IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received February 22, 2022, accepted March 15, 2022, date of publication March 22, 2022, date of current version April 1, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3161568

Estimation of the Canopy Height Model From Multispectral Satellite Imagery With Convolutional Neural Networks

SVETLANA ILLARIONOVA[1], DMITRII SHADRIN[1,2], VLADIMIR IGNATIEV[1], SERGEY SHAYAKHMETOV[3], ALEXEY TREKIN[1], AND IVAN OSELEDETS[1]

[1]Skolkovo Institute of Science and Technology, 121205 Moscow, Russia
[2]Institute of Information Technology and Data Science, Irkutsk National Research Technical University, 664074 Irkutsk, Russia
[3]ESG Department, PJSC Sberbank of Russia, 117997 Moscow, Russia

Corresponding author: Svetlana Illarionova (s.illarionova@skoltech.ru)

The work was supported by the Analytical center under the RF Government (subsidy agreement 000000D730321P5Q0002, Grant No. 70-2021-00145 02.11.2021).

ABSTRACT The canopy height model (CHM) is a representation of the height of the top of vegetation from the surrounding ground level. It is crucial for the extraction of various forest characteristics, for instance, timber stock estimations and forest growth measurements. There are different ways of obtaining the vegetation height, such as through ground-based observations or the interpretation of remote sensing images. The severe downside of field measurement is its cost and acquisition difficulty. Therefore, utilizing remote sensing data is, in many cases, preferable. The enormous advances in computer vision during the previous decades have provided various methods of satellite imagery analysis. In this work, we developed the canopy height evaluation workflow using only RGB and NIR (near-infrared) bands of a very high spatial resolution (investigated on WorldView-2 satellite bands). Leveraging typical data from airplane-based LiDAR (Light Detection and Ranging), we trained a deep neural network to predict the vegetation height. The provided approach is less expensive than the commonly used drone measurements, and the predictions have a higher spatial resolution (less than 5 m) than the vast majority of studies using satellite data (usually more than 30 m). The experiments, which were conducted in Russian boreal forests, demonstrated a strong correlation between the prediction and LiDAR-derived measurements. Moreover, we tested the generated CHM as a supplementary feature in the species classification task. Among different input data combinations and training approaches, we achieved the mean absolute error equal to 2.4 m using U-Net with Inception-ResNet-v2 encoder, high-resolution RGB image, near-infrared band, and ArcticDEM. The obtained results show promising opportunities for advanced forestry analysis and management. We also developed the easy-to-use open-access solution for solving these tasks based on the approaches discussed in the study cloud-free composite orthophotomap provided by Mapbox via tile-based map service.

INDEX TERMS Artificial intelligence, artificial neural networks, computer vision, data analysis, digital elevation models, remote sensing, forestry, transfer learning.

I. INTRODUCTION

Canopy height model (CHM) estimation has a long history, but advances in computer vision and satellite sensing technologies have opened new opportunities in this area. The height can be effectively utilized in different applications and broadens the surface's two-dimensional representation in the visible spectrum. There are both natural [1]–[5] and anthropogenic [6], [7] objects of landcover to be explored. The present study is focused on natural types of landcover, especially wild forest areas. Landcover height characteristics can be used in various applications, such as biomass evaluation [8]–[10], improving the accuracy of tree species classification [11], [12], and correlated vegetation properties extraction [13].

There are three frequently reported sources of canopy height information: 1) field-based measurements; 2) Unmanned Aerial Vehicle (UAV)-based approaches;

The associate editor coordinating the review of this manuscript and approving it for publication was Tallha Akram.

and 3) satellite remote sensing data. All aforementioned approaches have advantages and limitations connected with acquisition time and cost (Fig 1). The first data source is forest inventory documents, usually treated as field-based observations. They are available for some regions and useful in addressing forest owners', governmental, and independent organizations' needs [14]. However, these data do not cover all regions of practical interest [15]. Furthermore, such data actualization is time-consuming and cost intensive in difficult-to-access areas. An alternative approach is to use remote sensing data.

The remote sensing approach draws on both active and passive sensing technologies. During active sensing such as Light Detection and Ranging (LiDAR) measurements, the sensor measures time between the emitted light time and its return time to estimate the distance of an object (a surface). This technology allows digital elevation models to be produced. Passive remote sensing measures radiation that is emitted or reflected by the object in different spectral wavelengths. Spectral bands obtained this way can be used for future analysis and to calculate the height value in landcover extraction.

A common approach builds on UAV assessment. A UAV with LiDAR sensors is a powerful tool for forest height estimation. It obtains canopy height data with minor errors, meeting the precision requirements for almost all forestry tasks. However, such equipment is more expensive than a spectral aerial camera system, thus there remains the challenge of obtaining the same information using low-cost methods [16]. Many works use LiDAR data as a reference and aim to find a cheaper height data source. A detailed review of the alternative approaches to LiDAR sensing is presented in [17], [18]. Thus, most of the current studies in the sphere of canopy height estimation use UAVs with optical aerial systems [19]–[24]. Despite its advantages over field-based observations, when large regions have to be processed, the labor involved in working with vast and remote areas is problematic. Satellite data address this issue, providing a cheaper option for forest monitoring [17]. Point cloud data that is useful for estimation of the canopy height can also be derived from satellite imagery using photogrammetry approach. The comparison of such photogrammetry approach and high-density LiDAR measurements is presented in [25], where authors showed photogrammetry method is slightly less accurate (difference in R^2 is about 0.07) compare to the LiDAR method for height measurements of the forest region in New Zealand. The important benefit of the photogrammetry method is that it could provide information for the larger scale compare to the LiDAR method, however it requires special high resolution imagery which is not always available for the particular region. The other limitation of the photogrammetric method is that it is able to characterise only the upper canopy and is not able to perform vertical characterisation of the forest such as can be done by laser scanning. The comparison of the photogremmetry obtained by unmanned aerial systems and areal laser scanning for the forest inventory in Oregon was presented in [26], where authors stated that photogrammetry is slightly less accurate compare to laser scanning (difference in R^2 for height estimation is about 0.15). However photogrammetry is easier to integrate to existing forest monitoring methodologies.

Our work is focused on using satellite images for CHM estimation as it is more preferable data source than LiDAR derived measurements in terms of cost and spatial coverage. Neural networks allows us to conduct image processing automatically. We set up the hypothesis that neural networks can extract significant spatial features from very high-resolution RGB images of 1 m to improve performance of CHM estimation. It was expected that developing a satellite-based solution compatible with a high-resolution UAV approach would further enable the prediction of advanced forest characteristics. Thus, this study's objectives and contributions were:

1) to develop a method for vegetation height estimation utilizing deep neural networks and different configurations of input data varying spectral compound (reducing to Blue, Green, and Red), spatial resolution and by adding topography features;
2) to assess the generated height map, conducting a further investigation into the classification of dominant forest species (conifer and deciduous). For this, multispectral imagery was incorporated with height data;
3) to create the software toolchain to train a neural network to predict CHM using single satellite non-stereo imagery.
4) to develop the easy-to-use open-access solution for the community which is now available by the following resource [27]. The underlying code for CNN model training is shared:
https://github.com/LanaLana/forest_height.

II. RELATED WORK

For canopy height estimation studies, spectral satellite imagery can be distinguished by the following characteristics: spatial resolution, spectral range, and availability. The majority of works use a spatial resolution much higher than 20 m to tackle the canopy height evaluation problem. This approach is justified for particular tasks when large-scale maps are produced. In [28], they conducted a 30 m spatial resolution canopy height evaluation with Landsat imagery and showed the dynamics over 29 years in the Darwin region. In [29], they employed Landsat 7 and 8 time-series data (30 m spatial resolution) to estimate tree heights in Africa. GLAS (Geoscience Laser Altimeter System) height measurements from the ICESat satellite were used as reference data ($60 - 70$ m spatial resolution). The same height data source was mentioned in [30]. In [31], they used Sentinel-2 images that were resampled to a 20 m pixel size to predict Mangrove forest canopy height. Other studies involving Sentinel-2 data are reported in [32]–[34]. In [35], they assessed SAR images from ALOS PALSAR, and upsampled them from 30 to 5 m as a LiDAR elevation model. The cases of very high

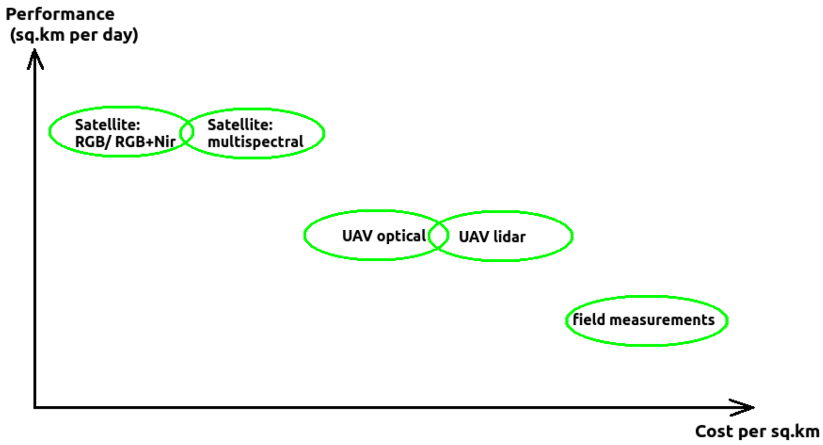


FIGURE 1. Cost comparison of different forest height measurement approaches (diagram is not to scale).

spatial resolution (3.7 m) images from the Planet Dove implementation are presented in [36]. However, the target height map spatial resolution for that study was 1 hectare. Very high spatial resolution (2 m) WorldView-2 satellite imagery was used in [37], but the working spatial resolution was adjusted to 5 m.

Another important data characteristic is the spectral range and the number of channels. A wider wavelength range is available for satellites with low spatial resolutions (Landsat, Sentinel) than for some very high spatial resolution satellites. For instance, Planet (3–5 m spatial resolution) and GeoEye (2 m spatial resolution) satellites have Blue, Green, Red, and NIR bands; RapidEye (6 m spatial resolution) has Red Edge. The GeoEye panchromatic channel has a 0.4 m spatial resolution and allows RGB to be enhanced. WorldView-2 provides eight spectral bands with a spatial resolution of 2 m. An additional source of very high remote sensing data is Basemaps, with RGB bands such as those provided by Maxar one [38]. Nevertheless, the majority of works focus on using only the wide multispectral range (more than eight bands), sacrificing the spatial resolution. From the aforementioned satellite-based studies, the minimal number of spectral bands (Blue, Green, Red, NIR) was only considered in [36]. However, the goal of the work was the creation of a large-scale country wide map, so the spatial resolution of the analysis was 1 hectare. Therefore, the issue of minimizing the number of required satellite bands for forest height estimation has not yet been well studied.

