Spatial and temporal deep learning methods for deriving land-use following deforestation: A pan-tropical case study using Landsat time series

Robert N. Masolele [a,*], Veronique De Sy [a], Martin Herold [a], Diego Marcos [a], Jan Verbesselt [a], Fabian Gieseke [b], Adugna G. Mullissa [a], Christopher Martius [c]

[a] Laboratory of Geo-Information Science and Remote Sensing, Wageningen University & Research, Droevendaalsesteeg 3, 6708 PB Wageningen, the Netherlands
[b] Department of Information Systems, University of Münster, Leonardo-Campus 3, 48149 Münster, Germany
[c] Center for International Forestry Research (CIFOR), Germany GmbH, 53113 Bonn, Germany

ABSTRACT

Assessing land-use following deforestation is vital for reducing emissions from deforestation and forest degradation. In this paper, for the first time, we assess the potential of spatial, temporal and spatio-temporal deep learning methods for large-scale classification of land-use following tropical deforestation using dense satellite time series over six years on the pan-tropical scale (incl. Latin America, Africa, and Asia). Based on an extensive reference database of six forest to land-use conversion types, we find that the spatio-temporal models achieved a substantially higher F1-score accuracies than models that account only for spatial or temporal patterns. Although all models performed better when the scope of the problem was limited to a single continent, the spatial models were more competitive than the temporal ones in this setting. These results suggest that the spatial patterns of land-use within a continent share more commonalities than the temporal patterns and the spatial patterns across continents. This work explores the feasibility of extending and complementing previous efforts for characterizing follow-up land-use after deforestation at a small-scale via human visual interpretation of high resolution RGB imagery. It supports the usage of fast and automated large-scale land-use classification and showcases the value of deep learning methods combined with spatio-temporal satellite data to effectively address the complex tasks of identifying land-use following deforestation in a scalable and cost effective manner.

1. Introduction

Land-use change is the second-largest contributor to greenhouse gas (GHG) emissions globally (IPCC, 2013), and in total, 24% of global greenhouse gas emissions come from deforestation activities(FAO, 2014). In response, the United Nations Framework Convention on Climate Change (UNFCCC) established a framework to reduce emissions from deforestation and forest degradation and enhance carbon stocks (REDD+) by result-based payments (UNFCCC, 2017). Before payments are made countries are required to show that emissions were reduced through a clear methodological and well-documented Measuring, Reporting, and Verification system (MRV) (IPCC, 2013; UNFCCC, 2018). A robust deforestation monitoring system can also support more informative and effective land-use policies and measures (UNFCCC, 2018) by monitoring what land-use activities drive deforestation. These land-use activities, i.e., proximate or direct drivers of deforestation (Geist and

Lambin, 2001), can be assessed using Earth Observation Technologies (EOT) to help provide spatially explicit and temporal information on land-use (Curtis et al., 2018; De Sy et al., 2015, 2019). However, these studies detect land-use following deforestation at coarse thematic, spatial and temporal scales or use time-consuming methods (i.e., visual interpretation of satellite imagery) which makes these approaches less suited for national level operational monitoring. Recent advances in Earth Observation (EO), computing technology and deep learning methods provide opportunities for automated large-scale assessment of land-use following tropical deforestation at more detailed spatial and temporal scales.

The latest advances and investments in Earth Observation Programmes (EOP) for global environmental data acquisitions, such as the European Copernicus EOP (Sentinel-1, Sentinel-2A, —2B) and joint NASA and U.S. Geological Survey program (Landsat 1–5, 7 & 8), have allowed the assessment of global forest cover change (Hansen et al.,

2010a; Kim, 2010), classification of land cover (Balzter et al., 2015; Cole et al., 2018), and the analysis of time series of satellite data to study changes on the land surface (Arévalo et al., 2019; Pandey et al., 2018; Verbesselt et al., 2010a). The EOP's are accompanied by free data policies for data retrieval and use (i.e., Landsat, MODIS, Sentinel). The EOP provides an opportunity to advance science in monitoring land-use essential for REDD+ (Boriah, 2010). On top of that, the increase in the spatio-temporal resolution of EO data enables the detection and classification of smaller and more subtle changes on the land surface with greater accuracy (Zhang et al., 2019).

Traditional land-use classification approaches with remote sensing images use the standard supervised/unsupervised machine learning methods such as K nearest neighbors, maximum likelihood estimators, support vector machines, or random forests. This is a mono-temporal approach exploiting the spectral aspect of land-use data (Curtis et al., 2018; Hansen et al., 2010a; Nguyen et al., 2018; Silva et al., 2018). However, these methods are not designed to work with time-series data nor with pattern recognition of land-use (Arévalo et al., 2019; Comber et al., 2016; Curtis et al., 2018; Huang et al., 2018; Mas et al., 2017; Pelletier et al., 2019) and, therefore, ignore the temporal and spatial dependency which are essential in retaining information to differentiate land-uses (Curtis et al., 2018; Castilla and Hay, 2007; Pelletier et al., 2019).

Recently there have been advances in computing technology, cloud computing, and high-performance computing paralleled with those in advanced artificial intelligence (machine learning and, particularly, deep learning) algorithms (Helber et al., 2017; Lecun et al., 2015; Luus et al., 2015; Marmanis et al., 2016; Neagoe et al., 2012; Tracewski et al., 2017). Deep learning enables the automatic detection of complex spatial and temporal patterns in environmental data such as time series of satellite images based on training data (Hughes et al., 2018; Kit and Lüdeke, 2013; Minh et al., 2018; Pelletier et al., 2019). In contrast to the traditional focus on spectral information in remote sensing imagery, the ability to capture spatial-temporal patterns allows the detection and differentiation of types of land cover and land-use with closely resembling spectral signatures (Comber et al., 2016). Despite the success of machine learning and deep learning (Castelluccio et al., 2015; Rußwurm and Körner, 2018; Zhang et al., 2019; Zhu et al., 2017) in extracting spectral, spatial, and temporal features from remote sensing data, the methods are not widely used for large-scale remote sensing analyses. However, promising results have been reported on small-scale problem applications for satellite image classification, land-cover/land-use mapping and land cover/land-use change detection (Castelluccio et al., 2015; Helber et al., 2017; Interdonato et al., 2018; Kong et al., 2018; Liu et al., 2016; Song et al., 2018; Uba, 2016). Additional challenges arise when extrapolating results to a large-scale, as models trained on small areas and evaluated over large areas will often not perform satisfactorily (Kellenberger et al., 2018; Zhu et al., 2017; Yuan et al., 2020).

Furthermore, the assessment of land-use is often rather generic and at coarse levels (Curtis et al., 2018; Hansen et al., 2010b), and does not fully differentiate land-use from land-cover (Bp et al., 2015; Campbell et al., 2005; Mas et al., 2017; Zhao et al., 2012; Fritz et al., 2017). The lack of land-use/cover differentiation generates confusion when forest cover change statistics derived from satellite imagery are compared directly against land-use statistics reported by governments in their national statistics (Curtis et al., 2018). In fact, they are used interchangeably in many land-cover/use maps. However, land-use describes how the land is used in a given location, such as forestry, residential, agricultural, industrial, forestry, and recreational, while land-cover refers to what is physically on the earth surface such as forests, grasslands, or water (Castilla and Hay, 2007). In other words, land-use can contain multiple land-covers (Herold et al., 2003; Fritz et al., 2017). Identifying land-use with remote sensing is challenging (IPCC, 2013) because we measure vegetation indices (VI) for land-cover and greenness and not directly use (Hansen et al., 2010b; Mas et al., 2017; van Asselen and

Verburg, 2012; Verbesselt et al., 2010b; Vogelmann et al., 2012; Nguyen et al., 2018). The assessment of land-use can be done via visual interpretation, but this is subjective and time-consuming.

In addition, land-use assessment approaches over regional or global scales are still problematic because they do not capture spatial heterogeneity (Curtis et al., 2018; FAO & JRC, 2012; Hansen et al., 2010b; Pandey et al., 2018; Pelletier et al., 2019). Here, the spatial heterogeneity refers to the variation in the patterns of land-use classes in spatial, spectral and spatial domains within and between regions. The heterogeneity of land-use may be observed as both the intra-regional variability (e.g., diverse patterns of same land-use, i.e. small-scale agricultural practice within the country) and the inter-regional variability (e.g., diverse patterns of same land-use i.e. small-scale agricultural practice, between the countries or regions) as observed in Fig. 1. In land-use classification, this heterogeneity of land-use patterns can result in the inability of the training data to represent the diversity of each class of the land-use. Not taking into account spatial heterogeneity introduces uncertainties and sources of error when evaluating the methods over different geographical locations, even if they present similar land-use. This results in an increase in uncertainty of the accuracy of land-use maps to estimate and measure the national forest emission reductions contributions for REDD+ (World Resources Institute, 2016). The critical question is whether integrating spatial-temporal information of land-use can achieve a reproducible large-scale land-use classification based on the diverse geographically sampled data as opposed to traditional approaches, where the sampled data are either relatively small (Olofsson et al., 2012), sampled in a non systematic way (Pengra et al., 2015; Fritz et al., 2017; Descals et al., 2021), sampled from a one time step (Fritz et al., 2017; Irvin et al., 2020), contain relatively few land-use classes (Descals et al., 2021; Irvin et al., 2020) or are sampled from small regions (Doggart et al., 2020; Irvin et al., 2020).

While recent advances in EOT, computing technology and deep learning methods hold promise for large-scale assessment of land-use following deforestation, or Follow-up Land-use (FLU), in the tropics, there is a need for new approaches that integrate the spatio-temporal information in dense satellite time series with deep learning methods. In this work, we therefore assess the potential of several deep learning methods for identifying FLU in the tropics using dense satellite time series. Specifically, the following two objectives are addressed:

1. We assess the performance of multiple deep learning approaches on a held out test dataset for classifying FLU after tropical deforestation using the spatial and temporal information from dense time series of satellite imagery.
2. We use the same procedure to assess the classification performance of continental models versus a pan-tropical model.

We concentrate on the deep neural network architectures that account for either (a) spatial-temporal information, i.e., Hybrid Recurrent convolutional neural network, 3D-convolutions, ConvLSTM, and the novel CNN + Multi Head Self-Attention model, or (b) only spatial information, i.e., 2D-convolutions, (c) only temporal information, i.e., Long short term memory (LSTM). Our main assumption is that the choice of model architecture is crucial to increase the classification performance on land-use following deforestation. Although some of these deep learning models with the exception of CNN + Multi-Head Self-Attention, are well-established techniques for remote sensing applications, to the best of our knowledge, we are the first to experiment and combine them for large-scale time series analysis of land-use in the remote sensing community (Ma et al., 2019; Reichstein et al., 2019; Yuan et al., 2020; Zhang et al., 2017; Rußwurm and Körner, 2020).

2. Method and materials

The study focuses in the tropical regions of Latin America, Africa, and Asia (Fig. 2). The heterogeneity of land-use following deforestation
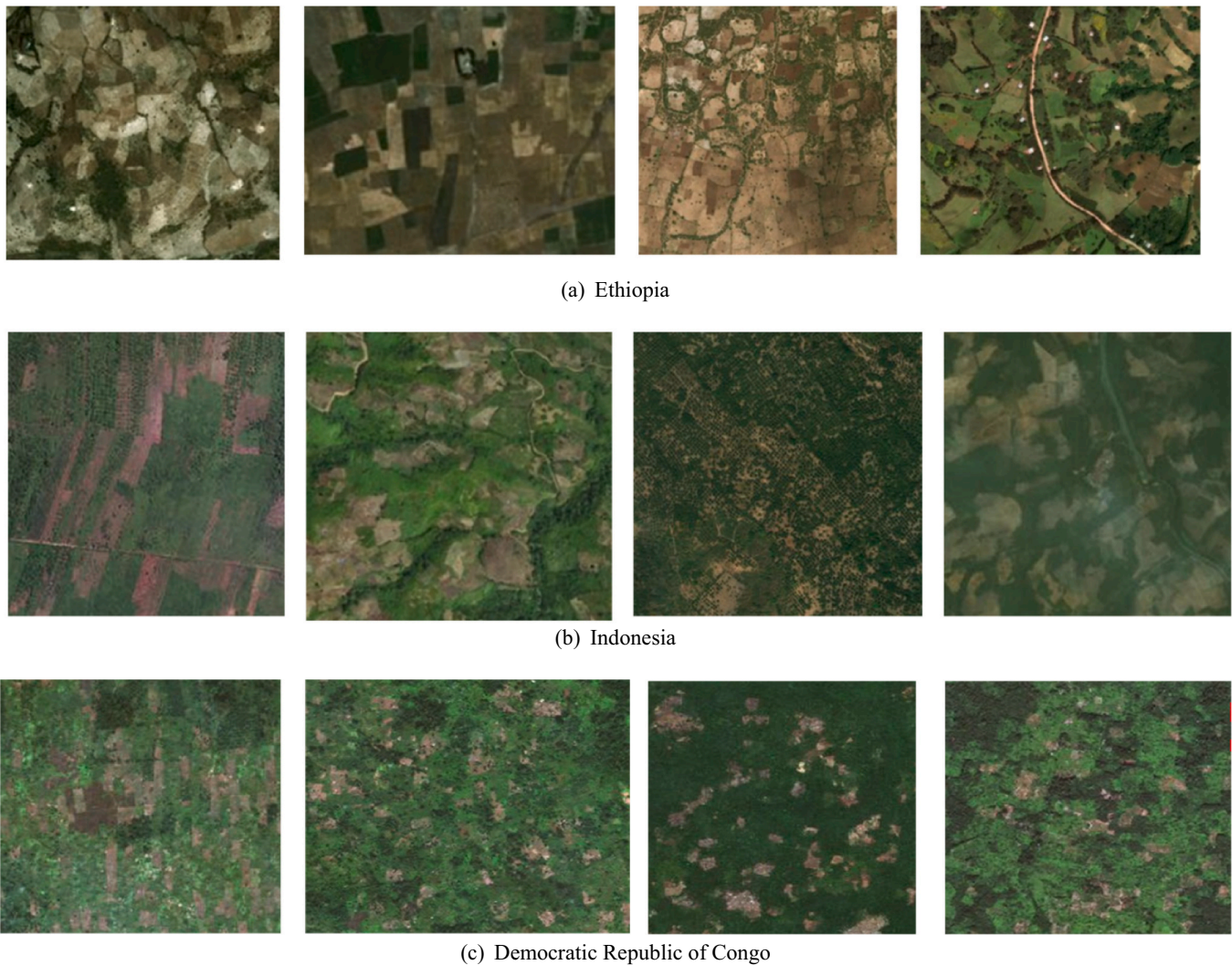
(a) Ethiopia

(b) Indonesia

(c) Democratic Republic of Congo

Fig. 1. Spatial heterogeneity of small-scale cropland land-use pattern within and between countries of (a) Ethiopia, (b) Indonesia and (c) the Democratic Republic of Congo - DRC, respectively. The heterogeneity is observed as both intra-variability (e.g., diverse patterns of small-scale agriculture within the country) and the inter-variability (e.g., diverse patterns of small scale agriculture between the countries or regions). Example imagery retrieved from Google Earth Engine.

or follow-up land-use (FLU) on these continents makes these regions suitable for methodological research on large-scale assessment of land-use following deforestation (Müller et al., 2015).

