ISPRS Journal of Photogrammetry and Remote Sensing 156 (2019) 14–26

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

journal homepage: www.elsevier.com/locate/isprsjprs

Mapping dead forest cover using a deep convolutional neural network and digital aerial photography

Jean-Daniel Sylvain*, Guillaume Drolet, Nicolas Brown

Direction de la recherche forestière, Ministère des Forêts, de la Faune et des Parcs du Québec, 2700 rue Einstein, Québec, Québec G1P 3W8, Canada

ARTICLE INFO

Keywords:
Remote sensing
Tree mortality
Machine learning
Deep learning
Convolutional neural network
Ensemble learning

ABSTRACT

Tree mortality is an important forest ecosystem variable having uses in many applications such as forest health assessment, modelling stand dynamics and productivity, or planning wood harvesting operations. Because tree mortality is a spatially and temporally erratic process, rates and spatial patterns of tree mortality are difficult to estimate with traditional inventory methods. Remote sensing imagery has the potential to detect tree mortality at spatial scales required for accurately characterizing this process (e.g., landscape, region). Many efforts have been made in this sense, mostly using pixel- or object-based methods. In this study, we explored the potential of deep Convolutional Neural Networks (CNNs) to detect and map tree health status and functional type over entire regions. To do this, we built a database of around 290,000 photo-interpreted trees that served to extract and label image windows from 20 cm-resolution digital aerial images, for use in CNN training and evaluation. In this process, we also evaluated the effect of window size and spectral channel selection on classification accuracy, and we assessed if multiple realizations of a CNN, generated using different weight initializations, can be aggregated to provide more robust predictions. Finally, we extended our model with 5 additional classes to account for the diversity of landcovers found in our study area. When predicting tree health status only (live or dead), we obtained test accuracies of up to 94%, and up to 86% when predicting functional type only (broadleaf or needleleaf). Channel selection had a limited impact on overall classification accuracy, while window size increased the ability of the CNNs to predict plant functional type. The aggregation of multiple realizations of a CNN allowed us to avoid the selection of suboptimal models and help to remove much of the speckle effect when predicting on new aerial images. Test accuracies of plant functional type and health status were not affected in the extended model and were all above 95% for the 5 extra classes. Our results demonstrate the robustness of the CNN for between-scene variations in aerial photography and also suggest that this approach can be applied at operational level to map tree mortality across extensive territories.

1. Introduction

Tree mortality is an important ecological process playing a critical role in determining the structure, composition and productivity of forest ecosystems (Caspersen et al., 2011; Franklin et al., 1987). Tree death results from a range of biophysical, climatic or anthropogenic factors which, alone or in combination, can contribute to killing a tree. Under typical conditions, tree mortality will occur mostly as quasi-random events in which dead trees will be found isolated and diffused across the landscape (Larson et al., 2015; Hurst et al., 2012). Under more intense environmental constraints (e.g. fires, insect, drought), tree mortality may increase above background levels and start to exhibit more structured patterns in which dead trees are found in isolated or extensive patches scattered across the land (Clyatt et al., 2016). In all cases, the assessment of tree mortality over large areas (regions, landscape) is a difficult task, yet it is required in order to improve our understanding of this important ecological process and, ultimately, to improve sustainable forest management practices.

Traditionally, the assessment of tree mortality has been mostly limited to field and aerial survey approaches. In field surveying methods, the proportion of dead trees is usually derived from random sampling of plots within which live and dead trees are recorded. Mortality estimates are then reported at the plot level, or aggregated at larger spatial levels (e.g. stand, ecoregion). The level of representativity of mortality estimates from this type of approach depends on the design and intensity of field sampling. In most field measurements campaign, the dimensions of the sampling unit is small (< 1 ha) and the number of replicates is not suitable for obtaining a systematic portrait of the

* Corresponding author.
E-mail address: jean-daniel.sylvain@mffp.gouv.qc.ca (J.-D. Sylvain).

https://doi.org/10.1016/j.isprsjprs.2019.07.010
Received 23 December 2018; Received in revised form 25 July 2019; Accepted 27 July 2019
Available online 02 August 2019

spatial distribution of mortality over the sampling domain. Furthermore, in field measurements the spatial coordinates of trees (live and dead) is seldom recorded, preventing detailed studies of tree mortality spatial patterns or of spatial correlations with environmental variables. Aerial survey approaches, on the other hand, allow to obtain a spatially-continuous portrait of tree mortality over large areas, potentially providing useful information on the mechanisms involved in tree mortality (Breshears et al., 2009). Aerial photo interpretation (API) is a method that is well-suited to obtain individual tree attributes and coordinates. However, API is a resource-consuming approach, which prohibits its application at regional and continental levels.

To overcome the limitations of field and aerial surveying methods, previous studies have combined field and remote sensing observations to detect tree mortality and assess its spatial distribution. For example, the use of high spatial resolution multispectral imagery for detecting dead trees has been the focus of many recent studies (Kellner and Hubbell, 2017; Larson et al., 2015; Van Gunst et al., 2016; Wang et al., 2016). Despite all the progress made in the detection and assessment of tree mortality, there still exists no operational method that rely solely on remote sensing data to detect and map tree mortality over large regions (e.g. 100–1000 km^2) (Latifi et al., 2018) The absence of such methods can be partly explained by the large between-scene variations (e.g., due to differences in acquisition periods, atmospheric conditions, etc.) inherent to large datasets of remote sensing scenes required to map extensive territories (Gueguen and Hamid, 2015). Furthermore, most standard classification methods assume that image pixels follow well-defined statistical distributions; this is rarely true in practice (Olson, 2009; Olson and Ma, 1989).

Deep convolutional neural networks (CNNs) are deep learning models which have achieved unprecedented performance in object recognition and classification tasks, with applications in fields such as medicine, self-driving cars, image search, or mapping. The great performance of CNNs is due to their ability to enhance the shape, texture and spatial relationships present in images, and to use that information to detect generic structures in new images (Szegedy et al., 2014; Simonyan and Zisserman, 2015; He et al., 2015; Huang et al., 2016; Russakovsky et al., 2015; Zoph et al., 2017). This capacity results mainly from the application of sequential convolutional filters that extract general and local image patterns. Extracted patterns (features) are used to condition a neural network that is trained to associate learned features with the properties of target classes (Simonyan and Zisserman, 2015; Szegedy et al., 2014). Recent work demonstrated that accuracy levels achieved by CNNs far exceed those achieved by traditional classification methods (Audebert et al., 2018; Chen et al., 2018; Marmanis et al., 2016). Therefore, the use of CNNs could provide a reliable and objective way to assess tree mortality over extensive territories. Moreover, the combination of API and CNNs would improve the efficiency and reproducibility of API projects, which in turn would enhance our ability to quantify tree mortality and monitor it over time.

The fitting procedure of a neural network involves the introduction of a stochastic component during weight initialization. This stochastic component may lead to different final weights and affect the classification from one iteration of a model to another (Kourentzes and Petropoulos, 2016; Marmanis et al., 2016) Few studies in machine learning studied the effect of this stochastic component on CNN accuracy although it may alter prediction stability and accuracy. To overcome this limitation and improve the accuracy and robustness of model predictions, it has been proposed to use ensemble learning, an approach by which a neural network is trained multiple times to generate multiple predictions (one from each trained model) for a given observation. The resulting predictions are then aggregated to provide a deterministic estimation of the real values (e.g. probabilities), thereby reducing the uncertainty associated with the stochastic component of the neural network (Kourentzes and Petropoulos, 2016; Brochero et al., 2015; Marmanis et al., 2016). Such an approach would also lead to better assessment of the uncertainty in the predictions; e.g. regions showing low adequation between predictions from different iterations of a model would be interpreted as more uncertain.

In this study, we aimed to evaluate the potential of CNNs to obtain a spatially-continuous pixel based coverage of forest trees functional type (broadleaf, needleleaf) and health status (dead, live) using high-resolution digital aerial photography acquired under a broad range of acquisition conditions (e.g. illumination, topography). To achieve this goal, we 1) built a CNN based on a compact version of the VGG model architecture and used it to predict tree health status (live, dead) and functional type (broadleaf/needleleaf) using multispectral orthoimages, 2) evaluated how window size and spectral channels combinations affect prediction accuracy, and 3) assessed the effect of the stochastic component of CNNs on prediction accuracy. Finally, we extended our 4-class model (live needleleaf, live broadleaf, dead needleleaf, dead broadleaf) to include 5 more land cover classes (water, road, wetland, timber harvest, building) and we applied it to a mosaic of 43 aerial orthoimages to assess the ability of the CNN to correctly classify pixels from images acquired under a range of acquisition conditions.

2. Data and methods

2.1. Study area

2.1.1. Vegetation type

The study area is located in the south-central part of the Province of Quebec, Canada, and encompasses three vegetation sub-zones: the temperate deciduous, the mixed-wood and the boreal forests (Fig. 1). Temperate deciduous forests are mostly located in the southern part of the study area and are characterized by the presence of tree species such as sugar maple (*Acer sacharrum* (Marsh.)), yellow birch (*Betula alleghaniensis* (Britt.)), and trembling aspen (*Populus tremuloides* Michx.)). The mixed forest mostly cover the central part of the area and are dominated by mixed stands composed, for the most part, of balsam fir (*Abies balasamea* (L.) Mill), white birch (*Betula papyrifera* (Marsh.)) and trembling aspen. Boreal forests occupy the northern part of the study area and are dominated by black spruce (*Picea mariana* (Mill.)), jack pine (*Pinus banksiana* (Lamb.)) and white birch. The study area was selected because of 1) its diversity in forest structural attributes, which results mostly from forest management activities, and 2) the occurrence of important tree species for the Quebec forest industry.

