



Article

Improved LSTM Model for Boreal Forest Height Mapping Using Sentinel-1 Time Series

Shaojia Ge ¹, Weimin Su ¹, Hong Gu ¹, Yrjö Rauste ², Jaan Praks ³ and Oleg Antropov ^{2,*}¹ School of Electronic and Optical Engineering, Nanjing University of Science and Technology, Nanjing 210094, China² VTT Technical Research Centre of Finland, P.O. Box 1000, 00076 Espoo, Finland³ Department of Electronics and Nanoengineering, Aalto University, P.O. Box 11000, 00076 Aalto, Finland

* Correspondence: oleg.antropov@vtt.fi

Abstract: Time series of SAR imagery combined with reference ground data can be suitable for producing forest inventories. Copernicus Sentinel-1 imagery is particularly interesting for forest mapping because of its free availability to data users; however, temporal dependencies within SAR time series that can potentially improve mapping accuracy are rarely explored. In this study, we introduce a novel semi-supervised Long Short-Term Memory (LSTM) model, CrsHelix-LSTM, and demonstrate its utility for predicting forest tree height using time series of Sentinel-1 images. The model brings three important modifications to the conventional LSTM model. Firstly, it uses a Helix-Elapse (HE) projection to capture the relationship between forest temporal patterns and Sentinel-1 time series, when time intervals between datatakes are irregular. A skip-link based LSTM block is introduced and a novel backbone network, Helix-LSTM, is proposed to retrieve temporal features at different receptive scales. Finally, a novel semisupervised strategy, Cross-Pseudo Regression, is employed to achieve better model performance when reference training data are limited. CrsHelix-LSTM model is demonstrated over a representative boreal forest site located in Central Finland. A time series of 96 Sentinel-1 images are used in the study. The developed model is compared with basic LSTM model, attention-based bidirectional LSTM and several other established regression approaches used in forest variable mapping, demonstrating consistent improvement of forest height prediction accuracy. At best, the achieved accuracy of forest height mapping was 28.3% relative root mean squared error (rRMSE) for pixel-level predictions and 18.0% rRMSE on stand level. We expect that the developed model can also be used for modeling relationships between other forest variables and satellite image time series.



Citation: Ge, S.; Su, W.; Gu, H.; Rauste, Y.; Praks, J.; Antropov, O. Improved LSTM Model for Boreal Forest Height Mapping Using Sentinel-1 Time Series. *Remote Sens.* **2022**, *14*, 5560. <https://doi.org/10.3390/rs14215560>

Academic Editor: Arturo Sanchez-Azofeifa

Received: 29 September 2022

Accepted: 29 October 2022

Published: 4 November 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: synthetic aperture radar; Sentinel-1; image time series; irregular sampling; tree height; boreal forest; LSTM; semi-supervised learning

1. Introduction

Timely assessment and monitoring of forests forms the basis for the definition and implementation of preventive and corrective measures for sustainable forest management and forest restoration after disturbances [1]. The dynamics of forest structural variables provides information on forest status and forest changes and represents key information for forest management purposes [2,3]. Traditional forest inventory variables, such as tree height, basal area, diameter at breast height and others can be used as inputs for forest biomass and carbon stock estimation, etc. Furthermore, many users require information on traditional forest inventory variables as well, for example, private forestry companies with smaller areas of interest, to support their forest management decisions.

Satellite-based operational forest inventories often use satellite optical data augmented by reference plots to produce forest maps and estimates [4,5]. When the use of optical satellite data is compromised due to near-permanent cloud coverage, a possible solution is to use synthetic aperture radar (SAR) sensors, relying on longer and denser time series.

Radar-based monitoring offers flexibility in forest monitoring applications, when results are requested in a fixed (e.g., yearly) schedule. Primary SAR data used in forest inventorying are L-band data because of the smaller saturation of the biomass-to-backscatter relationship compared to shorter wavelengths [6–8]. The use of advanced SAR temporal and textural features and imaging modes can improve the accuracy of forest variable prediction [9,10].

The European Copernicus program has opened new opportunities in forest mapping with the launch of Sentinel-1 satellites thanks to their high spatial and temporal resolution, their ability to form long image time series and their data provision at no cost to users [11]. The Sentinel-1 mission consists of a constellation of two polar-orbiting satellites mounting a C-band SAR imaging system. They offer a repeat cycle of six days with all-weather and day-and-night monitoring capabilities.

Multitemporal C-band SAR data were extensively used for evaluating and monitoring growing stock volume of both boreal and tropical forests, as well as in thematic mapping purposes [12–20]. The present consensus is that further research is required on methods exploiting dense time series of C-band SAR measurements, including multitemporal approaches, to achieve performance similar to L-band SAR data [8]. Important issues that need improvement are the relatively poor prediction accuracy and the lack of consistent ways to use SAR time series data [9]. One popular approach is to use multivariable models where each measurement/observation from time series is treated as an independent classification feature (or predictor variable in regression tasks). Such approaches are quite popular within machine learning and statistical non-parametric methodologies and have already been demonstrated with Sentinel-1 time series (or similar C-band sensors) in forest variable prediction and classification [10,15,21–23]. However, such approaches ignore explicit temporal dependencies between consecutive images and at best use only multitemporal variability as an independent classification feature [23]. As C-band SAR data have pronounced multitemporal variability and seasonal dependencies in the boreal zone [24–26], our expectation is that introducing multitemporal dynamics into prediction models can improve prediction accuracies. One possible solution to introduce an explicit temporal context is using Long Short-Term Memory (LSTM) models that can capture temporal relationships between consecutive images (timestamps).

LSTMs were previously demonstrated in several remote sensing applications, particularly land cover mapping and crop monitoring [26–34]. However, to date, the use of LSTMs in forest attribute prediction utilising Earth Observation (EO) data was limited if at all reported with SAR image time series. Additionally, the conventional LSTM still suffers from the gradient vanishing problem, making it less suitable for learning relationships using long image time series [35]. The presence of irregular time intervals between consecutive acquisitions due to, e.g., sensor malfunction or maintenance makes modeling even more complicated. Another general issue is the lack of supervision data, which could require semi-supervised approaches in model training. In this context, while consistency regularisation has demonstrated a success in semi-supervised classification tasks [36,37], there appear to be only a few reports on its successful use in regression models for predicting continuous forest variables [38].

In this manuscript, our primary aim is to develop an improved LSTM-based model for producing forest variable predictions using time series of Sentinel-1 data. In particular, we consider non-regular sampling of image datatakes and introduce a novel LSTM model that naturally takes the varying time variable into account. We compare our approach with the classical LSTM model and several other pixel-based regression approaches that are often used in satellite-based forest inventory. We use forest tree height as a representative forest structural variable in our paper.

The main novel contributions of our work are:

1. We introduce timestamps as an additional feature to better capture the relationship between SAR backscatter and forest variables. To accommodate possible irregular intervals between image datatakes, we propose a novel Helix-Elapse (HE) projection to explicitly model the circular seasonality pattern of the long image time series.

2. We introduce Skip-LSTM, a hybrid LSTM block featuring with skip-link structure to better capture long-term dependencies in time series data.
3. We employ a novel semi-supervised strategy called Cross-Pseudo Regression (CPR) to improve the model prediction performance with limited reference data.
4. We benchmark the developed improved LSTM model with other state-of-the-art versions of LSTM, as well as more conventional machine learning and statistical models, for the purpose of most precise forest height predictions using Sentinel-1 time series data. To the best of our knowledge, this is the first use of LSTM modeling of any kind for the purpose of forest inventory mapping.

The paper is organised as follows. We describe our study site, acquired Sentinel-1 time series dataset and developed modeling approaches in Sections 2 and 3. The performance of the developed model is analysed and compared to several benchmark approaches in Section 4. The computational complexity of the model is discussed in Section 5, as well as the impact of different stand sizes and dominant tree species on stand-level predictions, while potential challenges and opportunities are outlined in Section 6.

2. Materials

2.1. Study Site

The study site featuring the Hyytiälä forestry station is located in Central Finland, centre coordinates $37^{\circ}2'N$, $6^{\circ}11'E$ (WGS84). It represents a square area covering 2500 km^2 . The location of the site is shown in Figure 1 along with an RGB color composite of three Sentinel-1 images. Typical southern boreal forest types are present in the study area, such as Norway spruce (*Picea abies*), Scots pine (*Pinus sylvestris*) and birch (*Betula pendula*, *Betula pubescens*). The terrain is generally flat with the elevation ranging from 95 m to 230 m above sea level. Frozen conditions often start in October in this area. The minimum temperature in winter can drop to -25°C . The first snow often falls in November and melts away completely by early May. The snow layer depth can reach from 20 cm to 70 cm depending on weather conditions. Weather conditions over the study area are shown in Figure 2 according to the information from Finnish Meteorological Institute.

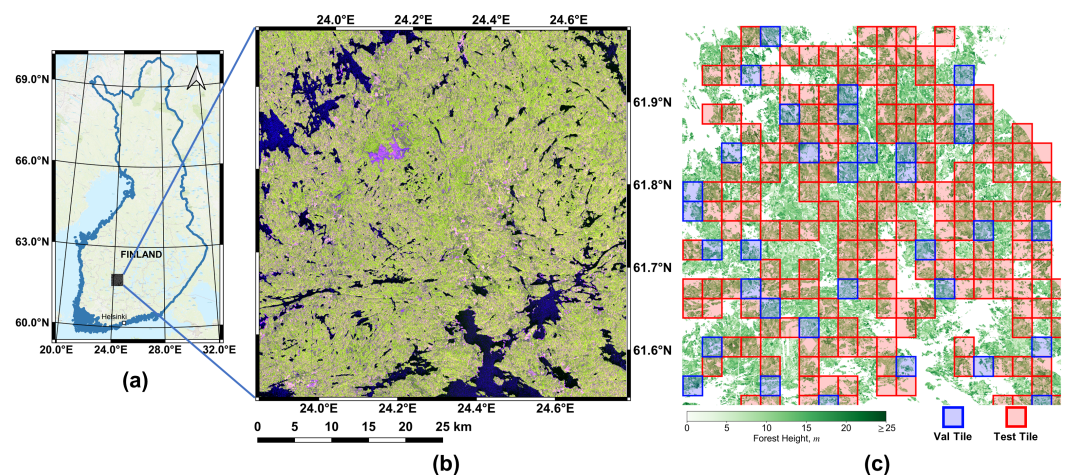


Figure 1. Study area: (a) study site location in Finland (WGS84), (b) RGB color-composite of 3 Sentinel-1 images (WGS84), (c) reference ALS-based forest height data with marked training and validation areas.

