INFORMATION-EXTREME MACHINE LEARNING OF AUTONOMOUS UAV FOR VIDEO MONITORING OF THE REGION

Dovbysh Anatolii
Doctor in Technical Science, Professor, Sumy State University, Ukraine
ORCID ID: 0000-0003-1829-3318

Myronenko Mykyta
Ph.D. in Technical Science, Sumy State University, Ukraine
ORCID ID: 0000-0001-5005-1672

Kovalevskyi Serhii
Ph.D. Student, Sumy State University, Ukraine
ORCID ID: 0000-0002-1332-7913

Kraskovskyi Roman
Ph.D. Student, Sumy State University, Ukraine
ORCID ID: 0000-0002-2908-7707

Introduction

The widespread use of unmanned aerial vehicles (UAVs) for video monitoring of the observation region is aimed at solving problems in various sectors of the socio-economic sphere of society, such as cartography, ecology, the agricultural sector, geodesy, urban planning, etc. Giving a UAV the property of autonomy when performing a planned mission allows for expansion of its functionality and ensures protection against unauthorized intrusions. The main way to develop an autonomous UAV is to use intelligent information technologies based on machine learning for information synthesis of the airborne recognition system (ABRS). Research related to the modeling of intelligent systems for various purposes has found relatively wide coverage in the scientific and technical literature thanks to the ideas and scientific achievements of Ivakhnenko O. G., Schlesinger M. I., Vasyliev V. I., Kuzmin I. V., Shevchenko A. I., Bodiansky E. V., Dovbysh A. S., Duda R., Hart P., Watt D. and other native and foreign scientists. At the same time, the issues of increasing the functional efficiency of machine learning still remain insufficiently researched due to scientific and methodological complications associated with incomplete data certainty, significant overlap of recognition classes in the feature space, and the large capacity of the feature dictionary and alphabet of recognition classes that characterize ground objects. Therefore, the current task, the solution of which the research is aimed at, is the creation of an intelligent information technology of machine learning of the onboard system of an autonomous UAV for video monitoring of the area via an optical-electronic surveillance channel.

The purpose of the scientific work is to increase the functional efficiency of the onboard system of an autonomous UAV for semantic segmentation of a digital image of the observed region by developing a machine learning method that is invariant to the multidimensionality of the dictionary of recognition features and the alphabet of recognition classes, which is flexible during retraining.

1. Current status and prospects for the development of autonomous UAVs

Autonomous UAVs are used for tasks that are not feasible with manned aircraft for various reasons. Such tasks include: monitoring of land and water surfaces and airspace, environmental control, mapping, cargo delivery to hard-to-reach regions, archaeology, agriculture, etc.

The most common methods for recognizing images of various objects are convolutional neural networks (CNN), which belong to the third generation of neural networks [1]. Thanks to the implementation of the principles of deep machine learning, CNNs have gained great popularity in solving various tasks in the field of computer vision [2]. At the same time, machine learning of large CNN networks is a complex task and requires large databases and large training matrices.

The paper [3] describes the examples of the use of UAVs for scheduled bridge inspections, disaster management, power line surveillance, and traffic surveys. It contains a detailed description of the procedure for CNN machine learning on a set of aerial photographs. The object recognition results show that, given a representative training sample, CNN can detect and classify objects with a high level of accuracy (97.5%). Unfortunately, data on operational efficiency and computational efficiency are not provided. In addition, it is worth noting that the general disadvantages of neural-like structures are their sensitivity to the multidimensionality of the recognition feature dictionary, the need for a large number of images of ground objects, and the interactive mode of machine learning, which significantly limits the application of neural-like structures for machine learning of autonomous UAVs.

The tasks of recognizing objects in an image include the task of semantic segmentation of digital images, which consists in highlighting local areas (segments) in the image that correspond to different recognition classes that characterize terrestrial natural and infrastructure objects. Semantic image segmentation is needed in a number of areas: automatic creation of terrain maps [4], georesource analysis [5], urban planning [6], land use analysis [7], etc. But despite the large number of known algorithms and methods for classifying objects in images, the task of developing methods and software tools that allow automating this process and increasing the functional efficiency of the UAV on-board system for recognizing ground objects is relevant.

Semantic segmentation of a digital image of the observed region is one of the important functions of an autonomous UAV when performing missions for various purposes. For example, in paper [8], a computer vision algorithm was implemented to provide reliable information about the landing site in the event of a failure of the global positioning system GPS. In paper [9] machine learning of UAVs is considered for pesticide spray area recognition for images obtained from aerial reconnaissance. Images were acquired from low (5 m) and high (15 m) elevations for crops and orchards, respectively. 74.4% accuracy was achieved in recognizing pesticide-treated and untreated crop areas.

The capabilities of deep CNN machine learning in the task of constructing semantically segmented high-resolution maps of Arctic vegetation from hyperspectral satellite data were researched in paper [10]. The analysis of the results used existing vegetation maps of the west coast of Alaska, which contain tundra and forested areas. The constructed deep CNN allowed for hierarchical production of efficient generalized features for semantic classification from input satellite images. As a result, it was possible to achieve semantic segmentation accuracy for a given alphabet from four recognition classes from 66% to 96%, i.e. an average of 81%. It has been noted that more detailed hyperspectral databases are needed to increase the accuracy of semantic segmentation of Arctic vegetation. At the same time, information synthesis of "very deep" CNNs has high time costs.

The application of well-known methods of data mining, including CNN, for information synthesis of the autonomous UAV ORS does not always ensure a successful solution to the problem due to the following scientific and methodological limitations:

• arbitrary initial conditions for forming images of recognizable objects on the ground, determined by different aerial photography angles, UAV flight altitude, location of the ground object, etc.;

• intersection of recognition classes characterizing images of ground objects in the recognition feature space;

• multidimensionality of the feature dictionary and the alphabet of recognition classes;

• the influence of uncontrolled factors related, for example, to changes in weather conditions, lighting, camouflage, etc.

Thus, it is possible to draw conclusions:

1) Modern UAVs from the world's leading developers are still used mainly as repeaters of images of ground objects, which are analyzed by ground control station operators to solve assigned tasks.

2) The main way to create autonomous UAVs for recognizing terrestrial natural, infrastructure and small-sized objects is to develop new methods of intelligent data analysis based on

machine learning. In addition, solving this problem will increase the functional efficiency of the onboard system of an autonomous UAV for recognizing navigational obstacles and air threats.

It is known that the USA company "General Atomics" is developing a project with the code name GS-2, aimed at developing an intelligent autonomous UAV. The completion of this project is planned for 2030, which confirms the relevance and complexity of developing autonomous UAVs based on machine learning and pattern recognition.

To date, there is no agreed or legal definition of an “autonomous UAV”. The most common view is that an autonomous UAV is one that is capable of independently choosing its own course of action. In the technological aspect, the autonomy of a UAV is understood to mean the presence of an intellectual component capable of building decision-making rules through machine learning and even independently developing possible courses of action in response to new problems. Taking into account the functional capabilities of a UAV, we propose the following definition of successive levels of accumulation for autonomy:

1) The first level of UAV autonomy is ensured by the presence of an on-board autopilot connected to global positioning networks such as GPS.

2) The second level of autonomy is the ability of the UAV onboard system to recognize ground, surface, underwater, and aerial objects based on the results of machine learning with decision-making rules and transmit the relevant information via a crypto-protected channel to the ground control station.

3) The third level is the ability of the UAV to perform autonomously programmed actions.

4) The fourth level of autonomy is the ability of the UAV onboard system to self-learn to recognize ground, surface, underwater and aerial objects..

5) The fifth level is the ability of the UAV to perform autonomous video navigation along ground natural and infrastructure landmarks without the autopilot's connection to the global environment in order to increase the accuracy and efficiency of machine learning and bring its conditions closer to real ones.

When creating an autonomous UAV, the question of choosing a machine learning method inevitably arises. The selected machine learning method must meet the following basic requirements:

1) relevant input mathematical description of the learning ORS.

2) high functional efficiency of the machine learning method, the main components of which are accuracy, efficiency, practical invariance of decision rules to the multidimensionality of the feature dictionary and the alphabet of recognition classes.

Analysis of existing methods of intelligent data analysis shows that they do not ensure the adaptability of decision rules built based on the results of machine learning to arbitrary conditions for forming images of ground objects, the flexibility of ORS to retraining, and invariance to increasing the power of the feature dictionary and the alphabet of recognition classes due to the following main reasons of a scientific and methodological nature:

- arbitrary initial conditions for forming images of objects on the ground, determined by different aerial photography angles, aircraft heights, and the orientation and position of the object in the geospatial scene;
- intersection of recognition classes characterizing images of ground objects in the recognition feature space;
- multidimensionality of the feature dictionary and the alphabet of recognition classes;
- the influence of uncontrolled factors related, for example, to changes in weather conditions, lighting, camouflage, etc..

One of the promising ways of analyzing and synthesizing autonomous UAVs capable of learning for semantic segmentation of digital images of the observation region is the use of ideas and methods of information-extreme intelligence technology (IEI technology) of data analysis, which is based on maximizing the information capacity of the recognition system in the process of its machine learning [11, 12].

