Neural-Based Hierarchical Approach for Detailed Dominant Forest Species Classification by Multispectral Satellite Imagery

Svetlana Illarionova, Alexey Trekin, Vladimir Ignatiev, and Ivan Oseledets

***Abstract*—Among different forest inventory problems, one of the most basic is defining dominant species. These data are crucial in forest management to determine forest category, and a cheaper remote sensing-based approach would be a useful supplement to field surveys. We used WorldView multispectral satellite imagery to address this problem as an image segmentation task dividing the image into regions with particular dominant species. Neural networks have recently become one of the most useful tools for this kind of problem, including incomplete or erroneous training labels. However, it is still challenging to distinguish between such similar patterns as different forest compositions. To handle this, we represented the multiclass forest classification problem as a hierarchical set of binary classification tasks, which allowed us to reach better results with both high- and medium-resolution satellite imagery. We also examined supplementary data, such as tree height, to improve the species classification results for wider tree age diversity. We conducted experiments considering six neural network architectures to find the best one for each task in the hierarchical decomposition. The proposed approach was tested on sample territories in Leningrad Oblast of Russia, for which the field-based observations were acquired and made publicly available as a single dataset. The proposed approach showed significantly better results (average F1-score 0.84) than multiclass classification (average F1-score 0.7).**

***Index Terms*—Convolutional neural network (CNN), forest species classification, remote sensing, semantic segmentation.**

I. Introduction

ALGORITHMIC analysis of remote sensing data allows for solving a wide range of tasks that previously required high professional skills and were time consuming. One of these challenges is forest species classification, which is commonly considered a dominant species classification problem. A forest's dominant species is the one that includes the majority of the timber stock of the stand, and forest management depends on this as a primary characteristic.

The industrial approach to the forest inventory still consists of several methods, including manual and partly automated satellite mapping, LIDAR data analysis, and ground-based surveys. Since the beginning of computer vision method development, many works have aimed to replace some stages with automatic remote sensing imagery analysis.

It is challenging to compare the performance of different methods proposed in the papers due to their region specificity and evaluation data inaccessibility for other researchers. Therefore, a comparison of the declared metrics cannot often explain what is better, and a literature survey is mostly qualitative. The presented work partly addresses this issue, as we provided the training markup and the images' IDs to compare achieved results in future studies.

A common choice of remote sensing data is medium-resolution multispectral satellite imagery (Landsat or Sentinel), which is freely available and has a good revisit time. This allows researchers to obtain images for any region of interest with relative ease. The multispectral channels in visible and infrared wavelengths provide a good deal of information about surface reflection properties. This data type is used in many research works both for single satellite images [1]–[3] and time series [4], [5]. Although, it makes it possible to automatically process the data for vast territories and with decent accuracy, it does not produce high-resolution semantic maps, which can be useful for the precise estimation of timber stock.

A significant number of works cover the usage of airborne multispectral or hyperspectral sensing for forestry inventory classification [6]–[9], and many of these works leverage a combination with LIDAR scans. It allows for evaluating different forest biomass components [10] and estimating timber stock [11]. In [12], they addressed the challenge of savanna tree species classification in South Africa. The basic premise is the heterogeneous nature of the considered region. Therefore, tree height was utilized as structural information to make classification more robust. However, this is not suitable for the preliminary large-area examination due to the high costs of the data and the need for expeditions to the area of interest for imagery acquisition.

A significant source in terms of information depth and availability is very high spatial resolution satellite imagery, such as WorldView satellite data (about 2-m spatial resolution). In [13], they classified deciduous-dominated forest species through three-seasonal WorldView images. In [14], they leveraged a single WorldView high-resolution satellite image for species and age classification. The scope of the work included both object- and pixel-based approaches. Sunlit areas of tree crowns

Manuscript received June 15, 2020; revised November 30, 2020 and December 19, 2020; accepted December 28, 2020. Date of publication December 31, 2020; date of current version January 21, 2021. *(Corresponding author: Svetlana Illarionova.)*

The authors are with the Skolkovo Institute of Science and Technology, 121205 Moscow, Russia (e-mail: s.illarionova@skoltech.ru; a.trekin@skoltech.ru; v.ignatiev@skoltech.ru; I.oseledets@skoltech.ru).

Digital Object Identifier 10.1109/JSTARS.2020.3048372

presented dataset objects. For such a polygon, a particular species class was ascribed, keeping each object's homogeneity. Moreover, only instances of approximately the same age were chosen for the study, making the samples within a class less diverse. In [15], they used QuickBird images (a 2.44-m spatial resolution) to classify forest species. Still, a relatively small number of works have given preference to such high-resolution satellite data instead of unmanned aerial vehicle (UAV) images.

Although classical machine learning methods, such as support vector machine [16] and random forest [17], are used in many remote sensing classification studies [5], [12], [14], [18], [19], other works consider newer approaches. In recent years, convolutional neural networks (CNNs) have become a principal method for many computer vision problems, including image classification, segmentation, and object detection. CNNs are applicable in different spheres, and the remote sensing area is no exception [20], [21]. Deep neural networks showed accurate results in the task of deciduous and coniferous classification [20], [22] and other forest inventory characteristic estimation [23] using LIDAR sensing data.

Hierarchical problem decomposition can often be implemented in various applied tasks of a particular nature containing subclasses. It has performed successfully in medical problems [24], [25]. In [26], they implemented a hierarchical multilabel classification for diatom images using a single predictive clustering tree. Just a few studies considered the hierarchical approach for forest species classification [27], [28]. However, in these works, UAV or airborne data was used with a spatial resolution higher than 0.3 m. The classification approaches were maximum likelihood classification techniques and object-based image classification [29]. Thus, all considered forest species classification studies based on satellite images rely exactly on the classical multiclass classification approach [5], [9], [14], [15].

The goal of this work is to enhance the spatial detail of dominant forest species estimation using the high-resolution WorldView satellite imagery (2 m per pixel). We have chosen this kind of remote sensing data because it can combine the high availability of satellite imagery (though it is not as high as with moderate resolution) and the spatial precision of aerial imaging. In contrast with most of the work in this area, we did not only concentrate on homogeneous forest stands of approximately the same age. Thus, we aimed to provide a more robust solution applicable to real-life conditions.

We aimed to make the following contribution.

1) To improve forest species multiclass image segmentation by splitting the problem into a hierarchy of binary segmentation problems.
2) To study the forest height maps usefulness as supplementary data for the forest species classification problem.
3) To prepare an open-source dataset for the dominant species segmentation problem—the lack of relevant markup causes obstacles in this sphere of study, so open-access data are crucial.

