ENVIRONMENTAL RESEARCH
LETTERS

LETTER • OPEN ACCESS

Mapping drivers of tropical forest loss with satellite image time series and machine learning

To cite this article: Jan Pišl *et al* 2024 *Environ. Res. Lett.* **19** 064053

IOP Publishing Environ. Res. Lett. 19 (2024) 064053 https://doi.org/10.1088/1748-9326/ad44b2

ENVIRONMENTAL RESEARCH
LETTERS

LETTER

Mapping drivers of tropical forest loss with satellite image time series and machine learning

Jan Pišl[1,*], Marc Rußwurm[1,2], Lloyd Haydn Hughes[1], Gaston Lenczner[1], Linda See[3], Jan Dirk Wegner[4] and Devis Tuia[1]

[1] École Polytechnique Fédérale de Lausanne (EPFL), Sion, Switzerland
[2] Wageningen University, Wageningen, The Netherlands
[3] International Institute for Applied Systems Analysis (IIASA), Laxenburg, Austria
[4] University of Zurich, Zurich, Switzerland
[*] Author to whom any correspondence should be addressed.

E-mail: jan.pisl@epfl.ch

Keywords: remote sensing, earth observation, machine learning, deep learning, time series, deforestation, tropical forest

OPEN ACCESS

RECEIVED
12 November 2023

REVISED
15 March 2024

ACCEPTED FOR PUBLICATION
30 April 2024

PUBLISHED
29 May 2024

Abstract

The rates of tropical deforestation remain high, resulting in carbon emissions, biodiversity loss, and impacts on local communities. To design effective policies to tackle this, it is necessary to know what the drivers behind deforestation are. Since drivers vary in space and time, producing accurate spatially explicit maps with regular temporal updates is essential. Drivers can be recognized from satellite imagery but the scale of tropical deforestation makes it unfeasible to do so manually. Machine learning opens up possibilities for automating and scaling up this process. In this study, we developed and trained a deep learning model to classify the drivers of any forest loss—including deforestation—from satellite image time series. Our model architecture allows understanding of how the input time series is used to make a prediction, showing the model learns different patterns for recognizing each driver and highlighting the need for temporal data. We used our model to classify over 588′000 sites to produce a map detailing the drivers behind tropical forest loss. The results confirm that the majority of it is driven by agriculture, but also show significant regional differences. Such data is a crucial source of information to enable targeting specific drivers locally and can be updated in the future using free satellite data.

1. Introduction

1.1. Drivers of deforestation

Tropical forests are amongst the most valuable ecosystems on the planet. They are epicenters of biodiversity [1], can store more carbon than any other land ecosystem [2], and provide drinking water, shelter, and wood to hundreds of millions of people [3]. However, they also experience consistently higher rates of deforestation than any other type of forests [2]. For example, 95% of the estimated 79 mHa of global deforestation that occurred between 2001 and 2015 was located in tropical regions [4].

There is a number of driving forces, *drivers*, behind tropical deforestation that vary regionally and in time [5]. For example, the most widespread deforestation driver is pasture in the Brazilian Amazon, oil palm plantation in Indonesia, and subsistence farming in the Congo Basin [5, 6]. Tackling deforestation effectively requires solutions and policies tailored to individual drivers [5]. The Amazon Soy Moratorium (ASM) in Brazil serves as a prime example. By signing the ASM, soy traders agreed not to purchase soy from recently deforested lands. Together with other complementary policies, this has resulted in a decline in deforestation in the Brazilian Amazon by 84% [5, 7]. To design such targeted policies, the underlying drivers must be known.

Satellite-based remote sensing opens up possibilities for monitoring forests at a high spatial resolution on a global scale. Products such as the Global Forest Change (GFC) [8] can detect any forest loss at a 30 m resolution. This encompasses deforestation (the conversion of forest to a different land use) and also

temporary disturbances such as a clearance at a forest plantation, where the forest is then left to regrow. In the tropics, deforestation is the primary concern, but in the absence of comprehensive deforestation-only data, forest loss data serves as a crucial starting point for better understanding the dynamics of tropical forest change as it can be used to identify areas where further investigation is needed. Once the drivers of forest loss are known, it can be determined which forest loss events are temporary (e.g. drivers such as wildfire or clearing of a forest plantation) and which correspond to deforestation (e.g. drivers associated with land use change such as agriculture or mining). Recently, the dataset of Vancutsem *et al* (2021) [9] largely closed this gap by providing data directly on deforestation. However, the dataset only covers moist tropical forests.

1.2. Mapping drivers

Given the scale and rate of tropical forest loss, it is not feasible to visit each site on the ground to determine the driver. Estimates have been produced using land-balancing models, international trade data, or reports from individual governments [10, 11], but they are limited in accuracy [12] and can only describe drivers at a national or sub-national level, which limits the effectivity of the policies based on such data [5].

Similar to mapping forest loss itself, remote sensing has become a vital tool for spatially explicit attribution of forest loss to drivers. The high spatial resolution of missions such as Landsat and Sentinel-2 has enabled the recognition of drivers across diverse landscapes [13–16]. Manual interpretation of the images is possible but only a small fraction of the detected forest loss can be analyzed in this way. Automation is necessary to scale up driver mapping and reduce the amount of human effort needed.

1.3. Machine learning for automatic recognition of drivers

Curtis *et al* (2018) [4] showed the potential of machine learning for the automatic recognition of drivers. They trained an ensemble of decision trees on population and remote sensing-based datasets to classify the dominant driver for the period 2001–2015 in every 10×10 km cell, producing the first global, spatially explicit map of drivers. This map remains widely used but comes with a set of limitations. In addition to its coarse spatial and temporal resolution, the model is trained on manually crafted features from a specific set of datasets. The model therefore relies on these datasets and may not be applied to other time periods if these datasets are not available.

More recently, deep learning (DL)-based approaches have been proposed that recognize drivers directly from satellite images. The end-to-end learning paradigm of DL enables the models to learn descriptive, problem-specific features from raw input data, alleviating the need for manual feature engineering and opening up the possibilities for finer-scale driver recognition. Among DL approaches, convolutional neural networks (CNNs) are most commonly used to extract visual features from images. CNNs are designed to take into account the spatial dimension of image data, making them well-suited for such a task. Once extracted, the visual features can be used directly as input into a classifier [17], in some cases augmented with manually extracted features from auxiliary datasets [18]. A concurrent forest loss segmentation and driver classification has been proposed by Mitton *et al* (2021) [19], where the visual features are classified into driver categories and also upsampled to the original image dimensions to produce a forest loss map. Vision Transformers [20] have also been shown to match CNNs for this task [21].

Despite promising results, mapping drivers from single images has its limits. Distinct spatial patterns associated with individual drivers appear at different points in time. For example, wildfire can be best recognized immediately after it occurs but it may take several years before certain crops (particularly tree crops) have grown enough to be recognizable. There may be features that appear even before the forest loss, such as a logging road before forest clearing. Therefore, exploiting the temporal information seems crucial. The high revisit frequency of satellite remote sensing makes it possible to replace single images with time series as inputs. However, using time series also brings an additional layer of complexity. The spatio-temporal nature of the input needs to be reflected in the model architecture to learn feature representations useful for the recognition of forest loss drivers. This is an active research topic in satellite image processing and many approaches have been proposed, mostly for crop type classification, using 1D or 3D convolution [22, 23], combining convolutional and recurrent modules [24–26] or utilizing the attention mechanism [27–29]. The adoption of spatio-temporal DL for mapping forest loss drivers is in its infancy, with a single study showing a significant increase in accuracy when compared to single-image approaches [17].

