



Contents lists available at ScienceDirect

International Journal of Applied Earth Observations and Geoinformation

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

Classifying surface fuel types based on forest stand photographs and satellite time series using deep learning

Pia Labenski^{a,*}, Michael Ewald^a, Sebastian Schmidlein^a, Fabian Ewald Fassnacht^{a,b}^a Institute of Geography and Geoecology, Karlsruhe Institute of Technology (KIT), Kaiserstr. 12, 76131 Karlsruhe, Germany^b Department of Remote Sensing and Geoinformation, Institute of Geographic Sciences, Freie Universität Berlin, Malteserstraße 74-100, 12249 Berlin, Germany

ARTICLE INFO

Keywords:

Fuel types
 Deep neural networks
 Image classification
 Convolutional neural networks
 Long short-term memory
 Sentinel-2

ABSTRACT

With the increasing threat of wildfires globally, improving the availability of accurate, spatially explicit fuel type information is critical for fire behavior predictions that can support management decisions to mitigate fire hazards. Since mapping surface fuel types using airborne or spaceborne sensors relies on ground truth data from laborious field assessments, here we propose a novel proximate sensing-based approach for classifying surface fuel types from in-forest RGB photographs using convolutional neural networks (CNNs). We test different configurations of deep learning models that integrate photographs of the forest stand and the forest floor as well as time series of multispectral satellite data from Sentinel-2 using long short-term memory (LSTM), and compare their performance in classifying understory and litter fuel types of Central European forests. We also investigate how ensemble approaches based on majority voting can help to improve classification results. We found that understory fuel types were classified with highest accuracy after cross-validation (0.78) using a combination of horizontal stand photos and forest floor photos. This accuracy was further improved by post-classification decision fusion of model predictions on multiple photographs of a forest stand and by considering the model's confidence in its predictions (0.85). Litter fuel type classification based on forest photographs resulted in lower overall accuracy (0.60), but using model ensemble predictions on both photographs and Sentinel-2 time series significantly improved the results (0.72). We found that the accuracy of our models was mostly limited by naturally smooth transitions between the defined fuel type classes and the co-occurrence of multiple fuel types in a photograph. This study shows that deep learning methods can provide an efficient means to assess fuel types from GNSS-located photos of forest stands as a basis for generating and validating fuel type and finally fire risk maps. The necessary data can be readily collected by forest managers or citizen scientists.

1. Introduction

Forests of Central Europe are becoming increasingly vulnerable to wildland fires as a consequence of global warming (de Rigo et al., 2017; Forzieri et al., 2021). Higher temperatures, more frequent and intense droughts in combination with other abiotic and biotic stressors affect the health of temperate forests and increase vegetation flammability (Millar and Stephenson, 2015; Spinoni et al., 2018; IPCC, 2019). In the drought years 2018 and 2019, several Central European countries reported a higher number of fires and a few exceptionally large burnt areas compared to the 10-year average from 2008 to 2017 (European Commission, 2021), indicating a link between extreme droughts and enhanced wildfire activity. This is in accordance with future projections of climate-driven wildfire activity that predict higher fire probabilities

in highly productive, previously flammability-limited regions due to longer fire weather seasons (Jolly et al., 2015; Abatzoglou et al., 2019). Model simulations suggest a lengthening of mid-latitude and boreal fire season by up to three months by the end of the 21st century (Veira et al., 2016), leading to an expansion of fire-prone regions in Europe. Wildfires can pose a serious threat to environment and society, in addition to causing major damage to timber volume and loss of carbon stocks, if adaptation measures are not taken (Seidl et al., 2014; Khabarov et al., 2016). As fire behavior is strongly determined by fuel characteristics, suitable management practices to reduce forest fire hazard require spatially explicit information about forest fuel availability, composition and structure (Keane, 2013).

The complex arrangement of fuels in a forest is often vertically stratified into canopy, surface and ground fuels. Surface fuels by

* Corresponding author.

E-mail address: pia.labenski@kit.edu (P. Labenski).

<https://doi.org/10.1016/j.jag.2022.102799>

Received 28 February 2022; Received in revised form 1 April 2022; Accepted 21 April 2022

Available online 29 April 2022

1569-8432/© 2022 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

definition comprise all biomass within two meters above the ground surface: senesced leaves, needles and other nonwoody discarded plant material (litter), fine and coarse woody debris from trees and shrubs (twigs, branches and logs), vascular plant biomass (grasses, herbs, forbs, shrubs and young trees) as well as lichens and mosses. These fuel components are each characterized by a specific particle size and shape, mineral and heat content, and are arranged with a certain compactness and continuity, thus showing distinct combustion properties (Countryside, 1964; Chuvieco et al., 2003). Fuel composition and structure strongly vary across spatio-temporal scales due to different environmental conditions and management practices. For simplification, it is common among practitioners to summarize the fuel properties relevant for fire hazard estimation of a forest stand as “fuel types”, which are usually determined by the dominant fuel component in an area (Keane, 2015). A fuel type is assumed to “exhibit characteristic fire behavior under defined burning conditions” (Merrill and Alexander, 1987), as manifested by ease of ignition, rate of spread, fireline intensity and fuel consumption (Varner et al., 2015). Commonly distinguished are i) herbaceous fuel types that form loosely packed fuelbeds that are easy to ignite and foster rapid fire spread, ii) shrub fuel types with diverse size and distribution of fuel particles which can burn at high intensities depending on species composition and compactness, iii) litter fuel types that dry quickly and ignite easily, but burn at low intensities and iv) woody fuel types dominated by dead woody fuel particles with different rates of drying depending on particle size, that can foster intense surface fires (Sandberg et al., 2001; Keane, 2015). The detailed numerical description of the physical properties of a fuel type is referred to as “fuel model” (Andrews and Queen, 2001) and is often used as a set of inputs to fire behavior models (Finney, 2006; Andrews, 2014) to help forest managers predict potential fire behavior and decide for effective fuel management options.

Accurate spatial information on surface fuels is fundamental for appropriate forest and fire management strategies, but mapping surface fuel types remains a difficult task. Traditional mapping methods based on recordings of fuel situations in the field are very time consuming and costly; nevertheless, such field surveys are still required as primary source of data and as ground reference for fuel type maps produced using other datasets, including those collected by remote sensing (Arroyo et al., 2008). Recording fuel types can be aided by the use of photographs of representative fuel types that can be matched by the observer in the field to the forest stand situation encountered to facilitate classification (Keane, 2015). Extensive photo series have been developed for fuelbeds across the USA (Vihnanek et al., 2009; Wright et al., 2010), and also in other countries (Ottmar et al., 2004; Morfin-Rios et al., 2008). These even allow to estimate fuel component loadings, but the technique is prone to assignment errors and limited repeatability across observers has been reported (Sikkink and Keane, 2008; Keane, 2015). Fuel type maps are often generated using other land classifications such as vegetation maps by assigning fuel types to existing map categories (McKenzie et al., 2007); however, fuels are not always related to vegetation categories and map resolutions can be much coarser than the scale of fuel variation (Keane, 2015). Remote sensing methods offer another means to create fuel type maps across large areas: multispectral and hyperspectral data from passive sensors like Landsat TM, ASTER, AVIRIS and Hyperion have been extensively used in classification approaches (Riaño et al., 2002; Jia et al., 2006; Lasaponara and Lanorte, 2007; Keramitsoglou et al., 2008), many of them again relying on associations with vegetation categories. In terms of mapping surface fuel types, the main drawback of passive optical sensors is their incapability to penetrate the forest canopy. Active sensors like LiDAR systems partly overcome the problem and have been successfully used to extract information about vertical fuel structure (Riaño, 2003; Erdody and Moskal, 2010; Botequim et al., 2019), often in combined approaches with multispectral data (Mutlu et al., 2008; García et al., 2011; Chirici et al., 2013; Domingo et al., 2020). However, acquisition costs still limit the availability of LiDAR data across large areas. Moreover, LiDAR data

hardly provide information about the type of fuel encountered beneath the tree crown, which is yet essential to fire behavior predictions.

Field photographs obtained within forest stands capture the relevant information about surface fuel types and are often used by fuel researchers as ancillary information to determine ground truth and validate fuel type maps (Mutlu et al., 2008; García et al., 2011; Alonso-Benito et al., 2016; Botequim et al., 2019). However, visual interpretation of photos carried out by humans is time-consuming and subjective, whereas automated interpretation of images by deep neural networks can significantly reduce the time required for this task and also increase the repeatability of the interpretation. Such deep learning-based models allow to operationalize expert knowledge and make this knowledge available to interested parties as for example demonstrated in several projects providing deep learning models in applications to automatically identify plants and animals, e.g., Pl@ntNet or BirdNET (Goëau et al., 2013; Kahl et al., 2021).

In this study, we apply convolutional neural networks (CNNs), a class of deep learning models that are particularly suited for analyzing image data. CNNs process images through multiple layers of convolutional filters, thereby extracting contextual 2D spatial features of varying levels of abstraction, allowing the models to effectively learn features relevant to a specific task in an end-to-end training directly from the data. They have been applied with great success in computer vision tasks such as image classification (Sladojevic et al., 2016; Krizhevsky et al., 2017), object detection (Tompson et al.; Chen et al., 2014), and semantic segmentation (Long et al., 2015; Chen et al., 2018), but have only recently been explored in ecology and vegetation science (see reviews by Christin et al., 2019; Kattenborn et al., 2021). Vegetation properties such as species information and plant traits have been successfully extracted from plant photographs (Wäldchen and Mäder, 2018; Schiller et al., 2021), while highly accurate vegetation mapping has been achieved on different types of remote sensing data (Langford et al., 2019; Guirado et al., 2020; Schiefer et al., 2020). Most studies applying CNNs to field photographs are found in agriculture, e.g., for identification of crop types (Ringland et al., 2019; Wang et al., 2020) or the detection of weed infestations (Gao et al., 2020), as well as in land use or land cover classifications (Xu et al., 2017; Cao et al., 2018), while rather few studies from the field of forest ecology exist: these have attempted, for example, to detect the regrowth of woody vegetation (Bayr and Puschmann, 2019), classify tree species and estimate stock volume by segmentation (Liu et al., 2019), monitor plant phenology stages (Correia et al., 2020) or estimate defoliation of forest trees from ground-level images (Kálin et al., 2019). Despite the increasing use of deep learning models in ecological research, few studies currently aim to understand the behavior of a network and, thus, increase the interpretability and trustworthiness of the predictions; although this would also help to better evaluate the potential and limitations of deep learning models for these applications. Moreover, it has rarely been assessed how ground-based imagery can be coupled with remote sensing data to harness multiple data sources to make more reliable predictions for a task. In the context of fuel research in forest ecosystems, the ubiquitous availability of multispectral satellite data with high spatiotemporal resolution provided by the Sentinel-2 satellites provides an excellent opportunity to test whether time series of Sentinel-2 data can complement field-level information derived from forest photographs to predict surface fuel types in Central European forests. Since multi-temporal satellite data have proven useful to classify tree species and crops based on their different phenological cycles using varieties of recurrent neural networks (RNN) such as Long Short-Term Memory (LSTM) (Zhong et al., 2019; Campos-Taberner et al., 2020; Xi et al., 2021), they also hold the potential to differentiate between fuel types, which are influenced by dominating tree species and stand density. In this work, we present a new approach for classifying surface fuel types using RGB imagery from within a forest stand in combination with Sentinel-2 time series in a deep learning framework. We specifically address the following research questions:

- i. How accurately can we classify surface fuel types from different types of within stand forest photographs using CNNs?
- ii. Does the integration of Sentinel-2 satellite time series with LSTM improve classification results and does it have the potential to be used as a stand-alone methodology?
- iii. Do ensemble approaches help to improve the results?
- iv. Which image regions of forest photographs and which spectral and temporal features from Sentinel-2 time series are important for classifying surface fuel types?

2. Methods

2.1. Study area

We collected surface fuel data from 278 plots in temperate forests in Germany from May to October in 2020 and 2021, focusing on two main study areas. One is located in south-western Germany, encompassing lowland pine-dominated (*Pinus sylvestris*) mixed forests on sandy soils of the upper Rhine plain (Fig. 1-1), submontane mixed forests with beech (*Fagus sylvatica*), oak (*Quercus petraea*) and douglas fir (*Pseudotsuga menziesii*) in the hilly landscape of the Kraichgau (Fig. 1-2), submontane beech, spruce (*Picea abies*) and silver fir (*Abies alba*) forests in the northern Black Forest (Fig. 1-3), and dry submontane pine forests on sandstones of the Palatine Forest (Fig. 1-4). The other study area is located in the state of Brandenburg in north-eastern Germany and consists of lowland pine forests on very dry sandy soils (Fig. 1-5), which are the most fire-affected forest sites in Germany. We thus covered the six main overstory tree species in Central Europe, but also included less frequently occurring *Larix decidua*, *Quercus rubra*, *Carpinus betulus* and *Robinia pseudoacacia* stands. We attempted to cover different age classes and stand structures, ranging from very young stands consisting only of regenerated trees with heights of less than 5 m, to stands with larger trees and closed canopies, to old stands with low tree density and more open canopies.