2.1. Data

In this study we used two data sources: 1) the ground truth/reference data, manually annotated using very high resolution imagery and provided by De Sy et al. (2019), containing the land-use following deforestation classes and 2) Landsat satellite time series data cubes over the same geographical extent.

2.1.1. Reference data

We used a large reference dataset containing land-use following deforestation from 1990 to 2005 in tropical regions of Latin America, Africa and Asia (De Sy et al., 2019). From the reference dataset, only the land-use data from 2000 to 2005 were used in this study to match with deforested areas from Hansen et al. (2013). The basis for the reference dataset is the FAO 2010 global Remote Sensing Survey (FAO FRA-2010 RSS) which used a systematic sampling design with sampling units spaced in a 10 km by 10 km grid on each degree latitude-longitude conFLUence point (FAO & JRC, 2012) to assess forest (change). This dataset consists of three main land-use classes, which are: forest, other wooded land, and other land (FAO & JRC, 2012). De Sy et al. (2015)

further classified land-use following deforestation into more detailed land-use classes (Table 1), through expert knowledge and visual interpretation of land-use using publicly available medium to high resolution satellite imagery. The definition of each of the FLU classes can be found in (De Sy et al., 2019). For this study we focus on 6 land-uses as the dominant drivers of deforestation, as indicated in (Table 1), i.e., large-scale and small-scale cropland, pasture, mining, tree-crops and other land with tree cover. The six selected land-use classes cover 95.0%, 97.2%, and 93.2% of the total reference dataset for Latin America, Africa and Asia respectively. In Fig. 2, we show the spatial distribution of reference/ground data used in this study. The reference data are based on the forest area loss per follow-up land-use in the pan-tropics for the six selected follow-up land-use classes.

2.1.2. Satellite data

For this study, we used Landsat 5 and 7 satellite imagery. This imagery has a spatial resolution of 30 m and a temporal resolution of 16 days. The Landsat data was chosen as it is the only satellite dataset with sufficient spatial and temporal coverage for the entire study period of 2000 to 2005. Two median composite images (January–June, July–December) were collected for each sample location from dense Landsat image time series for each year from 2000 to 2005 using Google Earth Engine. The collected dataset was filtered for clouds using the quality assessment band of Landsat SR data (Google Earth Engine,
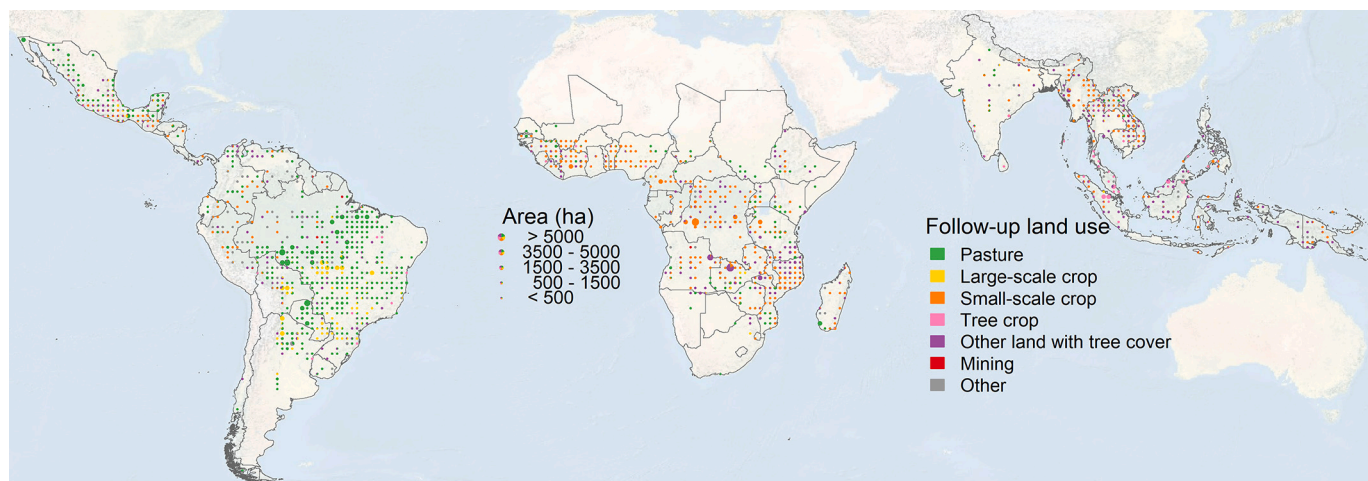
Fig. 2. Shows the spatial distribution of the reference data on land-use following deforestation across the pan-tropics for the six selected FLU classes. The size of each circle represents the forest loss area (in ha), and the colored pies within the circle represent the proportion of land-use following deforestation. The grey-lined countries are study area countries.

2020). Only images which contained cloud cover less than 50% were used to create a composite, this made sure we obtained enough images as possible but also we didn't lose more images. For each median composite in the time series, the normalized vegetation index (NDVI) and the normalized difference moisture index (NDMI), were computed. For each year, we collected at least 2 median composite images showing different phenological stages and texture, which contributes to differentiate FLU. Each composite image consisted of 7 spectral bands (Blue, Green, Red, Near-Infrared, Shortwave infrared 1, Thermal-infrared and shortwave infrared 2) and two vegetation indices (NDVI and NDMI), making a total of 9 bands for each Landsat based image composite in a time series collection. Overall every collected dataset from all sampled locations in the study area consist of a time series of 12 composites with 7 spectral bands (plus 2 vegetation indices) from 12-time steps from 2000 to 2005.

2.2. Deep learning models for FLU classification

Six neural network architectures with varying focus on spatial and/ or temporal features were tested to characterize FLU using dense time series of Landsat imagery:

a. A two-dimensional convolution neural network (2D CNN), which exploits the spatial spectral-correlations of neighboring pixels of each image in the time series data (Castelluccio et al., 2015; Huang et al., 2018).

2. A long short-term memory neural network (LSTM), a type of recurrent neural network (RNN) which focuses on the temporal dynamics of the time series data by explicitly controlling the flow of information through time (Wang et al., 2019; Zhong et al., 2019).

3. A three-dimensional convolutional neural network (3D CNN), which treats the time dimension as an additional spatial dimension and can thus exploit spatial-temporal-spectral correlations in the data (Kumar et al., 2019; Li et al., 2017; Xu et al., 2018).

4. A hybrid of 2D-CNN and LSTM, which aims to benefit from the exploitation of spatial information by using a 2D-CNN and controlling the temporal information flow with an LSTM (Hu et al., 2018; Yang et al., 2020; Zhu et al., 2020).

5. A convolutional long short-term memory neural network (ConvLSTM), a type of recurrent neural network with internal matrix multiplications replaced by convolution operations that can thus simultaneously exploit spatial-temporal correlations in the data (Shi et al., 2015).

6. A CNN + Multi-Head Self-Attention model (CNN-MHSA) or CNN-Transformer, an attention mechanism which focuses on certain parts of the input sequences of images to allow for more flexible

Table 1

land-use classes contained in the reference data. The land-use with * symbol denotes the land-use selected for our study.

Main classes	subclasses	Description
Agriculture	Mixed agriculture	● Mix of agricultural land-uses
	Small-scale cropland*	● Land under cultivation for crops, characterized by very small (<0.5 ha) to small field sizes (0.5–2 ha)
	Large-scale cropland*	● Land under cultivation for crops, characterized by medium (2–20 ha) to large (>20 ha) field sizes
	Tree-crops*	● Miscellaneous tree-crops (e.g. coffee, palm trees), orchards and groves
	Pasture (grazing land)*	● Land used predominantly for grazing; in either managed/cultivated (pastures) or natural (grazing land) setting; includes grazed woodlands
Infrastructure	Urban, settlements	● Residential areas
	Roads, built-up areas	● Transport, industrial and commercial infrastructures
Mining*		● Land used for extractive subsurface and surface mining activities (e.g. underground and strip mines, quarries and gravel pits), including all associated surface infrastructure
Water		● Natural (river, lake etc) or man-made water bodies (e.g. reservoirs)
Other	Bare land	● Exposed soil, sand, or rocks
	Other land with tree cover*	● Land not classified as forest, spanning more than 0.5 ha; with trees higher than 5 m and canopy cover of 5%–10%, or trees able to reach these thresholds in situ, or with a combined cover of shrubs, bushes and trees above 10%. It does not include land that is predominantly under agricultural or urban land-use.
	Grass and herbaceous	● Land covered with (natural) herbaceous vegetation or grasses
	Wetlands	● Areas of natural vegetation growing in shallow water or seasonally flooded environments. This category includes Marshes, swamps, and bogs.
Unknown land-use		● All land that cannot be classified (e.g. due to low resolution imagery)

interactions between the different time steps, applied in conjunction with a CNN feature extractor (Bazi et al., 2021; Rußwurm and Körner, 2020).

Implementation details for each model are discussed in the following section and summarized in Table. 2. The diagrams of individual model architectures are provided in appendix A.

2.3. Implementation details

The hyper-parameters used in these models were chosen based on the F1-score performance on a randomly chosen validation set consisting of 10% of the data. The dataset for each continent was split three times into training 75%, validation 10%, and 15% testing subsets for 3-fold cross-validation (Fig. 3). In order to select the model parameters (see below Table 2), we use Bayesian optimization using one of the folds (see details in Appendix B), in which the models have been trained on the training set and the best-performing parameters for each model type (2D-CNN, LSTM,CNN-LSTM, 3D-CNN, ConvLSTM, and CNN-MHSA) have been selected based on the accuracy achieved on the validation set. For each run, the final model was then evaluated on the held-out test set. The accuracies reported in this article are the averages and standard deviations on the test sets over the three folds. The final assignments of best parameters are reported in Table 2 (the same assignment was used for each of the three runs mentioned above).

All models were implemented by using the Keras library and TensorFlow as back-end. All network architectures were trained for 30 epochs with a batch size of 512. All convolutional layers were preceded by a padding operation that ensured that the spatial extent of the output stayed the same as the input and followed by a ReLU non-linearity (Table 2). Batch normalization was used to normalize the features in the second convolutional and the following two dense layers, while a dropout rate of 0.1 was used to regularize the second convolutional layer and third dense layer. All models were optimized by using Stochastic

Table 2

Architecture of 2D-CNN, LSTM, 3D-CNN, 2D-CNN-LSTM, ConvLSTM, and CNN-MHSA model.

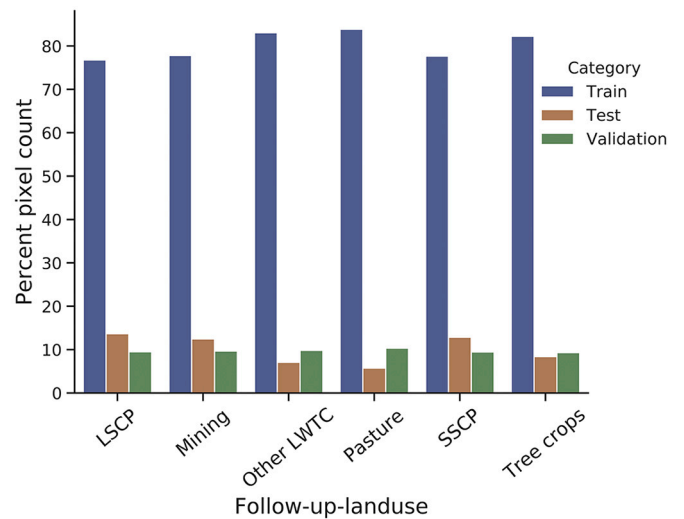| Parameters	Model Type and Dimensions					
	2D-CNN	LSTM	3D-CNN	2D-CNN-LSTM	ConvLSTM	CNN-MHSA
Input shape	$12 \times 9 \times 9 \times 9$	12×729	$12 \times 9 \times 9 \times 9$	$12 \times 9 \times 9 \times 9$	$12 \times 9 \times 9$	$12 \times 9 \times 9 \times 9$
No. Conv layers	6	–	6	6	6	6
No. filter	32, 32, 64, 64, 128, 128	–	32, 32, 64, 64, 128, 128	32, 32, 64, 64, 128, 128	32, 64, 128	32, 32, 64, 64, 128, 128
Filter size	3×3	–	$3 \times 3 \times 3$	3×3	3×3	3×3
Padding	same	–	same	same	same	same
Pool size	2×2	–	$2 \times 2 \times 2$	2×2	$1 \times 2 \times 2$	2×2
Strides	1×1	–	$1 \times 1 \times 1$	1×1	1×1	1×1
LSTM Units	–	32, 128	–	32, 128	–	–
No. Dense layer	2	2	2	2	2	2
Learning rate	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
Hidden dims	–	–	–	–	–	128
Hidden layer	–	–	–	–	–	32
No. of heads	–	–	–	–	–	4
No. Classes	–	–	–	–	–	6

Fig. 3. The percentage of pixels per FLU for the training, validation, and test sets of one of the random splits over the full tropical dataset. Where LSCP, other LWTC, and SSCP stands for large-scale cropland, Other land with tree cover, and small-scale cropland, respectively.