1.4. Challenges

Despite these advancements, the availability of driver maps across the tropics remains limited. Existing datasets only cover a single region [18, 30], have a coarse resolution [4], or are only rough estimates based on a limited number of samples [31, 32]. A comparison between different driver maps has shown major discrepancies [12].

The reasons are manifold. The size and heterogeneity of the tropics make robust recognition of drivers difficult, as the visual appearance of a class may vary depending on its location [17]. To be robust to such variations, DL models need to be trained with large amounts of annotated data. Currently, driver

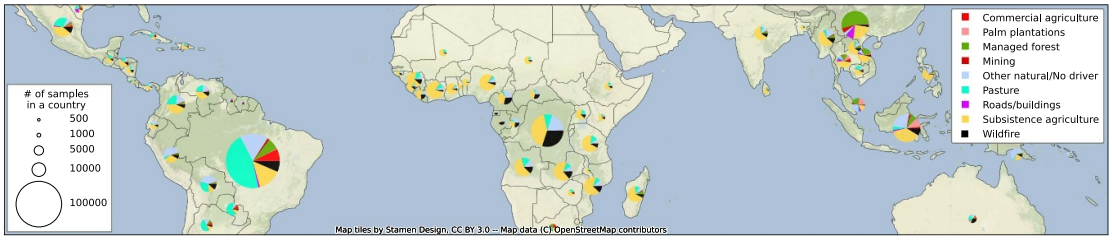


Figure 1. Forest loss drivers per country.

annotation is scarce, often focusing on a single region [18, 30] or sparse sampling [31]. Additionally, as discussed above, the distinct visual features of each driver may appear at different points in time. This can be modeled by approaches based on time series as they can contain features for all drivers.

Disentangling and tackling these challenges with DL models is complicated by the non-transparent way in which they work. The models are trained on large amounts of raw data and often contain tens of millions of parameters. As a result, they are capable of learning complex mappings from input to output, connecting, in our case, satellite images to forest loss drivers, but they do not provide any explanations as to why they made a particular prediction, and act as *black boxes*. This makes it difficult to understand how they work and therefore how to improve them. For example, while there is evidence that using a time series results in higher accuracy, it is not clear why, how many images should form a time series, and how the images should be sampled.

1.5. Contributions

In this paper, we propose and implement a new, DL-based method for the classification of forest loss drivers from time series of Sentinel-2 images. This method allows for a dense and spatially explicit mapping of forest loss drivers. We deploy it to classify over half a million sites of forest loss across the tropics to produce a map of drivers for the period 2017–2020, shown in figure 1. We analyze the produced data and compare it with a widely used driver map of Curtis *et al* (2018), uncovering new forest loss patterns thanks to the more recent data, higher spatial resolution and more fine-grained driver categorization used in our work. Thanks to the high revisit rate of Sentinel–2, the resulting map can be regularly updated without human intervention.

The DL method is described in detail in appendix A. We designed the model with a temporal attention module that promotes explainability, allowing the most important features to be identified for accurate driver classification, where they appear in time, and how this varies for different drivers. To train our DL model, we curated a dataset utilizing a recent crowd-sourced campaign which we augmented with examples from other sources. We describe the dataset in detail in appendix B. While our primary motivation is deforestation, we consider all types of forest loss. Once the forest loss driver is known, it is straightforward to determine whether a given forest loss event corresponds to deforestation or to a temporary disturbance.

2. Experiments

The experiments were carried out to understand (i) if using a time series yields better results compared to a single image in time, (ii) how long the time series should be, and (iii) what role the model architecture plays. We trained the proposed architecture and the baselines, which are described in appendix A, with inputs varying from a single image to 12 images. The process of constructing the time series is detailed in appendix C and details on the experimental setup can be found in appendix D.

2.1. Inference

We used the best-performing model to produce a pantropical map of drivers. To do so, we sampled 588′000 sites of forest loss between 30°N and 30°S using GFC and predicted the driver for each 1 km^2 site. We only sampled sites where Curtis *et al* (2018) predicted a driver to allow for a comparison. We considered it important to compare our results to Curtis *et al* (2018) because their work is widely used as a reference.

However, the comparison can remain only qualitative, given the differences in methodology and time period:

- Curtis *et al* (2018) classified drivers for the period 2001-2015 while we focused on the period 2017–2020,
- our model predicts drivers for individual forest loss sites at a spatial resolution of 1 km^2, while Curtis *et al* (2018) assigned the major driver to each 10 km^2 cell in a grid,
- the label space used in Curtis *et al* (2018) only contains 5 classes (*commodity-driven deforestation*, *shifting agriculture*, *forestry*, *wildfire*, *urbanization*) while we used 9 classes.

3. Results

In this section, we first report on how the different models compare when trained on time series of varying length and we identify the best-performing model from all the experiments. Then, we analyze the attention scores produced by this model to better understand its behavior. Finally, we present the results of classifying 588′000 sites of forest loss sampled across the tropics.

3.1. Model comparison

The proposed spatio-temporal models outperform the *CNN* baseline if multiple images are used, as shown in figure 2. Adding images from the year before the forest loss does not have a significant positive impact on the F1 score. The results per class are shown in table 1, confirming that the spatio-temporal models are more accurate than the *CNN* baseline for all classes. The model '*CNN-Attention-LSTM*' outperforms '*CNN-LSTM*', suggesting that the temporal attention module contributes to the correct recognition of the drivers.

Figure 3 shows the F1 score disaggregated per class. The first row shows three classes that exhibit significant differences. With respect to the class *managed forest*, all three models benefit from longer sequences. The models still show improvement when images from the year *before* the estimated forest loss are also used as input. We hypothesize that this may be because indicators of forest management are often visible continuously, not only after the forest loss event.

In contrast, for recognizing the class *pasture*, the best performance by all models is reached when the amount of input data is limited. The models still benefit from using time series, but the performance starts to deteriorate when using more than 8 images (i.e. when using images acquired before the forest loss). We believe that this may be due to the forest loss patterns associated with this driver—they are often large-scale, one-time clearances that are not preceded by any indicators that the model could recognize on images *before* the clearance. The distinct features of this class may lie in the spectro-temporal response which can be seen after the forest loss event.

When looking at the class *mining*, the importance of a dedicated spatio-temporal model becomes apparent. Both *CNN-LSTM* and *CNN-Attention-LSTM* benefit from an increasing number of images, while the performance of the CNN baseline degrades significantly.

Overall, most classes can be recognized better with the dedicated spatio-temporal models proposed in this work. This is not true for *wildfire* which is associated with features very different from other classes as discussed in section 3. Also, most classes can be better recognized with more data but there are little or no benefits from including images from before the forest loss.

