IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received August 29, 2020, accepted September 16, 2020, date of publication September 30, 2020, date of current version October 12, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3028030

Improving the Efficiency of Deep Learning Methods in Remote Sensing Data Analysis: Geosystem Approach

STANISLAV A. YAMASHKIN[1], ANATOLIY A. YAMASHKIN[2], VICTOR V. ZANOZIN[3], MILAN M. RADOVANOVIC[4,5], AND ALEXANDER N. BARMIN[3]

[1]Institute of Electronics and Lighting Engineering, National Research Mordovia State University, 430005 Saransk, Russia
[2]Faculty of Geography, National Research Mordovia State University, 430005 Saransk, Russia
[3]Faculty of Geology and Geography, Astrakhan State University, 414056 Astrakhan, Russia
[4]Geographical Institute "Jovan Cvijic," Serbian Academy of Sciences and Arts, 11000 Belgrade, Serbia
[5]Institute of Sports, Tourism and Service, South Ural State University, 454080 Chelyabinsk, Russia.

Corresponding author: Stanislav A. Yamashkin (yamashkinsa@mail.ru)

This work was supported by Russian Foundation for Basic Research (RFBR) under Project 20-37-70055.

ABSTRACT The article proposes a solution for the problem of high-resolution remote sensing data classification by applying deep learning methods and algorithms in conditions of labeled data scarcity. The problem can be solved within the geosystem approach, through the analysis of the genetic uniformity of spatially adjacent entities of different scale and hierarchical level. Advantages of the proposed GeoSystemNet model rest on a large number of freedom degrees, admitting flexible configuration of the model contingent upon the task at hand. Testing GeoSystemNet for classification of EuroSAT dataset, algorithmically augmented after the geosystem approach, demonstrated the possibility to improve the classification precision in conditions of labeled data accuracy by 9% and to obtain the classification precision with a larger volume of training data (by 2%) which is slightly inferior in comparison with other deep models. The article also shows that synthesis of the geosystem approach with deep learning capabilities allows us to optimize the diagnostics of exogeodynamic processes, owing to the calculation of landscape differentiation regularities. Application of the presented approach enabled us to improve the accuracy in detecting landslides at the testing site "Mordovia" by 5% in comparison with the classical approach of using deep models for remote sensing data analysis. The authors advocate that application of the geosystem approach to improve the efficiency of remote sensing data classification through methods, proposed in the article, requires an individual project-based approach to source data augmentation.

INDEX TERMS Convolutional neural networks, deep learning, geospatial analysis, geosystems, image classification, machine learning.

I. INTRODUCTION

Development and experimental substantiation of new geo-information methods and algorithms for automated analysis of spatial data (satellite images, digital models and maps, attributive spatial-temporal information) instrumental in the analysis of the state of lands and prediction of natural and man-made emergencies is a pressing challenge of our times. Development of machine learning technologies, including those based on deep neural network models [1], enables us to perform a highly precise automated monitoring of natural resources management systems and to analyze regularities of occurrence of natural processes and phenomena.

In view of this, algorithms of large arrays of spatial-temporal data and software packages that function on their basis are becoming an integral part of digital spatial data infrastructures (SDI).

Automated analysis of spatial data can be made by both traditional hard computing and soft computing, based on a combined use of fuzzy logic, artificial neural networks and evolutionary modeling [2]. The first decade of the 21st century has seen the rise of deep learning [3], methods and principles relying on the use a variety of levels of the non-linear data processing for extraction and transformation of features, analysis and pattern classification.

Research into the deep learning methods and algorithms have contributed to solution of a range of issues, connected with automated analysis of large arrays of spatial-temporal

The associate editor coordinating the review of this manuscript and approving it for publication was Gerardo Di Martino.

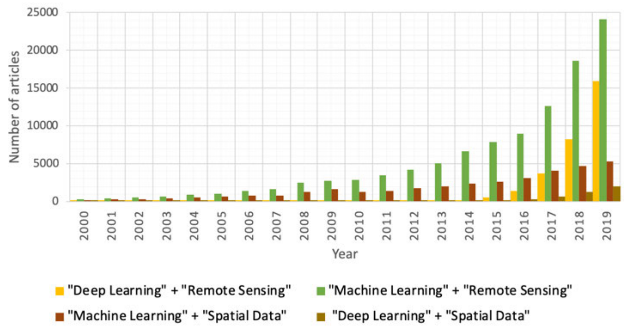

FIGURE 1. Number of articles with keywords from the problem area in the Google Scholar bibliographic database.(https://scholar.google.com).

data [4]. According to Google Scholar search engine [5], scientific interest in application of deep learning methods and algorithms to spatial data analysis has been growing for the last five years (Figure 1).

Application of deep neural network models should be problem-oriented: the processes of machine analyzer architecture design, selection of its hyperparameters, formulation of requirements for data output and consolidation of amounts of training, checking and testing samples are to be determined by specificity of the task to be tackled [6]. Quality of the documentation on application and flexible tuning of a select model does also matter.

The article proposes a solution for the problem of high-resolution remote sensing data classification relying on deep learning methods and algorithms in conditions of labeled data scarcity within the geosystem approach that implies analysis of the genetic uniformity of spatially adjacent entities of different scale and hierarchical level.

II. RELATED WORK

A. DEEP LEARNING IN GEOSPATIAL DATA ANALYSIS

Assessment and reasoning on the basis of spatial-temporal data deep analysis are applied in addressing multiple practical tasks – environmental monitoring, forecasting of development of different natural phenomena and processes. A great deal of publications on spatial data analysis looks into the issues of environmental incidents and natural disasters; a number of papers explore social and economic spatial processes.

Application of deep learning reduces the cost of conducted research owing to the possibility of fine interpolation and extrapolation of the measurements. Although most of the publications in the subject area are focused on the application of recurrent and convolutional neural networks, as well as autoencoders, other deep models are also used to address a variety of tasks. For example, deep belief networks can be applied to tackle the tasks of spatial anomaly detection [7] and classification [8] quite well; self-organizing maps are used for spatial data classification together with feed-forward networks [9].

The task of classification by means of deep learning mostly refers to remote sensing data analysis. Reference [10] describes multitier architecture, keyed to Land Cover classification by Landsat 8 and Sentinel multispectral satellite images. The proposed model rests on a spontaneously learning neural network, and the system is supplemented with an ensemble of supervised neural networks – multilayer perceptrons, decision trees and convolutional networks.

Deep learning methods and algorithms facilitate efforts to address the issue of feature extraction from multidimensional spatial data, which has been traditionally solved by means of labour-intensive expert work. Reference [11] proposes a deep model, based on the use of deep belief networks and restricted Boltzmann machines, as well as new functions of hierarchical feature extraction and image classification. Reference [12] describes a classification model, based on spectral-spatial features, which applies a combination of methods of dimension reduction and deep learning for collaborative automated extraction of spectral and spatial features from the sets of large-dimension hyperspectral data with the use of convolutional networks. Reference [13] proposes a feature extraction method with respect to unlabeled data, based on application of a sparse autoencoder for automatic extraction of spectral singularities and multi-scale spatial characteristics and support vector machines for further classification.

Autoencoders can be used for addressing the task of spatial data classification quite well. For example, [14] presents a deep learning method, based on an automated coder for remote sensing data classification, exemplified by a case study of the land cover in Africa. Methods of spectral-spatial classification of hyperspectral images with the use of autoencoders are described in [15]–[17]. Models that are based on a combined use of convolutional and recurrent models are successfully applied to remote sensing data classification. References [18]–[20] describe some methods of land cover classification on the basis of multitemporal spatial data with the use of deep recurrent neural networks.

Analysis of publications showed that application of deep learning methods and algorithms to the analysis of geospatial data comes across a number of open problems that need to be solved:

1) *deep model training in conditions of labeled data scarcity.* Despite the fact that deep models are capable of extracting information features from multidimensional data, the efficiency of this relies strongly on a large number of training samples. Nevertheless, creating banks of high-resolution labeled spatial data is costly and labour-intensive, that is why a task of preserving the deep model's ability to extract features on the basis of a smaller number of costly or hard-to-extract training samples remains open.

2) *internal complexity of the images, obtained from remote sensing data,* is determined by the fact that certain spatial objects are characterized by different sizes, spectral brightness, structural features and mutual position in relation to neighbor objects, and leads us to harder extraction of reliable and stable features from them. One should be careful with using the traditional methods of training data set augmentation that are based on applying primitive distortions in

FIGURE 2. Sample image patches of 10 classes covered in the EuroSAT dataset.

TABLE 1. Classification accuracy (%) on EuroSAT dataset of different training-test splits.

Method	Unit number	Dataset	Training time	Training-test split proportion								
				10/90	20/80	30/70	40/60	50/50	60/40	70/30	80/20	90/10
CNN (2 layers)	422 378	EuroSAT	164	75.88	79.84	81.29	83.04	84.48	85.77	87.24	87.96	88.66
GoogleNet	6 797 700	EuroSAT	902	77.37	90.97	90.57	91.62	94.96	95.54	95.70	96.02	96.17
DenseNet121	8 062 504	EuroSAT	1381	73.11	79.48	84.14	86.35	88.31	91.50	94.45	96.64	97.64
InceptionV3	23 851 784	EuroSAT	1044	75.03	89.13	92.20	93.01	**95.25**	**95.77**	96.02	**96.86**	**97.93**
ResNet50	25 636 712	EuroSAT	955	75.06	91.01	93.75	94.01	94.45	95.26	95.32	96.43	96.37
ResNet101	44 707 176	EuroSAT	1505	79.77	84.77	87.81	89.60	90.67	92.99	**96.17**	95.57	96.89
VGG16	138 357 544	EuroSAT	1134	76.53	85.59	88.23	89.88	91.86	93.02	94.66	96.65	96.08
GeoSystemNet	1 324 526	EuroSAT extended	895	**86.23**	**91.52**	**93.98**	**94.11**	94.29	94.35	94.41	94.65	95.30

analysis of remote sensing data: natural systems are often not indifferent to cardinal-direction orientation. For example, the slope exposure is an informative feature that should not be lost.