2.2. Orthoimages

2.2.1. Acquisition and metadata

We used a set of 990 very high resolution multispectral digital aerial photos acquired over the study area between July 2, 2007 and August 21, 2007. This type of imagery is routinely acquired by the Government of Quebec for forest mapping and planning purposes. Aerial photos were acquired with a Vexcel UltraCamD large format aerial camera designed for co-located acquisition of panchromatic and multispectral images. For our analyses, we used digital count values (8 bits) in the four spectral channels (red (R), green (G), blue (B), near-infrared (I)), that were pan-sharpened and stacked into multi-channel, 20-cm pixel resolution GeoTiff images. Finally, we used interior and exterior camera parameters provided by the image supplier, a triangulated irregular network (TIN) elevation dataset and the Summit Evolution photogrammetric workstation to create multispectral RGBI orthoimages.

2.3. Tree database

2.3.1. Labelled-images

Orthoimages were used as the main data source for building a database of labelled-trees, to be used for calibration and validation of the CNNs. We started by randomly selecting 315 undisturbed forest polygons from a set of candidate polygons distributed over the study area. Candidate polygons were extracted from the 2007 1:20 k Quebec forest

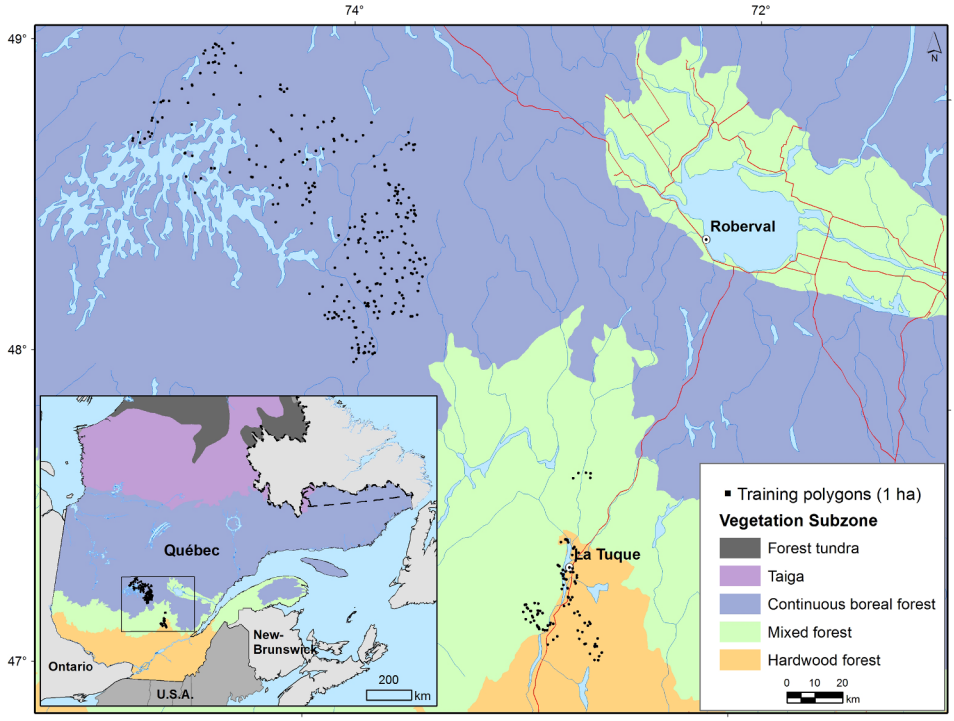

Fig. 1. Map of the study area. Black squares show the locations of 1-ha polygons used for photo-interpretation of tree positions (x, y), functional type (needleleaf, broadleaf) and heath status (live, dead).

map (Direction des inventaires forestiers, 2009) using the following criteria: years since last disturbance (> 25), stand age class (> 30 years old), and forest cover type (coniferous, mixed, deciduous). For each selected polygon, we defined a 1-ha square area (100 m × 100 m) around its centroid, thus defining boundaries for image interpretation. Using 3D stereo rendering, a senior photointerpreter characterized all trees taller than 7 m and located within the 1-ha polygons. Each tree was characterized in terms of the following attributes: plant functional type (PFT: broadleaf, needleleaf), health status (HS: dead, live), and location (X and Y coordinates) (Fig. 2). Tree locations were represented by the crown center of mass. Slight differences could occur between a point coordinates and its tree true center of mass. In our study, this was not considered critical since the data augmentation scheme we used at a later step (Section 2.4) applied random translations to the tree coordinates. In fact, in our case the use of translations resulted in increased prediction accuracies. Defoliated trees were classified into live or dead trees according to a 50% defoliation threshold, meaning that trees with a percentage of crown defoliation greater than or equal to 50% were labelled as dead. Since highly defoliated trees are spectrally very similar to dead trees, we defined that a 50% defoliation threshold was representative of dead forest cover. All along the photo-interpretation process, an independent photo-interpreter with 15 years of professional experience randomly selected 10% of the 1-ha polygons for quality checking. Verified polygons in which the proportion of trees with errors (localization, plant functional type, health status, and defoliation level) exceeded 5% were sent back for corrections.

The interpretation of all images resulted in a database of more than 290,000 georeferenced trees, each labelled with one of 4 classes: live broadleaf (LB), dead broadleaf (DB), live needleleaf (LN), and dead needleaf (DN) (Table 1). Finally, for each tree in the database we extracted image values in the four spectral channels using a two dimensional window (patch) centered around the tree spatial coordinates (Fig. 3). The resulting patches, which we refer to as “labelled-images” in this paper, were used as inputs for trainining and testing the CNNs. Labelled-image dimensions varied depending on the tested model configuration (see Sensitivity Analyses section below).

2.3.2. Extension of the base model to 9 classes

To increase the potential use of our CNNs (e.g. by forest planners), we added 5 land cover classes to the training dataset: wetland, timber harvest, water, road and building. These classes were chosen because they represent, altogether, the majority of land cover classes found in the study area. Training samples for these classes were created by first selecting, from the same forest map that we used for building the tree database, any polygon labelled as one of our 5 land use classes. In a second step, we visually inspected all extracted polygons by overlaying them over the aerial photos and evaluating if their image content was representative of its associated land cover class. When required, we edited polygon geometries to maximize their spatial and spectral homogeneity. The edition of polygon geometries also aims at removing geometrical alignment errors and help increasing the representativeness of the elements associated with each land use class. Wetland polygons were manually edited to keep only open areas that were easily distinguishable from adjacent forest cover. Roads and water features were edited in the database using SQL queries. For road segments, we first applied a 5-m buffer that converted linear features into polygons, from which we manually removed all regions covered by vegetation. For timber harvest (e.g. clear-cut), we retained only polygons in which harvest occurred at most 4 years prior to image acquisition. Surface water bodies were shrunk using a 10-m interior buffer to remove potential bordering vegetation or beaches. Buildings were all manually positioned using aerial photography. The edition of polyon geometries aimed at removing geometrical alignment errors and also at increasing the representativeness of the end-member for each land cover. Finally, we generated randomly distributed points across all polygons and we sampled 20,000 points for each land use class. Using the same approach used for trees, we extracted image values from RGBI multispectral channels using two-dimensional windows centered around the class points. These additional labelled-images (20,000 × 5 classes = 100,000 labelled-images) were merged with tree labelled-images to create an extended database that we used to train the 9-class model (VGG16S-RGBI-41px-9cl).

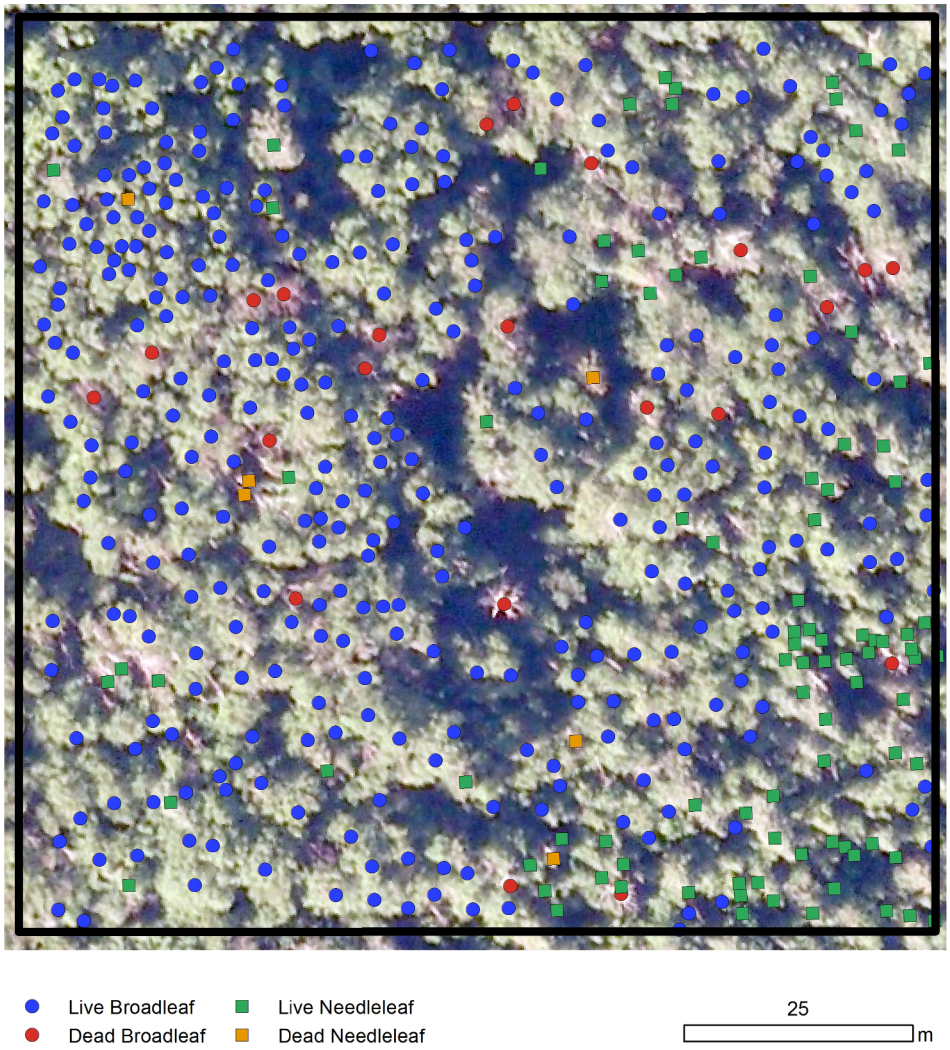

Fig. 2. Example of a 1-ha square polygon (black contour line) showing the spatial distribution of tree functional type and health status classes (represented by symbols and colors, respectively) overlayed on a true color (RGB) high-resolution orthoimage. These points were used to extract RGBI count values from orthoimages and to train the CNNs. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 1
Number of trees in the original database, by health status and plant functional type. Numbers in parentheses represent the proportion of database records labelled with a given combination of plant functional type and health status classes.