Overall, the proposed model learned to recognize the drivers related to agriculture and mining relatively accurately. However, it is less accurate with the classes *roads/buildings*, *other natural/no driver* and *wildfire*. As for the first, the problem may lie in insufficient and noisy annotations. As for *other natural/no driver*, we believe that this is a difficult class for the model to recognize because it combines multiple relatively rare drivers. It may be beneficial to divide this class into multiple subclasses such as *water*, *windthrow* (trees uprooted by wind), and *no driver*.

The class *wildfire* proves difficult to recognize even with high-quality examples. We believe that this is caused by *wildfire* having distinct visual features which are very different from those of other classes. This is discussed in more detail in section 3.2.

3.2. Attention score analysis

We used the best-performing variant of the CNN-Attention-LSTM model that was trained on time series of 7 images. Here we analyze the attention scores that the model produces when making correct predictions. A quantitative analysis can be found in appendix E.

Figure 4 shows three examples of an input time series and the corresponding scores. These scores, corresponding to the elements of vector **a** in equation (A.2), indicate which images the model attended to. Figure 4(a) shows that the model can attend to multiple images when needed, arguably because temporal patterns are important to classify that particular time series. In other cases, such as the one shown in figure 4(b), the model mostly relies on a single image as it is enough to predict the driver. This is most common with the driver *wildfire* that is often associated with distinct burnt areas that appear shortly after the fire. More examples of input time series and the corresponding attention scores are presented in figure E4.

Figure 4(c) shows that the model has learned to implicitly ignore cloudy images. This is because, during training, the model learns to identify features from the input images that are associated with individual classes and to use these features to make a prediction. For example, a set of rectangular shapes with bright colors and sharp edges may be associated with buildings and urban structures. Similarly, the model also learns which features do *not* correspond to any particular class and therefore are not useful for the classification task. This includes clouds since any forest loss driver can be covered by clouds. As a result, the model learns to ignore cloudy images, which alleviates the need for cloud masking algorithms when preprocessing images.

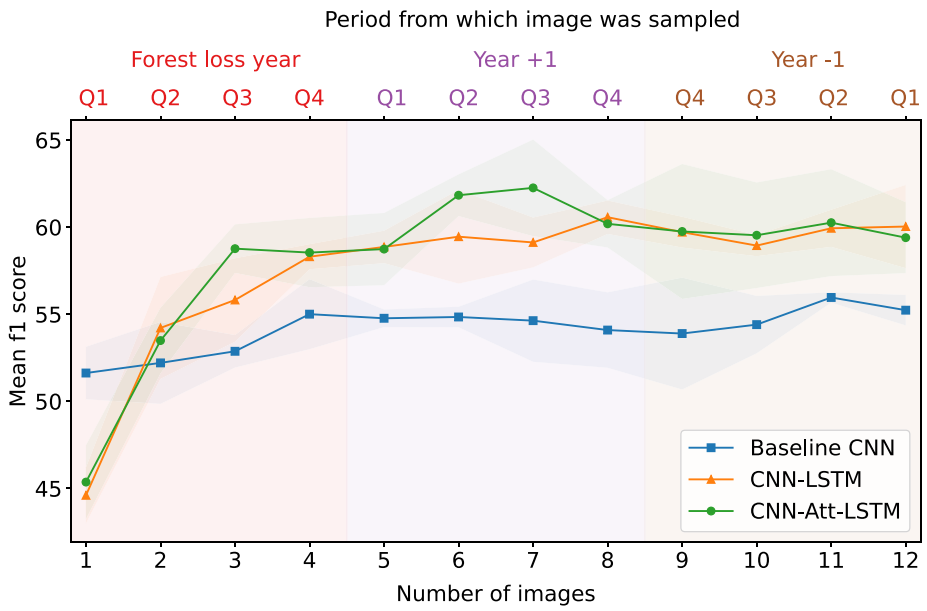


Figure 2. F1 score for time series of different lengths; from left to right, each time series contains all the images from the previous shorter time series plus one additional image from a period shown at the top *x*-axis; the sampling is detailed in appendix C.

Table 1. Per-class precision, recall and F1 score achieved by the three tested architectures; for each architecture, we chose the best performing model; the best F1 score for each class is in bold.

	CNN			CNN-LSTM			CNN-Att-LSTM		
	Prec.	Recall	F1	Prec.	Recall	F1	Prec.	Recall	F1
Commercial agr.	0.79	0.37	0.50	0.72	0.44	0.55	0.72	0.46	**0.56**
Palm plantations	0.84	0.53	0.65	0.87	0.56	**0.68**	0.91	0.51	0.66
Managed forest	0.61	0.64	0.62	0.71	0.60	**0.65**	0.66	0.60	0.63
Mining	0.82	0.74	0.78	0.91	0.88	0.89	0.90	0.93	**0.91**
Other/no driver	0.51	0.39	0.44	0.60	0.34	0.44	0.66	0.43	**0.52**
Pasture	0.58	0.63	0.61	0.62	0.66	0.64	0.62	0.72	**0.67**
Roads/buildings	0.61	0.35	0.43	0.71	0.47	**0.56**	0.73	0.39	0.51
Subsistence agr.	0.51	0.69	0.58	0.60	0.67	0.62	0.58	0.78	**0.66**
Wildfire	0.33	0.60	0.43	0.26	0.76	0.39	0.38	0.67	**0.49**
Macro Average	0.62	0.55	0.57	0.67	0.60	0.60	0.68	0.61	**0.62**

3.3. Inference

In this section, we present and discuss the results of the driver predictions at 1 km^2 resolution across the tropics, both aggregated by country (figure 1) and by latitude and longitude (figure 5). For the latter, we also present the results of Curtis *et al* (2018) processed in the same way for qualitative comparison. To ease the comparison, we describe which classes from both datasets correspond to each other in table 2.

The results we obtained confirm that the majority of tropical forest loss is human-induced. In the Americas, both datasets agree that commodity production drives forest loss in the Brazilian Amazon and small-scale agriculture is more prevalent in Colombia, Peru and Ecuador as well as in Central America. Additionally, our dataset shows that the major driver in the Brazilian Amazon is mostly pasture, which agrees with other existing data [12].

Our data also shows commercial agriculture being more prevalent in the Southeast direction from the Amazon, in the regions of Cerrado and the Atlantic rainforest. While we do not recognize individual crop types, we believe this is largely soy cultivation, which we also confirmed by visual interpretation of multiple samples in the region. It has been documented that after the Soy Moratorium was signed in 2006, by which commodity traders agreed not to purchase soy originating in lands where the Amazon rainforest had been cleared, soy production has shifted to these regions [33]. However, since the dataset of Curtis *et al* (2018) does not distinguish between pasture and crop cultivation, it is not possible to evaluate whether earlier deforestation in the Brazilian Amazon was driven more by crops such as soy, as opposed to pasture.