3) *defining hyperparameters of the model* in the analysis of complex spatial data is an open problem. Deep models can learn a larger number of features, but they are strongly liable to overfitting. Not the least is the factor of high resource intensity when training excessively intensive models, which makes one embrace for either high prices of experimental facilities or significant time expenditures.

4) *the problem of adaptation to a new set of data* is also relevant when finishing the deep model training for classification of lands of a new spatial area. The possibility to reuse the trained model for deciphering satellite images of another area can significantly increase the profitability and speed of conducted work, and the ways to obtain it require further systemic research.

The trained models can be tested when solving the task of testing data classification. A method of error matrices, built upon the classification results, can be chosen as a mathematical apparatus to calculate objective metrics of the models efficiency [21]. To test and fine-tune new algorithms for Earth remote sensing data analysis, open labeled datasets are created. The EuroSAT open dataset [22], created for training and testing machine training models to effectively deal with the problem of classification of land use and vegetation cover systems on the basis of Sentinel-2 satellite images, was applied for primary testing of the proposed method. The dataset is uniformly labeled by 10 classes and consists of 27,000 images, containing information on land plots, distributed across the European Union, in 13 spectral ranges. The size of each element of the dataset is 64 × 64 pixels with the spatial resolution of 10 meters per pixel, and is characterized by geographical reference (Figure 2).

The authors of EuroSAT dataset [23] describe the classification precision characteristics based on different ratios between training and test samples (Table 1). ResNet-50 neural network model demonstrates the precision of 96.43% (in spliting training and test samples at the ratio of 80:20) and 75.06% (at the ratio of 10:90). A shallow two-layer CNN reaches the precision of 87.96% (at the ratio of 80:20) and 75.88% (at the ratio of 10:90). Note also that convolution-layer deep models mostly demonstrated higher precision than support vector machines did.

Therefore, modern deep convolutional networks demonstrate excellent precision of satellite image classification with a relatively large size of the training sample of EuroSAT dataset, but the presented approaches have a significant precision loss in case of training data scarcity. At the same time, improving the precision of methods and algorithms of spatial data analysis with account of their scarcity is a topical problem [24].

One has to search for a decision to the outlined problems not only through the improvement of deep model architectures but also through the development of methods and algorithms of optimal enrichment of training data sets. It is proposed to complete this task on the basis of the geosystem approach.

B. GEOSYSTEM APPROACH

The geosystem approach [25] relies on a hypothesis that the geographical envelope, landscape sphere, population, environment, economic sectors and territorial production complexes are intertwined with internal deep interconnections that form a basis for drawing conclusions on object's origin and state.

Geosystems include holistic entities, sets of interconnected components, the properties and state of which depend on their spatial positioning and qualities of the environment. The notion of geosystem is used to designate quite a wide range of spatial objects, geographical and territorial production complexes. Any geosystem and its structural components are based on a functioning element that exercises specific functions and is indivisible in dealing with the task to be solved.

The introduction of the geosystem approach rests on examination of the complex objects' structure, formed from simpler ones that are organically interconnected, and makes it possible to deal with the issues of geoinformation support of regional schemes and projects, which leads to development of the systemic mapping.

Hierarchic structuring of geosystems for land analysis and classification should be based on marking out typological units (taxa). For example, to study the state and regularities of development of certain areas, it is worthwhile to use taxa of geosystems, proposed in [26] – system, class, types, genera and kind of geosystems:

1) *System (rank) of geosystems* (GS_1) is an essential classification category that is marked out by peculiarities of a power base – the water and heat balance. Systems that are marked out by peculiarities of the microclimate determine the specificity of development of geoecological processes like weathering, exogeodynamics of geological and geomorphological processes, hydrogeodynamics and hydrogeochemistry of groundwater, hydrological and soil-forming processes and biological cycle; they determine the possibility of climatogenic emergency situations.

2) *Classes of geosystems* (GS_2) are determined by heat and moisture redistribution under the action of lithogenous bases – following this, plain and mountain classes are marked out. Differences in lithogenous bases determine peculiarities of landscape zonation – horizontal zonation on the plains and the vertical in the mountains. The same features, though this time more specific ones, are used to mark out subclasses of geosystems. Subclasses of high, lowland and low-lying plains are marked out in the plain geosystem class with regard to the genesis and history of development. The landscape metric diagnoses peculiarities of manifestations of the newest and modern tectonics, exogeodynamic processes, localization of areas with progressing and prevailing processes of denudation and accumulation.

3) *Types of geosystems* (GS_3) are marked out by soil and bioclimatic features and are essential classification categories in terms of analysis of landscape development processes. The geosystem type is a category that is singled out with account to specificity of the development of soil and biological processes: 1) of chemical elements and compounds getting onto the parent rock with atmospheric precipitation, soil animals and plants; 2) transformation, movement and accumulation of chemical elements along the soil profile and formation of genetic horizons; 3) chemical elements' carry-over out of the soil profile with atmospheric precipitation. Analysis of the area at the level of geosystem types diagnoses such geoecological processes as transit or accumulation of technogenesis products in soils, development of slope processes and erosion.

4) *Genera of geosystems* (GS_4) are marked out on the basis of morphosculptural forms of the relief and their composing sediments and are a result of the activity of exogeodynamic processes (types of the erosion relief, forms of karst, suffusion, etc.).

5) *Kinds of geosystems* (GS_5) are determined by a factorial and dynamical structure of the sites and vegetation on the topological level and are an elementary territorial unit that can be associated with the land cover class.

Following the geosystem approach, the state and properties of each territorial unit are determined by 1) peculiarities of its interaction with neighbor objects of the same hierarchical level; 2) characteristics of the host geospatial system of a higher hierarchical level, and 3) interaction between the objects of a lower hierarchical level that the area under analysis consists of.

Specific features of every area are largely determined by its location in the structure of the environment. Hierarchical nature and self-dependence are typical features of geosystems, that is why large volume of information on spatial object's belonging to a certain class can be potentially contained in the data on the geosystem of a higher level of the hierarchy that hosts the classified area and influences it.

III. METHODS OR METHODOLOGY

A. DATA PREPARATION

In view of this, it is possible to formulate a hypothesis that it is possible to improve the precision of land classification, based on a remote sensing database, if the classifying model takes into account and analyzes not only the properties of a certain area but also specific features of the geosystems that it interacts with and, in particular, that it is a part of.

In order to test the hypothesis, it is necessary to prepare several sets of data to train the models – basic (consisting of expertly labeled samples of the areas, captured by the satellite) and extended (augmented with the compared data, containing information on neighbour and host geosystems). After this, it is necessary to propose a deep model that receives input data on the area and the associated geosystems, analyzes this information and makes a decision on classifying this area. If the classification precision is improved and the preparation process becomes slightly more expensive and complicated, one may speak about reasonability of application of the geosystem approach in the training of deep neural network models.

The solution to the problem of classification of remote sensing data through deep learning using the geosystem approach should be based on the preliminary augmentation of the data and the creation of a deep model that is capable of analyzing these data efficiently. Under ''classification''

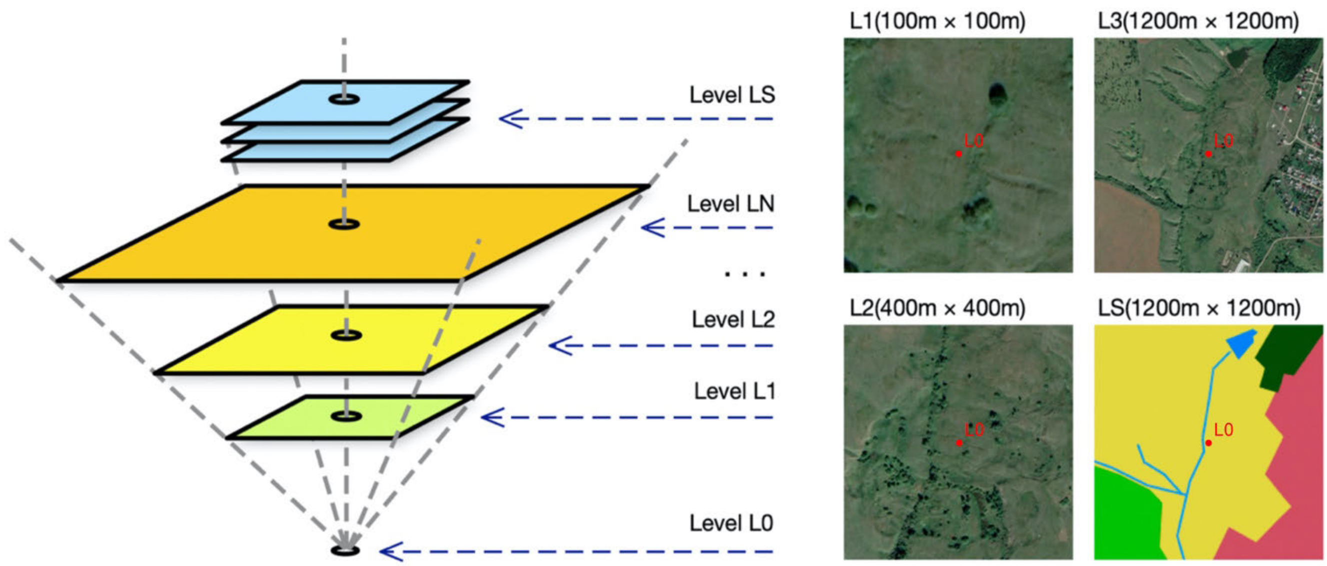

FIGURE 3. **Geosystem model of the area. The following illustrations are given as an example: the layers L_1, L_2 and L_3 are host geosystems of different hierarchical levels in the visible spectrum, while the layer L_S is a fragment of the landscape map of the area.**

we shall basically mean the function f, carried out by the model M that has the experience E and allows us to associate a local object, characterized by the set of parameters $\mathbf{x}_{Local}$ and directly interconnected with the geosystems, determined by the property vector $\mathbf{x}_{Geosystem}$, to a certain class mark y:

$$y = f\left(\left\langle \mathbf{x}_{Local}, \mathbf{x}_{Geosystem} \right\rangle, M, E\right) \tag{1}$$