Satellite data are frequently accompanied with data of other sensing techniques. In [39], they combined four Kompsat-3 multispectral bands and PALSAR-1 radar images resampled into 2.8 m to train a neural network. Few studies have implemented this into self-contained spectral satellite data [33], [40]–[42]. However, the spatial resolution of the Sentinel and Landsat images (lower than 10 m) considered in these studies is not high enough to extract small details on the surface. Thus, the satellite spatial resolution of 1-m per pixel is still beyond the scope of the majority of studies.

Data availability is also a significant aspect of implementation in practice. Sentinel and Landsat images are available free of cost, while WorldView, Planet, and RapidEye are commercial and contain a greater amount of the spatial information required in applied tasks.

After data acquisition, the obvious question of data processing arises. Computer vision algorithms enable high-quality automatic satellite imagery analysis. Such methods are usually based on key feature extraction from input spectral bands to describe some object, which can be a pixel or set of pixels. Then, the algorithm aims to ascribe a label (for classification tasks) or a value (for regression tasks) to the object. The processing methods for expansive forestry areas using satellite images are classical machine learning models, such as Random Forest [43] or Support Vector Machine [44]. Their main advantages are simplicity and straightforward interpretation in the case of linear models. Generally, spatial characteristics are not taken into consideration, and an algorithm relies on spectral values or precalculated vegetation indices. In [28], a combination of 14 vegetation indices and spectral bands were used in the Random Forest model to predict the canopy height using Landsat images. Moreover, the strong correlation between the normalized difference vegetation index (NDVI) and canopy height has been well emphasized in aerial photography [16], [35]. Despite the importance of spectral data, other vital features can also be processed. For instance, there is a strong correlation between forest height and canopy width, as discussed in [32], in which the canopy volume was

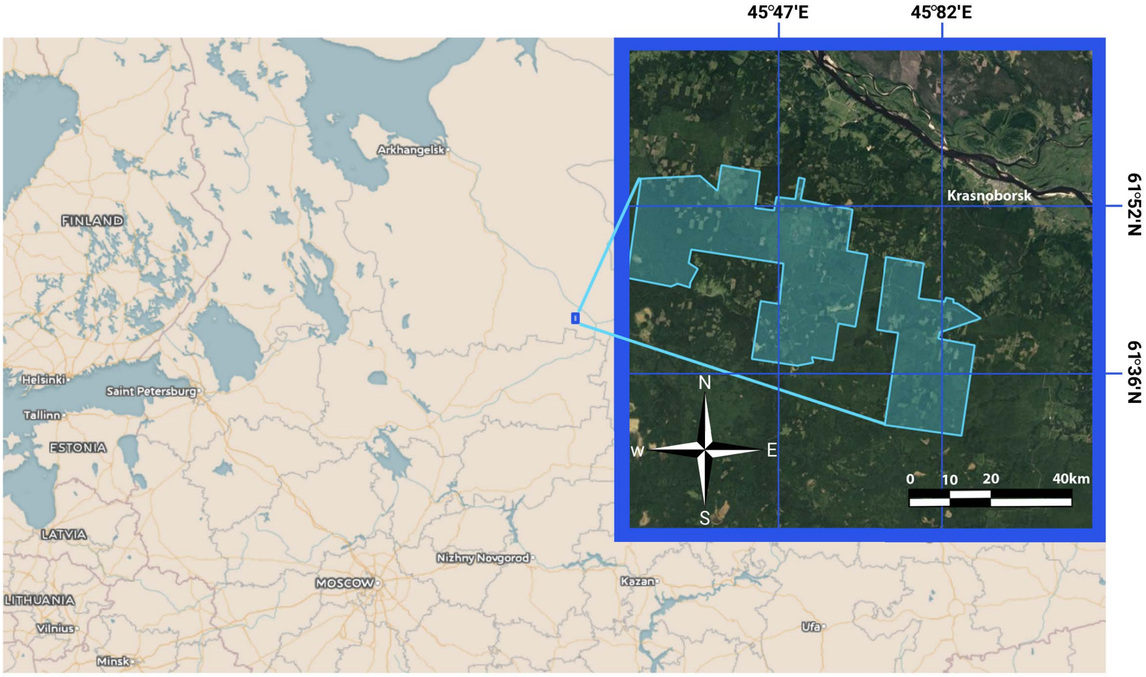


FIGURE 2. Region of interest.

estimated using only the crown projected area and the crown diameter combined in a particular regression equation. The deep neural network-based approach is more capable than classical machine learning methods for the following reasons: the texture and spatial features extracted by the neural networks include sufficient information about landcover; it not only handles spectral values, but also the aforementioned spatial characteristics of an object available, for instance, in UAV-based tasks [45].

Tree height is correlated with tree diameter for each forest species [46]. In [47], tree height was estimated from the exponential equation, including diameter at breast height value. The crown form depends on the tree species; accompanied by the crown diameter, it can provide important features for a neural network. Tree height can also be derived from spectral information only, as it depicts meaningful vegetation characteristics such as chlorophyll content [48].

III. MATERIALS AND METHODS

A. STUDY AREA

The study area is located in the Arkhangelsk region of northern European Russia with coordinates between $45^\circ 16'$ and $45^\circ 89'$ longitude and between $61^\circ 31'$ and $61^\circ 57'$ latitude (Fig 2). The investigated territory belongs to the middle boreal zone. The region's climate is humid, with the warmest month being July when the temperature rises to $17^\circ C$. The topography is flat, with a height difference in a range between 170 and 215 m above sea level [49]. The main species present in the region are pine, spruce, aspen, and birch.

B. REFERENCE DATA

We used forest inventory and LiDAR-derived data covering the area of about 50 thousand hectares. LiDAR measurements were continued in the end of August of 2017 and 2018 by Leica ALS 80 HP scanner. Then the Canopy Height Model (CHM) with a 1 m spatial resolution was generated from LiDAR-derived point clouds.

The inventory data were collected in accordance with the official Russian inventory regulation in 2018 and 2019 [50]. It included such characteristics as canopy height, species percentage distribution, and age. This data was organized as a set of individual stand coordinates with appropriate characteristics based on the assumption that the crop was homogeneous. A species class markup was used in additional experiments presented as a raster map of dominant conifer and deciduous classes. The statistics of this data are shown in Table 3.

However, the shift in geo-referencing between the satellite data and LiDAR-derived measurements makes the target at 1 m spatial resolution less useful. As the typical shift lies between 2 and 3 m, the high spatial resolution CHM will show erroneous value for the particular point in the satellite image. This forced us to downsample the height map to 5 m to make the target value for each point represent the mean value of the area including the true location.

The distribution of the height over the study region is shown in the Figure 4. Although, height is usually represented as a continues value, height categories are essential for practical use in power lines services. Height classes are often required instead of continues values for decision making

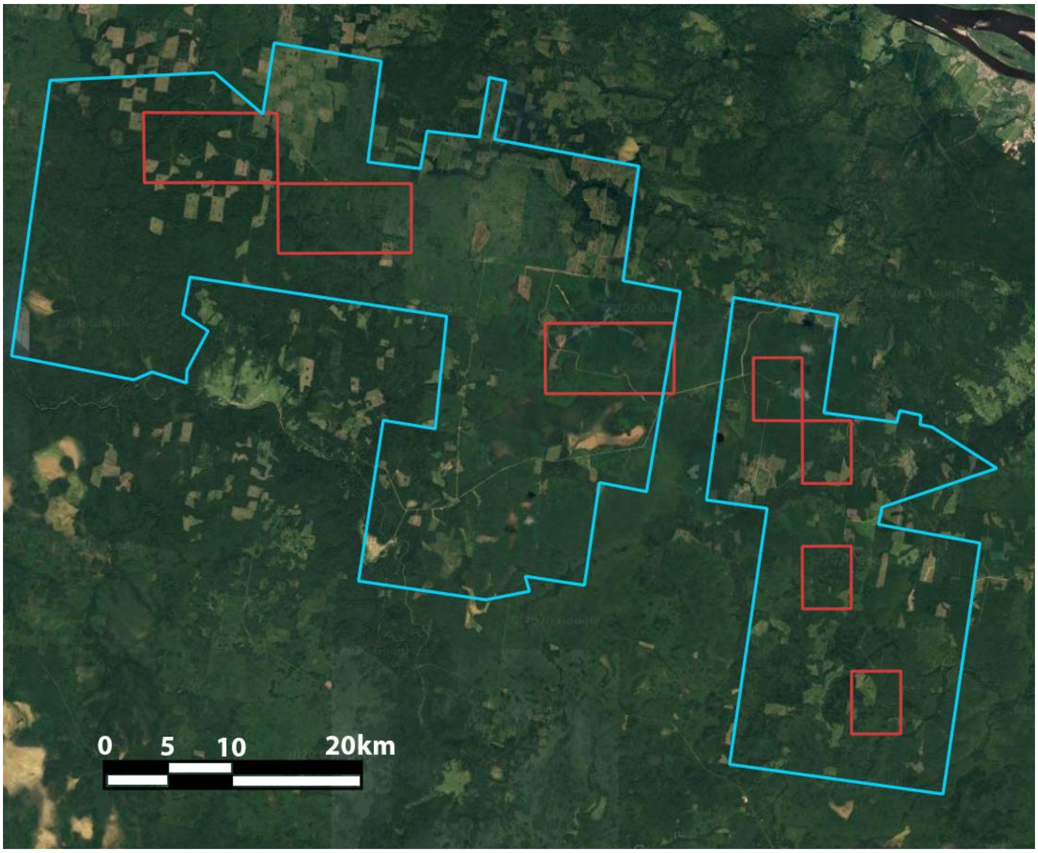


FIGURE 3. The blue lines define the study area with LiDAR measurements. The red squares are the test regions.

TABLE 1. WorldView images.

	Image ID	Date	Off-nadir angle
0	1030010056130F00	05.30.16	14
1	103001005683F200	05.30.16	14
2	1030010031934700	06.08.14	7
3	1030010032660800	06.08.14	7

TABLE 2. Sentinel images.

	Image ID	Date
0	L2A_T38VNP_A005695_20160725T082012	07.25.16
1	L2A_T38VNP_A007297_20180730T081559	07.30.18
2	L2A_T38VNP_A010986_20170730T082009	07.30.17
3	L2A_T38VNP_A013017_20190903T081606	09.03.17
4	L2A_T38VNP_A015748_20180628T082602	06.28.18
5	L2A_T38VNP_A016606_20180827T083208	08.27.18

TABLE 3. Dataset statistics for conifer and deciduous classification.

