UDC 519.6 DOI: 10.18698/0236-3933-2020-2-75-84

DEVELOPMENT OF A NEUROCOMPUTER MODULAR INFORMATION SYSTEM FOR CANCEROUS DISEASES DIAGNOSTICS IN ANIMALS

N.A. Staroverova[1] nata-staroverova@yandex.ru
M.L. Shustrova[1] shu.ma@bk.ru
S.A. Staroverov[2] staroverovsergey@me.com
L.A. Dykman[2]

[1] **Kazan National Research Technological University, Kazan, Republic of Tatarstan, Russian Federation**
[2] **Institute of Biochemistry and Physiology of Plants and Microorganisms, Russian Academy of Sciences, Saratov, Russian Federation**

Abstract

Information system was developed in the form of a web application that makes it possible to identify microscopic images of cytological samples, to establish an initial diagnosis and to provide recommendations for its further confirmation based on additional data. Approaches, assumptions and prerequisites adopted in the information system development are described. It is proposed to use neural networks as the information system element in sample identification and making the initial diagnosis. Patient data, affected area images and microscopic images of cytological samples are planned to be collected in the information system database under creation. Cytological sample images serve as the input data for neural networks operation. Cytological picture assessment is based on the use of the following characteristic features: preparation background, number and location of cells, size and shape of cells, nucleus, presence of multinucleated cells and fission entities (atypical mitoses), etc.

Keywords

Information system, neural networks, cancer, cytology, machine learning

Received 26.11.2019
Accepted 23.12.2019

Work by S.A. Staroverov and L.A. Dykman in regard to cytological studies was supported by the RFBR project no. 19-14-00077

Introduction. Cancer is considered the most complicated disease that humanity is facing. More than 200 forms of cancer were described, each of them is characterized by different molecular profiles that require unique therapeutic strategies. Cancer is accompanied by genetic alterations unique for each type of the disease [1]. Therefore, the task of improving diagnostics, treatment and prevention of cancer based on studying genetic alterations in tumor became topical in the recent years. The Cancer Genome Atlas (TCGA) was founded in 2005, and the International Cancer Genome Consortium (ICGC) was created in 2008; and these are the main projects involving acceleration of the comprehensive cancer genetics study using innovative genome analysis technologies and aimed at creating new cancer treatment methods, diagnostic methods and preventive strategies [2].

Animal in modern society is considered to be a member of the family; therefor, today veterinary medicine is presenting high demands to development of the efficient cancer diagnostics methods. Visual diagnostics methods, which are using cytological and histological research, are currently the main techniques in establishing the diagnosis.

At the first stage of our work, we are focusing on visual diagnostics and differentiation. In the future, it is supposed to extend this approach to other neoplasias. It should be noted that spontaneous mammary tumors in intact adult female dogs are the most common neoplasia, which results in malignant tumors constituting up to 50 % of cases [3, 4].

Cell-specific differentiation markers are usually preserved during mammary gland oncogenesis. Affected area morphological assessment in combination with immunophenotypic (in this case, visual diagnostics methods are considered) differences in healthy and diseased tissue could be used to establish an accurate diagnosis [5, 6].

The work objective is to develop an information system in the form of a web application that allows identification of the cytological samples microscopic images, establishing an initial diagnosis and providing recommendations for its further confirmation on the basis of additional data. Such information system should provide users with prompt access to information from microscopes connected to the Internet, interact with the system in real time and add new information to the database.

Open-source information tools that allow analyzing and processing cytological materials are currently missing. A large number of Internet resources are known that contain rich databases of experimental (scientific) materials on cytology and diagnostics. However, most of these resources are associated with the medicine area, but not with veterinary.

Work [7] demonstrates relevance of design, development and application of intelligent information systems that make it possible to improve the quality of diagnoses; an example of solving this problem by using the Eidos-X++ artificial intelligence system is presented. Studies performed showed prospects of developing an information system for identifying cytological samples in order to diagnose cancerous diseases based on the neural network technologies.

This area of information system design and development is currently being expanded in medicine [8–13] and veterinary [14, 15]. For example, work [14] presented a computer neuroimitator of animal internal non-infectious diseases showing a fairly high reliability of the diagnosis.