If $\mathbf{x}_{Geosystem}$ is an empty set, we may speak about classification without involvement of geosystem data. The set of features of the local object $\mathbf{x}_{Local}$ will be formed on the basis of remote sensing data and can be of different formats. For example, the area can be classified on the basis of pixel analysis (pixel-based classification) and through extraction of features from different-size fragments of the area (patch-based classification). Besides, data on the area are described by different spatial, spectral and radiometric resolutions.

The set of features of a local object which can be packed in different-dimension tensors by itself, determines the level L_0 of the created geospatial model of the area (Fig. 3). According to the geosystem approach, the host geosystem has a significant effect on the area's properties. Remote sensing data are a helpful source of information on it. However, while rigid requirements are set for the L_0 data on the object $\mathbf{x}_{Geosystem}$ (they must be obtained strictly at a certain time and have high resolution) and, as a consequence, they are rather expensive, requirements to data of the levels starting from L_1 can be relaxed, ensuring simplification and reduction in cost of the process of their receiving. Currently, remote sensing data of medium and high spatial resolution are openly supplied by various providers through the Internet, and some of them provide handy application programming interfaces (API) [27] for their quick receiving. The fact that these data have low temporal resolution (we often unable to choose a certain data of satellite imagery) explains their low cost.

At the same time, they still are a helpful source of information on $\mathbf{x}_{Geosystem}$ concerning the host geosystems of different hierarchical levels.

The process of receiving data of the levels $L_1, L_2, \ldots, L_N$ is subject to potentially full automation:

with data on the geographical coordinates (latitude and longitude) of the classified area, one is able to make a request to the API of the spatial data provider for a fragment of the territory's satellite image with above coordinates, required scale and resolution.

Therefore, we have the possibility of algorithmic augmentation of the training data set by importing satellite imagery fragments, describing the geosystems of a higher hierarchical level and hosting the classified area.

Not only remote sensing data of a certain scale are able to characterize geosystems of different hierarchical levels – this function can be well assigned to synthetic digital maps too. Land Cover, Land Use, electronic landscape and other thematic maps traditionally being a final artifact of remote sensing data analysis and classification, contain a significant volume of information on the properties of the areas that are part of it, and, as a consequence, they can be used for creating input tensors of additional information in the set $\mathbf{x}_{Geosystem}$. Despite the fact that these maps often have relatively low resolution, their high abstraction level allows us to speak of their high potential for enrichment of information on classification of a small-size region, located in the area of geosystems that are discernible only on a smaller scale. Synthetic digital maps form the level L_S of a geospatial model of the area become another source for augmentation of the auxiliary data set $\mathbf{x}_{Geosystem}$.

For classification on the basis of the data tuple $\langle \mathbf{x}_{Local}, \mathbf{x}_{Geosystem} \rangle$, let us move to a design of a deep neural system which receives input tensors of data of different hierarchical levels about the classified area (L_0) and its host geosystems (L_1, L_2, …, L_N, L_S) and gives a hypothesis on belonging of this area to a certain class.

B. METHODS FOR MODEL SEARCH

It is proposed to design a deep model by the following optimizing algorithm:

1) to form a system of requirements to the model: to determine inputs and outputs, performance and precision rating;

2) to determine basic architecture of the model on the basis of the unit approach that describes general organization of the classifier;

3) to decompose top-level units into linear or branching structures;

4) to manage the problem of the classification precision dilution and overfitting through heuristic configuration of deep model hyperparameters and addition of normalization, subsampling and regularizing layers;

5) to optimize the model by the principle ''small is better than big'' – the process of training large models is computing resource intensive and, which is even more important, deep neural networks are inclined to overfitting;

6) to train the model with testing different precision measures, optimization algorithms, loss functions and the number of training epochs;

7) to analyze the model training process through calculation of the dependence of the expected value and standard deviation of classification precision on the training epoch on the basis of a series of experiments;

8) to assess the obtained result quality through building error matrices and estimation of accuracy and fallibility metrics upon results of the model functioning;

9) to draw a conclusion on correspondence of the obtained model to objective and subjective requirements.

The proposed chain of actions results in obtaining a model sample with specific properties, and if they are met, the search can be completed. If the parameters of the designed and trained model do not meet the pre-set requirements, it is necessary to roll a few steps back along the deep model creation path (down to the first stage if the formulated requirements turned out to be unachievable) and re-search in a heuristically adjusted direction.

As a result, the process of seeking an efficient classification model may be formalized as a tree, the root node of which precedes the first stage of the search algorithm and corresponds to the task of study problem statement. The tree nodes determine a variant of model statefulness at the i-th stage of the efficient model search algorithm. Terminal nodes (leaves) of the tree correspond to a particular solution of the task of searching for an optimal model, ready for using a deep classifier.

Obtained particular solutions can be compared with split-testing, based on comparison of objective numerical metrics of model efficiency with subjective expert assessment of classification quality.

C. DEEP MODEL FOR GEOSYSTEM ANALYSIS OF SPATIAL DATA

From the black box perspective, a deep classification model, based on application of the geosystem approach (GeoSystemNet), is a functional element that receives input satellite images of the area (L_0) and its host geosystems (L_i), as well as synthetic maps (L_S). The number of inputs may vary depending on the number of levels of the area's geosystem model, but one should be particularly careful with their growth because it will inevitably lead to the necessity for increasing the model capacity. The model has one output in the form of a vector, each i-th element of which determines predicted probability of area's belonging to the i-th class. A final hypothesis on area's belonging to a certain class is put forward on a the-winner-takes-all principle, when the object belongs to a class for which the model predicts the maximum probability. Figure 4 illustrates the decomposition of the described model.

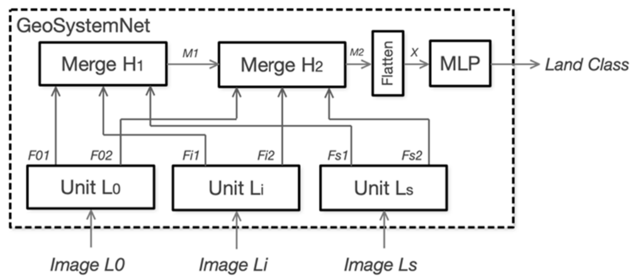

FIGURE 4. General architecture of GeoSystemNet model with 3 inputs.

Unit L_x, that extracts hierarchical features F_{xi} of different levels $i = \overline{1, N}$ from the source image L_x is introduced to initial extraction of features on the basis of the data of each input $L_x \in \langle L0, \ldots, Li, \ldots, Ls \rangle$. Unit L_x is decomposed into N data extraction units, each one having an external output. The structure of each unit is a chain of layers. The first layer, performing the operation of depthwise separable 2D convolution [28], allows the extraction of features from the source image and, in contrast to the application of a traditional convolution layer, facilitates us to make a deep model more compact and, consequently, overfitting-tolerant.

An operation of two-dimensional convolution with the kernel W of the size K, as the basis of the layer functioning, is linear transformation where each value $y_{i,j}$ of the output matrix Y is calculated on the basis of the values x of the initial matrix X according to the following equation:

$$y_{i,j} = W * X = \sum_{a=0}^{K-1} \sum_{b=0}^{K-1} W_{a,b} x_{i+a,j+b} \quad (2)$$

The convolution operation has a number of important properties: it preserves the input structure and geometry and is characterized by sparsity and a multiple use of the same weight. The operation of depthwise separable convolutions works not only with spatial measuring but also with depth measuring, for example, with image channels, and, in contrast to the classical convolution, implies the use of individual convolution kernels, on the basis of which two convolutions are sequentially applied to the initial tensor – depthwise and pointwise.

Noteworthy is that when solving the test tasks on classification described below, we carried out a split-testing of models with classical convolution layers and with depthwise separable 2D convolution layers, which confirmed the efficiency of the second approach. The batch normalization layer [29], that allows us to achieve the model regularization and stability, was the next layer of the feature extraction unit, the efficiency of which had been experimentally tested.

A rectified linear unit, which carries out transformation of the type $x = \max(0, x)$, was chosen to perform the activation operation. A subsampling layer with external outputs, performing the application of the operation of max pooling for 2D spatial data [30] to reduce the size of the obtained representations, completes the feature extraction unit.