2. Basic principles of information-extreme intelligent data analysis technology

The main idea of data mining methods within the framework of IEI technology, as in artificial neural networks, is to adapt the input mathematical description in the process of machine learning to the maximum full probability of making correct classification decisions. But the advantage of information-extreme machine learning methods is that, unlike neural-like structures, they are developed within the framework of a functional approach to modeling cognitive processes inherent in humans when forming and making classification decisions, that is, they directly model the natural decision-making mechanism. This approach, unlike structural methods, allows information-extreme machine learning methods to provide flexibility when retraining the system by expanding the alphabet of recognition classes. Building decision rules within the framework of a geometric approach practically solves the problem of multidimensionality of the recognition feature dictionary, since modern computers are capable of processing structured vectors consisting of 2^{85} recognition features. In addition, such decision rules are characterized by high efficiency in making classification decisions when the system is operating in operational mode.

The conceptual idea of IEI technology is to defuzzify, in the process of machine learning, the a priori fuzzy partition of the feature space into recognition classes into a clear partition (Fig. 1).

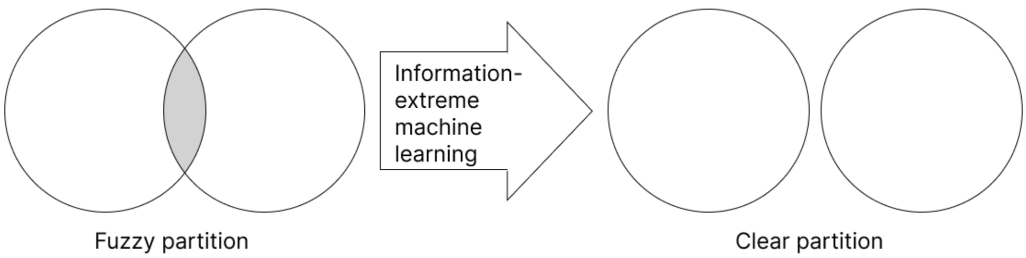


Fig. 1. *Conceptual model of information-extreme machine learning*

Machine learning methods within the framework of IEI technology belong to the class of radial basis methods for constructing separate hypersurfaces of recognition classes, which will be referred to in the text as containers.

The information-extreme machine learning algorithm is implemented using a multi-cyclic procedure for finding the global maximum of the information criterion averaged over the alphabet of recognition classes:

$$g_{\xi}^{*} = \arg\max_{G_{\xi}}\{\max_{G_{\xi-1}}\{...\{\max_{G_1 \cap G_E} \frac{1}{M}\sum_{m=1}^{M} E_m\}...\}\}, \tag{1}$$

where E_m is an information criterion for optimizing machine learning parameters of recognition class X_m^o; G_{ξ} is a permissible range of values of the ξ-th recognition feature; G_E is a working (permissible) area of determining the information criterion function for optimizing machine learning parameters.

In procedure (1), the inner loop implements a machine learning algorithm of the first level of depth, the functions of which are to calculate the information criterion at each step of machine learning, search for the global maximum of its function, and determine the optimal geometric parameters of the recognition class containers. At the same time, the following main restrictions are imposed on the information-extreme machine learning algorithm (1):

$$(\forall X_m^o \in \tilde{\Re}^{|M|})\left[X_m^o \neq \varnothing\right], \tag{2}$$

where $\tilde{\Re}^{|M|}$ is a fuzzy partition of the feature space into recognition classes, the cardinality of which is given by a cardinal number $Card\,\tilde{\Re} = M$;

$$(\exists X_k^o \in \tilde{\Re}^{|M|})(\exists X_l^o \in \tilde{\Re}^{|M|})\left[X_k^o \neq X_l^o \rightarrow X_k^o \cap X_l^o \neq \varnothing\right]; \tag{3}$$

$$(\forall X_k^o \in \tilde{\Re}^{|M|})(\forall X_l^o \in \tilde{\Re}^{|M|})\left[X_k^o \neq X_l^o \rightarrow KerX_k^o \cap KerX_l^o \neq \varnothing\right], \tag{4}$$

where $KerX_k^o$ is the kernel of recognition class X_k^o; $Ker\,X_l^o$ is the kernel for class X_l^o, closest to the recognition class X_k^o;

$$(\forall X_k^o \in \tilde{\Re}^{|M|})(\forall X_l^o \in \tilde{\Re}^{|M|})\left[X_k^o \neq X_l^o \rightarrow [(d_k^* < d(x_k \oplus x_l))\,\&\,(d_l^* < d(x_k \oplus x_l))]\right], \tag{5}$$

where d_l^* is the optimal radius of the container of recognition class X_l^o; $d(x_k \oplus x_l)$ is the code distance between vector x_k, which is averaged over the ensemble of feature vectors of the recognition class X_k^o, and the corresponding vector x_l of class X_l^o;

$$\bigcup_{X_m^o \in \tilde{\Re}} X_m^o \subseteq \Omega_B, \tag{6}$$

where Ω_B is a binary Hamming space.

For (3)–(5), $k \neq l$ і $k,l,m = \overline{1,M}$.

The depth of information-extreme machine learning is characterized by the number of system operating parameters that are optimized according to the information criterion. At the same time, the inner loop of procedure (1) optimizes the geometric parameters of the recognition class containers, which are used to build decision rules for classifying machine learning objects.

Thus, in the process of information-extreme machine learning, a purposeful search for the global maximum of a multi-extreme function of the statistical information criterion in the working (allowable) domain of its definition is carried out with the simultaneous restoration of optimal containers of recognition classes, which are built in the radial basis of the Hamming binary space of recognition features.

3. Formalized statement of the problem of information-extreme machine learning using the basic algorithm

The basic information-extreme algorithm for optimizing the spatiotemporal parameters of the functioning of an intelligent system is implemented in the first and second internal cycles of the machine learning procedure (1), i.e. it has two levels of depth.

The purpose of the inner loop of procedure (1) is:

- optimization of geometric parameters of recognition class containers;
- calculation of the information criterion for optimizing the parameters of the machine learning system;

- search for the global maximum of the information criterion in the working (permissible) domain of its function.

The purpose of the information-extreme machine learning algorithm of the second level of depth is to optimize the control tolerances for recognition features. Control tolerances within the framework of IEI technology play the role of quantization levels of recognition features of the input training matrix. As a result, by means of binary interval coding, the input training matrix is transformed into a working binary matrix specified in the Hamming feature space, which in the process of information-extreme machine learning is adapted to its maximum accuracy. The following recommendations are generally accepted for assessing the accuracy of machine learning, which is understood as the full probability of making correct classification decisions:

– if Precision = 0,5, then it's equal to a coin toss;
– if 0,5<=Precision<0,7, then it's poor classification;
– if 0,7<=Precision<0,8, it's an acceptable classification;
– if 0,8<=Precision<0,9, it's excellent classification;
– if Precision >= 0,9, it's an outstanding classification.

The basic algorithm of the second level of depth is a mandatory procedure of information-extreme machine learning. If the implementation of the basic algorithm does not allow to build error-free decision rules according to the training matrix, then it is necessary to increase the level of depth of information-extreme machine learning by optimizing additional parameters of the system's functioning, including the parameters of forming the input mathematical description.

Let's look at a formalized statement of the information synthesis problem of a learning-capable onboard system of an autonomous UAV for semantic segmentation of a digital image of a region within the framework of IEI technology. It's assumed that there is an alphabet $\{X_m^o \mid m=\overline{1,M}\}$ of recognition classes which characterize various terrestrial natural, infrastructural and other objects, including vehicles. For each recognition class, a three-dimensional training matrix $\|y_{m,i}^{(j)}\|$ of the brightness of the pixels of the receptor field of the image frames of the region was constructed , in which the row $\{y_{m,i}^{(j)} \mid i=\overline{1,N}\}$, where N is a number of recognition features, serves as an implementation for the recognition class, and the matrix column is a random training sample $\{y_{m,i}^{(j)} \mid j=\overline{1,N}\}$, consisting of n random values for the i-th feature of class X_m^o.

For a given level of depth ξ of information-extreme machine learning for ORS, the set $\{g_m\}$ of structured vectors of functioning parameters (hereinafter referred to as machine learning parameters) is presented. For the basic algorithm, the structure of the ORS machine learning parameter vector for a given alphabet of recognition classes $\{X_m^o\}$ has the form:

$$g_m = < x_m, d_m, \delta >, \tag{7}$$

where x_m is an averaged structured binary vector of features for class X_m^o; d_m is a radius for hyperspherical container of recognition class X_m^o, which is represented in the radial basis of the recognition feature space; δ is a parameter of the control tolerance field for the recognition feature, which is equal to half of the symmetric control tolerance field for the recognition features.

When performing missions of various purposes, the range of values of the radius of the container of recognition class X_m^o is given by the inequality $d_m < d(x_m \oplus x_c)$, where $d(x_m \oplus x_c)$ is an intercenter distance for recognition class X_m^o and its closest neighbor X_c^o, which is determined as a code distance between the corresponding averaged implementations x_m and x_c.

In the process of information-extreme machine learning for ORS it is necessary to:

1) Determine the optimal values of the machine learning parameters of vector (7) that provide the maximum value of the alphabet-averaged recognition of the information criterion classes, calculated in the working (allowable) region of its functions domain:

$$\overline{E}^* = \frac{1}{M}\sum_{m=1}^{M} \max_{G_E \cap G_d} E_m(d), \tag{8}$$

where $E_m(d)$ is information criterion for optimizing the parameters of the vector (7), calculated at the current radius d of the hyperspherical container for the recognition class X_m^o, which is represented in the radial basis of the Hamming feature space; G_d is a permissible doman of values for container radiuses.

