

## ON THE MATHEMATICAL METHODS IN BIOLOGY AND MEDICINE

*O. M. KLYUCHKO*

Kavetsky Institute of Experimental Pathology, Oncology and Radiobiology  
of the National Academy of Sciences of Ukraine, Kyiv

*E-mail: kelenaxx@rambler.ru*

Received 12.03.2017

The aim of the work was to analyze the range of mathematical methods and to choose the most prospective ones from the point of view of application in biology and medicine. After analyzing of approximately 200 current publications, a list of respective methods was completed. This list includes both the most recent, intensively developed methods as well as traditionally used ones — mathematical statistics, stochastic methods, regression analysis, and others. From the first group the methods of cluster analysis, artificial neural networks and image processing were subdivided. A description of each of these methods and examples of their application in practice are given. A separate group is dedicated to complex modern works, in which the problems requiring the complex application of several methods are present. In conclusions a brief assessment of the methods of cluster analysis, artificial neural networks, image processing methods are given as well as recommendations for their practical application.

**Key words:** methods, processing, data bases.

The practice of mathematical and program methods using in biotechnology (as in biology and medicine in general) has become widespread [1–21]. Moreover, the progress in these areas, including the works on biosensor elaboration, studies of molecular biological bases of bioprocesses, etc. need the development of new types of technical information systems based on electronic databases (DB) and the latest advances in information computer technologies (ICT). Actuality of such approaches is due to a number of factors specific to the development of biosensors and medical biotechnology in general, since they have a direct access to the practice of treatment and prevention of the development of many diseases.

However, the use of ICT in biotechnology meets the same difficulties that are characteristic for this practice in biology and medicine in general [1]. The large number of elements in living systems, complexity of the links between elements in such systems, their multifactor peculiarities, the unpredictability of interactions and influences of numerical

factors — all this limits the application of methods developed for technical branches and encourages researchers to look for new methods and to develop existing ones. There are also some subjective reasons for the development of new ICT methods in biology and medicine — the requirement of non-invasiveness for many diagnostic methods, the complexity of the process of results obtaining in clinics, in biotechnological studies, their multiplicity, since each modern scientific observation, monitoring, gives tens or even hundreds of results [1].

A large number of mathematical methods have been widely used in biotechnology. They enable the fast processing of large amounts of data (up to hundreds of thousands), which is especially important for modern biotechnology [1–20]. Modern modifications of such methods were carried out in parallel with the progress of modern ICTs and the development of biological and medical databases. This is especially important for contemporary practice, when the work becomes great scaled; for example, when it is necessary to decipher

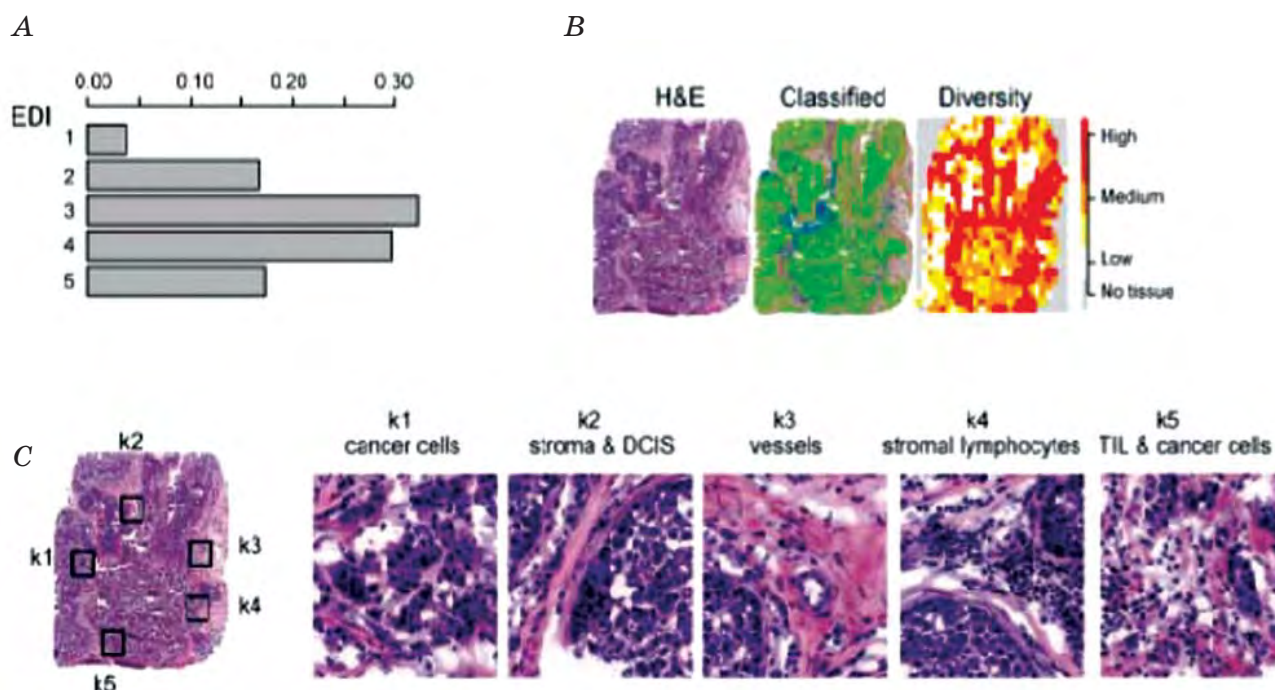
gene sequences, to process the images on hundreds or thousands of histological sections, to investigate the effects of many biologically active substances, etc. During the observation of numerical contemporary scientific sources we defined a set of mathematical analytical methods for data processing. Some of types of analysis have already become traditional: factor, dispersion, correlation, discriminant ones; as well as methods of classification, other methods of statistical data processing [1, 2]. Simultaneously, among the modern widely used methods, particular attention is paid to methods of cluster analysis [3–9], image processing [10–14], neural networks [15–20]. The development of diagnostic methods for image analysis and the spreading of video techniques, the recordings of number of histological photos in the database of modern medical information systems (ICs) contributed to the development of a powerful direction in mathematics and technology — the development of image processing methods [10–14]. Since in the format of this article it is impossible to consider all of the above mentioned methods, let's observe deeply the most popular data processing methods: cluster analysis, neural networks, and image processing. Here is a brief description of each of these methods and examples of its application.

