Deep learning for forest inventory and planning : a critical review on the remote sensing approaches so far and prospects for further applications

Hamedianfar, Alireza

2022-10

Hamedianfar , A , Mohamedou , C , Kangas , A & Vauhkonen , J 2022 , ' Deep learning for forest inventory and planning : a critical review on the remote sensing approaches so far and þÿprospects for further applications ' , Forestry , vol. 95 , no. 4 , pp. 45

http://hdl.handle.net/10138/347352
https://doi.org/10.1093/forestry/cpac002

Forestry An International Journal of Forest Research

Institute of Chartered Foresters

Forestry 2022; **95**, 451–465, https://doi.org/10.1093/forestry/cpac002
Advance Access publication 1 February 2022

Deep learning for forest inventory and planning: a critical review on the remote sensing approaches so far and prospects for further applications

Alireza Hamedianfar[1], Cheikh Mohamedou[1], Annika Kangas[2] and Jari Vauhkonen[1,*]

[1]*Department of Forest Sciences, University of Helsinki, Latokartanonkaari 7 (P.O. Box 27), FI-00014 Helsinki, Finland*
[2]*Bioeconomy and Environment Unit, Natural Resources Institute Finland (Luke), Yliopistokatu 6 B, FI-80100 Joensuu, Finland*

*Corresponding author Tel: +358 50 4303895; E-mail: jari.vauhkonen@helsinki.fi

Received 14 June 2021

Data processing for forestry applications is challenged by the increasing availability of multi-source and multi-temporal data. The advancements of Deep Learning (DL) algorithms have made it a prominent family of methods for machine learning and artificial intelligence. This review determines the current state-of-the-art in using DL for solving forestry problems. Although DL has shown potential for various estimation tasks, the applications of DL to forestry are in their infancy. The main study line has related to comparing various Convolutional Neural Network (CNN) architectures between each other and against more shallow machine learning techniques. The main asset of DL is the possibility to internally learn multi-scale features without an explicit feature extraction step, which many people typically perceive as a black box approach. According to a comprehensive literature review, we identified challenges related to (1) acquiring sufficient amounts of representative and labelled training data, (2) difficulties to select suitable DL architecture and hyperparameterization among many methodological choices and (3) susceptibility to overlearn the training data and consequent risks related to the generalizability of the predictions, which can however be reduced by proper choices on the above. We recognized possibilities in building time-series prediction strategies upon Recurrent Neural Network architectures and, more generally, re-thinking forestry applications in terms of components inherent to DL. Nevertheless, DL applications remain data-driven, in contrast to being based on causal reasoning, and currently lack many best practices of conventional forestry modelling approaches. The benefits of DL depend on the application, and the practitioners are advised to *ex ante* subject their requirements to operational data availability, for example. By this review, we contribute to the technical discussion about the prospects of DL for forestry and shed light on properties that require attention from the practitioners.

Introduction

Various forestry applications are suggested to benefit from autonomous machines and systems that re-configure themselves upon an introduction of new components or information (e.g. Uusitalo *et al.*, 2006; Nuutinen *et al.*, 2011; Pukkala *et al.* 2021). As reviewed by Müller *et al.* (2019), artificial intelligence could potentially allow for autonomous decision-making in the planning and implementation of forestry operations by learning from observations and experiences. Yet, to date, the main use for artificial intelligence has related to translating data from remote sensing into forest attributes (Müller *et al.*, 2019). Although this is valuable, future forestry applications could benefit from discovering complex or unconventional relationships in different scales. The more frequent availability of multi-source forest data due to new remote sensing methods (Kangas *et al.*, 2018) may open possibilities on the above, but also generate challenges related to considerations of data from multiple sources and (temporal and spatial) scales (see Seidl *et al.*, 2013, for a review on scaling issues).

Machine learning is a form of artificial intelligence, in which a computer is algorithmically trained to perform a task such as prediction or classification (Hastie *et al.*, 2009; Schmidhuber, 2015). The dictionary definition by Oxford Languages stresses the ability 'to learn and adapt without following explicit instructions, by using algorithms and statistical models to analyse and draw inferences from patterns in data'. Deep Learning (DL) (LeCun *et al.*, 1989; see e.g. Graupe, 2016; Deng and Yu, 2013, for overviews) is a form of machine learning that is based on the neural network concept that resembles the function of brains. Each network is composed of many layers that transfer the input to output by progressively learning higher level features

(Hatcher and Yu, 2018). These layers in between are called 'hidden layers' and a network with a sufficiently high number of hidden layers can be considered deep (Schmidhuber, 2015; Litjens *et al.*, 2017), in contrast to more shallow neural networks. DL can perform learning tasks without the need of human-derived explanatory variables (LeCun *et al.*, 2015; Schmidhuber, 2015), by virtue of which it has more potential to learn abstract features from data (Shao *et al.*, 2017). The ability to use DL with all kinds of data, including numbers, images and audio, has paved its way to a dominant role in the development of predictive systems for regression and classification problems (Hatcher and Yu, 2018).

The feasibility of machine learning approaches was recently reviewed for applications such as Earth observation (Salcedo-Sanz *et al.*, 2020), change detection (Shi *et al.*, 2020) and fire management (Jain *et al.*, 2020). Rammer and Seidl (2019) and Reichstein *et al.* (2019) provided perspectives of using DL to enhance the modelling of biotic damages (namely, bark beetle outbreaks) and geoscientific processes, respectively. There are additional implementation-specific reviews of DL for image analysis and segmentation (Hoeser and Kuenzer, 2020; Hoeser *et al.*, 2020) that are applicable to the aforementioned domains and remote sensing-based inventories (see also Kattenborn *et al.*, 2021). Finally, Diez *et al.* (2021) meritoriously reviewed DL algorithms both in general and with respect to applicability in various forest inventory tasks based on imagery acquired using unmanned aerial systems. Nevertheless, the existing reviews are not informative on the applicability of DL for forest inventory and planning applications. In the reviews, DL was predominantly applied to analyses of remotely sensed images, whereas its benefits with other forestry data types are unknown and may depend on factors such as operational data availability in an application. Supervised learning to map the input data to meaningful labels has been extensively studied using linear and non-linear models, nearest neighbour search, support vector machines and decision trees such as the random forest. A forest modeller would likely benefit from guidance for choosing between DL and these approaches, some of which may have more conservative training data demands.

This article aims to present the state-of-the-art of DL as applicable to various forestry applications, mainly those related to forest inventory. The main article text is intended as an easy-to-approach counsel to especially point out DL counterpoints to conventional forestry modelling. The article is augmented by a Supplementary data file with technical definitions on DL architectures. Those are described with a special focus on their importance and implementation in these applications. A literature review of studies employing DL in forest inventory and planning was conducted to figure out the effect of architecture and input data-specific parameters on the obtained results. Based on the review, strengths and opportunities of applying DL are identified and discussed with current problems and challenges. The text is structured as follows:

- A review of generic concepts and inherent properties of DL that need attention when applied; in particular, to briefly explain the key principles of commonly used approaches.
- A quantitative review of DL approaches currently used in forestry applications; specifically, their architecture and input data-specific factors and other aspects affecting their applicability and feasibility.
- A qualitative review of current forestry applications of DL; specifically, how the inherent aspects of DL methods are currently realized.
- Summing up ways forward and concluded recommendations.

Generic concepts and rationale behind DL

The interest in DL arouse especially since 2012, when the Convolutional Neural Network (CNN)-based AlexNet-architecture of Krizhevsky *et al.* (2012) outperformed traditional machine learning algorithms in image object detection and classification (Ma *et al.*, 2019). We therefore elaborate principles of a CNN as a baseline technique for the further review. The adoption of a CNN can be rationalized by juxtaposing to conventional supervised learning, in which the workflow from data to labels consists of feature generation, feature selection, classifier design and evaluation (e.g. Theodoridis and Koutroumbas, 2008). These steps require hand-crafted adjustments that need to be done application- and data-specifically involving manual processing and noticeable expertise (Sothe *et al.*, 2020). In Figure 1, for example, the image can be analysed by means of statistics extracted from its histogram or, if information on the co-occurrence of the image tones is needed, specific second-order statistics (see, e.g. Niemi and Vauhkonen, 2016). The analyses are complicated in particular in the latter case due to considerations of the relevant feature types, extraction scales, and that selected features are still optimal to model the phenomenon in question when computed from an image rotated slightly differently, for instance. The extracted features should later comply with the assumptions of the classifier or regressor such as normality or linearity with the training data. Considering three-dimensional data or data from multi-temporal or multiple sources makes the considerations even more complex, which, on one hand, suggests that avoiding tasks related to the manual feature selection and classifier design is an asset. On the other hand, human-interpretable features designed based on causal reasoning can also be an asset.