	Training (individual stands)	Testing (individual stands)	Full dataset
Conifer	1219	534	25913 hectares
Deciduous	756	341	24397 hectares

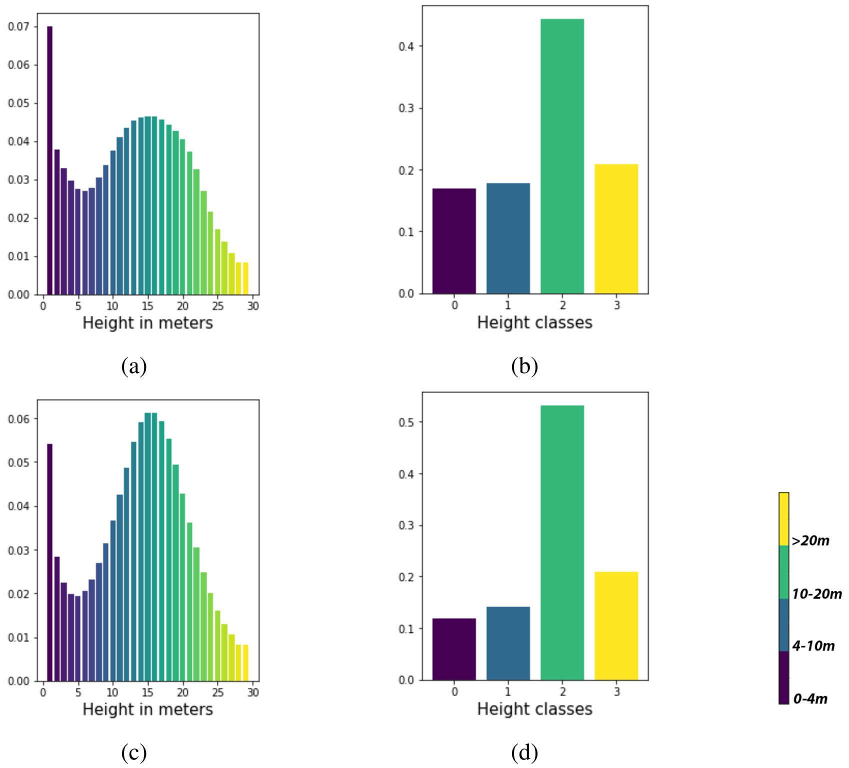
within protected areas [51]. The reason is that different categories (dangerous vegetation overgrowing) have different importance and estimation in particular categories have to be more precise to reduce accidents on power lines corridors.

C. THE TEST REGION SELECTION

The training and test area was from the same satellite images, but without overlapping. The test region was manually chosen to include a diversity of height classes. The total test area was equal to 13% of the initial dataset. The spatial location is presented in the Fig 3. The height distribution through the test areas is presented in the Fig 4.

D. SATELLITE DATA

We used Sentinel-2 and WorldView-2 satellite imagery to check the high and very high spatial resolution data sources. The boreal location of the study area resulted in a lack of cloudless images. All images were from the boreal growing season (from May to August). Image IDs and dates are presented in tables 1, 2. WorldView imagery was downloaded from GBDX [52]. For the height estimation task, we used Red, Green, Blue, and Near-Infrared bands, while for the species classification problem, all eight bands were considered. The resolution of the WorldView images was 1, 2, or 5 m depending on the experiment statement. For CNN-based tasks, image values in the range from 0 to 1 are usually used [53], [54]. Therefore, pixel values were brought into a range between 0 and 1 using Equation 3. For the spatial resolution adjustment, the pansharpening procedure was implemented using a panchromatic band which was



(a) (b)

(c) (d)

FIGURE 4. Reference LiDAR-derived height (Canopy height model (CHM) values) distribution for the study area. (a-b) Training dataset. (c-d) Test dataset. These height categories are the important ones for power lines services in Russia.

obtained in the imagery bundle with multispectral data from the data vendor. We did not consider any predefined cloud mask for WorldView. However, during training, pixels with particular properties were eliminated from consideration (see subsection III-G). This allowed us to clean the dataset from erroneous labels.

$$m = max(0, mean(I) - 3 * std(I)) \tag{1}$$

$$M = min(max(I), mean(I) + 3 * std(I)) \tag{2}$$

$$I' = (I - m)/(M - m) \tag{3}$$

where $mean, std$ are the mean and standard deviation of the image. In equations 1, 2, we calculate m and M (minimum and maximum of the preserved dynamic range). 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 [55].

For the additional analysis, freely available Sentinel data were downloaded in L1C format from EarthExplorer USGS [56] and preprocessed using Sen2Cor [57] to an L2A format. Pixel values were brought into a range between 0 and 1 using Equation 3. We used the $B02, B03, B04, B05, B06, B07, B08, B11, B12$, and $B8A$ bands, which were adjusted to a 10 m resolution. 60m bands were discarded as they are more affected by atmosphere than the land surface. 20 to 10 m bands were upsampled with the nearest neighbor method to avoid initial data corruption (they can be unambiguously downsampled back to exactly initial 20m data).

Both for Sentinel and WorldView, each image covered the entire study area, and images were considered separately without any spatial averaging (the same as in [58]).

As supplementary features, we used a freely available high-resolution digital elevation model (DEM), Arctic-DEM [59], covering boreal regions (Fig 5). It provides a resolution of 2 m. For some experiments, the resolution was upsampled to 1 m by interpolation (see the section III-E).

Both the satellite and LiDAR data were co-registered through geo-referencing, the same as in [37].

We used cloud-free composite orthophotomap provided by Mapbox [60] via tile-based map service as an example of free-available high-resolution RGB data-source. This image covered the same test region and was used just for the developed model assessment. We chose this data-source, because model implementation without expensive input data demands is crucial for open-access platform that can handle a more available images. The spatial resolution was 1 m per pixel, and the preprocessing was the same as for WorldView data.

E. FEATURE SELECTION FOR DEEP NEURAL NETWORK

Convolutional neural networks take a tensor as an input. The feature selection to create this tensor is fundamental. To find

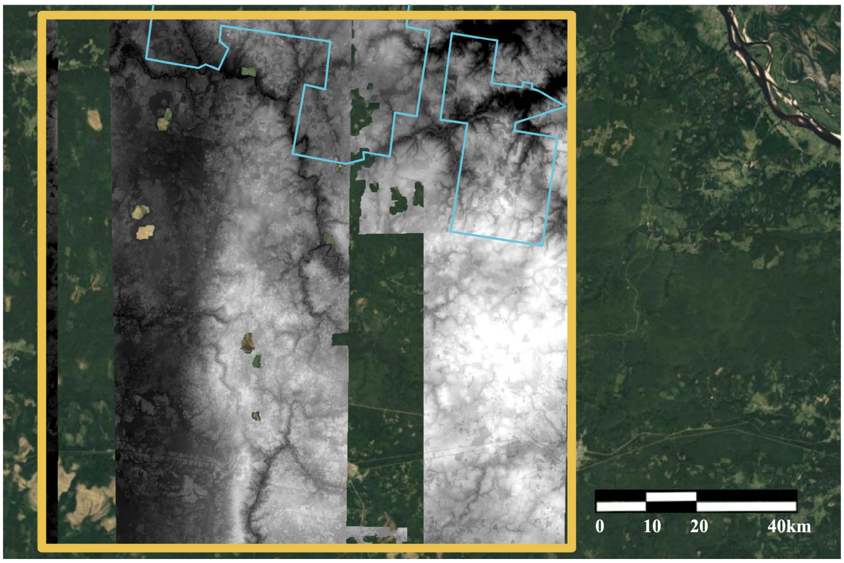


FIGURE 5. One of the ArcticDEM tiles (yellow square) with an overlay of the studied area (blue lines). Even in boreal regions, ArcticDEM layer can have some missing data.

the best input data representation for the CHM estimation problem, we established a range of experiments. Firstly, we conducted a study with the WorldView bands.

The workflow of our research is shown in the Fig 6. For each experiment, the RGB bands were used constantly. The variable part concerned the resolution changing and the supplementary features (NIR and ArcticDEM), which were combined with the RGB channel in a single input tensor for the neural network model.

We studied the original (2 m), pansharpened (1 m), and downsampled (5 m) images. For the experiments with the 1 m resolution, bands were upsampled to the target resolution by bilinear interpolation. We used bilinear interpolation for image resampling to avoid aliasing emerging in nearest neighbor and halo inherent to higher-order interpolation methods, which are more problematic for neural networks than bilinear interpolation. A reference CHM was used during the training procedure to estimate the model's error. To minimize data mismatches, reference and predicted height maps were intersected with the forest cover mask before the loss function calculation stage. Therefore, we conducted the following experiments for the WorldView images:

1) RGB original resolution 2 m;
2) RGB pansharpened to 1 m;
3) RGB pansharpened to 1 m + ArcticDEM upsampled to 2 m;
4) RGB + NIR original resolution 2 m;
5) RGB + NIR original resolution 2 m + ArcticDEM upsampled to 2 m;
6) RGB pansharpened to 1 m + NIR upsampled to 1 m;
7) RGB pansharpened to 1 m + NIR upsampled to 1 m + ArcticDEM upsampled to 1 m;
8) RGB downsampled to 5 m resolution.

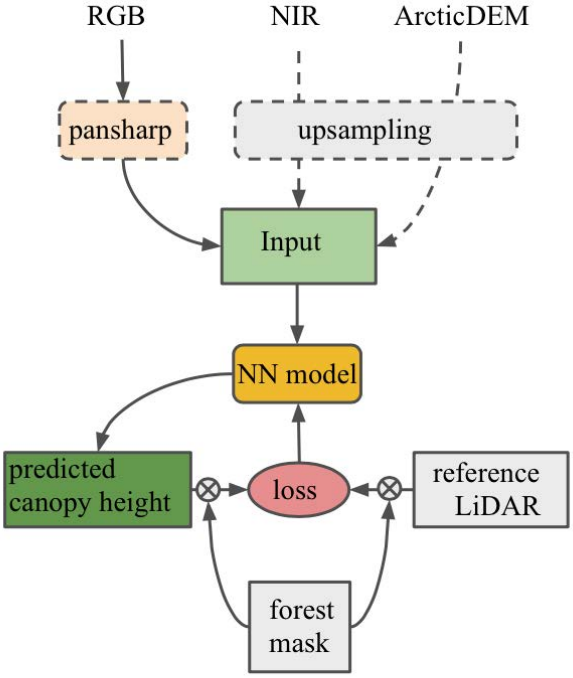


FIGURE 6. Experiment workflow for canopy height estimation with RGB WorldView bands. The dotted lines show optional steps for input tensor creation.

For experiments 1, 2, there was three-band raster; for experiments 3, 4, 6, we used four-band raster; and for experiments 5, 7, five-band raster was considered.

To assess the importance and restriction of the spatial resolution, we also checked the model's performance for the WorldView RGB bands downsampled to 5 m.

We conducted the following study to compare model's performance for high-resolution RGB images and less detailed but richer in terms of the spectral information Sentinel data with 10 bands, upsampled to 10 m. There were two experiments:

1) Multispectral bands;
2) Multispectral bands + ArcticDEM downsampled to 10 m.