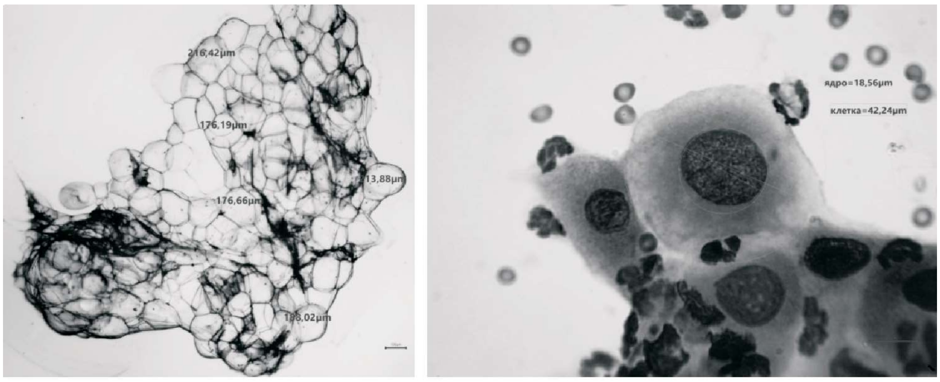
The system under development is designed to solve such problems in diagnosing cancer and would allow specialists regardless of their location to conduct analysis and establish a diagnosis. Development and implementation of such information systems in practice would make it possible to create the cytological samples database accessible to scientists and being constantly replenished, thereby expanding the capabilities of diagnostics, research and analysis of data on cancerous diseases.

Problem statement. Database of the information system should store data about patients, images of the affected area and microscopic images of cytological samples. Initial data that are entered into the information system for a preliminary diagnosis should include animal species, tumor location, tumor shape and its condition, as well as additional data. Information system should not only contain a database on existing patients, diagnoses established and prescribed treatment, but also provide an opportunity for the disease initial diagnosis (tumor nature identification) based on the patient anamnesis and cytological testing. Currently, the system database is being created on the basis of treatment practice of two veterinary clinics in the City of Saratov.

Since the initial diagnosis establishment is a task that is solved to a greater extent on the basis of the doctor intuition (there are no precise markers in the veterinary medicine, and the diagnosis is based on the doctor conclusions) and is difficult to formalize, it is planned to use neural networks in the information system to solve this problem. Input data for neural networks are images of cytological samples, which examples are shown in Fig. 1. Cytological picture assessment is based on the following characteristic features: preparation background, number and location of cells, size and shape of cells, nucleus, presence of multinucleated cells, fission entities (atypical mitoses), etc.

In identifying malignancy of a tumor, cell and nucleus characteristic features are primarily informative (Table 1). It is on the basis of analyzing these features that the initial conclusion is made about malignancy or non-malignancy of the tumor. Then, other features are evaluated (Table 2), which were considered significant. On this basis, the final conclusion is made on the type of tumor.

Input images are processed into black and white. This is because the RGB color space does not distinguish color information from other information,



a *b*

Fig. 1. Examples of cytological samples used in the information system: *a* are cells obtained by biopsy and taken from the mammary gland non-cancerous tumor (lipoma diagnosis); *b* are neoplastic cells obtained by biopsy and taken from the mammary gland tumor (preliminary diagnosis of adenocarcinoma)

such as lighting. If RGB is used to represent the image, all three channels should be taken into account in calculations, and this increases the number of layers, and at the same time brings excessive information that is not required in the framework of the problem being solved. In this case, it would be more appropriate to work with the HSV color space, which isolates color information into a single channel. Thus, the hue (H) channel is responsible for the color. We achieve our tasks in this way without loading the network under development with excessive data.

Table 1

Sample characteristic features for the ANN neural network ANN (example)

Object	Characteristic feature	Feature type
Tissue cell taken at biopsy	Cell size exceeds cell size of tissue that was the source of tumor growth	Yes/No
	Compliance with the form inherent to normal tissue cells	Yes/No
	Form	
	Round	Yes/No
	Spindle-shaped	Yes/No
	Nucleoplasmatic ratio	Calculated
	Polynuclear cells presence	Yes/No
Nucleus	*Form*	
	Roundish	Yes/No
	Oval	Yes/No

End of the Table 1

Object	Characteristic feature	Feature type
Nucleus	Polygonal (irregular)	Yes/No
	Size	
	Small (lymphocyte size approximately)	Yes/No
	Medium	Yes/No
	Location	
	Center	Yes/No
	Eccentrically	Yes/No
	Occupies actually the entire cell	Yes/No
	Naked nucleus, naked nucleus cell (cytoplasm is practically not identifiable)	Yes/No