*1. Use of cluster methods for data analysis* [3–9]. Cluster methods of data analysis can be used to solve biotechnology tasks concerning determination whether there are two (or more) biological objects form one object, or two (some) different objects in terms of mathematics (Fig. 1, 2). Application of cluster methods for DB elaboration in modern medical and biological research is quite common [1, 4–9]. Thus, the feasibility of these methods use was substantiated for computer diagnostics of diseases requiring the processing of several hundreds or more cell images, with the need to distinguish between cells with weak differences (in normal and pathological cases). These methods began to be used in biology since the end of XX century. However, at first these methods were used in very limited manner, a few specific tasks were solved with their help only. The idea of cluster methods application for the solution of tasks for object distinguishings during DB creating is quite attractive and new; for example when it is necessary to input information in different table fields about the objects that differ slightly one from another .

Let's analyse some examples. In D'haeseleer [4] studies was stated that clustering is often the first step in the analysis of gene expression. The investigated genes are subdivided using this method into smaller categories, the analysis of which makes it possible to reveal a comprehensive picture of studied phenomena. In Karpov et al. publication [5] a cluster analysis was done for the investigation of the similarity of associated with microtubules and serine-threonine protein kinase cell cycles for human and their plant homolog. 191 plant homolog of human proteinurases have been registered, which are involved in phosphorylation of microtubule proteins and cell cycle regulation. The protein kinases similarity was analyzed using the method of neighbors junction (NJ).

Separately let's observe the experience of elaborating of computer system for disease diagnostic, in which algorithms the cluster methods are used [1, 6]. In this case due to clustering methods, you can generate data split even without knowing of the details about DB domain or labels (for example, the names of diseases given by doctors during the diagnostics). By such approaches, in some cases, knowing about these new generated classes, you can diagnose the disease beginning. In publications [1, 6] the successful application of cluster methods was demonstrated, and various types of successful applications for data set were analyzed. Comparison of these methods was carried out by assessing of successful division of brain cell on groups; some cells had symptoms (characteristics) of poliomyelitis — differences in neuron nuclear materials. For comparison, 4 cluster methods were selected and evaluated in terms of application for this purpose: single-, and complete-linkage agglomerative hierarchical clustering, Ward's method and rough clustering. For comparison for each method single similarity measure, linear combination of the Mahalanobis distance for numerical attributes and Hamming distance for nominal attributes were suggested. The importance of cluster methods use was estimated through the quality of generated clusters, correlation between attributes used for high quality cluster generation and clinical experience.

*Characteristics of attributes during the elaboration of medical DB for objects with slight differences.* DB for clinical laboratory analysis is complicated, multi dimensional one with different types of attributes [1,6]. In general attributes of such DB are subdivided



**Fig. 1. Cluster analysis of histological images of tumors and cells in microsurrounding:**

*A* — diagram of different type cell numbers; *B* — results of images processing: sequences of stages, *C* — selection of 5 image fragments for processing and further demonstration of their increased versions [12]

on 2 types: numerical attributes and categorical attributes. The last ones it is possible to subdivide on ordered attributes and nominal attributes. For example such values as 12.114 mg, 0.15 unit/l — are numerical attributes with own origin. So, one can find the distance for two values basing on their origin. Such values as “dark”, “moderate”, “rough” belong to such attributes for which one can determine their order but cannot calculate distances in numerical values. Such values as “positive” (+), “negative” (–) are nominal attributes; one can describe them but cannot find their order or distance. It mean that for analysis of such DB it is necessary to have suitable similarity measures for characterization of differences between objects during diagnostics. Previously it was studied how similarity measures may be used in biological practice. Such often used similarity measures were studied [1, 6]:

- 1) Mahalanobis distance for numerical attributes;
- 2) Hamming distance for nominal attributes;
- 3) linear combination of Mahalanobis distance and Hamming distance for mixed attributes with four types of cluster methods.