F. STRATEGIES FOR HEIGHT PREDICTION AND EVALUATION METRICS

Regression may naturally lead to richer (continuous) estimations for practical implementations than rigid class-based output maps. Therefore, we considered both regression and classification tasks for a comparative analysis. The regression problem statement means that we ascribe each pixel with a particular value corresponding to the height parameter. Then, the loss can be estimated as an error between real height value (CHM value) and the predicted value. The considered metrics are root mean square error (RMSE), mean absolute error (MAE), and mean bias error (MBE):

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}} \tag{4}$$

$$MAE = \frac{\sum_{i=1}^{n}|y_i - \hat{y}_i|}{n} \tag{5}$$

$$MBE = \frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)}{n} \tag{6}$$

where $\bar{y}$ is the mean target value among all pixels (mean CHM value), $\hat{y}_i$ is the predicted value of the i^{th} pixel, y_i is the target value of the i^{th} pixel (CHM value), and n is the pixel number. Test regions results were computed for all images in WorldView or Sentinel datasets.

Using the same reference data we can also solve classification task. When we formalized the problem as a classification task, we divided the continuous values of height into various classes. The choice of such a division often depends on an applied task's demands. For our study, we chose intervals $0 - 4$, $4 - 10$, $10 - 20$, and > 20 m. We rely on the amount of classes and intervals of height that described [61]. We slightly shifted the boundaries of the height intervals, described in [61] according to the suggestion inventory data provider from Arkhangelsk region. After splitting the continuous dataset to the aforementioned classes we can compute the portion of the wrong estimated pixel classes and use F1-score [62] for evaluation of the trained classification models.

$$precision = \frac{TP}{TP + FP} \tag{7}$$

$$recall = \frac{TP}{TP + FN} \tag{8}$$

$$F1 = \frac{2 * precision * recall}{precision + recall} \tag{9}$$

where TP denotes true positive, FP denotes false positive, and FN denotes false negative. The above formulas were applied in a per-class basis. To compute results, test regions from all images were used.

This refers to the area assessment, while in terms of regression, we strove to optimize each pixel value. Therefore, these two approaches can lead to a different local optimum. For example, if we split heights between 0 and 30 m into the following buckets: $0 - 4$, $4 - 10$, $10 - 20$, and $20 - 30$, then it is not important that we do not ascribe the exact values but some value from the correct bucket to some pixels. Then, it is clear that regression predictions can also be represented in terms of classification.

For the classification task, the multiclass weighted cross-entropy loss function was used to make the predictions more balanced even for classes with fewer representatives. The same approach was implemented for the regression task. We compared the simple RMSE loss (Equation 11) and the weighted RMSE loss (Equation 11). For heights with fewer representatives, the penalty for the wrong prediction was increased by predefined weights. The weights were inversely proportional to the height distribution. There was also a threshold for the height when the weight was equal to 1 (no extra penalty). The range of weights and the threshold were chosen empirically, as shown in the Figure 7.

$$\text{RMSE loss} = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{N}} \tag{10}$$

$$\text{Weighted RMSE loss} = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 * weights(y_i)}{N}} \tag{11}$$

where $\hat{y}_i$ is the predicted value of the i^{th} pixel, y_i is the target value of the i^{th} pixel, N is the number of relevant (non-masked) pixels, $weights(y_i)$ is the extra penalty depending on the target value of the i^{th} pixel.

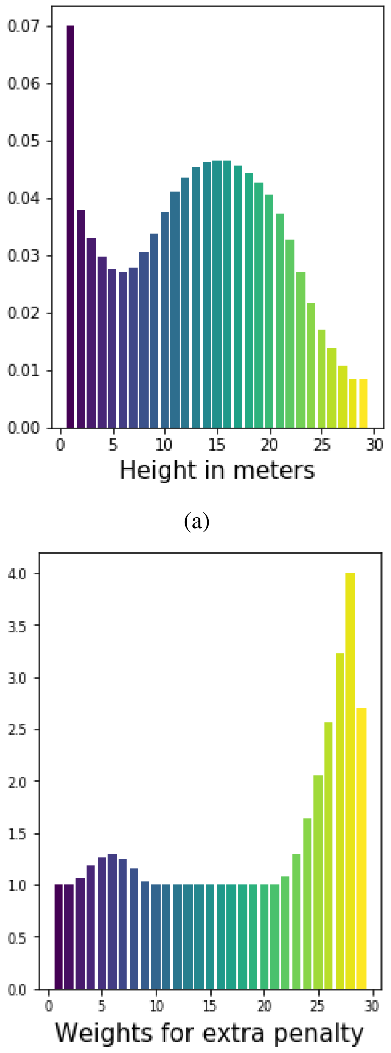
We needed to manage the temporal mismatch (such as logging) between LiDAR scanning and satellite imagery. To do so, we used two heuristics. The first one was that pixels labeled as forest by the forest cover model but with a height of less than 1 m were considered to be a forest logging. The forest cover model classifies pixels covered with clouds as non-forested. Therefore, the second heuristic was that pixels not labeled as forest but with $CHM > 5$ m were considered clouds. Reference and predicted height values for these pixels were not used in the loss function calculation during the training procedure (they were treated as masked). Thus, the mask of relevant pixels was defined by the following equations:

$$logging = (height\_map < 1) * forest\_mask \tag{12}$$

$$cloud = (height\_map > 5) * (forest\_mask == 0) \tag{13}$$

$$height\_mask = (logging == 0) * (cloud == 0) \tag{14}$$

where forest mask was obtained by the neural network model trained to predict forest cover with a high accuracy, especially in terms of small details using RGB bands. The model was implemented in the GeoAlert service [63].



(a)

(b)

FIGURE 7. **LiDAR-derived height distribution (a) and penalty weights for errors on corresponding height values (b). These weights are used during loss computation.**

G. EXPERIMENTAL SETTINGS

For all the neural network models, training was performed on the Skoltech supercomputer Zhores [64], using Keras [65] with a Tensorflow [66] backend. The source code containing the implementation details is available in the aforementioned repository.

Both for the regression and classification task, U-Net [67] with an Inception-ResNet-v2 [68] encoder was used (Figure 8). U-Net is a popular CNN architecture in the remote sensing domain which has shown high performance in various problems [69], [70]. The upsampling layers follow the U-Net's downsampling layers. Skip connections between layers allow the convolutional neural network to manipulate vital information at large spatial scales avoiding losing local information. Skip connections also facilitate gradient flow during the training procedure that was highlighted in [71]. We substituted the original VGG encoder with a ResNet-based one as it has shown high results in various works [72]. Residual connections in the Inception-ResNet-v2 encoder support shortcuts leading to better prediction quality [73] and enabling substantial simplification of the Inception blocks. We used the original U-Net decoder, where every step consists of an upsampling of the feature map followed by a 2 × 2 convolution. That halves the number of feature channels. The expansive path also includes concatenation with the cropped feature map from the contracting path and two 3 × 3 convolutions followed by a ReLU. The total number of parameters in the neural network is 62M where the encoder includes 54M. The decoder has 5 blocks, while the encoder part consist of 8 blocks. The models' implementation was based on opensource library [74].

Each model was trained 25 epochs for 200 training and 100 validation steps with a decreasing learning rate from 0.001 using RMSprop [75] optimizer and early stopping with patience 5 epochs. For the classification task as an activation function for the last layer, the softmax function was chosen. As an activation function for the last layer's regression model, we used linear function.

For all models, geometrical augmentation was implemented. This involves random rotations, and vertical and horizontal flipping. For models using the RGB channels only, we implemented color transformation. For this task, the albumentations library [76] was used.

H. CLASSICAL MACHINE LEARNING METHODS

We also conducted experiments with classical machine learning methods to compare different approaches in canopy height estimation. Two approaches were considered: Random Forest (RF) [43] and Gradient Boosting (GB) [77]. These approaches are widely used in the remote sensing domain due to relatively high performance in various tasks. For the RF method, we implemented 300 decision trees with maximum depth equal to 8, as these parameters shown the best quality. We also compared it to decision tree numbers 100, 200, 400, 500, 600, and maximum depth values equal to 4, 5, 6, 7, 8, 9, 10. In the GB method the parameters were 200 estimators with learning rate equal to 0.1, and maximum depth equal to 7, that were also set empirically (the same grid was considered to choose number of trees and maximum depth as in the RF case). For both two methods the implementation was used from scikit-learn [78]. A proper feature space is essential for machine learning algorithms, namely in classical one. The features were selected according to the study described in [79] as more relevant for vegetation properties estimation from Sentinel images. Therefore, the following vegetation indices were computed and accomplished initial multispectral bands resulting in Sentinel-derived features: the Normalized Difference Vegetation Index (NDVI), the Simple Ratio Index (SRI), the red-edge Normalized Difference Vegetation Index (RENDVI), and the Anthocyanin Reflectance Index 1 (ARI1). Thus, each pixel was considered as an input

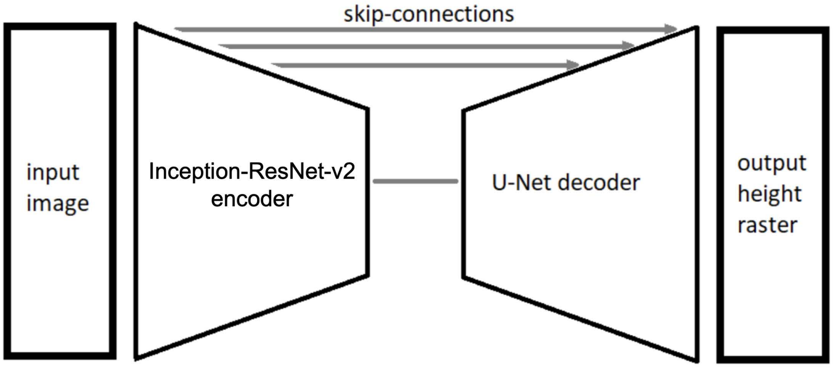


FIGURE 8. U-Net model with Inception-ResNet-v2 encoder.

sample with 14 features (10 Sentinel bands and 4 vegetation indexes) for a machine learning algorithm.

The following experiments were performed:

1) RF + Sentinel-derived features (Sentinel resolution 10 m)
2) RF + Sentinel-derived features (Sentinel resolution 10 m + ArcticDEM)
3) GB + Sentinel-derived features (Sentinel resolution 10 m)
4) GB + Sentinel-derived features (Sentinel resolution 10 m + ArcticDEM)

I. FOREST-TYPE CLASSIFICATION MODEL

To estimate the quality of the developed models, we considered a forest-type classification problem. To train the neural network model to predict two species (conifer and deciduous), we leveraged both WorldView and Sentinel imagery. The problem was defined as the per-pixel semantic segmentation task. Forest inventory characteristics were used as reference data. Eight WorldView bands were intersected with the forest mask. Both for the Sentinel and WorldView imagery, a height map or age map was used as an additional channel. This was done to make the model more robust in terms of species diversity resulting from different forest ages. Therefore, the neural network input was formed of 10 bands.

As mentioned above, there are two familiar sources of height values: LiDAR-derivied data and forest inventory characteristics. The difference is in the data representation. Forest inventory characteristics establish height for each individual stand (small region joined according to some similar value of features such as tree species, age, density). Although real height within each stand can differ for each pixel, all pixels corresponding to a particular stand have the same height value. Thus, for this experiment we used both inventory- and LiDAR-derived height data.