2.2. Field measurements

We recorded overstory tree species and cover, understory tree/shrub cover and average height, herb cover and height, moss and litter cover

on 278 circular plots with a radius of 7.5 m (176.6 m²). We also recorded the litter type and the presence of fine woody fuels. In 179 of the plots, we sampled all surface fuel components (seedlings, shrubs, herbaceous species, dead woody fuels, litter) following an established protocol of the USDA Forest Service (Woodall and Monleon, 2008), to later calculate fuel loadings for each component. Details of the sampling procedure and data preparation can be found in the [supplementary material](#). Before sampling, we systematically photographed all plots. Twelve horizontal photos were taken from a circle with 10 m radius, facing the center of the plot, with a spacing of 30° between the photos. We also photographed the transects along which dead woody fuels were measured (see [supplementary material](#)), from four directions at 90° to each other, obtaining 12 forest floor photos per plot.

2.3. Fuel type classification

Unsupervised k-means clustering was performed on the fuel loading data to identify the most important clusters in the data. The data were then presented along with the photographs to two fuel experts, who related fuel and species information to effects on fire behavior. The final fuel type classification and respective thresholds to separate between classes were based on field data and expert opinion. Understory and litter type were considered most decisive to fire behavior and were thus used as sub-classification systems to constitute a fuel type.

Seven understory fuel types with expected different effects on fire behavior were identified (Fig. 2): 1) Broadleaved tree or shrub understory (hereafter referred to as *shrub-broadleaf*) mainly encountered as regeneration of *Fagus sylvatica*, *Carpinus betulus* and *Prunus serotina*, with large leaves that have high surface area-to-volume (SAV) ratio and water content, and the largest share of biomass allocated in stem and coarse branch wood. 2) Needle-leaved trees in the understory (*shrub-needle*) from regeneration of *Abies alba*, *Picea abies*, *Pseudotsuga menziesii* or *Pinus sylvestris*, that have leaves with smaller SAV ratio, higher lignin and terpenoid content (Perry et al., 1987; Scott and Binkley, 1997; Bohlmann and Keeling, 2008) and generally more biomass allocated in fine plant parts. 3) Herbaceous nongrassy species (*forb*) with high water content (we also included low-growing *Rubus fruticosus* agg. in this group due to its high moisture), which have a lower fire hazard than 4) grass species (*grass*) such as *Brachypodium sylvaticum* or *Deschampsia flexuosa*,

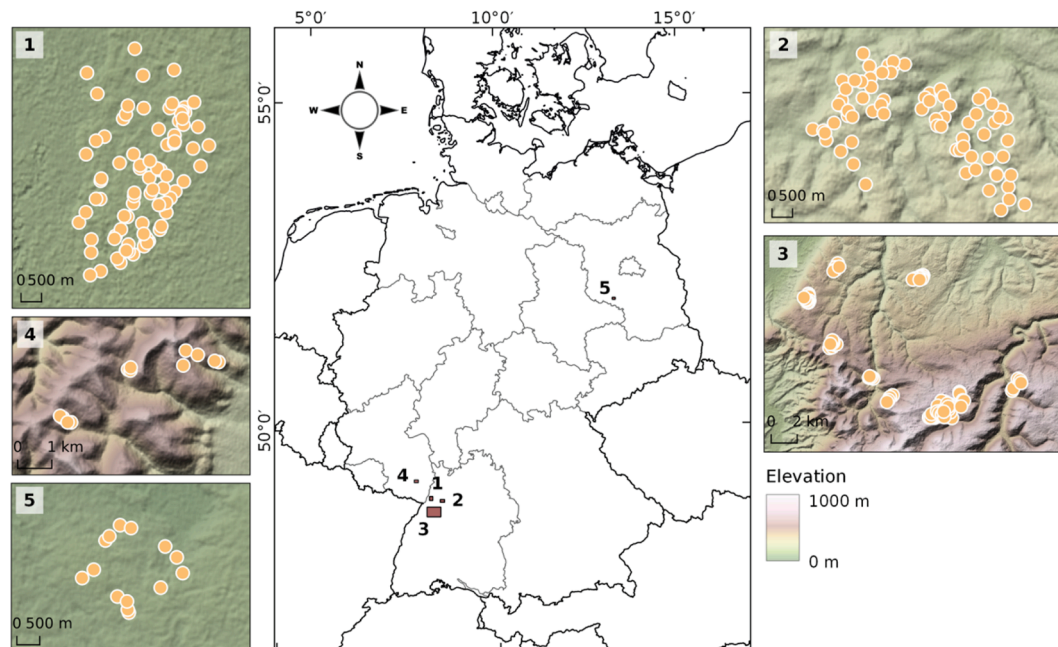


Fig. 1. Study areas for fuel type sampling in south-western and north-eastern Germany: 1) upper Rhine valley 2) Kraichgau 3) northern Black Forest 4) Palatine Forest 5) Brandenburg.



Fig. 2. Different understory fuel types. From top left to bottom right: Two examples of *shrub-broadleaf* (a-b), two examples of *shrub-needle* (c-d), then one example each of *forb* (e), *grass*(f), *dwarf-shrub* (g), *moss*(h) and *litter*(i).



Fig. 3. Examples of forest floor mosaics for a) broadleaf (*bl*), b) short-needle (*sn*), c) long-needle (*ln*) and d) short-needle fine woody debris (*sn-fwd*) litter.

especially after curing of the latter at the end of the season, 5) dwarf shrubs (*dwarf shrub*), in particular *Calluna vulgaris* and *Vaccinium myrtillus*, which can burn well even when green and 6) thick, continuous moss layers (*moss*) of species such as *Polytrichum formosum* or *Pleurozium schreberi* that can dry to very low moisture contents and provide significant fuel loadings. Cover of at least 50 % (within 2 m above the ground) of the respective understory type was considered necessary to achieve significant loading and continuity that would impact fire spread and was therefore chosen as threshold for the class assignment. If none of the aforementioned understory types was present with sufficient cover, the plots were assigned to the 7) litter (*litter*) fuel complex.

The litter fuel types relevant to fire behavior were distinguished based on leaf morphology of the litter, assuming that its relation with the compactness of the litter layer strongly influences the availability of oxygen in the combustion process. We therefore distinguished between broadleaf (*bl*), short-needle (*sn*) and long-needle (*ln*) litter (Fig. 3). We also assumed that the different chemical composition, especially of broadleaf and coniferous litter (Philpot, 1970; Scott and Binkley, 1997), affects the combustion properties. As mixtures between these litter types are very common in Central European forests, we also included the mixed classes broadleaf-short-needle (*bl-sn*) and broadleaf-long-needle

(*bl-ln*), assuming altered combustion properties compared to stands with pure litter types. In our study areas, we rarely encountered a mix of long-needle and short-needle litter and therefore assigned these plots to the dominating litter type. We found very high loads of fine woody debris in some short-needle stands, which could strongly alter the intensity of a fire, and therefore defined a separate litter type (*sn-fwd*). A simplified litter classification with only four different litter types, achieved by combining the classes *bl-ln* with *bl-sn* and *sn* with *sn-fwd*, was also tested. Table 1 provides an overview of the fuel type classifications and the number of plots surveyed for each class.

2.4. Image data preprocessing

Our dataset consisted of 3336 horizontal forest stand photos (12 per plot) and the same amount of forest floor photos, each 4000×3000 pixels in size. Single missing or damaged photos were replaced by duplicating a randomly selected photo from the same plot to ensure equal sample size for each plot. Horizontal photos were resized to 512×512 pixels before feeding them into the model and pixel values were normalized to the interval (0, 1) to allow faster convergence of the model. During model training, on-the-fly data augmentation was

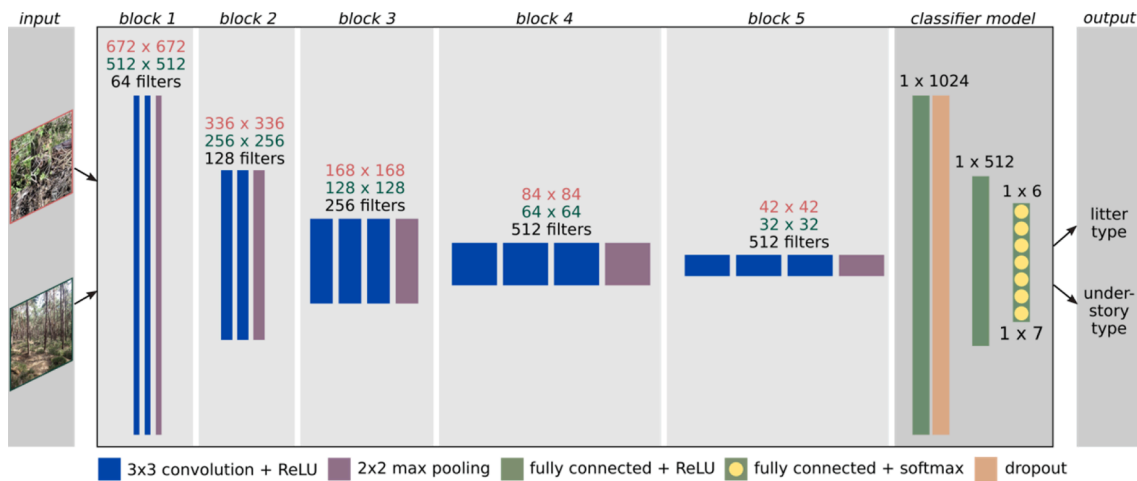


Fig. 4. Architecture of the CNN model based on VGG-16 for the classification into litter and understory fuel types. The model accepts either horizontal or forest floor photos as inputs. These are processed through five blocks of convolutional layers with increasing number of filters, while their resolution is decreased through pooling operations. The last fully-connected layer in the classifier model consists of 6 neurons in litter classification and 7 neurons in understory classification which give the probability values for a photo belonging to the respective classes.

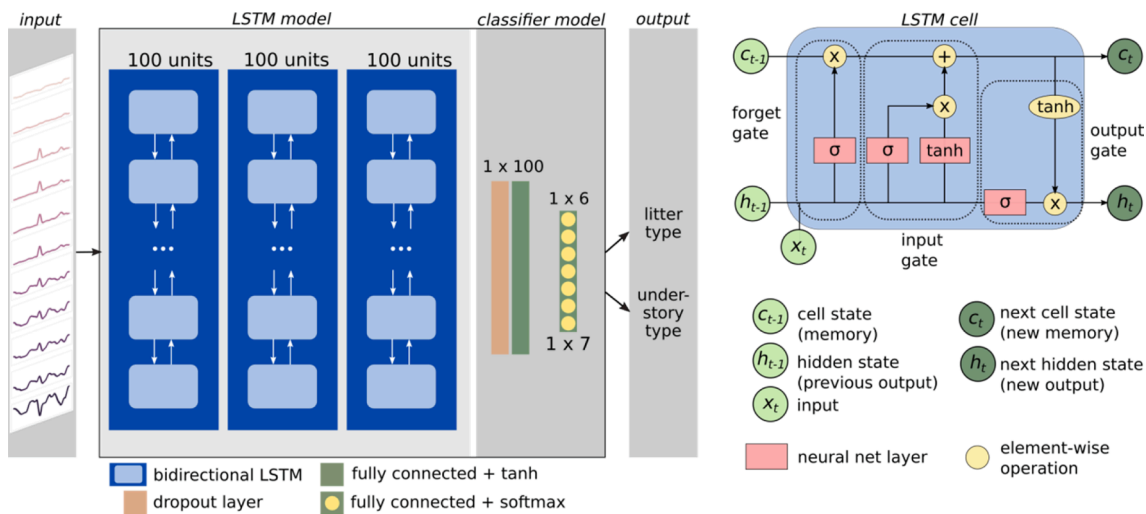


Fig. 5. Architecture of the LSTM model (left) and structure of a LSTM cell (right). The LSTM model accepts the 11 time series (features) derived from Sentinel-2 images, with 72 time steps in each series. The time series data is processed through 3 bidirectional LSTM layers, each of which contains a LSTM cell as repeating module that passes the (filtered) information from each time step in forward and backward direction and outputs a vector of length 100 (hidden state vector). Based on this output, the time series are assigned to different understory and litter fuel types.

Table 1
Overview of the recorded plots in each fuel type classification.

understory fuel type	no. ofplots	litter fuel type	no. ofplots	simplified litter fuel type	no. ofplots
shrub-broadleaf	39	bl	73	bl	73
shrub-needle	33	bl-ln	45	ln	44
forb	34	bl-sn	37	sn	79
grass	32	ln	44	mixed	82
moss	28	sn	61		
dwarf-shrub	19	sn-fwd	18		
litter	93				

performed, i.e. slight transformations were applied to the photos to increase the variation in the dataset during each epoch of training. These transformations included small image rotations, horizontal and vertical shifts, random horizontal flips and brightness changes within a range of values that was previously identified to produce realistic results. Forest floor photos were processed differently to avoid a loss of details when resizing the images to smaller sizes processible by the model: We randomly cropped 9 small image patches (224×224 pixels) from the forest floor photos and reassembled them to a 3×3 mosaic with a size of 672×672 pixels. Strong illumination variations within an image due to shadow effects were reduced by applying contrast limited adaptive histogram equalization (CLAHE) (Pizer et al., 1987) to each image before cropping. Similar to the horizontal photos, pixel values of the mosaics were normalized to the interval (0, 1). During training, forest

floor mosaics were randomly rotated by 90° degrees, but received no further transformations.