The subsampling operation, applied to the elements $x_{i,j}$ of the initial matrix X, leads to obtaining the matrix Y, for which the value of each element $y_{i,j}$ under the subsampling window size d is calculated according to the equation (3). Experiments showed that application of the operation on reaching the maximum yielded the best result.

$$y_{i,j} = \max_{\substack{0 \le a < d \\ 0 \le b < d}} (x_{i+a,j+b}) \tag{3}$$

Also of relevance is that we propose to choose the number of output filters in the convolution and the convolution kernel size according to the principle of minimization of these values to keep the classification adequately precise. It is recommended to increase the number of output filters for the depthwise separable 2D convolution with each new step of extracting features of the next level. Figure 5 presents a block diagram of Unit L_x, allowing the extraction of hierarchical features of two levels from the image L_x.

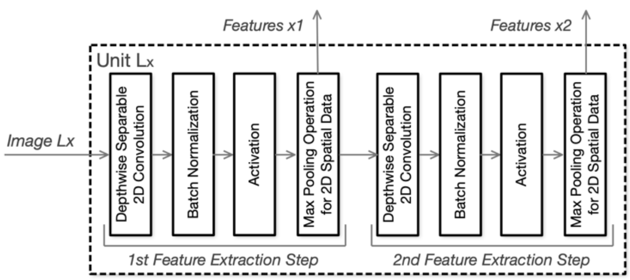

FIGURE 5. Module Unit Lx of extracting hierarchical features of two levels.

The next component unit of GeoSystemNet model is a feature merging module, presented in Figure 6. It receives input features of the level N, extracted from the classified area's image and geosystems' images associated with it. The merging modules of the second and subsequent levels also

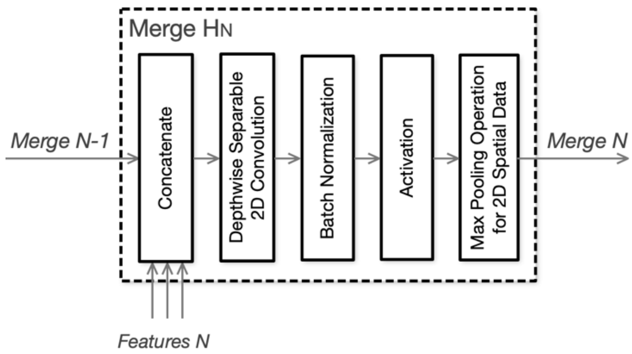

FIGURE 6. Feature merging module of level Merge H_N.

receive the output data of the previous merging module. The overall number of feature merging modules equals to the number of levels of hierarchical feature extraction in sections of Unit L_x.

All input data are concatenated in a single tensor and processed by a feature extraction pipeline that has a structure similar to the one used in the module Unit Lx. It consists of such layers as depthwise separable 2D convolution, batch normalization, activation and max pooling for 2D spatial data, and it is proposed to opt for a larger number of output filters in the convolution for the unit Merge H_N than the filter dimension in the process of feature extraction at the respective level N in the module Unit L_x.

The feature merging module output is transformed into a vector through the flatten operation and is input to a multilayer perceptron (Figure 7).

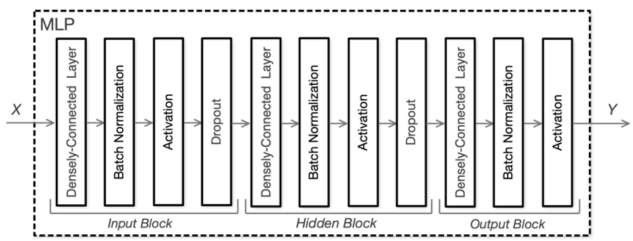

FIGURE 7. Module of geosystem classification based on hierarchical features.

The number of tightly concatenated layers of the multilayer perceptron and their power are chosen following the principle of minimization of these parameters to maintain sufficient precision of the classification. Besides, in order to solve the problem of overfitting, we recommend to apply operations of batch normalization and dropout to the outputs of a tightly concatenated layer to activate the output of input and hidden layers; for the output layer, it was a sigmoid unit for the binary classification and a softmax unit for the multi-class classification.

When training GeoSystemNet classifier, the root mean square propagation algorithm (RMSProp), based on the method of stochastic gradient descent, was used as an optimizer, while the cross entropy served as a loss function. Peculiarities of a specific classification task have an impact on the fine tuning of GeoSystemNet model.

IV. EXPERIMENTS AND RESULTS

A. QUALITY ASSESSMENT ON EuroSAT DATASET

Availability of spatial data on the host geosystem has paramount importance. We have developed an algorithm for training dataset augmentation that allows us to download different-scale images of the host area from MapBox API, using coordinates of the element from EuroSAT dataset. Therefore, the basic dataset (level L_0) was extended with additional levels of information without substantial costs. The final dataset (EuroSAT extended) got the following structure:

- *Level* L_0. Labeled data of EuroSAT dataset (64 × 64 images in the visible spectral range, natural colours). Training and test samples are splited at the ratio of 10:90 to simulate a data scarcity situation.

- *Levels* L_1, L_2, L_3. Fragments of open satellite images in the visible spectral range, obtained automatically from MapBox online map provider through the application of programming interface (API) in the zoom level of 8, 12 and 14 respectively.

The initial dataset augmentation led to each classified section's being represented by four different-scale images of the area.

The experiment was conducted on the equipment with the following key characteristics: CPU – Skylake X (14-Core 3.30 GHz Intel Core i9-9940X), RAM – (3000 MHz DDR4 32 GB), GPU – Nvidia Titan RTX based on Turing architecture (576 tensor cores, 24 GB GDDR6).

A software code, implementing GeoSystemNet model and a possibility to conduct experiments and comparative analysis, was written in Python with involvement of TensorFlow framework and Keras library [31].

The module of extraction of hierarchical features Unit L_x is implemented at two levels; experiments prove that the values of the number of output filters in convolutions for these levels are equal to 64 and 128 (decrement of these values resulted in lower precision, while their increment led to instability of the training process at the early epochs and to the increase of its resource intensity). The convolution filters are square with the side size of 3, pooling window – 2.

The first-level feature merging module Merge H_N has a layer performing the operation of depthwise separable 2D convolution with the number of output filters equal to 128, of the second level – 256. The convolution filters are square with the side size of 3.

The multi-layer perceptron, meant for geosystem classification on the basis of hierarchical features, have layers with the capacity of 128, 64 and 10 elements. The dropout layer coefficient was adjusted to be equal to 0.4. The deep model was trained during 50 epochs.

Comparison values of the precision of the proposed model GeoSystemNet and modern deep learning models (GoogleNet, DenseNet121, InceptionV3, ResNet50, ResNet101, VGG16) are given in Table 1.

The model, proposed in the article, demonstrates the best result with extraction of the testing data from EuroSAT dataset at the ratio of 40% and lower, and the relative efficiency increases with a decrease in the training sample up to 10% (86.23% against 79.77% in the second result (ResNet101)). As long as the training data grow in their size, GeoSystemNet model becomes inferior to other deep models ResNet-50, but this gap falls within the range of 3%.

It should be noted that GeoSystemNet model obtained such results owing to the analysis of EuroSAT dataset, extended after the geosystem approach (while other models were trained with and analyzed EuroSAT initial dataset). The difference in the experiment conditions was leveled with the low cost and quickness of a fully automated process of the training set augmentation, as well as with lower capacity of GeoSystemNet model – 1.3 million units against 138 million in case of VGG16. Therefore, the advantages in deep model training under data scarcity conditions were gained owing to 1) low-cost automated augmentation of the dataset on the basis of the geosystem approach and 2) creation of an efficient deep model for its analysis.

Analysis of the GeoSystemNet training process is of interest. Since neural network training is a random process, a series of 10 experiments was conducted, resulted in constructing dependence of the expectation E of the testing data classification precision on training epochs. Results of the experiment under training data scarcity (training - test samples' split at the ratio of 10:90) are presented in Figure 8.

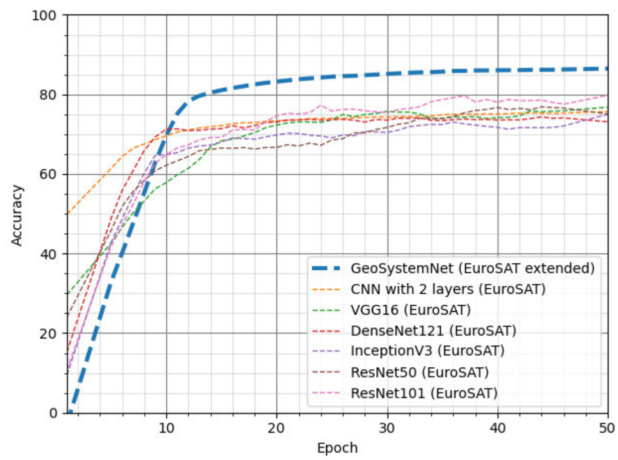

FIGURE 8. Visualization of the expectation E of the classification accuracy of the testing dataset dependence (90% of the total set) on the number of training epochs.

GeoSystemNet model demonstrates low precision of the extended set classification at early stages of the training process, which starts growing almost from the zero value. A light two-layer CNN and deep models reach the precision of more than 40% from the first epoch. Nevertheless, GeoSystemNet outstrips other models after the 10th training epoch, reaching the desired precision of 86%. Of significance is a

TABLE 2. Classification metrics (%) of different classification models trained in condition of data scarcity.

