of Intelligent Systems and Applications (IJISA) 9 (2017) 29-36. https://doi.org/10.5815/ijisa.2017.09.04.