2.5. Satellite data preprocessing

We constructed time series of multispectral Sentinel-2 satellite data from the Level-2A surface reflectance product using Google Colab as Python interface to Google Earth Engine (GEE). Therefore, we selected all available scenes with less than 70 % cloud cover above our study areas in the 3-year period from July 2018 to June 2021. We used the Sentinel-2 cloud probability product and near-infrared reflectance to mask out cloud and cloud shadow pixels in the individual scenes. We then extracted pixel reflectance from 10 spectral bands with 10 to 20 m spatial resolution (visible, red-edge, near-infrared, shortwave-infrared bands) at our field plot locations. In addition to the spectral bands, we calculated the normalized difference vegetation index (NDVI) as an indicator of photosynthetic activity / vegetation greenness for each observation. To obtain a time series dataset equal in size to our photo dataset (3336 samples, 12 per plot) and increase the variability among time series of one plot, we binned the Sentinel-2 observations into 14-day intervals and randomly selected one observation from each 14-day interval to construct 12 slightly different time series for each of the 11 features per plot, with 72 time steps in each feature. Due to the cloud masking procedure, we obtained varying amounts of valid data points in the time series depending on the plot location. We linearly interpolated missing observations and smoothed the time series using a Savitzky-Golay filter (Savitzky and Golay, 1964). The final input to the LSTM model was a 72×11 matrix (time steps \times features) for each sample.

2.6. CNN architecture

We tested different CNN architectures typically used for image classification tasks, including VGG, Inception and EfficientNet. We achieved the best results using VGG-16 (Simonyan and Zisserman, 2015) with weights pre-trained on the ImageNet dataset as backbone. VGG-16 uses 5 blocks of consecutive 2D convolutions with a filter size of 3×3 and Rectified Linear Unit (ReLU) activation. Each block is followed by a max-pooling layer with stride 2 that reduces the resolution of the layers, allowing the transition from lower-level to higher-level image feature extraction. To reduce the number of trainable parameters in the model, we froze the layers in the first two convolutional blocks, i.e. their weights were not updated during training, so that low-level image features such as edge detectors were directly adopted from the pre-trained model. The convolutional layers deeper in the model were retrained on our dataset to allow the model to learn the higher-level concepts specific to our problem. The outputs from the VGG-16 backbone were then summarized in a global average-pooling layer and processed through a classifier model consisting of two fully connected layers and a final classifier with softmax activation, computing the class probabilities for the litter and understory fuel types, respectively. To limit overfitting and improve the model's ability to generalize, a 50 % dropout layer and L2 weight regularization with the regularization rate set to 0.01 were used. A multi-input model was constructed using two VGG-16 branches, which were concatenated before the final classifier model, to process horizontal and forest floor photos in parallel. A summary of the architecture of the CNN model is provided in Fig. 4.

2.7. LSTM architecture

We used a long short-term memory network (LSTM) to classify litter and understory fuel types based on the time series extracted from Sentinel-2 acquisitions. LSTM can learn long-term dependencies in sequences of data without suffering from the vanishing gradient problem that can occur when training normal recurrent neural networks (RNN) (Hochreiter and Schmidhuber, 1997). This is achieved by enforcing a

constant error flow through the network by regulating the information flow through LSTM units called cells. The memory content of a cell (cell state c) is controlled and protected by three sigmoid gate units (σ): the forget, input and output gates. The forget gate takes the output of the previous cell (h_{t-1}) and the current input (x_t) and decides which part of the memory content of the cell (c_{t-1}) will be thrown away. The input gate similarly uses the inflowing information to decide which parts of the memory will be updated, and a tanh layer gives weights to the respective values to be added to the current state. The new cell state (c_t) is then passed through another tanh layer (to scale the values between -1 and 1) and finally through the output gate, which decides what part of the cell state will be passed on to other cells (output values h_t). In this way the cells effectively discriminate between currently useful and irrelevant memory contents while ensuring constant error backpropagation to bridge even extended time intervals. As an extension of normal LSTMs, bidirectional LSTMs look at a time series from both forward and backward directions, allowing them to learn temporal dependencies using information from past *and* future time steps (Schuster and Paliwal, 1997). We used three bidirectional LSTM layers with 100 hidden units and 20 % dropout each to process the time series of the 10 spectral bands and NDVI from Sentinel-2, followed by a fully-connected layer and a final softmax layer to compute the class probabilities for the desired outputs. A summary of the architecture of the LSTM model is provided in Fig. 5.

2.8. Model training

The dataset was split into a training/validation set and an independent test set using stratified 10-fold cross validation. The test set thus contained 336 samples (photos / time series) from plots that the model had never seen during training, and it had the same distribution of classes as the full dataset. The training/validation set was split in a ratio of 80/20, resulting in 2400 samples for training and 600 for validation. Litter and understory fuel types were converted to one-hot encoded target variables before being fed into the network. The network was trained for a maximum of 50 epochs with a batch size of 32, i.e. 32 samples of the training dataset were shown to the network before the parameters were updated, while the entire training dataset was shown to the network a maximum of 50 times. To account for the imbalanced distribution of classes in our dataset, class weights were calculated by inversely relating occurrences per class to the total number of samples and used in training to weight up underrepresented classes. We tested five different combinations of input data with our models (Table 2): 2 CNNs were trained with only horizontal forest photos or only forest floor photos, respectively. Another CNN was trained with both horizontal and forest floor photos in two parallel VGG-16 branches. The LSTM model was trained with the Sentinel-2 time series, and a combined CNN-LSTM model was trained with all three data sources simultaneously in three parallel branches. The individual branches of the multi-input models were concatenated before the final classifier to arrive at a single joint prediction. For the CNN models and the multi-input models, we used the robust Adam optimizer with a learning rate of 0.0001 as optimization algorithm. The LSTM was optimized using RMSprop with a learning rate of 0.0001 and momentum set to 0.8, as determined by a hyperparameter grid search. The loss function to be minimized was categorical cross-entropy for all outputs. The learning rate was reduced during training when validation loss stopped improving for two epochs and training was stopped early if the loss did not improve for four epochs.

Model development and training were implemented in Python version 3.8 (van Rossum and Drake, 2009) using the Keras library (Chollet, 2015) as interface to the TensorFlow backend. For model training, we used 4 NVIDIA Tesla V100 GPUs provided by the bwUniCluster 2.0 within the Baden-Württemberg High Performance Computing (bwHPC) framework.

Table 2
Overview of the 5 model configurations tested.

input	forest floor photos	horizontal forest photos	Sentinel-2 time series	forest floor photos, horizontal photos	forest floor photos, horizontal photos and Sentinel-2 time series
model architecture	single-input CNN	single-input CNN	LSTM	multi-input CNN with two VGG-16 branches	multi-input CNN-LSTM model with two VGG-16 branches and one LSTM branch
output	litter fuel type: simplified litter fuel type: understory fuel type:		<i>bl, bl-ln, bl-sn, n, sn, sn-fw</i> <i>bl, ln, mixed, sn</i> <i>dwarf-shrub, grass, forb, litter, moss, shrub-broadleaf, shrub-needle</i>		

2.9. Model evaluation

Model performance was evaluated by calculating overall accuracy (1) and class-wise precision (2), recall (3) and f1-score (4) (harmonic mean of precision and recall) as well as Cohen's kappa (5) for predictions on the independent test sets that were generated using stratified 10-fold cross validation. All photos in the test set were considered as independent samples in these calculations. Confusion matrices were computed to gain further insights into which classes are difficult to separate using the models. Class prediction probabilities output by the final softmax layer of the best-performing model were examined for their informative value about the confidence of the model predictions. Python libraries Pandas (McKinney, 2010) and Scikit-learn (Pedregosa et al., 2011) were used for all computations.

$$accuracy = \frac{truepositives + truenegatives}{truepositives + truenegatives + falsepositives + falsenegatives} \quad (1)$$

$$precision = \frac{truepositives}{truepositives + falsepositives} \quad (2)$$

$$recall = \frac{truepositives}{truepositives + falsenegatives} \quad (3)$$

$$f1 - score = 2 \times \frac{precision \times recall}{precision + recall} \quad (4)$$

$$Cohen's\ kappa = \frac{observedagreement - expectedagreement}{1 - expectedagreement} \quad (5)$$

2.10. Ensemble approaches to improve classification results

Due to our data structure with 12 different photos acquired from one plot (forest stand), we had the opportunity to test the effect of decision fusion methods to improve final classification results. Two different approaches based on majority voting were tested: First, we aggregated the predictions from the same model on multiple photographs of the same forest stand and determined the final class label based on the most frequently predicted class. When two classes appeared to have the same number of votes, the final class was randomly chosen from the two. We additionally tested the effect of considering only the most certain model predictions by setting a threshold for the minimum required probability of the predicted class (tested values were 80 % and 90 %). Second, we aggregated the predictions from the single-input models that used the three different available data sources forest floor photos, horizontal photos and Sentinel-2 time series individually. Final class assignment was similarly based on majority voting from the ensemble of model predictions, and prediction probabilities were taken into account as described above.

2.11. Assessment of model explainability

2.11.1. Feature importance via random permutation in LSTM model

We assessed the relative importance of different spectral bands and

different acquisition times to the classification of understory and litter fuel types by using feature permutations: We randomly permuted the reflectance values of one band across all samples in the test set, applied the trained LSTM model to the modified data and recorded the change in accuracy compared to the baseline performance of the model on the unperturbed test set. Likewise, we permuted the reflectance values of all bands from each acquisition month across all samples in the test set and recorded the change in classification accuracy. Hence, the importance of each feature (band or month) was calculated as the decrease in classification accuracy of model predictions when the feature was permuted, normalized with respect to the most important feature.

2.11.2. Importance of image regions via Grad-CAM in CNN model

We used Gradient-weighted Class Activation Mapping (Grad-CAM) (Selvaraju et al., 2017) to visualize the image regions that were important for the classification decision of the CNN model. Grad-CAM computes the gradient of the score (raw output before the softmax) for any class with respect to the activations of the feature maps produced by a convolutional layer to derive the weight for each feature map. A weighted combination of feature maps is computed and followed by a ReLU operation to emphasize only pixels that have a positive influence on the class of interest. The output is a coarse localization map, which is upsampled to the resolution of the input image to highlight the pixels that were important for the class decision. We computed Grad-CAM heatmaps for randomly selected, correctly predicted images of each class based on the activations of the last two convolutional layers of our model.

3. Results

3.1. Model training

All models converged within 50 epochs of training or earlier for both outputs (Fig. 6 and Fig. 7). Multi-input models and the LSTM model generally showed slower convergence compared to the single-input CNN models. Except for the LSTM model, slight overfitting to the training data was observed for all models especially in litter classification and when using forest floor photos, despite the regularization techniques applied. When classifying understory, a single CNN-model trained on horizontal photos got stuck in a local minimum at an early epoch and remained at near-zero accuracy throughout training. Training and validation loss stabilized for all models after about 20 epochs.

3.2. Accuracy assessment

Average classification accuracy differed only marginally among the models using different input data, except for a lower understory classification accuracy (0.41) of the LSTM model using Sentinel-2 data only (Fig. 8). Highest accuracy for understory classification was achieved using the combined CNN model with horizontal and forest floor photos as input (OA = 0.78, f1-score = 0.76).

For the full litter classification with 6 litter types, different input data yielded very similar results, but the models using multiple data sources

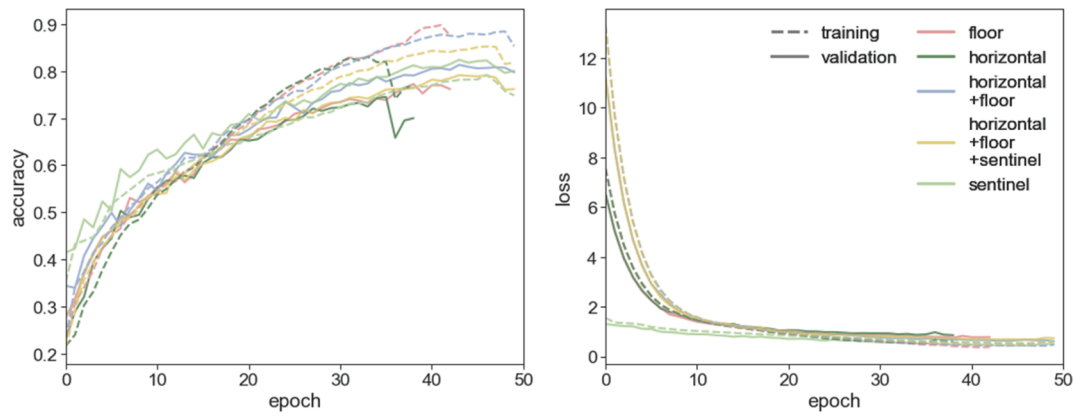


Fig. 6. Averaged evolution of training and validation accuracy and loss across 10 cross-validation runs for litter classification models.