Gradient Descent (SGD) with a learning rate of 10^{-4}, learning rate decay of $3.3 \cdot 10^{-6}$ at each iteration, and momentum of 0.9. The optimized loss was a multi-class cross-entropy between the post-softmax scores and the one-hot label corresponding to the class of the central pixel of the patch. All models are designed to receive input tensors of shape (12 time steps $\times$ 9 width $\times$ 9 height $\times$ 9 bands) and to predict the FLU class of the central pixel of every data cube, such that the output is always a vector with the six scores, one per FLU class, followed by a softmax activation function.

2.3.1. 2D Convolutional Neural Network (2D-CNN)

We use a six-layer 2D-CNN classification architecture applied to each (9 width $\times$ 9 height $\times$ 9 bands) temporal slice of the input tensor, with shared weights. The six layers of the 2D-CNN consists of 32, 32, 64, 64, 128, 128 filters respectively, all with kernel size 3×3 and a stride of 1. A 2×2 max-pooling operation is applied on the second and fourth convolution layers. Each convolution operation is followed by a rectified linear unit (ReLU) activation function and batch normalization. These CNNs are wrapped inside a TimeDistributed layer in Keras that allows for sequential operations to every 9 width $\times$ 9 height $\times$ 9 bands image in a time series. The resulting feature maps are concatenated and then flattened into a vector of size 6144, followed by three linear layers with 1024, 512, and 6 neurons to provide FLU class predictions.

2.3.2. Long Short-Term Memory Recurrent Neural Network (LSTM)

Designed to control the flow of time-series information over arbitrarily long or short time intervals in sequential data, the LSTM model was used to leverage the temporal patterns in the time series input that are relevant to produce the final labels of FLU. The LSTM model was optimized by using two LSTM layers with 32, and 128 units respectively, to extract relevant non-linear temporal dependencies present in the remote sensing time series. The input shape for the LSTM layer was (12 time steps $\times$ 729 pixels) feature vectors flattened over 9 width $\times$ 9 height $\times$ 9 bands images. The LSTM layer was used jointly with three fully connected layers of size 1024, 512 and 6, batch normalization, and dropout of 0.1.

2.3.3. 3D Convolutional Neural Network (3D-CNN)

The 3D-CNN model treats the time series data as volumetric data by extracting features in three dimensions, one temporal and two spatial. We used an input image patches of size (12 time steps $\times$ 9 width $\times$ 9

height × 9 bands). The filter size of (3 × 3 × 3) was used to convolve in all spatial and temporal dimensions in all convolution layers. The resulting feature maps from the convolution layers are concatenated and then flattened into a vector of size 1536, followed by three linear layers with 1024, 512 and 6 neurons.

2.3.4. Hybrid CNN-LSTM

As an alternative to the 3D-CNN to account for spatial-temporal patterns in the imagery, we used an LSTM that resorts to convolutional filters in order to also account for spatial patterns. At each of the 12 time steps, a 2D-CNN is used to extract a feature vector of size 32, 64, 128 from the corresponding 9 width × 9 height × 9 bands image patch. A two-layer LSTM module then receives these vectors sequentially, keeping a state vector of size 12. This architecture enabled our model to extract spatial features associated with each FLU from CNN layers, then cascading the features into the LSTM layer, which extracted the non-linear temporal dependencies present in the remote sensing time series. Similarly to the 2D-CNN and LSTM methods, after the LSTM layer, a fully connected layer, dropout and batch normalization.

2.3.5. ConvLSTM

We use a ConvLSTM network to map FLU using spatial-temporal features from the input sequence of satellite images of size 12 time steps × 9 width × 9 height × 9 bands. The first part of the network includes three ConvLSTM layers followed by batch normalization and max pooling layers. The ConvLSTM layers include 32, 64, and 128 filters, respectively, with a kernel size of 3 × 3 used to perform convolutions. Finally, max-pooling with pooling window size of 2 × 2 is used to reduce the size of the output ConvLSTM features. The output of the last ConvLSTM layer is followed by three fully-connected layers with 128, 64, and 6 neurons, respectively.

2.3.6. CNN Multi-Head Self-Attention

Is another spatial-temporal model comprising a hybrid network of CNN and Multi Head self-attention block. The CNN block is used to determine spatial features and the self-attention block to determine temporal features from the CNN layer. The input data has a sequence of image of size 12 time steps × 9 width × 9 height × 9 bands. The first block of the network includes six convolutional layers of filter size 32, 32, 64, 64, 128, and 128 followed by ReLU activation function, and batch normalization. Max pooling layers of pool size 2 × 2 is applied on the second and fourth convolution. The kernel size of 3 × 3 used to perform convolution operation to determine spatial features respectively. Finally the convolution operation is followed by the global average pooling operation.

The second block of the network includes the transformer block which encodes features from the CNN block with 4 attention heads, 128 hidden dimension, and 32 hidden layer size in the feed forward inside the transformer. The transformer block is followed by global average pooling which takes the mean across all time steps and feed into 2 dense layers.

2.4. Evaluation of performance deep learning models

The performamce of the deep learning models in classifying the FLU was evaluated using a held out test set and following a 3-fold strategy. The standard error of the F1-score on test set was computed as $\sigma_{\bar{x}} = \sigma/\sqrt{n}$, where $\sigma\bar{x}$ is the standard error, σ standard deviation and $n = 3$ is the number of test F1-scores. The evaluation metrics computed were the classwise F1-scores, micro-average of F1-scores and macro-average of F1-scores. The F1-score is the harmonic mean of precision and recall, $F1 = 2(P * R)/(P + R)$ (Rußwurm and Körner, 2017). The precision, $P = TP/(TP + FP)$, is a measure of result relevancy, while recall, $R = TP/(TP + FN)$, is a measure of how many relevant results are returned.. TP, FP, and FN stands for the number of true positives, false positives, and false negatives of each of the predicted FLU classes. In addition, the confusion

matrix of the predicted samples and true samples were computed using the unseen test set. We used the F1-score of each class as an indicator of the model's capability to identify each single class of FLU. At the same time, the micro-average and macro-average of F1-score was used as an indicator of the general classification capability of the model (Zhang et al., 2019; Pan et al., 2017). The macro-average of the F1-score is the average of all classwise F1-scores and covers class imbalance of each class, giving more importance to rare classes (Johnson and Khoshgoftaar, 2019; Pan et al., 2017). The micro-average F1-score computes the aggregated contribution of all classes by using precision and recall values averaged across all samples. This puts emphasis on the more abundant classes in the data, since it gives each sample the same importance (Pan et al., 2017).

3. Results

In this chapter, we present the classification results using spatial, temporal, and spatio-temporal models for identifying the FLU in the pan-tropics (namely 2D-CNN, LSTM, 3D-CNN,Hybrid CNN-LSTM, ConvLSTM, and CNN-MHSA). In the following sections we present the FLU classification results for the continental models (Section 3.1) and for the pan-tropical models (Section 3.2). Finally, in Section 3.3 we present a spatial comparison of predicted versus reference dominant FLU per sample location using the best continental models and pan-tropical model from section 3.1, and 3.2.

3.1. Classification results – Continental models

The results in Fig. 4 show the per-class F1-scores, micro-averaged F1-score and macro-averaged F1-score for the six deep learning methods, both for the continental and pan-tropical models. All methods obtain comparable levels of accuracy in the contenental setting except for the LSTM, which tends to trail the other methods by a large margin. The 2D-CNN performs almost as well as the hybrid spatio-temporal methods (Fig. 4).

In Fig. 5(a), 5(b), 5(c) and 5(d) we show the confusion matrix for the FLU's classification of one of the best performing model (Hybrid CNN-LSTM) over Latin America, Africa, Asia and for full pan-tropical test datasets, respectively. For Latin America, there is high recall rate of over 78% for large-scale cropland, 90% pasture, and 91% mining. Other classes such as small-scale cropland, tree-crops and other land with tree cover attained low recall rate of over 45%, 24%, 26% respectively. Most of the small-scale cropland and other land with tree cover tends to be confused with the pasture while tree-crops tend to be confused with other land with tree cover as in Fig. 5(a). In Africa (Fig. 5(b)), pasture and tree-crops tend to be confused with small-scale cropland (55%, 38%). In Asia (Fig. 5(c)), almost every FLU class has a recall rate higher than 80% with the exception of tree-crops (54%) which tends to be confused with SSCP (26%) and other land with tree cover (19%).

3.2. Classification result – pan-tropical model

Unlike in the continental setting, the Hybrid CNN-LSTM, ConvLSTM and CNN-MHSA outperform all the other methods in the pan-tropical setting by a substantial margin, as shown in Fig. 4. The pan-tropical Hybrid CNN-LSTM, and CNN-MHSA model reached a micro-average F1-score of (58%, 66%) and a macro-average F1-score of (53%, 61%). Among all the follow-up land-use types, larger-scale cropland (42%), mining (96%), small-scale cropland (95%) and other land with tree cover (75%) are the most distinguishable FLUs (Fig. 5(d)). Larger-scale cropland, pasture, tree-crops and other land with tree cover are some likely to be confused with small-scale cropland by (30%, 62%, 87%, 24%), respectively. Larger-scale cropland is also confused with pasture (22%), while pasture is confused with other land with tree cover (11%). The pan-tropical model's confusions are similar to the continental models confusion, where tree-crops in Latin America and Asia is the
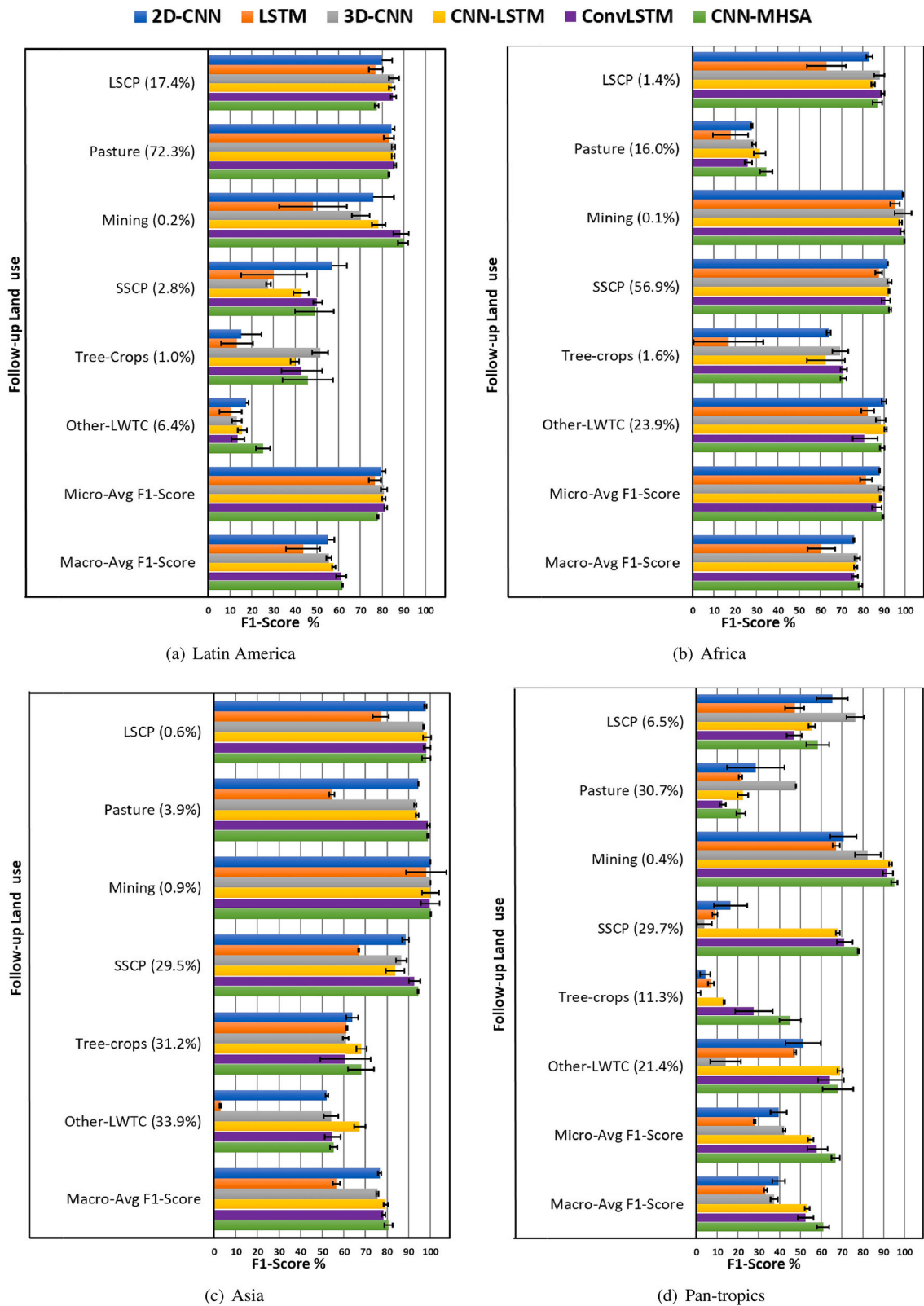
Fig. 4. F1-scores of FLU classification (in percentages) for deep learning methods for continental and pan-tropical models for (a) Latin America, (b) Africa, (c) Asia, and (d) pan-tropics, respectively. LSCP stands for large-scale cropland, SSCP for small-scale cropland, and Other LWTC for other land with tree cover. In brackets are the percentages of each FLU class present on each continent. The error bars are the standard deviation on F1-scores.
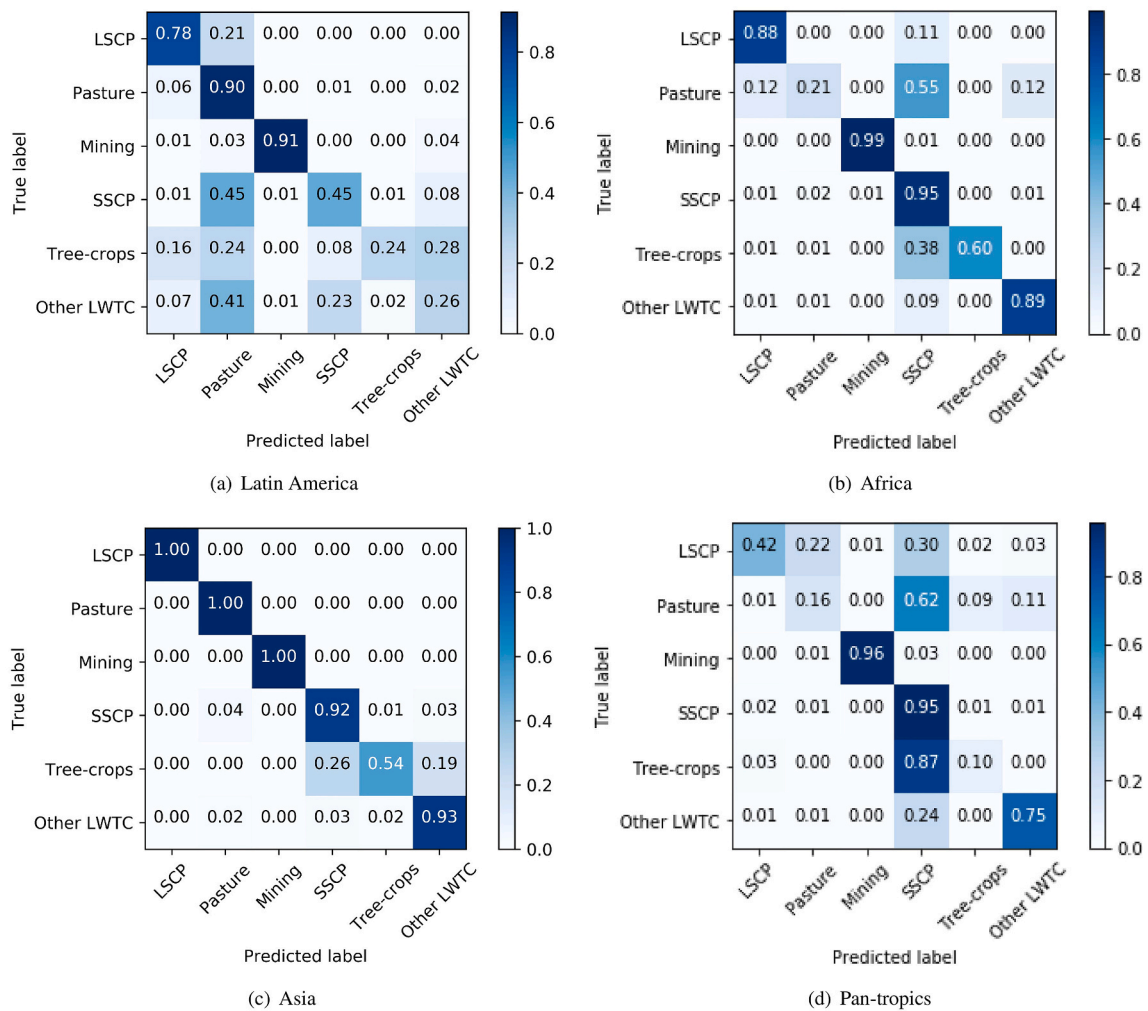
Fig. 5. The confusion matrix showing the percentage of correct predicted and false predicted FLU's using one of the spatio-temporal deep learning method with highest score (CNN-LSTM) for (a) Latin America, (b) Africa, (c) Asia, and (d) pan-tropics, respectively. LSCP stands for large-scale cropland, SSCP for small-scale cropland, and Other LWTC for other land with tree cover.