Metric	Model	Dataset	Class: Annual crop	Forest	Herbaceous	Highway	Industrial	Pasture	Permanent crop	Residential	River	Water
Precision	Shallow CNN	EuroSAT	85.40	90.28	54.99	68.02	82.63	73.04	67.45	90.46	67.15	93.36
	GoogleNet	EuroSAT	85.90	81.17	57.22	56.15	91.03	81.75	69.50	83.93	57.12	96.91
	DenseNet121	EuroSAT	71.12	75.39	52.39	51.54	**94.52**	63.11	57.23	79.74	78.34	91.82
	InceptionV3	EuroSAT	86.80	80.27	56.32	57.15	91.03	82.71	68.51	84.96	56.18	97.90
	ResNet50	EuroSAT	82.87	**92.79**	67.11	53.29	71.54	81.98	65.25	85.88	**67.51**	93.41
	ResNet101	EuroSAT	83.61	91.62	72.54	57.70	73.07	61.95	59.18	87.08	53.55	92.04
	VGG16	EuroSAT	79.03	80.69	50.27	66.38	86.64	58.16	58.03	**93.77**	66.82	89.74
	GeoSystemNet	EuroSAT ext.	**93.30**	90.40	**89.29**	**70.38**	91.74	**89.85**	**84.47**	92.52	67.30	**98.32**
Recall	Shallow CNN	EuroSAT	83.41	93.44	65.90	50.50	91.24	68.78	63.26	96.03	65.73	90.06
	GoogleNet	EuroSAT	76.45	98.03	71.80	46.21	70.04	55.20	55.68	95.06	79.91	82.41
	DenseNet121	EuroSAT	85.67	92.40	64.38	53.81	40.39	74.97	53.33	95.66	53.88	7218
	InceptionV3	EuroSAT	77.58	97.54	72.94	44.00	71.30	56.84	56.76	**96.47**	78.79	83.40
	ResNet50	EuroSAT	83.01	90.01	48.88	**62.61**	**93.34**	60.27	73.63	84.03	66.21	**94.07**
	ResNet101	EuroSAT	81.32	93.88	45.90	53.24	86.07	68.28	44.73	94.62	59.21	15.18
	VGG16	EuroSAT	80.45	92.33	56.33	49.97	83.01	63.84	54.09	92.16	66.91	86.66
	GeoSystemNet	EuroSAT ext.	**89.21**	**98.18**	**85.44**	58.05	90.59	**84.29**	**87.84**	94.82	**83.43**	92.62
F1 Score	Shallow CNN	EuroSAT	84.39	91.84	59.95	57.97	86.72	70.85	65.29	93.16	66.43	91.68
	GoogleNet	EuroSAT	80.39	89.25	62.52	48.69	82.71	66.35	64.18	92.01	64.34	89.41
	DenseNet121	EuroSAT	77.72	83.03	57.77	52.65	57.29	68.53	55.21	86.98	63.85	80.82
	InceptionV3	EuroSAT	81.93	88.07	63.56	49.72	79.97	67.38	62.08	90.35	65.59	90.07
	ResNet50	EuroSAT	82.94	91.38	56.56	57.58	81.00	69.47	69.19	84.94	66.85	93.74
	ResNet101	EuroSAT	79.42	63.19	58.18	44.15	79.04	74.27	50.95	88.95	56.24	76.06
	VGG16	EuroSAT	79.73	86.12	53.12	57.02	84.78	60.87	55.99	92.96	66.87	88.18
	GeoSystemNet	EuroSAT ext.	**91.21**	**94.13**	**87.32**	**63.62**	**91.16**	**86.98**	**86.12**	**93.65**	**74.50**	**95.38**
Fbeta Score (β=0.5)	Shallow CNN	EuroSAT	84.99	90.90	56.87	63.61	84.22	72.15	66.57	91.52	66.86	92.68
	GoogleNet	EuroSAT	84.18	82.38	61.97	53.01	87.31	74.01	67.37	86.93	61.54	95.50
	DenseNet121	EuroSAT	73.62	78.27	54.42	51.98	76.50	65.18	56.40	82.48	**71.82**	87.08
	InceptionV3	EuroSAT	84.78	83.22	59.01	53.93	86.26	75.81	65.79	87.03	59.60	94.61
	ResNet50	EuroSAT	82.90	**92.22**	62.45	54.93	75.04	76.47	66.77	85.50	67.24	93.54
	ResNet101	EuroSAT	86.32	90.82	63.34	40.04	75.34	78.63	55.59	80.04	54.59	45.74
	VGG16	EuroSAT	79.31	82.78	51.37	62.29	85.89	59.21	57.20	92.44	66.84	89.11
	GeoSystemNet	EuroSAT ext.	**92.45**	91.86	**88.49**	**67.51**	**91.51**	**88.68**	**85.12**	**92.97**	70.01	**97.12**

low standard deviation from the dependence mathematical expectation, typical for GeoSystemNet when being trained with a small dataset. This points to a higher stability of the deep model training process and high capability of the correct generalization of information on the analyzed features.

Informative relative metrics were calculated on the basis of the absolute measures of the classification process: TP_{class} – number of hits (correctly detected class objects), FP_{class} – Type I error or false positive (it shows how many times the objects were incorrectly classified as belonging to the class), FN_{class} – Type II error or miss (it shows how many times the class objects were classified incorrectly).

They served the basis for determining such relative metrics as precision (4), recall (5), F1 Score (F_{1class}) (6) and Fbeta Score ($F_{\beta class}$) (7) for each class of the land surface according to the following equations.

$$precision_{class} = \frac{TP_{class}}{TP_{class} + FP_{class}} \tag{4}$$

$$recall_{class} = \frac{TP_{class}}{TP_{class} + FN_{class}} \tag{5}$$

$$F_{1class} = \frac{precision_{class} \cdot recall_{class}}{precision_{class} + recall_{class}} \tag{6}$$

$$F_{\beta class} = (1 + \beta^2)\frac{precision_{class} \cdot recall_{class}}{\beta^2 \cdot precision_{class} + recall_{class}} \tag{7}$$

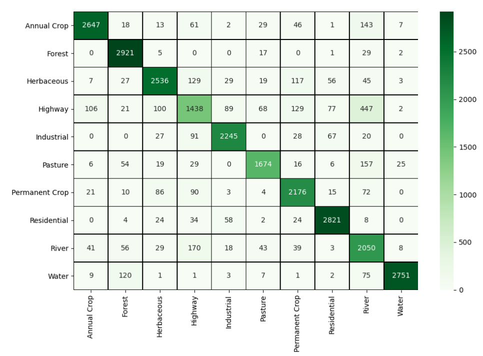

FIGURE 9. Error matrix for GeoSystemNet, trained under data scarcity conditions (10:90 training-test splitting range) on EuroSAT dataset, extended according to the geosystem approach.

Result values of the metrics are given in Table 2.

It is seen that GeoSystemNet, trained in conditions of source data scarcity, demonstrated a far better result in many cases, up to the rate increase by more than 30%. Note the model weaknesses too – lower efficiency indices, fixed in extracting forest and human-induced geosystems (it ranks below the leader by 5% at most).

Figure 9 presents error matrix for GeoSystemNet models within the conducted experiment in conditions of training dataset scarcity. Values of the main diagonal elements show that the value of metric "true positive" for GeoSystemNet model is relatively low only for Industrial and Water classes.

It is connected with the fact that consideration of images of the host geosystems during the classification process inevitably leads to the increase in the analyzed data, and higher capacity of the model may be required to analyze these data.

Besides, in some cases additional images can misinform the deep model: in terms of an algorithm, it is certainly easier to classify one uniform image of the water surface than an image, supplemented by several fragments of a smaller scale that include coastal areas.

Therefore, the augmentation of EuroSAT dataset after the geosystem approach and the development of GeoSystemNet model allowed to raise the classification precision in conditions of training data scarcity (training-test samples splitting at the ratio from 10:90 to 40:60) and to demonstrate results that exceed precision values of deep models in the EuroSAT dataset classification.

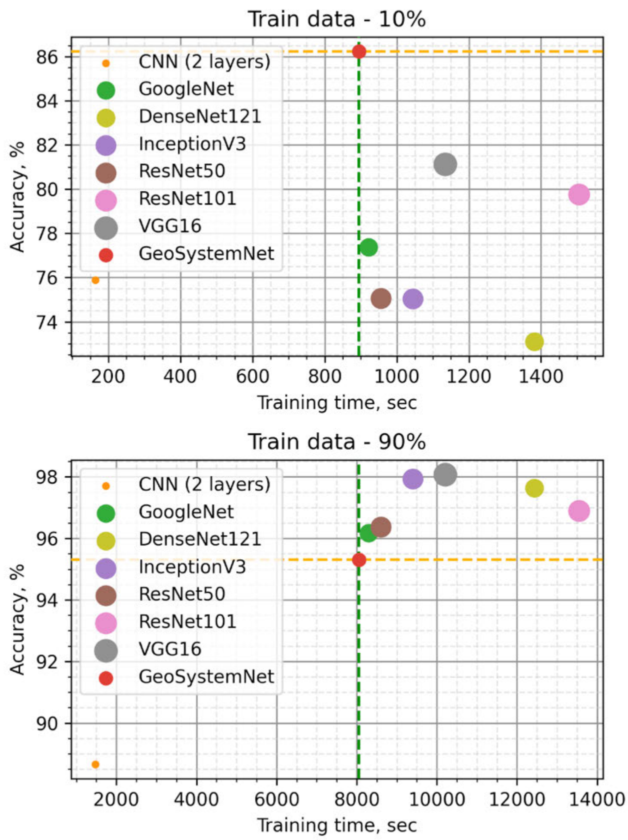

FIGURE 10. A scatter plot visualizing the accuracy and training time characteristics of compared models under conditions of different training data sizes.