Methods of cluster analysis were studied from the point of view of the best suitability for bioobject classes formation [1, 6]. Below is the list of mathematic methods — cluster methods — that were analyzed:

- 1) single-linkage agglomerative hierarchical clustering (AHC);
- 2) complete-linkage agglomerative hierarchical clustering (AHC);
- 3) Ward’s method;
- 4) rough clustering.

In process of experiments with classes the authors compared the differences between theoretically generated classes and classes which were diagnosed in practice [1, 4]. The necessity of such similarity measures was estimated due to such reasons:

- 1) quality of generated clusters;
- 2) importance of attributes for clinics which were used for the generation of high quality clusters.

Results of investigation of objects from observed DB with 140 objects and 32 attributes demonstrated that results of Ward’s method were the best from the point of view of clinical practice; used attributes were the most important for such practice. Below we suggest brief review of the work with practical use of Ward’s method for early diagnostics.

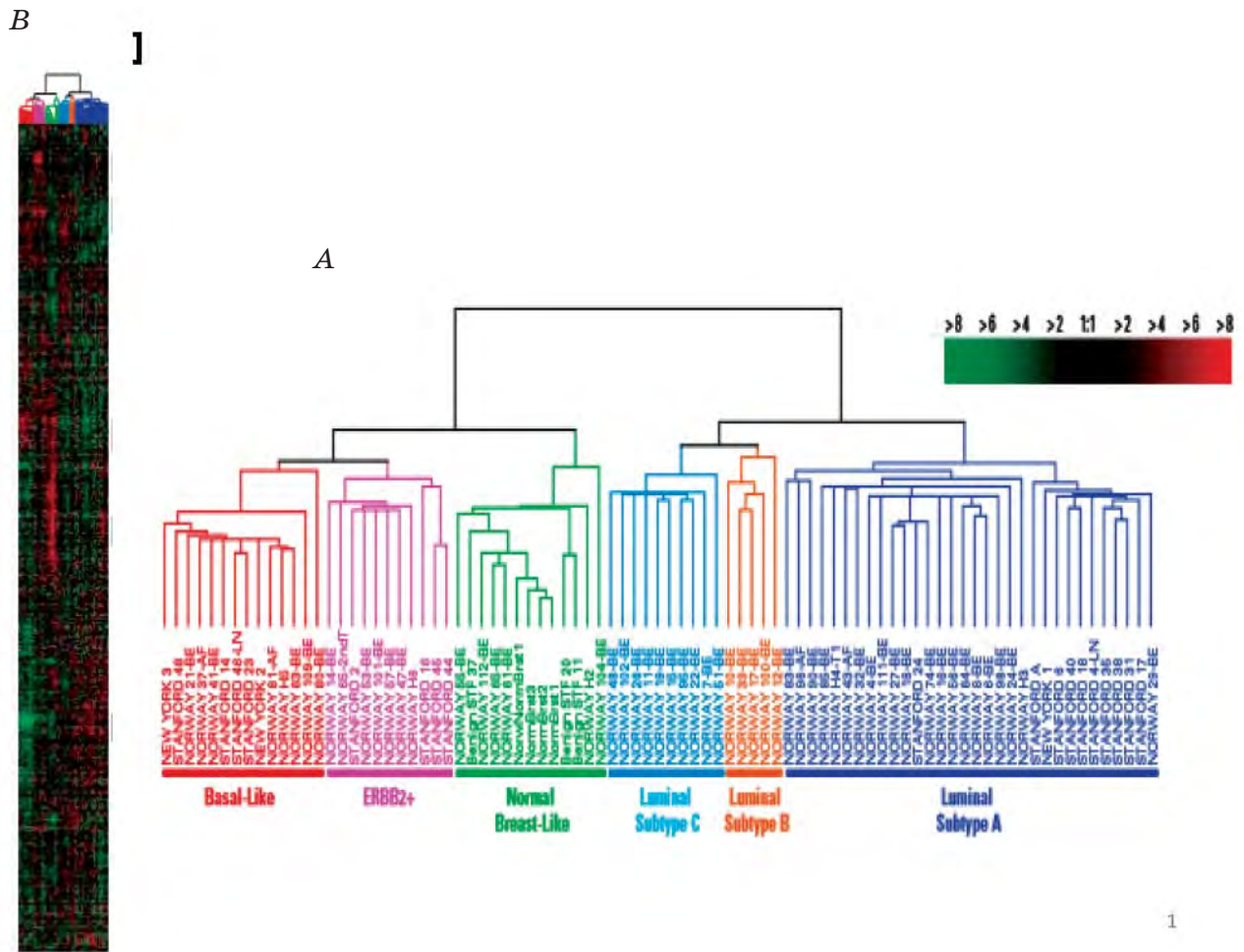


Fig. 2. Simultaneous use of 2 methods — image processing and cluster analysis: molecular classification and disease forecasting. Patterns of gene expressions are presented (85 samples). A — subdivision onto groups by cluster analysis, B — representation of experimental results [14]