most misclassified follow-up land-use class, reaching the F1-score of (10%) (Fig. 5(d)). Similarly, the tree-crops is the follow-up land-use class with the lowest recall rate in both continental and pan-tropical models. This can be attributed to the spatial scale and spectral similarity of tree-crops with small-scale cropland and large-scale cropland across the tropics.

3.3. Comparison of dominant FLU per sample location using the hybrid CNN-LSTM method

Using one of the best performing spatio-temporal methods in the pan-tropical setting (Hybrid CNN-LSTM), we compare the most dominant FLU per sample location for the predicted and reference FLU dataset, for both the continental (Figs. 6(a) and pan-tropical models (Figs. 6(b). In Latin America, for the continental models' predictions (Fig. 6a), pasture and large-scale cropland are the most dominant FLUs in both prediction and reference datasets. For sample locations in Brazil, Paraguay, Bolivia, and Uruguay these two dominant FLUs are mostly similar for predicted and reference FLU. However, we also observe that large-scale cropland tends to be confused with pasture, particularly in Argentina, Paraguay, northern Mexico, and eastern Brazil (mainly in the state of Bahia). Other confusions are also observed in Mexico between other land with tree cover and small-scale cropland as well as other land with tree cover and pasture (Figs. 6(a).

In Africa, small-scale cropland is the most dominant driver in both prediction versus reference dataset (Fig. 6(a). For small-scale cropland, predicted dominant FLU matches well in parts of central Africa (i.e., Democratic Republic of Congo -DRC, Congo), southern Zimbabwe and Mozambique, Ethiopia, and parts of West Africa (i.e., Nigeria, Cameroon, and Cote d'Ivoire). Nevertheless, we also observe confusion between other land with tree cover and small-scale cropland in central Mozambique, northern Zambia, and Angola. Another notable confusion is between pasture and other land with tree cover as well as small-scale cropland and pasture in Tanzania, Kenya, and some parts of West Africa.

In Asia, tree-crops and small-scale cropland tend to be the most dominant driver in both prediction versus reference dataset (Figs. 6(a). Tree-crops are predicted well as dominant FLU in Malaysia and Indonesia. Small-scale cropland are predicted well as dominant FLU in central parts of Myanmar, some parts of Thailand, Laos, and Vietnam. On the other hand, in parts of Thailand, Cambodia, India, and Indonesia there is mismatch between predicted and reference dominant FLU as small-scale cropland, pasture, and tree-crops tend to be confused with other land with tree cover.

For the pan-tropical model similar spatial patterns of the reference versus predicted dominant FLU could be observed as in the continental models' prediction results (Figs. 6(a) for Latin America and Asia. In Africa, larger-scale cropland was over-predicted as dominant FLU at the expense of small-scale cropland, especially in parts of west and east Africa.
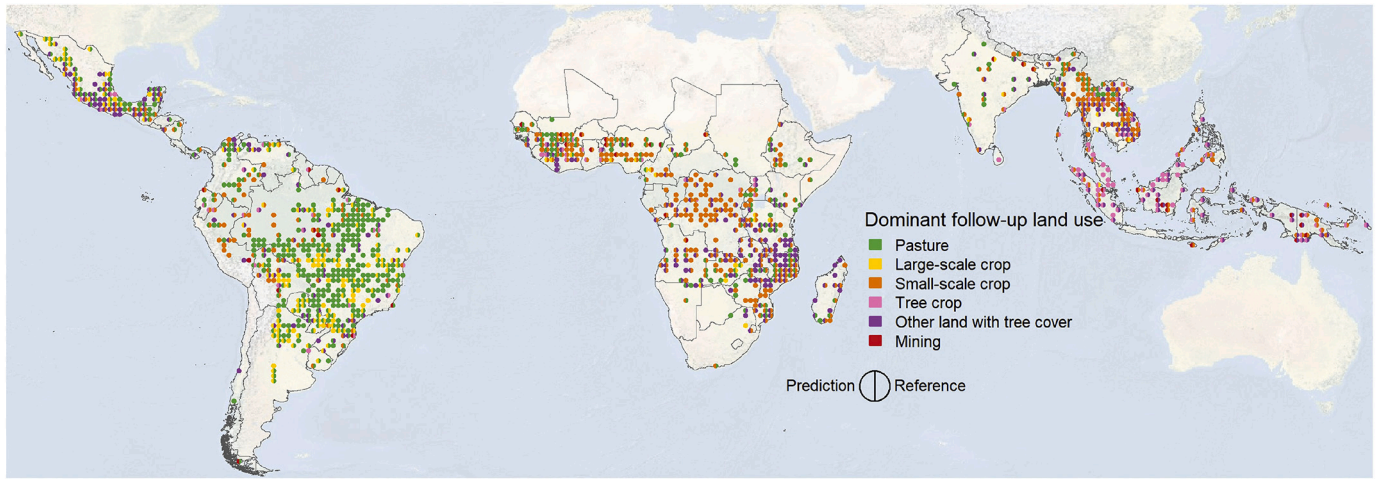
(a) Continental FLU predictions
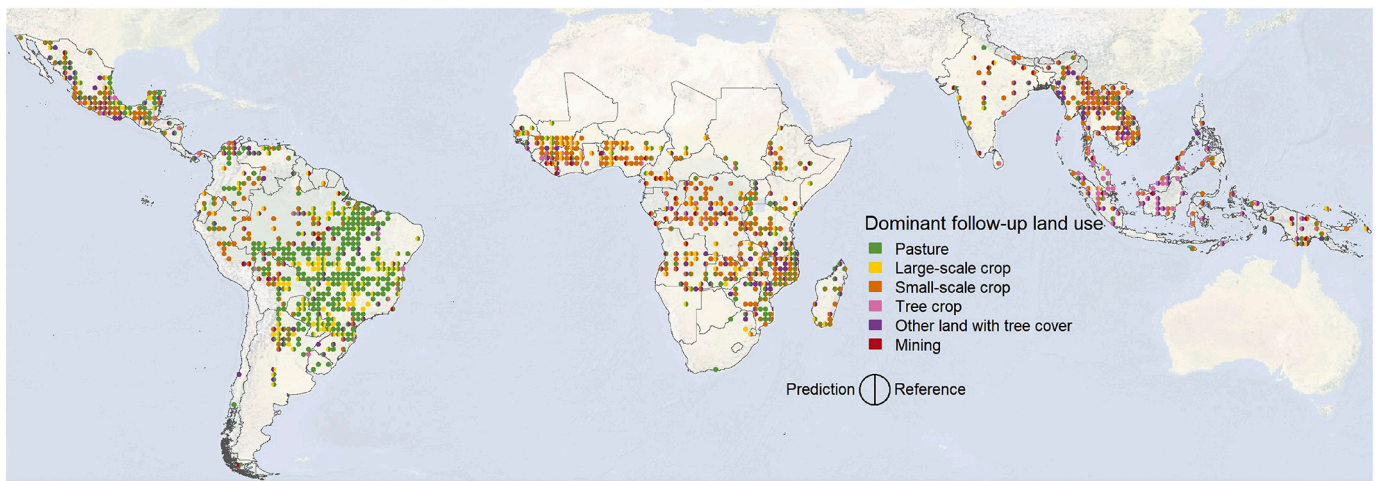
(b) Pan-tropical FLU predictions

Fig. 6. Dominant follow-up land-use for each sample location for prediction and reference FLU dataset for (a) continental models (CNN-LSTM), and (b) pan-tropical model (CNN-LSTM). Each circle represents the dominant FLU for a sample location, with the left half of the circle showing the predicted dominant FLU and the right half showing the reference dominant FLU. Best viewed in the digital version and by zooming in.

3.4. Computational considerations

In terms of computation time, the optimal architectures for LSTM and ConvLSTM resulted in more computationally intensive models at training time compared to the other four models (Table 3). This can be due to the sequential feed forward and back propagation between hidden vectors of each timesteps, since the subsequent steps in LSTM and ConvLSTM cells depends on previous ones resulting in more time and computational resource demand. This is different from other models such as 2D-CNN, 3D-CNN and CNN-MHSA where the feature computations happen in parallel. In the CNN-LSTM the CNN part is also computed in parallel, providing a better image representation than the LSTM without a CNN feature extractor, resulting in a more efficient architecture. There were only marginal differences in testing time between the models. This is useful, as, in the future, we can re-use the saved deep learning models when making predictions of FLU on recent deforested locations. Thus, there will be no need to re-train the model from scratch (Castelluccio et al., 2015). All experiments were completed in a secured cloud-based computing environment (SEPAL 2.0) that is part of the Amazon cloud with instance type g8, NVIDIA Tesla M60 GPU 32GB RAM.

Table 3

The training time and test time in minutes (m) for Latin America, Africa, and Asia datasets using 2D-CNN, LSTM, 3D-CNN, ConvLSTM, 2D-CNN-LSTM and 2D-CNN-Multi-Head Self-Attention methods.

*Study Regions	2D-CNN		LSTM		3D-CNN		ConvLSTM		CNN-LSTM		CNN-MHSA	
	Train(m)	Test(m)	Train(m)	Test(m)	Train(m)	Test(m)	Train(m)	Test(m)	Train(m)	Test(m)	Train(m)	Test(m)
Latin America	113.13	3.26	212.58	1.45	115.39	4.55	433.23	24.50	112.30	3.37	108.43	3.29
Africa	62.5	3.52	149.72	1.56	66.5	5.32	346.5	26.51	66.17	4.5	152.5	3.44
Asia	67.5	1.18	212.5	0.84	84.5	1.52	343.33	9.26	86.1	1.23	175.55	1.16

4. Discussion

The scope of this work was designed to evaluate the use of different spatio-temporal deep learning models for classifying land-use following deforestation at a large scale in continental Latin America, Africa, and Asia. Large-scale experiments were conducted on a real-world dataset acquired from freely available time series of Landsat satellite data (2000–2005) together with reference data provided by FAO 2010 global Remote Sensing Survey (?). The results from our experiments (in Sections. 3.1, 3.2) have shown that simultaneously accounting for spatial and temporal patterns in the model design decisions results in better performances compared to models which by design can solely rely only on spectral, spatial (2D-CNN) or temporal information (LSTM) (Zhang et al., 2019). The higher accuracies attained by the spatio-temporal methods confirms that 2D-CNN (spatial), LSTM (temporal), methods alone are less able to differentiate the FLUs. We also observed that the 2D-CNN achieved higher classification performance on the continental scale compared to the LSTM, as seen in Fig. 4(a), 4(b) and 4(c). This shows that spatial patterns of land-use are more useful for land-use characterization than temporal ones when considering the problem at the regional level. This may indicate a higher level of heterogeneity in terms of the temporal patterns that characterize each FLU, possibly due to regional differences in seasonality and land-use practices. The performance of all models in identifying tree-crops and other land with tree cover in Latin America and Asia, as well as pasture in Africa, is relatively low compared to other FLU classes, also suggesting a more heterogeneous spatio-temporal pattern in these cases. On the other hand, pasture achieves above average performance in Latin America and Asia, even though in the latter it is a minority class, suggesting that more homogeneous spatial patterns are associated with pasture in these continents than in Africa.