We compared model predictions according to the next strategies of data leveraging:

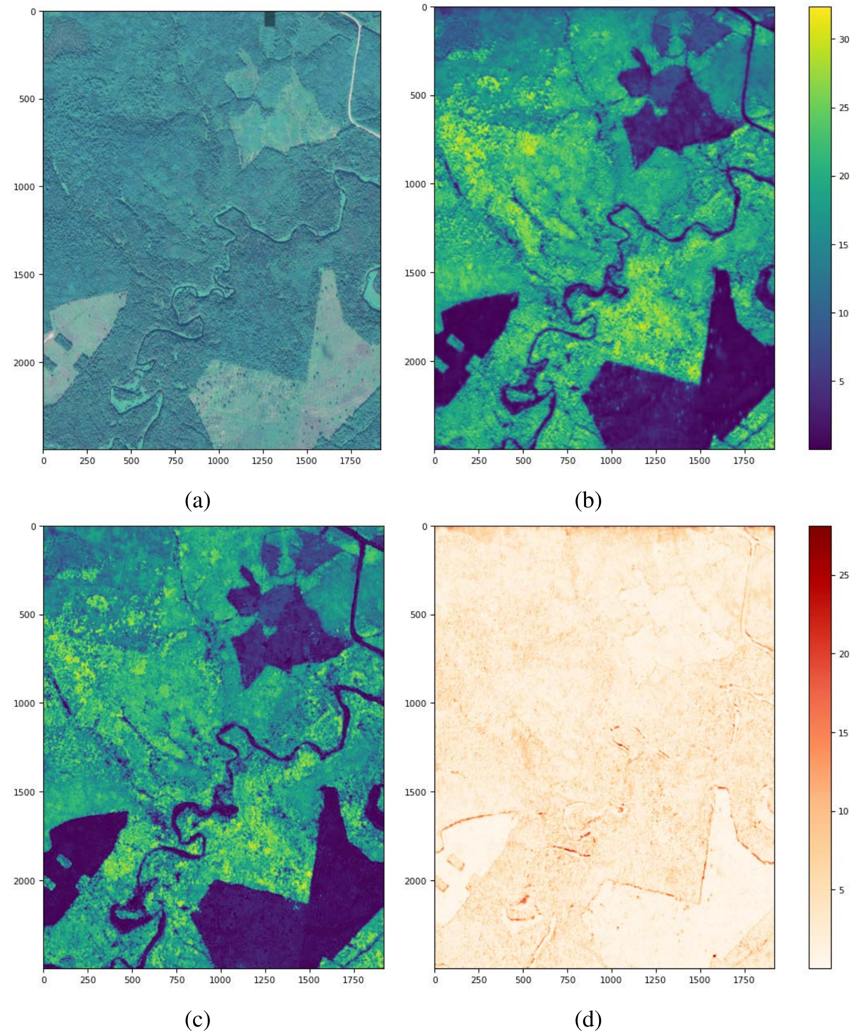
1) just multispectral data;
2) multispectral data and CHM data;
3) multispectral data and inventory height data;
4) multispectral data and inventory age data;
5) multispectral and artificially generated CHM by the best model height.

For these experiments, we trained a smaller U-Net model with the Resnet-34 encoder [80]. Individual stands from the dataset were randomly split into a training and testing set shown in Table 3. During training, the cross-entropy loss function was computed in a per-pixel manner. For testing, the F1-score was estimated for each individual stand. The predicted class for the individual stand was defined as a dominant class among all pixels within the stand. Each forest classification model was trained 25 epochs for 200 training and 100 validation steps with a decreasing learning rate from 0.001 using RMSprop [75] optimizer and early stopping with patience 5 epochs. The activation function for the last layer was soft-max.

IV. RESULTS

The achieved metrics for the regression models are shown in Table 4. The best quality predictions, using WorldView imagery with MAE 2.47 m (Exp. 9), were achieved with a combination of Red, Green, Blue pansharpened bands, the NIR band, and the supplementary ArcticDEM raster with resolution upsampled to 1 m (Fig 9). The smaller region is presented in the Fig 10. For the Sentinel imagery, only two experimental modes were considered: with ArcticDEM and without ArcticDEM. For both the Sentinel and WorldView data, ArcticDEM usage allowed us to improve the prediction results (for Sentinel, the MAE improved from 4.1 to 3.9 m, and for WorldView, the MAE improved from 2.9 to 2.58 m). The pansharpening procedure also contributed to the final result, decreasing the error from 3.3 to 3.1 m (Exp. 1 and Exp. 2) for the WorldView RGB model. The NIR band usage demonstrated an error reduction from 2.9 to 2.58 m (Exp. 3 and Exp. 7). This effect is linked to vegetation condition, which is reflected by the NIR wavelength. Additional weights during the loss computation reduced the MAE from 2.58 to 2.47 m (Exp. 7 and Exp. 9).

In Table 5, we can see a comparison between the regression model and the classification model (Fig 12). These two



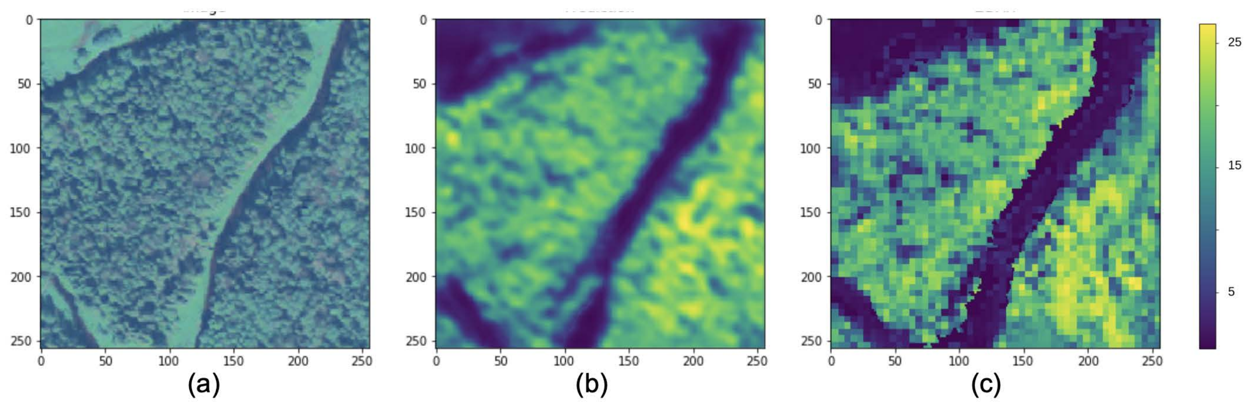
(a)

(b)

(c)

(d)

FIGURE 9. Input RGB WorldView image from test regions (a), generated CHM (b), LiDAR-derived height (c), error (d). Height measurements are in m.



(a)

(b)

(c)

FIGURE 10. Input RGB WorldView image from test regions (pansharpened to 1 m) (a), generated height (b), LiDAR height (downsampled to 5 m) (c).

models were trained using the same input data. The regression model's prediction was split into four appropriate height classes and the F1-score was calculated. This confirmed the assumption that after training the model to predict continuous values, the final results were not worse than the discreet ones (F1-score: 0.68 and 0.67). Moreover, the regression spectrum

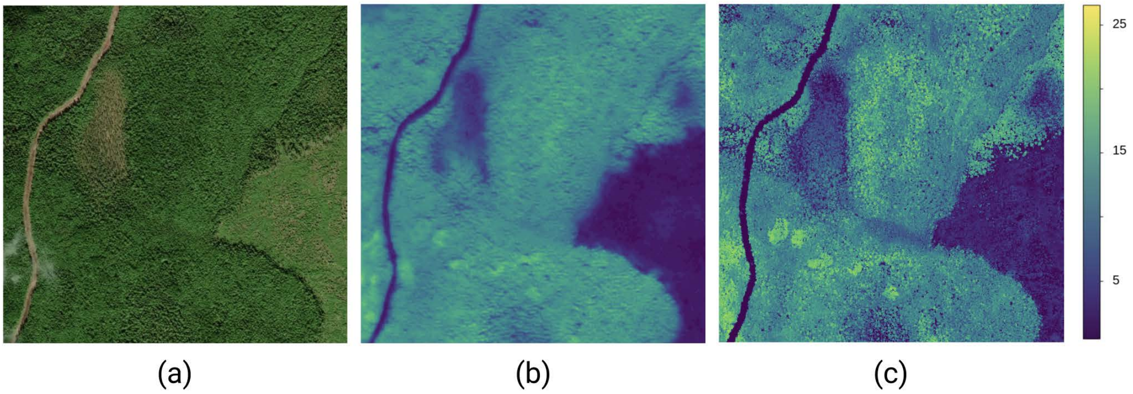


FIGURE 11. Input RGB Mapbox image from test regions (a), generated height (b), LiDAR height (c).

TABLE 4. Regression models (error in m).

Exp.	Description	MAE	RMSE	MBE
1	RGB (original resolution 2 m)	3.3	4.5	0.024
2	RGB (pansharpened to 1 m resolution)	3.1	4.3	0.009
3	RGB (pansharpened to 1 m resolution + ArcticDEM)	2.9	4.1	-0.75
4	RGB+NIR (original resolution 2 m)	2.9	4.1	-0.661
5	RGB+NIR (original resolution 2 m + ArcticDEM)	2.8	4	-0.132
6	RGB+NIR (RGB pansharpened to 1 m resolution)	2.9	4.1	-0.8
7	RGB+NIR (RGB pansharpened to 1 m resolution + ArcticDEM)	2.58	3.8	-0.99
8	RGB (downsampled to 5 m resolution)	4.4	5.9	0.65
9	Weighted RMSE RGB+NIR (RGB pansharpened to 1 m resolution + ArcticDEM)	**2.47**	**3.6**	**-0.267**
10	Multispectral (Sentinel resolution 10 m)	4.1	5.7	0.79
11	Multispectral (Sentinel resolution 10 m + ArcticDEM)	3.9	5.4	0.32
12	RF + Sentinel-derived features (Sentinel resolution 10 m)	4.3	5.6	0.91
13	RF + Sentinel-derived features (Sentinel resolution 10 m + ArcticDEM)	4.1	5.4	0.82
14	GB + Sentinel-derived features (Sentinel resolution 10 m)	4.2	5.5	-0.82
15	GB + Sentinel-derived features (Sentinel resolution 10 m + ArcticDEM)	4.	5.4	-0.78

TABLE 5. Classification task (F1-score in m). Exp. 1: Weighted RMSE RGB+NIR (RGB pansharpened to 1 m resolution + ArcticDEM). Exp. 2: Classification model RGB+NIR (RGB pansharpened to 1 m resolution + ArcticDEM).