Table 2

Sample characteristic features for the ART-1 neural network (examples)

Object	Characteristic feature	Feature type
Tissue cell taken at biopsy	*Cytosis*	
	High	Yes/No
	Low	Yes/No
	Cellular complexes	
	Glandular neoplasm presence	Yes/No
	Tubular neoplasm presence	Yes/No
Cytoplasm	*Volume*	
	Abundant	Yes/No
	Scarce	Yes/No
	Moderate	Yes/No
	Inclusions	
	Grains	Yes/No
	Pulverous granulosity	Yes/No
	Foamy cytoplasm	Yes/No
	Boundary accuracy	
	Accurate	Yes/No
	Irregular	Yes/No

Neural networks used. To solve the problems of medical diagnostics and forecasting, multilayer perceptron, which is a direct distribution network where neurons of one layer are sequentially connected to neurons of adjacent layers without recurrent connections, is most often used as a model of artificial neural networks. The most effective learning algorithm for a multilayer perceptron is the error back propagation algorithm [16].

Convolution type artificial neural networks (ANN) are usually used to evaluate images. The first convolution neural network was introduced by Yann LeCun [17]. This neural network consists of different types of layers: convolutional layers, subsampling layers and layers of the "normal" neural network, i.e., the perceptron. Neural networks of this type are used in optical image identification, image classification, object detection, semantic segmen-tation and other tasks [18, 19].

Another type of neural network architecture is the discrete neural network of the ART-1 adaptive resonance theory [20]. One of its advantages is the ability of additional learning in its functioning process, while storing previously collected information, as well as the ability to identify new information on its inputs that is not stored in its memory. Fig. 2 [20] presents a typical ART-1 diagram. There are three layers of neurons in the ART-1 neural network: S-neurons perceiving input information in the form of black and white images; interface Z-neurons; identifying Y-neurons. R, G_1, G_2 neurons are the governing ones.

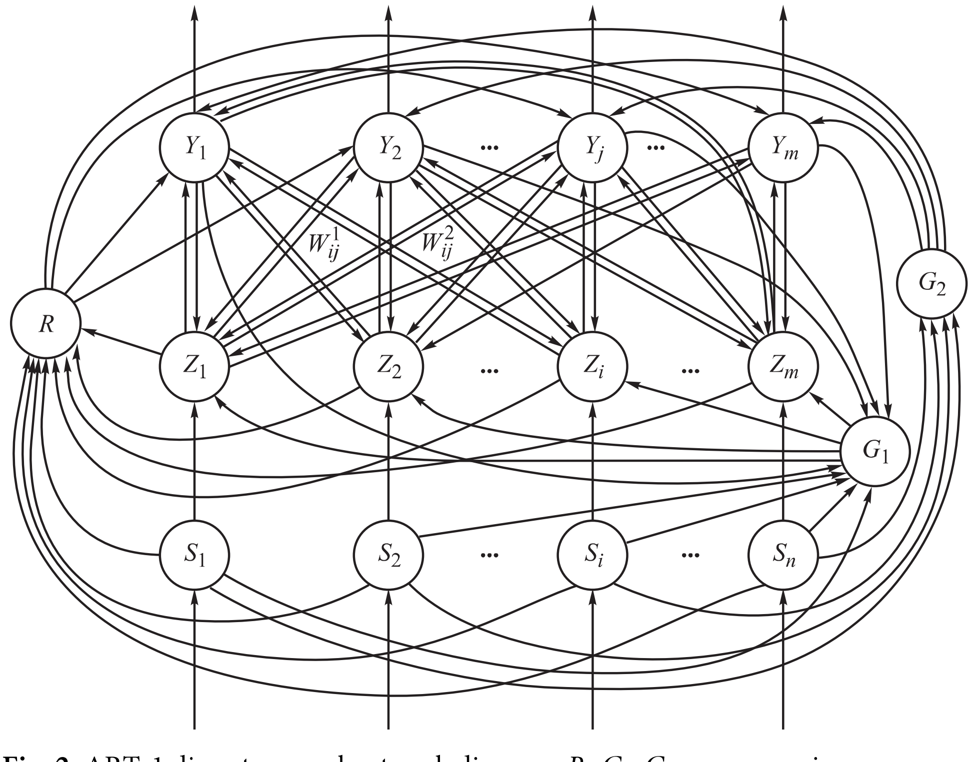


Fig. 2. ART-1 discrete neural network diagram: R, G_1, G_2 are governing neurons; S is input neurons; Z is interface neurons; Y is identifying neurons [19]

Fast learning method is usually used in discrete neural networks learning, where neuron connections equilibrium weights are determined by a single presentation of the input image [20].