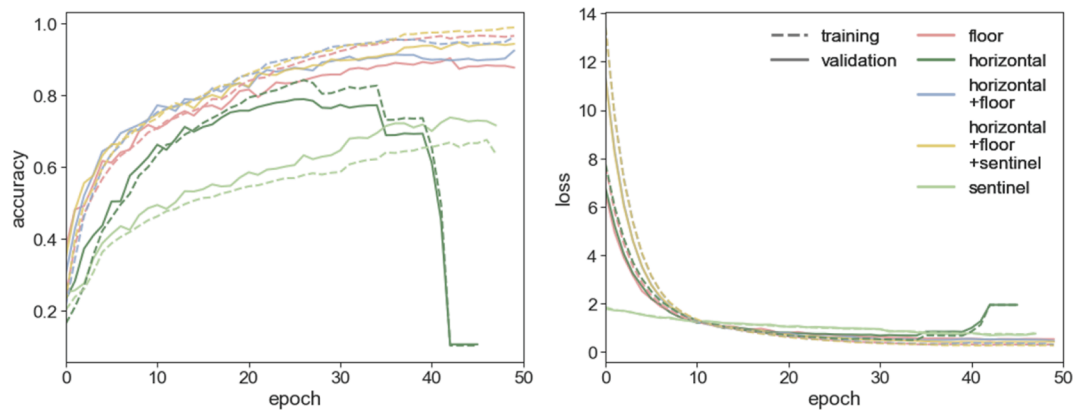


Fig. 7. Averaged evolution of training and validation accuracy and loss across 10 cross-validation runs for understory classification models. The curves for the CNN trained on horizontal photos are strongly distorted by a model stuck in a local minimum during training.

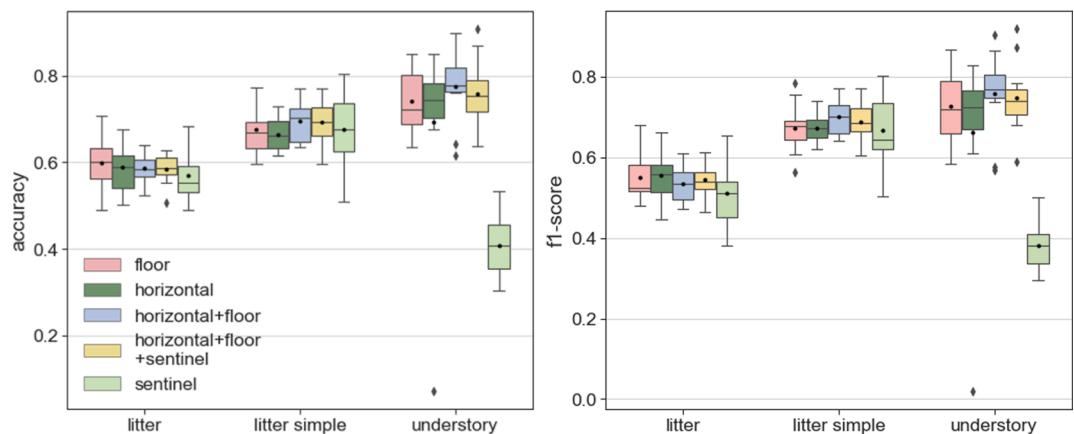


Fig. 8. Test accuracy (left) and f1-score (right) after cross-validation for different input data for original (6 classes) and simplified litter fuel types (4 classes), and understory fuel types (7 classes).

had lower variance compared to the single-input models (Fig. 8). Highest accuracy was obtained using forest floor photos (OA = 0.60, f1-score = 0.55). Simplifying the litter classification to only four different classes resulted in increased overall accuracy. Highest accuracy was achieved using the combination of horizontal and forest floor photos (OA = 0.70, f1-score = 0.70). In contrast to the understory classification, the LSTM based on Sentinel-2 data performed only slightly worse than the CNN models in litter classification, yet with high variability especially for the simplified litter fuel types (Fig. 8). Integrating Sentinel-2

data with the forest photos into a multi-input model improved overall classification accuracy only marginally for both understory and litter fuel types.

Confusion matrices (Fig. 9) show that pure litter types *bl*, *ln* and *sn* were highly distinguishable based on either forest photos or Sentinel-2 time series, whereas mixed litter classes were difficult to separate from each other and the pure litter types included in the mixtures. Class-wise precision, recall and f1-scores can be found in Table S1 in the supplementary material. In understory classification, *dwarf-shrub* was

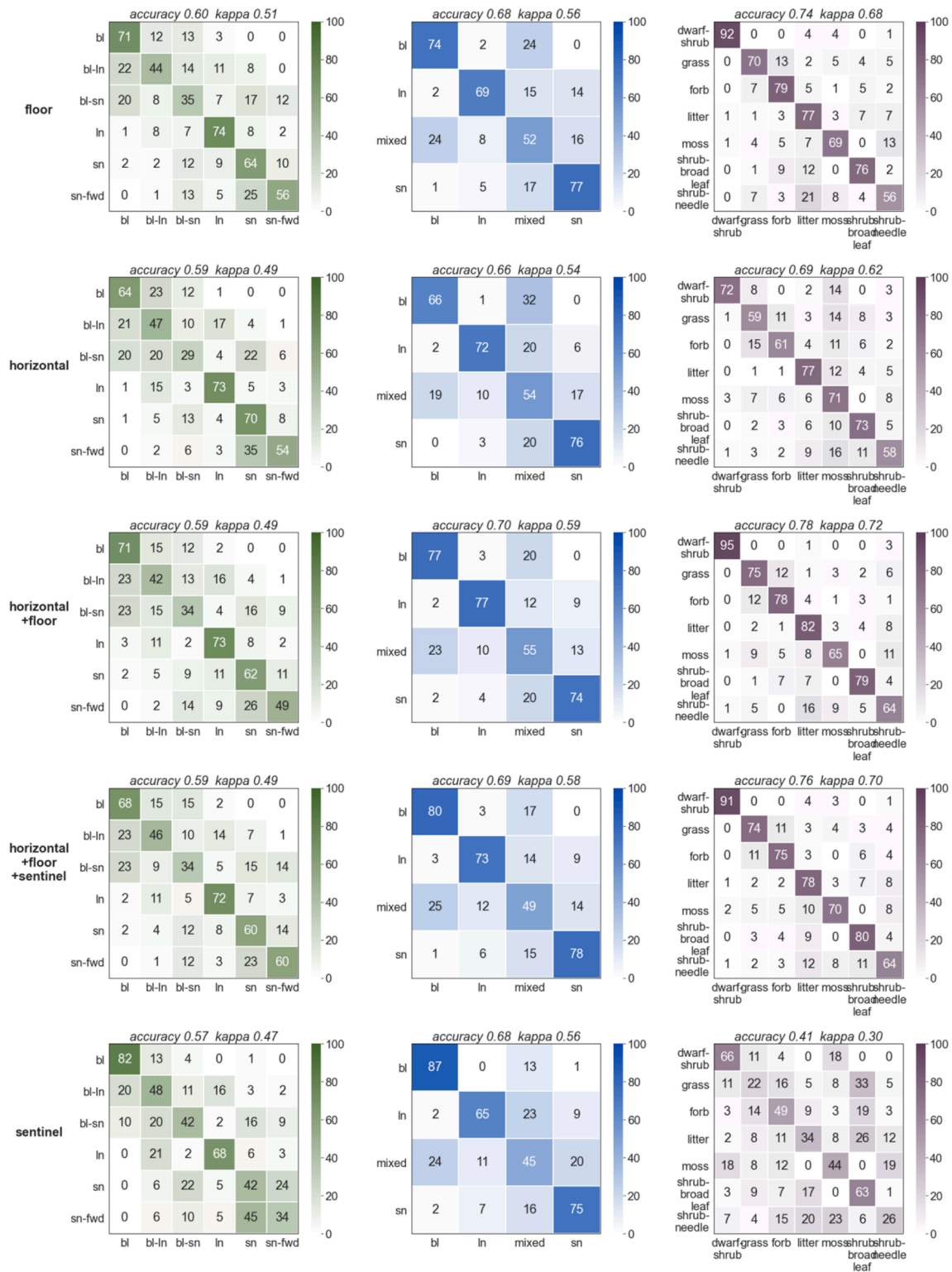


Fig. 9. Confusion matrices for litter classification (left column: 6 classes, middle column: 4 classes) and understory classification (right column). The matrices were averaged across all 10 cross-validation folds and normalized so that each row (true classes) sums to 100 for easier comparison across the imbalanced dataset. Note that training of a CNN on horizontal photos in one fold did not succeed and the model predicted all instances into the moss class.

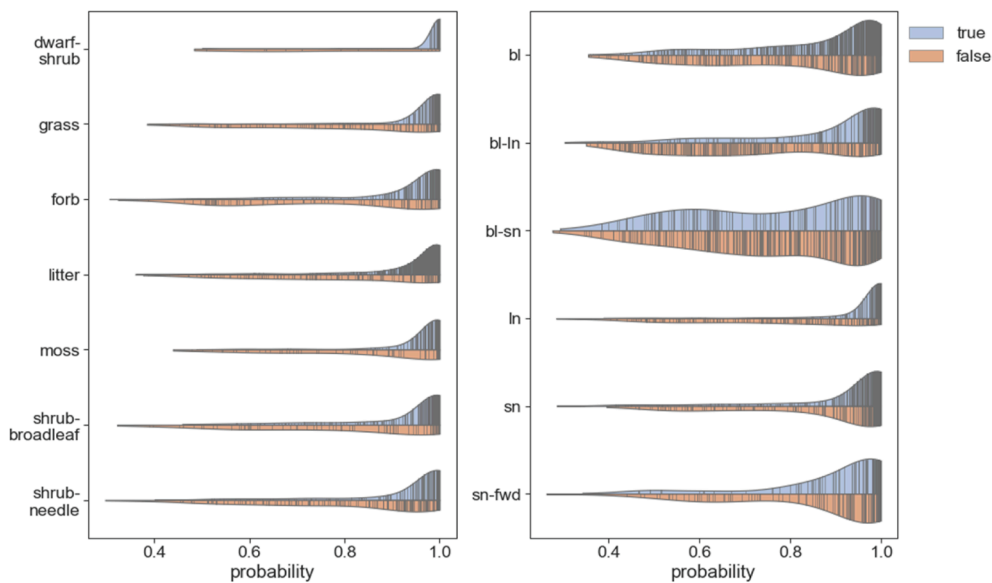


Fig. 10. Violin plots showing the distribution of prediction probabilities for understory fuel types (left) and litter fuel types (right) from the multi-input CNN trained on horizontal and forest floor photos. Blue: true predictions, orange: false predictions. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

classified correctly in almost all predictions based on forest photos. The other understory types were also identified well based on forest photos and in combination with Sentinel-2 time series, with a few confusions between *grass* and *forb* and between the classes *shrub-needle*, *moss* and *litter*.

3.3. Assessment of prediction probabilities

Correctly classified photos generally received a higher class assignment probability score than misclassified photos, indicating the model’s confidence for a correct class prediction (Fig. 10). Mean prediction probability was 95 % for correct understory predictions, whereas incorrect predictions had a mean probability of 80 %. Class-wise prediction probabilities were partly in line with the results from accuracy assessment, showing that classes with high f1-scores such as *dwarf-shrub* were more confidently predicted (99 % for true labels) than classes that were confused more frequently such as *forb* (93 % for true labels). For the classification of litter fuel types, prediction probabilities were generally lower, with an average of 88 % for correct and 79 % for incorrect predictions. Probability distributions in litter classification

highlight the model’s uncertainty with respect to mixed litter types and *sn-fwd*, suggesting difficulties in finding appropriate decision boundaries for the class assignment.

3.4. Effect of ensemble approaches on classification results

Post-classification aggregation procedures provided a means to improve final classification results. While understory classification improved up to an accuracy of 0.85 (baseline 0.78), litter classification improved up to an accuracy of 0.72 (baseline 0.60). Our analysis showed that understory fuel type classification was best improved using majority voting from multiple photographs of the same stand and additionally using only the most certain predictions (Fig. 11). Ensemble predictions based on the predictions from three single-input models (2 CNNs and LSTM) failed to improve classification results in case of understory (Table 3) when uncertain predictions were not omitted, due to the low accuracy of the Sentinel-2 predictions. In contrast, the ensemble prediction of litter types (Table 3) clearly outperformed the prediction resulting from majority voting based on multiple photographs (Fig. 11).