Use of cluster analysis method (Ward's method) for early diagnostics. Above described results of analytic investigation were used for medical diagnostics because permitted to distinguish different types of objects. For example, Bozhenko [8] used Ward's method for finding, registration, and analysis of differences in complex sets of biochemical blood indices for mice with induced tumors (teratocarcinoma T-36), and for healthy ones in control. Analysis done by this method permits to distinguish sampling of biochemical blood indices (respectively, and all studied mice) according to analytical indices to 2 classes; healthy mice were in the first, and mice with transmitted tumors — in the second. Further, the last class of experimental mice was subdivided onto 2 subclasses according term after tumor transmission: before 30 days (beginning of mice death)

and with long terms of surviving. In such a way were demonstrated that standard laboratory blood indices permit to distinguish experimental mice groups with induced tumor at early stages of its development.

II. Methods for image processing [10–14]. One of the most common areas of application of these methods is diagnosis [1, 6]. There are two main approaches for describing of images (Fig. 2) [6, 7, 14]. The image can be described as a set of specific primitives from which it is formed. If the image is described by a two-dimensional array and each element represents a certain color description, then this image is called "raster". The bit of raster image is a pixel. A raster image can be represented as a rectangular matrix of points of different colors (Fig. 2). Matrix size is determined by the number of rows and columns of the image. During digital image

processing in experimental biology, it deals with raster images, for example, when working with histological section images or with chromatograms. Actually, any 2D array can be viewed as an image, and respective functions can be used for their transformations from corresponding software packages. In order to solve the application problems of image processing in biology, wavelet (VL) analysis is often used. VL-processing of optical signals in making of images provides the possibility of a sufficiently effective compression of such signals and their recovery with low quality losses, as well as solving of the problems of filtering signals. As a result, it is possible to obtain high-quality computer images that are necessary for high-quality diagnostics during operation planning. VL-analysis [6, 14] — is a special type of linear transformation of signals and physical data about the processes and physical properties of natural media and objects displayed by these signals. The basis of own functions, according to which VL-decomposition is carried out, has many specific properties and capabilities. VL-functions of basis give an opportunity to concentrate attention on those or other local features of analyzed phenomena. The fundamental value is the ability of wavelets to analyze nonstationary signals with the change of component content in time or space. The theory of wavelets is not a fundamental physical theory, but it provides a convenient and effective tool for many practical tasks, such as described below problem of pathology detecting as a result of image analysis [14, 15].

*Using of image processing methods for data analysis in chromatography and for samples identification.* One of the most progressive practical use of modern method — images processing is computer image analysis for chromatographic examination of biochemical sample contents [1, 6, 14]. As “samples” may be understood samples of genetic material, samples of characteristic proteins, “marked” or no, and etc. The sets of such images are ordered in DB, where “control” samples are also present. As “pairs” for analysis may be, for example, the results of chromatographic examination of biological liquids for healthy person or patient, etc. In case of specific differences in biochemical sample contents they will be revealed by chromatographic images. At following step of computer analysis this difference can be revealed; it

become the argument for diagnosis. Other usage of such methodic — sample identification in case of specific protein presence as well as marker substances, and etc. In such a way the problem of detecting of oncology cells among normal ones can be solved for early diagnostics in oncology.

*The use of method of images processing for the analysis of histological sections.* Detailed description of such analysis in [1, 6, 14] is given. Authors tried to optimize visual images analysis using the method of video frames texture color processing. The cells which computer system can identify usually have different sizes, forms, other characteristics. Under such conditions it is impossible to distinguish their details with enough quality by computer system. It was suggested to estimate image parts according to characteristic “normal”/“abnormal” one for different resolutions for diagnostic of high quality. In [6] the analysis according to many characteristics was done, using 2-dimensional discrete VL-transformation (2D-DVLT) for great number of images — colonoscopy video frames. Below the algorithm of such analysis is given.

*Algorithm of colonoscopy video frames analysis is given.* This method is based on the estimation of statistic measure covariations of the second order on DVLT of each video frame channel [4]. Different researchers suggested different covariation methods for the analysis of colored textures of biological object images; in majority of these cases the statistical information of the first order was taken into account. algorithm is realized in following four stages.