Convolutions and CNNs

A CNN uses convolutions to internally extract and analyse data features and, therefore, avoids manual feature generation and extraction. A convolution is a linear operation based on multiplying input data vectors with weight values given by kernels, similar to what convolving an image with a filter does in an image analysis context. Figure 2(a) shows the principle of how a convolved feature vector is obtained by sliding a kernel over input data and calculating weighted values based on the weights. Figure 2(b) shows how the resulting feature map depends on the number of kernels and each kernel's dimensions (determined by width, height and depth depending on the input data) and stride, which is a number of input unit that controls how the kernel is moved from one position to the next position. The CNN learns the weights of the kernels on its own, resulting in a feature map that represents features rather than pixel values. The use of many different convolution layers results in different feature maps: for example, in image analyses, the 1D convolutions may extract spectral information and 2D convolutions

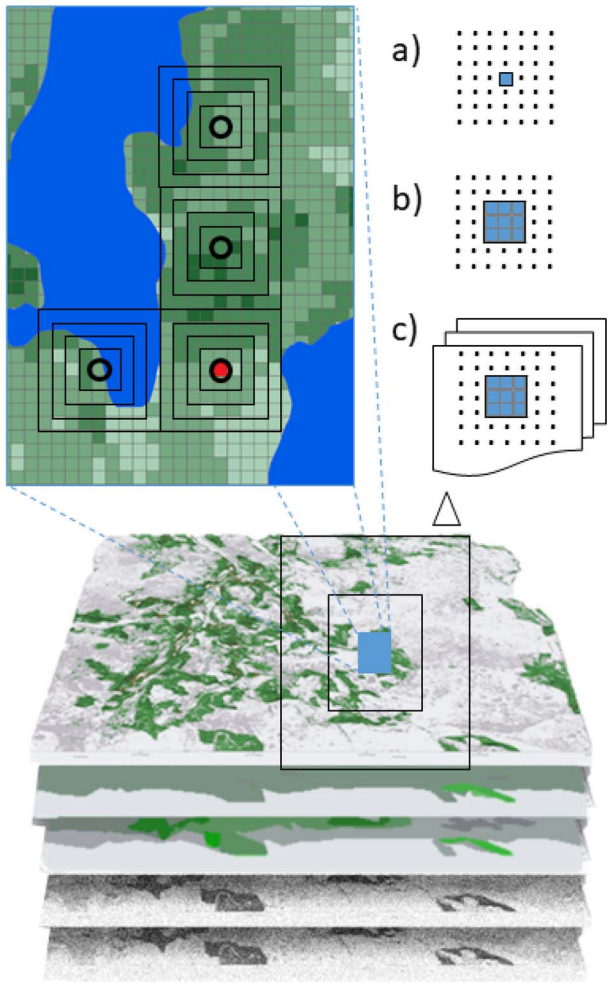

Figure 1 A schematic example of the unambiguity of feature extraction for four circular plots to predict a forest attribute for the plot filled by red. Besides extracting the pixel value of the plot, the co-occurrence of image tones in the nested squares drawn around the plots could be informative depending on the application. Relevant textural information can therefore be (a) 1-dimensional (based on a feature extractor of 1×1 pixels), (b) 2-dimensional ($n \times n$ pixels, where n is the size of the pixel neighbourhood) or (c) multi-dimensional ($n \times n \times m$ pixels, where m is the multi-source or multi-temporal data depth). In CNN, the feature extraction can ideally be replaced by the internal process to move convolution kernels corresponding to a–c over the entire multi-dimensional data stack and learn optimal weights for the obtained features based on representative training data.

spatial features, whereas 3D convolutions can exploit both these feature types. For instance, Mäyrä *et al.* (2021) found that 3D convolutions outperformed 2D ones as they were capable of learning spatial–spectral features in contrast to just spatial ones, respectively, when classifying trees of different species based on hyperspectral images. Stacking convolutional layers with different convolution kernels is reasoned as a convolution exploits data from its extent, and different convolutions may detect different types of features. As the image data contains local correlation, the local links within convolutions aim to exploit it for better features (Liu, 2020).

A typical CNN architecture includes a convolution and pooling mechanism, which breaks the image down into features analysed internally within the CNN, fully connected layers to weight the outputs of convolution/pooling, and prediction (Figure 2c). A fully connected neural network comprises a sequence of fully connected layers that link each neuron in one layer to neurons of another layer. The main advantage of fully connected network is that it does not consider any particular assumption for the input data. The major drawback of fully connected networks is that they are computationally expensive and could be prone to overfitting (Goodfellow *et al.*, 2016; Chollet, 2018). Pooling is a subsampling technique and does not implement any weights. A max-pooling takes a maximum of input elements in a region defined by the filter. It reduces the feature map dimensions while maintaining important information needed for the prediction task (Akhtar and Ragavendran, 2020). The convolution and pooling are usually operated in several rounds so that the former performs filtering to derive information, and the latter reduces and focuses the information. Altogether, these steps extend the field of view from the local computation unit to relationships all over the data and may mitigate the negative impacts of overfitting (Singh and Majumder, 2020).

Taken together, the whole CNN can, in principle, learn unique, hierarchical patterns in multi-dimensional data. In turn, the learning performance of a CNN depends on several parameters, affecting the data requirements, processing and computation burden, and the complexity of the resulting architecture, which fundamentally needs to be fine-tuned per each application. Figure 2(c) points out the needs for labelled data, or other knowledge depending on an application to optimize these parameters.

Inherent properties that require considerations by a practitioner

Different CNN architectures will be formed depending on how the components described in the previous section are arranged. When constructing a CNN-based analysis, the operator quickly comes across questions related to CNN architectures, training and optimal parameterization. To assist in comprehending these tasks, we have prepared a Supplementary data file that describes choices related to adopting a CNN at a technical level. Further, based on overviews such as Schmidhuber (2015) and Nweke *et al.* (2018), we identified Recurrent Neural Network (RNN), Autoencoders and Restricted Boltzmann Machine as the main additional DL methods with relevance to forestry. In the Supplementary data file, we provide a basic understanding of these approaches' methodological principles and differences.

When juxtaposed with any potential application, the aforementioned choices need to be considered with respect to the volume and cohesion of data available to train the DL framework for the predictions. One prerequisite to apply DL is access to large enough amounts of training data which is required to sufficiently fine-tune the large number of network parameters and at the same time to prevent overfitting. Regarding forestry applications, the following data aspects were identified as critical to consider already when pre-evaluating the suitability of a DL approach for different applications:

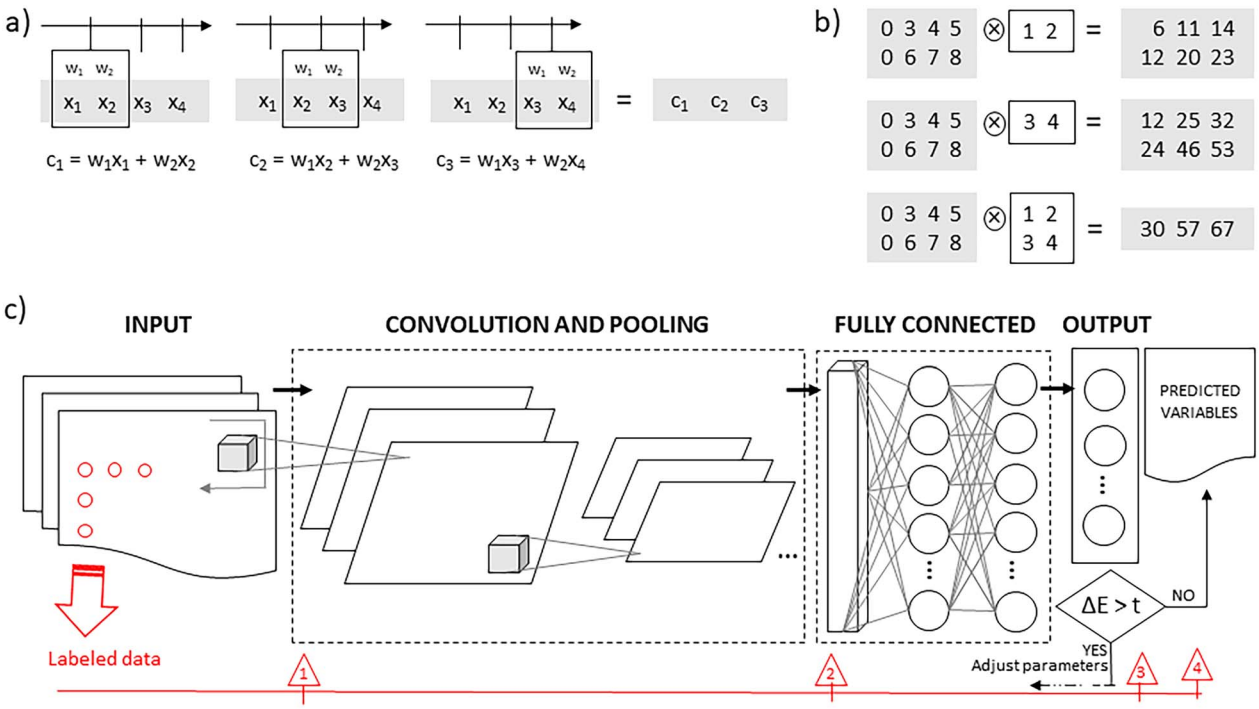

Figure 2 (a) The principle of obtaining a convolved feature vector [c_1, c_2, c_3] from an input vector [x_1, x_2, x_3, x_4] by sliding a kernel with weights w_1 and w_2. (b) Examples of convolving the input image with two rows and four columns by different kernels delineated by solid line. (c) A schematic diagram of components and required considerations of a Convolutional Neural Network (CNN). The different CNN architectures are based on concatenating varying numbers of convolution and pooling, fully connected and prediction layers. The three-dimensional box depicts the convolution kernel and arrows around it illustrate its operation. Unlike similar representations in other literature sources, the figure highlights where labelled data or other knowledge should be used to (1) inform on the form of the relationships using (piecewise) linear or non-linear activation functions; (2–3) adjust (hyper-)parameters according to the prediction performance; or (4) validate predictions. Steps 2–3 are based on comparing the change of an expected loss (E) between predicted and training features to a threshold (t). These steps differ depending on the operating mode of the CNN (forward-pass or back-propagation).

- Although the amount of required data samples cannot be generically instructed, we note that in the forestry context, sufficient data to train an efficient and robust DL architecture corresponds to a number of observations collected in wide-scale inventories (e.g. national forest inventories). An alternative is to employ data collected from various remote sensing platforms. Nevertheless, obtaining sufficient data may require merging several distinct acquisitions over time, resulting in potentially significant variations in data quality over the entire area (see further insights in Kangas *et al.*, 2019).
- Data diversity is critical to build an effective deep learning model with reliable prediction capability (Wong *et al.*, 2016). With that regard, data augmentation and transfer learning are DL-related techniques to ease dealing with scarce data conditions. Data augmentation aims to expand the number of training examples by generating new synthesized data based on various transformations applied to available training data sets (Shorten and Khoshgoftaar, 2019). For instance, the square windows drawn to Figure 1 could be rotated or processed utilizing numerous other image processing techniques to account for varying imaging conditions, thereby augmenting data. Although the new samples based on data augmentation are not independent, augmenting the existing data have proven to help to improve a DL architecture's performance and generalization potential. This approach has been rarely considered in forestry applications so far. In transfer learning, knowledge obtained from a training process of one problem is applied to another related problem (Pan and Yang, 2010). Transfer learning consists of training a DL architecture with a huge amount of data and then using the pre-trained model for fine-tuning or calibrating the network using a restricted number of case-specific training samples. This approach can be faster and more accurate than training the network from scratch. It is usually executed by removing the last units of the trained model and performing the training process with new units for the new problem (Pan and Yang, 2010).
- Target variables in forestry often require predictions for multiple time points or, in other words, the prediction of dynamics over time instead of a single state. Although time-series analyses based on weighting data from different sources and time points according to the associated uncertainty have been well-defined in the context of Bayesian filtering (Särkkä, 2013), many conventional machine learning methods applied in forestry have been extensively explored for single time variable prediction with past data, and broadening this

perspective with the use of DL methods provides an interesting possibility.