Scatter diagrams (Figure 10) allow a comparative evaluation of the efficiency of the considered models. The larger the radius of each point, the greater the number of units of the corresponding model. The first diagram describes the results of the experiment in the conditions of training models on 10% of the labeled data, the second - on 90%. It can be seen that under the conditions of a scarcity of training data, the GeoSystemNet model showed the maximum accuracy, while the time spent on its training was less than that of other models (with the exception of Shallow CNN, which showed a relatively low accuracy). In conditions of a sufficient amount of labeled data, the GeoSystemNet model is inferior to the known deep models within 3% accuracy, however, the model learns faster and is characterized by fewer units. The increase in the speed of training and a lesser size with a slight loss of data classification accuracy is an advantage that matters because of the high cost of GPUs.

Experiments also showed that application of the geosystem approach according to the method, described in the article, is not a panacea for the problem of improving remote sensing data classification efficiency in conditions of data scarcity. Each specific task of geospatial data analysis requires an individual approach to the choice of a data model and organization of the process for its classifier. One of the advantages of GeoSystemNet model is a large number of freedom degrees, ensuring flexible configuration when dealing with tasks at hand.

Variable parameters of the model include 1) number of model inputs, 2) number of levels of feature extraction by Unit L_x, 3) capacity and number of feature merging modules Merge H_N, 4) hyperparameters of the multi-layer perceptron that makes a result decision. Finally, the process of selecting sources of information on the host geosystems has an infinite number of variants: these can be both remote sensing data of different scale and resolution and fragments of electronic maps of different types. It is necessary to be fully aware of the responsibility in selecting data on host geosystems – a wrong decision will lead to a situation when an extended training set will rather misinform than raise the GeoSystemNet training model efficiency and even lower metrics of model efficiency and classification quality in general.

B. PRACTICAL CASE: LANDSLIDE DETECTION

The next stage in testing the proposed approach was implemented when dealing with detection of landslides on the remote sensing database.

Test area "Mordovia", located within the coordinates of 42.16"E, 46.78"E from westward and eastward and 53.62"N, 55.21"N from southward and northward, was taken as a basis. The area is located in subboreal semi-humid (forest-steppe) geosystems of the bedded-layer Volga Upland. The forest-steppe geosystems contrastively extend into the forest province of aquaglacial plains of the layered Oka-Don Lowlands.

The relevance of assessment of geological environment stability and exogeodynamic process forecasting in the chosen area is determined by an increasing number of complex natural and man-made emergencies. The exogeodynamic process forecasting (EPF) rests on the assumption that the future natural and man-made emergencies highly likely take place under the same conditions they did in the past. That is why assessment of the spatial correlation between factors of

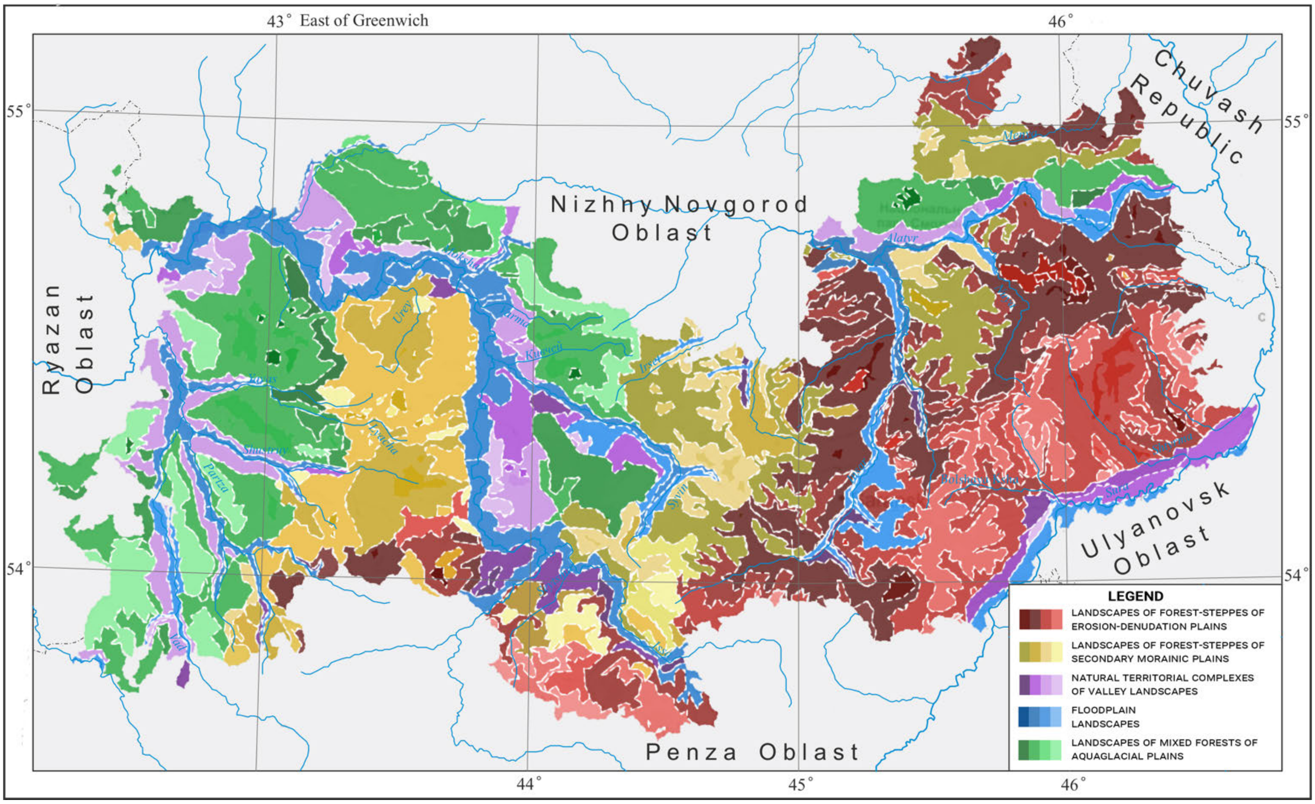

FIGURE 11. Landscape map, plotted on the basis of Mordovia GIS.

different nature and the previous episodes of occurrence plays an important role.

The factors that reflect EPF development include morphometrical (profile slope, altitude or curvature) and morphological features of the relief, tectonics, geology and hydrogeology, climate and hydrologic behaviour of surface water, type of soil and vegetation cover, land use, geotechnical systems and density of these objects.

Multi-zone satellite images, digital models of the relief and synthetic landscape maps constituted the main sources of information for analysis.

The geosystem model of data is as follows:

- *Level* L_0. Data, obtained from Sentinel-2 satellite and presented as a 32×32 px fragment in three spectral channels – 2 (blue), 4 (red), 12 (short-wave infrared range). Spring (April-May) images of the test area were taken as a basis since they demonstrate landslide process manifestation in the most pronounced way.

- *Level* L_1. Fragments of satellite images in the visual spectral range with the size of 32×32 px, algorithmically obtained through MapBox API in the 14th scale of tile representation 8.

- *Level* L_{SL}. Fragments of a digital landscape map of Mordovia GIS, corresponding to square areas with the side of 1 km and discretized to 32×32 px screens.

- *Level* L_{SE}. Fragments of a normalized altitude map, plotted on the basis of Google Elevation API and having the shape of a 32×32 px square with the resolution of 10 meters per pixel.

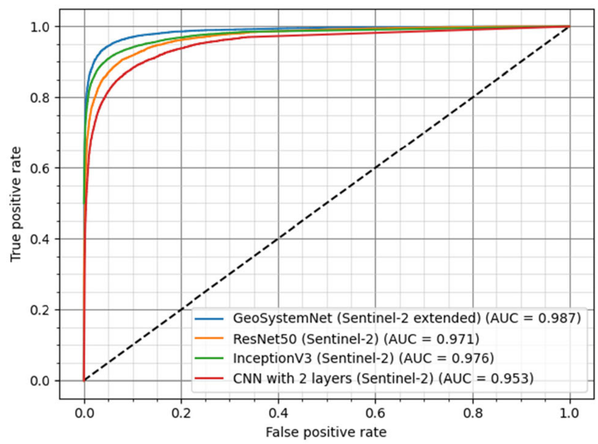

FIGURE 12. ROC curve of the used algorithms of extracting areas with development of landslide processes.

Figure 11 presents a digital landscape map, plotted on the basis of Mordovia GIS. The geosystems that are presented on the map and extracted in the process of landscape mapping are described by common origin and development, uniformity of interaction between superficial deposits, relief forms, water and geochemical regimes and, as a consequence, by similar morphological structures of the soil profile, water-air and thermal conditions of soils, content and reserves of humus and nutrient substances. The main mapping objects are tracts and geographical localities, united in landscapes and systematized in typological complexes – classes, groups, types, genera and kinds.