In Africa, shifting agriculture is by far the most common driver. The dataset of Curtis *et al* (2018) predicts almost exclusively this driver throughout the

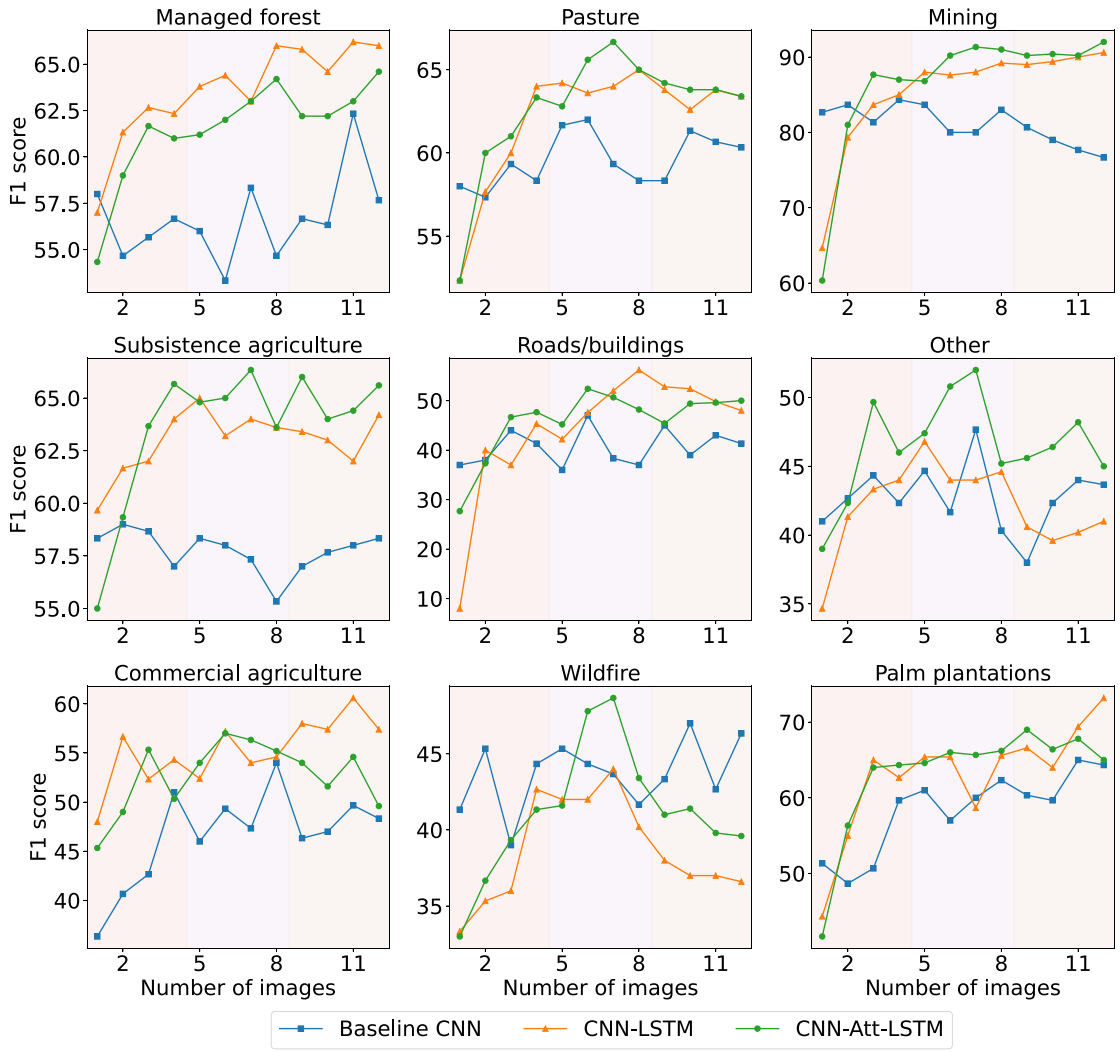


Figure 3. F1 score of the proposed architectures for different numbers of input images; disaggregated per class; the sampling of the time series is detailed in appendix C.

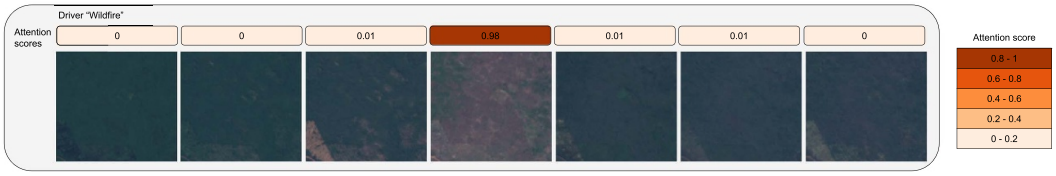
entire continent. Our model predicts a substantial portion of wildfire as a driver as well as other natural disturbances. However, our model has limited accuracy with respect to wildfire and the definition of the class *other natural causes/no driver* contains false positives in the GFC dataset, as we describe below. For these reasons, we believe that the results of Curtis *et al* (2018) might be more accurate for the African continent.

Both datasets predict forestry to be dominant in higher latitudes in Asia, especially in China. Compared to the map of Curtis *et al* (2018), our model predicts relatively little *oil palm plantation* in Indonesia and Malaysia, where it is the most common driver [6]. We see three possible explanations. First, the forest loss due to oil palm has decreased between the study period of Curtis *et al* (2018) and ours, as reported by [15]. Second, our model confuses the class forestry with palm oil plantation because of their visual similarity. Third, palm trees may take several years before they grow to such size they can be recognized from remote sensing images. Therefore, the young palm plantation may be missed by the model.

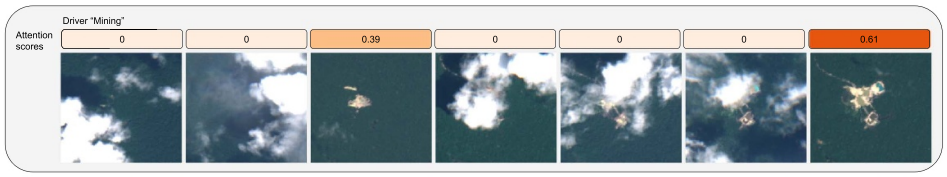
Across the study area, the class *other natural causes/no driver* is predicted relatively often, especially at lower latitudes as seen in figure 5(a). In other driver datasets, this is a minor class with only a few percent [15] or it is not considered at all [4, 17]. We believe that this may be because this class also includes cases where there was no driver visible. Given the false positive rate of GFC estimated at 13% [8], both our training and inference datasets likely contain examples with no real forest loss and the high occurrence of this class can also be related to this, as opposed to natural disturbances as a driver.