In the proposed study, six neural network architectures were considered to find the best one for each classification task. In addition, to confirm the developed approach's applicability, we tested our method modifications on moderate-resolution data (Sentinel images), which has lower resolution but is more available due to its being freely available for download.[1]

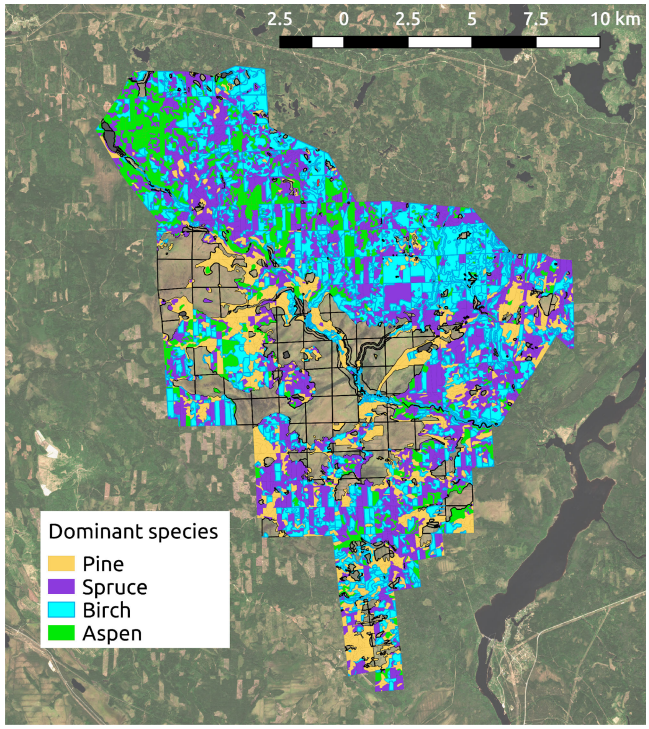

Fig. 1. Classes markup of study area.

II. Dataset

A. Study Area

The dataset for this work was created using ground-based observations of Leningrad Oblast of Russia during the 2018 year (see Fig. 1). The total area is around 20 000 hectares. The coordinates of this region are between $33^{\circ}42'$ and $33^{\circ}76'$ longitude and between $60^{\circ}78'$ and $61^{\circ}01'$ latitude. The region's climate is humid. The coldest day of the year is in February, with a temperature between 13 and 24°F [30]. The topography is flat. The vegetation cover is mixed and includes deciduous and conifer tree species.

B. Reference Data

The study area was split into small regions representing individual forest stands. The term "forest stand" in forest inventory instructions defines a contiguous forested area sufficiently uniform in essential characteristics to distinguish it from adjacent communities. Each stand is described by several aspects; the most important for this work are the following.

1) Forest composition, i.e., the percentage of each tree species, denoted with a stride of 10% of the relative timber volume (the composition is given in percentage points, each representing 10% of the total timber volume).

[1]Code for experiments and labeled data is available at https://github.com/LanaLana/forest_species

TABLE I
DATASET STATISTICS FOR INDIVIDUAL REGIONS

	area, ha	percentage
aspen	1270.1	14%
birch	2407.7	26%
spruce	3567.2	38%
pine	2063.8	22%

TABLE II
WORLDVIEW IMAGES

	Image ID	Date	Off-nadir angle
0	10300100812E1700	29.07.2018	21
1	1030010081253D00	29.07.2018	29
2	10300100828A7D00	19.07.2018	26
3	103001008067D100	19.07.2018	22
4	1030010080790B00	18.07.2018	22
5	10300110829C9600	18.07.2018	32
6	103001007DCF9400	12.05.2018	14
7	103001007ECC6B00	12.05.2018	18

2) Average tree height for each of the primary forest components in the forest composition.
3) Average tree age for each of the primary forest components in the forest composition.

The rest of the parameters leveraged for the forest analysis are not considered in the current research.

The dominant species is the one that has the highest percentage, and it is the target value that we want to evaluate in this work. Of course, there are situations when two or more forest species have the same or a similar percentage. This case is defined when the difference between the dominant and the second species is not greater than 1% point, and these stands are treated as "mixed forests." The composition of mixed forests is beyond this article's scope, so such stands were excluded from both training and test sets.

The dataset contains forest stands with four classes of dominant species: 38% spruce (*Picea spp.*), 14% aspen (*PÓpulus tremula spp.*), 26% birch (*Betula spp.*), and 22% pine (*Pónus spp.*) (see Table I). The rest of the study area species are less distributed and do not compose the stands as a dominant species. It is worth noting that the "dominant species" in forestry does not exactly match the biological term "species" and is connected mostly with the timber class and quality. In this research, the existing forest inventory standards were followed, and this inventory does not distinguish between species within a genus and treats the whole genus as a single class.

C. Satellite Data

WorldView 2 and 3 multispectral imagery with eight spectral bands was downloaded from GBDX [31]. The spatial resolution was about 2 m per pixel. Sentinel imagery with 13 spectral bands and a spatial resolution of about 10 m per pixel was downloaded from SentinelHub [32]. All images were from the high vegetation period from May to August. Image acquisition dates and catalogue IDs are presented in Tables II and III.

Dataset consists of georeferenced satellite images in the format of 8-b TIFF files and forestry inventory data converted into raster per pixel masks for each class.

TABLE III
SENTINEL IMAGES

Image	ID	date
0	L1C_T36VWN_A007126_20180718T092026	18.07.18
1	L1C_T36VWN_A016206_20180730T090554	30.07.18

The additional challenge was posed by the temporal mismatch between imagery and markup. Current forest inventory information is sparsely available. Thus, some forest areas were felled after the ground-based observations. To deal with this, we utilized a previously trained neural network that performs forest segmentation. It produces an up-to-date forest mask for the images and excludes the derived nonforested areas from the training and validation sets. We additionally cleaned the test set manually.

III. METHODS

A. Problem Definition

As described in Section II, we treated an individual forest stand as a homogeneous region with a common characteristic within its area. The aim was to develop a method that could produce high-resolution semantic maps outlining forest stands. Thus, the problem was formulated as image segmentation: to assign a species class to every pixel in the image. The background classes were excluded from the dataset before training and did not appear at the test time. The following fact complicated the problem. Forest stands can have inconsistency and include visible parts of the nondominant species. These parts should be segmented as a separate stand of another dominant species, but the training data do not support it, as the markup is completely standwise.

B. Neural Networks for Image Segmentation

As the most recent computer vision advances are connected with the novel neural network architectures, it is vital to select a suitable one for the given task and available computational resources. Since the task was formulated as a multiclass image segmentation problem, a fully convolutional architecture was considered, such as U-Net [33] or a feature pyramid network (FPN) [34]. Both of them show good image segmentation performance, including remote sensing data, with FPN being more suitable for multiclass segmentation. These architectures are constructed in an encoder–decoder fashion with skip connections, which allows us to use various convolutional encoders. Modern architectures outperformed the original VGG encoder used in [33], so the first variant was ResNet [35], used by Lin *et al.* [34]. As counterparts, we used Inception-ResNet-v2 [36] and EfficientNet [37] as one of the most resent and advanced architectures, showing state-of-the-art results at the ImageNet benchmark [38]. To comply with computational resource restrictions, the model size was limited to ResNet-34 and EfficientNet-B3 correspondingly. The models' architecture implementation was based on Yakubovskiy [39].