2) Construct decision rules based on the optimal geometric parameters of the recognition class containers determined at the machine learning stage;

3) In the functional testing mode of the ORS machine learning algorithm, check the accuracy of the decision rules built based on the machine learning results using the training matrix;

4) At the examination stage, verify the functional efficiency of ORS machine learning and, if necessary, increase the depth of machine learning by optimizing other parameters of the system's functioning, including the parameters for forming the input mathematical description.

Thus, the task of information-extreme synthesis of the autonomous UAV ORS is to optimize the machine learning parameters (7) by finding the global maximum of the information criterion (8).

4. Informational criteria for optimizing parameters of information-extreme machine learning

Among the logarithmic statistical information criteria, the most widespread are the Shannon entropy measure and the Kullback-Leibler measure [13, 14].

The entropy criterion has the following normalized form [13]:

$$E = \frac{H_0 - H(\gamma)}{H_0}, \tag{9}$$

where H_0 is a priori (unconditional) entropy:

$$H_0 = -\sum_{l=1}^{M} p(\gamma_l)\log_2 p(\gamma_l); \tag{10}$$

$H(\gamma) -$ a posteriori conditional entropy, which characterizes the residual uncertainty after decision-making:

$$H(\gamma) = -\sum_{l=1}^{M} p(\gamma_l)\sum_{l=1}^{M} p(\mu_m / \gamma_l)\log_2 p(\mu_m / \gamma_l), \tag{11}$$

where $p(\gamma_l)$ is a priori probability of accepting the hypothesis γ_l; $p(\mu_m/\gamma_l)$ is the posterior probability of occurrence for event μ_m given the acceptance of hypothesis γ_l; M is a number of alternative hypotheses.

In practice, the following assumptions may apply::

- the solution is two-alternative (M=2);
- since the learning system operates under conditions of data uncertainty, the Bernoulli-Laplace principle justifies the adoption of equally probable hypotheses:

$p(\gamma_1) = p(\gamma_2) = 0{,}5$.

Then criterion (9) taking into account expressions (10) and (11) takes the form

$$E = 1 + \frac{1}{2}\sum_{l=1}^{2}\sum_{m=1}^{2} p(\mu_m/\gamma_l)\log_2 p(\mu_m/\gamma_l) \quad (12)$$

In the two-alternative solution (M=2), we will take the main hypothesis γ_1 about finding the value of the controlled recognition feature in the tolerance field δ and the alternative hypothesis γ_2. Let's divide the set of feature values into the domains μ_1 and μ_2. The field μ_1 contains values that are within δ, and μ_2 has values that are not within tolerance. Then it can be said that:

- Type I error $\alpha = p(\gamma_2/\mu_1)$;
- Type II error $\beta = p(\gamma_1/\mu_2)$;
- First confidence level $D_1 = p(\gamma_1/\mu_1)$; ;
- Second confidence level $D_2 = p(\gamma_2/\mu_2)$.

Since the first confidence and the first type error constitute one group of events, and the second confidence and the second type error constitute another group, the following relations hold:

$$D_1 + \alpha = 1;\ D_2 + \beta = 1. \quad (13)$$

Let's express the posterior probabilities $p(\mu_m/\gamma_l)$ through a priori ones using Bayes' formula, assuming that $p(\mu_1)=p(\mu_2)=0{,}5$, according to the Bernoulli-Laplace principle:

$$p(\mu_2/\gamma_1) = \frac{p(\mu_2)\rho(\gamma_1/\mu_2)}{p(\mu_1)p(\gamma_1/\mu_1) + p(\mu_2)p(\gamma_1/\mu_2)} = \frac{\beta}{D_1+\beta};$$
$$(\mu_1/\gamma_2) = \frac{p(\mu_1)\rho(\gamma_2/\mu_1)}{p(\mu_1)p(\gamma_2/\mu_1) + p(\mu_2)p(\gamma_2/\mu_2)} = \frac{\alpha}{\alpha+D_2};$$
$$p(\mu_2/\gamma_2) = \frac{p(\mu_2)\rho(\gamma_2/\mu_2)}{p(\mu_1)p(\gamma_2/\mu_1) + p(\mu_2)p(\gamma_2/\mu_2)} = \frac{D_2}{\alpha+D_2}. \quad (14)$$

After substituting conditional probabilities (14) into formula (12), we obtain the formula for calculating Shannon's entropy criterion:

$$E = 1 + \frac{1}{2}\left(\frac{\alpha}{\alpha+D_2}\log_2\frac{\alpha}{\alpha+D_2} + \frac{D_1}{D_1+\beta}\log_2\frac{D_1}{D_1+\beta} + \frac{\beta}{D_1+\beta}\log_2\frac{\beta}{D_1+\beta} + \frac{D_2}{\alpha+D_2}\log_2\frac{D_2}{\alpha+D_2}\right) \quad (15)$$

It is known that the Kullback-Leibler information measure is considered as the product of the logarithm of the likelihood ratio by the measure of deviations of probability distributions. In the following, the Kullback-Leibler measure will be called by the surname of the first author. Now the connection of the Kullback measure with the accuracy characteristics of information-extreme machine learning is shown in form (16) according to the concept of IEI technology:

$$K_m(d) = [P_{m,t}(d) - P_{m,f}(d)]\log_2 \frac{P_{m,t}(d)}{P_{m,f}(d)}, \quad (16)$$

where $P_{m,t}(d)$ is the total probability of making correct classification decisions when recognizing recognition class X_m^o implementations; d is a variable value of the radii of the recognition class containers; $P_{m,f}(d)$ is the total probability of making erroneous classification decisions when recognizing implementations of recognition class X_m^o.

Since machine learning methods within the framework of IEI technology implement the nearest neighbor principle, the full probabilities $P_{m,t}(d)$ and $P_{m,f}(d)$, respectively, will be represented for two alternative solutions through the accuracy characteristics of classification solutions:

$$P_{m,t}(d) = p_m D_{1,m}(d) + p_c D_{2,m}(d);$$
$$P_{m,f}(d) = p_m \alpha_m(d) + p_c \beta_m(d), \quad (17)$$

where p_m is an a priori probability of accepting the hypothesis that the implementation belongs to its recognition class X_m^o; $D_{1,m}(d)$ is a first confidence level calculated for the current radius d of hyperspherical container of recognition class X_m^o; p_c is an a priori (unconditional) probability of accepting the hypothesis that the implementation belongs to its neighboring recognition class X_c^o; $D_{2,m}(d)$ is a second confidence level; $\alpha_m(d)$ is a type I error; $\beta_m(d)$ is a type II error.

Since the first reliability and the first type error constitute one group of events, and the second reliability and the second type error constitute another group, the following relationship exists between the accuracy characteristics for two alternative solutions:

$$D_1 + \alpha = 1;\ D_2 + \beta = 1. \quad (18)$$

According to the Bernoulli-Laplace principle, we assume the prior probabilities to be the same, i.e. $p_m = p_c = 0{,}5$. Then, for two alternative systems of decision evaluations, the modified Kullback criterion (16) takes the form:

$$K_m(d) = 0{,}5\{[D_{1,m}(d) + D_{2,m}(d)] - [\alpha_m(d) + \beta_m(d)]\} \times \log_2 \left\{ \frac{D_{1,m}(d) + D_{2,m}(d) + 10^{-\lambda}}{\alpha_m(d) + \beta_m(d) + 10^{-\lambda}} \right\}, \quad (19)$$

where $10^{-\lambda}$ is a sufficiently small number that is entered to avoid division by zero.

Taking into account relations (18), formula (19) can be used in practice in the following modifications:

$$K_m(d) = \left[D_{1,m}(d) - \beta_m(d)\right]\log_2\left\{\frac{1+[D_{1,m}(d)-\beta_m(d)]+10^{-\lambda}}{1-D_{1,m}(d)+\beta_m(d)+10^{-\lambda}}\right\}; \quad (20)$$

$$K_m(d) = \left[D_{1,m}(d) + D_{2,m}(d) - 1\right]\log_2\left[\frac{D_{1,m}(d)+D_{2,m}(d)+10^{-\lambda}}{2-D_{1,m}(d)-D_{2,m}(d)+10^{-\lambda}}\right]; \quad (21)$$

$$K_m(d) = \{1-[\alpha_m(d)+\beta_m(d)]\}\log_2\left\{\frac{2-[\alpha_m(d)+\beta_m(d)]+10^{-\lambda}}{\alpha_m(d)+\beta_m(d)+10^{-\lambda}}\right\}. \quad (22)$$

Criteria (19) – (22) can be presented in normalized form:

$$E = \frac{K_m(d)}{K_{\max}}, \quad (23)$$

where $K_{\max}$ is a value of criteria (19) – (22) for $D_1^{(k)} = D_2^{(k)} = 1$ і $\alpha^{(k)} = \beta^{(k)} = 0$.

When analyzing the results of optimizing machine learning parameters, normalization of criteria is advisable, as it allows for comparative analysis of research results on the same measurement scale.

Let's look at the procedure for calculating the information criterion.

Since the information criterion is a measure of the diversity of machine learning objects, its calculation requires a training matrix consisting of implementation vectors of two recognition classes: $\{x_1^{(j)} \mid j=\overline{1,n}\} \in X_1^o$ і $\{x_2^{(j)} \mid j=\overline{1,n}\} \in X_2^o$.

$$D_{1,m}(d) = \frac{K_{1,m}(d)}{n_{\min}}; \quad \alpha_m(d) = \frac{K_{2,m}(d)}{n_{\min}}; \quad \beta_m(d) = \frac{K_{3,m}(d)}{n_{\min}}; \quad D_{2,m}(d) = \frac{K_{4,m}(d)}{n_{\min}}, \quad (24)$$

where $K_{1,m}(d)$ is the number of events that indicate that the implementation belongs to its recognition class X_1^o; $K_{2,m}(d)$ is the number of events indicating that the implementation of the class X_1^o does not belong to it; $K_{3,m}(d)$ is the number of events indicating that the implementation of the class X_1^o belongs to another class; $K_{4,m}(d)$ is the number of events indicating that the implementation of class X_1^o does not belong to another class; $n_{\mathbf{min}}$ is the minimal size of representative training sample.