DL approaches used in forestry applications

Types of applications

Early studies that pioneered the use of DL for forestry focused on various applications including estimation of forest biophysical parameters using autoencoders which exploit high-level feature representations of image data based on decoders and encoders (García-Gutiérrez *et al.*, 2016; Shao *et al.*, 2017); interpretation and extraction of LiDAR features using CNN (Ayrey and Hayes, 2018; Contreras *et al.*, 2019); plant pattern identification and classification (Guan *et al.*, 2015; Mizoguchi *et al.*, 2017; Zou *et al.*, 2017; Carpentier *et al.*, 2018; Hamraz *et al.*, 2019; Zortea *et al.*, 2018; Dos Santos *et al.*, 2019; Fricker *et al.*, 2019; Liu *et al.*, 2019; Marrs and Ni-Meister, 2019; Narine *et al.*, 2019; Martins *et al.*, 2019; Pelletier *et al.*, 2019; Windrim and Bryson, 2019), tree attribute prediction (Ercanlı, 2020) and semantic segmentation (Chen *et al.*, 2020). Classification tasks have included, inter alia, forest pests and diseases (Safonova *et al.*, 2019), forest fire monitoring (Chen *et al.*, 2019), wind damage (Hamdi *et al.*, 2019), and dead wood as a proxy of forest health or biodiversity from aerial imagery (Sylvain *et al.*, 2019; Jiang *et al.*, 2019).

There was a notable recent focus on image and object analyses to detect trees and their species using various remote sensing data. Comparing the earlier review of Zhu *et al.* (2017) to Kattenborn *et al.* (2021), who already found about a hundred studies related to these themes from 2017–2020, this trend could be assumed to continue strong. We therefore consciously placed a stronger focus on estimating forestry characteristics and time-series dynamics (Table 1), which were not covered by the earlier reviews. In our summary of tree segmentation and classification tasks (Table 2), we focused on the archetypes of these studies and direct the reader to Kattenborn *et al.* (2021) for more applications.

Additionally, DL has been used for fusing data sources. For instance, Chang *et al.* (2019) fused data to combine methods to simultaneously classify forest cover types and estimate different forest variables using aerial, satellite, terrain and climate data at varying resolutions. Shah *et al.* (2020) used a synergy of Landsat imagery and Lidar data to produce a canopy height model for a forested area using CNN.

Su *et al.* (2019) predicted tree heights over time using the long short-term memory (LSTM). Extending RNN, LSTM networks are specialized for processing sequential data as the LSTMs 'remember' their data inputs over a longer time and improve the performance of traditional RNN which is only capable of maintaining the short-term memories.

Relationships between accuracy and input data

Our specific objective for this review was to determine whether and how DL methods improve the current estimation of forest variables of interest in terms of an error metric such as the root mean squared error (RMSE). We aimed at a meta-analysis between evaluation measures or other performance statistics and the parameterization of the approach. However, as both of these were rarely described in numerically comparative terms, we include as many details as possible that can explain the performance of the DL and hence its feasibility for the respective task.

Regarding the estimation of continuous forest inventory variables (Table 1), different numbers of field plots were utilized for the training and test datasets. The total number of the field plots varied from 60 to 17 537. Most of the studies focused on predicting above-ground biomass, and the best RMSE was 11 per cent based on 236 field plots for a study area covering 100 km^2 (Zhang *et al.*, 2019). García-Gutiérrez *et al.* (2016), reporting an RMSE of 15 per cent, similarly applied autoencoders in a smaller area (4 km^2), but they provide no information about the amount of training and validation data. Ayrey and Hayes (2018) used the highest number of field plots split to 15 537 samples for training and additional 1000 for testing and 1000 for validating CNN-based architectures.

In the structure of DL methods, the size of input data (tiles), spatial resolution, filters and batch size (the number of training examples utilized in one iteration) may effectively contribute to the predictive accuracy, efficiency and training time of the model. For instance, filter size influences the number of trainable parameters, and the size of the output depends on the size of inputs. Pelletier *et al.* (2019) reported the impact of different filter sizes (3, 5, 9, 17 and 33) and batch sizes (8, 16, 32, 64 and 128) for crop and forest species detection using temporal CNN applied on Formosat-2 satellite images resulting in better accuracy for filter size 9 and batch size 32. Schiefer *et al.* (2020) examined the impacts of different spatial resolutions and tile sizes for tree species discrimination using U-net semantic segmentation model on UAV-derived canopy height models. The authors claimed that tile size did not represent a meaningful effect on model accuracy. However, they witnessed that larger tiles could impact the accuracy of those classes with a low number of samples because increasing the size of input tiles reduces the number of samples for underrepresented classes. Other studies focusing on comparable target variables did not evaluate (or at least do not report) multiple filter and batch sizes. Moreover, the size of input data and spatial resolution is likely to influence the performance of the DL architecture. Current literature lacks a thorough assessment of these parameters for forestry prediction tasks. In general, DL hyperparameters may need to be optimized context-dependently for each application.

Relationships between accuracy, architecture and methodology

The architecture choice may affect DL results as compared with conventional machine learning methods for forest variable estimation. Zhang *et al.* (2019) investigated the use of autoencoders for forest biomass estimation on Landsat8 and LiDAR data sets. The autoencoders outperformed traditional k-nearest neighbour, random forest, support vector regression and multiple stepwise linear regression approaches by 1 per cent to 7 per cent in terms of the relative RMSE. According to Ayrey and Hayes (2018), Inception-V3 or GoogleNet (see the Supplementary data file) were the most successful CNN architectures for forest above-ground biomass estimation from LiDAR data with RMSE of 26 per cent or 27 per cent and bias of 0.7 per cent or 2.1 per cent,

Table 1 A summary of the studies related to continuous forest variable estimation.

DL method	Task	Data	Architecture	Optimizer	Learning rate	Variable (unit)	Sample size	Training data size	Input data size	Error	Time points	Reference
Autoencoder	Regression	Landsat8, Sentinel-1, LiDAR	Stacked auto autoencoder	stoch. gradient descent	-	AGB (t. ha^{-1})	1400 plots	800 plots	-	14.4%	1	Shao *et al.* (2017)
		Landsat8 LiDAR	Stacked auto autoencoder	stoch. gradient descent	-	AGB (t. ha^{-1})	236 plots	177 plots	-	11.4%	1	Zhang *et al.* (2019)
CNN	Semantic segmentation	LiDAR	FCN	-	-	tree diameter	-	NA	-	-	1	Chen *et al.* (2020)
	Classification	LiDAR	Inception-V3	-	-	AGB (t. ha^{-1})	17 537 plots	15 537 plots	$7 \times 7 \times 18$ pixels (3D CNN)	48.1%	1	Ayrey and Hayes (2018)
		Landsat 8, LiDAR	-	stoch. gradient descent	-	Canopy Height	-	NA	30×30	0.98 m	1	Shah *et al.* (2020)
R-CNN	Classification and Regression	Aerial image, Landsat 7 time-series, Topography, Climate	-	Adam	0.1	Tree species, AGB (t. ha^{-1})	9967 plots	7724 plots	80×80	36%	2	Chang *et al.* (2019)
DNN	Regression	LiDAR, ICESat-2 profile, Landsat image metrics	DNN	RMS Prop	0.1 0.01 0.0001	AGB (t. ha^{-1})	-	1448 pixels	32×32	15.5–15.6 (Mg/ha)	1	Narine *et al.* (2019)
		Plot data		NA	0.999	Height and diameter relation	150 plots	NA	-	4.95%	1	Ercanlı (2020)
Autoencoder		LiDAR	Autoencoder	NA	-	AGB (t. ha^{-1})	39 + 54 plots	NA	-	15%	1	García-Gutiérrez *et al.* (2016)
RNN		Tree age, temperature rainfall, soil, slope position	Autoencoder with LSTM	RMS Prop	0.001	Height growth	1000 (sample type not mentioned)	500 (data type not mentioned)		0.0706%	1	Su *et al.* (2019)

Table 2 A summary of the studies related to segmentation and classification.

DL method	Tasks	Data	Architecture	Optimizer	Learning rate	Variable	Input Data Size	Labelled Data	OA%	Reference
Deep Boltzmann machines	Classification	Mobile LiDAR	deep Boltzmann machines	stochastic gradient descent	-	tree species	-	50 000 tree samples from 10 different tree species	86.1	Guan *et al.* (2015)
Deep Belief Network		LiDAR		stochastic gradient descent	-		-	-	95.6	Zou *et al.* (2017)
CNN	patch-based	Hyperspectral and LiDAR	FCN	stochastic gradient descent	0.0001	tree species	-	713 trees	86.67	Fricker *et al.* (2019)
	Semantic segmentation	Formosat-2 images	TempCNN	Adam	-	Land cover and tree species	32 × 32	1419 polygons	93.45	Pelletier *et al.* (2019)
	Classification	UAV	ResNet-50 with SLIC and SVM	stochastic gradient descent	0.01	Tree detection	32 × 32 pixels	-	89.01	Martins *et al.* (2019)
	patch-based	public dataset (named BarkNet 1.0)	resnet34	Adam	0.0001	tree species	32 × 32 pixels	-	97.81	Carpentier *et al.* (2018)
	patch-based	UAV	VGG-16	Adam	0.0001	Forest damage	224 × 224 pixels	200 image-patches	More than 90	Safonova *et al.* (2019)
	Semantic segmentation	Aerial image (V-NIR)	U-Net	Adam	0.01 0.001 0.0005 0.00001	Forest damage	256 × 256 pixels	1525 tiles	92	Hamdi *et al.* (2019)
	patch-based	Aerial image	VGG16	stochastic gradient descent	0.001	Tree mortality	21 × 21 41 × 41 pixels	315 polygons	94	Sylvain *et al.* (2019)
		Ground photography	UNET	Adam	0.001	Tree species	224 × 224 pixels	64 field plots	96.03	Liu *et al.* (2019)
	Semantic segmentation	Aerial image (V-NIR)	FCN-DenseNet	Adam	-	Tree mortality	512 × 512pixels	261 fallen dead trees and 305 standing dead trees	NA	Jiang *et al.* (2019)

respectively. Inception-V3 was the top-performing architecture, and it outperformed linear mixed model and random forest predictions by 3–5 per cent difference in RMSE. In another study, Pelletier *et al.* (2019) investigated crop and tree species classification using Formosat-2 time-series data and found CNN to outperform an RNN alternative.