C. Image Preprocessing

As Sentinel images were contrast-enhanced and had a value range of $[0 : 255]$ in each channel, they were scaled as

$$I' = I/255 \tag{1}$$

where I and I' are intensities before and after the normalization, respectively.

To ensure relative brightness uniformity for different images, we performed minimum–maximum brightness normalization to the range [0, 1], as in [40].

The WorldView images have a wider dynamic range, different for each channel, so contrast enhancement was also included in the scaling formula to suppress the darkest and the brightest regions that lie beyond three standard deviations from the mean value

$$m = \max(0, \text{mean}(I) - 3 * \text{std}(I)) \tag{2}$$

$$M = \min(I\_\text{max}, \text{mean} + 3 * \text{std}) \tag{3}$$

$$I' = (I - m)/(M - m) \tag{4}$$

where mean and std are the mean and standard deviation of the image, respectively. In (2) and (3), we calculate *m* and *M* (minimum and maximum of the preserved dynamic range). In (4), values are scaled to 0 and 255 linearly. The values outside the [*m*, *M*] range are clipped. The standardization of the imagery according to the whole dataset statistics proves profitable for the neural network training compared to a simple scaling of the entire value range [41].

There are two ways to compute mean and standard deviation values: for all channels simultaneously or individually for each band. The advantage of the first type is that ratios between the channel values stay constant, which might be necessary for a more in-depth nature processes evaluation. On the other hand, when statistics within each channel are computed, connections between the same channels of different images are more robust, and it can be useful for algorithm adaptability.

D. Dataset Augmentation

The dataset augmentation is a common technique that can improve the robustness of the neural network. In the considered case, the spatial transforms were applied to the training images with 50% probability: rotation with a 90° step, a vertical and horizontal flip, and a zoom-in and -out within 20%.

E. Oversampling

To handle the class imbalance, we added extra weights for the smaller classes during loss computation. For this variant, weighted cross-entropy (WCE) was computed, and optimal weights were estimated according to the class distribution in the training set.

The other problem was that the label for the dominant species property was the same, whether there were 50% or 90%. Still, the former represented a more "dirty" markup for the segmentation, as about half of the pixels represented nondominant species. We managed to enforce the training on more clean samples by increasing the probability of the samples with a higher dominant species percentage. Species distribution is provided for each forest region in Table IV.

TABLE IV
DATASET STATISTICS FOR INDIVIDUAL REGIONS (DOMINATED SPECIES BY THRESHOLD), AREA IN HECTARE

threshold	pine	birch	aspen	spruce
0.5	2063.8	2407.7	1270.1	3567.2
0.6	1781.9	951.9	659.6	2390.6
0.7	1540.9	463.5	235.8	1350.2
0.8	1234.3	178.7	84.2	643.7

F. Problem Decomposition

The baseline approach used multiclass segmentation, where the output layer of the neural network had a number of outputs that was equal to the number of classes. The argmax (arguments of the maximum) of these values was treated as a class label for a pixel.

The approach modification was based on the fact that forest species classification has an explicit hierarchy: classes are divided into coniferous and deciduous tree species. Therefore, it was reasonable to decompose the problem. The hierarchical solution represented the multiclass segmentation as a set of binary segmentation problems. The multiclass segmentation map was obtained by consistently applying the method and aggregating the results (see Fig. 2).

The stages scheme of the hierarchical segmentation process is depicted in Fig. 3. We used the "parent" data obtained from the previous stages of the processing at each step. For example, to segment coniferous and deciduous forest stands, the forest mask was utilized to exclude nonforest regions from the observation. During the model training, this "parent" data were used as a mask for the loss function computation. The training loss was calculated within the parent class areas only because, for the same example, there was no need to rely on the nonforested regions to distinguish between the forest types. During the inference, the result of the binary segmentation was multiplied by the "parent" mask.

We also compared this approach with "one versus all" classification, where a set of separate neural network models is trained to predict just one class. All predictions are then aggregated, and the most likely label is ascribed to each pixel.

G. Height Data

It is worth noticing that a part of the intro-class variance is connected with the forest height or age, with a high correlation. The same forest species at different ages shows different patterns (see Fig. 4). As the height data could be obtained from separate sources, we studied the height data's effect on the dominant species classification. The input data were extracted from the same forest inventory characteristics used for training, and it was used as an additional raster band in the network input. This modification also contributed to the method performance in both multiclass tasks and binary segmentation cases.

TABLE V
RESULTS FOR MULTICLASS CLASSIFICATION WITHOUT HEIGHT (F1-SCORE) FOR WORLDVIEW AND SENTINEL (BASELINE) ON VALIDATION. BOLD NUMBERS — THE BEST SCORE (THE CORRESPONDING MODEL WAS CHOSEN FOR THE FINAL RESULTS AGGREGATION)

WorldView						
	Unet + Resnet34	Unet + EfficientNet	Unet + Inceptionresnetv2	FPN + Resnet34	FPN + EfficientNet	FPN + Inceptionresnetv2
aspen	0.39	0.26	0.385	0.35	0.2	0.35
birch	0.79	0.548	0.781	0.76	0.18	0.71
spruce	0.759	0.743	0.754	0.75	0.68	0.76
pine	0.868	0.847	0.859	0.87	0.81	0.859
average	**0.702**	0.599	0.695	0.682	0.47	0.66
Sentinel						
	Unet + Resnet34	Unet + EfficientNet	Unet + Inceptionresnetv2	FPN + Resnet34	FPN + EfficientNet	FPN + Inceptionresnetv2
aspen	0.367	0.356	0.417	0.361	0.219	0.372
birch	0.713	0.687	0.738	0.694	0.258	0.681
spruce	0.717	0.708	0.658	0.721	0.669	0.722
pine	0.841	0.845	0.83	0.853	0.813	0.845
average	0.659	0.649	**0.66**	0.657	0.489	0.655

TABLE VI
RESULTS FOR MULTICLASS CLASSIFICATION WITH HEIGHT (F1-SCORE) FOR WORLDVIEW AND SENTINEL ON VALIDATION. BOLD NUMBERS — THE BEST SCORE (THE CORRESPONDING MODEL WAS CHOSEN FOR THE FINAL RESULTS AGGREGATION)

WorldView						
	Unet + Resnet34	Unet + EfficientNet	Unet + Inceptionresnetv2	FPN + Resnet34	FPN + EfficientNet	FPN + Inceptionresnetv2
aspen	0.38	0.43	0.39	0.42	0.38	0.39
birch	0.78	0.80	0.79	0.79	0.80	0.79
spruce	0.8	0.78	0.76	0.79	0.77	0.74
pine	0.87	0.87	0.85	0.85	0.82	0.84
average	0.707	**0.72**	0.697	0.712	0.692	0.69
Sentinel						
	Unet + Resnet34	Unet + EfficientNet	Unet + Inceptionresnetv2	FPN + Resnet34	FPN + EfficientNet	FPN + Inceptionresnetv2
aspen	0.426	0.414	0.415	0.419	0.371	0.474
birch	0.726	0.733	0.712	0.726	0.748	0.772
spruce	0.745	0.75	0.736	0.753	0.763	0.78
pine	0.837	0.851	0.847	0.844	0.846	0.864
average	0.68	0.687	0.677	0.685	0.682	**0.72**