	Health Status		
Plant Functional Type	Live	Dead	$Total_{PFT}$
Broadleaf	73,356 (0.25)	1039 (< 0.01)	73,795
Needleaf	218,190 (0.74)	3561 (0.01)	218,279
$Total_{HS}$	287,474	4600	292,074

2.4. Data preprocessing

2.4.1. Data splitting

The labelled-images database was split into calibration (60%), validation (10%) and test (30%) datasets. The calibration dataset was used to train the model. To avoid overfitting and assess the expected prediction error, we used the validation dataset to track model performance during the training process. The test dataset, which is completely independent from the calibration process, was used to assess the generalization error of the model. The assessment of performance on an independent dataset is extremely important in practice, as it provides a measure of the quality of the expected predictions on a dataset that was never previously seen by the model. We then assessed the optimism of the final model by calculating the difference between expected and generalization errors (Hastie et al., 2009).

2.4.2. Spatial constraints

The validation and test datasets were created by randomly sampling an equal amount of labelled-images in each class. To avoid overfitting and ensure a good assessment of the generalisation error, we imposed a spatial constraint that eliminated from the calibration dataset any tree that was located within 8 m of a tree contained in the validation or test datasets. The 8 m distance is required to ensure independence between train and validation samples. we removed from the training dataset all labelled-images that overlapped 5% or more of a validation or test window. The minimal distance used to subset validation and test dataset ensured that a labelled-image used for training the CNN could not have more than 5% of its area occupied by pixels from a validation or test labelled-image. To guarantee a 0% overlap would have required a 45% increase in the minimum distance between trees, which would have led to losing many additional training samples, including some in the minority classes. This operation resulted in a 4% decrease in the number of labelled-images from the minority class (DN) in the calibration dataset.

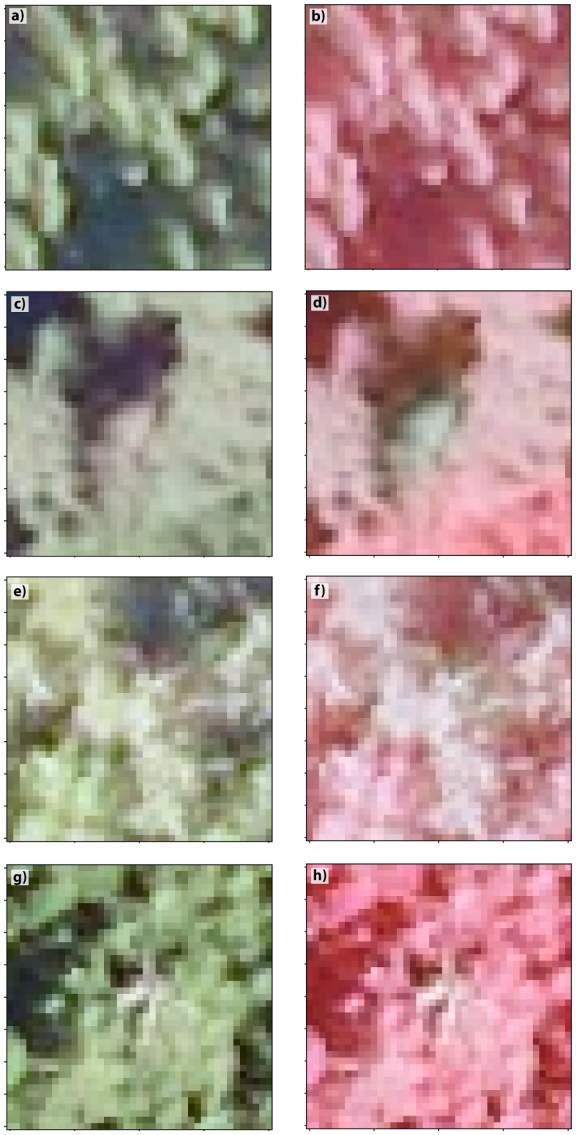

Fig. 3. Examples of 41 × 41-pixel labelled-images. True (RGB) and false (IRG) color images are shown on the left and right hand side panels, respectively. Each row shows images for one of the base model 4 classes: live needleleaf (LN, a-b), dead needleleaf (DN, c-d), live broadleaf (LB, e-f), and dead broadleaf (DB, g-h).

2.4.3. Undersampling

The resulting calibration dataset was largely dominated by trees from the LN (170,278) and LB (63,422) classes (versus dead tree classes, DN (2672) and DB (850)), which resulted in an extremely large class imbalance. In our study, the small number of samples from minority classes resulted from the zero-inflated distribution of highly defoliated and dead trees, whose occurrences are considered as rare events outside of catastrophic events such as insect outbreaks or large-scale diseases (Franklin et al., 1987). To obtain a balanced dataset, Buda et al. (2017) recommend oversampling minority classes until they meet the number of samples from the majority class. In this study, oversampling would have led to an important duplication of dead trees and to an increase in computation time during training. For these reasons, we chose instead to undersample the number of calibration labelled-images from LN and LB classes to 40,000 using random selection. To avoid the effect of a large class imbalance on performance assessment, we also randomly undersampled all but the minority class of the validation and test datasets so that they equalled the number of samples from the minority classes (i.e. DB_{val} = 85 and DB_{test} = 255).

2.4.4. Data augmentation

Due to the heterogeneous spatial distribution of tree mortality and, to a lesser extent, to our sampling plan, the number of labelled-images from dead tree classes (DN, DB) in our database was limited compared to live tree classes. Class imbalances can result in predictions that are biased toward majority classes and also lead to poor generalization of the models. In our case, this could have led to reduced proportions of dead forest cover when applying the CNN over large areas. To increase the ability of our CNNs to recognize dead forest cover in new images, we performed data augmentation on our training dataset (Krizhevsky et al., 2012). Data augmentation was realized by randomly applying the following transformations on labelled-images from the undersampled classes (LB, DN, DB): shift by 0.4 meters horizontally (left or right) or vertically (up or down), flip (horizontal or vertical) or rotation (90° increments), or a combination of any two transformations from this list. In total, these transformations yielded to a potential of 12 unique samples per labelled-image. Transformations were applied to the undersampled classes by randomly selecting a labelled-image and applying a (combination of) transformation until the number of images from the target class reached 40,272. It is important to note that data augmentation was applied only to labelled-images from the calibration dataset; validation and test datasets comprised only original samples.

2.4.5. Normalisation

Finally, we standardized pixel values in all labelled-images by subtracting the calibration dataset mean and dividing by its standard deviation. The final calibration, validation and test datasets contained 161,088, 340 and 1020 labelled-images, respectively.

2.5. CNN Configuration

2.5.1. CNN structure

Standard CNNs use stacks of convolutional and max-pooling layers as an image feature extractor. Extracted features, which take into account the spatial context of an image object, are then usually used as inputs to a fully-connected neural network for classification. VGG (Visual Geometry Group, Oxford University, Simonyan and Zisserman (2015)) is a model architecture that has won 2nd place at the 2014 ImageNet Large Scale Visual Recognition Competition (ILSVR). Today, VGG is still considered a state-of-the-art CNN image feature extractor, although it was recently outperformed by other models such as GoogleNet (Szegedy et al., 2014), ResNet (Wu et al., 2015), DenseNet (Huang et al., 2016), and SENet (Hu et al., 2017). In this study, we used a simplified version of VGG with 16 layers (VGG16) because it is more adapted to small objects and it has shown great performance for the detection of small objects or part of object such as trees (Simonyan and Zisserman, 2015)). Preliminary analyses (not shown) using a trimmed down version of VGG16 (Table 2) allowed us to drop the last 2 conv-maxpool filter blocks without affecting prediction accuracy. This

Table 2
Comparison between VGG16 and VGG16S architectures.

Step	Number of layers	VGG16	VGG16S
Extraction	2	Conv64	Conv32
	1	Maxpool	Maxpool
	2	Conv128	Conv64
	1	Maxpool	Maxpool
	3	Conv256	Conv128
	1	Maxpool	Maxpool
	3	Conv512	
	1	Maxpool	
	3	Conv512	
	1	Maxpool	
Classification	1	Flatten	Flatten
	2	Dense 4096	Dense 512
	1	Dense (softmax)	Dense (softmax)

modification reduced by half the number of convolutional filters and reduced the number of neurons in the final classifier by a quarter. These modifications resulted in a simplified model which had 22 times fewer trainable parameters than the original VGG16, which significantly decreased GPU time. This simplified configuration, named VGG16S, is the model that we used in subsequent analyses.