#### Stage 1

Let's suppose that  $I$  is a primary multychannel signal which exist on separated channels  $C_i$ ,  $i = 1, 2, \dots, c$ . Examined model which describes the colors of images has maximal numbers of channels equal to  $c = 3$ . Each of this channel was scanned in raster manner with window which has fixed dimensions equal to slide square.

#### Stage 2

In each window  $K$ -level of 2D-DVLT ( $j_0 = K$ ) is used in accordance with equation of VL-decomposition. This is a transformation of the results in new representation of primary window that consists on: sub-windows, according to different VL-bands  $D_{j_1}, D_{j_2}, D_{j_3}$ ,  $1 \leq j \leq j_0$ .

If to mark each band as  $B_b(k)$ , where  $b = 0, 1, 2, 3$  for  $k = K$  and  $b = 1, 2, 3$  for  $k < K$ , so, to these bands correspond  $B_0(k) = L_K$ ,  $B_b(k) = D_{Kb}$ ,  $b = 1, 2, 3$  for  $k = K$ , and  $B_b(k) = D_{jb}$ ,  $1 \leq j < j_0$ ,  $b = 1, 2, 3$ , for  $k < K$ , respectively.

### Stage 3

Cooccurrence measures are calculated for each sub-window  $B_b(k)$ ,  $b = 1, 2, 3$ ,  $k = 1, 2, \dots, K$ . Obtained in the result number of measures correspond to different channels and VL-bands. For one level of VL decomposition of colored window one will obtain 144 measures (16 cooccurrence measures  $\times$  3 VL band  $\times$  3 color channels) (16 cooccurrence measures  $\times$  3 VL band  $\times$  3 color channels), which includes 144-dimensional characteristic space.

### Stage 4

Necessary reduce of dimensions for characteristic space can be obtained if to suppose that covariation of these characteristics between different channels  $C_i$ ,  $i = 1, 2, \dots, c$ . For  $c = 3$  we will find covariation of colors VL for VL-band  $B_b(k)$ ,  $b = 1, 2, 3$ ,  $k = 1, 2, \dots, K$ , between two colors of channels  $C_l$  and  $C_m$ .

For example, automatic identification of early stages of cancer using markers may be suggested to be performing in following way [6]. First, ones can receive a video frame corresponding to the triangular signal with the values according to the additive color model ( $RGB$ ). These values are then pre-treated and transformed into a multicolored  $C_1C_2C_3$  model, which increases their ability to characterize the "normal" / "abnormal" areas of examined organism with suspicion of tumor.

The texture information present in each of the  $C_1$ ,  $C_2$ , and  $C_3$  channels in the video frame is computed in VL-domain, and then covariance measures of VL-colors (CCVL) are determined. The resulting measures form a set of characteristic vectors, which input the classifier of vector support machine (VSM). The result of the classification is a finite artificially generated frame formed by overlapping of windows that correspond to the original video frames of the studied tissue areas. The windows that were considered "normal" or "abnormal" respectively, are painted black and white. The same procedure is repeated for other received video frames.

*Peculiarities of images registration during examination of organism surface cell* [6]. This procedure is performed using a

standard video recording system for rectum cancer diagnostics. If ones receive color images of organism area suspected of having a tumor, such images provide important information for diagnosing of various types of pathology. A standard camera for such research consists of three light sensitive sensors known as  $i = R, G, B$  ( $RGB$  color processing systems, where  $R$  (red),  $G$  (green),  $B$  (blue)). Each of these sensors is characterized by a spectral frequency response, the function  $S_i(\lambda)$ , which shows sensor sensitivity to waves of different length  $\lambda$ . The spectral frequency response of these sensors corresponds to intervals overlapping in the visible spectrum with wavelength maxima in areas: red ( $R$ ), green ( $G$ ) and blue ( $B$ ). This overlap provides a correlation between  $RGB$  components. Image colors output by the endoscope depends on several factors, including: 1) spectral distribution of radiation  $E(\lambda)$ , which characterizes the energy emitted by the light source in the endoscope for each wavelength  $\lambda$ ; 2) spectral sensitivity of  $S_i(\lambda)$  sensors  $i = R, G, B$ , which characterizes the light energy sensitivity for each wavelength  $\lambda$ , 3) spectral way of images through lenses  $L(\lambda)$ , and 4) spectral reflexion of mucosal surface  $O(\lambda)$ .