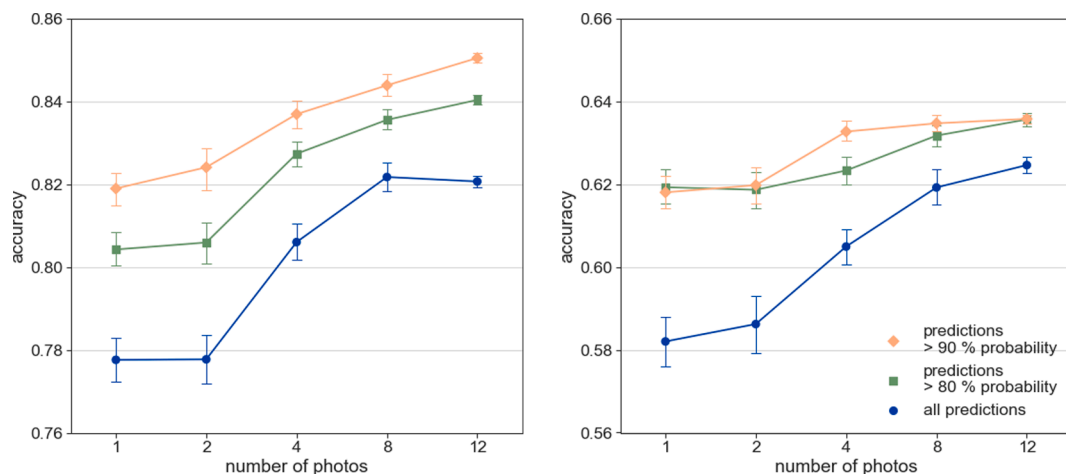


Fig. 11. Average classification accuracy depending on different numbers of photos used to determine class assignment for a forest stand by majority voting for understory fuel types (left) and litter fuel types (right, original classification with 6 litter types).

Table 3
Average classification accuracy for model ensemble predictions with and without filtering based on prediction probability.

	understory		litter	
	overall accuracy	no. of predicted instances	overall accuracy	no. of predicted instances
all predictions	0.72	3336	0.65	3336
predictions > 80 % probability	0.78	3175	0.70	2897
predictions > 90 % probability	0.81	2982	0.72	2530

3.5. Feature importance in LSTM model

Time series of Sentinel-2's shortwave infrared band (SWIR, B11) were most important in litter classification, followed by a narrow near infrared band (narrow NIR, B8A) and the second SWIR band (B12) (Fig. 12). Patterns changed for the simplified litter classes, where the blue band (B2), NDVI and SWIR (B11) were most important. Little importance in litter classification was given to the red band (B4), vegetation red edge bands (B5-7) and NIR band (B8). Understory classification relied on vegetation red edge (B8A), SWIR bands and NDVI.

Summer acquisitions were more important in all classification tasks than winter acquisitions. Litter classification strongly relied on the months July and August, while the most important dates for understory classification were slightly earlier in the year, in May and June.

3.6. Importance of image regions in CNN model

In many cases, the most salient image regions in understory classification coincide with the occurrence of (or parts of) the particular understory fuel type in the image, especially in forest floor photos (Fig. 13 a)). Color and texture appear to be important for the distinction of relevant from irrelevant image content (see for example *dwarf shrub, forb, litter*). However, sometimes only small regions are highlighted even though the fuel type covers large parts of the image (Fig. 13 b), for example *grass, forb, shrub-broadleaf*, or even the fuel type is not highlighted at all, but another image feature is (e.g., grass blade instead of *moss*). In most horizontal photos of *dwarf-shrub, grass* and *litter* fuel types (Fig. 13 c)), the bottom parts of the image are correctly identified as the relevant regions the model has to look for. The salient image regions in photos of *forb* fuel types seem to follow no clear patterns, while for *moss* either forest floor regions or stems are highlighted. In case of *shrubs-broadleaf* and *shrubs-needle*, mostly foreground image features such as branches and leaves seem to be relevant. For *litter* and *shrubs-needle*,

however, also stems in the image background can be a decisive feature (note the clear demarcation from the forest floor or foreground vegetation).

4. Discussion

4.1. Potential of forest photographs to classify surface fuel types

Our results showed that forest photographs are suitable to classify litter fuel types with moderate overall accuracy (60 %) and understory fuel types with fairly high accuracy (78 %) using CNNs. The small differences in performance using horizontal stand photos compared to forest floor photos indicate that both can be used for surface fuel type classification, depending on which of the two is available; and combining both can stabilize results and improve accuracy in case of understory fuel type classifications.

4.1.1. Litter fuel types

The good discrimination between the three basic types of short-needle, long-needle and broadleaf litter by our models show that CNNs are able to extract the necessary information that is relevant to estimate surface fire spread for the included forest types based on photographs. However, our results revealed that it is difficult to correctly identify mixtures of different litter types: The challenge lies in the almost continuous transition from litter accumulations consisting of only one type of litter to few mixed-in leaves of, e.g., broadleaf litter, to more balanced mixtures between different litter types, where all components are assumed to have an effect on fire behavior. Leaves of broadleaf litter in particular have a disproportionate influence on the appearance of a photograph compared to their actual abundance, leading to misclassifications also by human observers. This explains the frequent confusions of mixed types *bl-ln* and *bl-sn* with *bl*. However, the influence of mixtures of different litter types on the combustion process, and thus the level of detail required for litter characterization, remains to be investigated.

Learning critical features for litter discrimination, particularly from forest floor photos, is difficult also when the litter layer itself is not visible due to continuous understory vegetation. Although there are some relationships of litter types with understory vegetation, e.g., a continuous moss or herbaceous layer is rarely encountered underneath broadleaved trees in our target region Central Europe, we cannot be sure whether a model learns these patterns. It can be argued that in such cases litter is also less relevant for fire behavior than the understory fuel type; however, there are situations where both are important, for example a pine forest with grass understory will burn more intense than an oak forest with grass understory due to the greater heat release from

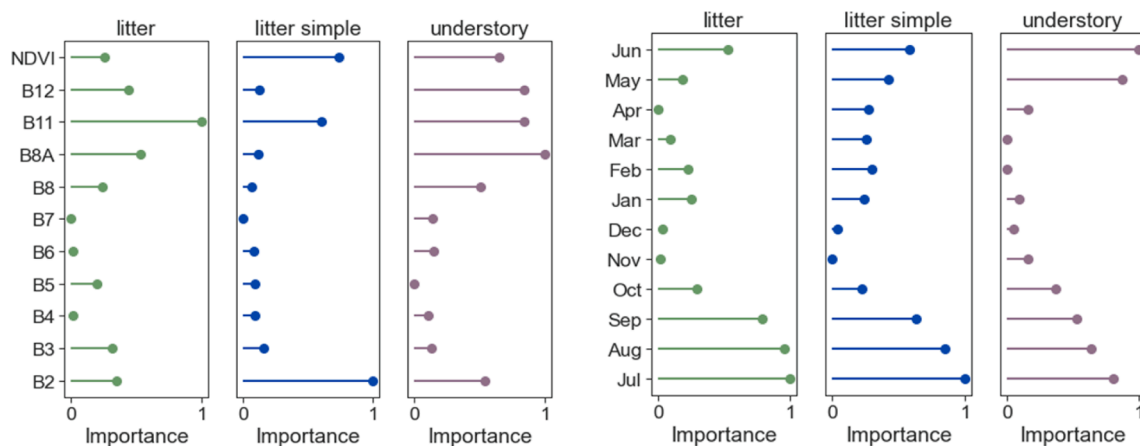


Fig. 12. Normalized relative feature importances calculated from decreases in overall classification accuracy when reflectance values of a Sentinel-2 band (left) or acquisition month (right) were randomly permuted.

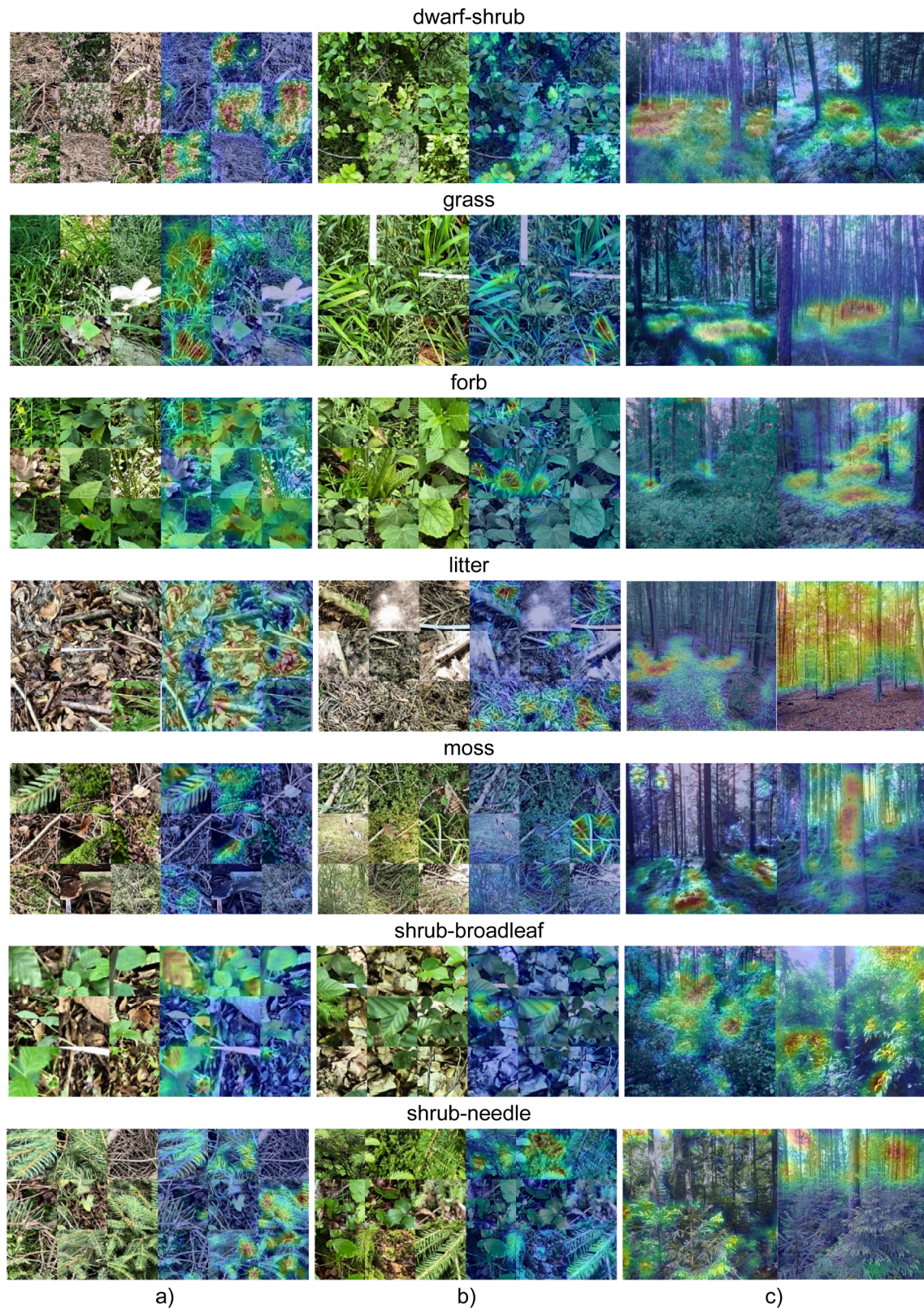


Fig. 13. Most salient image regions (colored in red) in the last convolutional layers of the CNN models for understory fuel type classification. Columns a) and b) show forest floor photos in the original (left) and overlaid with a Grad-CAM heatmap (right); column c) shows horizontal photos overlaid with a Grad-CAM heatmap. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

long-needle pine litter (Hough, 1969). Nevertheless, litter classification based on forest floor photos is expected to improve when excluding photos where the litter layer itself is not visible, whereas correct classifications based on horizontal photos may be still possible based on indirect relationships, e.g., to stems and crown morphologies of different tree species.

4.1.2. Understory fuel types

Understory fuel types were easier to distinguish based on forest photographs than litter fuel types. One reason for this might be that understory can be readily identified in an image and can have a unique appearance, such as in the case of *dwarf-shrub*. Since *dwarf-shrub* is the potentially most fire-prone understory fuel type included in this study (severe fires can occur in *Calluna vulgaris* habitats, e.g., Davies et al., 2010), its reliable detection by the CNN allows for the successful identification of high-risk forest areas. Fires can also spread rapidly through cured *grass* fuel types; here a better discrimination from the moister *forb* fuel types would be required, which is likely to be feasible with more training data. In other cases correct class attribution was more difficult because different understory fuel types appeared within the same photograph, e.g., *moss* and *shrub-needle*. This type of missclassification also occurred in the study by Xu et al. (2017), where a single land cover label was used for each photo. In the context of our study, the “confusion” of classes by the model merely reflects real-world conditions, if both fuel types contribute significantly to the fuel complex, and raises the question of whether a separation is actually meaningful in this case; or whether classification by presence/absence for different fuel types is more appropriate. However, a meaningful threshold for the minimum abundance of a fuel type to be effective in the context of fire behavior needs to be defined. Area-based thresholds like a minimum cover as in our study are commonly used in fuel type classifications (e.g., Arroyo et al., 2008), but these may be easier to detect from an aerial than a horizontal perspective. This is where photo interpretation (CNN-based or by humans) reaches its limits: multiple branches in the foreground of an image or a photograph taken from a path where there is sufficient light for understory vegetation to grow, will result in the visual impression of high understory cover, but this is not necessarily representative of the forest stand behind. Therefore, standardized requirements for photo acquisitions are needed to ensure representativeness. Avoiding acquisitions close to occluding objects, however, can also result in subjective and potentially biased sampling. Until other well-established means are available to assess in-forest understory vegetation from a more nadir perspective, variation in the “footprint” of a photograph with understory density and height needs to be taken into account. One way to overcome this problem in the future may be under-canopy drone acquisitions, which have recently been introduced (Kuzelka and Surový, 2018; Krisanski et al., 2020).