Information system structure. Information system functioning algorithm diagram has the following form (Fig. 3): initial data input; cytological sample processing, initial data analysis and tumor malignancy determination; in the case tumor malignancy is diagnosed, repeated sample identification and its analysis based on additional data; diagnosis establishment and treatment recommendations.

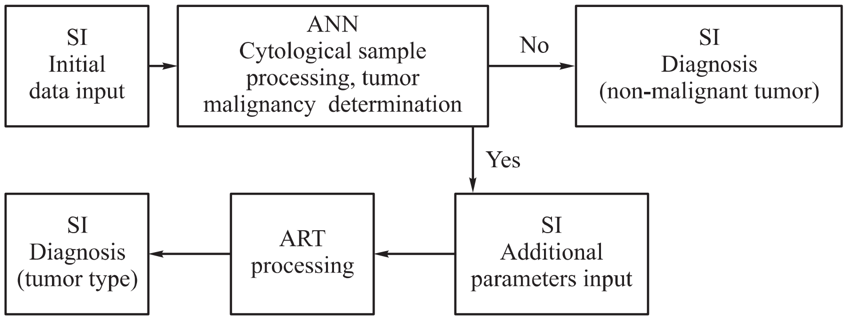


Fig. 3. Information system: SI is system interface; ANN is convolutional neural network; ART is adaptive resonance theory neural network

Diagnosis is established in three stages. At the first stage, image is identified and tumor malignancy is determined. At the second stage, repeated recognition and sample assignment to one of the most likely types of cancer are carried out. At the third stage, more accurate tumor classification, diagnostics and search for similar cases in the existing databases are performed.

First stage. ANN network used at this stage belongs to the convolutional neural networks (Fig. 4). ANN neural network determines, whether a tumor is malignant or not. If "No", then a tumor is diagnosed as non-cancerous, and its possible cause is determined (lipoma, etc.). If "Yes" (malignant tumor), then the image analysis based on additional data continues.

Two blocks could be distinguished in the ANN network. The first block is responsible for selecting characteristic features and consists of alternating

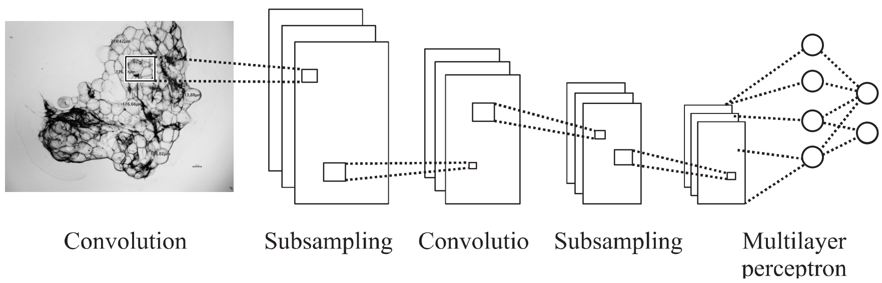


Fig. 4. ANN convolutional neural network structure: convolution are convolutional layers; subsampling are subdiscretization layers

convolution layers and subsampling. Each neuron of the network perceives a signal from the previous layer through the susceptibility field, which is a matrix of adjacent neurons in the previous layer connected to the neuron of the current layer with adjustable weights. These layers generate maps of abstract features for each of the classes in the learning process.

The second block solves classification problem and is a multilayer perceptron. Other neural networks, such as SVM or RBF, could be used [19].

Table 1 presents examples of features used in the ANN neural network learning. Features are encoded with Boolean values, and the encoded feature vector is sent to the neural network input.

The task of the ANN neural network is to identify malignancy of the tumor, the network output layer consists of two neurons. The number of hidden layers and neurons in these layers is determined during the entire system setup.