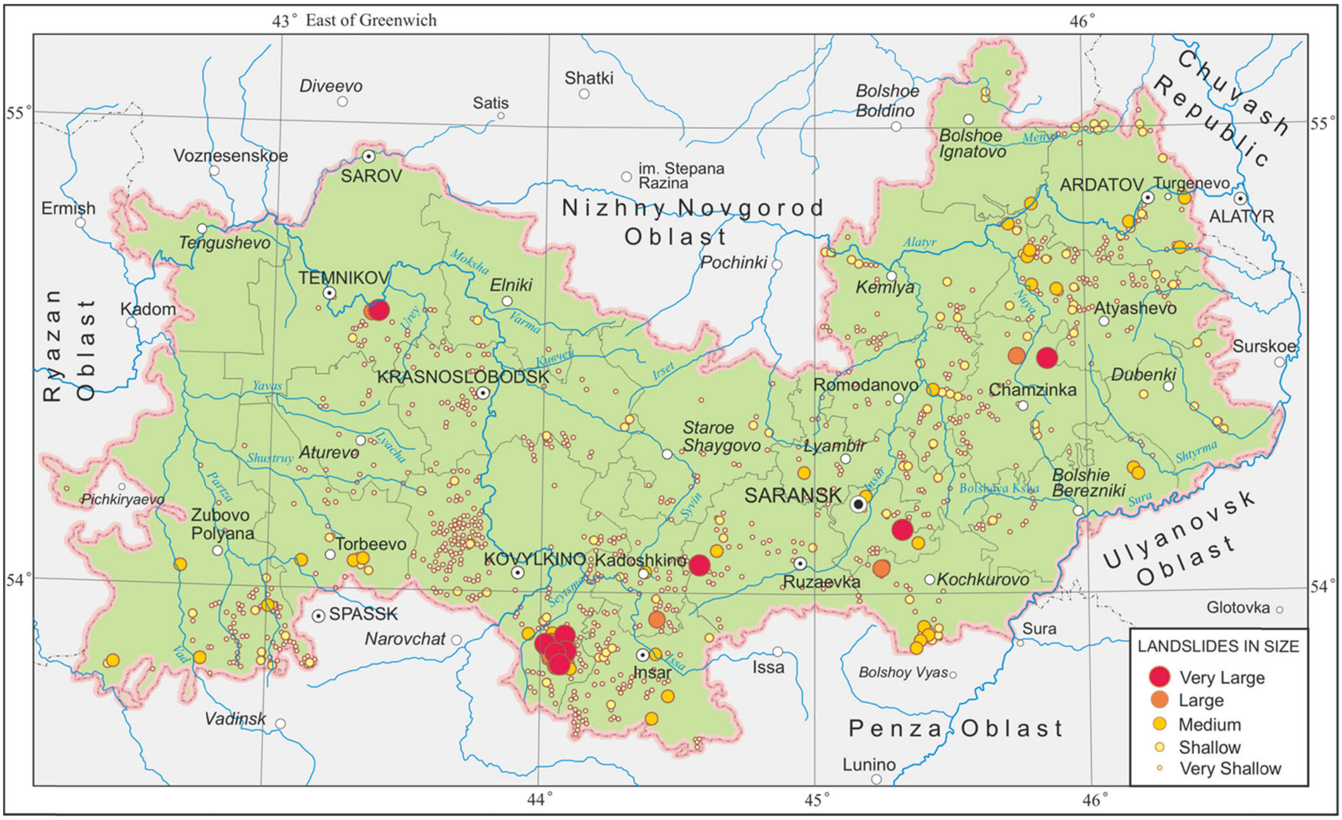

FIGURE 13. Landslide map, plotted upon results of the geospatial data analysis.

24 landscapes are marked out on the Mordovia landscape map – their structure is formed by 19 types of localities, divided into 43 genera of tracts. Landscapes of broadleaved forests and forest-steppes of secondary morainic and erosion-denudation plains are characterized by a slope change of localities and tracts from forest cameo and water-dividing areas and areas near the water-divide to meadow-steppe geocomplexes of lower (near-valley) slope parts and river valleys. On the aquaglacial plains, a cell spatial structure of mixed-wood landscapes is prevalent due to poor development of slope processes.

Machine analysis was to solve a binary classification task of finding out if the area is exposed to landslides. Source data of the level L_0 were expertly labeled on the basis of materials of the satellite imagery Sentinel-2, and then they were algorithmically extended by data of the levels L_1, L_{SL}, L_{SE}.

Figure 12 presents ROC curves, that reflect correlation between sensitivity and specificity of the classification algorithm under variation of the decision rule threshold, for GeoSystemNet, trained on the geosystem data model described above, and for ResNet50, InceptionV3 and Shallow CNN, trained on Sentinel-2 data.

Application of the geosystem approach to the task of detecting lands that are characterized by development of landslide processes with deep learning technologies allowed to raise the classification precision by 5%.

Classification results allow to plot a map of landslide process development in the Republic of Mordovia (Fig. 13). A total of 1,370 landslides were detected. The following types were singled out by sets of features on the basis of expert analysis: block slides, debris flows, topples, viscoplastic landslides, flow slide landslides and slides.

Mapping results showed that highland (higher than 245 m) cameo and water-dividing masses of an axial region of the Volga Upland are exposed to landslide formation least of all – there, landslides are few and far between.

Most of the landslide geosystems are located within the altitude interval from 120 to 250 metres with the maximum concentration on the absolute elevation of 141 – 210 m. It should be noted that distribution of springs has a similar regularity. Highland areas (more than 250 m) have low spring run-off.

The most evident morphostructural elements on the multi-zone satellite images are lineaments. Geodiagnostics of EPF development considers them as areas with high fracture of rock formations and collectors of groundwater flow. In order to identify regularities of natural differentiation, a map of lineament density is plotted upon results of their interpretation, followed by a comparison with a structural and tectonic map, plotted on the basis of geophysical data interpretation.

For the training area under study, significant activity of EPF is typical for regions with prevailing orthogonal systems of lineaments. Prevalence of short dashes on a satellite image may indicate that this region has a high probability of active development of the processes. It is important to mark out some common regularities of landslide propagation – confinedness to the slopes of southwest, southern and southeast expositions and stem slopes of river valleys that form discharge areas of interlayer water.

Groups of geosystems represent the elements of groundwater filtration areas – regions of intake, transit and discharge. In the structure of natural differentiation, it is expressed in formation of geosystems with different degrees of moisture.

Regularities of landscape cover differentiation at the level of geosystem genera are determined by morphosculptural forms of the relief and the deposits they are formed of.

V. CONCLUSION

The findings, presented in the article, allow us to draw the following conclusions:

1) The main importance of the approach to geospatial data analysis by means of deep learning, presented in the paper, rests on in the use of the geosystem approach for profitable augmentation of the training dataset and development of GeoSystemNet deep model that is capable of efficient analysis of these data. The presented approach gains the main advantages in conditions of geospatial training data scarcity. It also allows to approximate to the precision of deeper models by means of a certain-capacity model through analysis of additional automatically obtained information.

2) One of the advantages of the presented GeoSystemNet model rests on a large number of freedom degrees, admitting its flexible configuration contingent upon the task to be solved. Variable parameters of the model include the number of model inputs, the number of levels of feature extraction by Unit Lx, the capacity and number of feature merging modules Merge H_N and hyperparameters of the multi-layer perceptron that makes a resulting decision.

3) Application of GeoSystemNet model for classification of EuroSAT, algorithmically augmented within the geosystem approach, allowed to raise the classification precision in conditions of training data scarcity (splitting the training set into training and test ones at the ratio from 10:90 to 40:60) by 9%, and to show the classification precision with larger volume of training data (by 3%) which is slightly inferior in comparison with different deep models.

4) Synthesis of the geosystem approach with deep learning capabilities enables to optimize the process of online diagnostics of exogeodynamic process development, establishment of landscape differentiation regularities and development of exogeodynamic processes. Detecting lands that are characterized by development of landslide processes in the test area "Mordovia" showed that in the area of interaction of forest-steppe geosystems of the bedded-layer Volga Upland and forest landscapes of the layered Oka-Don Lowlands, the main regularities of the development of exogeodynamic processes are observed in the area of meadow-steppe and forest landscapes. Application of the presented approach allowed to raise the precision of extracting landslides by 5% in comparison with the classical approach of using deep models for remote sensing data analysis.

5) Application of the geosystem approach following the methods, presented in the article, to the task of improving efficiency of remote sensing data classification in conditions of data scarcity requires an individual approach to model configuration and organization of its training. The process of selecting sources of information on host geosystems has a large number of solutions – these can be both remote sensing data of different scale and resolution and fragments of electronic maps of different type. It is necessary to fully realize the responsibility when selecting data on host geosystems – a wrong decision will lead to a situation when an extended training set will rather misinform than raise GeoSystemNet training model efficiency and even lower metrics of model efficiency and classification quality in general.