4.1.3. Comparison with other studies

We found no studies that have used forest photographs in a similar task before. Several studies have used CNNs to classify road view images in an agricultural context. For example, Ringland et al. (2019) characterized food production along roads in Thailand by using Google Street View (GSV) panoramas and achieved an overall accuracy of 83.3 % for seven different plant species. Yan and Ryu (2021) similarly employed GSV imagery to generate ground truth data for crop type mapping in the Central Valley in California, with an accuracy of 92 % for seven different crop types. Both studies used a considerably larger dataset than ours, with >2,000 images per class in the first study, and 500 to 1,000 images per class in the second study. Xu et al. (2017) used CNN-based feature extraction from 30,000 geo-tagged field photos in a multinomial logistic regression model to classify 19 land cover types, and achieved an accuracy of 48.4 % for top-1 prediction and 76.3 % for top-3 prediction (true class matches one of the three most probable predicted classes). Few studies focused on categorization problems using ground-taken imagery in a more ecological context. Habitat classification is one of

such tasks and has been addressed by extracting visual features and contextual information from ground photographs, feeding them into a random forest classifier and adding information about geographical closeness of the geo-referenced images (Torres and Qiu, 2016). Reported accuracy metrics range from f1-scores of about 0.2 for heathland to 0.7 for woodland and scrub habitats. Understory density has been estimated from understory images by distinguishing between vegetation-covered and background pixels using logistic regression on spectral variables (Campbell et al., 2018) or CNN-based segmentation (Abrams et al., 2019). However, these studies require that an artificial background is used during data collection to separate understory from background areas in the photographs.

Although the aforementioned studies differ substantially from our work in terms of research context, specific aims and employed learning algorithms, we assume that model results strongly rely on the dataset size available for the task, on data cleaning procedures and on the human effort involved in the correct annotation of the training data. The highly complex and heterogeneous data from natural environments further complicate the correct interpretation of images, even for human surveyors. Reducing this complexity by categorizing data allows more effective characterization and comparison, but class boundaries need to be set artificially and yet often remain fuzzy, making it difficult to clearly identify and separate classes. We consider this the most important limitation of our approach and it has remained largely unexplored how CNNs deal with such complexity outside of simple object recognition. We will discuss this further in chapter 4.4.2.

4.2. Effect of integrating Sentinel-2 time series and utility as stand-alone methodology

Our results indicate that Sentinel-2 time series alone are of limited use for surface fuel type classifications: While they were similarly useful as forest photographs for classifying litter fuel types, they were of little value for distinguishing understory fuel types. Integrating them with the forest photographs in a multi-input model did not notably improve classification results for both litter and understory.

4.2.1. Litter fuel types

Sentinel-2 predictions of litter fuel types rely on the spectral reflectance of the pixel(s) covering the field plots, which is dominated by overstory tree species (see also chapter 4.4.1). While tree species classification has been performed with good accuracies on multi-temporal Sentinel-2 data (Persson et al., 2018; Grabska et al., 2019), our study showed that predicting litter fuel types based on tree species information alone is difficult: small understory trees and shrubs can contribute significantly to the litter layer, and especially broadleaf litter from neighboring stands can be blown into a stand. Data that was recorded during field work showed that litter fuel types cannot be perfectly predicted based on the basal areas of the tree species present using a random forest classifier (OA = 0.68). This may explain why the litter fuel type classifications from Sentinel-2 time series achieved only moderate accuracy (OA = 0.59).

4.2.2. Understory fuel types

Understory characterization, especially species classification, based on remote sensing data is a challenging task, as shown by the few attempts that have been made so far (Hall et al., 2000; Korpela et al., 2008; Landry et al., 2020). The fact that Sentinel-2 time series are insufficient to distinguish understory fuel types is due to multiple reasons: first, the same fuel type may occur in the understory (e.g., *litter* or *grass*) under completely different tree species in the overstory; and even in case of open stands, the spectral signal from the understory is superimposed by overstory reflectance, resulting in a complex mixture of reflectance values contributing to the final pixel reflectance (Kobayashi et al., 2018; Singh and Gray, 2020). Second, it is not clear whether the spectral signal from small understory trees, e.g., regeneration of beech, differs

significantly from a closed-canopy of mature beech trees, which could explain confusions between *shrub-broadleaf* and *litter* understory fuel types. In such cases, integrating information on vertical forest structure derived from active remote sensing systems such as LiDAR would help to distinguish between overstory and understory vegetation. Including Sentinel-2 data can lead to small improvements in case of understory fuel types such as *moss* that are related with a certain type of overstory (mostly coniferous); yet effects are too small to justify the additional effort.

4.3. Improving classification results by ensemble approaches

Ensemble approaches helped notably to improve base model predictions. Our results showed that a forest stand can be characterized more reliably using multiple photographs from different perspectives and additionally using only the most certain predictions.

Our findings also showed that aggregating the predictions of several single-input models is more useful than using a multi-input model from the start, if all inputs have similar predictive power, such as in the case of litter classification. This could also be due to the greater difficulty in finding optimal hyperparameters for a complex model with multiple inputs, e.g., with respect to the best optimization algorithm, which may be different for the CNN and the LSTM branch of a model. In this sense it is recommended to optimize the smaller and less computationally demanding single-input models and then aggregate their predictions. However, there is still room for experimentation with different fusion schemes, as the increasing availability of multiple, heterogeneous datasets with different scales and dimensions for a given task has recently driven advances in deep multimodal learning (see review by Bayouh et al., 2021).

Using prediction probabilities as additional filter criterion further improves the results from decision fusion approaches, but always needs to be weighed against the associated discarding of data. At the same time, it can be worth to have a closer look at the more ‘unsure’ predictions: often the prediction probabilities contain much additional information, such as when a forest stand is actually better represented by a mixture of different fuel types than by a single one (Fig. 14). In any case, providing the prediction probabilities along with the predictions helps in assessing the reliability of the prediction.

4.4. Assessment of model explainability

4.4.1. Feature importance in LSTM model

Our results on variable importance of Sentinel-2 bands for classifying litter fuel types in the LSTM model (SWIR, blue band, see Figure S1 in supplementary material and NDVI) are consistent with previous studies that have classified tree species using multi-temporal Sentinel-2 data: Immitzer et al. (2016) similarly identified the SWIR band (related to leaf water content) and the blue band (absorbed by chlorophyll) as the two most important bands for tree species mapping in Germany, while the study of Persson et al. (2018) also ranked the red edge bands very high for tree species classification in Sweden. The latter were found to be

insignificant in our study, which could be related to the high correlation between these bands. Grabska et al. (2019) confirmed the importance of SWIR bands, red-edge bands, blue and red bands for the discrimination of tree species in the Polish Carpathians, while Ottosen et al. (2020) found that similar features (blue, green, red-edge, SWIR bands) were also most suited to map tree cover in Europe based on Sentinel-2 images, indicating that these bands are generally useful for mapping and differentiating canopy characteristics. Understory discrimination in our study relied somewhat more on NIR and SWIR bands, but a detailed discussion is omitted due to the rather low accuracy of the classification. The aforementioned studies mostly agreed that late spring and early summer acquisitions were most helpful for tree species discrimination, while our study revealed that midsummer acquisitions were more suitable for litter fuel type classifications; potentially due to fully developed tree canopies at this time of the year. Understory, however, is better identified earlier in the year, when phenological variations of the undergrowth may be more pronounced and better sensed through a less dense canopy. The choice of spectral variables in this study was guided by the aforementioned studies that attempted to map tree species and tree cover. However, other spectral indices have been found to be more sensitive to vegetation structure, such as the tasseled cap indices (especially the wetness feature) or the Normalized Difference Moisture Index (NDMI) (Cohen and Spies, 1992; Jin and Sader, 2005). Therefore, we trained another LSTM model on Sentinel-2 time series, adding tasseled cap wetness, tasseled cap greenness and NDMI, but did not observe any improvement in classification results (see Figure S2 in supplementary material).

4.4.2. Importance of image regions in CNN model

Due to the great heterogeneity of the input data in this study, it is challenging to assess what information from an image the CNN uses for its classification decision. Although it seems that the model generally responds to the parts of a photograph that also appear relevant to a human observer, there are still many cases where an (for the human observer) irrelevant image region drives the model towards the correct class decision. The concepts the model learns may be entirely different from what we expect in the first place; for example, we cannot be sure whether the decision for a *moss* or *litter* fuel type in a horizontal photo is actually driven by the texture and color of these two types, or whether the model is responding to coarse deadwood on the forest floor that is barely visible in photos of other understory fuel types. Since Grad-CAM heatmaps as well as other feature attribution algorithms are specific to the input image, displayed material will always reflect only a (potentially human-biased) minimal portion of the data, making it difficult to find generalizable rules. Visualizing the features the model responds to by generating synthetic images that maximize the activations of a particular convolutional filter reveals that the model mainly learns small-scale geometric features, even in late convolutional layers (Figure S3 in supplementary material). Although one might suspect that some of them resemble the shapes of leaves, small twigs or the texture of moss; such interpretations should be taken with caution (Kattenborn et al., 2021). Filter activations showed that maximally activated filters



Fig. 14. Predicted probabilities in understory fuel type classification based on horizontal photos. a) *grass* 0.59, *shrub-broadleaf* 0.23 and *forb* 0.18. b) *litter* 0.63 and *shrub-needle* 0.34. c) *litter* 0.44, *moss* 0.56.

are very similar for photos of different fuel types, since all contain plant parts, but in slightly different compositions.

4.5. Outlook

We have taken a first step towards the application of deep learning methods to classify surface fuel types for fire behavior and fire risk assessment from forest photographs and satellite time series. As with all deep learning problems, the availability of labeled training data is a bottleneck. To improve the capabilities of the model and to apply it to a larger geographical area, the dataset should be further expanded using (crowd-sourced) photos annotated by trained individuals. In our study, we identified the most common surface fuel types in temperate forests of Central Europe; however, the targeted fuel type classification scheme could be arbitrarily detailed, provided a large enough data set. For the validation of fuel type maps across larger areas, the challenge will be to obtain sufficient imagery also from remote locations and ensure quality in terms of geolocation accuracy. Incorporating point cloud data from ALS, TLS or drones, if available, could further improve or even refine fuel type classification, and also forest inventory data or biophysical factors could be included. The model itself could be improved by leveraging a more efficient architecture that requires less parameters, which would speed up training and inference times. Testing alternative approaches such as segmentation of, for example, understory vegetation on forest photographs is laborious, but could help the model to learn the relevant features and not be distracted by artifacts. Another exciting area of research would be to explore whether it is also possible to move away from classifications and retrieve quantitative information such as estimates of fuel loadings from a photograph. In addition, many other interesting use cases for forest photos are conceivable, just to mention forest health and biodiversity assessments, which have been already examined from photographs using other methodical approaches (e.g., Gyllin and Grahn, 2015; Murray et al., 2018).

In terms of practical applications, GNSS-located photos of forest stands obtained by local forest managers or through citizen science can be used not only to validate and improve fuel type maps, but also to provide forest practitioners and firefighters with immediate information about potential fire behavior at their location, for example via a cloud-based smartphone application: The extracted fuel type information could be used to approximate the available burnable biomass and to derive relevant physical properties that determine the combustion process in order to calculate fire behavior in a forest stand, e.g., under different moisture scenarios. Knowledge from fire experts could also be incorporated to help practitioners decide, for example, whether understory vegetation needs to be removed to reduce fire hazard in critical areas, or to understand the extent to which moist green vegetation can even serve as fire barrier. This would greatly advance knowledge exchange on fuel-related forest fire risk, particularly in temperate forests of Central Europe, which have been poorly studied in this regard to date.