When the light comes out of the light source, it is reflected by the mucosal surface and enters finally the sensor through the lenses. So, the spectrum modifies and the spectral characteristics at these stages are multiplied. The spectrum of the light beam entering each sensor receives a weight coefficient on each sensor through its response function (frequency characteristics)  $S_i(\lambda)$ , and each sensor covers its own interval of signal wavelengths. The value of intensity  $V_i(x, y)$  of sensor response to this beam at the selected point with coordinates  $(x, y)$  is calculated for the spectrum of event taking into account the band filter  $S_i(\lambda)$ :

$$V_i(x, y) = \int_w E(\lambda) S_i(\lambda) L(\lambda) O(\lambda, x, y) d\lambda$$

where  $w$  — is the number of wavelengths for which the sensor has non-zero sensitivity (visible spectrum). Multispectral images of various objects demonstrate that more than 3 components are required to accurately reproduce of their spectra. It has been demonstrated that the spectra of reflexion of the rectal membrane can be adequately

evaluated using *RGB* — channels of electronic endoscope output without significant loss of information necessary for diagnosis. The analogue output signals of image sensor are further transmitted to the standard video recording system input, and video frames are digitized so that they can be processed further by computer systems [1, 6].

*III. Method of neural networks* [15–20]. Artificial *neural networks* (ANN) are called machines that are assembled from simple processing elements. These are machines that can easily adapt to various tasks [19]. Such tools in modern science are successfully used to solve images recognition problems. For example, the above-mentioned image processing tasks can be very well complemented by methods of artificial neural networks for further recognition of images, which have signs of oncological pathologies. There are many classes of medical tasks exists that are successfully solved by methods of neural networks. Artificial neural networks are parallel computing devices that are associated with numerous simple processors that interact with each other [19]. Each such simple processor functions by receiving of periodic signals on input, and it periodically sends output signals to the network. The neural network is a set of elements (*neurons*) that are interconnected to provide interaction. The computing capabilities of neurons (*simple processors*) are limited by a certain rule of combining input signals and the activation rule, which allows you to calculate the output signal depending on input signals. The output signal is transmitted to another element with a certain weight coefficient, depending on the weight the signal can either amplify or fade. An attractive feature of this method is that although the computational capabilities of each neuron are limited, the combination of their large number in the network essentially increases their capabilities. The structure of links between the network elements reflects how they are combined, and for which tasks they are designed to solve.

In [20] the method of neural networks (NN) was used for the prognostic studies of locally-spread cervical cancer (LSCC). It is noted that the modern, more objective method for predicting of this disease course, a method with high potential is the mathematical technology of artificial neural

networks (ANN), the peculiarity of which is the parallel processing of information about the features of tumor process. Over the past decade there has been publications on the successful use of ANN in oncology, including for the prognosis of prostate cancer. Taking into account the fact that the treatment of patients with LSCC is not only an actual, but also quite complicated problem of gynecology, which requires the development of individual treatment tactic taking into account the biological characteristics of the primary tumor, the authors believe that the method of neural networks is a modern approach to the personalization of treatment of patients with LSCC. Authors hope that this method will provide an individual prognosis for this disease and determine the effectiveness of radical treatment.

Properties of artificial neural networks [20].

1. Ability to study. ANN can change their behavior depending on various factors and provide the necessary response with a large number of training algorithms.

2. The ability to synthesize and to recognize images. A neural network trained on limited number of data is able to summarize obtained information and to show credible results on data that was not used in the learning process.

3. Ability for abstractions. Some of the ANMs have the ability to find the essence among input signals.

4. Stability and speed. ANN for a huge number of interneuronal connections significantly accelerates the process of information processing.

ANN found application in many branches of medicine — for differential analysis of different types of cells, diagnosis of diseases, denture construction, optimization of transplantation time, planning of hospital costs, consultations in the absence of specialists. In the era of huge number of drugs, the modern analytical methods are needed to detect hidden causative relationships between one or more responses and a large set of properties, and ANN method provides such opportunities. ANM principles are: approximation, classification and images recognition; prognostication; identification and evaluation; as well as ability to simulate complex nonlinear relationships of different parameters and to process data quickly and in parallel.

In order to classify and to recognize images, ANN accumulates primary the knowledge on the basic properties of these characteristics, further defines the differences between the features that make up the basis for making of classification decisions. According to the existing sequences of previous states, the ANN can predict its future behavior.

In [3] we give an example of ANN method application for estimation of tumor weight and blood biochemical parameters. The works are based on regression using the ANN method (so-called method of multi-level perceptron (MLP)) and the discriminant classification method. Using these methods, the author attempted to search for more optimal set and complex of biochemical and hematological indicators for diagnostic problems solution. He did not succeed in significantly improving of the model developed for this purpose, but he succeeded in successfully solving an alternative task — estimating of time period since the tumor was transplanted. The results analyzed by him indicate that applied discriminant classification method can effectively divide mice with tumor into groups, depending on the time passed after tumor initiation; the overall classification efficiency was 82%. Thus, the used discriminant analytic method is effective in determining of such an important characteristic as the term after tumor initiation. Tasks of objects classification — the presence of classes of normal mice and mice with tumor, as well as the presence of subclasses, such as those corresponding to the week after tumor initiation, and others, can also be effectively solved by methods of neural networks. The authors note that the percentage of correct classification in such tasks was very high, only 3 objects were classified incorrectly.