Looking at predicted versus reference dominant follow-up land-use (Figs. 6(a) and 6(b)), some spatial patterns could be observed. In Latin America, we do not see as much confusion in the so-called arc of deforestation (edge of Amazon forest) compared to the region more to the south, where the dominant FLUs (pasture and large-scale cropland) are more established. This might be related to the fact that, over time, the extensively managed large-scale cropland tends to be used as pastures during one or two rotational cycles, thus exhibiting a similar texture and spatial arrangement compared to newly established pasture/large-scale cropland in the arc of deforestation (Müller et al., 2015). Another spatial pattern of confusion is observed in eastern and mid-southern Africa, where most small-scale cropland and pasture tend to be confused with other land with tree cover. The confusion may arise from the fact that both pasture, small-scale cropland, and other land with tree cover are characterized by savanna and shrubland vegetation types and have no clear distinct spatial and temporal patterns to make them distinguishable, making it even challenging with visual identification (Müller et al., 2015). Another reason for this confusion can be attributed to the confidence in reference data, i.e., some samples of reference data used in our study from visual interpretation had low confidence level of belonging to the referenced class, which might have contributed to low performance of the model (De Sy et al., 2019) in these areas.

Overall, the Hybrid CNN-LSTM, ConvLSTM and CNN-MHSA models achieved better classification accuracies compared to the 2D-CNN, LSTM and 3D-CNN in the pan-tropical setting. This is in spite of the cloud cover affecting parts of the data. While other studies focus on filtering out cloudy data (Zhang et al., 2019) or on exploiting partly cloudy images (Oehmcke et al., 2019), our results suggest that models making explicit use of spatio-temporal patterns, in particular those using attention-based mechanisms such as CNN-MHSA, are able to cope with these high levels of cloud coverage. The observation that explicitly accounting for both spatial and temporal patterns in the data boosts the generalization performance is in line with the results of Zhong et al. (2019), who use a multi-temporal deep neural network for a classification task, and the Joint Deep Learning (JDL) model of Zhang et al.

(2019), which incorporates a multilayer perceptron (MLP) and convolutional neural network (CNN) for land cover and land-use classification. The approaches in the Zhang et al. (2019) are methodologically closest to the Hybrid CNN-LSTM and CNN-MHSA models used in this study and are comparable with our setting in terms of the type of land-use data used. However, their relatively small study area (Southampton and Manchester), together with their image data type (aerial photos), hinders a direct comparison. Indeed, one of the main contributions of this work is going beyond small-scale benchmarking tasks (Castelluccio et al., 2015; Li et al., 2018; Zhang et al., 2019; Doggart et al., 2020; Irvin et al., 2020). Also (Doggart et al., 2020; Irvin et al., 2020; Descals et al., 2021) achieved relatively higher accuracies compared to our study, although their studies were country based or covered fewer FLU classes. Overall, in all three study areas (Latin America, Africa, and Asia), the hybrid CNN-LSTM, ConvLSTM and CNN-MHSA models yielded the best results, thus making the use of these models for identifying the FLU in the pan-tropical more promising by reducing the work of using human interpreters in identifying land-use over large areas.

It is also important to note that the deep learning approaches in this study were tested and evaluated on data limited to the sample locations and dates in De Sy et al. (2019). Nevertheless, the deep learning model evaluated in this study could be applied to monitor land-use following deforestation wall-to-wall and for more recent time periods. Given the current and expected future availability of high resolution and accurate global deforestation data, these methods could be applied following open-source global deforestation data such as the Hansen et al. (2013) global forest loss and RADD alerts (Reiche et al., 2021) or to national deforestation data in the context of national forest monitoring systems. These provide an opportunity to implement the proposed methods to identify the land-use following deforestation in deforested areas while masking out non-deforestation areas using the updated forest loss layers. Relying on existing deforestation data makes the task of detecting land-use following deforestation at large-scale more tractable without compromising its applicability.

In addition, this study further validates that large-scale land-use classification tasks can be accomplished without the necessity of using spatio-temporal handcrafted features (expert knowledge) to account for spatio-temporal dependencies of land-use as in traditional machine learning models, i.e., random forests or support vector machines (Ma et al., 2019; Reichstein et al., 2019; Zhang et al., 2019; Zhu et al., 2017). Better classification results of deep learning versus traditional machine learning in terms of classification accuracy have also been obtained in (Huang et al., 2018; Zhang et al., 2019). Besides its success, technical challenges remain, such as (1) limited model interpretability due to lack of direct causal relationship between inputs and outputs, i.e., in identifying important and most useful variable for the classification problem (2) high computational requirements due to an increase in data volume as a result of an increase in the number of features and dimensions extracted by the models during the computational stage, resulting to higher training time and RAM requirement or usage, (3) high demand of reference data, caused by intraclass variability (heterogeneity) (Reichstein et al., 2019; Zhu et al., 2017). According to Reichstein et al. (2019), the possible challenge of deep learning models when working with heterogeneous data is that they may work well during training and on test datasets but perform poorer when extrapolating to other regions outside their valid domain. Despite these challenges, the deep learning field is still growing and has achieved better classification accuracies when compared to other traditional models (Zhang et al., 2019; Zhao and Du, 2016). Swift advancement are expected in the near future in the field of remote sensing, especially through the use of transfer learning and self-supervised learning (Zhu et al., 2017) for large-scale classification problems. For future work, we plan to expand our analysis to predict, validate, and monitor changes in land-use at the continental level while leveraging the pan-tropical data through meta-learning (Tseng et al., 2021; Rußwurm et al., 2020).

5. Conclusion

We have explored the potential of using spatio-temporal deep learning models for large-scale classification of land-use following deforestation using time series of satellite (Landsat) images. Six deep learning models were considered in this study, the 2D-CNN, LSTM, 3D-CNN, Hybrid CNN-LSTM, ConvLSTM, and CNN-MHSA. The models were evaluated in the tropical regions of Latin America, Africa, and Asia as continental models and pan-tropical model. We found that, for most land-use classes, the complementary spatio-temporal information extracted by the spatial-temporal models (Hybrid CNN-LSTM, 3D-CNN, ConvLSTM, CNN-MHSA) from the time series of Landsat images improved the accuracy of the model in classifying the FLU by a significant margin compared to the 2D-CNN and LSTM, which are designed to focus only on spatial or temporal features, respectively.

Nevertheless, the FLU classification from Landsat satellite imagery remains a challenging task due to the existence of spatial heterogeneity and spatio-temporal variability of the FLU over large-scale assessments. In some areas, the classification challenges were particularly caused by clouds and shadows that cover most tropical areas of Africa and Asia. Yet, the evaluated spatio-temporal models, were able to distinguish the FLU in most of these areas, particularly when trained on regional data.

We also found that models focusing on spatial patterns only performed competitively in the continental setting but not in the pan-tropical one, suggesting that land-use types within a region are more readily characterized by their spatial patterns than their temporal signature. This effect disappears at the pan-tropical level, pointing at different spatial patterns across the tropics. Indeed, this is the setting were it was most advantageous to use spatio-temporal models and, particularly, the attention-based CNN-MHSA. We hypothesize that the arbitrary temporal interactions allowed by attention-based models provide an edge against the noise derived from the high cloud cover characteristic of the tropics.

Therefore, given the size and coverage of this study, this work could be particularly useful for large-scale forest and land-use change monitoring in the context of REDD+, the global stocktake for the Paris Agreement and the Sustainable Development Goals. In this paper we address challenges associated with large-scale FLU assessments in the pan-tropics such as the heterogeneity of land-uses, and identifying land-use instead of just land cover. Our methodology can support a more detailed spatial and temporal assessment of where forests are lost, and the land-use activities driving it. This will allow targeting of REDD+ mitigation efforts toward specific proximate deforestation drivers in order to achieve more impact. Our approach could also be adopted for national forest monitoring systems as it uses open-source data and platforms, and can be calibrated with local or national data. Our method could be further developed toward more frequent and wall-to-wall monitoring of land-use following deforestation to identify hotspots and local patterns of land-use change. Other data source such as recent forest loss data and other satellite sources (e.g. high resolution imagery, radar data) could be included.

Declaration of Competing Interest

None.

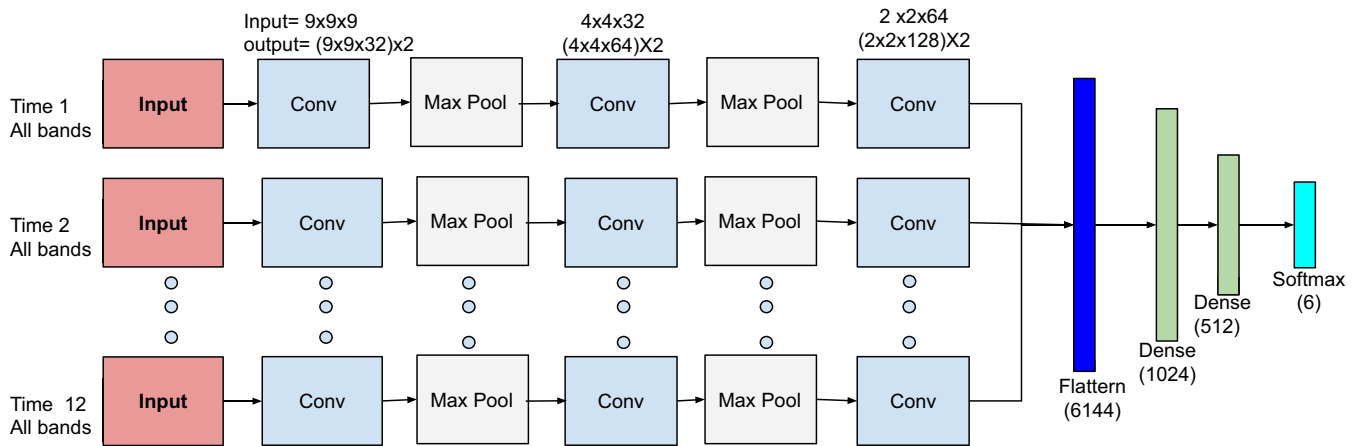## Appendix A. Diagrams of individual deep learning model architectures

Fig. 7. Schematic view of our convolution network. Time 1 to Time 12 are the tensors input along the temporal dimension each of size $9 \times 9 \times 9$. The colors represents, light red = Input images, light blue = Convolution operations, light grey = MaxPooling operation, blue = flatten operation, light green = feed forward network and cyan = output layer. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
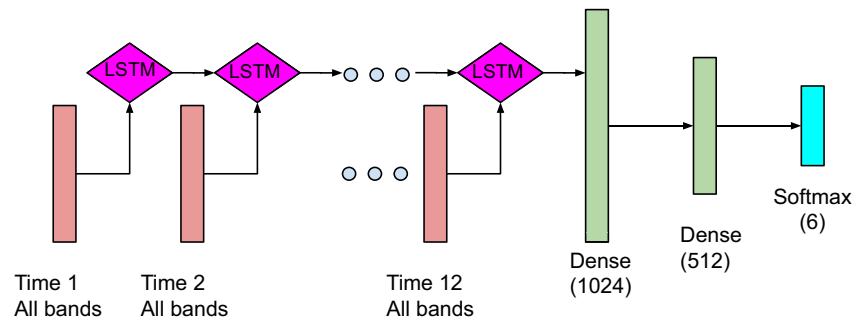
Fig. 8. Schematic view of our long-short term memory network.The colors represents, light red = Input array, magenta = LSTM cells, light green = feed forward network and cyan = output layer. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
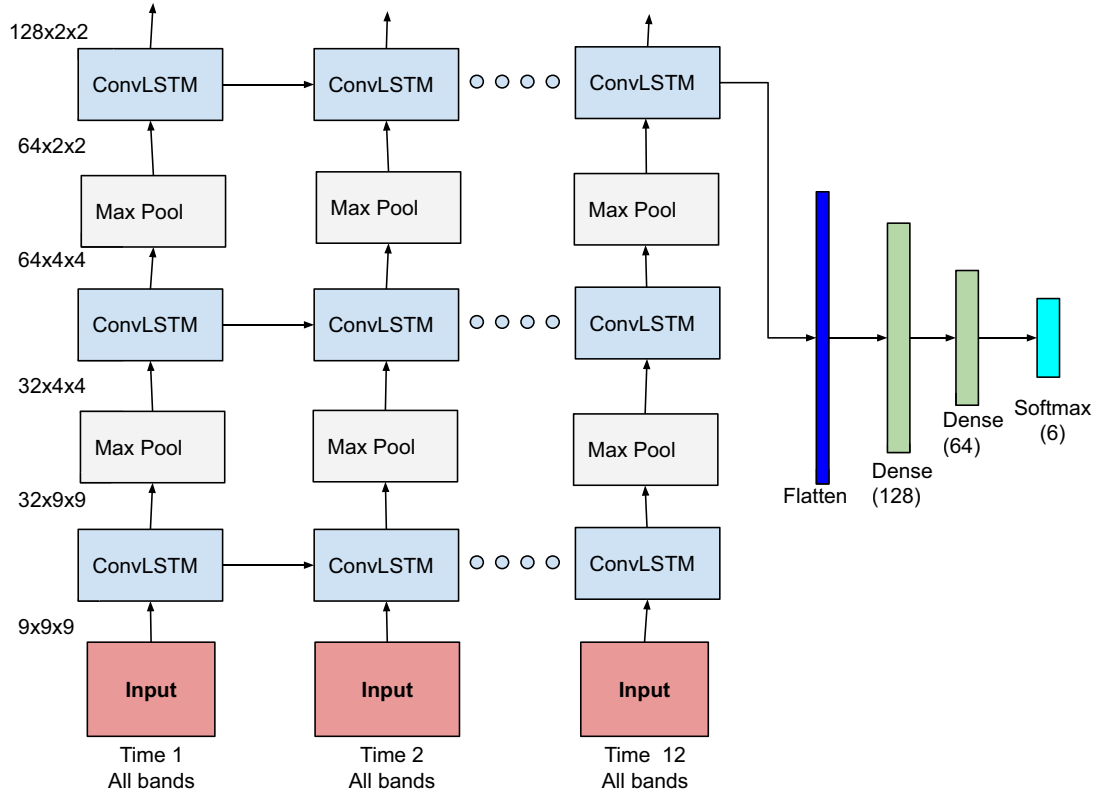