The two datasets show little agreement with respect to wildfire. According to Curtis *et al* (2018), wildfire as a dominant driver is found mostly at higher latitudes and specifically in Australia and Central America. In contrast, our model predicts wildfire more often in general and particularly around the equator. To some extent, this can be attributed to the increased amount of wildfires that

(a) Multiple images are combined to recognize *commercial agriculture* as a driver, possibly due to distinct phenological features

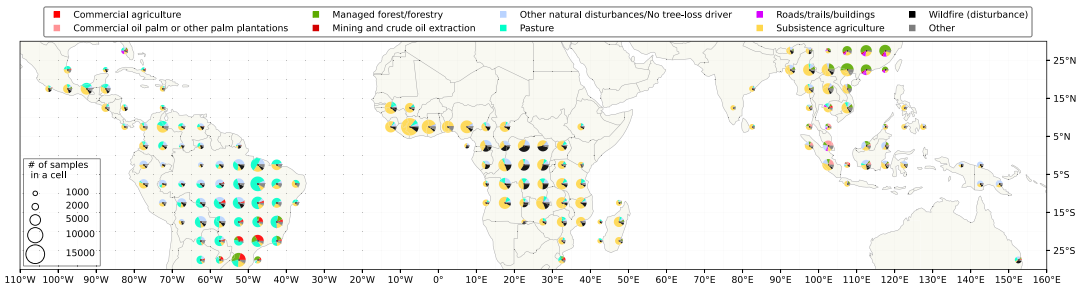


(b) The model mostly attends to a single image with visible burned area to identify *wildfire*

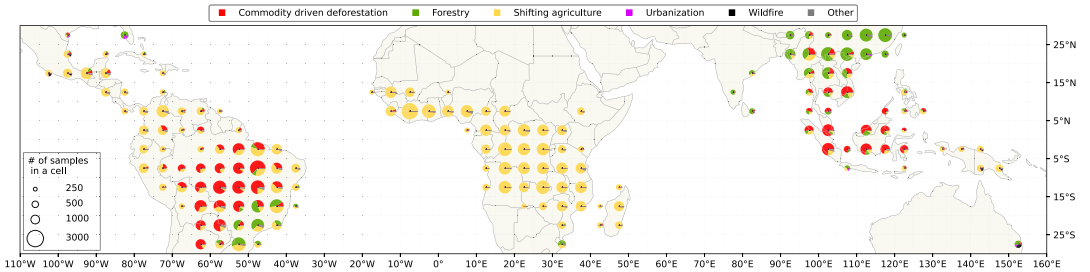


(c) The model learns to ignore cloudy images and attends to only two cloud-free images to identify driver *mining*

Figure 4. Examples of attention scores the model produced for different time series.



(a) Our results - tropical forest loss drivers in 2017-2020 (groups with less than 1000 predictions were disregarded)



(b) Results of Curtis et al. (2018) - forest loss drivers in 2001-2015 (groups with less than 250 predictions were disregarded)

Figure 5. All predictions were aggregated using a grid of 5° of longitude and 5° of latitude; the distribution of drivers within each group is drawn in the center of the corresponding cell; the size of the cell corresponds to the relative size of the group; the relatively smaller numbers of predictions in Curtis *et al* (2018) are due to the coarser spatial resolution of their analysis.

Table 2. Comparison of the classes used in this work and in Curtis *et al* (2018), using the definitions used in both works.

Class name in our work	Equivalent in Curtis *et al* (2018)
Subsistence agriculture	Shifting agriculture
Managed forest	Forestry
Pasture	Commodity driven deforestation
Roads/buildings	Urbanization
Commercial agriculture	Commodity driven deforestation
Wildfire	Wildfire
Oil palm plantations	Commodity driven deforestation
Mining	Commodity driven deforestation
Other natural causes/no driver	—

occurred during our study period 2017–2020 [34]. It is also likely due to our model overpredicting this driver, given the low precision values for this class in table 1.

4. Discussion

In this work, we presented, trained, and evaluated a DL model, 'CNN-Attention-LSTM', for the recognition of forest loss drivers from Sentinel-2 time series. We tackled the main challenges associated with this task—the lack of labeled training data, the heterogeneity of the tropical landscape, and the various temporal patterns in which drivers are visible from satellite imagery. We trained our model on a large, crowdsourced dataset, alleviating the need for manual image interpretation and providing a rich and diverse set of examples across the entire tropical region.

We showed that we can map forest loss drivers with time series of Sentinel-2 images and gain a significant performance boost over mapping from single images. Extensive experiments were performed to understand the optimal length and temporal sampling of the input time series and to compare different architectural variants. We designed our model to include a temporal attention module that enables the user to inspect which images from the time series were used to make a prediction, adding transparency and interpretability to the results. The analysis of these data shows how the model uses different strategies to recognize individual drivers and also learns to ignore cloudy images as they do not contain useful information.

Finally, we used the trained model to produce a pantropical map of drivers of recent forest loss. This map represents not only a temporal update compared to the widely adopted drivers map of Curtis *et al* (2018), but importantly demonstrates that using modern DL methods, drivers can be mapped at a high spatial resolution—the resolution of our method is two orders of magnitude higher compared to Curtis *et al* (2018). Given the high revisit rate of Sentinel-2, this map can be updated regularly with minimal user intervention. Moreover, since we have shown how the model implicitly learns to ignore cloudy images, the model can be used on raw Sentinel-2 data without requiring sophisticated preprocessing workflows. We hope this contributes to the wider adoption of the model.

The results confirm the trends of tropical forest loss—most is driven by the conversion of forests to agriculture. We demonstrate the importance of recognizing more driver categories as we show that the type of agriculture and the commodity produced differ widely. In the Amazon, cattle pastures dominate while subsistence agriculture is most common across Africa. Southeast Asia experiences a mix of different drivers, with oil palm plantations and subsistence agriculture most prevalent. Our results show that wildfire is becoming a more common driver even in tropical regions. This confirms the trends reported by Tyukavina *et al* (2022).

We would like to point out some limitations of our work. The accuracy of our model is limited. Specifically, considering the results presented in table 1, our model is likely to have overpredicted *wildfire* and *subsistence agriculture* and underpredicted the share of *commercial agriculture*, *roads/buildings* and *oil palm plantations*. This must be taken into account when using the data for decision-making. While the presented map demonstrates a significant spatial and temporal improvement over existing works, it should not be used as a single source of data when designing policies for tackling deforestation. The model could certainly be improved by using more training data, but manual high-quality annotation remains a crucial pre-requisite that requires significant human labor. Furthermore, although we recognize as many or more driver classes than all previous studies, we do not differentiate between individual agricultural commodities. This is important to support the development of policies targeted at specific industries and supply chains. An important extension of this line of work could therefore be to increase of the granularity of the driver maps, distinguishing between specific commodities. Finally, some of the driver classes are not mutually exclusive. While most correspond to the subsequent land use after the forest loss event, *wildfire* and *other natural causes/no driver* correspond to the forest disturbance type. Therefore, they can in principle appear in combination with any subsequent land use.

In future work, we would like to investigate the possibility of learning multiple regional models that would share a single, global backbone. Such models could specialize in each region while the backbone could utilize all the data.

Overall, our work demonstrated that DL models can be used for mapping drivers from free satellite images at a high spatial resolution and on a large scale. We showed that DL models can be designed to be more transparent and that the insights gained can be used to better understand how they work and how they can be improved. The data produced by our model serve as a timely update to identifying rapidly-evolving drivers to forest loss. We hope that the presented findings can support further investigations and contribute to reducing the rates of tropical deforestation.

Data availability statement

The data cannot be made publicly available upon publication because the cost of preparing, depositing and hosting the data would be prohibitive within the terms of this research project. The data that support the findings of this study are available upon reasonable request from the authors.