Compared to a single data source, data fusion-based feature enrichment by multi-source data may support applications that entail more temporally, spatially or spectrally varying properties than single data could deliver. With that regard, Su *et al.* (2019) applied joint autoencoder-RNN for tree height growth prediction upon integration of different datasets, such as tree age, temperature, rainfall, soil and slope position. Chang *et al.* (2019) developed a multi-task recurrent CNN to integrate various data sources, including aerial and satellite image time-series, topography, and climate data to classify different forest cover types and forest variable attributes, such as above-ground biomass, quadratic mean diameter, basal area and canopy cover. They concluded that a multi-task method outperformed support vector machine and random forest algorithms.

As summarized in Table 2, CNN has been a prominent method for tree detection and classification tasks such as tree species discrimination (Fricker *et al.*, 2019; Carpentier *et al.*, 2018; Liu *et al.*, 2019; Pelletier *et al.*, 2019), forest damage detection (Hamdi *et al.*, 2019; Safonova *et al.*, 2019) and tree mortality mapping (Sylvain *et al.*, 2019; Jiang *et al.*, 2019). The CNNs mainly were used for semantic segmentation and patch-based approaches, the definitions of which are elaborated in the Supplementary data file. Besides CNN, deep Boltzmann machines (Guan *et al.*, 2015) and deep belief network (Zou *et al.*, 2017) were applied for tree species recognition. A critical limitation on training a CNN architecture for image classification is the laborious process of preparing training sample labels, and consequently, no training data with wide representation and generality for species classification are available. For instance, transferring knowledge from labelled to unlabelled data has been evaluated for several classification tasks (e.g. Li *et al.*, 2017), but not representatively for classification tasks in areas within a continuous forest cover.

Despite the importance of minimizing the loss function, the related optimization is regularly overlooked and has not received much attention. Using a proper optimizer is essential in selecting the beneficial features for predicting the response variable and fine-tuning the model parameters. It is done by evaluating which optimizer is more effective and leads to better performance concerning evaluation criteria (Okewu *et al.*, 2019). As reported in Tables 1 and 2, most studies utilized classical stochastic gradient descent, and few studies used more efficient choices based on adaptive learning rate methods such as Adam and RMSProp (Mubin *et al.*, 2019).

Discussion

Considering the findings above and in Tables 1 and 2, it is evident that the applications of DL for forestry are in an early phase. The primary study line has related to comparing various CNN architectures between each other and against conventional machine learning techniques in the estimation of forest attributes. A better performance was often reported for CNN, but the reason for this was not explained in terms of simple or complex relationships between the attributes considered, availability of data to estimate those, or similar. Even though it is not yet possible to conclude which solution may perform better, in the discussion we aim at identifying some good practices and challenges in the current state-of-the-art, and at extending the discussion by qualitatively collating characteristics of forestry applications and inherent DL properties.

A selection of studies representing the current state-of-the-art of DL in forestry

Preceded by initial trials with autoencoders (García-Gutiérrez *et al.*, 2016; Shao *et al.*, 2017), Ayrey and Hayes (2018) is one of the pioneering studies in using CNN for forest variable estimation, and many later studies are based on the same principles. Ayrey and Hayes (2018) identified separate extraction of predictive metrics and related considerations (e.g. variation in acquisition parameters, multicollinearity) as weaknesses that could be circumvented by means of DL. Many CNN architectures were evaluated and compared with other approaches for LiDAR-based forest inventory. Many later studies justify the choice of a DL method based on the same argumentation that omitting the step of metric extraction is beneficial.

Chen *et al.* (2020) introduced a semantic segmentation approach for LiDAR cloud points for DBH predictions using the CNN architecture PointNet++. The method automatically produces tree diameter estimates from an analysis of point cloud data, indicating that conventional LiDAR and textural features would result in less accurate results. Although the developed algorithm and technique are unique to the case, the principles of semantic segmentation may also have many other applications in fields related to mimicking segmentation patterns satisfactory to the user of the data.

Among applications related to modelling forestry dynamics over time, Su *et al.* (2019) employed DL in predicting the height growth of large trees based on tree age, temperature rainfall, soil and slope data. Although developed for powerline safety assessment, the method showed potential for modelling the time-series of tree heights of fast-growing Eucalyptus species. The study hints at hyperparameter choices and configurations as critical factors influencing the accuracy of the DL network, which was illustrated by choosing one activation function over another and then comparing the resulting RMSE. The best-performing approach was to merge the extraction capabilities of an autoencoder with the forecasting capabilities of the LSTM. This is one of the hybrid ideas that we see can yield satisfactory results.

Chang *et al.* (2019) presented a DL approach that employed several methods and principles. The hybrid DL approach concurrently classified forest types and estimated forest parameters including above-ground biomass, basal area, canopy cover and mean diameter based on the openly available optical remote sensing, terrain and climate data. An RNN based on the LSTM constituted an umbrella to combine classification and regression and learn from a complexity of the data, showing here as a time-series. The interesting point in this research is the high efficiency of the RNN to combine (fuse) different data sources as input variables. Yet, this work can also be criticized for lacking details preventing the replication of the analyses (see below).

Mäyrä *et al.* (2021) evaluated the integration of hyperspectral remote sensing and LiDAR-derived canopy height model for boreal tree species classification using CNN. They compared 3D-CNNs based on the use of spectral–spatial information (contrasted to 2D-CNN that would only use spatial information) with benchmarks conventionally used for this task. The 3D-CNNs outperformed a support vector machine and an artificial neural network by 3–5 percentage point improvement to the overall classification accuracy and other benchmarked methods by a much wider margin. The CNN was trained from scratch, and data augmentation was implemented to overcome the negative impact of a limited number of labelled training samples. Upon analysis of input image patch sizes, the smallest (4 m) and largest (10 m) tested patches were found to result in the highest (87 per cent) and lowest (85 per cent) overall accuracy, respectively, but this cannot be considered a significant difference. Importantly, they analysed the CNN solutions to discover the spectral and spatial features that had positively impacted the classification, thereby providing a useful interpretation of the CNN result that can otherwise be considered as a black box.

Current challenges, reproducibility of the results and replicability of the methods

The crucial factor for the success of a DL model is the accessibility to a sufficient amount of training data. A specific challenge is the requirement to have annotated (labelled) data for the training (Padarian *et al.*, 2019). In many real-world forestry problems, it may be challenging to acquire massive amounts of such labelled information. Based on field inventories, collecting large amounts of observations is difficult and expensive, requiring extensive field campaigns over large areas. Although it is nowadays feasible to use various remote sensing techniques (Kangas *et al.*, 2018), it remains a problem that these observations usually need to be interpreted, i.e. refined to labelled information. In this aspect, forestry applications may differ from many fields where the application of the DL can employ databases of labelled data collected in huge amounts from cash and credit card transactions and similar registers or by means of social media and other crowdsourcing approaches.

No similar label generator can easily be identified for forestry applications, except possibly for harvester data (Uusitalo *et al.*, 2006), which however need to be linked to other data sources for prediction purposes. Although there is a great potential to use smartphone applications such as iNaturalist to collect taxa observations, those usually need to be annotated by experts for reliability (Lahti *et al.*, 2021) and accounted for location biases originating from the behaviour of the observers (Mononen *et al.*, 2018). Measuring forest parameters requires a sample to cover the entire forest variation under sufficient visibility for the (at the moment, proprietary) interpretation algorithms (Pitkänen *et al.*, 2021). Only public participation applications that collect opinions (e.g. Kangas *et al.*, 2015) could be used directly as the people are reliable sources on their own preferences. Those could be used as reference data for DL to learn generally appealing location, aesthetic and natural properties related to trees and other aspects based on multi-source data.

To compensate for the small number of data samples, Shao *et al.* (2017) used LiDAR data as synthetic data; however, no assessment was provided to indicate the accuracy improvement. Among the studies reviewed by Kattenborn *et al.* (2021), 60 per cent used visual interpretation to generate the training data, whereas the remaining 17 per cent or 22 per cent were based on pure *in situ* or combined visual and *in situ* observations, respectively. Further studies should obviously generate more efficient and robust means to the data generation. Another, more generic solution that appears under-utilized in forest variable estimation is the use of the transfer learning method, which is beneficial to propagate the knowledge learned from a large-scale dataset to a comparatively small-scale dataset (Zhang *et al.*, 2019). In addition, the use of data augmentation could possibly mitigate the negative impacts of issues related to training data limitation. Standard data augmentation techniques based on changing grey level values or mirroring datasets might not work or be sufficient when aiming to learn 3D-structural phenomena. With some data, such as synthetic aperture radar (SAR) reflectivity including both amplitude and phase components, one might even end up with very implausible results with standard data augmentation so that more research on proper techniques is needed.