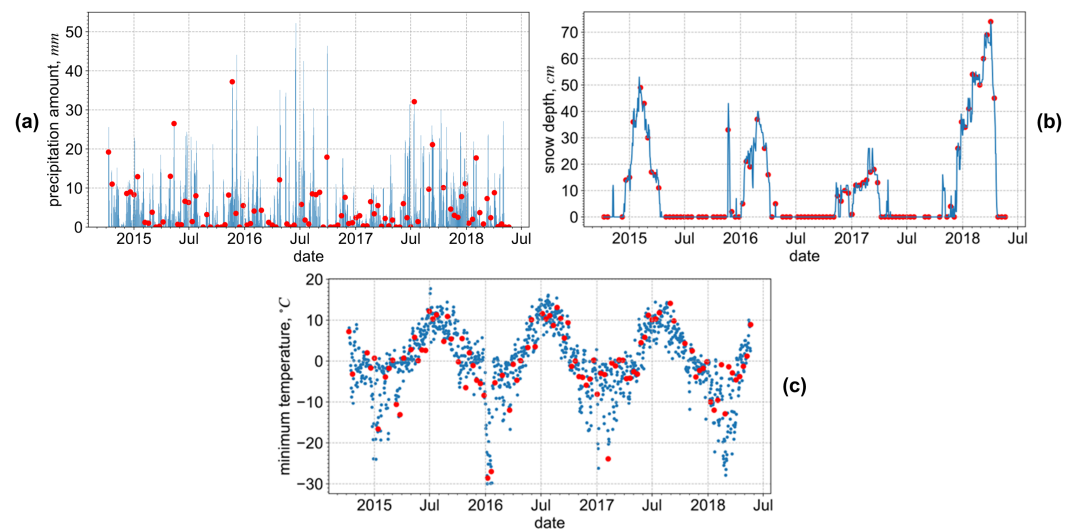


Figure 2. Daily weather history of the study site, from October 2014 to May 2018: (a) amount of precipitation within last three days, (b) snow depth, (c) minimum temperature. The red markers denote the timings of our Sentinel-1 datatime points.

2.2. SAR and Reference Data

Our dataset is represented by a long time series of Sentinel-1A IW-mode backscatter intensity images. Overall, there were 96 dual-polarisation (VV+VH) images acquired from 9 October 2014 to 21 May 2018. The time interval between adjacent observations ranges from 12 to 36 days. The original SAR data were acquired with Sentinel-1 satellites and initially preprocessed (focused and detected) and distributed by the European Space Agency (ESA) as ground range detected (GRD) products. SAR image orthorectification and radiometric normalisation were carried out using VTT in-house software [39]. Multilooking with a factor of 2×2 (range \times azimuth) was performed before orthorectification. Radiometric normalisation with respect to the projected area of the scattering element was performed to eliminate the topography-induced radiometric variation [40]. In this way, a time series of coregistered “gamma-naught” backscatter images was formed, with a pixel size of 20 m by 20 m. Each image in the stack has size of 2500 px \times 2500 px corresponding to an area of 50 km \times 50 km.

Airborne laser scanning (ALS) data collected by National Land Survey of Finland in summer 2015 are used as reference data. The forest height was computed by averaging relative heights of ALS cloud points after the ground removal within each mapping unit. The height ranges from 0 m to 25.5 m and the mean is 11.2 m. For additional comparisons with other conventional methods, stand-level estimates were also calculated from ALS data using a forest stand mask from the Finnish Forest Centre.

We split the dataset and corresponding reference into three subsets: training, validation and testing as shown in Figure 1c. Firstly, the whole area was equally divided into nonoverlapping tiles, the size of each tile was 128 px \times 128 px. Further, as depicted with red colour in Figure 1c, 50% of the tiles were randomly selected and pixels within them were extracted as the test subset. In a similar way, 10% of tiles were used to populate the validation subset and pixels from the remaining area composed the training dataset. The numbers of pixels are 1.5 mln, 0.375 mln and 1.778 mln for training, validation and testing subsets, respectively.

3. Methods

Here, we first briefly describe the fundamentals of LSTMs and introduce a Helix-Elapse (HE) projection concept to deal with the non-regular time interval between image timestamps. Further, we describe a semi-supervision regression and develop a model that combines several mentioned approaches. Baseline models that will be used in the benchmarking are also briefly described in this section.

3.1. Long Short-Term Memory Networks

Recurrent neural network (RNN) can capture temporal dependencies from a time sequence [41] thanks to its memory structure, in contrast to convolutional neural network (CNN) [42]. However, classic RNN structures suffer from gradient vanishing or explosion problems and fail to capture long-term dependencies. To handle this issue, a revised cell structure was proposed in [43]. As shown in the diagram in Figure 3c, the structure can be described as the “remember or forget” mechanism. Compared to the classical RNN cell, not only the previous hidden state h_{t-1} but also the previous memory c_{t-1} are fed into the current LSTM cell at timestamp t . They bring the history from previous timestamps into the current workflow. Then by integrating three gates (input gate i_t , forget gate f_t and output gate o_t), the flow of information is regulated and fused and a decision on whether to keep it is made using activation functions \tanh . Finally, the new current states, c_t and h_t , are calculated and fed forward into the next timestamp.

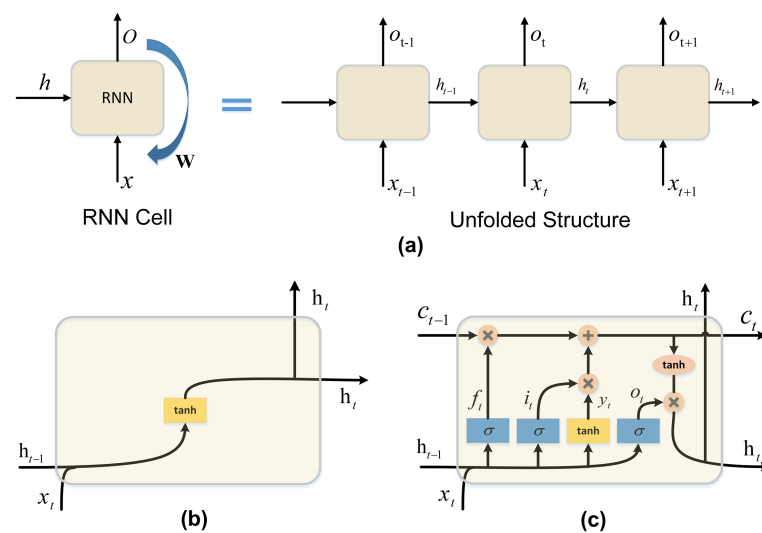


Figure 3. (a) Information flow diagram of RNN cell (on the left) and its unfolded structure (on the right). The two different structures of (b) the classic RNN cell and (c) the LSTM cell.

The corresponding equations are mathematically described as

$$i_t = \sigma(W_{ix}x_t + W_{ih}h_{t-1} + b_i), \quad (1)$$

$$f_t = \sigma(W_{fx}x_t + W_{fh}h_{t-1} + b_f), \quad (2)$$

$$o_t = \sigma(W_{ox}x_t + W_{oh}h_{t-1} + b_o), \quad (3)$$

$$y_t = \tanh(W_{yx}x_t + W_{yh}h_{t-1} + b_y), \quad (4)$$

$$c_t = i_t \odot y_t + f_t \odot c_{t-1}, \quad (5)$$

$$h_t = o_t \odot \tanh(c_t), \quad (6)$$

where σ and \tanh represent Sigmoid and Hyperbolic activation function separately, \odot denotes pointwise multiplication, W and b are weight matrix and bias vector.

3.2. Helix-Elapse Projection

Timestamps of acquisitions contain seasonal information and have explicit correlations with SAR image features, which may be considered as a priori knowledge for modeling. In particular, time series with irregular timestamps often happen in real world practice, for instance, the acquisition time interval differs from 12 to 36 in our case. How to alleviate the effect of irregular timestamps is an urgent problem. With timestamps as attributes, we can explicitly denote the acquisition interval differences and thus guide the training of the model.

We already knew that the seasonal pattern apparently exists in forest remote sensing images. Other than just bringing raw timestamps as linear attributes, we map them into a two-dimensional space, where the transformed timestamps can form a helix curve. The mapping is performed by an HE projection module, which can be mathematically described in the following equations:

$$\mathbf{t}_h = [t_1, t_2], \quad (7)$$

$$t_1 = t * \sin \frac{2\pi t}{n}, \quad (8)$$

$$t_2 = t * \cos \frac{2\pi t}{n}, \quad (9)$$

where t is the day index since 1 January 2014; n is the total number of days in a year—here, we simply assign it as 365; \mathbf{t}_h is the projected timestamp vector.

As visualised in Figure 4, let the origin point be the start date (1 January 2014); the circulating angle of the helix curve indicates the date of a year and the diameter indicates the growing year. The projected timestamps thus simulate the circulation of seasons in a heuristic way, as well as the year growing. We embed this HE projection module into our model and then stack the output, helix time attributes $\mathbf{T}_h = [\mathbf{T}_1, \mathbf{T}_2]$, as two additional vectors together with the original input.