After substituting the corresponding frequencies (24) into expression (12), we obtain a practical formula for calculating the entropy criterion for optimizing machine learning parameters for recognition class X_m^o under equally probable hypotheses of two alternative solutions:

$$E_m(d) = 1 + \frac{1}{2}\left(\frac{K_{2,m}(d)}{K_{2,m}(d)+K_4^{(k)}} \log_2 \frac{K_{2,m}(d)}{K_{2,m}(d)+K_4^{(k)}} + \frac{K_{2,m}(d)}{K_{2,m}(d)+K_{4,m}(d)} \log_2 \frac{K_{2,m}(d)}{K_{2,m}(d)+K_{4,m}(d)} + \right.$$
$$\left. + \frac{K_{3,m}(d)}{K_{1,m}(d)+K_{3,m}(d)} \log_2 \frac{K_{3,m}(d)}{K_{1,m}(d)+K_{3,m}(d)} + \frac{K_{4,m}(d)}{K_{2,m}(d)+K_{4,m}(d)} \log_2 \frac{K_{4,m}(d)}{K_{2,m}(d)+K_{4,m}(d)} \right). \qquad (25)$$

Accordingly, when substituting frequencies (23) into expressions (20 –22), we obtain practical formulas for calculating the modified Kullback criterion:

$$K_m(d) = \frac{1}{n_{\min}} \{K_{1,m}(d) - K_{3,m}(d)\} \log_2 \left\{ \frac{n_{\min} + [K_{1,m}(d) - K_{3,m}(d)] + 10^{-\lambda}}{n_{\min} - K_{1,m}(d) + K_{3,m}(d) + 10^{-\lambda}} \right\}; \qquad (26)$$

$$K_m(d) = \frac{1}{n_{\min}} [K_{1,m}(d) + K_{4,m}(d)] \log_2 \left\{ \frac{[K_{1,m}(d) + K_{4,m}(d)] + 10^{-\lambda}}{[2n_{\min} - K_{2,m}(d) - K_{3,m}(d)] + 10^{-\lambda}} \right\}; \qquad (27)$$

$$K_m(d) = \frac{1}{n_{\min}} [n_{\min} - (K_{2,m}(d) + K_{3,m}(d)] \log_2 \left\{ \frac{2n_{\min} - [K_{2,m}(d) + K_{3,m}(d)] + 10^{-\lambda}}{[K_{2,m}(d) + K_{3,m}(d)] + 10^{-\lambda}} \right\}. \qquad (28)$$

Thus, the information criteria considered above are functionals of both the accuracy characteristics of classification solutions and distance criteria, which allows them to be considered general validity criteria for machine learning, since they are a generalization of known statistical and distance proximity criteria.

5. Description of the basic algorithm for information-extreme machine learning based on linear data structure

The development of an information-extreme machine learning algorithm begins with the construction of its functional categorical model in the form of a directed graph of one-to-one mappings of sets used in the machine learning process by operators. The input mathematical description of an ORS trained on a linear structure of input data has the form

$$I_{ent} = < F, T, \Omega, K, Z, Y^{|M|}, X^{|M|}; f_1, f_2 > ,$$

where F is a space of factors that influence the machine learning object; T is a set of information retrieval points in time; Ω is a recognition feature space; K – a set of frames of a digital image of a region; Z is an alphabet of recognition classes; $Y^{|M|}$ is an input training matrix of size $Card\{X_m^o\} = M$; $X^{|M|}$ is a working binary training matrix; f_1 is an operator for forming a matrix $Y^{|M|}$ rom a source of information given by a Cartesian product $F \times T \times \Omega \times K \times Z$; f_2 is an operator for transforming a matrix $Y^{|M|}$ into a working matrix $X^{|M|}$.

Figure 2 shows a functional categorical model of information-extreme machine learning of the second level of depth with optimization of control tolerances for recognition features.

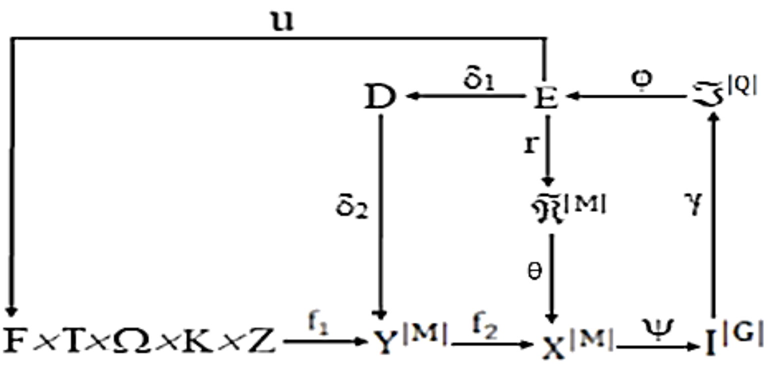


Fig. 2. *Functional categorical model of the basic information-extreme machine learning algorithm*

In Figure 2, the values of the information criterion, which are calculated at each step of machine learning, form the term set E. Operator $r: E \rightarrow \tilde{\Re}^{|M|}$ restores the containers of recognition classes in a radial basis of the binary feature space, which generally forms a fuzzy partition $\tilde{\Re}^{|M|}$. Operator θ projects the partition $\tilde{\Re}^{|M|}$ on a fuzzy partition of a priori classified implementations for training matrix $X^{|M|}$. Next, the operator $\psi: X \rightarrow I^{|G|}$, where $I^{|G|}$ is a set of hypotheses, checks the primary statistical hypothesis $\gamma_1 : x_m^{(j)} \in X_m^o$, operator γ determines a set of accuracy characteristics $\Im^{|Q\backslash}$, where $Q = G^2$, and operator φ calculates the value of information criterion E. The control tolerance optimization loop is closed through the term set D, the elements of which are the values of the control tolerances for the recognition features. Operator u regulates the machine learning process.

According to the functional categorical model (Fig. 2), the basic algorithm of informational=extreme machine learning of ORS with optimization of control tolerances is presented in the form of two cyclic procedures:

$$\delta^* = \arg\max_{G_\delta}\{\max_{G_E \cap G_d} \frac{1}{M}\sum_{m=1}^{M} E_m(d)\}, \qquad (29)$$

where G_d is a permissible domain for radiuses of recognition classes.

The implementation of the basic algorithm of information-extreme machine learning involves determining the basic recognition class, relative to the averaged brightness feature vector of which a system of lower and upper control tolerances for recognition features is set, which are optimized in the process of machine learning. There are several approaches to determining the basic recognition class, the main ones of which are the following:

1) Pragmatic, which consists in taking as the base the recognition class most desirable for the system developer. This approach is justified when solving control and diagnostic problems.

2) Heuristic, in which various hypotheses are tested regarding the choice of the base recognition class. An example of such an approach is the study of the influence of the variance of the training matrix of the brightness of the pixels of a digital image on the choice of the base recognition class.

3) The brute force method, which consists in the sequential implementation of the basic information-extreme machine learning algorithm for each recognition class from a given alphabet, which is taken as the basic one. In this case, the recognition class is taken as the basic one, for which the information criterion for optimizing machine learning parameters takes the largest maximum value in the working area of determining its function.

The input data of the basic information-extreme machine learning algorithm is an array of training matrices $\{y_{m,i}^{(j)} \mid m=\overline{1,M}; i=\overline{1,N}; j=\overline{1,J_{\max}}\}$ and the value of the parameter δ_H for field of normalized tolerances for recognition features, which specifies the range of control tolerance values. Assuming that the base class is X_m^o, the basic algorithm is as follows:

1) resetting the counter for classes: $m:=0$;
2) $m:=m+1$;
3) resetting the counter for changes in parameter δ: $\delta:=0$;
4) $\delta:=\delta+1$;
5) calculating the lower $A_{HK,i}$ and upper $A_{BK,i}$ level of control tolerances for the features:

$$A_{HK,i}=\overline{y}_{m,i}-\delta\text{; } A_{BK,i}=\overline{y}_{m,i}+\delta \tag{30}$$

where $\overline{y}_{m,i}$ is an averaged value for feature *i* of base class X_m^o;
6) resetting the counter for changes in a container radius: $k:=0$;
7) $k:=k+1$;
8) forming a tri-dimentional array for binary training matrix $\{x_{m,i}^{\{j\}}\}$, the elements of which are calculated by the principle

$$x_{m,i}^{(j)}[d]=\begin{cases}1, \text{ } if \text{ } A_{HK,i}[k]<y_{m,i}^{(j)}<A_{BK,i}[k];\\ 0, \text{ } if \text{ } else;\end{cases} \tag{31}$$

9) forming an binary array of averaged implementations $\{x_m\}$;
10) partitioning a vector set $\{x_m\}$ into pairs of nearest neighbors;
11) calculating the information criterion for optimization (1.2);
12) if $k<d(x_m\oplus x_c)$, then step 7, otherwise step 13;
13) if $\delta<\delta_H$, then step 4, otherwise step 14;
14) determining the maximal value of information criterion in bounds of its functions working (permissible) domain;
15) if $m\le M$, then step 2, otherwise step 16;
16) determining the global maximum of criterion $\overline{E}^*$ and optimal parameters: $\{x_m^* \mid m=\overline{1,M}\}$ are the averaged vectors for recognition class features; $\{d_m^* \mid m=\overline{1,M}\}$ are radiuses of recognition class containers; δ^* is a parameter of control tolerance field for the class features. According to (30) the lower $\{A_{HK,i}^* \mid i=\overline{1,N}\}$ and upper $\{A_{BK,i}^* \mid i=\overline{1,N}\}$ control tolerances are calculated;
18) STOP.