5. Conclusions

In this work, we investigated the usefulness of deep neural networks (CNNs and LSTM) to classify surface fuel types of Central European forests based on within-stand photographs and Sentinel-2 time series. Our results demonstrated that understory fuel types can be classified with good accuracy from a combination of horizontal stand photos and forest floor photos using CNNs. Litter fuel types were classified with moderate accuracy from both types of photographs. The main limitation of the approach was the occurrence of multiple fuel types within the same photograph, leading to confusions especially in litter classification. Our study further showed that Sentinel-2 time series alone are

insufficient for understory classification, but that they have potential for litter fuel type classifications both as additional predictor in ensemble approaches and as stand-alone methodology when photographs of a forest stand are not available. The decisive spectral features were reflectance differences associated with canopy characteristics, manifested primarily in NDVI, SWIR and blue bands during summer. From a practical perspective, our research showed that a forest stand can be better characterized the more photos are available, especially concerning understory fuel types. For litter fuel types, it has proven useful to make predictions on multiple types of data separately, i.e., photographs and satellite time series, and combine the predictions of all models by majority voting. Class prediction probabilities were found to be a useful filter criterion for the most reliable predictions and provided insights into the complexity of fuel type composition in a forest stand. While our study has demonstrated that artificial intelligence can help with classification problems in complex natural environments, it has also shown that the model's capabilities are limited by fuzzy class boundaries, as humans are; and although influential image regions in CNNs often contain features that appear relevant to the observer (i.e. the respective fuel), we are unable to fully comprehend the model's decisions. Translating the task into a regression problem to quantify individual fuel components could help deal with natural gradients, but would also require extensive collection of reference data. Nonetheless, results from this study indicate that automatic processing of within-stand photographs by CNNs has the potential to facilitate validation of fuel type maps and provide forest practitioners with the information needed to mitigate fire hazard. We hope that our work can contribute to opening a new field of research for deep learning-based applications to characterize forest fuels for fire behavior and risk assessment in light of the increasing threat of wildfires, even in temperate forests, under a changing climate.

CRedit authorship contribution statement

Pia Labenski: Conceptualization, Methodology, Formal analysis, Writing – original draft. **Michael Ewald:** Conceptualization, Writing – review & editing. **Sebastian Schmidlein:** Writing – review & editing. **Fabian Ewald Fassnacht:** Conceptualization, Writing – review & editing, Supervision, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was funded by the German Federal Ministry of Food and Agriculture and Federal Ministry for the Environment, Nature Conservation and Nuclear Safety as part of the project “ErWiN” (code 2219WK54A4). The authors acknowledge support by the state of Baden-Württemberg through bwHPC. We would like to thank Paulo Fernandes and Davide Ascoli for their valuable support in developing the fuel type classification. We also want to thank Gabriel Cabezas for his help with the fieldwork and Felix Schiefer for fruitful discussions on deep learning issues. Furthermore, we are grateful to the anonymous reviewers, who gave valuable comments to improve our manuscript. We acknowledge support by the KIT-Publication Fund of the Karlsruhe Institute of Technology.

Appendix A. Supplementary material

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jag.2022.102799>.

References

- Abatzoglou, J.T., Williams, A.P., Barbero, R., 2019. Global Emergence of Anthropogenic Climate Change in Fire Weather Indices. *Geophys. Res. Lett.* 46 (1), 326–336.
- Abrams, J.F., Vashishtha, A., Wong, S.T., Nguyen, A.N., Mohamed, A., Wieser, S., Kuijper, A., Wilting, A., Mukhopadhyay, A., 2019. Habitat-Net: Segmentation of habitat images using deep learning. *Ecol. Inf.* 51, 121–128.
- Alonso-Benito, A., Arroyo, L., Arbelo, M., Hernández-Leal, P., 2016. Fusion of WorldView-2 and LiDAR Data to Map Fuel Types in the Canary Islands. *Remote Sensing* 8 (8), 669.
- Andrews, P.L., 2014. Current status and future needs of the BehavePlus Fire Modeling System. *International Journal of Wildland Fire* 23 (1), 21–33.
- Andrews, P.L., Queen, L.P., 2001. Fire modeling and information system technology. *International Journal of Wildland Fire* 10 (4), 343–352.
- Arroyo, L.A., Pascual, C., Manzanera, J.A., 2008. Fire models and methods to map fuel types: The role of remote sensing. *For. Ecol. Manage.* 256 (6), 1239–1252.
- Bayouh, K., Knani, R., Hamdaoui, F., Mtibaa, A., 2021. A survey on deep multimodal learning for computer vision: advances, trends, applications, and datasets. *The Visual Computer* 1–32.
- Bayr, U., Puschmann, O., 2019. Automatic detection of woody vegetation in repeat landscape photographs using a convolutional neural network. *Ecol. Inf.* 50, 220–233.
- Bohmann, J., Keeling, C.I., 2008. Terpenoid biomarkers. *Plant J.* 54 (4), 656–669.
- Botequin, B., Fernandes, P.M., Borges, J.G., González-Ferreiro, E., Guerra-Hernández, J., 2019. Improving silvicultural practices for Mediterranean forests through fire behaviour modelling using LiDAR-derived canopy fuel characteristics. *International Journal of Wildland Fire* 28 (11), 823–839.
- Campbell, M.J., Dennison, P.E., Hudak, A.T., Parham, L.M., Butler, B.W., 2018. Quantifying understory vegetation density using small-footprint airborne lidar. *Remote Sens. Environ.* 215, 330–342.
- Campos-Taberner, M., García-Haro, F.J., Martínez, B., Izquierdo-Verdiguier, E., Atzberger, C., Camps-Valls, G., Gilabert, M.A., 2020. Understanding deep learning in land use classification based on Sentinel-2 time series. *Sci. Rep.* 10 (1), 17188.
- Cao, R., Zhu, J., Tu, W., Li, Q., Cao, J., Liu, B., Zhang, Q., Qiu, G., 2018. Integrating Aerial and Street View Images for Urban Land Use Classification. *Remote Sensing* 10 (10), 1553.
- Chen, L.-C., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A.L., 2018. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. *IEEE Trans. Pattern Anal. Mach. Intell.* 40 (4), 834–848.
- Chen, X., Xiang, S., Liu, C.-L., Pan, C.-H., 2014. Vehicle Detection in Satellite Images by Hybrid Deep Convolutional Neural Networks. *IEEE Geosci. Remote Sens. Lett.* 11 (10), 1797–1801.
- Chirici, G., Scotti, R., Montagni, A., Barbati, A., Cartisano, R., Lopez, G., Marchetti, M., McRoberts, R.E., Olsson, H., Corona, P., 2013. Stochastic gradient boosting classification trees for forest fuel types mapping through airborne laser scanning and IRS LISS-III imagery. *Int. J. Appl. Earth Obs. Geoinf.* 25, 87–97.
- Chollet, F., 2015. Keras. <https://github.com/fchollet/keras>.
- Christin, S., Hervet, É., Lecomte, N., Ye, H., 2019. Applications for deep learning in ecology. *Methods Ecol. Evol.* 10 (10), 1632–1644.
- Chuvieco, E., Riaño, D., van Wageningen, J., Morsdorf, F., 2003. Fuel Loads and Fuel Type Mapping. In: Chuvieco, E. (Ed.), *Wildland fire danger estimation and mapping. The role of remote sensing data*. World Scientific, Singapore, pp. 119–142.
- Cohen, W.B., Spies, T.A., 1992. Estimating structural attributes of Douglas-fir/western hemlock forest stands from landsat and SPOT imagery. *Remote Sens. Environ.* 41 (1), 1–17.
- Correia, D.L.P., Bouachir, W., Gervais, D., Pureswaran, D., Kneeshaw, D.D., de Grandpre, L., 2020. Leveraging Artificial Intelligence for Large-Scale Plant Phenology Studies From Noisy Time-Lapse Images. *IEEE Access* 8, 13151–13160.
- Countryman, C.M., 1964. Mass Fires and Fire Behavior. Pacific Southwest Forest and Range Experiment Station, Forest Service, U.S. Department of Agriculture.
- Davies, M.G., Smith, A.A., MacDonald, A.J., Bakker, J.D., Legg, C.J., 2010. Fire intensity, fire severity and ecosystem response in heathlands: factors affecting the regeneration of *Calluna vulgaris*. *J. Appl. Ecol.* 47 (2), 356–365.
- de Rigo, D., Libertà, G., Durrant, T.H., Vivanco, T.A., San-Miguel-Ayanz, J., 2017. Forest fire danger extremes in Europe under climate change: variability and uncertainty. *Diss., Publications Office of the European Union*.
- Domingo, D., de La Riva, J., Lamelas, M.T., García-Martín, A., Ibarra, P., Echeverría, M., Hoffrén, R., 2020. Fuel Type Classification Using Airborne Laser Scanning and Sentinel 2 Data in Mediterranean Forest Affected by Wildfires. *Remote Sensing* 12 (21), 3660.
- Erdody, T.L., Moskal, L.M., 2010. Fusion of LiDAR and imagery for estimating forest canopy fuels. *Remote Sens. Environ.* 114 (4), 725–737.
- European Commission, 2021. EFFIS Statistics. <https://effis.jrc.ec.europa.eu/apps/effis-statistics.portal/>. (Accessed 27 October, 2021).
- Finney, M.A., 2006. An Overview of FlamMap Fire Modeling Capabilities. In: Andrews, P. L., Butler, B.W. (Eds.), *Fuels Management - How to Measure Success: Conference Proceedings*. U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station, pp. 213–220.
- Forzieri, G., Girardello, M., Ceccherini, G., Spinoni, J., Feyen, L., Hartmann, H., Beck, P. S.A., Camps-Valls, G., Chirici, G., Mauri, A., Cescatti, A., 2021. Emergent vulnerability to climate-driven disturbances in European forests. *Nat. Commun.* 12 (1), 1081.
- Gao, J., French, A.P., Pound, M.P., He, Y., Pridmore, T.P., Pieters, J.G., 2020. Deep convolutional neural networks for image-based *Convolvulus sepium* detection in sugar beet fields. *Plant Methods* 16, 29.
- García, M., Riaño, D., Chuvieco, E., Salas, J., Danson, F.M., 2011. Multispectral and LiDAR data fusion for fuel type mapping using Support Vector Machine and decision rules. *Remote Sens. Environ.* 115 (6), 1369–1379.
- Goëau, H., Boujemaa, N., Molino, J.-F., Duché, G., Péronnet, A., Bonnet, P., Joly, A., Bakic, V., Barbe, J., Yahiaoui, L., Selmi, S., Carré, J., Barthélémy, D., 2013. Pl@ntNet mobile app. ACM Press, New York, New York, USA.
- Grabska, E., Hostert, P., Pflugmacher, D., Ostapowicz, K., 2019. Forest Stand Species Mapping Using the Sentinel-2 Time Series. *Remote Sensing* 11 (10), 1197.
- Guirado, E., Alcaraz-Segura, D., Cabello, J., Puertas-Ruiz, S., Herrera, F., Tabik, S., 2020. Tree Cover Estimation in Global Drylands from Space Using Deep Learning. *Remote Sensing* 12 (3), 343.
- Gyllin, M., Grahn, P., 2015. Semantic assessments of experienced biodiversity from photographs and on-site observations - a comparison. *Environment and Natural Resources Research* 5 (4), 46–62.
- Hall, R.J., Peddle, D.R., Klita, D.L., 2000. Mapping conifer understory within boreal mixedwoods from Landsat TM satellite imagery and forest inventory information. *The Forestry Chronicle* 76 (6), 887–902.
- Hocheiter, S., Schmidhuber, J., 1997. Long short-term memory. *Neural Comput.* 9 (8), 1735–1780.
- Hough, W.A., 1969. Caloric Value of Some Forest Fuels of the Southern United States. U. S. Department of Agriculture, Forest Service, Southeastern Forest Experiment Station, Asheville, NC.
- Immitzer, M., Vuolo, F., Atzberger, C., 2016. First Experience with Sentinel-2 Data for Crop and Tree Species Classifications in Central Europe. *Remote Sensing* 8 (3), 166.
- IPCC, 2019. *Climate Change and Land: an IPCC special report on climate change, desertification, land degradation, sustainable land management, food security, and greenhouse gas fluxes in terrestrial ecosystems*. In: Shukla, P.R., Skea, J., Calvo Buendia, E., Masson-Delmotte, V., Pörtner, H.O., Roberts, D.C., Zhai, P., Slade, R., Connors, S., van Diemen, R., Ferrat, M., Haughey, E., Luz, S., Neogi, S., Pathak, M., Petzold, J., Portugal Pereira, J., Vyas, P., Huntley, E., Kissick, K., Belkacemi, M., Malley, J. (Eds.), In press.
- Jia, G.J., Burke, I.C., Goetz, A.F.H., Kaufmann, M.R., Kindel, B.C., 2006. Assessing spatial patterns of forest fuel using AVIRIS data. *Remote Sens. Environ.* 102 (3–4), 318–327.
- Jin, S., Sader, S.A., 2005. Comparison of time series tasseled cap wetness and the normalized difference moisture index in detecting forest disturbances. *Remote Sens. Environ.* 94 (3), 364–372.
- Jolly, W.M., Cochrane, M.A., Freeborn, P.H., Holden, Z.A., Brown, T.J., Williamson, G.J., Bowman, D.M.J.S., 2015. Climate-induced variations in global wildfire danger from 1979 to 2013. *Nat. Commun.* 6, 7537.
- Kahl, S., Wood, C.M., Eibl, M., Klinck, H., 2021. BirdNET: A deep learning solution for avian diversity monitoring. *Ecol. Inf.* 61, 101236.
- Kälén, U., Lang, N., Hug, C., Gessler, A., Wegner, J.D., 2019. Defoliation estimation of forest trees from ground-level images. *Remote Sens. Environ.* 223, 143–153.
- Kattenborn, T., Leitloff, J., Schiefer, F., Hinz, S., 2021. Review on Convolutional Neural Networks (CNN) in vegetation remote sensing. *ISPRS J. Photogramm. Remote Sens.* 173, 24–49.
- Keane, R.E., 2013. Describing wildland surface fuel loading for fire management: a review of approaches, methods and systems. *International Journal of Wildland Fire* 22 (1), 51–62.
- Keane, R.E. (Ed.), 2015. *Wildland Fuel Fundamentals and Applications*. Springer International Publishing, Cham.
- Keramitsoglou, I., Kontoes, C., Sykioti, O., Sifakis, N., Xofis, P., 2008. Reliable, accurate and timely forest mapping for wildfire management using ASTER and Hyperion satellite imagery. *For. Ecol. Manage.* 255 (10), 3556–3562.
- Khabarov, N., Krasovskii, A., Obersteiner, M., Swart, R., Dosio, A., San-Miguel-Ayanz, J., Durrant, T., Camia, A., Migliavacca, M., 2016. Forest fires and adaptation options in Europe. *Reg. Environ. Change* 16 (1), 21–30.
- Kobayashi, H., Nagai, S., Kim, Y., Yang, W., Ikeda, K., Ikawa, H., Nagano, H., Suzuki, R., 2018. In Situ Observations Reveal How Spectral Reflectance Responds to Growing Season Phenology of an Open Evergreen Forest in Alaska. *Remote Sensing* 10 (7), 1071.
- Korpela, I., Tuomola, T., Tokola, T., Dahlin, B., 2008. Appraisal of seedling stand vegetation with airborne imagery and discrete-return LiDAR – an exploratory analysis. *Silva Fennica* 42 (5), 753–772.
- Krisanski, S., Taskhiri, M., Turner, P., 2020. Enhancing Methods for Under-Canopy Unmanned Aircraft System Based Photogrammetry in Complex Forests for Tree Diameter Measurement. *Remote Sensing* 12 (10), 1652.
- Krizhevsky, A., Sutskever, I., Hinton, G.E., 2017. ImageNet classification with deep convolutional neural networks. *Commun. ACM* 60 (6), 84–90.
- Kuzelka, K., Surový, P., 2018. Mapping Forest Structure Using UAS inside Flight Capabilities. *Sensors* 18 (7), 2245.
- Landry, S., St-Laurent, M.-H., Pelletier, G., Villard, M.-A., 2020. The Best of Both Worlds? Integrating Sentinel-2 Images and Airborne LiDAR to Characterize Forest Regeneration. *Remote Sensing* 12 (15), 2440.
- Langford, Z.L., Kumar, J., Hoffman, F.M., Breen, A.L., Iversen, C.M., 2019. Arctic Vegetation Mapping Using Unsupervised Training Datasets and Convolutional Neural Networks. *Remote Sensing* 11 (1), 69.
- Lasaponara, R., Lanorte, A., 2007. Remotely sensed characterization of forest fuel types by using satellite ASTER data. *Int. J. Appl. Earth Obs. Geoinf.* 9 (3), 225–234.