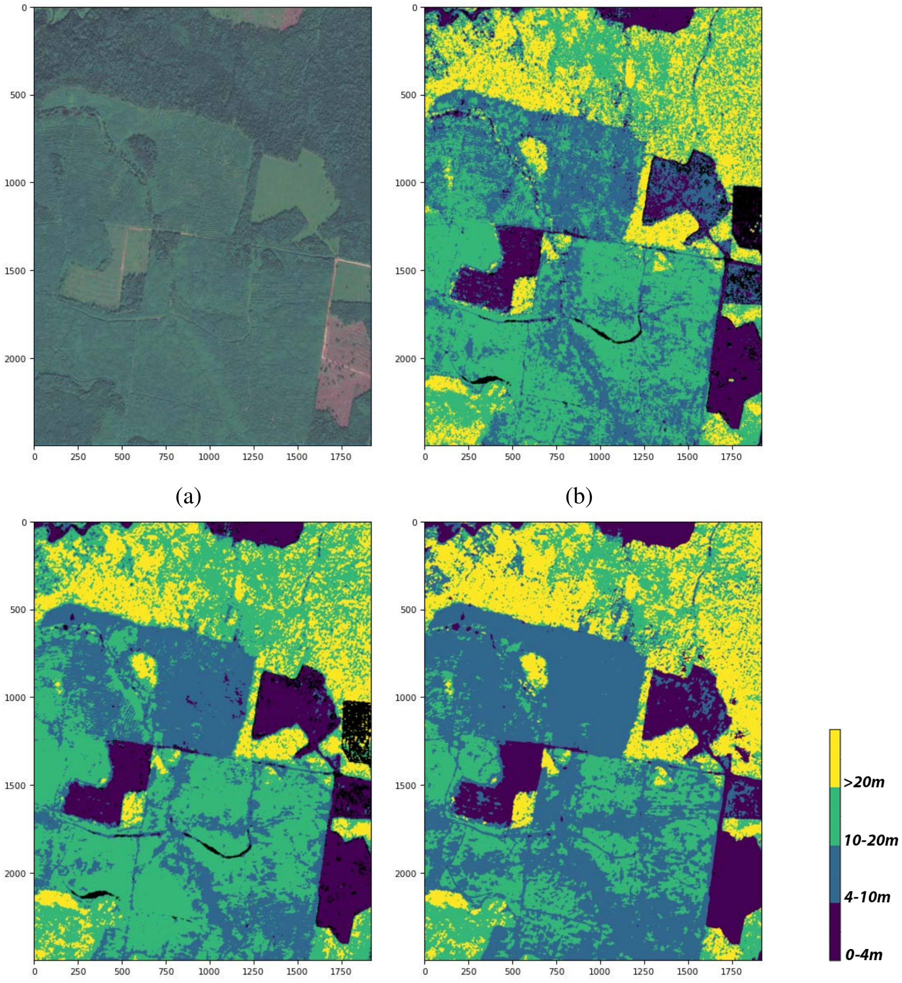
Exp.	0-4	4-10	10-20	> 20	Average F1
1	0.79	0.51	0.84	0.6	0.68
2	0.79	0.49	0.78	0.62	0.67

TABLE 6. Forest-type classification (average for all classes F1-score) for WorldView and Sentinel imagery. Generated height is derivied from the best model predictions (Exp. 9 Weighted RMSE RGB+NIR (RGB pansharpened to 1 m resolution + ArcticDEM).

Description	WorldView	Sentinel
multispectral	0.87	0.88
multispectral + CHM	0.9	0.92
multispectral + inventory height	0.9	0.93
multispectral + inventory age	0.93	0.94
multispectral + generated	0.89	0.90

of values makes the model more flexible, e.g., other classes can be presented and it does not require extra training for new splitting into target classes. This approach would be of potential interest for use in other forest characteristics computations.

The recognition class that is most difficult to process is the height between $4 - 10$ m. This is mainly caused by the spatial distribution specificity of the class, and it often occurs due to the small regions between crowns and depends dramatically on the satellite and LiDAR geo-reference data. For this study, we used LiDAR data downsampled to 5 m, while the WorldView imagery resolution was 1 or 2 m. This allowed us to save high-resolution spatial surface characteristics.



(a)

(b)

(c)

(d)

FIGURE 12. **Input RGB WorldView image from test regions (a), original height classes (b), generated height classes in regression (c) and classification (d) problem statement.**

To assess the importance of texture information, we experimented with RGB bands downsampled to 5 m 4. The MAE for this case was 4.4 m. This result is lower than that of the Sentinel images (4.1 m) and confirms that when we reduced the spectral information, we faced stricter demands for spatial resolution.

We checked the generated height in the forestry task of species classification. The results are presented in Table 6. The first objective of the experiment was to show how supplementary features can enhance the quality of applied tasks. Both LiDAR and inventory data helped to improve classification in comparison with simple multispectral data. The second goal was to show that the generated height is of sufficient quality to beat the base model using just satellite data. We did not intend to conduct a comparison between WorldView and Sentinel sources. For this reason, in both experiments, observation dates were not equal in the data used. The superior results for the Sentinel imagery, as compared with the WorldView data, were partially due to the wider dataset.

We also evaluated the regression model trained using RGB WorldView (pansharpened to 1 m resolution) image on a cloud-free composite orthophotomap provided by Mapbox [60] and covering the same test area. For this experiment, the MAE was equal to 3.5, and the RMSE was 4.6. Prediction example is shown in Figure 11. This promising result allows cheaper CHM estimation for large areas using only high-resolution free-available satellite RGB data.

We conducted experiments with classical machine learning algorithms using Sentinel-derived features to compare this approach to the proposed one, namely the CNN-based with high-resolution data. The best results were achieved for the GB algorithm and combination of Sentinel-derived features with ArcticDEM, where MAE was equal to 4 and RMSE was equal to 5.4 4.

V. DISCUSSION

It is challenging to perform a fair comparison between the majority of studies related to height estimation for various reasons. The main reason is the difference in height distribution. For example, in [37], the predicted height was limited by 30 m, the spatial resolution was 5 m, and the final RMSE was 2.2 m. However, according to the presented plots, the mean value was less than 10 m, while in our study, it was about 15 m. In [28], the validation pixels range was defined as being from 0 to 25 m, with a mean value of 7 m. The model's spatial resolution was 30 m. For this height distribution, an RMSE from 2.3 to 4.1 m was achieved. In [31], they studied the ranges between 0 to 18 m and 3 to 15 m, by leveraging satellite (both spectral and radar) data with a 20 m resolution. In contrast to our work, field-based observations with a sampling frequency of the 10 largest trees per inventory plot were used as reference material. Therefore, the achieved result (an RMSE of 1.48 m) cannot be compared with our model's performance. Other obstacles impeding a fair comparison are the species diversity and regional conditions.

It is worth mentioning that although ArcticDEM provides a stable improvement in canopy height estimation (see table 4, Exp. 6 and Exp. 7), it does not cover central or southern regions. For these areas, more powerful base models need to be implemented, leveraging just satellite imagery.

We showed that high-resolution WorldView 3-bands images provided more significant features than low resolution Sentinel with 10 spectral bands (see table 4, Exp. 2 and Exp. 10). However, resolution adjustment from 2 m to 5 m for the same WorldView dataset leads to a loss of important information, in particular texture information (see table 4, Exp. 2 and Exp. 8). The aforementioned experiments, which was performed on the same dataset and using the same NNs with only one difference - the adjusted spatial resolution, showed that neural networks can extract additional spatial features from very high-resolution optical images of 1 m. Thus we experimentally confirmed the initial hypothesis that by using high resolution data it is possible to make CHM estimation more accurate.

Creating the model with only high-resolution RGB channels allows it to be implemented in more available satellite images, such as RGB mosaic basemaps (google, yandex, and Mapbox). Therefore, an opportunity to replace WorldView data with satellite images derived from other sources, making the provided model more universal. We made a prediction for cloud-free composite orthophotomap provided by Mapbox [60] using the CNN model trained on RGB 1 m bands. The achieved quality (MAE = 3.5) confirms the opportunity for further model application for basemaps analysis.

There are the following directions for future research. The first involves improving the co-registration between LiDAR and satellite data. Now the developed RGB-based model shows the ability to reconstruct the main patterns corresponding to the CHM (Fig 10); large individual trees and spots within forest are detected successfully. However, satellite data has a slight shift in comparison with LiDAR data. Improving co-registration would allow the model's performance to be assessed more accurately for resolutions of less or equal to 1 m and also could probably improve the poor performance for the class of 4–10 m.

The ability of the model to be transferred to new regions is another essential question. As we did not have data from other regions, it is impossible to judge the model robustness for new areas. Moreover, for some regions, the ArcticDEM layer is not available; therefore, additional training for new areas might improve prediction quality. However, the neural network approach has proven to be powerful enough to extract the necessary spatial information and adapt to changing natural conditions. Augmentation and image diversity are often applied to overcome this weakness in real-life applications.

Another possible objective for future research is a canopy height estimation for areas with complex topography. Neural network models rely on landcover's spectral and texture characteristics, making the initial approach promising even when topography is not flat. However, shadows on slopes pose additional challenges to the multispectral satellite image analysis. LiDAR data additional preprocessing is also considered for study areas with complex topography [81].

In this study, we used all available images both for training and testing (splitting them into training and testing regions) as it is a common choice in the remote sensing domain [82]. However, in the future work, image-based cross-validation techniques can be used and robustness for new environmental conditions can be considered [83].

VI. CONCLUSION

Overall, in this study we confirm the hypothesis that neural networks can extract significant spatial features from very high-resolution RGB images, which can be used for more precise canopy height estimation. We also checked whether it is possible to get an accuracy of canopy height estimation by using of satellite-based solutions compatible with measurements obtained by UAV approach. For checking our assumptions we analysed the potential of very high-resolution images with limited spectral information in the task of canopy height model estimation. We created a software toolchain based on a state-of-the-art neural network architecture that enable us to extract spatial features from very high-resolution images. The proposed approach led to a reduction in the mean absolute error to 2.4 m, while leveraging just four spectral bands and the supplementary features from ArcticDEM. However, in southern regions where ArcticDEM is not available and without other sufficiently accurate DEM, the model achieved an MAE of 2.9 m. We also examined how generated height can be successfully used in the forest classification task. Our canopy height model estimation results using RGB bands indicated the prospect of replacing expensive LiDAR sensing data with easily attainable satellite data. Depending on the region of study, our technique allows a customer to promptly collect all the necessary relevant forestry inventory information without ground-based observations. We also developed and shared the easy-to-use open source solution which gives a new possibilities for the community to solve similar tasks. In future works, we are planning to include texture data, indexes and other attributes that can be obtained using ArcticDEM in the modeling procedure.

ACKNOWLEDGMENT

The authors acknowledge the use of the Skoltech CDISE Supercomputer Zhores [64] for obtaining the results presented in this article.

REFERENCES

[1] N. M. Thomas, P. Baltezar, D. Lagomasino, S.-K. Lee, T. Fatoyinbo, J. Green, and M. Rahman, "Extent and canopy height maps of trees outside forest (ToF) for Bangladesh," in *Proc. AGUFM*, 2018.