The optimal machine learning parameters $\{x_m^*\}$, $\{d_m^*\}$, $\{A_{HK,i}^*\}$ and $\{A_{BK,i}^*\}$ are stored in the knowledge base for building decision rules.

Based on the optimal geometric parameters of hyperspherical containers of recognition classes obtained in the process of machine learning, implicative decision rules are constructed, which in predicative form have the form

$$(\forall x_t \in \widetilde{\Re}^{|M|})(\forall X_m^o \in \widetilde{\Re}^{|M|})\Big(if \ [(\mu_m > 0)\ \& \ (\mu_m = \max_{\{m\}}\{\mu_m\}] \\ then \ x_t \in X_m^o \quad else \ xt \notin X_m^o \Big), \tag{32}$$

where x_t is a test binary implementation that is under recognition; μ_m is a membership function of implementation x_t to recognition class X_m^o.

In expression (31), the membership function for the hyperspherical container of the recognition class X_m^o is determined by the formula

$$\mu_m = 1 - \frac{d(x_m^* \oplus x^{(j)})}{d_m^*}. \tag{33}$$

The accuracy of machine learning is checked when the system is operating in the functional testing and exam modes. The purpose of functional testing is to check the accuracy of the decision rules (32) based on the implementations of the input training matrix. The purpose of the functioning of the learning system in the exam mode is to check the accuracy of machine learning based on the test examination implementation formed under conditions different from machine learning. At the same time, the exam algorithm is similar to the algorithm for the system's operation in the functional testing mode. Based on the results of checking the accuracy of the decision rules, a decision is made to stop machine learning or continue it by optimizing additional parameters of the system's operation, including the parameters for forming the input mathematical description of the system.

Since the decision rules in information-extreme machine learning methods are built within the framework of a geometric approach, they are highly efficient compared to other methods due to their low computational complexity and, in addition, are practically invariant to the multidimensionality of the recognition feature dictionary, since modern computer complexes are capable of processing implementations of recognition classes that contain up to 2^{85} features.

6. Example of implementing the basic information-extreme machine learning algorithm

The information-extreme machine learning algorithm considered above is implemented on the example of identifying frames of an image of a terrain (Fig. 3) obtained from aerial photography [15].

***Fig. 3.** General look of the area*

To form the input training brightness matrix, the terrain image (Fig. 3) was divided into frames with a pixel size of 50×50. The areas of interest were selected as a highway, which was characterized as a recognition class X_1^o; forest – recognition class X_2^o; field – recognition class X_3^o and grass cover – recognition class X_4^o. Selected image frames are shown in Figure 4.

a b c d

***Fig. 4.** Image of frames of interest zones: a – recognition class X_1^o; b – recognition class X_2^o; c – recognition class X_3^o; d – recognition class X_4^o*

The formation of the input training matrix was carried out by sequentially reading the brightness values of the receptor field pixels of each frame in the Cartesian coordinate system. Machine learning for ORS was carried out according to the procedure (29) with parallel optimization of control tolerances, which changed simultaneously for all recognition features with a given step. The selection of the base class was carried out within the framework of a heuristic approach. For this purpose, a variation series was previously constructed from the images of the frames shown in Figure 4 in order of increasing their brightness. In this case, the hypothesis was adopted that the recognition class that is inside the variation series should be taken as the basic one. In this case, the recognition class X_4^o – grass cover, was chosen as a base class, and its vector of averaged features was used to set the system of control tolerances for recognition features.

Figure 5 shows a graph of the dependence of the normalized entropy criterion (25) on the parameter δ of the control tolerance field for recognition features, obtained in the process of information-extreme machine learning with parallel optimization of control tolerances.

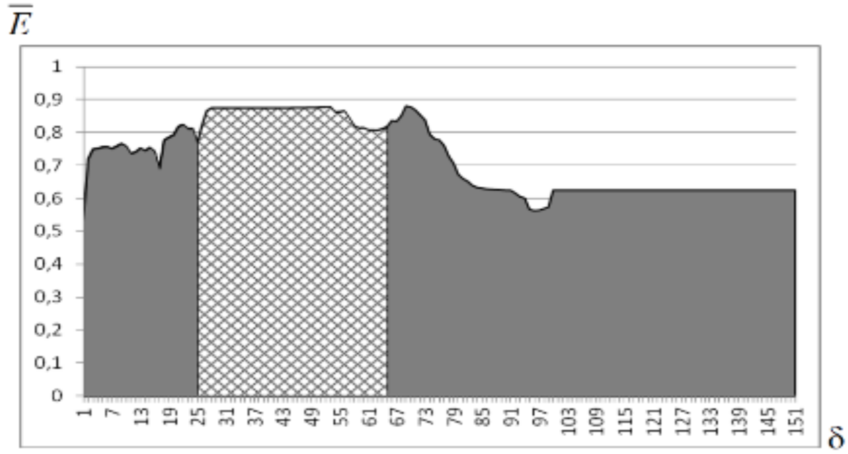


Fig. 5. *Graph of the dependence of the information criterion on the parameter of the control tolerance field*

In Figure 5, the shaded area of the graph indicates the working (permissible) area of determining the information criterion function for optimizing machine learning parameters, which meets the following conditions: $D_{1,m} > 0{,}5$ і $D_{2,m} > 0{,}5$, meaning that the first and second reliabilities exceed the first and second type errors, respectively. In addition, the right boundary of the working area is determined by preventing one class from "absorbing" another, meaning that $d_m < d(x_m \oplus x_c)$, which is a fundamental limitation when using radial basis functions. Analysis of Figure 5 shows that due to the presence of a "plateau" type area in the working area, the the maximum value of the information criterion (25) is ambiguous. Since the choice of the parameter δ for the control tolerance field significantly affects the degree of intersection of the recognition classes, in this case, to determine it, one should use the so-called distance relation proposed in [14] in the form:

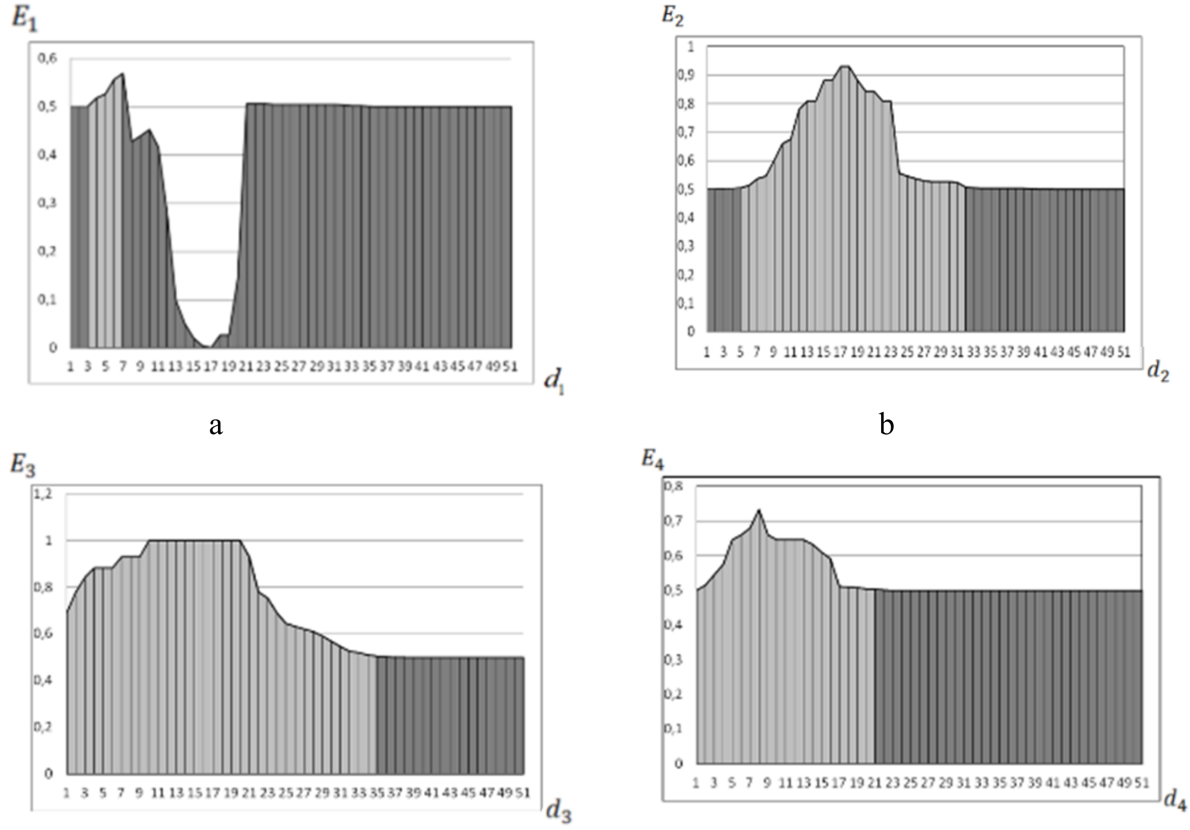
$$\eta_\delta = \frac{d_m^*}{d(x_m^* \oplus x_c^*)}, \tag{34}$$

where d_m^* is an optimal radius for the container of class X_m^o; $d(x_m^* \oplus x_c^*)$ is the code distance between the geometric centers of the nearest neighboring recognition classes X_m^o and X_c^o; x_m^* is an extreme value of the averaged structured feature vector of the recognition class X_m^o; x_c^* is an extreme value of the averaged structured feature vector of the recognition class X_c^o.