Acknowledgments

This project has received funding from the European Union's Horizon 2020 research and innovation programme Under the Marie Skłodowska-Curie Grant Agreement No. 945363.

Appendix A. Methods

In this section, we detail the DL approach we propose for driver classification from satellite image time series, named *CNN-Attention-LSTM*. This model uses the sequence nature of the time series and produces per-timestep importance scores (i.e. attention scores) showing which timesteps the model is using for its prediction. We also present two baseline models: a simple CNN, where the time information is provided as a concatenation of all images in the time series (CNN), and a spatio-temporal model *CNN-LSTM* that utilizes a CNN as well as a long short-term memory (LSTM) module [35]. *CNN-LSTM* exploits the fact that the images are in a temporal sequence, but does not provide any information about timestep importance.

We used Sentinel-2 images as input to all models. Sentinel-2 has a revisit time of 5 days and its sensors have 13 spectral bands, from which we used 4 that are available at 10 meter resolution (red, green, blue, and near-infrared). Nevertheless, the presented method is sensor-agnostic and can be used with images from other sensors and with other characteristics.

In the following, we denote the input to the model (a time series of remote sensing images) as $\mathbf{X}$, of shape $T \times C \times W \times H$ (T is the number of images in the time series, C is the number of image bands and W and H are image width and height, respectively). All models share the common structure of (i) an encoder that extracts features from $\mathbf{X}$ and (ii) a classifier that maps the encoded features to probabilities for each output class. All models have a single fully connected layer as a classifier and differ in the way the encoder is designed. To extract visual features from individual images, all models use a single ResNet34 [36] with the last two fully connected layers removed.

A.1. Baseline

A.1.1. CNN

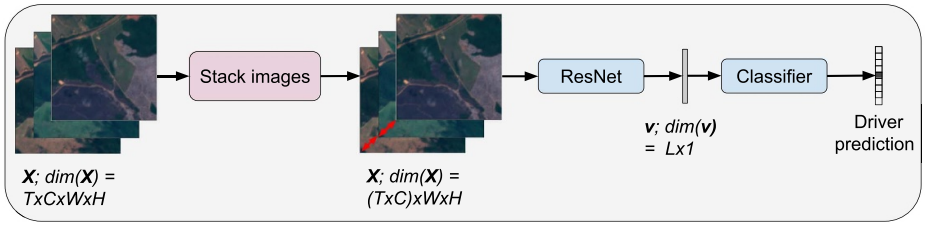
The encoder of the first baseline consists only of the ResNet34 backbone. To enable time series as input, we concatenate the images along the channel dimension and feed them to the encoder as a single datacube (figure A1). The model encodes the input in a single feature vector, denoted $\mathbf{v}$. We denote the size of $\mathbf{v}$ as L and set it to 512. This vector is then passed to the classifier.

A.1.2. LSTM

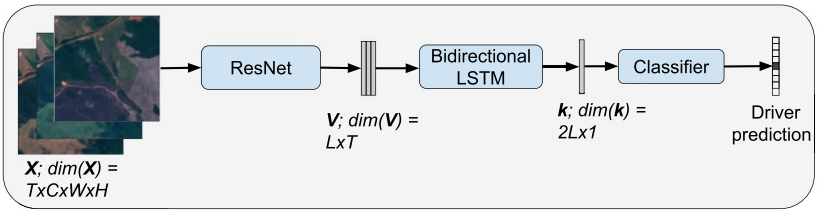
As shown in other studies [17, 37], concatenating images together is not the optimal way of handling time series. Including modules for extracting both visual and temporal features explicitly is important for effective learning from spatio-temporal data. Therefore, we employ LSTM for extracting temporal features in addition to the CNN. LSTM is a type of recurrent neural network (RNN) which is designed for handling sequential data. RNNs maintain a memory, called the hidden state, that is updated as the network processes the input sequence. At every step, the current item from the sequence (in our study, the vector v at time t) and the previous hidden state (i.e. the vector k_{t-1}) are used as inputs. This allows the model to keep information from previous steps and use it when processing the next elements in the sequence. In addition, LSTM also maintains another memory unit, the cell state, which improves the modeling of long-term dependencies.

A.1.3. CNN-LSTM

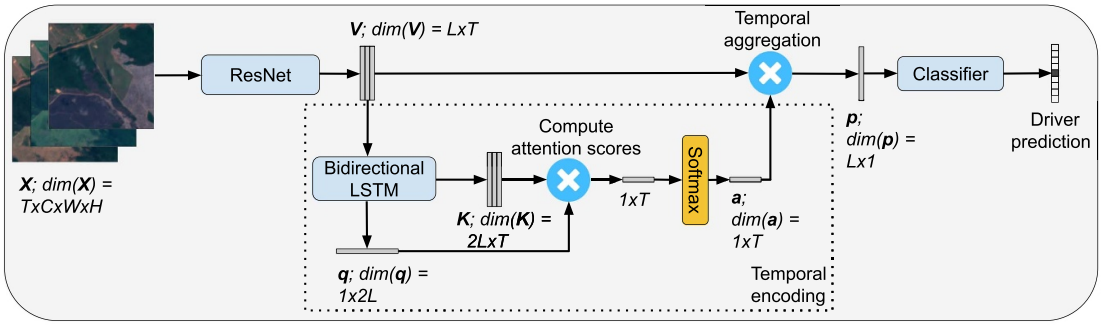
(b) The second baseline model '*CNN-LSTM*' is shown in figure A1. It extracts T feature vectors with the convolutional backbone, one vector per image. Maintaining the temporal order, these form a matrix $\mathbf{V}$ that is used as input to a bidirectional LSTM. A bidirectional LSTM refers to a module of two LSTM layers where each processes the input sequence in one direction, resulting in better representations of the sequence [38]. The last output of the LSTM in each direction is concatenated into a feature vector $\mathbf{k}$ of size $2L$, which is used by the classifier.



(a) Baseline (1): CNN. If the input is a single image, no stacking is done



(b) Baseline (2): CNN-LSTM



(c) Proposed: CNN-Attention-LSTM

Figure A1. Models for classification of forest loss drivers used in this work.

A.1.4. CNN-Attention-LSTM

The proposed model '*CNN-Attention-LSTM*' (figure A1(c)) adds an attention mechanism [27] inspired by the success of the Transformer [39]. The attention mechanism that we designed provides information about the importance of each image in the input time series to make a prediction. In the same way as '*CNN-LSTM*', the model extracts T visual feature vectors with the backbone, forming a matrix $\mathbf{V}$, which is passed into the bidirectional LSTM. All hidden states in both directions from the LSTM are used to form a matrix $\mathbf{K}$ of size $2L \times T$. The last cell states in each direction are concatenated together into a vector $\mathbf{q}$ of size $2L$.