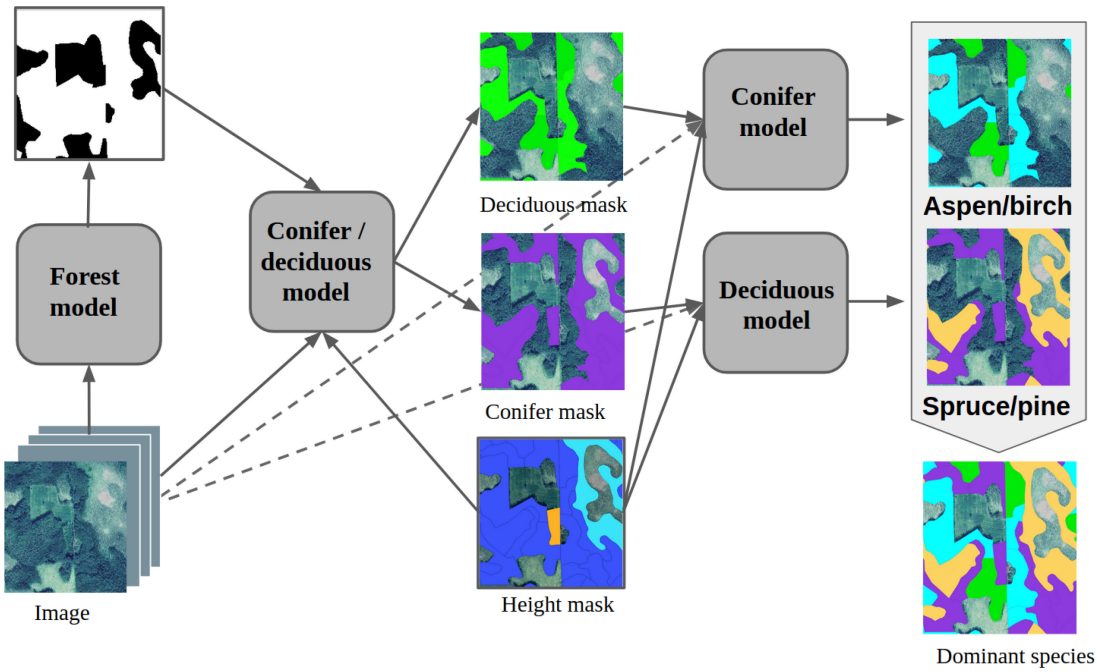

Fig. 2. Hierarchical model structure.

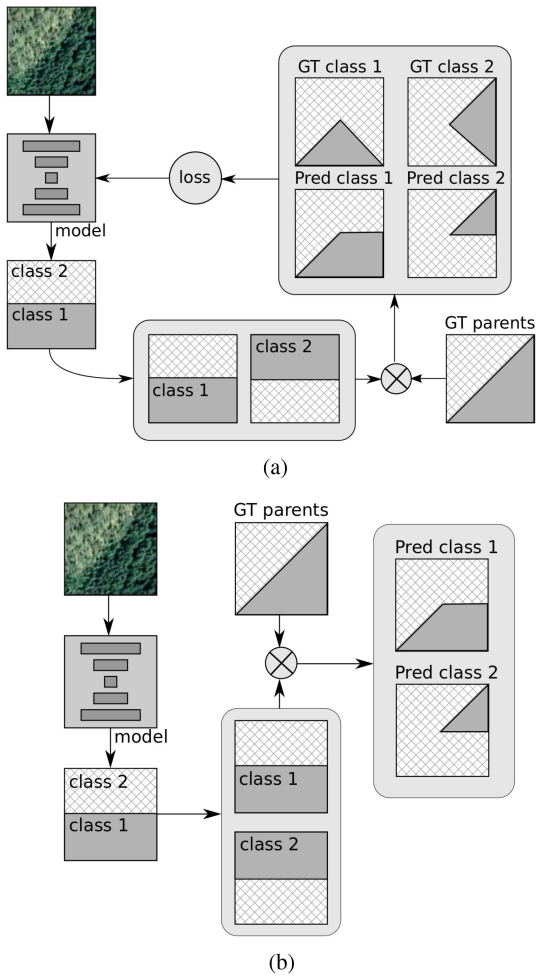

Fig. 3. Data flow through a level of the hierarchical process. (a) Model training. (b) Inference.

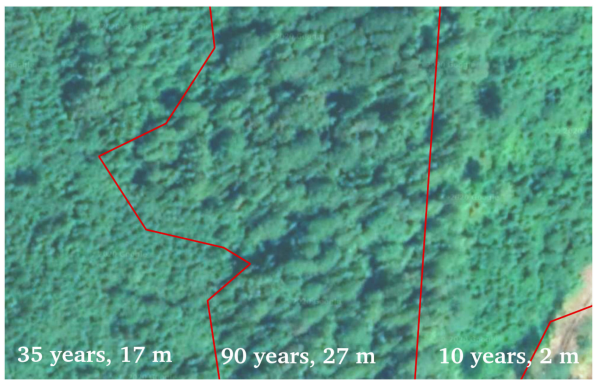

Fig. 4. Example of age and height variance within one species.

IV. Experiments

A. Training

The training of all the neural network models was performed on a PC with GTX-1080Ti GPUs.

The batch size varied from 16 to 30 depending on the architecture's memory restrictions.

During the binary segmentation models' training within the hierarchical segmentation approach, only two particular classes of the current stage were taken into account. Accordingly, the loss function was calculated only over the part of the image corresponding to the parent class of the current stage, as is shown in Fig. 3(a). The total loss for a training batch was normalized to the parent class area in the batch, as shown in in the following:

$$\text{loss} = \frac{\text{WCE} * w * h * b}{N} \tag{5}$$

where w and h correspond to the image crop size, b is the batch size (number of image crops), and N is the number of relevant pixels in the batch.

The final model combined all these approaches.

B. Medium Resolution Data

The same experiments were performed using widely spread in the forest inventory tasks Sentinel-2 data to compare the selected data to other possible sources.

The base model used 13 bands of Sentinel imagery at a spatial resolution from 10 to 60 m. These data are available for free download. The model was trained in the same manner as a model without height for WorldView data. The image crop size was reduced in batch from 256 to 64 to, by giving the field of view the same size, make the training procedure as similar as possible.

C. Evaluation

The dataset was split into training, validation, and test sets in the following proportion: 0.7, 0.15, and 0.15. The validation set was used to choose the best neural network parameters and architecture.

F1-score was utilized to measure the segmentation quality and compare the method variants, for the individual classes and averaged over all the classes.