2.5.2. CNN initialization

For model training, we used constant values for the hyperparameters defined in the previous section and we used VGG16 standard values for the other model parameters: dropout was set to 0.5 and L2 penalty multiplier to 5e-4. We initialized all weights using a Glorot uniform distribution (Glorot and Bengio, 2010) and we set the initial biases to zero. We used stochastic gradient descent (SGD) with Nesterov momentum of 0.9 to optimize the parameters. Apart from the last dense layer which used a softmax activation, all other layers in the model used the Rectified Linear Unit (relu) activation function. The total number of labelled-images used to train the model at each update (i.e. batch size) was set to 256. We used a time-based learning rate (lr) decay function, with lr initially set at 1e-3 and updated at each iteration according to the following formula, where 150 is the maximum number of epochs and iteration is the elapsed number of mini-batches (from 1 to a maximum of 38,400:

$$lrate(iteration) = lr * \frac{1}{1 + (lr * \frac{iteration}{150})}$$

2.5.3. Training and early stopping

We provided a maximum of 150 epochs for the algorithm to converge. In most cases, the model converged within about 100 epochs, amounting to a few hours of GPU usage per trained model. The limit of 150 epochs was never reached during model training. To avoid overfitting, the model was trained using an early stopping approach in which the training was stopped when the validation loss didn't decrease for 20 consecutive epochs and the final weights were those that provided the best overall validation accuracy. We did not apply any batch normalization (Ioffe and Szegedy, 2015) during the training process as we found it only increased training compute time with no significant gain in performance evaluation metrics.

2.6. Ensemble predictions

2.6.1. Deterministic prediction and uncertainty assessment

We hypothesized that ensemble predictions can be used to reduce some of the variability in CNN outputs that results from the stochastic component of CNNs. To evaluate the potential of ensemble predictions for increasing stability in model outputs, we trained each CNN 10 times for each combination of the studied parameters (2 window sizes × 3 band combinations = 6 CNNs. See next section). For each iteration of a CNN configuration, the model was initialized using a different random seed, which resulted in a modification of node weights initial values and, consequently, in differences in the final model. Each model iteration was then used to make predictions on the test dataset. Pixel-wise predictions from the 10 iterations were then aggregated using the modal class, i.e. the class with the highest frequency among the 10 predictions. As well as providing greater stability in CNN predictions, we also believe that ensemble predictions can be used to characterize the level of spatial uncertainty associated with predictions of specific classes and use this information to identify, for example, which type of mortality (e.g., broadleaf vs. deciduous trees, spatial patterns) is more hardly recognized by the model.

2.7. Sensitivity analyses

2.7.1. Window size

We studied the effect window size and spectral channel selection on CNN predictive accuracy. Window size defines the width and height of labelled-images, in number of pixels. Based on preliminary analyses using different window sizes (not shown), we selected the two sizes that were best adapted to the average crown diameter and area of needleleaf (4.2 m, 13.8 m^2) and broadleaf (8.2 m, 52.8 m^2) trees in the database: 21 × 21 (21px) and 41 × 41 (41px) pixels, respectively.

2.7.2. Availability of spectral channels

The choice and number of spectral channels available as model inputs depend on the camera used for the aerial survey, and this is likely to have an effect on model performance. To assess the magnitude of these effects on CNN predictions, we used different channel combinations as inputs to the CNNs and compared their respective predictive performance. We used two channel combinations that are commonly used in remote sensing-based forestry applications: true-color (RGB) and false-color (IRG) images. We also used a combination that included all 4 image channels (RGBI).

2.8. Performance assessment

The effect of parameter values (window size, channels) on model performance was assessed by comparative analyses of the global accuracy and of omission and commission errors. Global accuracy is a measure of the overall ability of the model and is defined as the number of accurate predictions relative to the total number of predictions made by the model. The omission error measures how often pixels from a given class were left out (omitted) in the classification (i.e. they were classified as other things than the class to which they belonged). Omission error is calculated as the ratio, for a given class, of the number of incorrect predictions to the total number of observations for that class. Conversely, the commission error measures how often, among all predictions made for a given class, these predictions were actually wrong. It is calculated as the ratio of the number of inaccurate predictions of a given class to the total number of times that class was predicted by the model. For the 4-class base model, we compared the performance of the CNNs globally but also when predicting HS alone, PFT alone, or all 4 combinations of HS and PFT (PFT + HS). The performance of the 9-class model was evaluated globally and for each class. All performance metrics were calculated from confusion matrices derived from predictions on the test dataset.

2.9. Inference

2.9.1. Dead forest cover mapping

To assess the effect of between-image variations due to differences in image acquisition conditions, we made predictions on a mosaic of 43 images (32 km^2) using the best model configuration as determined by sensitivity analyses. Images in the mosaic were acquired between 8:30AM and 6PM over the period from July 23 to July 26, 2007. During that period at that location (mosaic centered at 48.9 N, 74.6 W), sun zenith and azimuth angles varied between 30 and 105, and between 190 and 228 degrees, respectively. Dead forest cover maps were created using the sliding-window method to infer model classes on new orthoimages. This method extracts values in the new orthoimage at every N pixels, where each Nth pixel becomes the center of an small image to be classified by the CNN. In this study, we used a N value of 5 pixels, which resulted in 1-m resolution prediction maps.

2.9.2. Hardware and software

All CNNs and experiments in this study were implemented in the Python 3.5 language, using Tensorflow 1.4 (Martin et al., 2015) and Keras 2.1 (Chollet and Others, 2015). All our models were ran on a PC

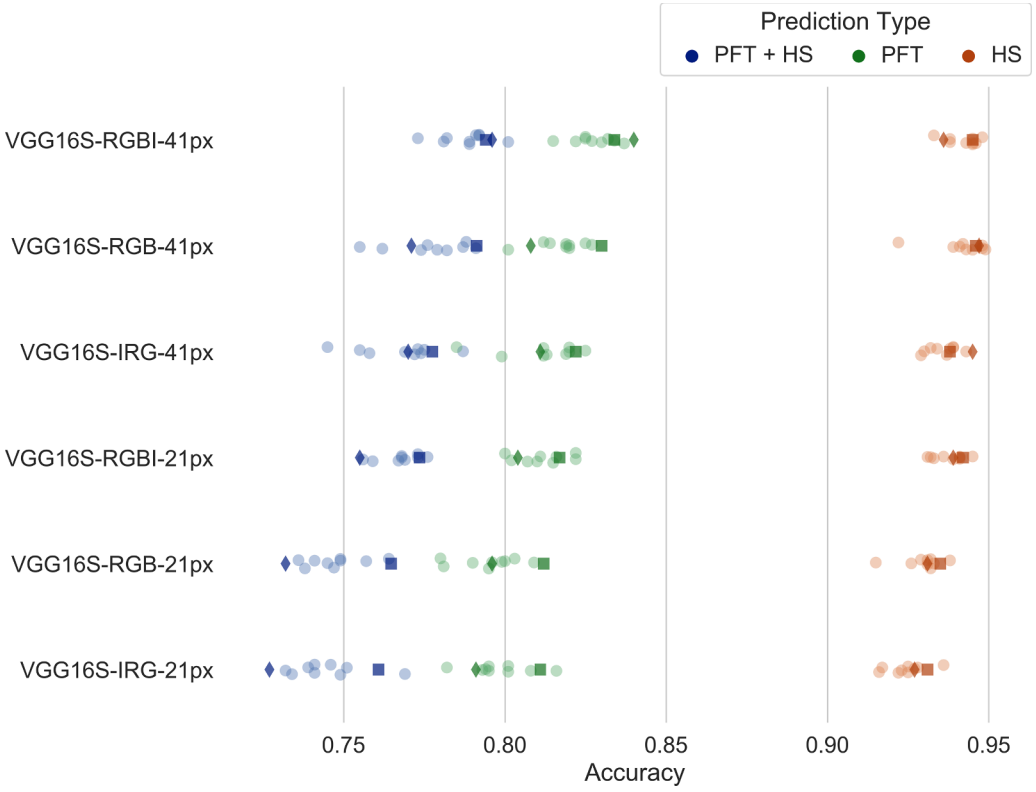

Fig. 4. Global accuracy values achieved on the test dataset, for each model configuration and prediction type. Each row shows a different model configuration. Prediction types are represented by different colors: red (plant functional type, PFT), green (health status, HS) and blue (4-class predictions, PFT + HS). Light-colored circles represent the accuracy values for 9 out of 10 model runs, each initiated with different seed value between 1 and 10. For each model configuration and prediction type, the dark diamond represents the model iteration that performed best on the validation dataset whereas the dark square represents the accuracy value derived from the confusion matrix of aggregated (modal) predictions. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

workstation equipped with a MSI Z270-A PRO motherboard, Intel Core i5-7600 CPU, 16 GB DDR4 memory and a Nvidia GTX1060 6 GB graphics card.

3. Results

3.1. Effect of the stochastic component of CNNs on model stability and performance

Fig. 4 presents the distribution of global accuracies achieved by 10 different weight initializations. Independently of model configuration and performance metrics, predictions of HS always resulted in accuracy values that were 8% to 23% higher than values obtained when predicting PFT (0.77–0.84) and PFT + HS (0.72–0.80) (Fig. 4). The distributions of accuracy values for PFT and PFT + HS also suggest higher uncertainties compared to HS predictions, as represented by their larger variances in accuracy values. This suggests that models showing lower performance are subject to greater levels of uncertainty in their predictions. In all model configurations but VGG16S-RGBI-21px, the performance of modal values outperformed the model iteration that showed the best validation accuracy during training (diamond symbol in Fig. 4). By training multiple CNNs with their own weight initialization values and using pixel-wise modal classes, we increased prediction accuracy and reduced the chance of reaching a local minimum during the training steps, thereby increasing the confidence in the classification. These results are in line with results from Kourentzes and Petropoulos (2016), Marmanis et al. (2016), who demonstrated that ensemble forecasts are more robust to misclassification. In addition, ensemble predictions also allow to map the agreement between iterations of a model, which can then be used as a tool to assess the level of uncertainty at the pixel level (Fig. 8b).