Fig. 9. Schematic view of our ConvLSTM network. The colors represents, light red = Input images, light blue = Convolution operations, light grey = MaxPooling operation, blue = flatten operation, light green = feed forward network and cyan = output layer. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
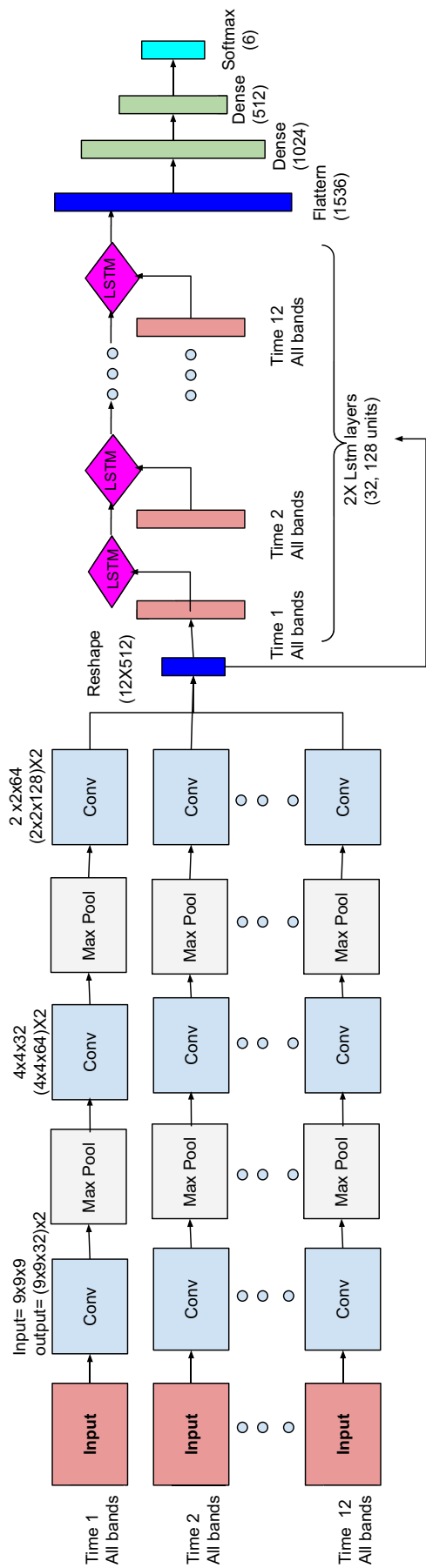
Fig. 10. Schematic view of our convolution+long-short term-memory network. Time 1 to Time 12 are the tensors input along the temporal dimension each of size $9 \times 9 \times 9$. The colors represents, light red = Input arrays, light blue = Convolution operations, light grey = MaxPooling operation, magenta = LITM cells, blue = reshape operation, light green = feed forward network and cyan = output layer. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
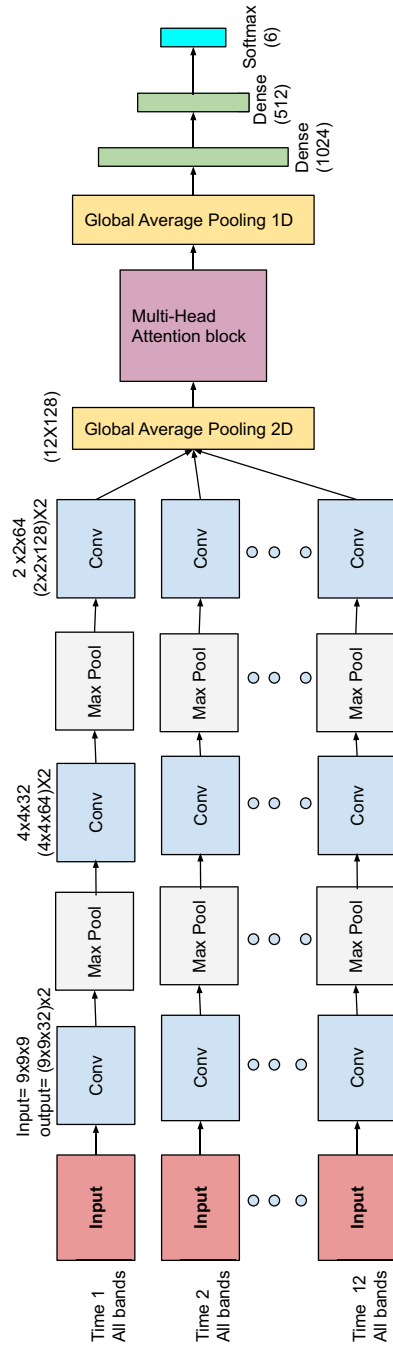

Fig. 11. Schematic view of our convolution+Multi-head Self-attention model. Time 1 to Time 12 are the tensors input along the temporal dimension each of size $9 \times 9 \times 9$. The colors represents, light red = Input images, light blue = Convolution operations, light grey = MaxPooling operation, light yellow = Average pooling operation, light green = Attention layer, light magenta = Attention layer, light magenta = Attention block and cyan = output layer. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Fig. 12. Schematic view of our 3D convolution network. The input tensors along the spatial and temporal dimension are of size $12 \times 9 \times 9$. The colors represents, light red = Input images, light blue = Convolution operations, light grey = MaxPooling operation, blue = flatten operation, light green = feed forward network and cyan = output layer. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
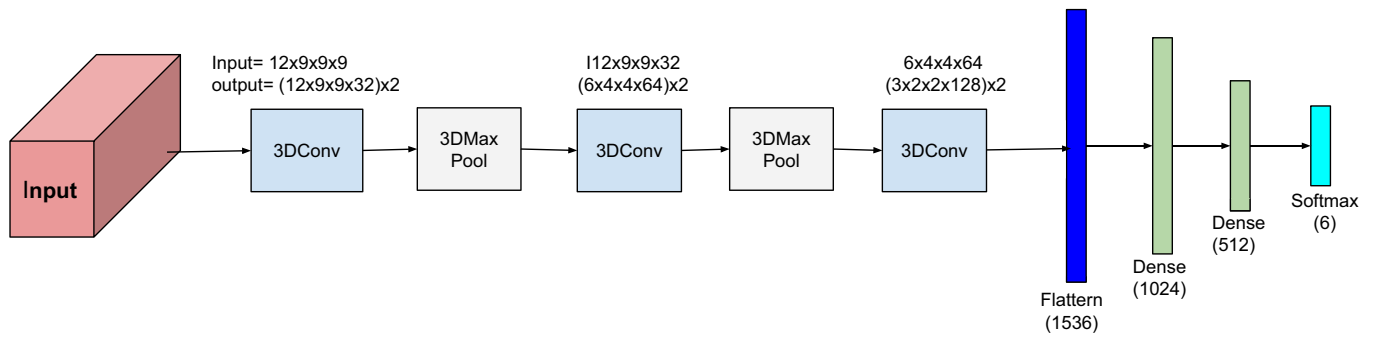
Appendix B. Hyper-parameter optimization using Bayesian optimization

We used Bayesian optimization from the scikit-optimize package (Gilles and Manoj, 2021) to find the best possible set of hyper-parameters that gives us high performance of the model on validation data. A total of 20 calls or runs was used to optimize the hyper-parameters Fig. 13. For most of the involved model hyper-parameters, a search space of parameters ranging from learning rate ($\{2^{-6}, ..., 2^{-2}\}$), number of filters ($\{16, ..., 128\}$), number of dense layers ($\{1, ..., 3\}$), number of dense nodes ($\{64, ..., 1024\}$), number of LSTM units ($\{32, ..., 128\}$), number of heads ($\{3, ..., 8\}$) have been used. However, as expected, the learning rate had significant impact on the validation accuracies compared to other hyper-parameters, refer to Fig. 14. In Fig. 15 we also show the sample distributions for each of the hyper-parameters during the Bayesian optimization and the order in which the samples were taken.
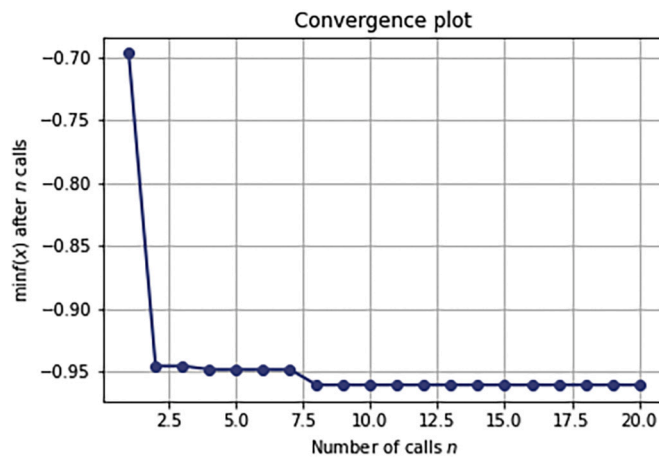