Currently, the literature lacks guidance on the appropriate amount, extent, geographical coverage and distribution, and so on, related to training data. Although specific guidelines have not been reported, obviously more geographically distributed observations could improve the predictive accuracy and generalization ability. Having sufficient data and choosing a proper approach to partition training, testing and validation subsets is particularly important. The training and validation subsets should be independent of each other, but represent the whole variation observed in the population. If the samples are divided randomly, the representativeness of divided samples will most likely vary. Therefore, the random splitting of the labelled data into train and test samples may cause overestimations of the accuracy. Some of the studies for forest variable estimation studies utilized the random sampling for dividing the ground truth data to train and test data (Su *et al.*, 2019; Zhang *et al.*, 2019; Narine *et al.*, 2019; Shah *et al.*, 2020). Although this approach is simple and widely used, it may limit the transferability of the model to new areas.

Increasing the number of neurons in DL has is a high chance of solving complicated problems. As a result, a deep network can highly adapt itself to the training data. However, adding more fully connected and convolutional layers leads in increasing the depth and complexity of the network, and as a result, the model could be prone to excessive running times and overfitting. The latter leads to degrading quantitative and qualitative accuracy. Selecting the appropriate number of neurons in the hidden layer and suitable hyperparameters provides the opportunities to optimally solve these problems. The learning rate is one well-known hyperparameter that relates to the rate of updating weights in the analysis. Most of the forest variable estimation studies utilized conventional optimizers of the hyperparameters based on the stochastic gradient descent algorithm, whereas more efficient approaches could be based on adaptive learning rate methods. Many studies report no information on the used optimizer (García-Gutiérrez *et al.*, 2016; Ayrey and Hayes, 2018; Ercanlı, 2020; Chen *et al.*, 2020), whereas few studies used dropout and batch normalization to reduce overfitting and improve the generalization capability in the training process (Ayrey and Hayes, 2018; Chang *et al.*, 2019; Shah *et al.*, 2020). There may be room to improve the model accuracy and generalization abilities by further studying the optimization of the parameters involved. Apart

from the importance of proper parameter optimization, it would be essential to investigate and evaluate the effect of varying subsample of input data for creating training, test and validation splits in accordance with the choice of filters and optimizers used in the training process. Moreover, although most of the studies only utilized limited numbers of field plots and achieved promising results over a single study area, the reproducibility of a DL result has not yet been investigated over new (independent) datasets and separate study areas.

Moreover, issues related to the replicability of the methods were noted especially when reviewing the earliest DL studies, but also among more recent ones. For instance, based on publications of Ayrey and Hayes (2018) and Chang *et al.* (2019), it is not possible to determine what was the exact modelling unit and method to derive the response variable for those units. Especially linking field data measured from circular plots with pixels corresponding to the plot is nontrivial (cf., Figure 1), but required when training convolutions for the response values to be predicted. It is possible that the reviewers of the early DL manuscripts might not have been familiar enough with the methodology to ask crucial questions on the implementations. Similar signs of undermaturation can be pointed out from the recent DL applications as there were in the proliferation of nearest neighbour imputations until exemplary guidance on feature selection and cross-validation (e.g. Packalen *et al.*, 2012). Our findings call for similar investigations and how-to-instructions on DL.

Perspectives of data dimensions on choosing a modelling approach

Over the last decades, various machine learning algorithms have been applied for regression and classification of tree and stand parameters. Comparisons of DL approaches with these algorithms cannot be considered satisfactory so far. This is especially true if the validation is extended to qualitative aspects such as the fitness of the DL method to the data availability and similar prerequisites of the given task. It is not clear all along the studies whether the reported excellent performances of DL can be attributed to the method itself or methodical aspects such as no independent validation, undetected overfitting or a lack of proper comparison to other feasible methods.

A study on using DL for estimating the tree height–diameter relationship (Ercanlı 2020) can be used as a cautionary example on potentially choosing an excessively complex approach to model a simple phenomenon. Although mentioning DL in its title, the algorithm described by Ercanlı (2020) was a common Multilayer Perceptron (MLP), the structure of which was grown to include 100 neurons within nine hidden layers in the best-performing version. Its performance was compared with nonlinear regression and mixed-effects modelling, indicating the MLP to outperform the conventional methods. Although there can be possible overfitting issues due to evaluating the network with specific data, Ercanlı (2020) is brought up here to proclaim the rationale of the modelling choices. The height–diameter modelling is a well-known and thoroughly studied problem in the field of forest biometrics. Because of measuring trees within plots and plots within stands, the data become nested with a hierarchical structure of errors. Therefore, a general recommendation is to adopt a mixed-effects modelling approach to manage the hierarchical errors, but also because in a practical case the parameters associated with the random effects are unknown and must be predicted. For the latter, the mixed-effects modelling approaches offer undisputed benefits due to calibration abilities employing the Best Linear Unbiased Prediction, which is elaborated by Mehtätalo *et al.* (2015). Ercanlı (2020) did not address the hierarchical data structure, whereas predicting the random stand effects or calibrating the predictions with a limited number of observations would have resulted in a comparative evaluation accounting for all the possibilities of the modelling methods in a practical situation. Even if DL approaches may not have a similar theoretical basis for these aspects as statistical analysis, accounting for the hierarchical structure that is visible in the data, for instance, must be developed in the future.

Apart from the above example, Mohamedou *et al.* (2019) did not find the MLP approach to add value over diameter increment predictions. As reasons, they suggest a better parameterization of the linear mixed-effects model according to causal reasoning on the biophysical phenomenon. Predictions of inventory totals or attributes of major species based on high-resolution auxiliary data are generally found challenging to improve, especially in boreal forest conditions. For instance, Niska *et al.* (2010) found the artificial neural networks and k-nearest neighbour predictions comparably accurate for the total attributes. The methods differed in accuracy for plot and stand levels, and it is typical to get similarly contradictory performances between methods in predicting species-specific and minor species' properties (Varvia *et al.*, 2019). Better performances of different methods can also be just by a chance because of the small proportion of the better-predicted phenomena in the data or similar.

We cautiously suggest that in all the above cases, the potential of a method may be related to the number of dimensions of the predicted phenomena. Estimating forest growing stock or biomass is a traditional modelling task, which is essentially doable with a 1D approach (i.e. using a vector of explanatory features based on a 1D feature extractor as in Figure 1(a)). The possibilities to extract additional information for 1D vectors with CNN are probably limited to what can be achieved by properly adding interaction terms etc. to conventional parametric or non-parametric modelling. On the contrary, tree detection, segmentation and species recognition are 2D tasks (Figure 1(b)) and correspond to generic image object detection and classification, for which the CNN first broke through. It is further possible that CNNs turn out more useful in tasks where the modelled phenomenon requires considering multiple scales or the time dimension (Figure 1(c); see below).

Future DL studies could benefit from lessons learned with other modelling approaches

Various DL approaches were applied to single tree-level inventory that constitute a chain of events of individual tree detection, feature extraction and estimation of tree attributes in terms of conventional methods. It is instructive to consider best practices that already exist for these steps in the literature without the DL intervention. There is a number of approaches to carry out individual tree detections as reviewed by Koch *et al.* (2014), Zhen *et al.* (2016) and Lindberg and Holmgren (2017). Among these techniques, integrating external knowledge into the image

analysis of remotely sensed data has been beneficial for the success rates regardless of whether the knowledge was obtained by learning from previous algorithm runs (Heinzel *et al.*, 2011), estimating the total stem number based on other data (Ene *et al.* 2012) or point processes (Kansanen *et al.*, 2016), employing prior knowledge on allowable tree dimensions (Lähivaara *et al.*, 2014; Swetnam and Falk, 2014; Sačkov *et al.*, 2017), or combining tree detection and size modelling (Kansanen *et al.*, 2019). All the listed considerations could logically be attempted by means of DL, but this ambition was missing from the studies we came up with.

Fassnacht *et al.* (2016) provide generally applicable recommendations on tree species classification from remotely sensed data. Although not covering DL methods, Fassnacht *et al.* (2016) concluded that '[m]ost studies followed data-driven approaches and pursued an optimization of classification accuracy, while a concrete hypothesis or a targeted application was missing in all but a few exceptional studies'. They further encourage more research on the causal understanding of the traits that affect the remotely sensed signal and therefore affect which tree species can or cannot be classified under given conditions. Even if providing an otherwise meritorious study of DL and other methods for species classification, Mäyrä *et al.* (2021) can be categorized as a purely data-driven and algorithm benchmarking study. Moreover, it could be questioned that since the CNN method required data augmentation, would the compared methods not have benefitted from the similar expansion of the training data? In principle, manually rotating the images (cf., Figure 1) or using appropriate textural features could have yielded the same result also with a more traditional approach, but obviously based on a much higher manual processing burden.

Instead of decimal improvements to RMSE figures, it may be more fruitful to use DL in applications that require producing something the conventional methods cannot. The CNN's ability to internally learn features from data may be considered as such. Applications specifically benefitting from that can be identified by juxtaposing with traits described in earlier literature based on other methods. First, discussing scale issues in the context of modeling ecosystem structure and functioning, Seidl *et al.* (2013) hinted that existing models could be reviewed to learn on scale-dependencies for various applications. Even in the case where tree-level results were aimed for, Maltamo *et al.* (2009) and Vauhkonen *et al.* (2010) found beneficial to, in addition to parameters extracted from the tree segments, also use multi-scale predictors such as those describing the area-level forest structure in the proximity of the tree. Because of the internal logic based on image convolution with varying kernel size (Figures 1–2), the CNNs could possibly infer appropriate scaling from the data. The data-drivenness of the CNNs could that way be soundly employed to learn multi-scale patterns for predictions described above and domains such as spatial ecology and forest ecosystem modeling, where the scale issues are of importance, but related analyses seem to be lacking.

Possibilities on DL of forestry dynamics and management scheduling over time

Although the reviewed studies presented promising DL methods and results for state variables of a single point in time, our review indicates that the current applications of DL for forest management and inventory largely lack predictions of forest dynamics and growth over time. As the time factor is essential for the forest dynamics of an area, a time-sensitive DL framework would be important for a better understanding of forest change and providing timely informed decisions. Developing DL frameworks that can handle time-series data is an essential aspect requiring innovative solutions. A few suggestions for a structured workflow to process forest inventory time-series using DL can be formulated based on literature from other fields, bearing in mind that no comprehensive standard DL procedure relevant to forest characteristics is currently sketched or tested.