Given the minimum value of the relation (34) on a plateau-type area, the optimal parameter of the control tolerance field for recognition features is $\delta^* = 28$ (hereinafter the brightness gradations of the pixels of the receptor field for digital images). In this case, the maximum normalized value of the optimization criterion is equal to $\overline{E}^* = 0{,}88$.

Since within the framework of IEI technology, decision rules (31) are constructed within the framework of a geometric approach, for their construction it is necessary to know the optimal geometric parameters of the recognition class containers.

Figure 6 shows the results of optimization in the process of information-extreme machine learning of the radii of recognition class containers with an optimal control tolerance system.



a b c d

Fig. 6. *Graphs of the dependence of the entropy criterion (24) on the radii of the containers: a – recognition class* X_1^o *; b – recognition class* X_2^o *; c – recognition class* X_3^o *; d – recognition class* X_4^o

Analysis of Figure 6 shows that the optimal radii of recognition class containers are equal to: $d_1^* = 7$ (hereinafter Hemming distance code units) for recognition class X_1^o, $d_2^* = 17$ for recognition class X_2^o, $d_3^* = 10$ for recognition class X_3^o and $d_4^* = 8$ for recognition class X_4^o. Optimal containers of recognition classes were constructed with the following values of the optimization criterion and accuracy characteristics: for the recognition class X_1^o – E_1^* = 0,58 (first confidence level $D_1^* = 0{,}82$, type II error $\beta^* = 0{,}09$; for recognition class X_2^o – E_2^*=0,92 ($D_1^* = 0{,}96$; $\beta^* = 0{,}02$),for the recognition class X_3^o – $E_3^* = 1{,}00$ ($D_1^* = 1{,}00$; $\beta^* = 0$) and for the recognition class X_4^o – $E_4^* = 0{,}73$ ($D_1^* = 0{,}86$; $\beta^* = 0{,}03$).

In the exam mode, in order to verify the functional efficiency of ORS machine learning, the identification of the frame region shown in Figure 4 was carried out on a digital image using decision rules (32).

Figure 7 shows an electronic map of the area formed during the frame identification process with zones of interest marked according to the recognition class numbers. In this figure, the frames are numbered according to the recognition class numbers: 1 – highway; 2 – forest; 3 – field and 4 – grass cover.

Fig. 7. *Segmented virtual map of the region*

Analysis of Figure 7 shows that the most reliable frames identified were “field” – 0.94 and “grass cover” – 0.92, while the reliability of identifying the frames as “forest” and “highway” is 0.86 and 0.84, respectively. At the same time, it is worth emphasizing that the main way to increase the accuracy of frame identification within the framework of IEI technology is to increase the depth of machine learning.

Thus, the synthesized ORS has the ability to determine a priority area of interest. If such an area is a “highway,” then an algorithm for recognizing, for example, a wanted ground vehicle can be launched.

7. Hierarchical information-extreme machine learning with recursive data structure

The construction of decision rules within the framework of the geometric approach makes them practically invariant to the multidimensionality of the recognition feature space. At the same time, in practice, as a rule, there is a need to solve the problem of the multidimensionality of the recognition class alphabet, the power of which can significantly increase in the process of functioning of the intelligent system. In the case of increasing the power of the recognition class alphabet with an unchanged dimension of the recognition feature space, the degree of intersection of the recognition classes will also increase, which leads to a decrease in the total probability of making correct classification decisions. The main way to reduce the impact of the multidimensionality of the recognition class alphabet on the accuracy of machine learning is to transition from a linear data structure to a hierarchical one.

Let us consider a hierarchical data structure in the form of a binary tree, the peculiarity of which is the transfer of attributes of the vertices of the upper tier to the lower one. In contrast to the recursive hierarchical structure, such a structure will be called decursive. The construction of a decursive binary tree is carried out according to the scheme:

1) a variational series of recognition classes is constructed using the proximity criterion;

2) the variational series is divided into two groups, which respectively define two branches of the decursive binary tree;

3) as attributes of the vertices of the upper (first according to dendrographic classification) tier of the decursive tree, the training matrices of the neighboring boundaries for each of the groups of recognition classes are selected;

4) attributes of vertices of the upper-tier stratum are transferred to vertices of the corresponding lower-tier branch stratums;

5) as attributes of other vertices of the lower tiers of each branch of the tree, the training matrix of the nearest neighbor in its group of recognition classes is selected;

6) The tree construction continues until the final strata are formed, which contain the training matrices of all recognition classes.

As a criterion for the proximity of recognition classes when constructing a variation series, the Euclidean distance between the averaged implementations of the input training matrix of image brightness shown in Figure 4 has been used.

Thus, the binary decursive tree constructed according to the above scheme splits the given alphabet of recognition classes into strata, each of which contains two nearest neighboring classes. As a result, for each final stratum, the above-considered linear information-extreme machine learning algorithm can be applied to the two nearest neighboring recognition classes.

The input information description of the system, which is trained according to a hierarchical structure in the form of a decursive binary tree, is represented as the structure

$$I = < F, T, \Omega, K, Z, Y, H, Y_{h,s}^{|2|}, X_{h,s}^{|2|}; g_1, g_2, g_3, g_4 >,$$

where $Y^{|M|}$ is an input training matrix for a given alphabet of recognition classes with the size of $Card\{X_m^o\} = M$; H is a decursive binary tree; $Y_{h,s}^{|2|}$ is an input training matrix of two recognition classes for the s-th stratum of h-th tier in the tree; $X_{h,s}^{|2|}$ is the working training matrix is given in Hamming space; g_1 is an operator for forming the input training matrix $Y^{|M|}$; g_2 is an operator for constructing a recursive binary tree H; g_3 is an operator for forming the training matrix $Y_{h,s}^{|2|}$; g_4 is an operator for forming the training matrix $X_{h,s}^{|2|}$.

The functional categorical model of information-extreme machine learning based on the hierarchical data structure is shown in Figure 8.

$$\begin{array}{ccccccccc}
F\times T\times\Omega\times K\times Z & \xleftarrow{u} & & E & & \xleftarrow{\varphi} & & \Im^{|Q|} \\
\downarrow g_1 & & \swarrow\delta_1 & & \downarrow r & & & \downarrow\gamma \\
& & D & & \Re_{h,s}^{|2|} & & & \\
& & \swarrow\delta_2 & & \downarrow\xi & & & \\
Y^{|M|} \xrightarrow{g_2} H & \xrightarrow{g_3} & Y^{|2|}_{h,s} & \xrightarrow{g_4} & X^{|2|}_{h,s} & & \xrightarrow{\Psi} & I^{|G|}
\end{array}$$

Fig. 8. *Functional categorical model of hierarchical machine learning based on decursive data structure*

Analysis of the functional categorical model (Fig. 8) shows that it differs from the model (Fig. 2) in the way of forming the input information description of the learning system. At the same time, the partition $\tilde{\Re}_{h,s}^{|2|}$, constructed in the process of information-extreme machine learning for recognizing each stratum of a decursive tree, covers binary implementations of the working training matrix in the feature space $X_{h,s}^{|2|}$ for corresponding stratum. Classification operator $\psi : \tilde{\Re}^{|2|} \to I^{|G|}$, where $I^{|G|}$ is a set G of statistical hypotheses, checks the primary statistical hypothesis about implementation $x_{,m,h,s_s}^{(j)}$ belonging to recognition class X_{m,h,s_s}^{o}.

The use of a hierarchical data structure in the form of a decursive binary tree allows the automatic split for alphabet with a high number of recognition classes into pairs of nearest neighbors. As a result, linear multi-class information-extreme machine learning is reduced to a sequence of two-class ones. In turn, the construction of error-free decision rules according to the training matrix is achieved by increasing the depth of machine learning through optimizing additional parameters of the ORS operation, including the parameters of the formation of the input mathematical description. At the same time, an important task arises: determining the limitations for automatic transition from linear to hierarchical information-extreme machine learning algorithms. To study this problem, the following working hypothesis was formed. Since, due to the nature of the information criterion, the minimum number of recognition classes for machine learning cannot be less than two, the alphabet is limited to three recognition classes. Next, for the selected alphabet of three recognition classes, the linear and hierarchical information-extreme machine learning algorithms of the same depth level are sequentially implemented with further comparison of the experimental results. If the results are the same, then the power of the initial alphabet should be increased by one recognition class and so on until the value of the alphabet-averaged information criterion for optimizing machine learning parameters according to the hierarchical structure starts increasing. If, for an alphabet of three recognition classes, the value of the averaged information criterion obtained according to the hierarchical structure exceeds the value obtained according to the linear structure, then it should be concluded that for any alphabet of recognition classes, it is advisable to carry out information-extreme machine learning of ORS using the hierarchical structure in the form of a decursive binary tree. The objects of machine learning considered are shown in Figure 4 recognition classes X_2^o, X_3^o і X_4^o.