Then, a vector of attention scores $\mathbf{a}$ is computed as

$$\mathbf{a} = \text{softmax}\left(\mathbf{q}^\top \mathbf{K}\right), \tag{A.1}$$

where $\top$ p denotes the transpose of $\mathbf{q}$. *Softmax* is used to normalize the attention vectors so that they all have the same scale and are comparable. Then, a single feature vector $\mathbf{p}$, which is passed to the classifier, is computed as a weighted average of the visual feature vectors, with attention scores used as weights:

$$\mathbf{p} = \mathbf{a}\mathbf{V}, \tag{A.2}$$

which constrains the model to learn a representation of the input time series that is a weighted average of individual images. We then analyzed the weights (and the images corresponding to them) to better understand what the important time steps are for recognizing individual drivers.

Appendix B. Data

B.1. Annotation

For acquiring the labels, we primarily used a large, publicly available reference dataset[5] collected using crowdsourcing via the Geo-Wiki platform[6] Around 150 K locations were randomly selected on land surfaces between 30°N and 30°S where any forest loss

[5] https://pure.iiasa.ac.at/id/eprint/17539/
[6] https://www.geo-wiki.org

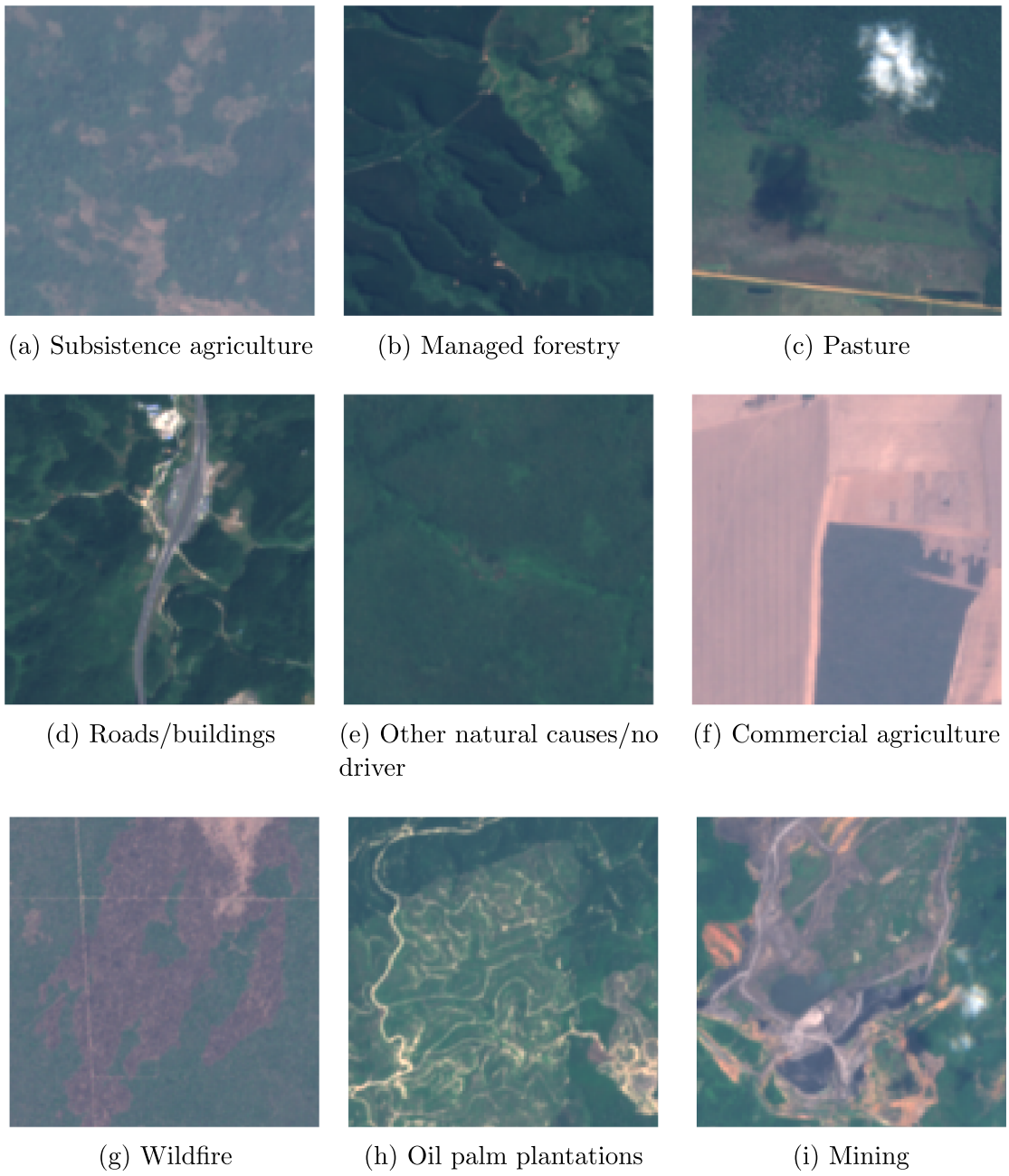
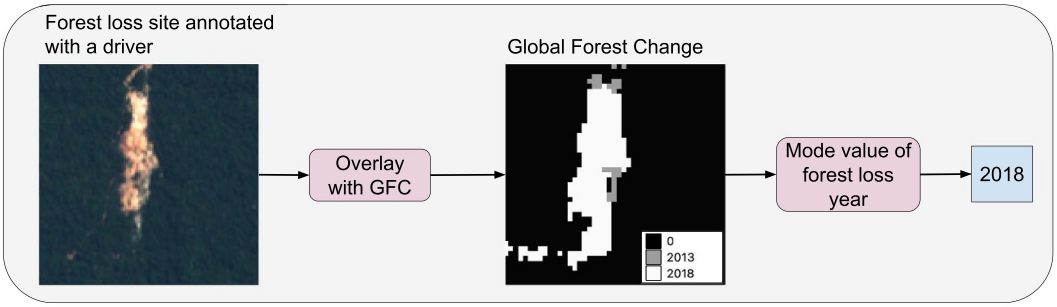


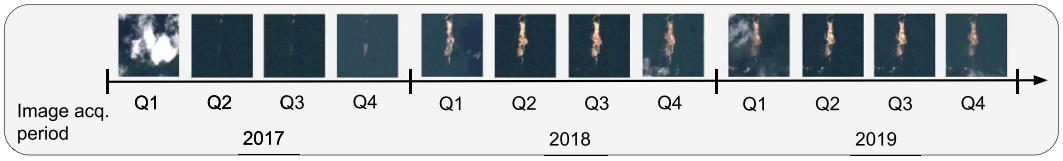
Figure B1. Examples of classes used in this work.

had occurred between 2008 and 2019 based on GFC [8]. These locations were then provided to Geo-Wiki, and data collection proceeded as a crowdsourcing campaign to label primary drivers of forest loss in each 1 km^2 area. Two other tasks were performed as part of the campaign (identifying secondary drivers and the presence of roads) but the corresponding data was not used in this study. A full description of the curation of this dataset is described in Bayas *et al* (2022) [40]. Examples of all classes are shown in figure B1.