According to observations made from other scientific fields with time-series data, we expect that RNNs such as the LSTM will take an influential role because of the ability to simultaneously consider both the present and past data (Hochreiter and Schmidhuber, 1997; Gers *et al.*, 1999). On a different occasion, Wan *et al.* (2019) criticize the LSTM and techniques as ineffective with aperiodic datasets and time-consuming with subsequent needs to develop its learning process over time. Although a combination of RNN–CNN has shown potential for time-varying image classification (Mou *et al.*, 2019) and multi-task learning for species detection and forest variable estimation (Chang *et al.*, 2019), such approaches should be further investigated for synergistic data fusion of multi-source data and field plots to develop a time-series forecast strategy of forest variable attributes.

Using prior data from models and historical observations together with current measurements has potential to both improve estimates and reduce the data collection burden. Little has been done to address the benefit from prior data by DL, compared with using more widely known Bayesian approaches for this purpose (e.g. Uusitalo *et al.*, 2006; Lähivaara *et al.*, 2014; Ehlers *et al.*, 2018; Varvia *et al.*, 2019). It is essential that the uncertainties of using pixels *vs* aggregated units are quantified for drawing informed decisions, where the LSTM-based solutions come conceptually close to the Bayesian filtering (e.g. Särkkä, 2013). A potential and logical solution would be to integrate DL with a Bayesian approach, where the DL would learn the weights for a Bayesian network. Those could possibly be further used to predict uncertainties, learn from them, and calibrate the predictions. Another example could be adding point estimates by credible intervals that allow uncertainty analyses, and subsequent computational advantages obtained by replacing an approach based on Bayesian linear assumptions (Varvia *et al.*, 2019). The correlation of multi-source or multi-temporal data sources is a challenge in fusing these data (Ehlers *et al.* 2018), and therefore it is essential to investigate the impact of this correlation on DL-based analyses. An integration of Bayesian filtering and prediction (or their non-parametric variants such as the Gaussian process regression of Varvia *et al.*, 2019) with DL would interestingly provide potential to analyze more than one static state and therefore more feasibly learn the dynamic nature of forestry attributes.

It is also possible to consider management scheduling in terms of DL by re-thinking the allocation of operations that is usually based on linear programming and its variants. Specifically, the management scheduling problem would be described as a discrete event stochastic system that is driven by Markov decision processes, which are in some state S at each time step. Moving into a new state S' is influenced by the chosen

action A, giving the decision-maker a corresponding reward R(S, A, S′). A policy is a rule that a decision-maker follows when selecting actions in each state. The task is then to formulate such action-value or reward functions that allow selecting the optimal policy to maximize the total reward over all successive steps. In other fields, shallow neural networks (Laguna and Marti, 2002; Jin *et al.*, 2004) and reinforcement learning have been tested for annealing manufacturing schedules (Stefán, 2003). Recently, Malo *et al.* (2021) tested reinforcement learning for optimizing forest management and found it to allow inclusions of stochastic events without discretizing state and control variables. The shallow version could be developed to deep reinforcement learning by means of recursive training with earlier solutions. In Hinton (2014), a layer of hidden units of a RBM was trained with earlier RBMs to cover the possible Markov decision processes of alternative treatments. The resulting deep hierarchy was then learned as a neural network (Hinton, 2014). This way of thinking corresponds to learning from the successes of previously solved forest planning problems, which is an interesting future outlook.

Conclusion

This review identified the current state, trends, challenges and research needs using DL for forestry applications. Several pioneering trials of tree species mapping, forest attribute estimation, health and disease determination, and fire monitoring have been presented. Nevertheless, this field remains relatively young, and it is expected to yield plentiful studies in the coming years.

DL provides the opportunity to learn from multi-source and multi-temporal data. The main asset of DL is the possibility to internally learn multi-scale features without an explicit feature extraction step. However, this asset can also be perceived negatively as DL models are currently hard to interpret, even if interpretations and visualizations of the patterns in the data identified by DL can be developed as in Mäyrä *et al.* (2021). Until better understanding, it is easily perceived as a black box approach with risks related to the generalization abilities of the predictions, for example. Essential factors for generating robust DL methods with low chance of overfitting are a sufficient amount of representative, labelled training data and the appropriate evaluation of various hyperparameters and optimization methods that depend on the selected architecture and available data. Consequently, the fitness of a DL method to an application depends on how extensively it can be parameterized under operational conditions. We note that lessons learned with conventional modelling approaches based on causal reasoning may turn out to be useful 'training data' for DL.

The prospects of DL are likely better realized, when the studies move forward from the current main study line related to comparing various CNN architectures between each other and against conventional machine learning techniques. It is possible that DL allows learning from observations and experiences, thereby improving forestry operations through more autonomous decision processes, as envisaged for machine learning in general by Müller *et al.* (2019). Meanwhile, we expect that the following applications are increasingly realized as intermediate steps to this overarching goal: (1) discovering new knowledge by novel combinations of data from multiple scales and sources including topographical surveys, weather and climate, historical maps, and taxonomic observations annotated by experts, in addition to conventional forestry and remote sensing data sources; (2) distinguishing species or sizes while segmenting tree or tree group instances using limited expert annotation of ground truths and semantic segmentation types of CNNs; (3) learning optimal weights for Bayesian probabilistic frameworks to account for stochastic features and thus better manage uncertainties and calibrate predictions accordingly; and (4) re-thinking management scheduling problems as deep reinforcement learning from databases containing information on forestry production possibilities and decision makers' preferences, allowing to learn from previously solved forest planning problems. Novel applications may be innovated based on considerations of which components inherent to DL optimally translate to forestry applications or (yet undiscovered) parts of them.

Supplementary data

Supplementary data are available at Forestry online.

Data availability statement

No new data were generated or analysed in support of this research.

Acknowledgements

We would like to thank the editor and two anonymous reviewers for exceptionally thoughtful comments that greatly improved the paper.

Conflict of interest statement

None declared.

Funding

The Academy of Finland [grant number 324193].

References

Akhtar, N. and Ragavendran, U. 2020 Interpretation of intelligence in CNN-pooling processes: a methodological survey. *Neural Comput. Appl.* **32**, 879–898.

Ayrey, E. and Hayes, D.J. 2018 The use of three-dimensional convolutional neural networks to interpret LiDAR for forest inventory. *Remote Sens.* **10**. 10.3390/rs10040649.

Carpentier, M., Giguere, P. and Gaudreault, J. 2018 Tree species identification from Bark Images using Convolutional Neural Networks. *IEEE Int. Conf. Intell. Robot. Syst.* 10.1109/IROS.2018.8593514

Chang, T., Rasmussen, B.P., Dickson, B.G. and Zachmann, L.J. 2019 Chimera: a multi-task recurrent convolutional neural network for forest classification and structural estimation. *Remote Sens.* **11**. 10.3390/rs11070768.

Chen, Y., Zhang, Y., Jing, X., Wang, G., Mu, L., Yi, Y., Liu, H., Liu, D. UAV image-based forest fire detection approach using Convolutional Neural Network.

In *Proceedings of the 2019 14th IEEE Conference on Industrial Electronics and Applications (ICIEA)*, Xi'an, China, 19–21 June 2019; pp. 2118–2123.

Chen, S.W., Nardari, G.V, Lee, E.S., Qu, C., Liu, X., Romero, R.A.F. and Kumar, V. 2020 SLOAM: semantic lidar odometry and mapping for Forest inventory. *IEEE Robot. Autom. Lett.* **5**, 612–619.

Chollet, F. 2018 *Deep Learning with Python*. Manning Publications Co., New York, NY, USA.

Contreras, J., Denzler, J. and Sickert, S. 2019 Automatically estimating forestal characteristics in 3D point clouds using deep learning. *IDiv Annual Conference*, Leipzig, Germany, *29-30 August 2019*. https://elib.dlr.de/133241/

Deng, L. and Yu, D. 2013 Deep learning: methods and applications foundations and trends R in signal processing. *Signal Proc.* **7**, 197–387.

Diez, Y., Kentsch, S., Fukuda, M., Caceres, M.L.L., Moritake, K. and Cabezas, M. 2021 Deep learning in forestry using UAV-acquired RGB data. *A Practical Review*. *Remote Sens.* **13**. 10.3390/rs13142837.

dos Santos, A.A., Marcato Junior, J., Araújo, M.S., Di Martini, D.R., Tetila, E.C., Siqueira, H.L., *et al.* 2019 Assessment of CNN-based methods for individual tree detection on images captured by RGB cameras attached to UAVS. *Sensors* **19**. 10.3390/s19163595.

Ehlers, S., Saarela, S., Lindgren, N., Lindberg, E., Nyström, M., Persson, H.J., *et al.* 2018 Assessing error correlations in remote sensing-based estimates of forest attributes for improved composite estimation. *Remote Sens.* **10**. 10.3390/rs10050667.

Ene, L., Næsset, E. and Gobakken, T. 2012 Single tree detection in heterogeneous boreal forests using airborne laser scanning and area-based stem number estimates. *Int. J. Remote Sens.* **33**, 5171–5193.

Ercanlı, İ. 2020 Innovative deep learning artificial intelligence applications for predicting relationships between individual tree height and diameter at breast height. *For. Ecosyst.* **7**. 10.1186/s40663-020-00226-3.

Fassnacht, F.E., Latifi, H., Stereczak, K., Modzelewska, A., Lefsky, M., Waser, L.T., *et al.* 2016 Review of studies on tree species classification from remotely sensed data. *Remote Sens. Environ.* **186**, 64–87.

Fricker, G.A., Ventura, J.D., Wolf, J.A., North, M.P., Davis, F.W. and Franklin, J. 2019 A convolutional neural network classifier identifies tree species in mixed-conifer forest from hyperspectral imagery. *Remote Sens.* **11**. 10.3390/rs11192326.

García-Gutiérrez, J., González-Ferreiro, E., Mateos-García, D. and Riquelme-Santos, J.C. 2016 A preliminary study of the suitability of deep learning to improve LiDAR-derived biomass estimation. *Lect. Notes Comput. Sci* **9648**, 588–596.