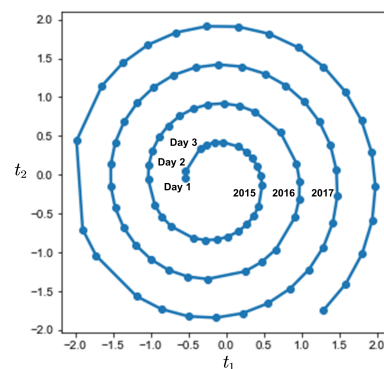


Figure 4. Helix time attributes for Sentinel-1 time series studied in the paper.

3.3. Skip-LSTM

In order to capture temporal dependencies within the SAR time series, we introduce LSTM as the backbone of our model. Despite the carefully designed memory structure, LSTM in practice often fails to capture very long-term correlations [35]. In the case of satellite observations spanning several years, the time series would be particularly long. This type of long-term dependency can hardly be captured by off-the-shelf recurrent units. Inspired by Dilated Convolution in CNN [44] and Recurrent-skip in LSTNet [35], we embed a Skip-LSTM module to alleviate long-term correlation issues. The structure of Skip-LSTM is shown in Figure 5. It consists of two parts: convolutional layer and Skip-LSTM layer.

Consider the input time series \mathbf{X} with a size of $T \times N_f$, where the length of the time series is T and the number of input channels is N_f . A convolutional layer is first applied to capture short-term features in the temporal domain as well as possible correlations between input channels. The convolutional layer is composed by multiple convolutional filters that sweep through the whole time series and extract short-term features from the original input. The kernel size of each filter is $h \times w$, where h decides the range of short-term features and w is equal to N_f . After a rectified linear unit (*ReLU*) activation, the output has a size of $T \times N_k$, where N_k is the number of filters.

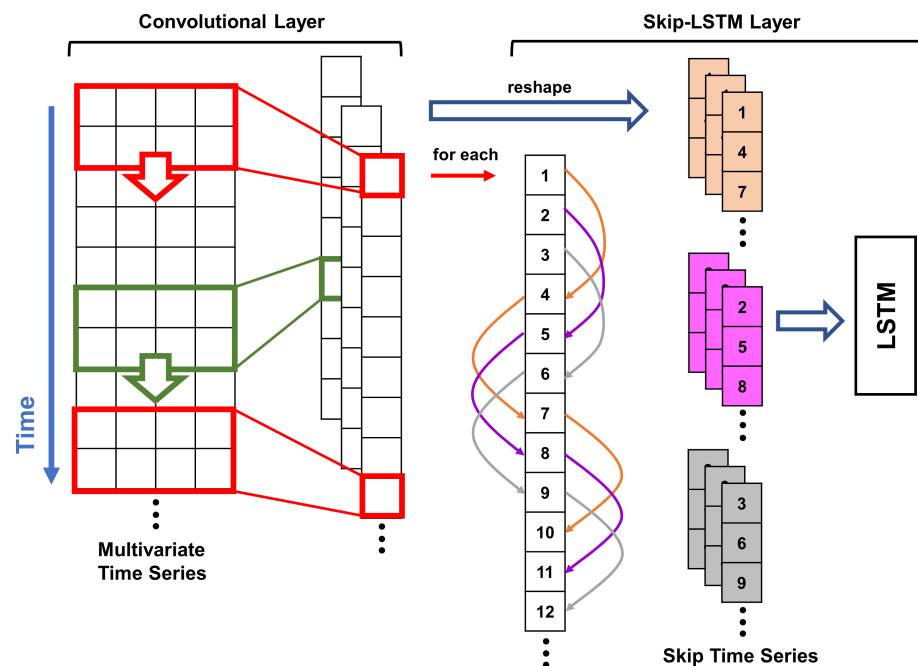


Figure 5. The structure of Skip-LSTM. The model consists of two layers: convolutional layer to fuse local temporal features and Skip-LSTM layer to extract large-scale temporal dependencies.

Further, a Skip-LSTM layer is applied to the product of the convolutional layer. As shown in Figure 5, the skip-link structure of Skip-LSTM can jump over multiple timestamps, thus shortening the length of the time series. Following the skip-link, the raw time series are converted into a group of skip time series. Within each skip time series, non-adjacent timestamps get appended together. The total number of converted series depends on the skip-factor s . Then the new grouped time series are fed into the backbone LSTM module as the input. By means of this “dilated receptive field”, long-term periodic patterns are captured in the temporal domain at a larger scale.

3.4. Cross-Pseudo Regression

A novel Cross-Pseudo Regression (CPR) strategy [45] is further converted to a wall-to-wall regression task to allow training the model in a semi-supervised way, thus compensating for the possible lack of training data. Both labeled and unlabeled data are included in the strategy. The model is constructed using two branches with the same structure but initialised differently, as shown in the flowchart of Figure 6.

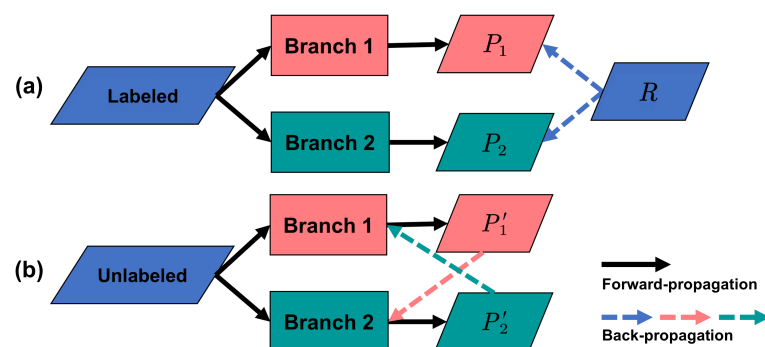


Figure 6. Cross-Pseudo Regression strategy. (a) The supervised step; (b) the unsupervised step.

Firstly, we train the two branches separately in a normal supervised way, as shown in Figure 6a; only labeled data are taken into the training. Specifically, we select mean squared

error (MSE) to measure the distance between the predictions and reference, as mathematically described in the following equation:

$$\text{Loss}(\hat{Y}, Y) = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2, \quad (10)$$

where \hat{Y} denotes the prediction and Y denotes the reference; n is the total number of the samples. So the supervised-loss can be denoted as

$$\ell_s = \text{Loss}(P_1, R) + \text{Loss}(P_2, R), \quad (11)$$

where P_1 and P_2 denote the predictions of Branches #1 and #2, respectively, and R represents the supervised reference.

Secondly, we ignore the label information and treat all the data as unlabeled, as illustrated in Figure 6b. The results predicted by one branch can be naturally treated as pseudolabels of the other, instead of the real reference. The backpropagation process can still be carried out according to the cross-pseudo-loss, which is defined as

$$\ell_c = \text{Loss}(P'_2, P'_1) + \text{Loss}(P'_1, P'_2), \quad (12)$$

where P'_1 and P'_2 denote the pseudolabels of Branches #1 and #2.

From the viewpoint of consistency regularisation [46,47], different initialisations of the branches would bring perturbations to the model. When fed the same input, both branches are encouraged to predict the same results, even though the perturbation is imposed. By minimising the cross-pseudo-loss, the discrepancy between predictions of both branches would also be minimised. In this way, a more compatible and representative feature space is learned by the model.

Finally, we combine supervised-loss and cross-pseudo-loss together. The combined loss of CPR can be defined as

$$\ell = \ell_s + \lambda_c \ell_c + \lambda_w \frac{1}{n_w} \sum_{j=1}^{n_w} (w_j)^2, \quad (13)$$

where λ_c controls the contribution of cross-pseudo-loss, which is simply set to 0.5 in our study. The l_2 regularisation item is added to the final loss to alleviate the overfitting, which is also known as the weight decay. w_j denotes the j -th weight in the model, n_w is the total number of weights and λ_w decides the trade-off. On the prediction stage, we simply select one of the trained branches, transfer all its weights to the task model and give the final regression of the forest height.

3.5. Overall Structure of CrsHelix-LSTM Model

We refer to the proposed Cross Helix LSTM method as CrsHelix-LSTM for short. The overall architecture of the model is shown in Figure 7.

The model consists of two branches based on CPR strategy. In each branch, the helix time attributes are firstly mapped and attached to the input data. Then, the input is fed into two separate components to obtain comprehensive features in the temporal domain. Basic LSTM is used to extract the overall invariant features of the time series. A convolutional layer is utilised to extract short-term features, followed by a Skip-LSTM layer to capture the seasonal pattern at a larger scope. At last, the feature maps from both sides are stacked together. The regression header finally projects the features to the forest height.

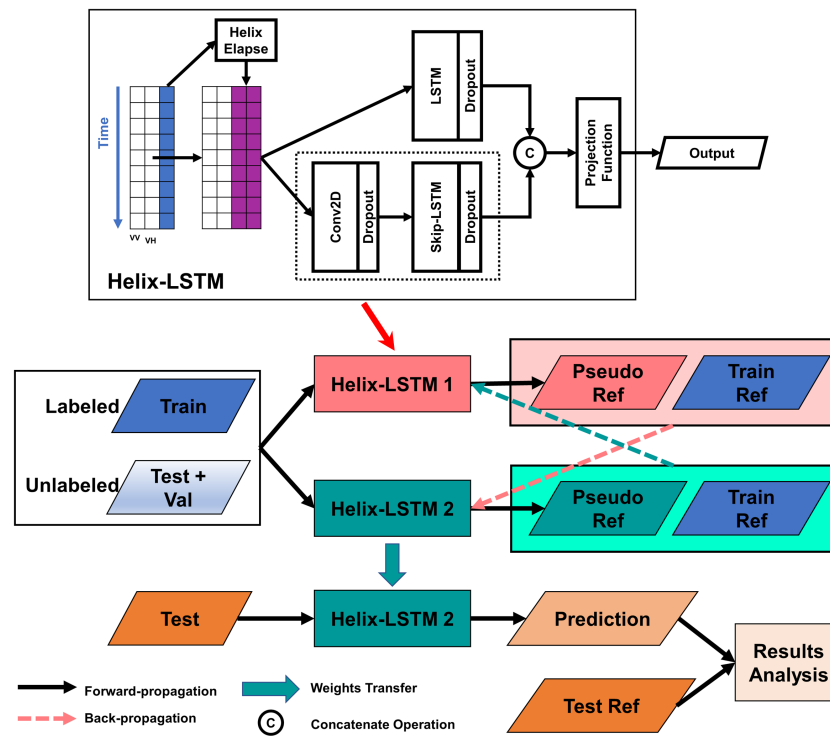


Figure 7. The architecture of the proposed CrsHelix-LSTM model.