The metric was computed in a pixelwise way

$$P = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{6}$$

$$R = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{7}$$

$$F1 = \frac{2 * P * R}{P + R} \tag{8}$$

where P is precision, R is recall, TP is true positive (the number of correctly classified pixels of a given class), FP is false positive (the number of pixels classified as a given class while being of another class), and FN is false negative (the number of pixels of a given class missed by the model).

F1-score was computed only for regions covered by species with a domination of more than 0.5, which was described in Section II. When the optimal in terms of the validation dataset architecture for each task had been found, the final models were evaluated using the test set of the images, which did not overlap with the training or validation sets. We also used confusion matrices, as this is a commonly considered accuracy assessment approach in remote sensing image classification [42].

TABLE VII
HIERARCHICAL CLASSIFICATION WITH HEIGHT DATA (F1-SCORE) FOR WORLDVIEW ON VALIDATION BEFORE THE RESULTS AGGREGATION. BLUE NUMBERS — THE BEST SCORE FOR MODELS WITHOUT HEIGHT, BOLD NUMBERS — THE BEST SCORE FOR MODELS WITH HEIGHT (THE CORRESPONDING MODELS WERE CHOSEN FOR THE FINAL RESULTS AGGREGATION)

	Unet + Resnet34		FPN + Resnet34		Unet + EfficientNet		FPN + EfficientNet		Unet + Inceptionresnetv2		FPN + Inceptionresnetv2	
species	without height	height	without height	height	without height	height	without height	height	without height	height	without height	height
aspen / birch	**0.76**	**0.78**	0.69	0.75	0.67	0.745	0.64	0.665	0.69	0.72	0.7	0.7
pine / spruce	**0.92**	0.926	0.9	**0.935**	0.918	0.922	0.876	0.911	0.9	0.916	0.896	0.908
conifer / deciduous	0.86	**0.906**	**0.863**	0.9	0.86	0.9	0.85	0.898	0.859	0.897	0.856	0.9

TABLE VIII
HIERARCHICAL CLASSIFICATION WITH HEIGHT DATA (F1-SCORE) FOR SENTINEL ON VALIDATION BEFORE THE RESULTS AGGREGATION. BLUE NUMBERS — THE BEST SCORE FOR MODELS WITHOUT HEIGHT, BOLD NUMBERS — THE BEST SCORE FOR MODELS WITH HEIGHT (THE CORRESPONDING MODELS WERE CHOSEN FOR THE FINAL RESULTS AGGREGATION)

	Unet + Resnet34		FPN + Resnet34		Unet + EfficientNet		FPN + EfficientNet		Unet + Inceptionresnetv2		FPN + Inceptionresnetv2	
species	without height	height	without height	height	without height	height	without height	height	without height	height	without height	height
aspen / birch	0.68	0.68	0.63	0.66	**0.69**	0.66	0.58	0.63	0.69	0.7	0.675	**0.724**
pine / spruce	0.895	0.9	0.885	0.922	0.895	**0.92**	**0.905**	0.919	0.885	0.915	0.905	0.92
conifer / deciduous	0.83	0.867	**0.84**	0.864	0.835	0.859	0.825	0.846	0.806	0.869	0.833	**0.874**

TABLE IX
HIERARCHICAL APPROACH (1) IN COMPARISON WITH "ONE VERSUS ALL" CLASSIFICATION AND (2) ON TEST DATA (BOTH APPROACHES USE HEIGHT DATA) FROM THE WORLDVIEW DATA

	1	2
aspen	0.72	0.75
birch	0.75	0.48
spruce	0.94	0.71
pine	0.92	0.82
average	0.836	0.69

TABLE X
HIERARCHICAL APPROACH (1) IN COMPARISON WITH "ONE VERSUS ALL" CLASSIFICATION AND (2) ON TEST DATA (BOTH APPROACHES USE HEIGHT DATA) FROM THE SENTINEL DATA

	1	2
aspen	0.79	0.46
birch	0.586	0.56
spruce	0.93	0.75
pine	0.789	0.88
average	0.77	0.667

TABLE XI
FINAL AGGREGATED RESULTS (F1-SCORE) FOR WORLDVIEW TEST DATA

	hierarchy + height	hierarchy	multi-class + height	multi-class
aspen	0.721	0.714	0.773	0.39
birch	0.751	0.649	0.469	0.796
spruce	0.947	0.954	0.764	0.759
pine	0.925	0.87	0.851	0.869
average	**0.836**	0.797	0.716	0.703

TABLE XII
FINAL AGGREGATED RESULTS (F1-SCORE) FOR SENTINEL TEST DATA

	hierarchy + height	hierarchy	multi-class + height	multi-class
aspen	0.79	0.612	0.608	0.586
birch	0.586	0.527	0.441	0.274
spruce	0.93	0.943	0.766	0.692
pine	0.789	0.792	0.855	0.791
average	**0.77**	0.72	0.668	0.58

V. RESULTS AND DISCUSSION

A. Hierarchical Decomposition

We compared hierarchical decomposition with two commonly used image semantic segmentation approaches: multiclass classification and "one versus all." All studies were conducted both for WorldView and Sentinel images to assess the proposed method using different data sources. The results of multiclass classification and hierarchical decomposition before aggregation are reported in Tables V–VIII. As shown in Tables XI and XII, which have the aggregated results, the hierarchical approach allows us to improve model performance in terms of the F1-score for WorldView from 0.716 to 0.836 and for Sentinel from 0.668 to 0.77. "One versus all" classification also shows lower results than those of the hierarchical decomposition depicted in Tables IX and X. For WorldView, there is decline in quality in the F1-score from 0.836 to 0, whereas for Sentinel, that decline is from 0.77 to 0.667. There is no significant difference between multiclass and "one versus all" classification. For WorldView, the difference is 0.716 and 0.69; for Sentinel, it is 0.668 and 0.667. Confusion matrices for WorldView and Sentinel data are shown in Fig. 5. The WorldView prediction quality is higher than that of Sentinel. Moreover, for the WorldView imagery, coniferous and deciduous subclasses are less often ascribed to the wrong parent class.

One of the important issues of the hierarchical approach is that, for each classification task, the most suitable neural network architecture can be chosen.

As is shown in Tables XI and XII, the accuracy of the classification of aspen and birch became more adequate, and the final performance is more satisfying in the context of available markup.