- Liu, J., Wang, X., Wang, T., 2019. Classification of tree species and stock volume estimation in ground forest images using Deep Learning. *Comput. Electron. Agric.* 166, 105012.
- Long, J., Shelhamer, E., Darrell, T., 2015. Fully convolutional networks for semantic segmentation. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3431–3440.
- McKenzie, D., Raymond, C.L., Kellogg, L.-K.-B., Norheim, R.A., Andreu, A.G., Bayard, A. C., Kopper, K.E., Elman, E., 2007. Mapping fuels at multiple scales: landscape application of the Fuel Characteristic Classification System. This article is one of a selection of papers published in the Special Forum on the Fuel Characteristic Classification System. *Can. J. For. Res.* 37 (12), 2421–2437.
- McKinney, W., 2010. Data Structures for Statistical Computing in Python. In: *Proceedings of the 9th Python in Science Conference*, pp. 51–56.
- Merrill, D.F., Alexander, M.E., 1987. Glossary of forest fire management terms, 4th ed. Canadian Committee on Forest Fire Management, National Research Council of Canada, Ottawa, Ontario.
- Millar, C.I., Stephenson, N.L., 2015. Temperate forest health in an era of emerging megadisturbance. *Science* 349 (6250), 823–826.
- Morfin-Rios, J.E., Alvarado-Celestino, E., Jardel-Pelaez, E.J., Vihnanek, R.E., Wright, D. K., Michel-Fuentes, J.M., Wright, C.S., Ottmar, R.D., Sandberg, D.V., Najera-Diaz, A., 2008. Photo series for quantifying forest fuels in Mexico: montane subtropical forests of the Sierra Madre del Sur and temperate forests and montane shrubland of the northern Sierra Madre Oriental. *Pacific Wildland Fire Sciences Laboratory Special Pub. No. 1*. Seattle, WA: University of Washington, College of Forest Resources. 93 p.
- Murray, J., Blackburn, G.A., Whyatt, J.D., Edwards, C., 2018. Using fractal analysis of crown images to measure the structural condition of trees. *Forestry: An International Journal of Forest Research* 91 (4), 480–491.
- Mutlu, M., Popescu, M., Stripling, C., Spencer, T., 2008. Mapping surface fuel models using lidar and multispectral data fusion for fire behavior. *Remote Sens. Environ.* 112 (1), 274–285.
- Ottmar, R.D., Vihnanek, R.E., Miranda, H.S., Sato, M.N., Andrade, S.M.A., 2004. Stereo photo series for quantifying biomass for the Cerrado vegetation in central Brazil. *Floresta* 34 (2).
- Ottosen, T.-B., Petch, G., Hanson, M., Skjøth, C.A., 2020. Tree cover mapping based on Sentinel-2 images demonstrate high thematic accuracy in Europe. *Int. J. Appl. Earth Obs. Geoinf.* 84, 101947.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E., 2011. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research* 12, 2825–2830.
- Perry, D.A., Choquette, C., Schroeder, P., 1987. Nitrogen dynamics in conifer-dominated forests with and without hardwoods. *Can. J. For. Res.* 17 (11), 1434–1441.
- Persson, M., Lindberg, E., Reese, H., 2018. Tree Species Classification with Multi-Temporal Sentinel-2 Data. *Remote Sensing* 10 (11), 1794.
- Philpot, C.W., 1970. Influence of Mineral Content on the Pyrolysis of Plant Materials. *Forest Science* 16 (4), 461–471.
- Pizer, S.M., Amburn, E.P., Austin, J.D., Cromartie, R., Geselowitz, A., Greer, T., ter Haar Romeny, B., Zimmerman, J.B., Zuiderveld, K., 1987. Adaptive histogram equalization and its variations. *Computer Vision, Graphics, and Image Processing* 39 (3), 355–368.
- Riano, D., 2003. Modeling airborne laser scanning data for the spatial generation of critical forest parameters in fire behavior modeling. *Remote Sens. Environ.* 86 (2), 177–186.
- Riano, D., Chuvieco, E., Salas, J., Palacios-Orueta, A., Bastarrika, A., 2002. Generation of fuel type maps from Landsat TM images and ancillary data in Mediterranean ecosystems. *Can. J. For. Res.* 32 (8), 1301–1315.
- Ringland, J., Bohm, M., Baek, S.-R., 2019. Characterization of food cultivation along roadside transects with Google Street View imagery and deep learning. *Comput. Electron. Agric.* 158, 36–50.
- Sandberg, D.V., Ottmar, R.D., Cushon, G.H., 2001. Characterizing fuels in the 21st Century. *International Journal of Wildland Fire* 10 (4), 381–387.
- Savitzky, A., Golay, M.J.E., 1964. Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Anal. Chem.* 36 (8), 1627–1639.
- Schiefer, F., Kattenborn, T., Frick, A., Frey, J., Schall, P., Koch, B., Schmidlein, S., 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.
- Schiller, C., Schmidlein, S., Boonman, C., Moreno-Martínez, A., Kattenborn, T., 2021. Deep learning and citizen science enable automated plant trait predictions from photographs. *Sci. Rep.* 11 (1), 16395.
- Schuster, M., Paliwal, K.K., 1997. Bidirectional recurrent neural networks. *IEEE Trans. Signal Process.* 45 (11), 2673–2681.
- Scott, N.A., Binkley, D., 1997. Foliage litter quality and annual net N mineralization: comparison across North American forest sites. *Oecologia* 111 (2), 151–159.
- Seidl, R., Schelhaas, M.-J., Rammer, W., Verkerke, P.J., 2014. Increasing forest disturbances in Europe and their impact on carbon storage. *Nat. Clim. Change* 4 (9), 806–810.
- Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D., 2017. Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. In: *Proceedings of the IEEE international conference on computer vision*, pp. 618–626.
- Sikkink, P.G., Keane, R.E., 2008. A comparison of five sampling techniques to estimate surface fuel loading in montane forests. *International Journal of Wildland Fire* 17 (3), 363–379.
- Simonyan, K., Zisserman, A., 2015. Very Deep Convolutional Networks for Large-Scale Image Recognition. In: *Conference Track Proceedings of the 3rd International Conference on Learning Representations*.
- Singh, K.K., Gray, J., 2020. Mapping Understorey Invasive Plants in Urban Forests with Spectral and Temporal Unmixing of Landsat Imagery. *Photogramm. Eng. Remote Sens.* 86 (8), 509–518.
- Sladojevic, S., Arsenovic, M., Anderla, A., Culibrk, D., Stefanovic, D., 2016. Deep Neural Networks Based Recognition of Plant Diseases by Leaf Image Classification. *Computational Intelligence and Neuroscience* 2016, 1–11.
- Spinoni, J., Vogt, J.V., Naumann, G., Barbosa, P., Dosio, A., 2018. Will drought events become more frequent and severe in Europe? *Int. J. Climatol.* 38 (4), 1718–1736.
- Tompson, J., Goroshin, R., Jain, A., LeCun, Y., Bregler, C. Efficient Object Localization Using Convolutional Networks. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 648–656.
- Torres, M., Qiu, G., 2016. Habitat image annotation with low-level features, medium-level knowledge and location information. *Multimedia Syst.* 22 (6), 767–782.
- van Rossum, G., Drake, F.L., 2009. Python 3 Reference Manual. CreateSpace, Scotts Valley, CA.
- Varner, J.M., Kane, J.M., Kreye, J.K., Engber, E., 2015. The Flammability of Forest and Woodland Litter: a Synthesis. *Current Forestry Reports* 1 (2), 91–99.
- Veira, A., Lasslop, G., Kloster, S., 2016. Wildfires in a warmer climate: Emission fluxes, emission heights, and black carbon concentrations in 2090–2099. *Journal of Geophysical Research: Atmospheres* 121 (7), 3195–3223.
- Vihnanek, R.E., Balog, C.S., Wright, C.S., Ottmar, R.D., Kelly, J.W., 2009. Stereo photo series for quantifying natural fuels. Volume XII: Post-hurricane fuels in forests of the Southeast United States. Gen. Tech. Rep. PNW-GTR-803, Portland, OR.
- Wäldchen, J., Mäder, P., Cooper, N., 2018. Machine learning for image based species identification. *Methods Ecol. Evol.* 9 (11), 2216–2225.
- Wang, S., Di Tommaso, S., Faulkner, J., Friedel, T., Kennepohl, A., Strey, R., Lobell, D.B., 2020. Mapping Crop Types in Southeast India with Smartphone Crowdsourcing and Deep Learning. *Remote Sensing* 12 (18), 2957.
- Woodall, C.W., Monleon, V.J., 2008. Sampling protocol, estimation, and analysis procedures for the down woody materials indicator of the FIA program. U.S. Department of Agriculture, Forest Service, Northern Research Station, Newtown Square, PA.
- Wright, C.S., Eagle, P.C., Olson, D.L., 2010. A high-quality fuels database of photos and information. *Fire Management Today*. Forest Service of the U.S. Department of Agriculture, Washington, DC. 7 p. 70(3) 27-31, 70 (3), 27–31.
- Xi, Y., Ren, C., Tian, Q., Ren, Y., Dong, X., Zhang, Z., 2021. Exploitation of Time Series Sentinel-2 Data and Different Machine Learning Algorithms for Detailed Tree Species Classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 14, 7589–7603.
- Xu, G., Zhu, X., Fu, D., Dong, J., Xiao, X., 2017. Automatic land cover classification of geo-tagged field photos by deep learning. *Environ. Modell. Software* 91, 127–134.
- Yan, Y., Ryu, Y., 2021. Exploring Google Street View with deep learning for crop type mapping. *ISPRS J. Photogramm. Remote Sens.* 171, 278–296.
- Zhong, L., Hu, L., Zhou, H., 2019. Deep learning based multi-temporal crop classification. *Remote Sens. Environ.* 221, 430–443.