[2] Ø. D. Trier, A.-B. Salberg, J. Haarpaintner, D. Aarsten, T. Gobakken, and E. Næsset, "Multi-sensor forest vegetation height mapping methods for Tanzania," *Eur. J. Remote Sens.*, vol. 51, no. 1, pp. 587–606, Jan. 2018.

[3] H. Huang, C. Liu, X. Wang, G. S. Biging, Y. Chen, J. Yang, and P. Gong, "Mapping vegetation heights in China using slope correction ICESat data, SRTM, MODIS-derived and climate data," *ISPRS J. Photogramm. Remote Sens.*, vol. 129, pp. 189–199, Jul. 2017.

[4] H. Zhang, Y. Sun, L. Chang, Y. Qin, J. Chen, Y. Qin, J. Du, S. Yi, and Y. Wang, "Estimation of grassland canopy height and aboveground biomass at the quadrat scale using unmanned aerial vehicle," *Remote Sens.*, vol. 10, no. 6, p. 851, May 2018.

[5] S. Illarionova, A. Trekin, V. Ignatiev, and I. Oseledets, "Neural-based hierarchical approach for detailed dominant forest species classification by multispectral satellite imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 1810–1820, 2021.

[6] L. Mou and X. Xiang Zhu, "IM2HEIGHT: Height estimation from single monocular imagery via fully residual convolutional-deconvolutional network," 2018, *arXiv:1802.10249*.

[7] A. N. Trekin, V. Y. Ignatiev, and P. Y. Yakubovskii, "Deep neural networks for determining the parameters of buildings from single-shot satellite imagery," *J. Comput. Syst. Sci. Int.*, vol. 59, no. 5, pp. 755–767, Sep. 2020.

[8] B. Wu, Y. Zeng, and D. Zhao, "Land cover mapping and above ground biomass estimation in China," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2016, pp. 3535–3536.

[9] Y. Sadeghi, B. St-Onge, B. Leblon, J.-F. Prieur, and M. Simard, "Mapping boreal forest biomass from a SRTM and TanDEM-X based on canopy height model and landsat spectral indices," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 68, pp. 202–213, Jun. 2018.

[10] D. Gwenzi, E. Helmer, X. Zhu, M. Lefsky, and H. Marcano-Vega, "Predictions of tropical forest biomass and biomass growth based on stand height or canopy area are improved by Landsat-scale phenology across Puerto Rico and the U.S. Virgin islands," *Remote Sens.*, vol. 9, no. 2, p. 123, Feb. 2017.

[11] T. Sasaki, J. Imanishi, K. Ioki, Y. Morimoto, and K. Kitada, "Object-based classification of land cover and tree species by integrating airborne LiDAR and high spatial resolution imagery data," *Landscape Ecol. Eng.*, vol. 8, no. 2, pp. 157–171, Jul. 2012.

[12] 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, Aug. 2012.

[13] T. Majasalmi, S. Eisner, R. Astrup, J. Fridman, and R. M. Bright, "An enhanced forest classification scheme for modeling vegetation–climate interactions based on national forest inventory data," *Biogeosciences*, vol. 15, no. 2, pp. 399–412, Jan. 2018.

[14] M. D. Venturas, H. N. Todd, A. T. Trugman, and W. R. L. Anderegg, "Understanding and predicting forest mortality in the western United States using long-term forest inventory data and modeled hydraulic damage," *New Phytologist*, vol. 230, no. 5, pp. 1896–1910, Jun. 2021.

[15] H. Haakana, "Multi-source forest inventory data for forest production and utilization analyses at different levels," *Dissertationes Forestales*, vol. 2017, no. 243, 2017.

[16] A. Matese, S. F. Di Gennaro, and A. Berton, "Assessment of a canopy height model (CHM) in a vineyard using UAV-based multispectral imaging," *Int. J. Remote Sens.*, vol. 38, nos. 8–10, pp. 2150–2160, May 2017.

[17] C. Stone, M. Webster, J. Osborn, and I. Iqbal, "Alternatives to LiDAR-derived canopy height models for softwood plantations: A review and example using photogrammetry," *Austral. Forestry*, vol. 79, no. 4, pp. 271–282, Oct. 2016.

[18] D. Lagomasino, T. Fatoyinbo, S. Lee, E. Feliciano, C. Trettin, and M. Simard, "A comparison of mangrove canopy height using multiple independent measurements from land, air, and space," *Remote Sens.*, vol. 8, no. 4, p. 327, Apr. 2016.

[19] S. Hartling, V. Sagan, P. Sidike, M. Maimaitijiang, and J. Carron, "Urban tree species classification using a WorldView-2/3 and LiDAR data fusion approach and deep learning," *Sensors*, vol. 19, no. 6, p. 1284, Mar. 2019.

[20] J. Marrs and W. Ni-Meister, "Machine learning techniques for tree species classification using co-registered LiDAR and hyperspectral data," *Remote Sens.*, vol. 11, no. 7, p. 819, Apr. 2019.

[21] Z. Lin, Q. Ding, J. Huang, W. Tu, D. Hu, and J. Liu, "Study on tree species classification of UAV optical image based on densenet," *Remote Sens. Technol. Appl.*, vol. 34, no. 4, pp. 704–711, 2019.

[22] S. Nezami, E. Khoramshahi, O. Nevalainen, I. Pölönen, and E. Honkavaara, "Tree species classification of drone hyperspectral and RGB imagery with deep learning convolutional neural networks," *Remote Sens.*, vol. 12, no. 7, p. 1070, Mar. 2020.

[23] H. M. Nguyen, B. Demir, and M. Dalponte, "Weighted support vector machines for tree species classification using LiDAR data," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2019, pp. 6740–6743.

[24] T. Swinfield, J. A. Lindsell, J. V. Williams, R. D. Harrison, Agustiono, Habibi, E. Gemita, C. B. Schönlieb, and D. A. Coomes, "Accurate measurement of tropical forest canopy heights and aboveground carbon using structure from motion," *Remote Sens.*, vol. 11, no. 8, p. 928, Apr. 2019.

[25] G. D. Pearse, J. P. Dash, H. J. Persson, and M. S. Watt, "Comparison of high-density LiDAR and satellite photogrammetry for forest inventory," *ISPRS J. Photogramm. Remote Sens.*, vol. 142, pp. 257–267, Aug. 2018.

[26] K. Fankhauser, N. Strigul, and D. Gatziolis, "Augmentation of traditional forest inventory and airborne laser scanning with unmanned aerial systems and photogrammetry for forest monitoring," *Remote Sens.*, vol. 10, no. 10, p. 1562, Sep. 2018.

[27] (2021). *Mapflow.AI*. Accessed: Feb. 10, 2022. [Online]. Available: https://docs.mapflow.ai/userguides/pipelines.html

[28] G. Staben, A. Lucieer, and P. Scarth, "Modelling LiDAR derived tree canopy height from Landsat TM, ETM+ and OLI satellite imagery—A machine learning approach," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 73, pp. 666–681, Dec. 2018.

[29] M. C. Hansen, P. V. Potapov, S. J. Goetz, S. Turubanova, A. Tyukavina, A. Krylov, A. Kommareddy, and A. Egorov, "Mapping tree height distributions in Sub-Saharan Africa using Landsat 7 and 8 data," *Remote Sens. Environ.*, vol. 185, pp. 221–232, Nov. 2016.

[30] S. Tao, Q. Guo, C. Li, Z. Wang, and J. Fang, "Global patterns and determinants of forest canopy height," *Ecology*, vol. 97, no. 12, pp. 3265–3270, Dec. 2016.

[31] S. M. Ghosh, M. D. Behera, and S. Paramanik, "Canopy height estimation using Sentinel series images through machine learning models in a mangrove forest," *Remote Sens.*, vol. 12, no. 9, p. 1519, May 2020.

[32] N. Verma, D. Lamb, N. Reid, and B. Wilson, "Comparison of canopy volume measurements of scattered eucalypt farm trees derived from high spatial resolution imagery and LiDAR," *Remote Sens.*, vol. 8, no. 5, p. 388, May 2016.

[33] N. Lang, K. Schindler, and J. D. Wegner, "Country-wide high-resolution vegetation height mapping with Sentinel-2," *Remote Sens. Environ.*, vol. 233, Nov. 2019, Art. no. 111347.

[34] S. Puliti, M. Hauglin, J. Breidenbach, P. Montesano, C. S. R. Neigh, J. Rahlf, S. Solberg, T. F. Klingenberg, and R. Astrup, "Modelling above-ground biomass stock over Norway using national forest inventory data with ArcticDEM and Sentinel-2 data," *Remote Sens. Environ.*, vol. 236, Jan. 2020, Art. no. 111501.

[35] W.-J. Lee and C.-W. Lee, "Forest canopy height estimation using multiplatform remote sensing dataset," *J. Sensors*, vol. 2018, pp. 1–9, Jan. 2018.

[36] O. Csillik, P. Kumar, and G. P. Asner, "Challenges in estimating tropical forest canopy height from planet dove imagery," *Remote Sens.*, vol. 12, no. 7, p. 1160, Apr. 2020.

[37] A. J. H. Meddens, L. A. Vierling, J. U. H. Eitel, J. S. Jennewein, J. C. White, and M. A. Wulder, "Developing 5 m resolution canopy height and digital terrain models from WorldView and ArcticDEM data," *Remote Sens. Environ.*, vol. 218, pp. 174–188, Dec. 2018.

[38] *Maxar Basemaps*. Accessed: 2020. [Online]. Available: https://www.maxar.com/products/imagery-basemaps

[39] Y.-S. Lee, S. Lee, W.-K. Baek, H.-S. Jung, S.-H. Park, and M.-J. Lee, "Mapping forest vertical structure in Jeju island from optical and radar satellite images using artificial neural network," *Remote Sens.*, vol. 12, no. 5, p. 797, Mar. 2020.

[40] X. Ni, M. Xu, C. Cao, W. Chen, B. Yang, and B. Xie, "Forest height estimation and change monitoring based on artificial neural network using geoscience laser altimeter system and Landsat data," *J. Appl. Remote Sens.*, vol. 14, no. 2, 2019, Art. no. 022207.

[41] Y.-S. Lee, S. Lee, and H.-S. Jung, "Mapping forest vertical structure in Gong-ju, Korea using Sentinel-2 satellite images and artificial neural networks," *Appl. Sci.*, vol. 10, no. 5, p. 1666, Mar. 2020.

[42] S. A. A. Shah, M. A. Manzoor, and A. Bais, "Canopy height estimation at Landsat resolution using convolutional neural networks," *Mach. Learn. Knowl. Extraction*, vol. 2, no. 1, pp. 23–36, Feb. 2020.

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

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

[45] H. He, Y. Yan, T. Chen, and P. Cheng, "Tree height estimation of forest plantation in mountainous terrain from bare-earth points using a DoG-coupled radial basis function neural network," *Remote Sens.*, vol. 11, no. 11, p. 1271, May 2019.