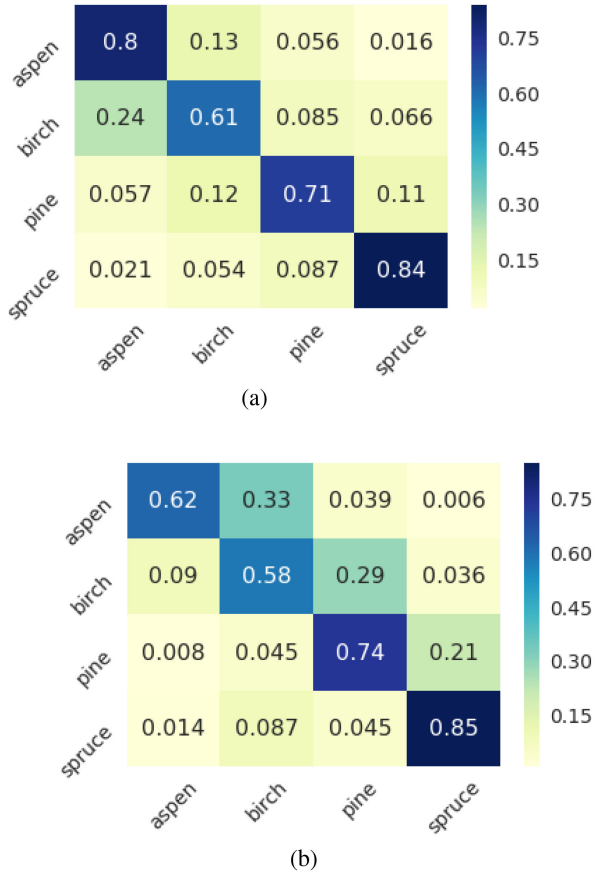

Fig. 5. Confusion matrices for the best aggregated hierarchical models with height data. (a) WorldView data. (b) Sentinel data.

The proposed work approach is only applicable when a hierarchy of classes is established. However, this approach can yield better results, as shown by utilizing the semantic connections between classes. It also helps to reduce computational costs in the case of a high number of classes (a binary logarithm instead of a linear one).

The computational overhead from the use of four models in the hierarchical approach instead of two in the multiclass baseline is not crucial since the problem is neither real time nor addressed to the mobile devices.

B. Supplementary Height Data

Aggregated results for experiments with height data are presented in Tables XI and XII. For the multiclass approach and hierarchical decomposition, height data usage improves model performance. WorldView hierarchical decomposition enhances the quality from 0.797 to 0.836. In multiclass classification, the F1-score without height is 0.703; with height, it is 0.716. The same trend is observed for the Sentinel data. Hierarchical decomposition with height improves the quality from 0.72 to 0.77; for multiclass classification, the scores are 0.58 and 0.668, respectively.

A sample of the test region with the ground truth markup and the predictions of the final hierarchical model with height supplementary data is presented in Figs. 6 and 7, which show a

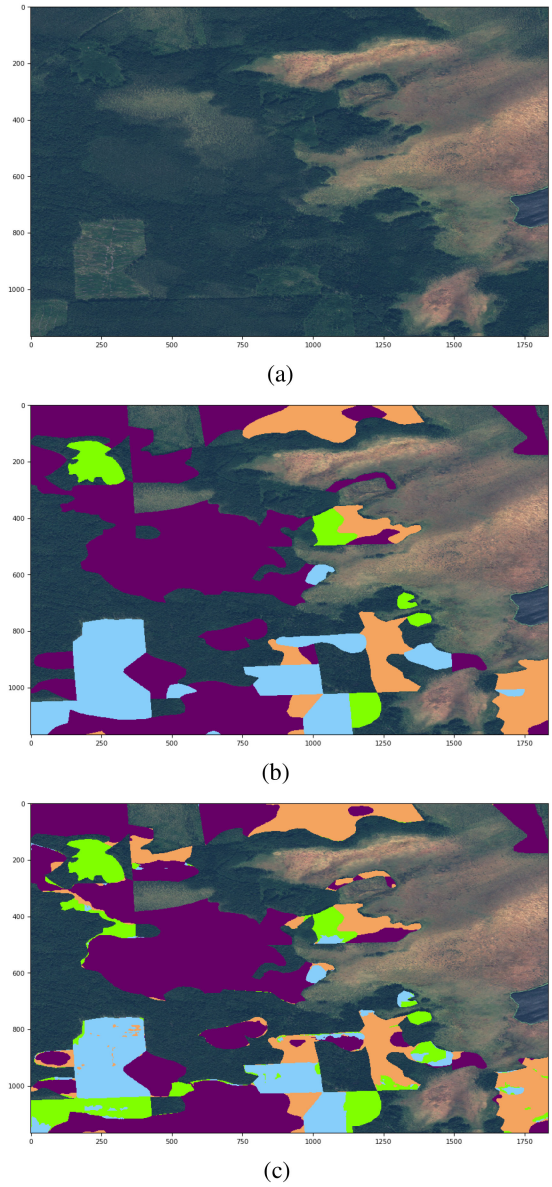

Fig. 6. Sample of the WorldView imagery for the test area. (a) Input image. (b) Ground truth. (c) Prediction.

significant intersection between real classes and the artificially estimated classes. Experiments with both high and medium resolution data confirmed the reliability of the chosen strategy.

C. Architecture Selection

We compared six neural network architectures (Unet with Resnte34 encoder, Unet with EfficientNet encoder, Unet with Inceptionresnetv2 encoder, FPN with Resnte34 encoder, FPN with EfficientNet encoder, and FPN with Inceptionresnetv2 encoder) for each of the classification tasks in the hierarchical decomposition and the multiclass approaches. Results are presented in Tables VII and VIII. Aggregated predictions were computed for the best models in each category. The batch size was limited by the available memory properties and was reduced for larger models for the WorldView data with a crop size of $256 * 256$

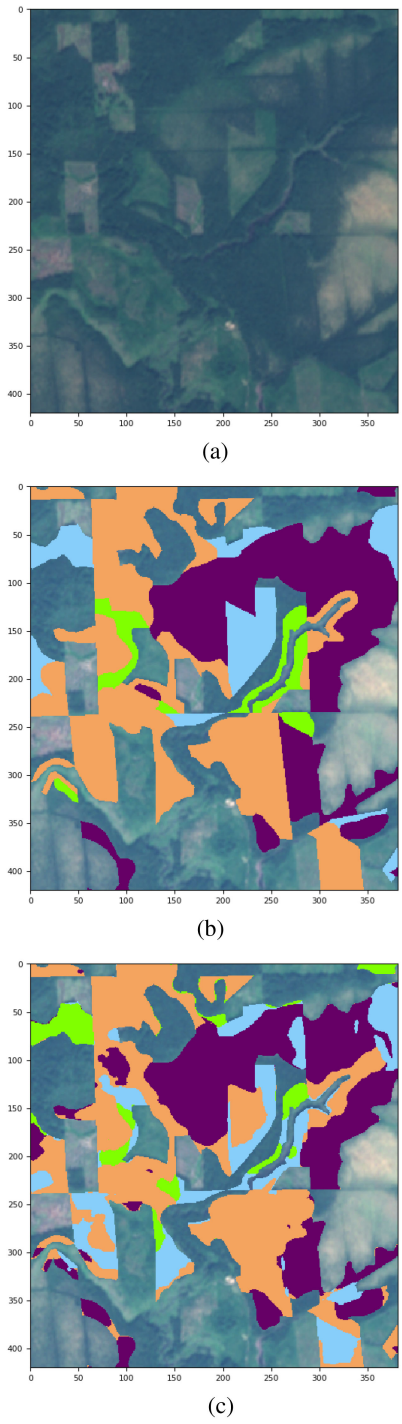

Fig. 7. Sample of the Sentinel imagery for the test area. (a) Input image. (b) Ground truth. (c) Prediction.

pixels. For Sentinel, the crop size was smaller ($64 * 64$ pixels); therefore, the batch size was the same for all experiments. The best models for WorldView are the smaller ones (Unet with Resnet34 encoder and FPN with the same encoder). However, for Sentinel experiments, the best architecture is considerably different. Both the WorldView and Sentinel studies show that the correct architecture for each task can adjust classification quality significantly.