The hidden sizes of LSTM and Skip-LSTM are set to 128 and for convolutional layer to 64. The kernel size of the CNN module is 5 and the skip-factor is set to 12. The total time step is 96, equivalent to the length of Sentinel-1 backscatter time series. To avoid potential overfitting, *Dropout* layers are implemented before the regression header. The *Dropout* coefficient is set to 0.5. At the training stage, the *Adam* optimisation algorithm is used to minimise the loss function. The batch size is 2048. The *OneCycleLR* learning rate strategy takes care of the training progress in cases of overfitting [48]. The training procedure is summarised in Algorithm 1.

Algorithm 1: Training Procedure of CrsHelix-LSTM

Input: The semi-labeled training time series $X = X_l \cup X_u$; The reference R of labeled subset X_l ;

Input: Iteration number K ;

- 1: Differently initialise the parameters of both Helix-LSTM branches θ_1, θ_2 .
 - 2: **for** $iteration = 0, \dots, K$ **do**
 - 3: **for** branch $i = 1, 2$ **do**
 - 4: Feed X into the branch;
 - 5: Extract features using Helix-LSTM, whose architecture is shown in Figure 7;
 - 6: Predict the inferences P_i for labeled subset X_l , \hat{P}_i for unlabeled X_u , respectively.
 - 7: **end for**
 - 8: Calculate the supervised-loss ℓ_s according to Equation (11):

$$\ell_s = \text{Loss}(P_1, R) + \text{Loss}(P_2, R).$$
 - 9: Let $P'_i = P_i \cup \hat{P}_i$, calculate the cross-pseudo-loss ℓ_c according to Equation (12):

$$\ell_c = \text{Loss}(P'_2, P'_1) + \text{Loss}(P'_1, P'_2).$$
 - 10: Summarise the combined loss of CPR according to Equation (13):

$$\ell = \ell_s + \lambda_c \ell_c + \lambda_w \frac{1}{n_w} \sum_{j=1}^{n_w} (w_j)^2.$$
 - 11: Back-propagate ℓ , update θ_1, θ_2 using gradient descent.
 - 12: **end for**
 - 13: **return** θ_1 (or θ_2) as the final model for testing.
-

3.6. Baseline Models

To access added value of developed models, several more traditional and widely used regression methods were used as baselines for comparison in our study. It should be noted that, similar to our models, all these methods operate on a pixel-level considering only temporal features. The following methodologies are included for comparison:

- **MLR, RF, LightGBM**
Multiple Linear Regression (MLR) and Random Forest (RF) are mature regression methods, which have been widely used in forest remote sensing tasks. Light Gradient Boosting Machine (LightGBM) [49] is a modern Gradient Boosting Decision Tree (GBDT) model. Since being proposed by Microsoft in 2017, its precision and efficiency in regression have been proven in different application areas [50,51].
- **LSTM, Attn-BiLSTM**
Since the proposed model is an improved version of LSTM, basic LSTM and its variant, Bidirectional LSTM with attention mechanism (Attn-BiLSTM) [33,52], are also included as baseline models for comparison. Bidirectional LSTM (BiLSTM) consists of two LSTMs with the same structure but opposite directions. Temporal dependencies are obtained from both directions. Furthermore, with the self-attention mechanism, attention weights establish the correlations between timestamps, which reportedly can better address the gradient vanishing problem and obtain long-term correlations [53]. Attn-BiLSTM combines both features and has been introduced into SAR remote sensing tasks [33].

To decrease the number of independent variables in MLR, a principal component analysis (PCA) module was applied before the modeling. This dimensionality reduction was not necessary for other methods, as they have built-in feature selection modules. All the models were fine-tuned with Optuna using 5-fold cross-validation. A 10% random sample of the training dataset was used in the fine-tuning.

3.7. Method Implementation

The overall methodology utilising the developed CrsHelix-LSTM model (as well as other benchmark models for comparison purposes) is illustrated by the flowchart shown in Figure 8. Sentinel-1 images are preprocessed as described in Section 2.2. The forest stand mask is applied to limit training and prediction to forested areas. ALS-based forest height reference data are used in model training and accuracy assessment. The training, validation and testing areas do not overlap after splitting is performed as described in Section 2.2. Importantly, even though Sentinel-1 images from the testing subset area is supplied into model training, those data are used as unlabeled for semi-supervised training purposes. Forest tree height predictions are compared to the reference data from testing areas on both pixel- and stand-levels.

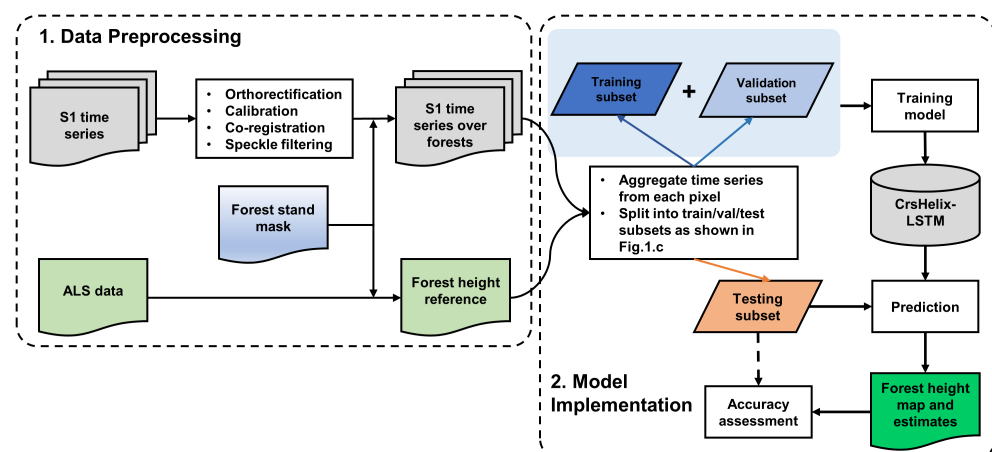


Figure 8. Flowchart of the proposed forest mapping methodology.

3.8. Model Performance Accuracy Assessment

The prediction accuracy of various regression models was calculated using the following accuracy metrics, including root mean squared error (RMSE), relative root mean squared error (rRMSE), the coefficient of determination (R^2), mean absolute error (MAE) and the index of agreement (IOA) [54]:

$$\text{RMSE} = \sqrt{\frac{\sum_i (y_i - \hat{y}_i)^2}{n}}, \quad (14)$$

$$\text{rRMSE} = \frac{\text{RMSE}}{\bar{y}} \cdot 100\%, \quad (15)$$

$$R^2 = 1 - \frac{SS_{\text{res}}}{SS_{\text{tot}}} = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y})^2}, \quad (16)$$

$$\text{MAE} = \frac{\sum_i |y_i - \hat{y}_i|}{n}, \quad (17)$$

$$\text{IOA} = \left[1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (|y_i - \bar{y}| + |\hat{y}_i - \bar{y}|)^2} \right] \cdot 100\%, \quad (18)$$

where y_i is the reference forest height of pixel i , \hat{y}_i is the predicted forest height, $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$ and n is the total number of samples. Stand-level estimates of forest height were calculated using spatial averaging at the extent of each stand (available from forest stand mask), with reported accuracies calculated using exactly the same Equations (14)–(18) for aggregated stand-level units. It is important to keep in mind that forest measurements for both training and accuracy assessment data were conducted in 2015, while SAR time series span more than three years. However, we primarily focus on relative performance of various prediction methods and exactly the same SAR and reference data are used for all benchmarked methods.

4. Experimental Results

4.1. Experimental Settings

The experiments were performed using Windows Server with Intel Xeon E5-2697 v4 CPU and NVIDIA GTX3060 GPU accelerated by CUDA 11.3 toolkit. LSTM, Attn-BiLSTM and the proposed model were built with a neural network library, Pytorch 1.11.0. MLR and RF were implemented with Scikit-learn machine learning toolbox. LightGBM was implemented with LightGBM Python-package provided by Microsoft.

4.2. General Performance Evaluation

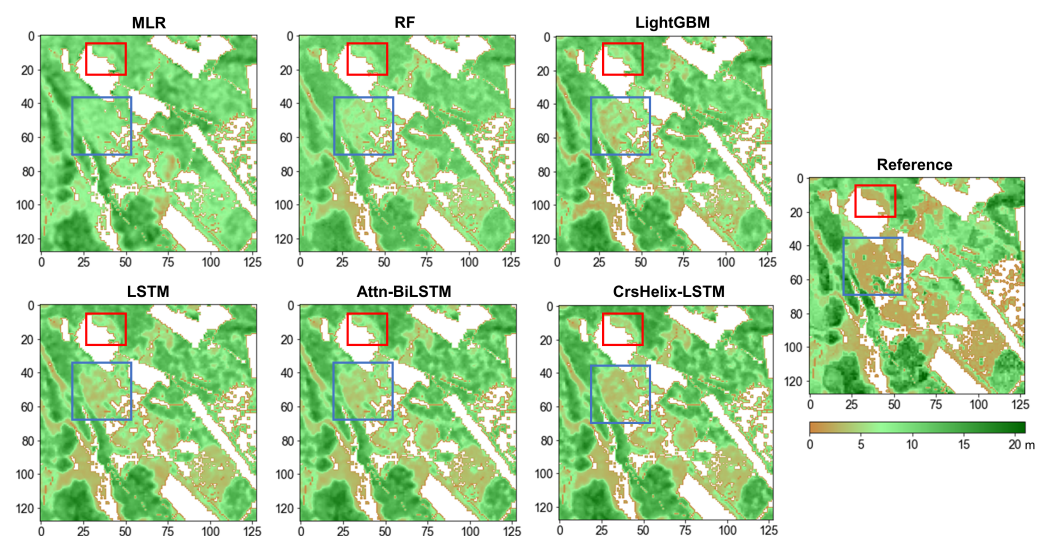
We examined the performance of the developed model versus several other LSTM-based models, as well as a representative set of benchmarking models often used in forest mapping, on both pixel and stand levels. The obtained ablation results for LSTM based models are gathered in Table 1 and the accuracy metrics for benchmarking models are shown in Table 2. Examples of forest maps produced by various examined approaches are shown in Figure 9. Prediction performances are gathered and illustrated with scatterplots between predicted and reference forest heights in Figure 10.