3.2. Base model performance

Table 3 shows accuracy values for predictions of PFT, HS and PFT + HS classes, for both validation and test dataset. Accuracy values were calculated using pixel-level modal predictions from the 10 model iterations presented in the previous section. Model performance was always higher on the validation dataset. However, the small differences between validation and test datasets (< 1%) suggest that our training process did not overfit our training dataset, and by extension, that the classification accuracy and the expected errors obtained from the validation dataset can be generalized on new datasets. When looking at performance on the test dataset only, CNN accuracies ranged from good (76%) to very good (95%) depending on the prediction type. The predictive performance for HS was systematically higher than for PFT or PFT + HS classes and this, regardless of the values of the configuration parameters (i.e. window size or spectral bands). Global accuracy of the best model for HS (VGG16S-RGBI-41px) was 12% and 16% higher than the accuracy achieved for PFT (83%) and PFT + HS (79%) for the same model, respectively.

Table 3
Global accuracy values calculated from predictions on the test and validation datasets, for each model configuration and prediction type. Values were calculated using aggregated predictions (pixel-wise modal class) from 10 model iterations.

	PFT		*HS*		*PFT + HS*	
	Val	*Test*	*Val*	*Test*	*Val*	*Test*
w = 41px						
RGBI	85.0	83.4	94.7	94.5	80.6	79.4
RGB	84.1	83.0	94.7	94.6	80.0	79.1
IRG	84.4	82.2	93.2	93.8	78.8	77.7
w = 21px						
RGBI	83.2	81.7	94.7	94.2	79.4	77.4
RGB	84.1	81.2	94.4	93.5	80.6	76.5
IRG	82.4	81.1	94.1	93.1	78.2	76.1

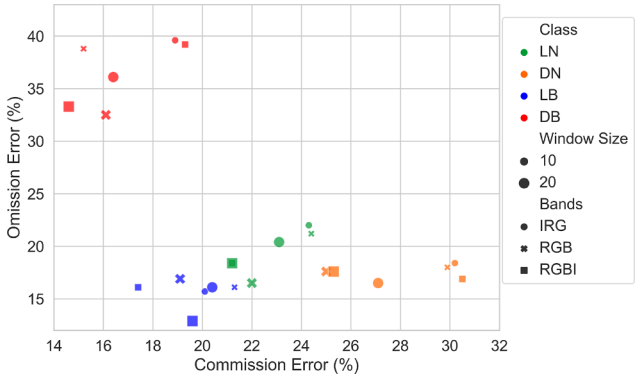

Fig. 5. Omission and commission errors (%) for each model configuration and for predictions of the 4 classes (plant functional type (HS, 2 classes), health status (HS, 2 classes)). Colored symbols show the omission and commission errors for a the different model configurations. Symbols represent channel combinations: circle = IRG, cross = RGB, and square = RGBI. Symbol sizes represent window sizes: small = 21px, large = 41px. Colors represent predicted classes (PFT + HS): green = live needleleaf (LN),orange = dead needleleaf (DN), blue = live broadleaf (LB), and red = dead broadleaf (DB). Omission and commission errors were calculated from the pixel-level aggregation (modal class), for a given model configuration, of 10 predictions on the test dataset (1 prediction = 1 of 10 iterations of a given configuration). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

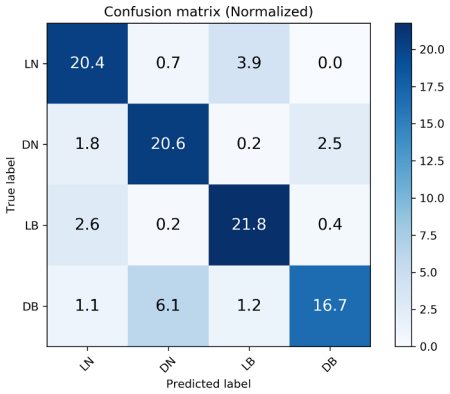

Fig. 6. Normalized confusion matrix for the aggregated (modal class) predictions on test dataset using 4-class model VGG16S-RGBI-41px. Numbers indicate the proportion of predictions for each combination of plant functional type and health status classes. The sum of the matrix equals 100%. All classes were represented by the same number of pixels (N = 255) in the test dataset.

3.3. Window size

The larger window size increased the ability of the CNNs to predict PFT (83% vs. 81% for 21px) and PFT + HS (79% vs. 77%) classes, but did not have much effect on HS classification (95% vs. 94%). It provided accuracy values equal to or higher than those from predictions by models that used the 21-pixel window, for all prediction types and channel combinations. These differences in accuracy suggest that, when using a smaller window size, CNNs cannot see the entire surface of tree crowns or do not see enough contextual information around the trees. The effect of window size may also be related to the minimum image size required by the CNN architecture which contains 3 maxpooling layers. Consequently, we recommend that, for similar applications, the window size used to extract labelled-images should be defined based on the dimensions of the largest target objects and the number of maxpooling layer. Window size should be chosen as to capture not only the full extent of the target object but also its spatial context. Window size also seemed to result in a decreased of the commission errors for predictions of dead tree classes (Fig. 5). By increasing the window size, we also increase the amount of spatial context used for making predictions. In our analysis, this led to decreases in commission errors for all classes. The dimensions and particular shape of deciduous trees required a greater window size than for needleleaf trees, as seen in the Fig. 3. Conversely, the overall effect of window size on the omission error seemed to be smaller than on commission errors, regardless of the class predicted and channel selection (Fig. 5).

3.4. Channel selection

Interestingly, the selection of spectral channels seemed to have a limited effect on global accuracies achieved by the different CNNs. Predictions were more accurate for configurations using RGBI and RGB than for those using IRG channels. Although the availability of spectral bands seemed to have a limited effect on global accuracy, the increase spectral richness of RGBI helped improving predictions for the PFT classes (Figs. 5 and 4) as well as decreasing variability in the predictions from the different model iterations (Fig. 4). The additional spectral richness provided by RGBI channels also helped to reduce ommission and commission errors for all prediction types (Fig. 5), but its impact was limited compared to that of the window size. Overall, although the effect of window size on accuracy was greater than the choice of spectral channels, the combination of the larger window with the four-channel image provided the best predictions for all prediction types (Fig. 4). Our results also show that, in our study, plant functional types were much more difficult to predict than tree health status, regardless of window size and channel selection.

3.5. Confusion matrix

Omission and commission errors calculated from the normalized confusion matrix presented in Fig. 6 were mostly related to predictions of PFT classes: 25% of DB pixels were classified as DN, 15% of LN pixels as LB, 10% of LB as LN, and 10% of DN as DB. Omission (OE) and commission (CE) errors associated with HS predictions were all under 5% except for DN/LN, that reach 7%. OEs and CEs associated with both PFT and HS predictions (e.g., DB pixels classified as LN) were negligible and varied between 0 and 1.5%. Overall, the live broadleaf class showed the best classification performance, followed by the live needleleaf class (OE: 7%, CE14%) while the dead broadleaf class showed the worst prediction successes (OE: 4%, CE: 35%).

3.6. CNN extended to 9 classes

Fig. 7 provides an overview of the accuracy of the 9-class model (VGG16S-RGBI-41px-9cl). The 9-class model achieved a global accuracy of 90%. The additional classes were predicted with a high level of accuracy, which varied between 95% and 98%. The introduction of new classes had a minor impact on the global accuracy and, generally, increased the ability of the model to predict all PFT + HS classes (DN (79% to 85%), LB (87% to 91%) and DB (67% to 74%)) but LN (82% to 79%). Again, the omission and commission errors were mostly associated with PFT classes: 15% of the LN pixels from the test dataset were classified as LB, while 18% of DB pixels were classified as DN. Based on the omission and commission errors derived from the matrix in Fig. 7, the LB class was the most easily classified of the 4 tree classes in the 9-class model (OE: 8%, CE: 19%), followed by LN (OE: 16%, CE: 19%), DN (OE: 19%, CE: 16%) and DB (OE: 9% and CE: 31%).

3.7. Mapping dead forest cover

3.7.1. Landscape scale patterns: 4-class model

Fig. 8 presents the result of an ensemble prediction for a zone representative of the mixed forests in our study area. The map results

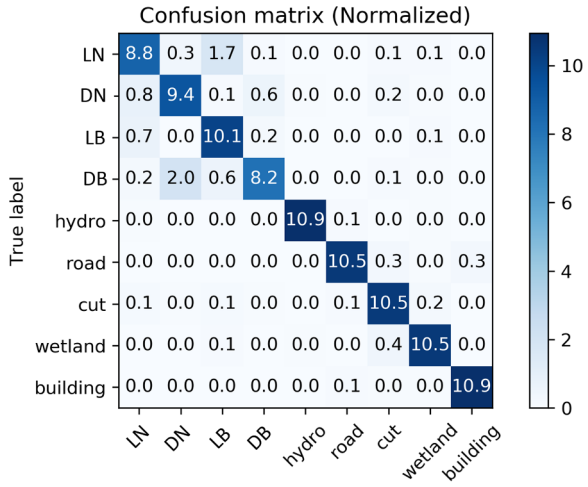

Fig. 7. Confusion matrix for the pixel-level aggregated (modal class) predictions made on the test dataset by the 9-class model (VGG16S-RGBI-41px-9cl). Numbers indicate the proportion of predictions for each combination of plant functional type, health status and land cover classes. The sum of the matrix equals 100%. All classes were represented by the same number of pixels (N = 255) in the test dataset.

from the pixel-wise aggregation (modal class) of predictions from 10 iterations of the VGG16S-RGBI-41px-9cl model applied to one orthoimage of our dataset. The figure allows visual comparisons between the original orthoimage (Fig. 8a), the level of agreement among classifiers (Fig. 8b), the predicted classes for iteration 1 of the model (seed = 1) (Fig. 8c), and the predicted classes from the ensemble predictions (Fig. 8d.). The area delimited by the orthoimage is mainly dominated by LN and LB pixels, and in less proportion by DB and DN pixels (Fig. 8). The general pattern of PFT classes show that classifications resulting from model iteration 1 and from ensemble predictions exhibit similar patterns. Both classifications provided a detailed view of the distribution of PFT + HS classes that is in agreement with the original image. Nevertheless, the map derived from the ensemble predictions shows a smoother spatial pattern and contains less speckles compared to the map from iteration 1. Consequently, ensemble predictions delineate more precisely the transitions between PFT + HS classes compared to the classification from iteration 1. The uncertainty associated with the speckle effect seems to be in accordance with the level of agreement among classifiers; the agreement is lower (red to yellow pixels) in the transition zone and in the areas dominated by broadleaf trees (Fig. 8b). Conversely, the agreement is generally higher in the needleleaf-dominated areas (yellow to blue pixels). These observations are in line with the confusion matrix in Fig. 6, which shows greater omission errors for broadleaf than for needleleaf pixels.