*IV. Combined application of several mathematical methods* [1, 3]. Quite a lot of problems solved in biological research, in general, are so complex that they require a complicated mathematical apparatus (Fig. 1, 2) [1, 3]. In this section, we will observe the works that demonstrate the combined application of the above described methods: cluster analysis, image processing and neural networks. Combined application of neural network and image processing methods. In publication, to which we have already addressed above: “Classification and search of biomarkers in proteomics” ([\[bioinformatics.ru/Raznoe/Klassifikatsia-i-poisk-biomarkerov-v-proteomike.html\]\(http://bioinformatics.ru/Raznoe/Klassifikatsia-i-poisk-biomarkerov-v-proteomike.html\)\) \[3\]. Methods of mathematics and bioinformatics have been used to identify proteins that are relevant to potential biomarkers. It is shown that one of the most widely used algorithms for image recognition in proteomics is the method of support vectors \(\*Support Vector Machines, SVM\*\) \[1, 3\].](http://</a></p></div><div data-bbox=)

Particular attention is paid not so much to the construction of the most accurate diagnostic rule as to identification of proteins that correspond to potential biomarkers, which allows obtaining new data on the molecular mechanisms of the disease development. To do this, you need to find a minimal surplus set of variables (peaks of the mass spectrum or gel spots), which nevertheless allows you to achieve the correct diagnostic accuracy. For this purpose the authors use feature selection methods that reduce the size of the feature space. In typical case the reducing of number of features is from several hundred to a couple of dozen.

An example of a method for selecting features can be the recursive removal of features (*Recursive Feature Elimination, RFE*) [3]. At each step of this procedure the classifier is taught (for example, *SVM*), and each variable assigns the weight in accordance with the constructed computing rule. Variables of the least weight are excluded from further analysis. Next, the classifier is taught on the remaining variables, the weights are calculated again, and the process continues until a complete set of attributes is exhausted. As a result, it is possible to define a small set of variables that classify samples with sufficient accuracy (Fig. 1, 2). It is considered that these variables (that mean “proteins”) are biomarkers and they are suitable for their identification [3].

Thus, the amount of data used in biology and medicine is so significant that they form whole data arrays and require appropriate approaches to their processing [1]. The review has proven that modern mathematical and computer methods, as well as interdisciplinary approaches to data analysis, are required for their successful use. A number of mathematical methods, successfully applied in modern biotechnology, have been demonstrated. Also complex works that requires the use of combined sets of mathematical methods have been described.