Fig. 13. The view of the progress of the hyper-parameter optimization on the validation data over 20 calls. The X-axis shows the number of calls and y-axis the minimization, or the convergence trace over n calls.
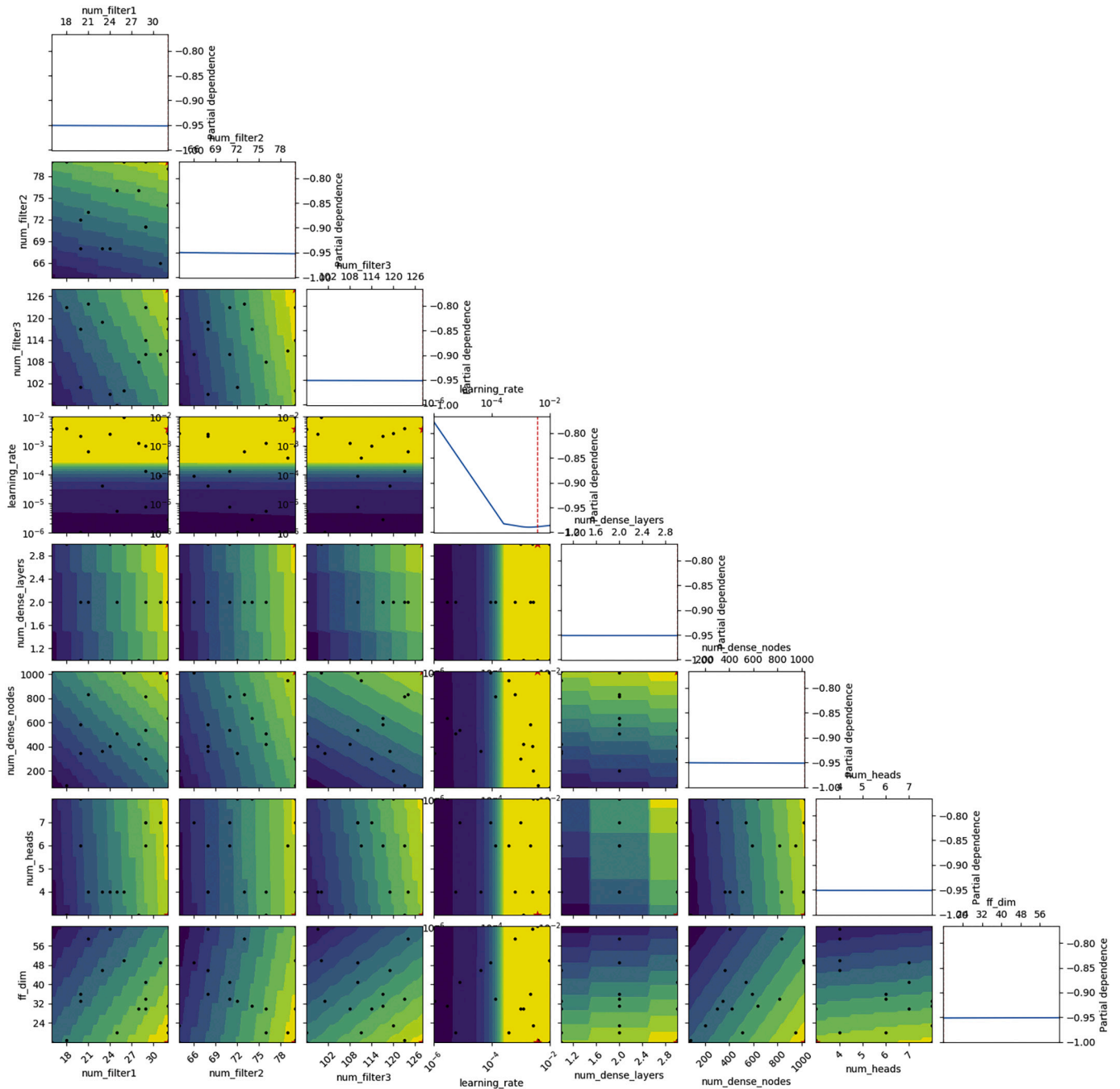
Fig. 14. The Partial Dependence plot showing a matrix-plot of all combinations of searched hyper-parameters during Bayesian optimization. This shows how the approximated fitness value changes when we are varying two dimensions simultaneously. The top diagonal subplots shows the influence of a single dimension on the accuracy. For each sub-plot the blue and yellow regions shows areas that gives us low and high performance hyper-parameters. The black dots show where the optimizer has sampled the hyper-parameter and the red dot shows the best hyper-parameters found. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
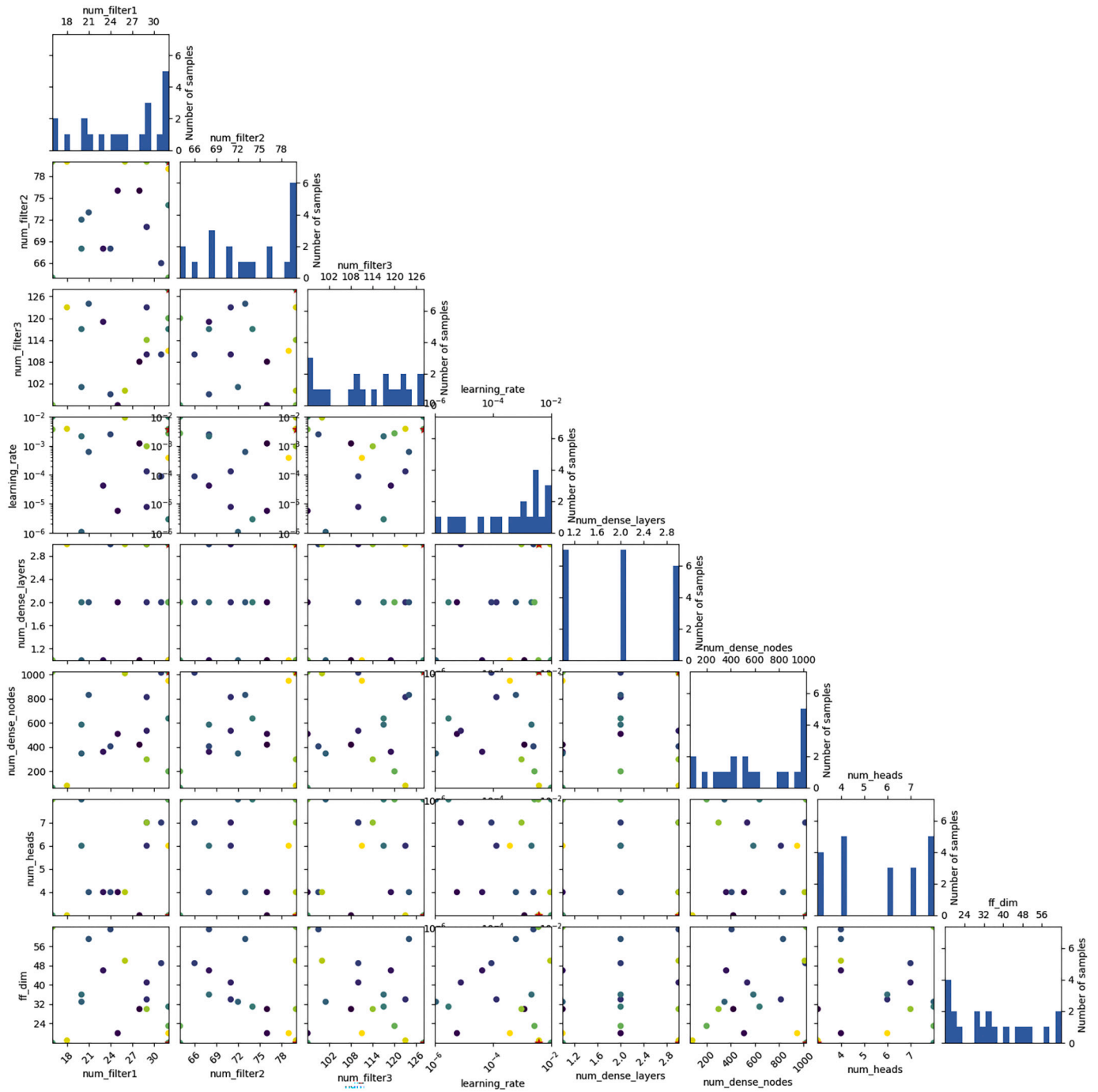
Fig. 15. The Matrix-plot. The top diagonal plots shows histograms of the sample distributions for each of the hyper-parameters during the Bayesian optimization.The plots below the top diagonal show the location of hyper-parameters in the search-space and the colour-coding shows the order in which the hyper-parameters were taken.The points in dark-purple corresponds to earlier hyper-parameters and lighter-yellow to later hyper-parameters. The points in red shows the location of the minimum or best hyper-parameter found by the Bayesian optimization process(Gilles and Manoj, 2021). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

References

Arévalo, P., Olofsson, P., Woodcock, C.E., 2019. Continuous monitoring of land change activities and post-disturbance dynamics from Landsat time series : A test methodology for REDD + reporting. Remote Sens. Environ. 1–14. URL. https://doi.org/10.1016/j.rse.2019.01.013.

Balzter, H., Cole, B., Thiel, C., Schmullius, C., 2015. Mapping CORINE land cover from Sentinel-1A SAR and SRTM digital elevation model data using random forests. Remote Sens. 7, 14876–14898. https://doi.org/10.3390/rs71114876.

Bazi, Y., Bashmal, L., Rahhal, M.M.A., Dayil, R.A., Ajlan, N.A., 2021. Vision transformers for remote sensing image classification. Remote Sens. 13, 516. URL. https://www.mdpi.com/2072-4292/13/3/516. https://doi.org/10.3390/rs13030516.

Boriah, S., 2010. Time Series Change Detection: Algorithms for Land Cover Change. Ph. D. thesis. University of Minnesota. URL. https://conservancy.umn.edu/handle/11299/90706.

Bp, G., Raju, A., Dwarakish, G.S., 2015. Different Approaches for Land Use Land Cover Change Detection : A Review Different Approaches for Land Use Land Cover Change Detection : A Review.

Campbell, D.J., Lusch, D.P., Smucker, T.A., Wangui, E.E., 2005. Multiple Methods in the Study of Driving Forces of Land Use and Land Cover Change : A Case Study of SE

Kajiado District, Kenya, 33, pp. 763–794. https://doi.org/10.1007/s10745-005-8210-y.

Castelluccio, M., Poggi, G., Sansone, C., Verdoliva, L., 2015. Land use classification in remote sensing images by convolutional neural networks. Computer Vision And Pattern Recognition 1–11. URL: http://arxiv.org/abs/1508.00092. https://doi.org/10.1021/cr8002505.

Castilla, G., Hay, G.J., 2007. Uncertainties in land use data. Hydrol. Earth Syst. Sci. Discuss. 11, 1857–1868. URL: www.hydrol-earth-syst-sci.net/11/1857/2007.

Cole, B., Smith, G., Balzter, H., 2018. Acceleration and fragmentation of CORINE land cover changes in the United Kingdom from 2006–2012 detected by Copernicus IMAGE2012 satellite data. Int. J. Appl. Earth Obs. Geoinf. 73, 107–122. URL: https://linkinghub.elsevier.com/retrieve/pii/S0303243418301314. https://doi.org/10.1016/j.jag.2018.06.003.

Comber, A., Balzter, H., Cole, B., Fisher, P., Johnson, S.C., Ogutu, B., 2016. Methods to quantify regional differences in land cover change. Remote Sens. 8, 1–19. https://doi.org/10.3390/rs8030176.

Curtis, P.G., Slay, C.M., Harris, N.L., Tyukavina, A., Hansen, M.C., 2018. Classifying drivers of global forest loss. Forest Ecol. 1111, 1108–1111.

De Sy, V., Herold, M., Achard, F., Beuchle, R., Clevers, J.G., Lindquist, E., Verchot, L., 2015. Land use patterns and related carbon losses following deforestation in South America. Environ. Res. Lett. 10, 124004. URL: https://doi.org/10.1088/1748-9326/10/12/124004.

De Sy, V., Herold, M., Achard, F., Avitabile, V., Baccini, A., Carter, S., Clevers, J.G., Lindquist, E., Pereira, M., Verchot, L., 2019. Tropical deforestation drivers and associated carbon emission factors derived from remote sensing data. Environ. Res. Lett. 14, 094022. URL: https://doi.org/10.1088/1748-9326/ab3dc6.

Descals, A., Wich, S., Meijaard, E., Gaveau, D.L.A., Peedell, S., Szantoi, Z., 2021. High-Resolution Global Map of Smallholder and Industrial Closed-Canopy Oil Palm Plantations. URL: https://doi.org/10.5194/essd-2020-159.

Doggart, N., Morgan-Brown, T., Lyimo, E., Mbilinyi, B., Meshack, C.K., Sallu, S.M., Spracklen, D.V., 2020. Agriculture is the main driver of deforestation in Tanzania. Environ. Res. Lett. 15, 034028. URL: https://doi.org/10.1088/1748-9326/ab6b35.

FAO, 2014. Agriculture, Forestry and Other Land Use Emissions by Sources and Removals by Sinks. URL: http://www.fao.org/3/i3671e/i3671e.pdf.

FAO & JRC, 2012. Global forest land-use change 1990–2005, , FAO Forestry PaperNo. 169. Food and Agriculture Organization of the United Nations and European Commission Joint Research Centre, Rome.

Fritz, S., See, L., Perger, C., McCallum, I., Schill, C., Schepaschenko, D., Duerauer, M., Karner, M., Dresel, C., Laso-Bayas, J.C., Lesiv, M., Moorthy, I., Salk, C.F., Danylo, O., Sturn, T., Albrecht, F., You, L., Kraxner, F., Obersteiner, M., 2017. A global dataset of crowdsourced land cover and land use reference data. Scientific Data 4, 1–8. URL: www.nature.com/sdata/. https://doi.org/10.1038/sdata.2017.75.

Geist, H.J., Lambin, E.F., 2001. What drives tropical deforestation? A meta-analysis of proximate and underlying causes of deforestation based on subnational case study evidence. In: Technical Report, p. 4.

Gilles, L., Manoj, K., 2021. Bayesian optimization with skopt — scikit-optimize 0.8.1 documentation. URL: https://scikit-optimize.github.io/stable/auto_examples/bayesian-optimization.html.

Google Earth Engine, 2020. USGS Landsat 7 Surface Reflectance Tier 1 | Earth Engine Data Catalog. URL: https://developers.google.com/earth-engine/datasets/catalog/LANDSAT_LE07_C01_T1_SR#description.

Hansen, M.C., Stehman, S.V., Potapov, P.V., 2010a. Quantification of global gross forest cover loss. Proc. Natl. Acad. Sci. 107, 8650–8655. https://doi.org/10.1073/pnas.0912668107.

Hansen, M.C., Stehman, S.V., Potapov, P.V., 2010b. Quantification of global gross forest cover loss. Proc. Natl. Acad. Sci. 107, 8650–8655. https://doi.org/10.1073/pnas.0912668107.

Hansen, M.C., Potapov, P.V., Moore, R., Hancher, M., Turubanova, S.A., Tyukavina, A., Thau, D., Stehman, S.V., Goetz, S.J., Loveland, T.R., Kommareddy, A., Egorov, A., Chini, L., Justice, C.O., Townshend, J.R., 2013. High-resolution global maps of 21st-century forest cover change. Science 342, 850–853. https://doi.org/10.1126/science.1244693.

Helber, P., Bischke, B., Dengel, A., Borth, D., 2017. EuroSAT: a novel dataset and deep learning benchmark for land use and land cover classification. EuroSAT 9.

Herold, M., Liu, X., Clarke, K.C., 2003. Spatial Metrics and Image Texture for Mapping Urban Land Use, 69. American Society for Photogrammetry and Remote Sensing, p. 1001. URL: http://www.geogr.uni-jena.de/~c5hema/pub/pers03_heroldetal.pdf.

Hu, Y., Wong, Y., Wei, W., Du, Y., Kankanhalli, M., Geng, W., 2018. A novel attention-based hybrid CNN-RNN architecture for sEMG-based gesture recognition. PLoS One 13. https://doi.org/10.1371/journal.pone.0206049 e0206049. URL:

Huang, B., Zhao, B., Song, Y., 2018. Urban land-use mapping using a deep convolutional neural network with high spatial resolution multispectral remote sensing imagery. Remote Sens. Environ. 214, 73–86. URL: https://doi.org/10.1016/j.rse.2018.04.050.

Hughes, L.H., Schmitt, M., Mou, L., Wang, Y., Zhu, X.X., 2018. Identifying corresponding patches in SAR and optical images with a pseudo-siamese CNN. IEEE Geosci. Remote Sens. Lett. 1–5. https://doi.org/10.1109/LGRS.2018.2799232.

Interdonato, R., Ienco, D., Gaetano, R., Ose, K., 2018. DuPLO: A DUal view Point deep Learning architecture for time series classificatiOn, 32. URL: http://arxiv.org/abs/1809.07589.

IPCC, 2013. Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Technical Report. Cambridge, United Kingdom and New York, NY, USA. URL: https://www.ipcc.ch/report/ar5/wg1/.

Irvin, J., Sheng, H., Ramachandran, N., Johnson-Yu, S., Zhou, S., Story, K., Rustowicz, R., Elsworth, C., Austin, K., Ng, A.Y., 2020. Forest-Net: Classifying Drivers of Deforestation in Indonesia using Deep Learning on Satellite Imagery, in: 34th Conference on Neural Information Processing Systems, Vancouver, p. 10. URL: https://stanfordmlgroup.github.io/projects/forestnet.

Johnson, J.M., Khoshgoftaar, T.M., 2019. Survey on deep learning with class imbalance. J. Big Data 6, 1–54. URL: https://doi.org/10.1186/s40537-019-0192-5.

Kellenberger, B., Marcos, D., Tuia, D., 2018. Detecting mammals in UAV images: best practices to address a substantially imbalanced dataset with deep learning. Remote Sens. Environ. 216, 139–153. https://doi.org/10.1016/j.rse.2018.06.028.

Kim, O.S., 2010. An Assessment of Deforestation Models for Reducing Emissions from Deforestation and Forest Degradation (REDD). Trans. GIS 14, 631–654. https://doi.org/10.1111/j.1467-9671.2010.01227.x.