[46] R. Özçelik, Q. V. Cao, G. Trincado, and N. Göçer, "Predicting tree height from tree diameter and dominant height using mixed-effects and quantile regression models for two species in Turkey," *Forest Ecol. Manage.*, vols. 419–420, pp. 240–248, Jul. 2018.

[47] R. P. Sharma, Z. Vacek, S. Vacek, and M. Kučera, "Modelling individual tree height–diameter relationships for multi-layered and multi-species forests in central Europe," *Trees*, vol. 33, no. 1, pp. 103–119, Feb. 2019.

[48] P. Rahimzadeh-Bajgiran, M. Munehiro, and K. Omasa, "Relationships between the photochemical reflectance index (PRI) and chlorophyll fluorescence parameters and plant pigment indices at different leaf growth stages," *Photosynthesis Res.*, vol. 113, nos. 1–3, pp. 261–271, Sep. 2012.

[49] T. Aakala, T. Kuuluvainen, T. Wallenius, and H. Kauhanen, "Tree mortality episodes in the intact picea abies-dominated Taiga in the arkhangelsk region of northern European Russia," *J. Vegetation Sci.*, vol. 22, no. 2, pp. 322–333, Apr. 2011.

[50] *Order of the Federal Forestry Agency (Rosleskhoz) of December 12, 2011 n 516 Moscow 'on Approval of the Forest Inventory Instruction'] 'Prikaz Federal'nogo Agentstva Lesnogo Hozyajstva (Rosleskhoz) ot 12 Dekabrya 2011 g. n 516 g. Moskva 'ob Utverzhdenii Lesoustroitel'noj Instrukcii*, Federal Forestry Agency, Moscow, Russia, 2012.

[51] D. W. Wanik, J. R. Parent, E. N. Anagnostou, and B. M. Hartman, "Using vegetation management and LiDAR-derived tree height data to improve outage predictions for electric utilities," *Electr. Power Syst. Res.*, vol. 146, pp. 236–245, 2017.

[52] *GBDX*. Accessed: 2020. [Online]. Available: https://gbdxdocs.digitalglobe.com/

[53] R. Vaddi and P. Manoharan, "Hyperspectral image classification using CNN with spectral and spatial features integration," *Infr. Phys. Technol.*, vol. 107, Jun. 2020, Art. no. 103296. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1350449520300876

[54] M. Debella-Gilo and A. K. Gjertsen, "Mapping seasonal agricultural land use types using deep learning on Sentinel-2 image time series," *Remote Sens.*, vol. 13, no. 2, p. 289, Jan. 2021.

[55] 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. (RTEICT)*, May 2016, pp. 1778–1781.

[56] *Earthexplorer USGS*. Accessed: 2020. [Online]. Available: https://earthexplorer.usgs.gov/

[57] *Sen2cor*. Accessed: 2020. [Online]. Available: https://step.esa.int/main/third-party-plugins-2/sen2cor/

[58] H. Astola, T. Häme, L. Sirro, M. Molinier, and J. Kilpi, "Comparison of Sentinel-2 and Landsat 8 imagery for forest variable prediction in boreal region," *Remote Sens. Environ.*, vol. 223, pp. 257–273, Mar. 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0034425719300252

[59] *ArcticDem*. Accessed: 2020. [Online]. Available: https://www.pgc.umn.edu/data/arcticdem/

[60] *Mapbox Service*. Accessed: 2021. [Online]. Available: https://www.mapbox.com/maps

[61] B. Peterson and K. Nelson, "Mapping forest height in Alaska using GLAS, Landsat composites, and airborne LiDAR," *Remote Sens.*, vol. 6, no. 12, pp. 12409–12426, Dec. 2014.

[62] C. Goutte and E. Gaussier, "A probabilistic interpretation of precision, recall and F-score, with implication for evaluation," in *Proc. Eur. Conf. Inf. Retr.* Berlin, Germany: Springer, 2005, pp. 345–359.

[63] Geoalert.IO. (2020). *Geoalert Analytics Platform*. [Online]. Available: https://www.geoalert.io/en-U.S./

[64] I. Zacharov, R. Arslanov, M. Gunin, D. Stefonishin, A. Bykov, S. Pavlov, O. Panarin, A. Maliutin, S. Rykovanov, and M. Fedorov, "'Zhores'–petaflops supercomputer for data-driven modeling, machine learning and artificial intelligence installed in skolkovo institute of science and technology," *Open Eng.*, vol. 9, no. 1, pp. 512–520, 2019.

[65] (2020). *Keras*. [Online]. Available: https://keras.io/

[66] (2020). *TensorFlow*. [Online]. Available: https://github.com/tensorflow/tensorflow

[67] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2015, pp. 234–241.

[68] 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.

[69] T. Kattenborn, J. Leitloff, F. Schiefer, and S. Hinz, "Review on convolutional neural networks (CNN) in vegetation remote sensing," *ISPRS J. Photogramm. Remote Sens.*, vol. 173, pp. 24–49, Mar. 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0924271620303488

[70] K. Li, G. Wan, G. Cheng, L. Meng, and J. Han, "Object detection in optical remote sensing images: A survey and a new benchmark," *ISPRS J. Photogramm. Remote Sens.*, vol. 159, pp. 296–307, Jan. 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0924271619302825

[71] M. Drozdzal, E. Vorontsov, G. Chartrand, S. Kadoury, and C. Pal, "The importance of skip connections in biomedical image segmentation," in *Deep Learning and Data Labeling for Medical Applications*. Cham, Switzerland: Springer, 2016, pp. 179–187.

[72] J. L. Arrastia, N. Heilenkötter, D. O. Baguer, L. Hauberg-Lotte, T. Boskamp, S. Hetzer, N. Duschner, J. Schaller, and P. Maass, "Deeply supervised UNet for semantic segmentation to assist dermatopathological assessment of basal cell carcinoma," *J. Imag.*, vol. 7, no. 4, p. 71, Apr. 2021.

[73] C. A. Ferreira, T. Melo, P. Sousa, M. I. Meyer, E. Shakibapour, P. Costa, and A. Campilho, "Classification of breast cancer histology images through transfer learning using a pre-trained inception ResNet V2," in *Proc. Int. Conf. Image Anal. Recognit.* Cham, Switzerland: Springer, 2018, pp. 763–770.

[74] P. Yakubovskiy. (2019). *Segmentation Models*. [Online]. Available: https://github.com/qubvel/segmentation_models

[75] S. N. G. Hinton and K. Swersky, "Lecture 6D—A separate, adaptive learning rate for each connection. Slides of lecture neural networks for machine learning," Dept. Comput. Sci., Univ. Toronto, Toronto, ON, Canada, Tech. Rep., 2012.

[76] A. Buslaev, V. I. Iglovikov, E. Khvedchenya, A. Parinov, M. Druzhinin, and A. A. Kalinin, "Albumentations: Fast and flexible image augmentations," *Information*, vol. 11, no. 2, p. 125, 2020. [Online]. Available: https://www.mdpi.com/2078-2489/11/2/125

[77] J. H. Friedman, "Stochastic gradient boosting," *Comput. Stat. Data Anal.*, vol. 38, no. 4, pp. 367–378, 2002.

[78] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, A. Müller, J. Nothman, G. Louppe, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, Oct. 2012.

[79] N. Puletti, F. Chianucci, and C. Castaldi, "Use of Sentinel-2 for forest classification in Mediterranean environments," *Ann. Silvicultural Res.*, vol. 42, no. 1, pp. 32–38, 2018.

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

[81] J. Liu, A. K. Skidmore, M. Heurich, and T. Wang, "Significant effect of topographic normalization of airborne LiDAR data on the retrieval of plant area index profile in mountainous forests," *ISPRS J. Photogram. Remote Sens.*, vol. 132, pp. 77–87, Oct. 2017.

[82] E. Saralioglu and O. Gungor, "Semantic segmentation of land cover from high resolution multispectral satellite images by spectral-spatial convolutional neural network," *Geocarto Int.*, vol. 37, no. 2, pp. 1–21, 2022.

[83] S. Illarionova, S. Nesteruk, D. Shadrin, V. Ignatiev, M. Pukalchik, and I. Oseledets, "MixChannel: Advanced augmentation for multispectral satellite images," *Remote Sens.*, vol. 13, no. 11, p. 2181, Jun. 2021.

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 pursuing the Ph.D. degree in computer science with the Skolkovo Institute of Science and Technology, Moscow. Her research interests include computer vision, deep neural networks, and remote sensing.

DMITRII SHADRIN received the M.S. degree in applied physics and mathematics from the Moscow Institute of Physics and Technology (MIPT), in 2016, and the Ph.D. degree in data science from the Skolkovo Institute of Science and Technology (Skoltech), Russia, in 2020. He is currently a Research Scientist at Skoltech. His research interests include data processing, modeling of physical and bioprocesses in closed artificial growing systems, machine learning, and computer vision. He involved in the development of approaches for monitoring and modeling of the carbon footprint at Skoltech. He is responsible for the experimental research and several projects in the research center in artificial intelligence in the direction of optimization of management decisions to reduce carbon footprint.

VLADIMIR IGNATIEV graduated from the Moscow Institute of Physics and Technology, in 2012. He received the Ph.D. degree, in 2017. He is currently a Research Scientist at the Skolkovo Institute of Science and Technology, Moscow, Russia. Before joining Skoltech, he worked at Dorodnitsyn CCAS and the Aerocosmos Research Institute. He leads the Aeronet Laboratory that is focused on various applications of the deep learning methods to remote sensing data. He has experience in different remote sensing data processing and forecasting models development.

SERGEY SHAYAKHMETOV graduated from the MIET National Research University of Electronic Technology, in 1999. He is currently the Executive Director at the ESG Department, Sberbank, Russia. He is also in charge of research and development activities and implementation of AI tools in field of ESG for the largest bank in Europe. He has a vast experience in IT technologies, including founding a successful healthcare AI startup.

ALEXEY TREKIN graduated from the Moscow Institute of Physics and Technology, in 2012, and the Ph.D. degree in computer science, in 2017. He is currently a Research Scientist at the Skolkovo Institute of Science and Technology, Moscow, Russia. He is also the Head of research at the Aeronet Laboratory. From 2011 to 2017, he worked at the Moscow Aerocosmos Research Institute on problems of remote sensing data processing, including work on wildfire monitoring and impact assessment.

IVAN OSELEDETS graduated from the Moscow Institute of Physics and Technology, in 2006. He received the Candidate of Sciences and the Doctor of Sciences degrees from the Marchuk Institute of Numerical Mathematics, Russian Academy of Sciences, in 2007 and 2012, respectively. In 2013, he joined Skoltech CDISE. 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. It resulted in publications in top computer science conferences, such as ICML, NIPS, ICLR, CVPR, RecSys, ACL, and ICDM. His current research interests include 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. He is an Associate Editor of *SIAM Journal on Mathematics in Data Science*, *SIAM Journal on Scientific Computing*, and *Advances in Computational Mathematics* (Springer).