Second phase. At this stage, ART-1 discrete neural network is used, which allows, first of all, to perform additional image analysis using a larger number of characteristic features. Due to its additional learning ability, ART network also makes it possible to expand base of identifiable types of cancer.

Table 2 shows examples of features that are used in image analysis by the ART-1 network. Since this network is the simplest network based on adaptive resonance theory, it takes only binary values at the input.

Information system is being developed as a web application. Ruby on Rails framework written in Ruby was selected as the development tool, as well as other required tools in creating web applications: Ruby on Rails, HTML, JavaScript, Bootstrap, git.

Conclusion. Information system is designed to solve simultaneously several problems including tumor malignancy diagnostics and determination of the tumor type. Possibility of the system self-learning is planned, and for this purpose the ART discrete neural network is used. It is also assumed that the system would implement searching in the Internet by cytological samples with description thereof and generating knowledge basis on this foundation.

At present, information system structure and interface are developed, architecture is defined, and prototypes used in the system of artificial neural networks were created. To ensure potentiality in expanding its functionality, the system is being constructed on a modular principle, which allows, in particular, to use ART-2 and ART-3 type networks instead of the APT-1 neural network.

Issues of preparing data for the neural networks learning and developing algorithms for their learning would be considered in the subsequent publications.

Translated by D.L. Alekhin

REFERENCES

[1] Gutman D.A., Cobb J., Somanna D., et al. Cancer digital slide archive: an informatics resource to support integrated in silico analysis of TCGA pathology data. *J. Am. Med. Inform. Assoc.*, 2013, vol. 20, no. 6, pp. 1091–1098.
DOI: https://dx.doi.org/10.1136%2Famiajnl-2012-001469

[2] Bogoslavskiy C.N. Scope of artificial neural networks and prospect of their development. *Nauchnyy zhurnal KubGAU* [Scientific Journal of KubSAU], 2007, no. 27 (in Russ.). Available at: http://ej.kubagro.ru/2007/03/pdf/27.pdf

[3] Sleeckx N., de Rooster H., Veldhuis Kroeze E.J., et al. Canine mammary tumours, an overview. *Reprod. Domest. Anim.*, 2011, vol. 46, no. 6, pp. 1112–1131.
DOI: https://doi.org/10.1111/j.1439-0531.2011.01816.x

[4] Sorenmo K. Canine mammary gland tumors. *Vet. Clin. North Am. Small. Anim. Practice*, 2003, vol. 33, no. 3, pp. 573–596.
DOI: https://doi.org/10.1016/S0195-5616(03)00020-2

[5] Hoang M.P., Sahin A.A., Ordonez N.G., et al. HER-2/neu gene amplification compared with HER-2/neu protein overexpression and interobserver reproducibility in invasive breast carcinoma. *Am J. Clin. Pathol.*, 2000, vol. 113, no. 16, pp. 852–859.
DOI: https://doi.org/10.1309/VACP-VLQA-G9DX-VUDF

[6] Sorenmo K.U., Rasotto R., Zappulli V., et al. Development, anatomy, histology, lymphatic drainage, clinical features, and cell differentiation markers of canine mammary gland neoplasms. *Vet. Pathol.*, 2011, vol. 48, no. 1, pp. 85–97.
DOI: https://doi.org/10.1177%2F0300985810389480

[7] Lutsenko E.V., Korzhakov V.E. Realization of tests and supertests for veterinary and medical diagnostics in the Eidos-X++ system of artificial intelligence without programming. *Nauchnyy zhurnal KubGAU* [Scientific Journal of KubSAU], 2013, no. 89 (in Russ.). Available at: http://ej.kubagro.ru/2013/05/pdf/14.pdf

[8] Vyucheyskaya M.V., Kraynova I.N., Gribanov A.V. Neural network technologies in medical diagnosis (review). *Zhurnal mediko-biologicheskikh issledovaniy* [Journal of Medical and Biological Research], 2018, vol. 6, no. 3, pp. 284–294 (in Russ.). DOI: https://doi.org/10.17238/issn2542-1298.2018.6.3.284

[9] Al-Shayea Q.K. Artificial neural networks in medical diagnosis. *IJCSI*, 2011, vol. 8, no. 2, pp. 150–154.