3.7.2. Stand level patterns: 4-class model

The upper right insets in panels of Fig. 8 highlight classification results for a small region of interest within a forest stand. The inset in panels a, c and d show that all dead trees were classified as such by the model. Again, we see that, in predictions from the iteration 1 model, many pixels at the edge of DN pixels were classified as broadleaf trees (commission error). These errors are not present in the map from the ensemble predictions (Fig. 8d). It is also interesting to note that some predictions by iteration 1 of the model were later overturned by the ensemble predictions: in the middle right section of the inset, one tree

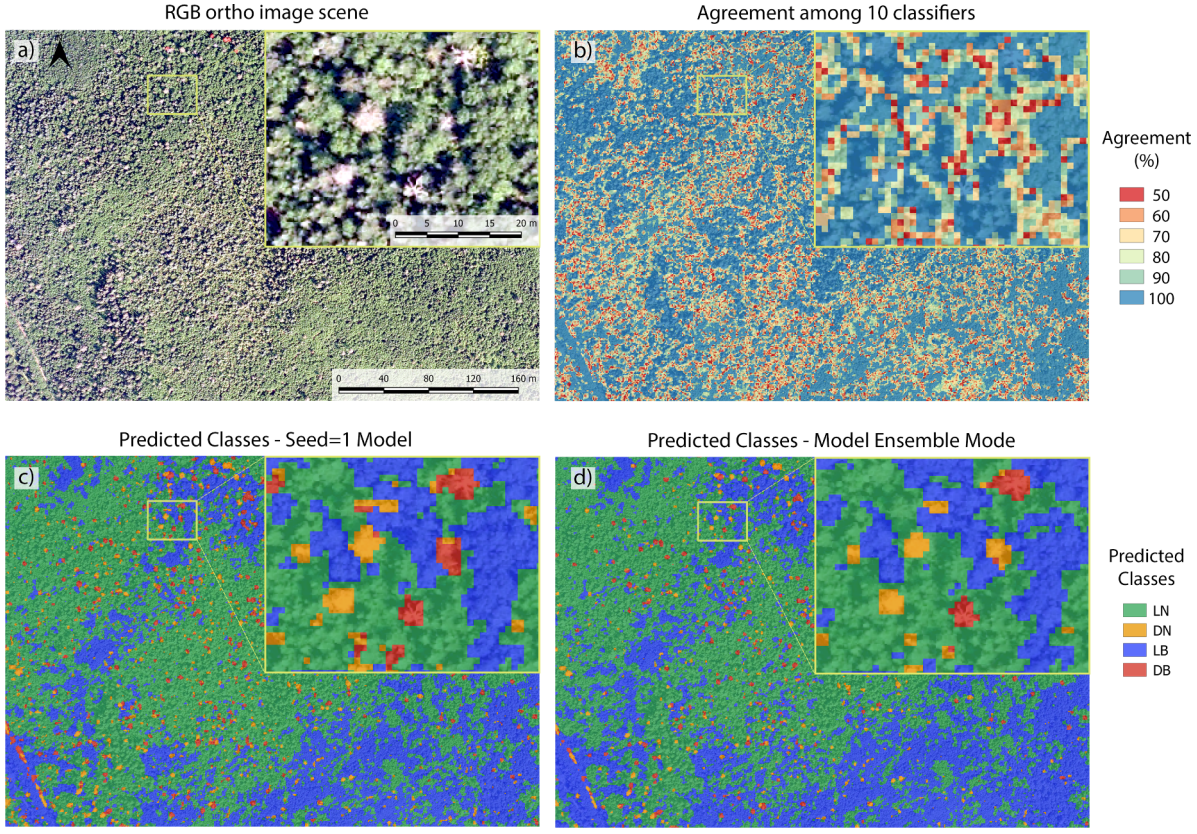

Fig. 8. Prediction of plant functionnal type (PFT) and health status (HS) on an RGBI orthoimages using VGG16S-RGBI-41px. a) Original RGB orthoimage, b) agreement among 10 classifiers, c) the predictions obtained using only one seed and 4) the predictions obtained using the aggregated value (modal value). In panel c and d, the color describes the combination PFT + HS: live broadleaf (LB, blue), live needleleaf (LN, green), dead broadleaf (DB, red), dead needleleaf (DN, orange). In panel b, the legend indicates the level of agreement among the 10 classifiers: red (low level, 0–50%), yellow (medium, 50–75%) and blue (high, 75–100%). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

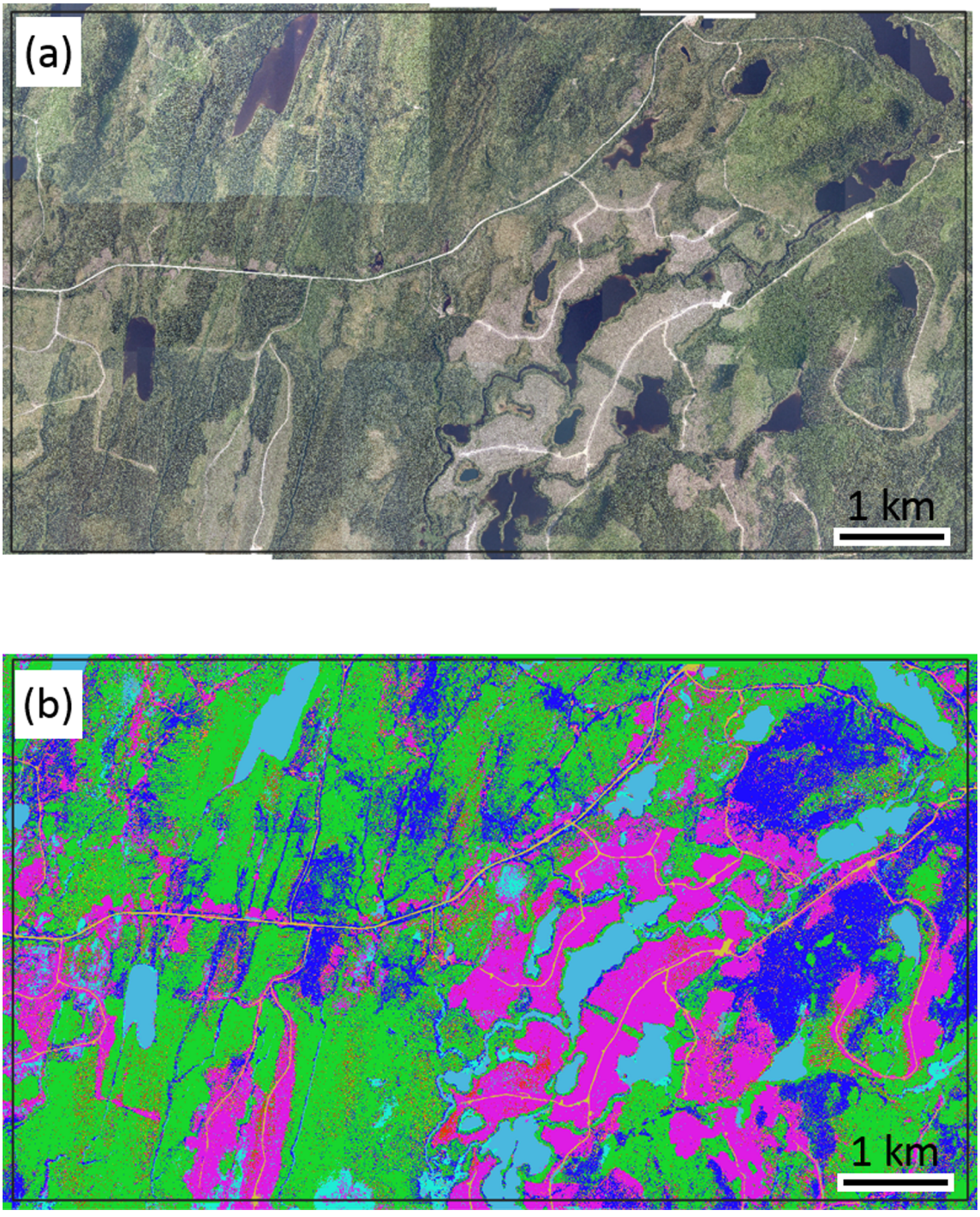

Fig. 9. (a) Mosaic of RGB orthoimages and (b) predictions of plant functional type (PFT) and health status (HS) classes, plus 5 additional land cover classes on a mosaic of 43 RGBI orthoimages. Predictions were made using iteration 1 of model VGG16S-RGBI-41px-9cl. Pixel colors in panel b represent: live broadleaf (LB, dark blue), live needleleaf (LN, green), dead broadleaf (DB, orange), dead needleleaf (DN, red), water (light blue), wetland (cyan), timber harvest (clear-cut, etc) (purple), road (yellow), and building (gray). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

that was classified as DB by the model from iteration 1 was later classified as DN by the ensemble approach. Fig. 8d also shows that most of the shaded canopy pixels were classified as needleleaf pixels, which can contribute to an overestimation of the number of needleleaf pixels in images acquired under low solar zenith angles or in mountainous areas. At the same time, based on visual examination of our image database, shaded areas are more likely to occur in areas covered with needleleaf trees than in those where the forest cover is mostly composed of deciduous broadleaf trees. Finally, some of the confusion observed among LN and LB classes in the confusion matrix could also be explained by our prediction method: small objects have been smoothed and enlarged by our prediction approach using the sliding window a 1-m intervals.