Kit, O., Lüdeke, M., 2013. Automated detection of slum area change in Hyderabad, India using multitemporal satellite imagery. ISPRS J. Photogramm. Remote Sens. 83, 130–137. URL: https://doi.org/10.1016/j.isprsjprs.2013.06.009.

Kong, Y.L., Huang, Q., Wang, C., Chen, J., Chen, J., He, D., 2018. Long short-term memory neural networks for online disturbance detection in satellite image time series. Remote Sens. 10, 13. https://doi.org/10.3390/rs10030452.

Kumar, S., Krishna, R., Shiv, R., Dubey, B.B., . HybridSN: Exploring 3D-2D CNN Feature Hierarchy for Hyperspectral Image Classification. IEEE Geoscience and Remote Sensing Letters. URL: www.ehu.eus/ccwintco/index.php/Hyperspectral. https://doi.org/10.1109/LGRS.2019.2918719.

Lecun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521, 436–444. https://doi.org/10.1038/nature14539.

Li, Y., Zhang, H., Shen, Q., 2017. Spectral–spatial classification of hyperspectral imagery with 3D convolutional neural network. Remote Sens. 9, 67. URL: http://www.mdpi.com/2072-4292/9/1/67. https://doi.org/10.3390/rs9010067.

Li, Z., Bagan, H., Yamagata, Y., 2018. Analysis of spatiotemporal land cover changes in Inner Mongolia using self-organizing map neural network and grid cells method. Sci. Total Environ. 636, 1180–1191. URL: https://doi.org/10.1016/j.scitotenv.2018.04.361.

Liu, X., Kang, C., Gong, L., Liu, Y., 2016. Incorporating spatial interaction patterns in classifying and understanding urban land use. Int. J. Geogr. Inf. Sci. 30, 334–350. URL: https://doi.org/10.1080/13658816.2015.1086923.

Luus, F.P., Salmon, B.P., Van Den Bergh, F., Maharaj, B.T., 2015. Multiview deep learning for land-use classification. IEEE Geosci. Remote Sens. Lett. 12, 2448–2452. https://doi.org/10.1109/LGRS.2015.2483680.

Ma, L., Liu, Y., Zhang, X., Ye, Y., Yin, G., Johnson, B.A., 2019. Deep learning in remote sensing applications: A meta-analysis and review. https://doi.org/10.1016/j.isprsjprs.2019.04.015.

Marmanis, D., Datcu, M., Esch, T., Stilla, U., 2016. Deep learning earth observation classification using ImageNet pretrained networks. IEEE Geosci. Remote Sens. Lett. 13, 105–109. https://doi.org/10.1109/LGRS.2015.2499239.

Mas, J.F., Lemoine-Rodríguez, R., González-López, R., Piña-Garduño, A., Herrera-Flores, E., 2017. Land use / land cover change detection combining automatic processing and visual interpretation. Eur. J. Remote Sens. 50, 626–635. URL: https://doi.org/10.1080/22797254.2017.1387505.

Minh, D.H.T., Ienco, D., Gaetano, R., Lalande, N., Ndikumana, E., Osman, F., Maurel, P., 2018. Deep Recurrent Neural Networks for Winter Vegetation Quality Mapping via Multitemporal SAR Sentinel-1. IEEE Geosci. Remote Sens. Lett. 6 https://doi.org/10.1109/LGRS.2018.2794581.

Müller, H., Rufin, P., Griffiths, P., Barros Siqueira, A.J., Hostert, P., 2015. Mining dense Landsat time series for separating cropland and pasture in a heterogeneous Brazilian savanna landscape. Remote Sens. Environ. 156, 490–499. https://doi.org/10.1016/j.rse.2014.10.014.

Neagoe, V.E., Neghina, M., Datcu, M., 2012. Neural network techniques for automated land-cover change detection in multispectral satellite time series imagery. Int. J. Math. Models Methods Appl. Sci. 6, 130–139.

Nguyen, L.H., Joshi, D.R., Clay, D.E., Henebry, M., 2018. Characterizing land cover / land use from multiple years of Landsat and MODIS time series : A novel approach using land surface phenology modeling and random forest classi fi er. Remote Sens. Environ. https://doi.org/10.1016/j.rse.2018.12.016.

Oehmcke, S., Thrysøe, C., Borgstad, A., Antonio, M., Salles, V., Brandt, M., Gieseke, F., 2019. Detecting hardly visible roads in low-resolution satellite time series data. In: 2019 IEEE International Conference on Big Data, IEEE, Los Angles. https://doi.org/10.1109/BigData47090.2019.9006251.

Olofsson, P., Stehman, S.V., Woodcock, C.E., Sulla-Menashe, D., Sibley, A.M., Newell, J. D., Friedl, M.A., Herold, M., 2012. A global land-cover validation data set, part I: fundamental design principles. Int. J. Remote Sens. 33, 5768–5788. URL: https://doi.org/10.1080/01431161.2012.674230.

Pan, W., Narasimhan, H., Protopapas, P., Kar, P., Ramaswamy, H.G., . Optimizing the Multiclass F-measure via Biconcave Programming, in: 2016 IEEE 16th International Conference on Data Mining (ICDM), IEEE, Barcelona. URL: https://ieeexplore.ieee.org/document/7837956. https://doi.org/10.1109/ICDM.2016.0143.

Pandey, B., Zhang, Q., Seto, K.C., 2018. Time series analysis of satellite data to characterize multiple land use transitions : a case study of urban growth and agricultural land loss in India. J. Land Use Sci. 13, 221–237. URL: https://doi.org/10.1080/1747423X.2018.1533042.

Pelletier, C., Webb, G.I., Petitjean, F., 2019. Temporal convolutional neural network for the classification of satellite image time series. Remote Sens. 1–22.

Pengra, B., Long, J., Dahal, D., Stehman, S.V., Loveland, T.R., 2015. A global reference database from very high resolution commercial satellite data and methodology for application to Landsat derived 30m continuous field tree cover data. Remote Sens. Environ. 165, 234–248. https://doi.org/10.1016/j.rse.2015.01.018.

Reiche, J., Mullissa, A., Slagter, B., Gou, Y., Tsendbazar, N.E., Odongo-Braun, C., Vollrath, A., Weisse, M.J., Stolle, F., Pickens, A., Donchyts, G., Clinton, N., Gorelick, N., Herold, M., 2021. Forest disturbance alerts for the Congo Basin using

Sentinel-1. Environ. Res. Lett. 16, 24005. URL. https://doi.org/10.1088/1748-93 26/abd0a8.

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., Prabhat, 2019. Deep learning and process understanding for data-driven Earth system science. Nature 566, 204. URL. www.nature.com/nature. https://doi.org/10.1038/s41586-019-0912-1.

Rußwurm, M., Körner, M., 2017. Multi-temporal land cover classification with long short-term memory neural networks. In: International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives, 42, pp. 551–558. https://doi.org/10.5194/isprs-archives-XLII-1-W1–551-2017.

Rußwurm, M., Körner, M., 2018. Multi-Temporal Land Cover Classification with Sequential Recurrent Encoders. ISPRS Int. J. Geo Inf. 7, 1–19. URL. https://doi.org/10.3390/ijgi7040129.

Rußwurm, M., Körner, M., 2020. Self-attention for raw optical Satellite Time Series Classification. ISPRS J. Photogramm. Remote Sens. 169, 421–435. URL. https://doi.org/10.1016/j.isprsjprs.2020.06.006.

Rußwurm, M., Wang, S., Körner, M., Lobell, D., 2020. Meta-learning for few-shot land cover classification. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, pp. 200–201.

Shi, X., Chen, Z., Wang, H., Yeung, D.Y., Wong, W.K., Woo, W.C., Kong Observatory, H., 2015. Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting. In: Proceedings of the 28th International Conference on Neural Information Processing Systems, pp. 802–810.

Silva, A.L., Alves, D.S., Ferreira, M.P., 2018. Landsat-based land use change assessment in the brazilian atlantic forest : forest transition and sugarcane expansion. Remote Sens. 20. https://doi.org/10.3390/rs10070996.

Song, A., Choi, J., Han, Y., Kim, Y., 2018. Change detection in hyperspectral images using recurrent 3D fully convolutional networks. Remote Sens. 22 https://doi.org/10.3390/rs10111827.

Tracewski, L., Bastin, L., Fonte, C.C., Group, F, 2017. Repurposing a deep learning network to filter and classify volunteered photographs for land cover and land use characterization. Geo-spatial Informat. Sci. 20, 252–268. URL. https://doi.org/10.1080/10095020.2017.1373955.

Tseng, G., Kerner, H., Nakalembe, C., Becker-Reshef, I., 2021. Learning to predict crop type from heterogeneous sparse labels using meta-learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, pp. 1111–1120. URL. https://doi.org/10.5281/zenodo.4680394.

Uba, N.K., 2016. Land Use and Land Cover Classification Using Deep Learning Techniques. Ph.D. thesis. ARIZONA STATE UNIVERSITY.

UNFCCC, 2017. UN Climate change: Annual Report 2017. Technical Report. United Nations Framework Convention on Climate Change. Bonn. URL. https://unfccc.in t/sites/default/files/resource/UNClimateChange_annualreport2017_final.pdf.

UNFCCC, 2018. The Katowice Climate Package: Making The Paris Agreement Work For All | UNFCCC. URL. https://unfccc.int/process-and-meetings/the-paris-agreement /katowice-climate-package.

van Asselen, S., Verburg, P.H., 2012. A Land System representation for global assessments and land-use modeling. Glob. Chang. Biol. 18, 3125–3148. https://doi.org/10.1111/j.1365-2486.2012.02759.x.

Verbesselt, J., Hyndman, R., Newnham, G., Culvenor, D., 2010a. Detecting trend and seasonal changes in satellite image time series. Remote Sens. Environ. 114, 106–115. URL. https://doi.org/10.1016/j.rse.2009.08.014.

Verbesselt, J., Hyndman, R., Newnham, G., Culvenor, D., 2010b. Detecting trend and seasonal changes in satellite image time series. Remote Sens. Environ. 114, 106–115. URL. https://doi.org/10.1016/j.rse.2009.08.014.

Vogelmann, J.E., Xian, G., Homer, C., Tolk, B., 2012. Monitoring gradual ecosystem change using Landsat time series analyses: Case studies in selected forest and rangeland ecosystems. Remote Sens. Environ. 122, 92–105. URL. https://doi.org/10.1016/j.rse.2011.06.027.

Wang, H., Zhao, X., Zhang, X., Wu, D., Du, X., 2019. Long time series land cover classification in China from 1982 to 2015 Based on Bi-LSTM Deep Learning. Remote Sens. 11, 1639. URL. https://www.mdpi.com/2072-4292/11/14/1639. https://doi.org/10.3390/rs11141639.

World Resources Institute, 2016. INSIDER: Global Forest Watch and the Forest Resources Assessment, Explained in 5 Graphics | World Resources Institute. URL. https://www.wri.org/blog/2016/08/insider-global-forest-watch-and-the-forest-resources-assessment -explained-5-graphics.

Xu, Z., Guan, K., Casler, N., Peng, B., Wang, S., 2018. A 3D convolutional neural network method for land cover classification using LiDAR and multi-temporal Landsat imagery. ISPRS J. Photogramm. Remote Sens. 144, 423–434. https://doi.org/10.1016/j.isprsjprs.2018.08.005.

Yang, R., Singh, S.K., Tavakkoli, M., Amiri, N., Yang, Y., Karami, M.A., Rai, R., 2020. CNN-LSTM deep learning architecture for computer vision-based modal frequency detection. Mech. Syst. Signal Process. 144, 106885. https://doi.org/10.1016/j.ymssp.2020.106885.

Yuan, Q., Shen, H., Li, T., Li, Z., Li, S., Jiang, Y., Xu, H., Tan, W., Yang, Q., Wang, J., Gao, J., Zhang, L., 2020. Deep learning in environmental remote sensing: Achievements and challenges. Remote Sens. Environ. 241, 111716. https://doi.org/10.1016/j.rse.2020.111716.

Zhang, Y., Yang, B., Liu, X., Wang, C., 2017. Estimation of rice grain yield from dual-polarization Radarsat-2 SAR data by integrating a rice canopy scattering model and a genetic algorithm. Int. J. Appl. Earth Obs. Geoinf. 57, 75–85. URL. http://linki nghub.elsevier.com/retrieve/pii/S0303243416302124. https://doi.org/10.1016/j.jag.2016.12.014.

Zhang, C., Sargent, I., Pan, X., Li, H., Gardiner, A., Hare, J., Atkinson, P.M., 2019. Joint Deep Learning for land cover and land use classification. Remote Sens. Environ. 221, 173–187. URL. https://doi.org/10.1016/j.rse.2018.11.014.

Zhao, W., Du, S., 2016. Learning multiscale and deep representations for classifying remotely sensed imagery. ISPRS J. Photogramm. Remote Sens. 113, 155–165. https://doi.org/10.1016/j.isprsjprs.2016.01.004.

Zhao, Y., Zhang, K., Fu, Y., Zhang, H., 2012. Examining Land-Use / Land-Cover Change in the Lake Dianchi Watershed of the Yunnan-Guizhou Plateau of Southwest China with Remote Sensing and GIS Techniques : 1974–2008, 3843–3865. https://doi.org/10.3390/ijerph9113843.

Zhong, L., Hu, L., Zhou, H., 2019. Deep learning based multi-temporal crop classification. Remote Sens. Environ. 221, 430–443. URL. https://doi.org/10.1016/j.rse.2018.11.032.

Zhu, X.X., Tuia, D., Mou, L., Xia, G.S., Zhang, L., Xu, F., Fraundorfer, F., 2017. Deep Learning in Remote Sensing: A Comprehensive Review and List of Resources. IEEE Geoscience and Remote Sensing Magazine, 5, pp. 8–36. URL. http://ieeexplore.ieee.org/document/8113128/. https://doi.org/10.1109/MGRS.2017.2762307.

Zhu, J., Chen, H., Ye, W., 2020. A hybrid CNN-LSTM network for the classification of human activities based on Micro-Doppler Radar. IEEE Access 8, 24713–24720. https://doi.org/10.1109/ACCESS.2020.2971064.