Table 1. Ablation Study. *LE* stands for Linear-Elapse projection, *HE* for Helix-Elapse projection, *Skip* for Skip-LSTM block and *CPR* for CPR strategy.

		<i>HE</i>	<i>Skip</i>	<i>CPR</i>	RMSE (m)	rRMSE (%)	R ²	MAE (m)	IOA (%)
Pixel-level	LSTM				3.26	29.16	0.49	2.51	80.53
	Attn-BiLSTM				3.22	28.87	0.50	2.48	81.11
	LSTM+LE	<i>LE</i>			3.26	29.16	0.49	2.52	80.28
	LSTM+HE	✓			3.22	28.88	0.50	2.49	80.99
	Helix-LSTM	✓	✓		3.19	28.59	0.51	2.46	81.76
	CrsHelix-LSTM	✓	✓	✓	3.16	28.31	0.52	2.42	82.46
Stand-level	LSTM				2.11	18.95	0.62	1.64	86.02
	Attn-BiLSTM				2.08	18.63	0.63	1.61	86.62
	LSTM+LE	<i>LE</i>			2.12	18.97	0.62	1.64	85.79
	LSTM+HE	✓			2.09	18.69	0.63	1.62	86.47
	Helix-LSTM	✓	✓		2.04	18.26	0.64	1.58	87.32
	CrsHelix-LSTM	✓	✓	✓	2.01	18.01	0.65	1.55	87.96

Table 2. Experimental results compared to benchmarks.

		RMSE (m)	rRMSE (%)	R ²	MAE (m)	IOA (%)
Pixel-level	MLR	3.50	31.38	0.40	2.74	75.40
	RF	3.45	30.91	0.42	2.74	72.73
	LightGBM	3.30	29.55	0.47	2.56	79.26
	LSTM	3.26	29.16	0.49	2.51	80.53
	Attn-BiLSTM	3.22	28.87	0.50	2.48	81.11
	CrsHelix-LSTM	3.16	28.31	0.52	2.42	82.46
Stand-level	MLR	2.37	21.22	0.52	1.85	81.21
	RF	2.38	21.36	0.51	1.90	77.68
	LightGBM	2.17	19.42	0.60	1.69	84.72
	LSTM	2.11	18.95	0.62	1.64	86.02
	Attn-BiLSTM	2.08	18.63	0.63	1.61	86.62
	CrsHelix-LSTM	2.01	18.01	0.65	1.55	87.96

**Figure 9.** Examples of forest height maps produced by various examined regression methods. Typical areas with significant improvements are highlighted with red and blue polygons.

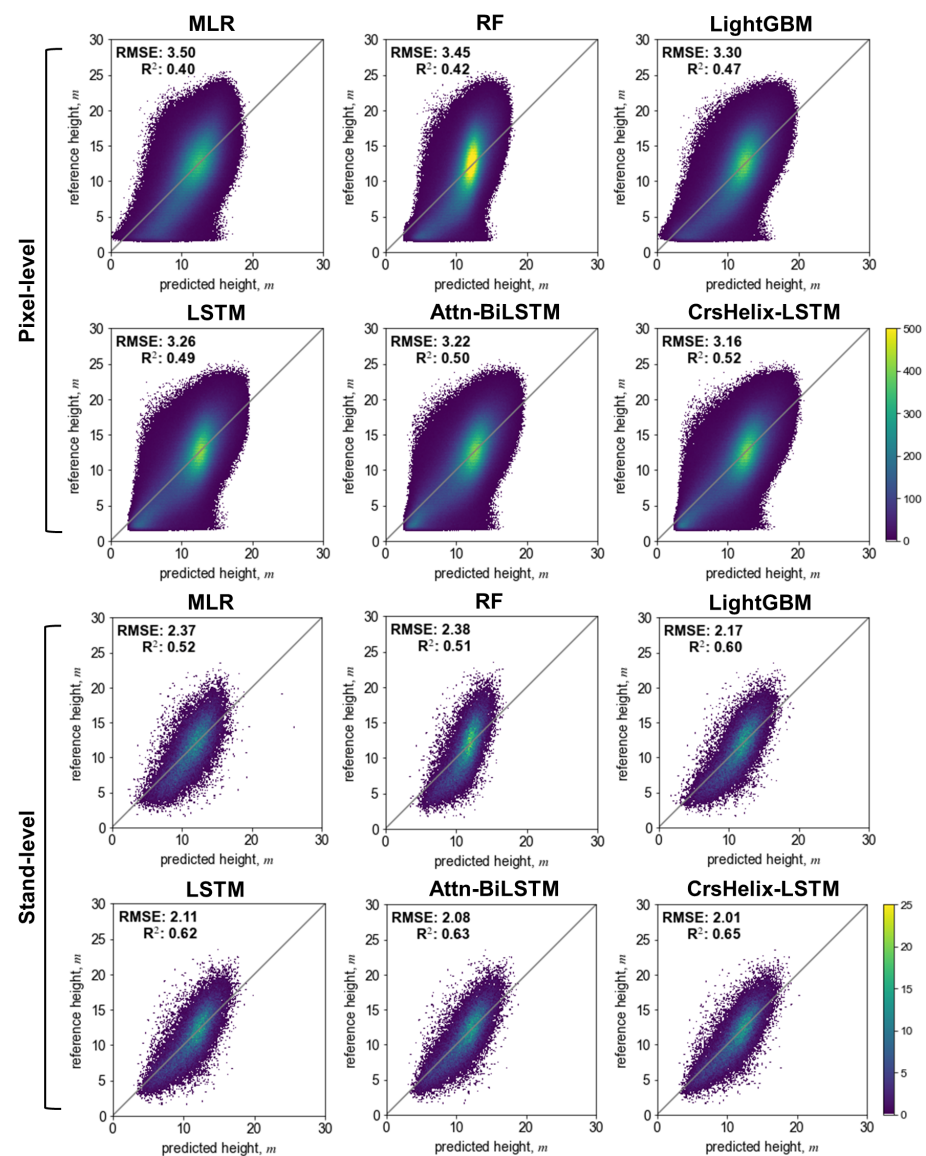


Figure 10. Pixel- (upper rows) and stand level (bottom rows) scatterplots for various studied regression methods: MLR, RF, LGBM, LSTM, Attn-BiLSTM and CrsHelix-LSTM.

4.2.1. Ablation Study

Firstly, we verified the effectiveness of different blocks in the ablation study, the results are shown in Table 1. In general, the forest height prediction at the stand level demonstrates a larger accuracy compared to the pixel level, taking advantage of averaging within the homogeneous stands. When we handled the irregular time intervals by using Linear-Elapse (LE) projection, the LE attribute was stacked onto the input as a new feature. The regression results were not improved compared to basic LSTM. This indicates the elapsing of time is not properly modeled in this case. On the contrary, by simply substituting the LE projection with our HE projection, RMSE is somewhat improved from 3.26 m to 3.22 m at the pixel level. The results are close to Attn-BiLSTM, which is considered a much stronger baseline when it comes to long time series tasks. This indicates that HE projection can better model the annual dynamics and help establish relationships between forest variables and seasonal patterns, thereby approximating the role of the attention mechanism at a lower computational cost.

When the Skip-LSTM block was embedded in the model, the regression performance of Helix-LSTM further improved at both the pixel and the stand level. The rRMSE is 28.59%, a 0.28% improvement from Attn-BiLSTM for the pixel level and 0.37% for the

stand level. Considering Attn-BiLSTM is only better than basic LSTM for 0.29% and 0.32%, the improvement is quantitatively considerable. Finally, by wrapping up two Helix-LSTM branches with the CPR strategy, CrsHelix-LSTM obtained the best regression performance in the ablation study. The best rRMSE is 28.31% for the pixel level and 18.01% for the stand level, the best MAE is 1.55 m. IOA also obtains the optimal in this case as 87.96% and R^2 of the final model is as high as 0.65. Note that its rRMSE is 0.28% further decreased compared to Helix-LSTM, even though the backbone models were the same. This indicates that a positive impact is imposed by the semi-supervised learning strategy, the model learns the forest representations from not only the labeled but also the unlabeled data.

4.2.2. Method Performance Comparison with Baseline Approaches

In Table 2, we also compare our methods to some existing machine learning models. MLR, which is most widely used in forest mapping, presents the most fundamental performance, which is 3.50 m of RMSE for the pixel level and 2.37 m for the stand level. Other approaches generally perform better than MLR; among them, the use of LSTM somewhat improves the prediction performance at both pixel and stand levels. While the developed CrsHelix-LSTM approach provided consistently better results compared to other benchmarking models. Its rRMSE for the pixel level reaches to 28.31%, 1.24% decreased compared to LightGBM. Note that the rRMSE of LightGBM is only 1.83% decreased compared to MLR. From this aspect, the improvement of our model is considerable. This is also confirmed by MAE, which is less sensitive to outliers. MAEs of CrsHelix-LSTM are 0.32 m improved for pixel level and 0.30 m for stand level compared to MLR.