[10] Das R., Turkoglu I., Sengur A. Effective diagnosis of heart disease through neural networks ensembles. *Expert Syst. Appl.*, 2009, vol. 36, no. 4, pp. 7675–7680.
DOI: https://doi.org/10.1016/j.eswa.2008.09.013

[11] Esteva A., Kuprel B., Novoa R.A., et al. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 2017, vol. 542, no. 7639, pp. 115–118.
DOI: https://doi.org/10.1038/nature21056

[12] Hirschauer T.J., Adeli H., Buford J.A. Computer-aided diagnosis of Parkinson's disease using enhanced probabilistic neural network. *J. Med. Syst.*, 2015, vol. 39, no. 11, art. 179. DOI: https://doi.org/10.1007/s10916-015-0353-9

[13] Aksenov S.V. An ensemble of convolutional neural networks for the use in video endoscopy. *Sovremennye tekhnologii v meditsine* [Modern Technologies in Medicine], 2018, no. 2, pp. 7–19 (in Russ.). DOI: https://doi.org/10.17691/stm2018.10.2.01

[14] Borisevich M.N. Computer neuroimitator of internal noncontagious diseases of animals. *Vestnik Vitebskogo gosudarstvennogo meditsinskogo universiteta* [Vestnik of Vitebsk State Medical University], 2017, vol. 16, no. 6, pp. 125–130 (in Russ.).

[15] Borisevich M.N. Neural network computer technologies in diagnostics of deseases. *Uchenye zapiski UO VGAVM*, 2006, vol. 42, no. 1-2, pp. 88–93 (in Russ.).

[16] Sikorskiy O.S. Review on convolution neural networks for problem of images classification. *Novye informatsionnye tekhnologii v avtomatizirovannykh sistemakh*, 2017, no. 20, pp. 37–42 (in Russ.).

[17] Svertochnaya neyronnaya set', chast' 1: struktura, topologiya, funktsii aktivatsii i obuchayushchee mnozhestvo [Convolution neural network, part 1: structure, topology, activation function and training set]. *habr.com: website.*
Available at: https://habr.com/ru/post/348000/ (accessed: 10.03.2019) (in Russ.).

[18] Zemlevskiy A.D. Study on architecture of convolution neural networks for problem of images recognition. *Vestnik nauki i obrazovaniya*, 2017, no. 6, pp. 36–43 (in Russ.).

[19] Dmitrienko V.D., Zakovorotnyy A.Yu. Diskette neural network ART, ideas immunocomputing. *Vestnik NTU "KhPI". Seriya: Informatika i modelirovanie* [Bulletin of the National Technical University "KhPI". Series of "Information and Modeling"], 2012, no. 62, pp. 25–63 (in Russ.).

[20] Abragin A.V. Prospects of development and application of neural networks. *Problemy sovremennoy nauki i obrazovaniya*, 2015, no. 12, pp. 12–15 (in Russ.).

Staroverova N.A. — Cand. Sc. (Eng.), Assoc. Professor, Department of Automated Data Acquisition and Processing Systems, Kazan National Research Technological University (K. Marksa ul. 72, Kazan, Republic of Tatarstan, 420015 Russian Federation).

Shustrova M.L. — Cand. Sc. (Eng.), Assoc. Professor, Department of Automated Data Acquisition and Processing Systems, Kazan National Research Technological University (K. Marksa ul. 72, Kazan, Republic of Tatarstan, 420015 Russian Federation).

Staroverov S.A. — Dr. Sc. (Biol.), Professor, Russian Academy of Sciences; Leading Research Fellow, Laboratory of Immunochemistry, Institute of Biochemistry and Physiology of Plants and Microorganisms, Russian Academy of Sciences (Entuziastov prospekt 13, Saratov, 410049 Russian Federation).

Dykman L.A. — Dr. Sc. (Biol.), Senior Research Fellow, Laboratory of Immunochemistry, Institute of Biochemistry and Physiology of Plants and Microorganisms, Russian Academy of Sciences (Entuziastov prospekt 13, Saratov, 410049 Russian Federation).

Please cite this article as:

Staroverova N.A., Shustrova M.L., Staroverov S.A., et al. Development of a neurocomputer modular information system for cancerous diseases diagnostics in animals. *Herald of the Bauman Moscow State Technical University, Series Instrument Engineering*, 2020, no. 2, pp. 75–84. DOI: https://doi.org/10.18698/0236-3933-2020-2-75-84