Gers, F.A., Schmidhuber, J. and Cummins, F. 1999 Learning to forget: Continual prediction with LSTM. *IEE Conf. Publ.* **2**, 850–855.

Goodfellow, I., Bengio, Y. and Courville, A. 2016 *Deep Learning*. MIT Press, Cambridge, MA, USA.

Graupe, D. 2016 Deep learning neural networks. *World Sci.* 10.1142/10190.

Guan, H., Yu, Y., Ji, Z., Li, J. and Zhang, Q. 2015 Deep learning-based tree classification using mobile LiDAR data. *Remote Sens. Lett.* **6**, 864–873.

Hamdi, Z.M., Brandmeier, M. and Straub, C. 2019 Forest damage assessment using deep learning on high resolution remote sensing data. *Remote Sens.* **11**. 10.3390/rs11171976.

Hamraz, H., Jacobs, N.B., Contreras, M.A. and Clark, C.H. 2019 Deep learning for conifer/deciduous classification of airborne LiDAR 3D point clouds representing individual trees. *ISPRS J. Photogramm. Remote Sens.* **158**, 219–230.

Hastie, T., Tibshirani, R. and Friedman, J. 2009 *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer-Verlag New York, ISBN 978–0–387-84857-0

Hatcher, W.G. and Yu, W. 2018 A survey of deep learning: platforms, applications and emerging research trends. *IEEE Access* **6**, 24411–24432.

Heinzel, J.N., Weinacker, H. and Koch, B. 2011 Prior-knowledge-based single-tree extraction. *Int. J. Remote Sens.* **32**, 4999–5020.

Hinton, G. 2014 Where do features come from? *Cogn. Sci.* **38**, 1078–1101.

Hochreiter, S. and Schmidhuber, J. 1997 Long short-term memory. *Neural Comput.* **9**, 1735–1780.

Hoeser, T. and Kuenzer, C. 2020 Object detection and image segmentation with deep learning on earth observation data: a review-part I: *evolution and recent trends*. *Remote Sens.* **12**. 10.3390/rs12101667.

Hoeser, T., Bachofer, F. and Kuenzer, C. 2020 Object detection and image segmentation with deep learning on earth observation data: a review—part II: applications. *Remote Sens.* **12**. 10.3390/rs12183053.

Jain, P., Coogan, S.C.P., Subramanian, S.G., Crowley, M., Taylor, S. and Flannigan, M.D. 2020 A review of machine learning applications in wildfire science and management. *Env. Rev.* **28**, 478–505.

Jiang, S., Yao, W. and Heurich, M. 2019 Dead wood detection based on semantic segmentation of VHR aerial CIR imagery using optimized FCN-Densenet. *Int. Arch. Photogramm. Remote. Sens. Spat. Inf. Sci.* **42**, 127–133.

Jin, C., Liu, X. and Gao, P. 2004 An intelligent simulation method based on artificial neural network for container yard operation. *Lect. Notes Comput. Sci* **3174**, 904–911.

Kangas, A., Astrup, R., Breidenbach, J., Fridman, J., Gobakken, T., Korhonen, K.T., *et al.* 2018 Remote sensing and forest inventories in Nordic countries–roadmap for the future. *Scand. J. For. Res.* **33**, 397–412.

Kangas, A., Rasinmäki, J., Eyvindson, K. and Chambers, P. 2015 A mobile phone application for the collection of opinion data for forest planning purposes. *Env. Manage.* **55**, 961–971.

Kangas, A., Räty, M., Korhonen, K.T., Vauhkonen, J. and Packalen, T. 2019 Catering information needs from global to local scales-potential and challenges with national forest inventories. *Forests* **10**. 10.3390/f10090800.

Kansanen, K., Vauhkonen, J., Lähivaara, T. and Mehtätalo, L. 2016 Stand density estimators based on individual tree detection and stochastic geometry. *Can. J. For. Res.* **46**, 1359–1366.

Kansanen, K., Vauhkonen, J., Lähivaara, T., Seppänen, A., Maltamo, M. and Mehtätalo, L. 2019 Estimating forest stand density and structure using Bayesian individual tree detection, stochastic geometry, and distribution matching. *ISPRS J. Photogramm. Remote Sens.* **152**, 66–78.

Kattenborn, T., Leitloff, J., Schiefer, F. and Hinz, S. 2021 Review on Convolutional Neural Networks (CNN) in vegetation remote sensing. *ISPRS J. Photogramm. Remote Sens.* **173**, 24–49.

Koch, B., Kattenborn, T., Straub, C. and Vauhkonen, J. 2014 Segmentation of forest to tree objects. In *Forestry Applications of Airborne Laser* Scanning M. Maltamo, E. Næsset and J. Vauhkonen (eds). Springer, Dordrecht, Managing Forest Ecosystems **27**. 10.1007/978-94-017-8663-8_5

Krizhevsky, A., Sutskever, I. and Hinton, G.E. 2012 Imagenet classification with deep convolutional neural networks. *Adv. Neural Inf. Proc. Syst.* **1**, 1097–1105.

Laguna, M. and Marti, R. 2002 Neural network prediction in a system for optimizing simulations. *IIE Trans.* **34**, 273–282.

Lähivaara, T., Seppänen, A., Kaipio, J.P., Vauhkonen, J., Korhonen, L., Tokola, T., *et al.* 2014 Bayesian approach to tree detection based on airborne laser scanning data. *IEEE Trans. Geosci. Remote Sens.* **52**, 2690–2699.

Lahti, K.M., Heikkinen, M., Juslén, A. and Schulman, L. 2021 Tackling data quality challenges in the Finnish Biodiversity Information Facility (FinBIF). *Biodiv. Inf. Sci. Standard* **5**. 10.3897/biss.5.75559.

LeCun, Y., Bengio, Y. and Hinton, G. 2015 Deep learning. *Nature* **521**, 436–444.

LeCun, Y., Boser, B., Denker, J.S., Henderson, D., Howard, R.E., Hubbard, W., *et al*. 1989 Backpropagation applied to handwritten zip code recognition. *Neural Comput*. **1**, 541–551.

Li, A., Lu, Z., Wang, L., Xiang, T. and Wen, J.R. 2017 Zero-shot scene classification for high spatial resolution remote sensing images. *IEEE Trans. Geosci. Remote Sens.* **55**, 4157–4167.

Lindberg, E. and Holmgren, J. 2017 Individual tree crown methods for 3D data from remote sensing. *Cur. For. Rep.* **3**, 19–31.

Litjens, G., Kooi, T., Bejnordi, B.E., Arindra, A., Setio, A., Ciompi, F., *et al*. 2017 A survey on deep learning in medical image analysis. *Medic. Image Anal*. **42**, 60–88.

Liu, H. 2020 Robot systems for rail transit applications. *Elsevier* . 10.1016/C2019-0-04615-8.

Liu, J., Wang, X. and Wang, T. 2019 Classification of tree species and stock volume estimation in ground forest images using deep learning. *Comput. Electron. Agr*. **166**. 10.1016/j.compag.2019.105012.

Ma, L., Liu, Y., Zhang, X., Ye, Y., Yin, G. and Johnson, B.A. 2019 Deep learning in remote sensing applications: a meta-analysis and review. *ISPRS J. Photogramm. Remote Sens.* **152**, 166–177.

Marrs, J. and Ni-Meister, W. 2019 Machine learning techniques for tree species classification using co-registered LiDAR and hyperspectral data. *Remote Sens*. **11**. 10.3390/rs11070819.

Martins, J., Junior, J.M., Menezes, G., Pistori, H., Santaana, D. and Goncalves, W. 2019 Image segmentation and classification with SLIC Superpixel and convolutional neural network in forest context. *IEEE Int. Geosci. Remote Sens. Symp.* **2019**, 6543–6546.

Malo, P., Tahvonen, O., Suominen, A., Back, P. and Viitasaari, L. 2021 Reinforcement learning in optimizing forest management. *Can. J. For. Res., in press* . 10.1139/cjfr-2020-0447.

Maltamo, M., Peuhkurinen, J., Malinen, J., Vauhkonen, J., Packalén, P. and Tokola, T. 2009 Predicting tree attributes and quality characteristics of scots pine using airborne laser scanning data. *Silva Fenn.* **43**, 507–521.

Mäyrä, J., Keski-Saari, S., Kivinen, S., Tanhuanpää, T., Hurskainen, P., Kullberg, P., *et al*. 2021 Tree species classification from airborne hyperspectral and LiDAR data using 3D convolutional neural networks. *Remote Sens. Environ*. **256**. 10.1016/j.rse.2021.112322.

Mehtätalo, L., de-Miguel, S. and Gregoire, T.G. 2015 Modeling height-diameter curves for prediction. *Can. J. For. Res*. **45**, 826–837.

Mizoguchi, T., Ishii, A., Nakamura, H., Inoue, T. and Takamatsu, H. 2017 Lidar-based individual tree species classification using convolutional neural network. *Proc. Videometrics, Range Imag. Appl.* **10332**. 10.1117/12.2270123.

Mohamedou, C., Korhonen, L., Eerikaïnen, K. and Tokola, T. 2019 Using LiDAR-modified topographic wetness index, terrain attributes with leaf area index to improve a single-tree growth model in south-eastern Finland. *Forestry* **92**, 253–263.

Mononen, L., Auvinen, A.P., Packalen, P., Virkkala, R., Valbuena, R., Bohlin, I., *et al*. 2018 Usability of citizen science observations together with airborne laser scanning data in determining the habitat preferences of forest birds. *For. Ecol. Manag.* **430**, 498–508.

Mou, L., Bruzzone, L. and Zhu, X.X. 2019 Learning spectral-spatialoral features via a recurrent convolutional neural network for change detection in multispectral imagery. *IEEE Trans. Geosci. Remote Sens.* **57**, 924–935.

Mubin, N.A., Nadarajoo, E., Shafri, H.Z.M. and Hamedianfar, A. 2019 Young and mature oil palm tree detection and counting using convolutional neural network deep learning method. *Int. J. Remote Sens.* **40**, 7500–7515.