By visualising the prediction results, as shown in Figure 9, we can better observe the prediction discrepancy over forests at different height levels. Compared to other methods, CrsHelix-LSTM is more sensitive to undergrowth forests whose height is within 5 m. The corresponding areas are showing as light brown according to the colormap, typical areas are highlighted with polygons in Figure 9. Similar observations can also be made from the analyses of the scatter plots shown in Figure 10, particularly on stand level. Compared to other benchmarking methods, the samples predicted by our CrsHelix-LSTM are more inclined to follow the diagonal line. Taller referenced samples around 20 m are less biased, which indicates a better prediction performance for taller forest stands.

5. Discussion

5.1. Computational Complexity Analysis

For computational complexity analysis, when using the floating point operations (FLOPs) as the evaluation criteria [55], for a dummy input tensor with dimensions $10,000 \times 96 \times 4$, the FLOPs of LSTM, Attn-BiLSTM and CrsHelix-LSTM are 66.85 G, 133.70 G and 312.68 G, respectively. The FLOPs of CrsHelix-LSTM are approximately five times as large as LSTM, due to its dual branches and unique architecture.

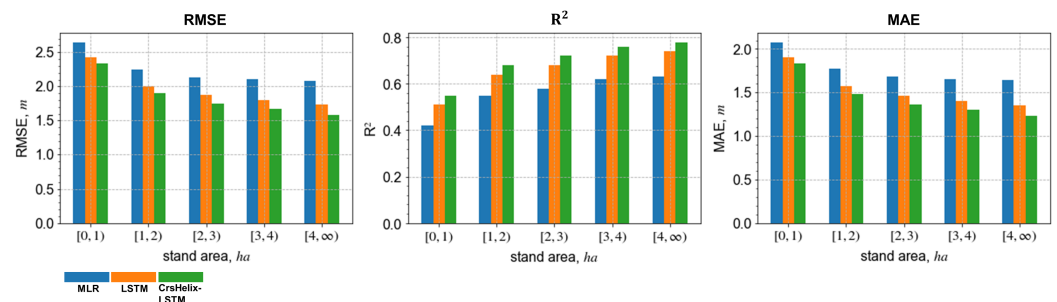
However, such analysis is not applicable to conventional baselines like MLR or RF, as those methods usually utilise CPU while deep learning models are running in parallel accelerated by GPU and CUDA. With this in mind, we compared the processing times of different methods directly, including both training and testing times, as presented in Table 3. Although the prediction performance of MLR is modest, it uses the least amount of time for training, only 37.39 s. RF and LightGBM are next. It is worth noting that their computation complexity may vary strongly depending on the hyperparameters. For example, RF with more decision trees, node leaves and bigger depth would take considerably more time. All LSTM-based methods are time consuming. After the same number of epochs, the training time is generally consistent with their FLOPs. The proposed CrsHelix-LSTM takes 4182.08 s for training purposes. However, there is no need to retrain the model from scratch every time the mapping is needed. By using an already trained CrsHelix-LSTM model, the testing time is only 72 s, which is even less than basic LSTM.

Table 3. Processing time of the proposed and baseline methods.

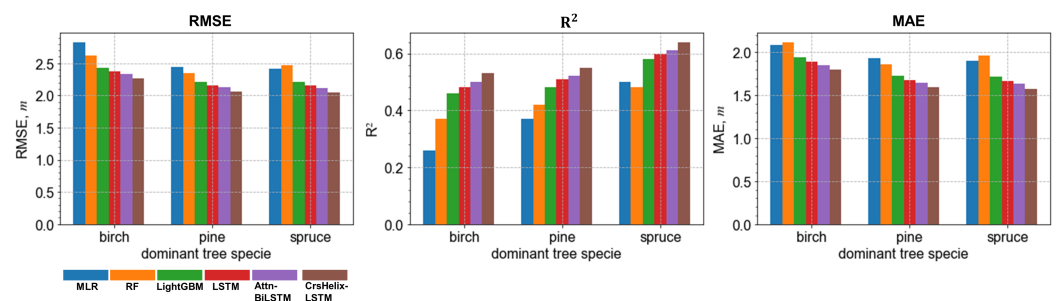
	MLR	RF	LightGBM	LSTM	Attn-BiLSTM	CrsHelix-LSTM
Training time, s	37.39	485.89	70.17	886.58	1326.10	4182.08
Testing time, s	0.036	33.21	40.17	80.13	66.23	72.05

5.2. Stand-Level Performance against Different Stand Sizes and Dominant Tree Species

Since stand-level averaging within the homogeneous stands can help improve the prediction performance, in this section we firstly analyse the effects of different stand sizes on stand-level prediction. As shown in Figure 11, we divide all stands into five subsets according to their size. There are 13,197, 7474, 4639, 2839 and 4912 stands for each area range, from less than 1 ha to over 4 ha. In general, the prediction accuracy improves with increasing stand size. However, the improvement becomes minor when the stand size is large enough. Such saturation is less pronounced with our proposed model. For each subset, CrsHelix-LSTM demonstrates the most accurate prediction accuracy, varying from 1.58 m RMSE for stands that are larger than 4 ha (R^2 of 0.78 and MAE of 1.23 m) to 2.34 m RMSE for smaller stands under 1 ha.

**Figure 11.** Stand-level performance against different stand area ranges.

Another factor that influences the stand-level prediction performance is the dominant tree species. In total, there are 1542 birch, 6701 pine and 12,194 spruce dominated stands in the testing area. The prediction performance for each strata is presented in Figure 12. Among the three studied tree species, tree height predictions for birch stands are somewhat worse compared to other species, while predictions of spruce dominated stands are the most accurate. Among all the methods, CrsHelix-LSTM still shows the best prediction accuracy for all studied tree species. It can be more clearly observed using R^2 as an evaluation criterion.

**Figure 12.** Stand-level performance against different forest species.

5.3. Comparison with Other Studies and Outlook

Observed experimental results are encouraging further investigations and are generally in line with other reported studies in boreal forest biome [3,8,56,57]. The obtained accuracies are notably higher than several other studies, in which Sentinel-1 or Sentinel-2 datasets or their combinations were used and compared well versus earlier multisensor EO data studies [3,10,21,56–59]. Several datasets with high potential for forest variable

retrieval, e.g., TanDEM-X, have relatively sparse coverage (both in geographic and temporal domains) and limited availability and thus are not fully suitable for large-area mapping and persistent monitoring purposes.

When using Sentinel-1 or Landsat data to study boreal forests, existing studies report their variable prediction accuracies within the range of 35–60% rRMSE [3,56], while the proposed model utilising Sentinel-1 time series data obtains an rRMSE as small as 18%. Our predictions obtained using traditional machine learning (ML) models were within the same accuracy range as in recently published studies where Sentinel-2 and Landsat data were used [60], while our predictions using different versions of LSTM models appear more accurate. The usage of attention mechanism and Helix attribute in Attn-BiLSTM and the proposed CrsHelix-LSTM are instrumental in providing larger accuracies compared to the basic LSTM model.

There is relatively limited literature using SAR data for forest height predictions, with predictions often reported on stand level or coarser resolution spatial units [8]. However, our stand-level tree height predictions were at the same accuracy level or even better than reported retrievals with TanDEM-X interferometric SAR data, even though TanDEM-X is deemed much more suitable for retrieving vertical forest structure [61–63]. To the best of our knowledge, the obtained accuracy level of 18% RMSE for boreal forest using CrsHelix-LSTM and SAR time series is superior to earlier results reported in the literature [8,56,60].

We expect that LSTM-based models and particularly CrsHelix-LSTM can be used to provide better predictions with Sentinel-1 time series (compared to conventional machine learning methods) also in other forest biomes where seasonal trends with C-band SAR data or irregular sampled image data are present, e.g., in hemiboreal or temperate forests.

6. Conclusions

Our study demonstrated the potential for applying LSTM approaches with Helix attribute for predicting forest structural variables such as forest height. Introducing temporal context into prediction models even using basic LSTM model has indicated improvement of prediction performance compared to scenarios when each SAR observation is treated independently (benchmarked MLR, RF and LightGBM models).

Furthermore, a novel LSTM model incorporating Helix-Elapse projection, Skip-LSTM and Cross-Pseudo Regression has been developed and tested in the study. The developed model was demonstrated using long time series of Sentinel-1 data but can be applicable to other SAR time series. The CrsHelix-LSTM model provided larger accuracies in boreal forest height mapping in Central Finland compared to other evaluated LSTM approaches and a set of representative machine learning approaches often used in forest mapping. At best, the achieved accuracy of forest height mapping was 28.3% rRMSE for the pixel level predictions and 18.0% rRMSE on the forest stand level.

Obtained results are generally much better than those reported in the literature with SAR datasets. Superior performance can be explained by better capturing temporal dependencies within SAR time series with skip-link based LSTM blocks. Better performance of CrsHelix-LSTM compared to other studied LSTM models indicates that seasonal patterns and irregular data can be better modeled by leveraging the Helix attribute. Additionally, we suggest taking advantage of semisupervised learning, such as CPR incorporated into the CrsHelix-LSTM model, to improve prediction performance with limited reference data.

Future work will concentrate on introducing other datasets particularly suitable for retrieving vertical structure of forests, such as Sentinel-1 interferometric SAR and TanDEM-X datasets, as well as studying other forest variables, such as growing stock volume and above-ground biomass.

Author Contributions: Conceptualisation, S.G. and O.A.; methodology, S.G. and O.A.; validation O.A.; data curation, Y.R. and O.A.; writing—original draft preparation, S.G. and O.A.; writing—review and editing, H.G., O.A. and Y.R.; supervision, O.A., W.S. and J.P. All authors have read and agreed to the published version of the manuscript.