3.7.3. Landscape scale patterns: 9-class model

Fig. 9 depicts the results of applying the VGG16S-RGBI-41px-9cl model on a mosaic of 43 orthoimages. The figure allows for a visual comparison of the original orthoimages (Fig. 9a) with the map resulting from applying iteration 1 of the model (Fig. 9b). The classification

achieved a good recognition of the areas dominated by LN (green) and LB (dark blue) pixels, large timber harvest areas (purple), lakes (light blue). The model was also able to recognize the occurrence of a wetlands (cyan) in the south central part of the area. The 9-class model is well suited to detect spatial patterns of dead needleleaf tree cover (DN, in red), as is seen in the lowerleft portion of the image and also near the southern section of a lake located in the upperleft part of the image. The model was also able to accurately delineate main and secondary roads (yellow pixels). The model was almost unaffected by the differences in image brightness levels, which are noticeable mostly in the northwestern part and in the southern central part of the orthoimage mosaic. Transition lines are almost unperceptible in the classification, which suggests that the CNN is less sensitive to changes in acquisition conditions between the images; this represents a significant advantage of CNNs compared to more traditional statistical approaches. These results also imply that CNNs are more influenced by the information about the spatial structure of the relative pixel values than by radiometric changes of the absolute pixel values. This particular property will provide greater robustness when processing many aerial orthoimages, which are in general more subject to important changes in environmental conditions during their acquisition.

4. Discussion

In this study, we demonstrated the potential of CNNs to map dead forest cover composed of broadleaf and deciduous trees using digital aerial photography acquired over a wide range of environmental conditions. We demonstrated that 1) the effect of window size on predictive accuracy is important Franklin et al. (2010) while 2) that of channel combinations is limited. Window size should then at least include the maximum size of the target and its full spatial context (Latifi et al., 2018; Byer and Jin, 2017; Meng et al., 2016; Zhang et al., 2014; Coops et al., 2006). In this context, higher spatial resolution imagery should therefore be prioritized for the assessment of tree health status when using CNN, as coarser resolution imagery will tend to smooth higher frequencies used by the CNN (Dash et al., 2017; Meng et al., 2016; Coops et al., 2006).

Our results also suggest that the aggregation of multiple predictions lead to more robust and accurate forecasts, which in turn lead to a better predictive accuracy (Kourentzes and Petropoulos, 2016; Marmanis et al., 2016). The ensemble approach also offers the advantage of generating uncertainty maps, which allow to spatialize the level of agreement among classifiers. In an operational context, uncertainty maps can be used to identify patterns that are common to misclassified labelled-images and thus help in refining the modelling approach.

It is also important to recall that all models developed in this study were trained and applied on aerial imagery digital numbers without any radiometric correction applied. The good performances of the 4- and 9-class CNNs suggest that this classification approach considers the spatial relationships between pixel values as much as their absolute values. This property makes CNNs less sensitive to changes in acquisition conditions between multiple scenes compared to traditional approaches, and will allow for their application on datasets having extensive spatial (mosaics) and temporal coverage (stacks). The visual comparison between the mosaic of orthoimages and the map from the 9-class model demonstrated the capacity of CNNs to manage orthoimages acquired under varying conditions (illumination and solar angles).

In its actual form, our best CNN performed very well for classifying a range of land cover, HS and PFT classes when compared to recent literature (Byer and Jin, 2017; Dash et al., 2017; Wang et al., 2016). However, the confusion matrices revealed relatively high rates of omission and commission errors for predictions of PFT classes. Also, in this study we limited the training dataset to a small number of land cover classes; it would be relevant to integrate more land cover types to increase the possible applications of the models to other territories containing, for example, more rocks, bare soil, bryophytes, lichens, ground, ericaceous, grasses, shrubs, crops, etc. Moreover, topographical gradients, tree morphology (size, height, shape) and camera parameters (spatial and spectral resolution) may introduce geometrical distortion in the aerial imagery. The features associated with these distortions may be limited in the dataset used for model training, as we used only a limited set of observations for the minority classes, even when including rotations, flips and translations. In future work, it would be well advised to integrate more transformations in the preprocessing steps, which may lead to a better characterization of the diversity encountered in aerial imagery acquired over different forest zones; scaling, perspective transform and brightness values are possible avenues to improve the training dataset. Consequently,these would also increase the ability of the models developed in this study to achieve similar accuracy levels on datasets acquired over a diversity of spatial domains and acquisition periods. Alternative CNN-based approaches (e.g., object-based CNN, semantic segmentation (FCN)) that could have the potential to map dead forest cover in forested ecosystems should be tested and compared based on their performance (Chen et al., 2018; Anwer et al., 2018). Based on the results we obtained with ensemble approach, we would also recommend to build a prediction model that combined the output of multiple CNN-based model approaches.

5. Conclusion

We extended the application of CNNs to map dead forest cover in deciduous, mixed and boreal forests using aerial photography. The models we developed provide a robust approach for obtaining spatially continuous forecasts of forests plant functional type and health status. The global accuracies achieved by the best 4-class model was 96% for the health status, 85% for plant functional type and 79% for the combined classes of plant functional type and health status. We also demonstrated that it is possible to increase the robustness and stability of CNN outputs by aggregating classes predicted using multiple model iterations, and we showed how multiple predictions can be used to derive uncertainty maps at landscape level.

Historically, tree mortality remained a process that has been poorly characterized and understood. From an operational and research perspective, our work provides a new tool to map temporal changes in plant functional type and health status at landscape to regional scales. The maps resulting from these new developments will provide a precise and accurate portrait of tree mortality at unprecedented spatial resolution (1×1 m). The spatial portrait derived using this approach will provide a level of detail that will benefit to a better understanding of the effect of abiotic and biotic factors on tree mortality. This information will also provide important information about the structure, composition and productivity of forest ecosystems. Decision makers may also find many practical uses for these models, namely for forest management and planning: delineation of forest stands, wetlands and flooded zones, assessment of forest composition, planning salvage cuts after wildfires or insect disturbances, etc. The application of our algorithm on time series aerial imagery will provide the opportunity to assess mortality rates over large territories. Cartographic products resulting from this study will also support, in combination with multisource datasets, future research aimed at understanding the effects of environmental conditions on forest dynamics and on tree mortality in particular.

Acknowledgements

This project was funded by the Ministère des Forêts, de la Faune et des Parcs du Québec (MFFP, project #142959251) and the Plan d'action 2013–2020 sur les changements climatiques (Fonds Vert). The authors want to thank Geneviève Auclair (tech. geom.) and Louis Lemieux (tech. for.), both from the Direction des inventaires forestiers (MFFP), for their great support with the production of aerial orthoimages and in writing photointerpretation guidelines. We are also grateful to our industrial partner, Gestion Forestière St-Maurice, for giving us access to their database of aerial photos.

Appendix A. Additional metrics

See Tables A1–A3.

Table A1
Recall values calculated from predictions on the test and validation datasets, for each model configuration and prediction type. Values were calculated using aggregated predictions (pixel-wise modal class) from 10 model iterations. For PFT + HS, recall is defined as the average recall for all 4 classes. $Recall = \frac{tp}{tp + fn} * 100\%$.

	PFT		*HS*		*PFT + HS*	
	Val	*Test*	*Val*	*Test*	*Val*	*Test*
$w = 41px$						
RGBI	84.1	86.9	93.5	91.6	80.6	79.5
RGB	85.9	87.3	93.5	92.2	80.0	79.1
IRG	84.1	86.7	91.2	91.6	78.8	77.7
$w = 21px$						
RGBI	85.3	87.5	93.5	92.9	79.5	77.4
RGB	87.6	86.5	92.4	90.8	80.6	76.5
IRG	82.4	86.1	92.4	91.0	78.3	76.1

Table A2
Precision values calculated from predictions on the test and validation datasets, for each model configuration and prediction type. Values were calculated using aggregated predictions (pixel-wise modal class) from 10 model iterations. For PFT + HS, precision is defined as the average precision for all 4 classes. $Precision = \frac{tp}{tp + fp} * 100\%$.

	PFT		*HS*		*PFT + HS*	
	Val	*Test*	*Val*	*Test*	*Val*	*Test*
$w = 41px$						
RGBI	85.6	81.3	95.8	97.3	80.6	79.8
RGB	83.0	80.5	95.8	96.9	80.0	79.5
IRG	84.6	79.5	95.1	95.9	79.0	78.3
$w = 21px$						
RGBI	81.9	78.4	95.8	95.4	79.4	77.9
RGB	81.9	78.2	96.3	96.1	80.9	77.3
IRG	82.4	78.3	95.7	95.1	78.3	76.6

Table A3
F1 score values calculated from predictions on the test and validation datasets, for each model configuration and prediction type. Values were calculated using aggregated predictions (pixel-wise modal class) from 10 model iterations. For PFT + HS, F1 score is defined as the average F1 score for all 4 classes. $F1 = 2 * \frac{Precision * Recall}{Precision + Recall} * 100\%$.

	PFT		*HS*		*PFT + HS*	
	Val	*Test*	*Val*	*Test*	*Val*	*Test*
$w = 41px$						
RGBI	84.9	84.0	94.6	94.3	80.6	79.3
RGB	84.4	83.7	94.6	94.5	80.0	79.0
IRG	84.4	82.9	93.1	93.7	78.8	77.6
$w = 21px$						
RGBI	83.6	82.7	94.6	94.1	79.4	77.2
RGB	84.7	82.1	94.3	93.3	80.5	76.3
IRG	82.4	82.0	94.0	93.0	78.1	75.8

References

Anwer, R.M., Khan, F.S., van de Weijer, J., Molinier, M., Laaksonen, J., 2018. Binary patterns encoded convolutional neural networks for texture recognition and remote sensing scene classification. ISPRS J. Photogramm. Remote Sens. 138, 74–85. https://doi.org/10.1145/3078971.3079001.