Müller, F., Jaeger, D. and Hanewinkel, M. 2019 Digitization in wood supply – a review on how industry 4.0 will change the forest value chain. *Comput. Electron. Agr*. **162**, 206–218.

Narine, L.L., Popescu, S.C. and Malambo, L. 2019 Synergy of ICESat-2 and landsat for mapping forest aboveground biomass with deep learning. *Remote Sens*. **11**. 10.3390/rs11121503.

Niemi, M.T. and Vauhkonen, J. 2016 Extracting canopy surface texture from airborne laser scanning data for the supervised and unsupervised prediction of area-based forest characteristics. *Remote Sens.* **8**. 10.3390/rs8070582.

Niska, H., Skön, J.P., Packalén, P., Tokola, T., Maltamo, M. and Kolehmainen, M. 2010 Neural networks for the prediction of species-specific plot volumes using airborne laser scanning and aerial photographs. *IEEE Trans. Geosci. Remote Sens.* **48**, 1076–1085.

Nuutinen, T., Berger, F., Karjalainen, A., Lempinen, R., Maltamo, M. and Siitonen, M. 2011 Request-driven generation of calculation chains for adaptive forest analysis. *Scand. J. For. Res.* **26**, 2–10.

Nweke, H.F., Teh, Y.W., Al-garadi, M.A. and Alo, U.R. 2018 Deep learning algorithms for human activity recognition using mobile and wearable sensor networks: State of the art and research challenges. *Expert Syst. Appl*. **105**, 233–261.

Okewu, E., Adewole, P. and Sennaike, O. 2019 Experimental comparison of stochastic optimizers in deep learning. *Lect. Notes Comput. Sci* **11623**, 704–715.

Packalen, P., Temesgen, H. and Maltamo, M. 2012 Variable selection strategies for nearest neighbor imputation methods used in remote sensing based forest inventory. *Can. J. Remote. Sens.* **38**, 557–569.

Padarian, J., Minasny, B. and McBratney, A.B. 2019 Using deep learning for digital soil mapping: a review aided by machine learning tools. *Soil* **5**, 79–89.

Pan, S.J. and Yang, Q. 2010 A survey on transfer learning. *IEEE Trans. Knowledge Data Eng.* **22**, 1345–1359.

Pelletier, C., Webb, G.I. and Petitjean, F. 2019 Temporal convolutional neural network for the classification of satellite image time series. *Remote Sens.* **11**. 10.3390/rs11050523.

Pitkänen, T.P., Räty, M., Hyvönen, P., Korhonen, K.T. and Vauhkonen, J. 2021 Using auxiliary data to rationalize smartphone-based pre-harvest forest mensuration. *Forestry*. 10.1093/forestry/cpab039.

Pukkala, T., Vauhkonen, J., Korhonen, K.T. and Packalen, T. 2021 Self-learning growth simulator for modelling forest stand dynamics in changing conditions. *Forestry* **94**, 333–346.

Rammer, W. and Seidl, R. 2019 Harnessing deep learning in ecology: an example predicting bark beetle outbreaks. *Front. Plant Sci.* **10**. 10.3389/fpls.2019.01327.

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J. and Carvalhais, N. 2019 Deep learning and process understanding for data-driven earth system science. *Nature* **566**, 195–204.

Sačkov, I., Hlásny, T., Bucha, T. and Juriš, M. 2017 Integration of tree allometry rules to treetops detection and tree crowns delineation using airborne lidar data. *IForest Biogeosci. For*. **10**, 459–467.

Safonova, A., Tabik, S., Alcaraz-Segura, D., Rubtsov, A., Maglinets, Y. and Herrera, F. 2019 Detection of fir trees (*Abies sibirica*) damaged by the bark beetle in unmanned aerial vehicle images with deep learning. *Remote Sens*. **11**. 10.3390/rs11060643.

Salcedo-Sanz, S., Ghamisi, P., Piles, M., Werner, M., Cuadra, L., Moreno-Martínez, A., *et al*. 2020 Machine learning information fusion in earth observation: a comprehensive review of methods, *applications and data sources*. *Inf. Fusion* **22**, 480–545.

Särkkä, S., 2013 *Bayesian Filtering and Smoothing*. Cambridge University Press, ISBN 9781139344203. 10.1017/CBO9781139344203

Schiefer, F., Kattenborn, T., Frick, A., Frey, J., Schall, P., Koch, B., *et al*. 2020 Mapping forest tree species in high resolution UAV-based RGB-imagery by means of convolutional neural networks. *ISPRS J. Photogramm. Remote Sens*. **170**, 205–215.

Schmidhuber, J. 2015 Deep learning in neural networks: an overview. *Neural Netw*. **61**, 85–117.

Seidl, R., Eastaugh, C.S., Kramer, K., Maroschek, M., Reyer, C., Socha, J., *et al*. 2013 Scaling issues in forest ecosystem management and how to address them with models. *Eur. J. For. Res*. **132**, 653–666.

Shah, S.A.A., Manzoor, M.A. and Bais, A. 2020 Canopy height estimation at Landsat resolution using convolutional neural networks. *Mach. Learn. Knowledge Extract.* **2**, 23–36.

Shao, Z., Zhang, L. and Wang, L. 2017 Stacked sparse autoencoder modeling using the synergy of airborne LiDAR and satellite optical and SAR data to map Forest above-ground biomass. *IEEE J. Selected Topics Appl. Earth Obs. Remote Sens.* **10**, 5569–5582.

Shi, W., Zhang, M., Zhang, R. and Chen, S. 2020 Change detection based on artificial intelligence: state-of-the-art and challenges. *Remote Sens.* **12**. 10.3390/rs12101688.

Shorten, C. and Khoshgoftaar, T.M. 2019 A survey on image data augmentation for deep learning. *J. Big Data* **6**, 1–48.

Singh, S.A. and Majumder, S. 2020 Short and noisy electrocardiogram classification based on deep learning. In *Deep Learning for Data Analytics*, H. Das, C. Pradhan and N. Deypp (eds). Elsevier. 10.1016/B978-0-12-819764-6.00002-8.

Sothe, C., de Almeida, C.M., Schimalski, M.B., la Rosa, L.E.C., Castro, J.D.B., Feitosa, R.Q., *et al*. 2020 Comparative performance of convolutional neural network, weighted and conventional support vector machine and random forest for classifying tree species using hyperspectral and photogrammetric data. *Gisci. Remote Sens*. **57**, 369–394.

Stefán, P. 2003 *Combined Use of Reinforcement Learning and Simulated Annealing: Algorithms and Applications*. Ph.D. Thesis, University of Miskolc, Department of Mechanical Engineering, Budapest, Hungary, 119. http://phd.lib.uni-miskolc.hu/JaDoX_Portlets/documents/document_5607_section_985.pdf

Su, C., Wu, X., Tang, X. and Hu, J. 2019 Growth height prediction for the trees under overhead lines based on deep learning algorithm. *Int. Conf. Power Syst. Tech.* **2018**, 3693–3699.

Sylvain, J.D., Drolet, G. and Brown, N. 2019 Mapping dead forest cover using a deep convolutional neural network and digital aerial photography. *ISPRS J. Photogramm. Remote Sens*. **156**, 14–26.

Swetnam, T.L. and Falk, D.A. 2014 Application of metabolic scaling theory to reduce error in local maxima tree segmentation from aerial LiDAR. *For. Ecol. Manag.* **323**, 158–167.

Theodoridis, S. and Koutroumbas, K. 2008. *Pattern Recognition*. 4th edn. Academic Press, ISBN: 9780080949123

Uusitalo, J., Puustelli, A., Kivinen, V.P., Nummi, T. and Sinha, B.K. 2006 Bayesian estimation of diameter distribution during harvesting. *Silva Fenn.* **40**, 663–671.

Varvia, P., Lähivaara, T., Maltamo, M., Packalen, P. and Seppänen, A. 2019 Gaussian process regression for forest attribute estimation from airborne laser scanning data. *IEEE Trans. Geosci. Remote Sens.* **57**, 3361–3369.

Vauhkonen, J., Korpela, I., Maltamo, M. and Tokola, T. 2010 Imputation of single-tree attributes using airborne laser scanning-based height, intensity, and alpha shape metrics. *Remote Sens. Environ.* **114**, 1263–1276.

Wan, R., Mei, S., Wang, J., Liu, M. and Yang, F. 2019 Multivariate temporal convolutional network: a deep neural networks approach for multivariate time series forecasting. *Electronics* **8**. 10.3390/electronics8080876.

Windrim, L. and Bryson, M. 2019 Forest tree detection and segmentation using high resolution airborne LiDAR. *Proc. IEEE/RSJ Int. Conf. Intell. Robot. Syst.* **2019**, 3898–3904.

Wong, S.C., Gatt, A., Stamatescu, V. and McDonnell, M.D. 2016 Understanding data augmentation for classification: when to warp?. *Proc. Int. Conf. Digit. Image Comput. Tech. Appl.* **2016**, 1–6.

Zhang, L., Shao, Z., Liu, J. and Cheng, Q. 2019 Deep learning based retrieval of forest aboveground biomass from combined LiDAR and Landsat 8 data. *Remote Sens.* **11**. 10.3390/rs11121459.

Zhen, Z., Quackenbush, L.J. and Zhang, L. 2016 Trends in automatic individual tree crown detection and delineation-evolution of LiDAR data. *Remote Sens.* **8**. 10.3390/rs8040333.

Zhu, X.X., Tuia, D., Mou, L., Xia, G.S., Zhang, L., Xu, F., *et al*. 2017 Deep learning in remote sensing: a review. *IEEE Geosci. Remote Sens. Mag.* **5**, 8–36.

Zortea, M., Nery, M., Ruga, B., Carvalho, L.B. and Bastos, A.C. 2018 Oil-palm tree detection in aerial images combining deep learning classifiers. *Proc. IEEE Int. Geosci. Remote Sens. Symp.* **2018**, 657–660.

Zou, X., Cheng, M., Wang, C., Xia, Y. and Li, J. 2017 Tree classification in complex forest point clouds based on deep learning. *IEEE Geosci. Remote Sens. Lett.* **14**, 2360–2364.