TABLE XIII
OVERSAMPLING EFFECT ON THE WORLDVIEW VALIDATION IMAGES (F1-SCORE)

	1	2
aspen	0.371	0.39
birch	0.746	0.79
spruce	0.732	0.759
pine	0.751	0.868
average	0.65	0.667

Note: (1) All stands with a dominant species content larger than 50% are used. (2) Special thresholds are defined for each class (0.7 for spruce and pine, 0.6 for birch, and 0.5 for aspen).

D. Augmentation and Oversampling

For all models, we implemented geometrical augmentations. This allowed us to achieve a higher diversity in the training dataset. As augmentation in neural network training is well studied, we assessed its contribution to the classification quality for only one architecture and one classification solution: the Unet with Resnet34 encoder in the multiclass problem definition, with WorldView images, and without supplementary height data. The F1-score without augmentation during training is 0.67 (for validation augmented data), whereas the augmentation procedure increases the quality to 0.7 (for the same validation augmented data). This effect is explained by the fact that a neural network treats any geometrical transformation as a new training sample.

We conducted class oversampling according to the thresholds defined in Table IV. Two strategies were compared: first, a dataset of forest stands was formed with a dominant species content of more than 50%, and second, a special threshold was defined for each class (0.7 for spruce and pine, 0.6 for birch, and 0.5 for aspen). The averaged results for a multiclass approach with WorldView images and without height data are presented in Table XIII. It shows that such an oversampling can increase model performance.

VI. CONCLUSION

We studied the applicability of the neural networks for the automatic extraction of forest inventory characteristics from satellite imagery and concentrated on the dominant species classification problem. We present the following contributions.

1) We provide a labeled dataset for dominant species classification, covering a part of Leningrad Oblast, Russia.
2) We developed a hierarchical pipeline for the neural network segmentation, which allows outperforming the basic network approach in the multiclass image segmentation problem. Applicability and relevance of our solution were proved on two data sources: Sentinel and WorldView satellites.
3) We investigated the effect of the supplementary height data, which increases the accuracy significantly.

This approach can be extended to other forest inventory problems and can be improved by a better training markup, both of which we are going to pursue in future work. Moreover, the results in this study are limited to dominant species classification

only. However, in future research, we are going to cover mixed forest cases, which will fall entirely into the hierarchical segmentation scheme. The other goal is to add more forest inventory characteristics, which can also be estimated from the satellite imagery.

REFERENCES

[1] M. Immitzer, F. Vuolo, and C. Atzberger, "First experience with Sentinel-2 data for crop and tree species classifications in central Europe," *Remote Sens.*, vol. 8, no. 3, 2016, Art. no. 166.

[2] M. Wessel, M. Brandmeier, and D. Tiede, "Evaluation of different machine learning algorithms for scalable classification of tree types and tree species based on Sentinel-2 data," *Remote Sens.*, vol. 10, no. 9, 2018, Art. no. 1419.

[3] M. Mngadi, J. Odindi, K. Peerbhay, and O. Mutanga, "Examining the effectiveness of Sentinel-1 and 2 imagery for commercial forest species mapping," Geocarto Int., vol. 36, pp. 1–12, 2019.

[4] M. Immitzer, M. Neuwirth, S. Bck, H. Brenner, F. Vuolo, and C. Atzberger, "Optimal input features for tree species classification in Central Europe based on multi-temporal Sentinel-2 data," *Remote Sens.*, vol. 11, no. 22, 2019, Art. no. 2599.

[5] D. Sheeren *et al.*, "Tree species classification in temperate forests using Formosat-2 satellite image time series," *Remote Sens.*, vol. 8, 2016, Art. no. 734.

[6] V. Kozoderov and E. Dmitriev, "Models of pattern recognition and forest state estimation based on hyperspectral remote sensing data," *Izvestiya, Atmos. Ocean. Phys.*, vol. 54, no. 9, pp. 1291–1302, 2018.

[7] V. Kozoderov, T. Kondranin, and E. Dmitriev, "Hyperspectral remote sensing imagery processing: An overview," *Climate&Nature*, vol. 1, no. 1, pp. 2–18, 2017.

[8] E. Shinzato, Y. Shimabukuro, N. Coops, P. Tompalski, and E. Gasparoto, "Integrating area-based and individual tree detection approaches for estimating tree volume in plantation inventory using aerial image and airborne laser scanning data," *iForest—Biogeosci. Forestry*, vol. 10, no. 1, pp. 296–302, 2017. [Online]. Available: https://iforest.sisef.org/contents/?id=ifor1880-009

[9] M. Dalponte, L. Bruzzone, and D. Gianelle, "Tree species classification in the Southern Alps based on the fusion of very high geometrical resolution multispectral/hyperspectral images and LiDAR data," *Remote Sens. Environ.*, vol. 123, pp. 258–270, 2012.

[10] A. Hernando *et al.*, "Estimation of forest biomass components using airborne LiDAR and multispectral sensors,"*iForest—Biogeosci. Forestry*, vol. 12, no. 2, pp. 207–213, 2019.

[11] S. Tuominen *et al.*, "Hyperspectral UAV-imagery and photogrammetric canopy height model in estimating forest stand variables," *Silva Fennica*, vol. 51, 2017, Art. no. 7721.

[12] L. Naidoo, M. A. Cho, R. Mathieu, and G. Asner, "Classification of Savanna tree species, in the Greater Kruger National Park region, by integrating hyperspectral and LiDAR data in a random forest data mining environment," *ISPRS J. Photogrammetry Remote Sens.*, vol. 69, pp. 167–179, 2012.

[13] Y. He, J. Yang, J. Caspersen, and T. Jones, "An operational workflow of deciduous-dominated forest species classification: Crown delineation, gap elimination, and object-based classification," *Remote Sens.*, vol. 11, no. 18, 2019, Art. no. 2078.

[14] M. Immitzer, C. Atzberger, and T. Koukal, "Tree species classification with random forest using very high spatial resolution 8-band WorldView-2 satellite data," *Remote Sens.*, vol. 4, no. 9, pp. 2661–2693, 2012.

[15] Y. Ke and L. J. Quackenbush, "Forest species classification and tree crown delineation using QuickBird imagery," in *Proc. Amer. Soc. Photogrammetry Remote Sens. Annu. Conf.*, 2007, pp. 7–11.

[16] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, 1995.

[17] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.

[18] Y. Guo, X. Jia, and D. Paull, "Effective sequential classifier training for SVM-based multitemporal remote sensing image classification," *IEEE Trans. Image Process.*, vol. 27, no. 6, pp. 3036–3048, Jun. 2018.