Audebert, N., Le, B., Lefèvre, S., 2018. Beyond RGB: Very high resolution urban remote sensing with multimodal deep networks. ISPRS J. Photogramm. Remote Sens. 140, 20–32. https://doi.org/10.1016/j.isprsjprs.2017.11.011.

Breshears, D.D., Myers, O.B., Meyer, C.W., Barnes, F.J., Zou, C.B., Allen, C.D., McDowell, N.G., Pockman, W.T., 2009. Research communications research communications Tree die-off in response to global change-type drought: Mortality insights from a decade of plant water potential measurements. Front. Ecol. Environ. 7, 185–189. https://doi.org/10.1890/080016.

Brochero, D., Hajji, I., Pina, J., Plana, Q., Sylvain, J.D., Vergeynst, J., Anctil, F., 2015. One-day-ahead streamflow forecasting via super-ensembles of several neural network architectures based on the Multi-Level Diversity Model. Eur. Geosci. Union (EGU), Geophys. Res. Abs. 0–1.

Buda, M., Maki, A., Mazurowski, M.A., 2017. A systematic study of the class imbalance problem in convolutional neural networks. pre-print abs/1710.0, 1–23.

Byer, S., Jin, Y., 2017. Detecting drought-induced tree mortality in Sierra Nevada forests with time series of satellite data. Remote Sens. 9, 1–23. https://doi.org/10.3390/rs9090929.

Caspersen, J.P., Vanderwel, M.C., Cole, W.G., Purves, D.W., 2011. How stand productivity results from size- and competition-dependent growth and mortality. PLoS One 6, e28660. https://doi.org/10.1371/journal.pone.0028660. URL: https://dx.plos.org/10.1371/journal.pone.0028660.

Chen, K., Fu, K., Yan, M., 2018. Semantic segmentation of aerial images with shuffling convolutional neural networks. IEEE Geosci. Remote Sens. Lett. 15, 173–177. https://doi.org/10.1109/LGRS.2017.2778181.

Chollet, F., Others, 2015. Keras. URL: https://keras.io.

Clyatt, K.A., Crotteau, J.S., Schaedel, M.S., Wiggins, H.L., Kelley, H., Churchill, D.J., Larson, A.J., 2016. Historical spatial patterns and contemporary tree mortality in dry mixed-conifer forests. For. Ecol. Manage. 361, 23–37. https://doi.org/10.1016/j.foreco.2015.10.049.

Coops, N.C., Johnson, M., Wulder, M.A., White, J.C., 2006. Assessment of quickbird high spatial resolution imagery to detect red-attack damage due to mountain pine beetle infestation. Remote Sens. Environ. 1, 67–80. https://doi.org/10.1016/j.rse.2006.03.012.

Dash, J.P., Watt, M.S., Pearse, G.D., Heaphy, M., Dungey, H.S., 2017. Assessing very high resolution UAV imagery for monitoring forest health during a simulated disease outbreak. ISPRS J. Photogramm. Remote Sens. 131, 1–14. https://doi.org/10.1016/j.isprsjprs.2017.07.007.

Direction des inventaires forestiers, 2009. Normes de cartographie écoforestière Troisième inventaire écoforestier.

Franklin, J.F., Shugart, H., Harmon, M.M.E., 1987. Tree death as an ecological process. Bioscience 37, 550–556. URL. http://www.jstor.org/stable/1310665.

Franklin, S.E., Wulder, M.A., Gerylo, G.R., 2010. Texture analysis of IKONOS panchromatic data for Douglas-fir forest age class separability in British Columbia. Int. J. Remote Sens. 1161, 2627–2632. https://doi.org/10.1080/01431160120769.

Glorot, X., Bengio, Y., 2010. Understanding the difficulty of training deep feedforward neural networks. In: Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics (AISTATS 2010), pp. 249–256 doi:10.1.1.207.2059.

Gueguen, L., Hamid, R., 2015. Large-scale damage detection using satellite imagery. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1321–1328.

Hastie, T., Tibshirani, R., Friedman, J.H., 2009. The elements of statistical learning: data mining, inference, and prediction, 2nd ed. Springer Series in Statistics Springer.

He, K., Zhang, X., Ren, S., Sun, J., 2015. Spatial pyramid pooling in deep convolutional networks for visual recognition. IEEE Trans. Pattern Anal. Mach. Intell. 37, 1904–1916. https://doi.org/10.1109/TPAMI.2015.2389824.

Hu, J., Shen, L., Sun, G., 2017. Squeeze-and-Excitation Networks. pre-print, 1–14, https://doi.org/10.1109/CVPR.2018.00745.

Huang, G., Liu, Z., van der Maaten, L., Weinberger, K.Q., 2016. Densely Connected Convolutional Networks. pre-print abs/1608.0.

Hurst, J.M., Stewart, G.H., Perry, G.L., Wiser, S.K., Norton, D.A., 2012. Determinants of tree mortality in mixed old-growth Nothofagus forest. For. Ecol. Manage. 270, 189–199. https://doi.org/10.1016/j.foreco.2012.01.029.

Ioffe, S., Szegedy, C., 2015. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. pre-print abs/1502.0. https://doi.org/10.1007/s13398-014-0173-7.2.

Kellner, J.R., Hubbell, S.P., 2017. Adult mortality in a low-density tree population using high-resolution remote sensing. Ecology 98, 1700–1709. https://doi.org/10.1002/ecy.1847.

Kourentzes, N., Petropoulos, F., 2016. Forecasting with multivariate temporal aggregation: The case of promotional modelling. Int. J. Prod. Econ. 181, 145–153. https://doi.org/10.1016/j.ijpe.2015.09.011.

Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. ImageNet classification with deep convolutional neural networks. In: In: Pereira, F., Burges, C.J.C., Bottou, L., Weinberger, K.Q. (Eds.), Advances In Neural Information Processing Systems, vol. 25. Curran Associates, Inc., pp. 1097–1105. https://doi.org/10.1016/j.protcy.2014.09.007.

Larson, A.J., Lutz, J.A., Donato, D.C., Freund, J.A., Swanson, M.E., HilleRisLambers, J., Sprugel, D.G., Franklin, J.F., 2015. Spatial aspects of tree mortality strongly differ between young and old-growth forests. Ecology 96, 2855–2861. https://doi.org/10.1890/15-0628.1.

Latifi, H., Dahms, T., Beudert, B., Heurich, M., Kübert, C., Dech, S., 2018. Synthetic RapidEye data used for the detection of area-based spruce tree mortality induced by bark beetles. GIScience Remote Sens. 55, 839–859. https://doi.org/10.1080/15481603.2018.1458463.

Marmanis, D., Wegner, J.D., Galliani, S., Schindler, K., Datcu, M., Stilla, U., 2016. Semantic segmentation of aerial images with an ensemble of cnns. ISPRS Annals of Photogrammetry, Remote Sensing and Spatial. Inf. Sci. III-3, 473–480. https://doi.org/10.5194/isprs-annals-III-3-473-2016.

Martin Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Jia, Y., Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dan Mané, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Mike Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, Xiaoqiang Zheng, 2015. TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems. URL: https://www.tensorflow.org/.

Meng, J., Li, S., Wang, W., Liu, Q., Xie, S., Ma, W., 2016. Mapping Forest Health Using Spectral and Textural Information Extracted from SPOT-5 Satellite Images. Remote Sens. 8, 1–20. https://doi.org/10.3390/rs8090719.

Olson, C., Ma, Z., 1989. Normality assumptions in supervised classification of remotely sensed terrain data. In: Quantitative Remote Sensing: An Economic Tool for the Nineties, IGARSS '89, pp. 1857–1859.

Olson, C.E., 2009. The fallacy of normality in remotely sensed data, in: ASPRS 2009 Annual Conference.

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A.C., Fei-Fei, L., 2015. ImageNet large scale visual recognition challenge. Int. J. Comput. Vision 115, 211–252. https://doi.org/10.1007/s11263-015-0816-y.

Simonyan, K., Zisserman, A., 2015. Very deep convolutional networks for large-scale image recognition, in: ICLR 2015, pp. 1–14. https://doi.org/10.1016/j.infsof.2008.09.005.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A., 2014. Going Deeper with Convolutions. pre-print abs/1409.4. https://doi.org/10.1109/CVPR.2015.7298594.

Van Gunst, K.J., Weisberg, P.J., Yang, J., Fan, Y., 2016. Do denser forests have greater risk of tree mortality: A remote sensing analysis of density-dependent forest mortality. For. Ecol. Manage. https://doi.org/10.1016/j.foreco.2015.09.032.

Wang, H., Zhao, Y., Pu, R., Zhang, Z., 2016. Mapping Robinia Pseudoacacia forest health conditions by using combined spectral, spatial and textural information extracted from Ikonos imagery, in: ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, pp. 1425–1429. https://doi.org/10.5194/isprs-archives-XLI-B8-1425-2016.

Wu, S., Zhong, S., Liu, Y., 2015. Deep residual learning for image recognition. Arxiv preprint arXiv abs/1512.0. https://doi.org/10.1007/s11042-017-4440-4.

Zhang, W., Hu, B., Woods, M., 2014. Mapping forest stand complexity for woodland caribou habitat assessment using multispectral airborne imagery, in: ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, pp. 179–185. https://doi.org/10.5194/isprsarchives-XL-2-179-2014.

Zoph, B., Vasudevan, V., Shlens, J., Le, Q.V., 2017. Learning Transferable Architectures for Scalable Image Recognition. https://doi.org/10.1126/science.1216744.