Funding: This study was supported by the National Natural Science Foundation of China (Grant No. 62001229, 62101264, 62101260) and by China Postdoctoral Science Foundation (Grant No. 2020M681604). O.A. was supported by Multico project funded by Business Finland and Forest Carbon Monitoring project funded by European Space Agency.

Data Availability Statement: Preprocessed EO images, reference data and produced maps presented in this study are available from the corresponding author upon request.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

ALS	airborne laser scanning
Attn-BiLSTM	Bidirectional LSTM with attention mechanism
BiLSTM	Bidirectional LSTM
CNN	convolutional neural network
CPR	Cross-Pseudo Regression
CrsHelix-LSTM	Cross Helix LSTM
EO	earth observation
ESA	European Space Agency
FLOPs	floating point operations
GBDT	Gradient Boosting Decision Tree
GRD	ground range detected
HE	Helix-Elapse
IOA	index of agreement
LE	Linear-Elapse
LightGBM	Light Gradient Boosting Machine
LSTM	Long Short-Term Memory
MAE	mean absolute error
ML	machine learning
MLR	Multiple Linear Regression
MSE	mean squared error
PCA	principal component analysis
ReLU	rectified linear unit
RF	Random Forest
RMSE	root mean squared error
RNN	recurrent neural network
rRMSE	relative root mean squared error
SAR	synthetic aperture radar

References

1. Herold, M.; Carter, S.; Avitabile, V.; Espejo, A.B.; Jonckheere, I.; Lucas, R.; McRoberts, R.E.; Næsset, E.; Nightingale, J.; Petersen, R.; et al. The role and need for space-based forest biomass-related measurements in environmental management and policy. *Surv. Geophys.* **2019**, *40*, 757–778. [[CrossRef](#)]
2. McRoberts, R.E.; Tomppo, E.O. Remote sensing support for national forest inventories. *Remote Sens. Environ.* **2007**, *110*, 412–419. [[CrossRef](#)]
3. Miettinen, J.; Carlier, S.; Häme, L.; Mäkelä, A.; Minunno, F.; Penttilä, J.; Pisl, J.; Rasinmäki, J.; Rauste, Y.; Seitsonen, L.; et al. Demonstration of large area forest volume and primary production estimation approach based on Sentinel-2 imagery and process based ecosystem modelling. *Int. J. Remote Sens.* **2021**, *42*, 9467–9489. [[CrossRef](#)]
4. Tomppo, E.; Olsson, H.; Ståhl, G.; Nilsson, M.; Hagner, O.; Katila, M. Combining national forest inventory field plots and remote sensing data for forest databases. *Remote Sens. Environ.* **2008**, *112*, 1982–1999. [[CrossRef](#)]
5. White, J.C.; Coops, N.C.; Wulder, M.A.; Vastaranta, M.; Hilker, T.; Tompalski, P. Remote sensing technologies for enhancing forest inventories: A review. *Can. J. Remote Sens.* **2016**, *42*, 619–641. [[CrossRef](#)]
6. Le Toan, T.; Beaudoin, A.; Riou, J.; Guyon, D. Relating forest biomass to SAR data. *IEEE Trans. Geosci. Remote Sens.* **1992**, *30*, 403–411. [[CrossRef](#)]
7. Imhoff, M.L. Radar backscatter and biomass saturation: Ramifications for global biomass inventory. *IEEE Trans. Geosci. Remote Sens.* **1995**, *33*, 511–518. [[CrossRef](#)]

8. GFOI. *Integrating Remote-Sensing and Ground-Based Observations For Estimation of Emissions and Removals of Greenhouse Gases in Forests: Methods and Guidance From the Global Forest Observations Initiative*; Group on Earth Observations: Geneva, Switzerland, 2014.
9. Schmullius, C.; Thiel, C.; Pathe, C.; Santoro, M. Radar time series for land cover and forest mapping. In *Remote Sensing Time Series*; Springer: Berlin, Germany, 2015; pp. 323–356. [\[CrossRef\]](#)
10. Tomppo, E.; Antropov, O.; Praks, J. Boreal forest snow damage mapping using multi-temporal Sentinel-1 data. *Remote Sens.* **2019**, *11*, 384. [\[CrossRef\]](#)
11. Torres, R.; Snoeijs, P.; Geudtner, D.; Bibby, D.; Davidson, M.; Attema, E.; Potin, P.; Rommen, B.; Floury, N.; Brown, M.; et al. GMES Sentinel-1 mission. *Remote Sens. Environ.* **2012**, *120*, 9–24. [\[CrossRef\]](#)
12. Thiel, C.; Cartus, O.; Eckardt, R.; Richter, N.; Thiel, C.; Schmullius, C. Analysis of multi-temporal land observation at C-band. In *Proceedings of the 2009 IEEE International Geoscience and Remote Sensing Symposium*, Cape Town, South Africa, 12–17 July 2009; IEEE: Piscataway, NJ, USA, 2009; Volume 3, p. III-318. [\[CrossRef\]](#)
13. Antropov, O.; Rauste, Y.; Väänänen, A.; Mutanen, T.; Häme, T. Mapping forest disturbance using long time series of Sentinel-1 data: Case studies over boreal and tropical forests. In *Proceedings of the 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, Beijing, China, 10–15 July 2016; IEEE: Piscataway, NJ, USA, 2016; pp. 3906–3909. [\[CrossRef\]](#)
14. Laurin, G.V.; Balling, J.; Corona, P.; Mattioli, W.; Papale, D.; Puletti, N.; Rizzo, M.; Truckenbrodt, J.; Urban, M. Above-ground biomass prediction by Sentinel-1 multitemporal data in central Italy with integration of ALOS2 and Sentinel-2 data. *J. Appl. Remote Sens.* **2018**, *12*, 016008. [\[CrossRef\]](#)
15. Stelmaszczyk-Górska, M.A.; Urbazhev, M.; Schmullius, C.; Thiel, C. Estimation of Above-Ground Biomass over Boreal Forests in Siberia Using Updated In Situ, ALOS-2 PALSAR-2, and RADARSAT-2 Data. *Remote Sens.* **2018**, *10*, 1550. [\[CrossRef\]](#)
16. Antropov, O.; Rauste, Y.; Praks, J.; Seifert, F.M.; Häme, T. Mapping forest disturbance due to selective logging in the Congo Basin with RADARSAT-2 time series. *Remote Sens.* **2021**, *13*, 740. [\[CrossRef\]](#)
17. Tomppo, E.; Ronoud, G.; Antropov, O.; Hytönen, H.; Praks, J. Detection of forest windstorm damages with multitemporal sar data—A case study: Finland. *Remote Sens.* **2021**, *13*, 383. [\[CrossRef\]](#)
18. Rüetschi, M.; Small, D.; Waser, L.T. Rapid detection of windthrows using Sentinel-1 C-band SAR data. *Remote Sens.* **2019**, *11*, 115. [\[CrossRef\]](#)
19. Hoekman, D.; Kooij, B.; Quiñones, M.; Vellekoop, S.; Carolita, I.; Budhiman, S.; Arief, R.; Roswintarti, O. Wide-area near-real-time monitoring of tropical forest degradation and deforestation using Sentinel-1. *Remote Sens.* **2020**, *12*, 3263. [\[CrossRef\]](#)
20. Hethcoat, M.G.; Carreiras, J.M.; Edwards, D.P.; Bryant, R.G.; Quegan, S. Detecting tropical selective logging with C-band SAR data may require a time series approach. *Remote Sens. Environ.* **2021**, *259*, 112411. [\[CrossRef\]](#)
21. Ge, S.; Tomppo, E.; Rauste, Y.; McRoberts, R.E.; Praks, J.; Gu, H.; Su, W.; Antropov, O. Using hypertemporal Sentinel-1 data to predict forest growing stock volume. *bioRxiv* **2021**. [\[CrossRef\]](#)
22. Santoro, M.; Beer, C.; Cartus, O.; Schmullius, C.; Shvidenko, A.; McCallum, I.; Wegmüller, U.; Wiesmann, A. Retrieval of growing stock volume in boreal forest using hyper-temporal series of Envisat ASAR ScanSAR backscatter measurements. *Remote Sens. Environ.* **2011**, *115*, 490–507. [\[CrossRef\]](#)
23. Dostálová, A.; Wagner, W.; Milenković, M.; Hollaus, M. Annual seasonality in Sentinel-1 signal for forest mapping and forest type classification. *Int. J. Remote Sens.* **2018**, *39*, 7738–7760. [\[CrossRef\]](#)
24. Pulliainen, J.; Mikhela, P.; Hallikainen, M.; Ikonen, J.P. Seasonal dynamics of C-band backscatter of boreal forests with applications to biomass and soil moisture estimation. *IEEE Trans. Geosci. Remote Sens.* **1996**, *34*, 758–770. [\[CrossRef\]](#)
25. Pulliainen, J.T.; Kurvonen, L.; Hallikainen, M.T. Multitemporal behavior of L- and C-band SAR observations of boreal forests. *IEEE Trans. Geosci. Remote Sens.* **1999**, *37*, 927–937. [\[CrossRef\]](#)
26. Ge, S.; Antropov, O.; Su, W.; Gu, H.; Praks, J. Deep recurrent neural networks for land-cover classification using Sentinel-1 InSAR time series. In *Proceedings of the IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium*, Yokohama, Japan, 28 July–2 August 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 473–476. [\[CrossRef\]](#)
27. Ienco, D.; Gaetano, R.; Dupaquier, C.; Maurel, P. Land cover classification via multitemporal spatial data by deep recurrent neural networks. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 1685–1689. [\[CrossRef\]](#)
28. Yuan, Y.; Lin, L.; Huo, L.Z.; Kong, Y.L.; Zhou, Z.G.; Wu, B.; Jia, Y. Using an attention-based LSTM encoder–decoder network for near real-time disturbance detection. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2020**, *13*, 1819–1832. [\[CrossRef\]](#)
29. Ma, L.; Liu, Y.; Zhang, X.; Ye, Y.; Yin, G.; Johnson, B.A. Deep learning in remote sensing applications: A meta-analysis and review. *ISPRS J. Photogramm. Remote Sens.* **2019**, *152*, 166–177. [\[CrossRef\]](#)
30. Xie, Y.; Huang, J. Integration of a Crop Growth Model and Deep Learning Methods to Improve Satellite-Based Yield Estimation of Winter Wheat in Henan Province, China. *Remote Sens.* **2021**, *13*, 4372. [\[CrossRef\]](#)
31. Hakim, W.L.; Nur, A.S.; Rezaie, F.; Panahi, M.; Lee, C.W.; Lee, S. Convolutional neural network and long short-term memory algorithms for groundwater potential mapping in Anseong, South Korea. *J. Hydrol. Reg. Stud.* **2022**, *39*, 100990. [\[CrossRef\]](#)
32. Lin, Z.; Zhong, R.; Xiong, X.; Guo, C.; Xu, J.; Zhu, Y.; Xu, J.; Ying, Y.; Ting, K.; Huang, J.; et al. Large-Scale Rice Mapping Using Multi-Task Spatiotemporal Deep Learning and Sentinel-1 SAR Time Series. *Remote Sens.* **2022**, *14*, 699. [\[CrossRef\]](#)
33. Sun, C.; Zhang, H.; Xu, L.; Wang, C.; Li, L. Rice Mapping Using a BiLSTM-Attention Model from Multitemporal Sentinel-1 Data. *Agriculture* **2021**, *11*, 977. [\[CrossRef\]](#)
34. Sun, C.; Zhang, H.; Ge, J.; Wang, C.; Li, L.; Xu, L. Rice Mapping in a Subtropical Hilly Region Based on Sentinel-1 Time Series Feature Analysis and the Dual Branch BiLSTM Model. *Remote Sens.* **2022**, *14*, 3213. [\[CrossRef\]](#)