REFERENCES

[1] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015, doi: 10.1038/nature14539.
[2] R. A. Schowengerdt, *Remote Sensing: Models and Methods for Image Processing*, 3rd ed. Orlando, FL, USA: Academic, 2006, pp. 387–456.
[3] Y. Bengio and Y. LeCun, "Scaling learning algorithms towards AI," *Large-Scale Kernel Mach.*, vol. 34, no. 5, pp. 1–41, 2007.
[4] X. X. Zhu, D. Tuia, L. Mou, G.-S. Xia, L. Zhang, F. Xu, and F. Fraundorfer, "Deep learning in remote sensing: A comprehensive review and list of resources," *IEEE Geosci. Remote Sens. Mag.*, vol. 5, no. 4, pp. 8–36, Dec. 2017, doi: 10.1109/MGRS.2017.2762307.
[5] *Google Scholar*. Accessed: Aug. 1, 2020. [Online]. Available: https://scholar.google.ru/
[6] L. Zhang, L. Zhang, and B. Du, "Deep learning for remote sensing data: A technical tutorial on the state of the art," *IEEE Geosci. Remote Sens. Mag.*, vol. 4, no. 2, pp. 22–40, Jun. 2016, doi: 10.1109/MGRS.2016.2540798.
[7] Z. Zhang, Q. He, J. Gao, and M. Ni, "A deep learning approach for detecting traffic accidents from social media data," *Transp. Res. C, Emerg. Technol.*, vol. 86, pp. 580–596, Jan. 2018, doi: 10.1016/j.trc.2017.11.027.
[8] W. Zhao and S. Du, "Spectral–spatial feature extraction for hyperspectral image classification: A dimension reduction and deep learning approach," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 8, pp. 4544–4554, Aug. 2016, doi: 10.1109/TGRS.2016.2543748.
[9] X. Hu and Q. Weng, "Estimating impervious surfaces from medium spatial resolution imagery using the self-organizing map and multi-layer perceptron neural networks," *Remote Sens. Environ.*, vol. 113, no. 10, pp. 2089–2102, Oct. 2009, doi: 10.1016/j.rse.2009.05.014.
[10] N. Kussul, M. Lavreniuk, S. Skakun, and A. Shelestov, "Deep learning classification of land cover and crop types using remote sensing data," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 5, pp. 778–782, May 2017, doi: 10.1109/LGRS.2017.2681128.
[11] Y. Chen, X. Zhao, and X. Jia, "Spectral–spatial classification of hyperspectral data based on deep belief network," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 6, pp. 2381–2392, Jun. 2015, doi: 10.1109/JSTARS.2015.2388577.
[12] C. Zhao, X. Li, and H. Zhu, "Hyperspectral anomaly detection based on stacked denoising autoencoders," *J. Appl. Remote Sens.*, vol. 11, no. 4, p. 1, Sep. 2017, doi: 10.1117/1.JRS.11.042605.
[13] C. Tao, H. Pan, Y. Li, and Z. Zou, "Unsupervised spectral–spatial feature learning with stacked sparse autoencoder for hyperspectral imagery classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 12, pp. 2438–2442, Dec. 2015, doi: 10.1109/LGRS.2015.2482520.
[14] W. Li, H. Liu, Y. Wang, Z. Li, Y. Jia, and G. Gui, "Deep learning-based classification methods for remote sensing images in urban built-up areas," *IEEE Access*, vol. 7, pp. 36274–36284, 2019, doi: 10.1109/ACCESS.2019.2903127.
[15] Z. Lin, Y. Chen, X. Zhao, and G. Wang, "Spectral-spatial classification of hyperspectral image using autoencoders," in *Proc. 9th Int. Conf. Inf., Commun. Signal Process.*, Tainan, Taiwan, Dec. 2013, pp. 1–5, doi: 10.1109/ICICS.2013.6782778.

[16] R. Dong, X. Pan, and F. Li, "DenseU-Net-based semantic segmentation of small objects in urban remote sensing images," *IEEE Access*, vol. 7, pp. 65347–65356, 2019, doi: 10.1109/ACCESS.2019.2917952.

[17] W. Li, H. Fu, L. Yu, P. Gong, D. Feng, C. Li, and N. Clinton, "Stacked autoencoder-based deep learning for remote-sensing image classification: A case study of african land-cover mapping," *Int. J. Remote Sens.*, vol. 37, no. 23, pp. 5632–5646, Dec. 2016, doi: 10.1080/01431161.2016.1246775.

[18] D. Ienco, R. Gaetano, C. Dupaquier, and P. Maurel, "Land cover classification via multitemporal spatial data by deep recurrent neural networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 10, pp. 1685–1689, Oct. 2017, doi: 10.1109/LGRS.2017.2728698.

[19] H. Wu and S. Prasad, "Convolutional recurrent neural networks forhyperspectral data classification," *Remote Sens.*, vol. 9, no. 3, p. 298, Mar. 2017, doi: 10.3390/rs9030298.

[20] L. Mou, P. Ghamisi, and X. X. Zhu, "Deep recurrent neural networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 7, pp. 3639–3655, Jul. 2017, doi: 10.1109/TGRS.2016.2636241.

[21] D. Powers, "Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation," *J. Mach. Learn. Technol.*, vol. 2, no. 1, pp. 37–63, Jan. 2011, doi: 10.9735/2229-3981.

[22] *EuroSAT: Land Use and Land Cover Classification with Sentinel-2*. Accessed: Aug. 1, 2020. [Online]. Available: https://github.com/phelber/EuroSAT

[23] P. Helber, B. Bischke, A. Dengel, and D. Borth, "EuroSAT: A novel dataset and deep learning benchmark for land use and land cover classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 7, pp. 2217–2226, Jul. 2019, doi: 10.1109/jstars.2019.2918242.

[24] D.-E. Kim, P. Gourbesville, and S.-Y. Liong, "Overcoming data scarcity in flood hazard assessment using remote sensing and artificial neural network," *Smart Water*, vol. 4, no. 1, p. 2, Jan. 2019, doi: 10.1186/s40713-018-0014-5.

[25] B. V. Sochava and V. V. U. O. geosistemah, *[Introduction to the doctrine of Geosystems]*. Novosibirsk, Russia: Nauka. Sibirskoe Otdelenie, 1978, pp. 11–28.

[26] V. A. Nikolaev, *Landshaftovedenie [Landscape Science]*. Moscow, Russia: MSU Publishing House, 2000, pp. 8–22.

[27] *MapBox–API*. Accessed: Aug. 1, 2020. [Online]. Available: https://docs.mapbox.com/api/

[28] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1251–1258, doi: 10.1109/CVPR.2017.195.

[29] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," 2015, *arXiv:1502.03167*. [Online]. Available: http://arxiv.org/abs/1502.03167

[30] Y. L. Boureau, J. Ponce, and Y. LeCun, "A theoretical analysis of feature pooling in visual recognition," in *Proc. 27th Int. Conf. Mach. Learn. (ICML)*, Haifa, Israel, Jun. 2010, pp. 111–118.

[31] *GeoSystemNet on GitHub*. Accessed: Aug. 28, 2020. [Online]. Available: https://github.com/yamashkinsa/geosystemnet

STANISLAV A. YAMASHKIN received the engineering degree (Hons.) in automated systems of information processing and control from Mordovia State University, in 2012, and the M.Sc. degree in ecology and nature management, in 2015. In 2016, he defended his Ph.D. thesis at Penza State University. He was a Project Manager in a number of grants funded by Russian Foundation for Basic Research in 2019, the Russian Geographical Society in 2015, 2016, and 2020, and winner of the UMNIK program administered by Fund for the Promotion of the Development of Small Forms of Enterprises in the Scientific and Technical Field in 2016. His research interests include spatial data analysis, system analysis, web-technologies, and software engineering.

ANATOLIY A. YAMASHKIN defended his Ph.D. thesis in geographical sciences, in 1985. He received the doctor's degree in geographical sciences, in 2000. In 2002, he also received the title of Full Professor. He is currently a Full Member of the Russian Geographical Society, the Chairman of the Branch of the Russian Geographical Society in the Republic of Mordovia, and the Dean of Geography Faculty of Mordovia State University. He is also an Honored Scientist of the Republic of Mordovia. In 2002, he founded the Department of Geoecology and Landscape Planning in Mordovia State University, which in 2014 was transformed into the Department of Land Management and Landscape Planning. He is known as the Chairman of the Editorial Board of the Atlas of the Republic of Mordovia and the ideologist of the publication. He was the scientific manager of grant projects of the Russian Humanitarian Science Foundation from 2002 to 2007, the Russian Foundation for Basic Research, from 2002 to 2004, and the Russian Geographical Society, from 2011 to 2012. He is also a specialist in the field of geosystem analysis, landscape planning of territories, and geoecological diagnostics of the conditions of natural-social-production systems.

VICTOR V. ZANOZIN received the M.Sc. degree in ecology. He is currently pursuing the Ph.D. degree with the Department of Geology and Geography, Astrakhan State University. He was a GIS Specialist in Astrakhan natural biosphere reserve. His research interests include cartography, landscape mapping, GIS, and web-GIS. The participant of numerous conferences (including those backed by the Roscosmos State Corporation for Space Activities, the United Rocket and Space Corporation, and the Space and Telecom Cluster of the Skolkovo Innovation Center). He is also a member of the Russian Geographical Society and the International Association for Landscape Ecology (IALE).

MILAN M. RADOVANOVIC is currently the Director and a Principal Research Fellow with the Geographical Institute "Jovan Cvijic", Serbian Academy of Sciences and Arts (SASA), Belgrade, defended a Doctoral Thesis, titled "Impact of Relief and Atmospheric Circulation on Climate Differentiation in Serbia" in 2001 at the Faculty of Geography, University of Belgrade. Decided by the National Council for the Geo-heritage in 2002, he was appointed the Head of the work group for climate elements of the geo-heritage. He has been an Appointed Member of the Administrative Committee of Geographical Institute "Jovan Cvijic", SASA, since 2003. He acquired the title of Senior Research Associate in 2011, has become the National Coordinator of the Regional Network of the Balkan, Black Sea and Caspian Sea countries for Space Weather Studies, in 2015. Decided by the Executive Board of the SASA at the meeting held in 2016, he was an Elected Member of the Board for the Study of Kosovo and Metohija. His research interests include physical geography and natural hazards and tourism.

ALEXANDER N. BARMIN was born in Astrakhan. He graduated from the Astrakhan State Pedagogical Institute, in 1983. From 1990 to 1993, he studied in the graduate school of the Institute of Ecology of the Volga Basin, Russian Academy of Sciences. He received the Ph.D. degree in biological sciences from Voronezh State University, in 1993, for the defense of the thesis on the theme "Dynamics of Grassy Vegetation of the Volga River Delta in Conditions of Increased Water Flow". In 1998, he was confirmed as an Assistant Professor. The scientific degree of Doctor of Geographical Sciences, specializing in physical geography and biogeography, geography of soils, and geochemistry of the landscape, was awarded after defending his doctoral dissertation, in 2002.