## REFERENCES

1. *Klyuchko O. M.* Information and computer technologies in biology and medicine. *Kyiv: NAU-druk.* 2008. 252 p. (In Ukrainian).
2. *Kondrashov S. N., Gorokhova M. N.* Development of an algorithm for optimal control of the process of formaldehyde production. *Vestnik PNIPU "Chemical technology and biotechnology"*. 2016 No. 1, P. 718. (In Russian).
3. "Classification and search of biomarkers in proteomics". Available at <http://bioinformatics.ru/Raznoe/Klassifikatsiia-i-poisk-biomarkerov-v-proteomike.html> (In Russian).
4. *D'haeseleer P.* How does gene expression clustering work? *Nature Biotechnology.* 2005, No. 23, 1499–1501. doi:10.1038/nbt1205-1499.
5. *Karpov P. A., Nadezhdina E. S., Emets A. I., Blum Y. B.* Cluster analysis of similarity of microtubule-associated and cell cycle of human serine-threonine protein kinases with their plant homologues. *Bulletin of the Moscow University. Series 16: Biology. Moscow.* 2010, No 4. (In Russian).
6. *Iakovidis D. K., Maroulis D. E., Karkanis S. A.* Texture multichannel measurements for cancer precursors' identification using support vector machines. *Measurement.* 2004, — V. 36. P. 297313.
7. *Brenton J. D., Carey L. A., Ahmed A. A.* Molecular classification and molecular forecasting of breast cancer: ready for clinical application? *J. Clin. Oncol.* 2005, 23 (29), 7350–7360.
8. *Bozhenko V. K.* Multivariable analysis of laboratory blood parameters for obtaining diagnostic information in experimental and clinical oncology. *The dissertation author's abstract on scientific degree editions. Dc. med. study. Moscow,* 2004. (In Russian).
9. *Tashkinov A. A., Wildeman A. V., Bronnikov V. A.* Application of the classification tree method to predict the level of development of motility in patients with impaired motor functions. *Russian Journal of Biomechanics.* 2008, 12 (4), 8495. (In Russian).
10. *Vecht-Lifshitz S. E., Ison A. P.* Biotechnological applications of image analysis: present and future prospects. *Biotechnol.* 1992, 23 (1), 118.
11. *Goldys E. M.* Fluorescence Applications in Biotechnology and the Life Sciences. *USA: John Wiley & Sons.* 2009, 367 p.
12. *Perner P., Salvetti O.* Advances in Mass Data Analysis of Images and Signals in Medicine, Biotechnology, Chemistry and Food Industry. *Proceedings of the third International Conference, Leipzig, (Germany): Springer,* 2008. 173 p.
13. *Gavrilovich M.* Spectraimage processing and application in biotechnology and pathology. *Dissertation for Ph.D. Acta Universitatis Upsaliensis. Upsala.* 2011, 63 p.
14. *Shutko V. M., Shutko O. M., Kolganova O. O.* Methods and means of compression of information. *Kyiv: NAU-druk,* 2012, 168 p. (In Ukrainian).
15. *Natrajan R., Sailem H., Mardakheh F. K., Garcia M. F., Tape C. J., Dowsett M., Bakal C., Yuan Y.* Micro environmental heterogeneity parallels breast cancer progression: a histology–genomic integration analysis. *PLoS medicine.* 2016, 13(2), e1001961. <https://doi.org/10.1371/journal.pmed.1001961>.
16. *Rebello S., Maheshwari U., Dsouza S., DSouza R. V.* Back propagation neural network method for predicting Lac gene structures in Streptococcus pyogenes M Group A Streptococcus strains. *Int. J. Mol. Biol. Res.* 2 (4), 6172.
17. *Moghaddam M.G., Ahmad F.B.H., Basri M., Rahman M.B.A.* Artificial neural network modeling studies to predict the yield of enzymatic synthesis of betulinic acid ester. *Electronic Journal of Biotechnology,* 2010, 13 (3), 915.
18. *Montague G., Morris J.* Neural-network contributions in biotechnology. *Trends Biotechnol.* 1994, 12 (8), 31224.
19. *Kallan G.* Basic Concepts of Neural Networks. *Moscow: Williams.* 2001. 268 p. (In Russian).
20. *Kruzhanivska A.Ye.* Local widespread cervical cancer. *Ph.D. dissertation abstract. Ivano-Frankivsk,* 2015. (In Ukrainian).
21. *Onopchuk Yu. M., Biloshitsky P. V., Klyuchko O. M.* Development of mathematical models based on the results of researches of Ukrainian scientists at Elbrus. *Visnyk NAU,* 2008, No. 3, P. 146155. (In Ukrainian).

## ПРО МАТЕМАТИЧНІ МЕТОДИ У БІОЛОГІЇ ТА МЕДИЦИНІ

*О. М. Ключко*

Институт експериментальної патології,  
онкології та радіобіології  
ім. Р.Є. Кавецького НАН України,  
Київ

*E-mail: kelenaxx@rambler.ru*

Метою роботи було проаналізувати спектр математичних методів та обрати найбільш перспективні для застосування у біології і медицині. Після аналізу близько 200 сучасних публікацій у галузі біотехнології було складено перелік відповідних методів. Цей перелік містить як новітні, що інтенсивно розвиваються, так і традиційно застосовувані методи — математичної статистики, ймовірнісні, регресійного аналізу тощо. З першої групи методів було виокремлено методи кластерного аналізу, штучних нейронних мереж та обробки зображень. Дано характеристику кожного зазначеного методу та наведено приклади застосування їх на практиці. Окремою групою виділено складні сучасні роботи, вирішення задач у яких потребує комплексного застосування кількох методів. У висновках подається стисла оцінка методів кластерного аналізу, штучних нейронних мереж, методів обробки зображень і рекомендації щодо їх застосування.

**Ключові слова:** математичні методи, програмне забезпечення, бази даних.

## О МАТЕМАТИЧЕСКИХ МЕТОДАХ В БИОЛОГИИ И МЕДИЦИНЕ

*Е. М. Ключко*

Институт экспериментальной патологии,  
онкологии и радиобиологии  
им. Р.Е. Кавецкого НАН Украины,  
Киев

*E-mail: kelenaxx@rambler.ru*

Целью работы было проанализировать спектр математических методов и выбрать наиболее перспективные для применения в биологии и медицине. После анализа около 200 современных публикаций в области биотехнологии был составлен перечень соответствующих методов. Этот перечень включает как новые, интенсивно развивающиеся, так и традиционные применяемые методы — математической статистики, вероятностные, регрессионного анализа и другие. Из первой группы методов было отдельно выделены методы кластерного анализа, искусственных нейронных сетей и обработки изображений. Дана характеристика каждого указанного метода и приведены примеры применения их на практике. Отдельной группой выделены сложные современные работы, решение задач в которых требует комплексного применения нескольких методов. В выводах приведена краткая оценка методов кластерного анализа, искусственных нейронных сетей, методов обработки изображений и даны рекомендации относительно их применения.

**Ключевые слова:** математические методы, программное обеспечение, базы данных.