As a criterion for optimizing machine learning parameters according to the hierarchical data structure, we will use the modified Kullback information measure (27). Initially, ORS machine learning was carried out using a linear information-extreme machine learning algorithm according to the iterative procedure (29) with parallel optimization of control tolerances for recognition features. In the process of machine learning, ORS was previously defined as a basic recognition class X_2^o (grass cover), with a system of control tolerances set in relation to the averaged feature vector, according to the algorithm described above. Figure 9 shows a graph of the dependence of the normalized information criterion (27) on the parameter δ of the control tolerance field for recognition features, obtained in the process of information-extreme machine learning using the linear procedure (29).

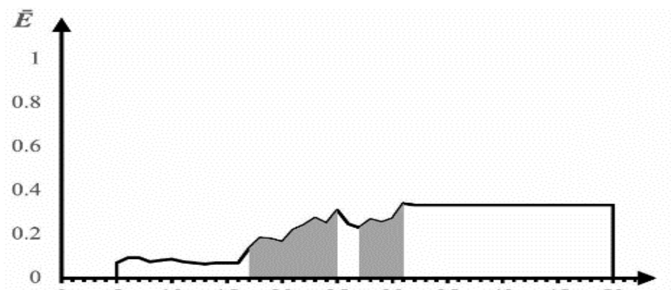


Fig. 9. *Graph of the dependence of the information criterion on the parameter of the control tolerance field*

In Figure 9 and further in the text, the dark areas of the graph indicate the working (permissible) area of determination of the information criterion function (27), which meets the conditions: $D_{1,m} > 0{,}5$ i $D_{2,m} > 0{,}5$, meaning that first and second confidence levels are higher that type I and type II errors accordingly. In addition, the right boundaries of the working areas are determined by preventing the "absorption" of the nearest neighbor kernel by one class, i.e., it is true that $d_m < d(x_m \oplus x_c)$.

To implement the hierarchical information-extreme machine learning algorithm, a variational series of images was constructed by increasing their average brightness as it corresponds to their position in Figure 4. Then, according to the above algorithm for constructing a decursive binary tree, the variation series was divided into two groups, which included recognition classes $\{X_2^o; X_3^o\}$ and $\{X_4^o\}$, accordingly. Figure 10 shows a decursive binary tree constructed for a given alphabet of recognition classes.

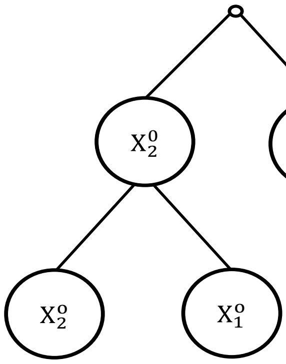


Fig. 10. *Recursive binary tree for three recognition classes*

Figure 11 shows a graph of the dependence of the averaged normalized criterion (27) on the parameter of the control tolerance field for recognition features for the recognition classes of the stratum of the upper tier of the decursive tree (Fig. 10), obtained in the process of two-class information-extreme machine learning for ORS.

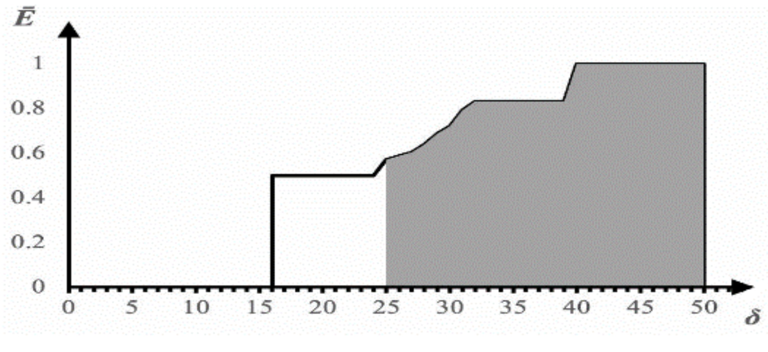


Fig. 11. *Graph of the dependence of the information criterion (27) on the parameter of the control tolerance field on recognition features for the upper tier stratum*

Analysis of Figure 11 shows that due to the presence of a “plateau” type area in the working area, to determine the optimal parameter δ of the control tolerance field, the relation (34) should be minimized. In this case, the optimal parameter of the control tolerance field for recognition features equals $\delta^* = 42$. At the same time, the normalized averaged optimization criterion is equal to the maximum limit value $\overline{E}^* = 1$.

To construct decision rules (31) for the upper stratum, knowing of the geometric parameters of the recognition class containers is necessary. Figure 12 shows the dependence of the normalized criterion (27) on the radii of the recognition class containers X_2^o and X_3^o.

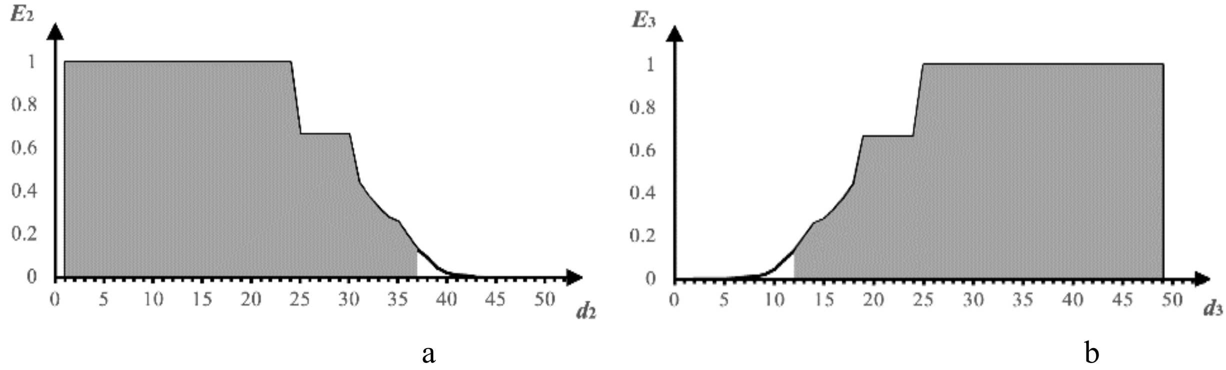


Fig. 12. *Graph of the dependence of the normalized criterion (27) on the radii of the containers for the upper stratum: a – recognition class* X_2^o *; b – recognition class* X_3^o

Since the graphs in Figure 12 have plateau-like areas, the minimum value of relation (34) is achieved at optimal container radii of $d_2^* = 19$ and $d_3^* = 32$ respectively. Figure 13 shows a graph of the dependence of the averaged criterion (27) on the parameter of the control tolerance field on the recognition features of the recognition classes of the lower tier stratum.

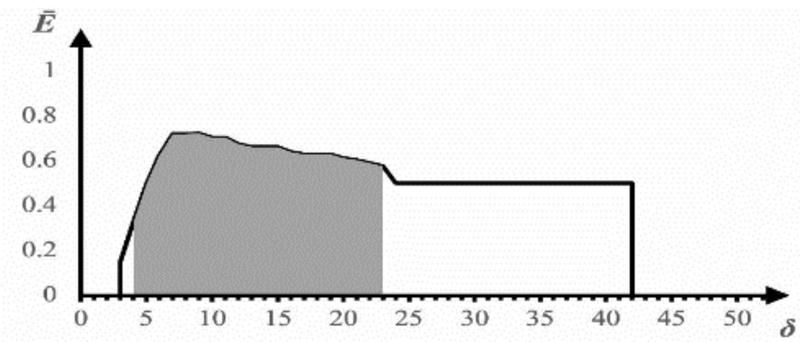


Fig. 13. *Graph of the dependence of criterion (27) on the parameter of the control tolerance field on the recognition features of the recognition classes of the lower tier stratum*

Analysis of Figure 13 shows that the information criterion does not reach its maximum limit value during machine learning, which makes it necessary to increase the depth of machine learning. For this purpose, a two-class machine learning algorithm with sequential optimization of control tolerances was implemented for the lower-tier stratum recognition classes according to the procedure

$$\{\delta_i^* \mid i = 1, N\} = \arg\left[\bigotimes_{l=1}^{L} \max_{G_\delta}\{\max_{G_E \cap G_d} \overline{E}_l(d)\}\right], \tag{35}$$

where $\overline{E}_l(d)$ is the average value of the information criterion for optimizing machine learning parameters of an intelligent system, calculated during the optimization of control tolerances for i -th recognition feature at l -th iteration of procedure for optimizing the control tolerance system; G_{δ_i} is a field of permissible values for control tolerances of *i*-th feature; $\otimes$ is the symbol of repetition; L is the number of iterations for optimizing the control tolerances; N is the number of recognition features.

Following are the main steps for implementing the algorithm (35) for sequential optimization of control tolerances for recognition features.

1. Initialization of the run counter of the machine learning parameter optimization procedure: $l := 0$.
2. $l := l + 1$.
3. Initialization of the recognition feature counter: $i := 0$.
4. $i := i + 1$.
5. Determining the extreme value of the control tolerance field parameter $\delta_i^*(l)$ according to the procedure (7.3)
6. Comparison: if $i \le N$, then step 4, otherwise step 7.
7. Calculating the value of information criterion $\overline{E}_l(d)$, averaged for the alphabet of recognition classes.
8. If $\{[\overline{E}^{(s)} < E_{\max}] \,\&\, (l < L)\}$, where L is a set number of iterations, then step 2, otherwise step 9.
9. Optimization for the field of control tolerances $\{\delta_i^*(L) \mid i = \overline{1, N}\}$ for recognition features.
10. The optimal lower and upper control tolerances for recognition features are calculated using the formulas:

$$\{A_{HK,i}^* \mid i = \overline{1, N}\} = A_{0,i} - \delta_i^*(L); \quad \{A_{BK,i}^* \mid i = \overline{1, N}\} = A_{0,i} + \delta_i^*(L),$$

where $A_{0,i}$ is a nominal averaged value for the i -th recognition feature.