35. Lai, G.; Chang, W.C.; Yang, Y.; Liu, H. Modeling long-and short-term temporal patterns with deep neural networks. In Proceedings of the 41st International ACM SIGIR Conference on Research & Development in Information Retrieval, Ann Arbor, MI, USA, 8–12 July 2018; pp. 95–104. [\[CrossRef\]](#)
36. Wang, Y.; Albrecht, C.; Braham, N.A.A.; Mou, L.; Zhu, X. Self-Supervised Learning in Remote Sensing: A Review. *IEEE Geoscience and Remote Sensing Magazine*, 5 September 2022. [\[CrossRef\]](#)
37. Wang, C.; Gu, H.; Su, W. SAR image classification using contrastive learning and pseudo-labels with limited data. *IEEE Geosci. Remote Sens. Lett.* **2021**, *19*, 1–5. [\[CrossRef\]](#)
38. Ge, S.; Gu, H.; Su, W.; Praks, J.; Antropov, O. Improved semisupervised unet deep learning model for forest height mapping with satellite sar and optical data. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2022**, *15*, 5776–5787. [\[CrossRef\]](#)
39. Rauste, Y.; Lonnqvist, A.; Molinier, M.; Henry, J.B.; Hame, T. Ortho-rectification and terrain correction of polarimetric SAR data applied in the ALOS/Palsar context. In Proceedings of the 2007 IEEE International Geoscience and Remote Sensing Symposium, Barcelona, Spain, 23–28 July 2007; IEEE: Piscataway, NJ, USA, 2007; pp. 1618–1621. [\[CrossRef\]](#)
40. Small, D. Flattening gamma: Radiometric terrain correction for SAR imagery. *IEEE Trans. Geosci. Remote Sens.* **2011**, *49*, 3081–3093. [\[CrossRef\]](#)
41. Graves, A.; Mohamed, A.-r.; Hinton, G. Speech recognition with deep recurrent neural networks. In Proceedings of the 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, Vancouver, BC, Canada, 26–31 May 2013; IEEE: Piscataway, NJ, USA, 2013; pp. 6645–6649. [\[CrossRef\]](#)
42. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. *Commun. ACM* **2017**, *60*, 84–90. [\[CrossRef\]](#)
43. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, *9*, 1735–1780. [\[CrossRef\]](#) [\[PubMed\]](#)
44. Yu, F.; Koltun, V. Multi-scale context aggregation by dilated convolutions. *arXiv* **2015**, arXiv:1511.07122. [\[CrossRef\]](#)
45. Chen, X.; Yuan, Y.; Zeng, G.; Wang, J. Semi-supervised semantic segmentation with cross pseudo supervision. In Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 20–25 June 2021; IEEE: Piscataway, NJ, USA, 2021; pp. 2613–2622. [\[CrossRef\]](#)
46. Bachman, P.; Alsharif, O.; Precup, D. Learning with pseudo-ensembles. *Adv. Neural Inf. Process. Syst.* **2014**, *27*. Available online: <https://proceedings.neurips.cc/paper/2014/hash/66be31e4c40d676991f2405aaecc6934-Abstract.html> (accessed on 28 September 2022).
47. Zhang, H.; Zhang, Z.; Odena, A.; Lee, H. Consistency regularization for generative adversarial networks. *arXiv* **2019**, arXiv:1910.12027. [\[CrossRef\]](#)
48. Smith, L.N.; Topin, N. Super-convergence: Very fast training of neural networks using large learning rates; Artificial intelligence and machine learning for multi-domain operations applications. In Proceedings of the SPIE Defense + Commercial Sensing, Baltimore, MD, USA, 14–18 April 2019; Volume 11006, pp. 369–386. [\[CrossRef\]](#)
49. Ke, G.; Meng, Q.; Finley, T.; Wang, T.; Chen, W.; Ma, W.; Ye, Q.; Liu, T.Y. Lightgbm: A highly efficient gradient boosting decision tree. *Adv. Neural Inf. Process. Syst.* **2017**, *30*. Available online: <https://proceedings.neurips.cc/paper/2017/hash/6449f44a102fde848669bdd9eb6b76fa-Abstract.html> (accessed on 28 September 2022).
50. Ju, Y.; Sun, G.; Chen, Q.; Zhang, M.; Zhu, H.; Rehman, M.U. A model combining convolutional neural network and LightGBM algorithm for ultra-short-term wind power forecasting. *IEEE Access* **2019**, *7*, 28309–28318. [\[CrossRef\]](#)
51. Sun, X.; Liu, M.; Sima, Z. A novel cryptocurrency price trend forecasting model based on LightGBM. *Financ. Res. Lett.* **2020**, *32*, 101084. [\[CrossRef\]](#)
52. Li, W.; Qi, F.; Tang, M.; Yu, Z. Bidirectional LSTM with self-attention mechanism and multi-channel features for sentiment classification. *Neurocomputing* **2020**, *387*, 63–77. [\[CrossRef\]](#)
53. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention is all you need. *Adv. Neural Inf. Process. Syst.* **2017**, *30*. Available online: <https://proceedings.neurips.cc/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html> (accessed on 28 September 2022).
54. Valbuena, R.; Hernando, A.; Manzanera, J.A.; Görgens, E.B.; Almeida, D.R.; Silva, C.A.; García-Abril, A. Evaluating observed versus predicted forest biomass: R-squared, index of agreement or maximal information coefficient? *Eur. J. Remote Sens.* **2019**, *52*, 345–358. [\[CrossRef\]](#)
55. Molchanov, P.; Tyree, S.; Karras, T.; Aila, T.; Kautz, J. Pruning convolutional neural networks for resource efficient inference. *arXiv* **2016**, arXiv:1611.06440. [\[CrossRef\]](#)
56. Astola, H.; Seitsonen, L.; Halme, E.; Molinier, M.; Lönnqvist, A. Deep neural networks with transfer learning for forest variable estimation using sentinel-2 imagery in boreal forest. *Remote Sens.* **2021**, *13*, 2392. [\[CrossRef\]](#)
57. Rees, W.G.; Tomaney, J.; Tutubalina, O.; Zharko, V.; Bartalev, S. Estimation of boreal forest growing stock volume in russia from sentinel-2 msi and land cover classification. *Remote Sens.* **2021**, *13*, 4483. [\[CrossRef\]](#)
58. Huang, W.; Min, W.; Ding, J.; Liu, Y.; Hu, Y.; Ni, W.; Shen, H. Forest height mapping using inventory and multi-source satellite data over Hunan Province in southern China. *For. Ecosyst.* **2022**, *9*, 100006. [\[CrossRef\]](#)
59. Lang, N.; Schindler, K.; Wegner, J.D. Country-wide high-resolution vegetation height mapping with Sentinel-2. *Remote Sens. Environ.* **2019**, *233*, 111347. [\[CrossRef\]](#)
60. Astola, H.; Häme, T.; Sirro, L.; Molinier, M.; Kilpi, J. Comparison of Sentinel-2 and Landsat 8 imagery for forest variable prediction in boreal region. *Remote Sens. Environ.* **2019**, *223*, 257–273. [\[CrossRef\]](#)

-
61. Praks, J.; Antropov, O.; Hallikainen, M.T. LIDAR-aided SAR interferometry studies in boreal forest: Scattering phase center and extinction coefficient at X-and L-band. *IEEE Trans. Geosci. Remote Sens.* **2012**, *50*, 3831–3843. [[CrossRef](#)]
 62. Olesk, A.; Praks, J.; Antropov, O.; Zalite, K.; Arumäe, T.; Voormansik, K. Interferometric SAR coherence models for characterization of hemiboreal forests using TanDEM-X data. *Remote Sens.* **2016**, *8*, 700. [[CrossRef](#)]
 63. Kugler, F.; Schulze, D.; Hajnsek, I.; Pretzsch, H.; Papathanassiou, K.P. TanDEM-X Pol-InSAR performance for forest height estimation. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 6404–6422. [[CrossRef](#)]