[19] M. Belgiu and L. Drăguţ, "Random forest in remote sensing: A review of applications and future directions," *ISPRS J. Photogrammetry Remote Sens.*, vol. 114, pp. 24–31, 2016.

[20] Y. Li, H. Zhang, X. Xue, Y. Jiang, and Q. Shen, "Deep learning for remote sensing image classification: A survey," *Wiley Interdisciplinary Rev., Data Mining Knowl. Discovery*, vol. 8, no. 6, 2018, Art. no. e 1264.

[21] L. Dong *et al.*, "Very high resolution remote sensing imagery classification using a fusion of random forest and deep learning technique-subtropical area for example," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.,* vol. 13, pp. 113–128, 2019.

[22] H. Hamraz, "Automated tree-level forest quantification using airborne LiDAR," Ph.D. dissertation, Dept. Comput. Sci., Univ. Kentucky, Lexington, KY, USA, 2018.

[23] E. Ayrey and D. J. Hayes, "The use of three-dimensional convolutional neural networks to interpret LiDAR for forest inventory," *Remote Sens.*, vol. 10, no. 4, 2018, Art. no. 649.

[24] S. Shen, S. X. Han, D. R. Aberle, A. A. Bui, and W. Hsu, "An interpretable deep hierarchical semantic convolutional neural network for lung nodule malignancy classification," *Expert Syst. Appl.*, vol. 128, pp. 84–95, 2019.

[25] C.-S. Huang, C.-L. Lin, L.-W. Ko, S.-Y. Liu, T.-P. Sua, and C.-T. Lin, "A hierarchical classification system for sleep stage scoring via forehead EEG signals," in *Proc. IEEE Symp. Comput. Intell., Cogn. Algorithms, Mind, Brain*, 2013, pp. 1–5.

[26] I. Dimitrovski, D. Kocev, S. Loskovska, and S. Deroski, "Hierarchical classification of diatom images using ensembles of predictive clustering trees," *Ecological Informat.*, vol. 7, no. 1, pp. 19–29, 2012.

[27] G. Gerylo, R. Hall, S. Franklin, A. Roberts, and E. Milton, "Hierarchical image classification and extraction of forest species composition and crown closure from airborne multispectral images," *Can. J. Remote Sens.*, vol. 24, no. 3, pp. 219–232, 1998.

[28] O. S. Ahmed *et al.*, "Hierarchical land cover and vegetation classification using multispectral data acquired from an unmanned aerial vehicle," *Int. J. Remote Sens.*, vol. 38, no. 8–10, pp. 2037–2052, 2017.

[29] T. Blaschke, "Object based image analysis for remote sensing," *ISPRS J. Photogrammetry Remote Sens.*, vol. 65, no. 1, pp. 2–16, 2010.

[30] Weather Spark, 2020. [Online]. Available: https://weatherspark.com/

[31] GBDX, 2020. Accessed: Aug. 17, 2020. [Online]. Available: https://gbdxdocs.digitalglobe.com/

[32] Sentinel Hub, 2020. Accessed: Aug. 17, 2020. [Online]. Available: https://www.sentinel-hub.com/explore/eobrowser/

[33] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assisted Intervention*, 2015, pp. 234–241.

[34] T.-Y. Lin, P. Dollr, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 2117–2125.

[35] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.

[36] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi, "Inception-v4, inception-ResNet and the impact of residual connections on learning," in *Proc. 31st AAAI Conf. Artif. Intell.*, 2017, pp. 4278–4284.

[37] M. Tan and Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in *Proc. Int. Conf. Mach. Learn.*, 2019, pp. 6105–6114.

[38] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2009, pp. 248–255.

[39] P. Yakubovskiy, "Segmentation models," 2019. [Online]. Available: https://github.com/qubvel/segmentation_models

[40] T. Jayalakshmi and A. Santhakumaran, "Statistical normalization and back propagation for classification," *Int. J. Comput. Theory Eng.*, vol. 3, no. 1, pp. 1793–8201, 2011.

[41] K. K. Pal and K. S. Sudeep, "Preprocessing for image classification by convolutional neural networks," in *Proc. IEEE Int. Conf. Recent Trends Electron., Inf. Commun. Technol.*, 2016, pp. 1778–1781.

[42] G. M. Foody, "Status of land cover classification accuracy assessment," *Remote Sens. Environ.*, vol. 80, no. 1, pp. 185–201, 2002.

Svetlana Illarionova received the bachelor's and master's degrees in computer science from Lomonosov Moscow State University, Moscow, Russia, in 2017 and 2019, respectively. She is currently working toward the Ph.D. degree in computer science with the Skolkovo Institute of Science and Technology, Moscow, Russia.

Her research interests include computer vision, deep neural networks, and remote sensing.

Alexey Trekin graduated in applied mathematics and physics from the Moscow Institute of Physics and Technology, Moscow, Russia, in 2012, and received the Ph.D. degree in computer science, in 2017 from the Higher School of Economics, Moscow, Russia.

He is currently a Research Scientist with the Skolkovo Institute of Science and Technology, Moscow, Russia. He is the Head of Research with the Aeronet Laboratory. From 2011 to 2017, he was with Moscow Aerocosmos Research Institute on problems of remote sensing data processing, including work on wildfire monitoring and impact assessment.

Vladimir Ignatiev graduated in applied mathematics and physics from the Moscow Institute of Physics and Technology, Moscow, Russia, in 2012, and received the Ph.D. degree from Higher School of Economics, Moscow, Russia, in 2017.

He is currently a Research Scientist at the Skolkovo Institute of Science and Technology, Moscow, Russia. He leads the Aeronet Laboratory that is focused on various applications of the deep learning methods to remote sensing data. Before joining Skoltech, he was with Dorodnitsyn CCAS and Aerocosmos Research Institute. He has experience in different remote sensing data processing and forecasting models development.

Ivan Oseledets graduated in applied mathematics and physics from the Moscow Institute of Physics and Technology, Dolgoprudny, Russia, in 2006, received the Candidate of Sciences and Doctor of Sciences degrees in numerical mathematics from the Marchuk Institute of Numerical Mathematics, Russian Academy of Sciences, Moscow, Russia, in 2007 and 2012, respectively.

He joined Skoltech CDISE in 2013. His research covers a broad range of topics. He proposed a new decomposition of high-dimensional arrays (tensors)—tensor-train decomposition, and developed many efficient algorithms for solving high-dimensional problems. His current research focuses on development of new algorithms in machine learning and artificial intelligence, such as construction of adversarial examples, theory of generative adversarial networks, and compression of neural networks. It resulted in publications in top computer science conferences, such as ICML, NIPS, ICLR, CVPR, RecSys, ACL, and ICDM.

Prof. Oseledets is an Associate Editor for the *SIAM Journal on Mathematics in Data Science*, *SIAM Journal on Scientific Computing*, and *Advances in Computational Mathematics* (Springer).