11. Optimal machine learning parameters are memorised:

$\{x_m^* \mid m = \overline{1, M}\}$ are the optimal average feature vectors of recognition classes from a given alphabet;

$\{d_m^* \mid m = \overline{1, M}\}$ are the optimal radii of recognition class containers;

$\{A_{HK,i}^* \mid i = \overline{1, N}\}, \{A_{BK,i}^* \mid i = \overline{1, N}\}$ are the optimal lower and upper control tolerances for recognition features.

12. STOP.

At the same time, control tolerances for recognition features obtained as a result of parallel optimization were taken as starting ones. Figure 14 shows a graph of the change in the averaged normalized criterion (27) in the process of machine learning with sequential optimization of control tolerances, in which the quasi-optimal control tolerances obtained through their parallel optimization as starting values.

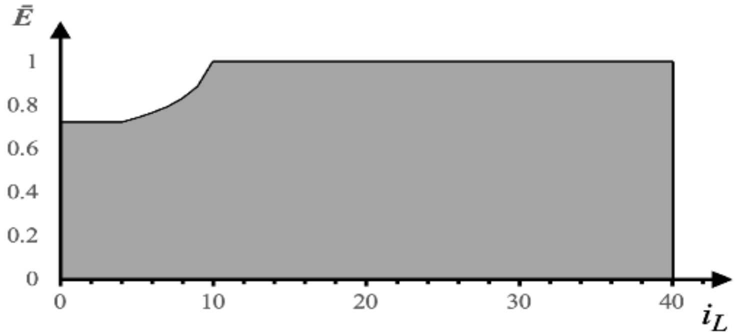


Fig. 14. *Graph of the change in the information criterion in the process of sequential optimization of control tolerances for features for the recognition classes of the lower tier stratum*

Analysis of Figure 14 shows that the information optimization criterion has already reached its maximum limit value of $\overline{E}^* = 1{,}00$ on the first iteration, which allows a decision about the end of machine learning for the lower-tier stratum recognition classes.

Figure 15 shows the graphs of the dependence of the information criterion (27) on the radii of the containers of the lower-tier stratum recognition classes, obtained based on the results of two-class information-extreme machine learning with parallel-sequential optimization of control tolerances.

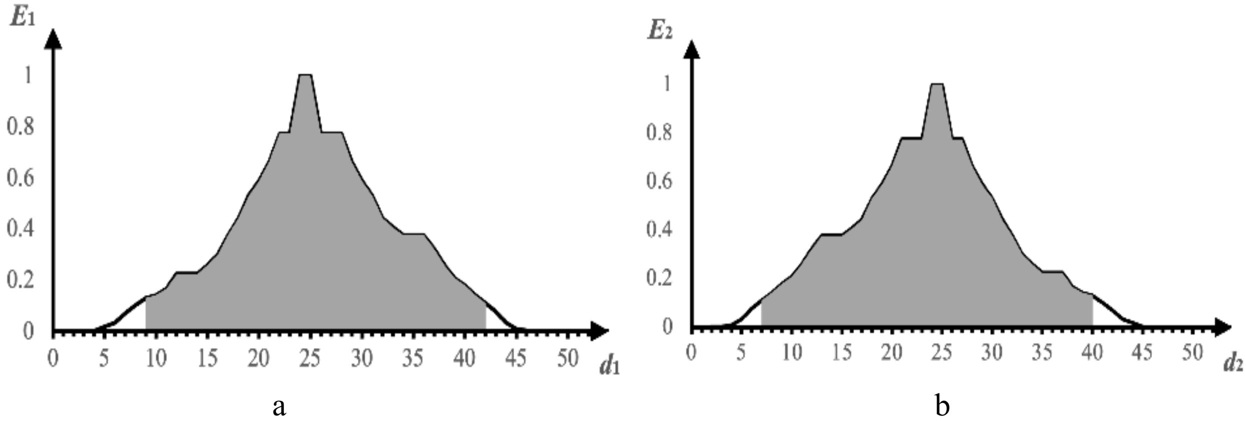


Fig. 15. *Graphs of the dependence of criterion (27) on the radius of containers of the recognition classes of the lower tier stratum: a – recognition class* X_1^o *; b – recognition class* X_2^o

Analysis of Figure 15 shows that at the minimum ratio (34) the optimal container radius for recognition class X_1^o is $d_1^* = 25$ and the optimal container radius for recognition class X_2^o is $d_2^* = 24$.

A comparison of the results obtained from machine learning of the upper and lower tier strata recognition classes shows that the optimal values of the container radii of the recognition class X_2^o are different. Therefore, according to the minimum distance principle, when constructing decision rules (31) for the recognition class X_2^o, it is necessary to take a smaller radius value, i.e. $d_2^* = 19$ code units of the Hamming distance.

Thus, it has been experimentally proven that with an alphabet power of more than two classes, in the general case, it is advisable to perform information-extreme machine learning of the ORS of an autonomous UAV using a hierarchical structure of input data in the form of a decursive binary tree.

Conclusions

1. An important scientific and practical task of developing an information intelligent machine learning technology for an on-board system of an autonomous UAV for video monitoring of the terrain under the condition of incomplete data certainty within the framework of a functional approach to modeling cognitive processes of natural intelligence has been solved.

2. The method of information extreme machine learning for an autonomous UAV for video monitoring of the terrain has been improved using a hierarchical data structure in the form of a decursive binary tree, which allows constructing decision rules that are error-free according to the training matrix in the process of machine learning with a given depth.

3. The results of computer modeling have proven that when the number of recognition classes is more than two, it is advisable to switch to information extreme machine learning using a hierarchical data structure in the form of a decursive binary tree, which allows reducing multi-class machine learning to two-class.

REFERENCES

1. J. Long, E. Shelhamer, and T. Darrell, Fully convolutional networks for semantic segmentation, *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR),* pp. 115–127, 2015.
2. Andrej Karpathy et al. "Large-scale video classification with convolutional neural networks", *IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* 2014.
3. A. Krizhevsky, I. Sutskever and G. E. Hinton, "Imagenet classification with deep convolutional neural networks", *Advances in neural information processing* systevolms, pp. 1097–1105, 2012.
4. Z. Zhen and G. Wang. "Learning discriminative hierarchical features for. object recognition", *IEEE Signal Process. Lett.*, vol. 21, is. 9, pp. 1159–1163, 2014.
5. Tankred Ott and Ulrich Lautenschlager, "GinJinn2: Object detection and segmentation for ecology and evolution", *Methods in Ecology and Evolution*, vol. 13, is. 3, pp. 603–610, 2022. DOI: https://doi.org/10.1111/2041-210X.13787.
6. L. Mou, P. Ghamisi, and X. Zhu, "Deep recurrent neural networks for hyperspectral image classification", *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 7, pp. 3639–3655, 2017.
7. Tankred Ott and Ulrich Lautenschlager, "GinJinn2: Object detection and segmentation for ecology and evolution", *Methods in Ecology and Evolution*, vol. 13, is. 3, pp. 603–610, 2022. DOI: https://doi.org/10.1111/2041-210X.13787.
8. Mladen Jurišić, Dorijan Radočaj, Ante Šiljeg, Oleg Antonić and Tihomir Živić, "Current status and perspective of remote sensing application in crop management", *Journal of Central European Agriculture*, vol. 22, is. 1, 2021. DOI: https://doi.org/10.5513/JCEA01/22.1.3042.
9. P. Gao, Y. Zhang, L. Zhang, R. Noguchi, and T. Ahamed, "Article development of a recognition system for spraying areas from unmanned aerial vehicles using a machine learning approach," *Sensors (Switzerland)*, vol. 19, is. 2, 2019. DOI: 10.3390/s19020313.
10. Z. L. Langford, J. Kumar, and F.M. Hoffman, "Convolutional Neural Network Approach for Mapping Arctic Vegetation using Multi-Sensor Remote Sensing Fusion". *IEEE International Conference on Data Mining Workshops*, 2017. DOI: 10.1109/ICDMW.2017.48.
11. Dovbysh A.S. Osnovy proektuvannia intelektualnykh system: Navchalnyi posibnyk / A.S. Dovbysh.– Sumy: Vydavnytstvo SumDU, 2009.– 171 s.
12. Dovbysh A.S. Intelektualni informatsiini tekhnolohii v elektronnomu navchanni / A.S. Dovbysh, A.V. Vasyliev, V.O. Liubchak. – Sumy: Vydavnytstvo SumDU. – 2013. – 172 s.
13. Kuzmin I. V. Osnovy teorii informatsii ta koduvannia: pidruchnyk / I. V. Kuzmin, I. V. Trotsyshyn, A. I. Kuzmin; za red. I. V. Kuzmina. – Vyd. 3-tie, pererobl. ta dopov. — Khmelnytskyi: KhNU, 2009. — 373 s.
14. Dovbysh A.S., Budnyk M.M., Piatachenko V. Yu., Myronenko M. I. Information-Extreme Machine Learning of On-Board Vehicle Recognition System. Cybernetics and Systems Analysis, 2020, 56(4),pp 534-543. DOI:10.1007/s10559-020-00269-y.