There were nine primary drivers to choose from, which were partly based on those used by Curtis *et al* (2018) [4] and literature on drivers of deforestation [5, 41]. The first four are agriculture-based, i.e. subsistence agriculture, commercial or commodity-driven agriculture, oil palm as a separate commodity, and pasture. Oil palm and pasture are recognized separately from other commercial agriculture, as they are the two most widespread drivers of tropical deforestation [6]. As such, it is particurly important to be able to recognize them. The next three are other human activities—plantation forestry, roads/urban expansion and mining. The final two consider drivers that are both natural and human-induced, i.e. wildfires, and other natural causes. The last category also included no driver if it was not possible to attribute a driver to the forest loss based on



(a) For every annotated forest loss location, the year of forest loss is estimated as the most common GFC value (excluding 0 which corresponds to no forest loss)



(b) Sentinel-2 images are sampled around the estimated forest loss year, one image per quarter, resulting in 12 images for each site

Figure B2. TS acquisition and construction workflow.

visual interpretation. Multiple responses were collected at each location for quality assurance purposes. In our study, we disregarded data points where no driver had a majority of respondents votes (i.e. where there was little or no consensus between the responses).

As shown in figure B2(a), the year of forest loss was estimated for every 1 km^2 location using the GFC [8] which specifies the forest loss year (if any occurred) for every pixel at 30 m resolution. We estimated the year of forest loss as the mode value within the annotated location. We only used locations where we estimated the year of forest loss between 2017 and 2020. This is due to the availability of Sentinel-2 images from 2015 and the fact we used time series of lengths of up to 3 years. Note that in the crowdsourced campaign, the locations were annotated with respect to the period 2008-2019. We assumed that the primary driver remained constant and the annotations were valid for our target period.

The distribution of the dataset is shown in figure B3. The dataset obtained from the crowdsourced campaign was augmented with additional examples to improve the class balance because earlier work showed that the underrepresented classes were classified with significantly lower accuracy [37]. For this reason, we added examples of classes *mining* and *wildfire*, utilizing the data of Maus *et al* (2022) [42] and Tyukavina *et al* (2022) [34], respectively. Both datasets were overlaid with GFC to identify locations of mines or wildfires where forest loss occurred in the target period 2017–2020. All examples were manually verified. In total, 435 examples of *mining* and 323 examples of *wildfire* were added, so the total dataset contained 10'395 examples.

The dataset was split into train, validation, and test sets of sizes 9083, 136, and 1176, respectively. The class distributions for each of the three datasets are shown in table B1. To avoid bias resulting from spatial autocorrelation, the study area was divided by degrees of longitude. For every 5 degrees of longitude, examples located in the first three degrees were assigned to the training set, the fourth to the validation set, and the fifth to the test set. This is shown in figure B4.

B.1.1. Images

We used Sentinel-2 images as input features. Sentinel-2 is a constellation of two identical satellites that offers a revisit time of 5 days. The sensors have 13 spectral bands with a spatial resolution 10-60 meters. In our study, we used 4 bands (red, green, blue, and near-infrared) that have the highest spatial resolution of 10 meters. Experiments with more bands did not yield any improvement in accuracy while increasing the volume of data and computation. We used the Level-1 C product (top-of-atmosphere).

For each location, we downloaded a 3-year time series of Sentinel-2 images, one image per quarter (3 months). As shown in figure B2(b), we selected the time period to cover the estimated forest loss year, one year after and one year before. We selected the least cloudy image for each period. We did not perform any further cloud filtering as we designed our model architectures to learn to ignore cloudy images.

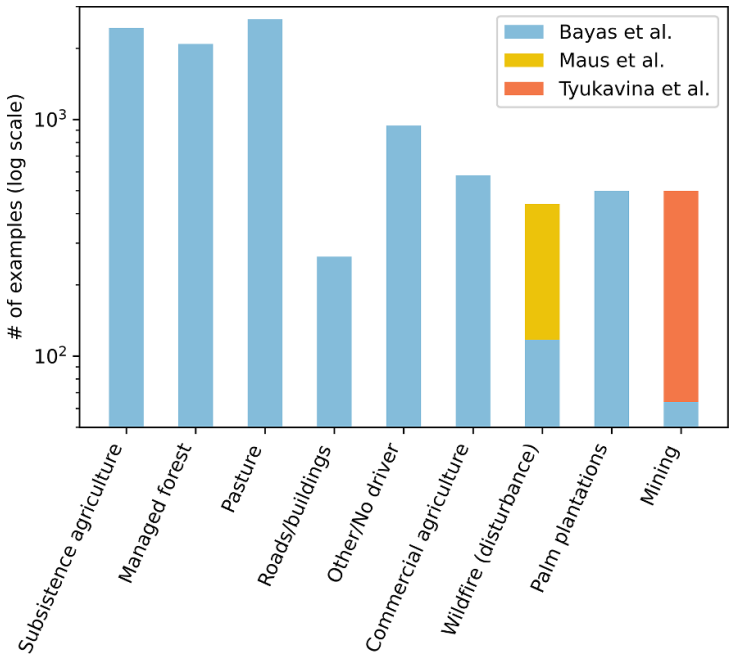


Figure B3. Dataset distribution.

Table B1. Class distributions for each of the training, validation, and test subsets of the dataset.

	Commercial agr.	Palm plantations	Managed forest	Mining	Other/no driver	Pasture	Roads/ buildings	Subsistence agr.	Wild-fire
Train	447	391	1891	359	803	2412	193	2235	350
Val.	21	11	11	12	11	14	22	16	18
Test	112	97	183	128	127	224	48	185	72
Total	580	499	2085	499	941	2650	263	2436	440

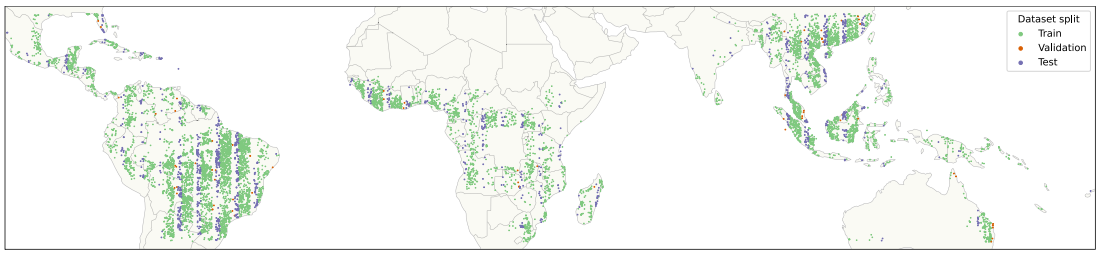


Figure B4. Spatial data split.

To predict forest loss drivers across all tropical forests, we also prepared an equivalent Sentinel-2 dataset by sampling forest loss areas detected by GFC between 2017 and 2020, which resulted in a total of 588k km^2 footprints. For these locations, the deforestation driver was unavailable and was predicted by our model.

Appendix C. Constructing time series

The time series used in this study were always formed by a set of consecutive images. For time series of less than 12 images, there are multiple options of how to construct them because we can shift the beginning (and accordingly the end) of the time series. We assumed that the most important period for recognizing drivers is the year in which forest loss took place, followed by the year *after*. We expected the images from the year *before* forest loss would be the least helpful for the model. The assumptions were based on visual inspection of the forest loss sites and literature on forest loss [5, 12, 40].

As shown in figure C1, we designed our experiments accordingly. When training models on a single

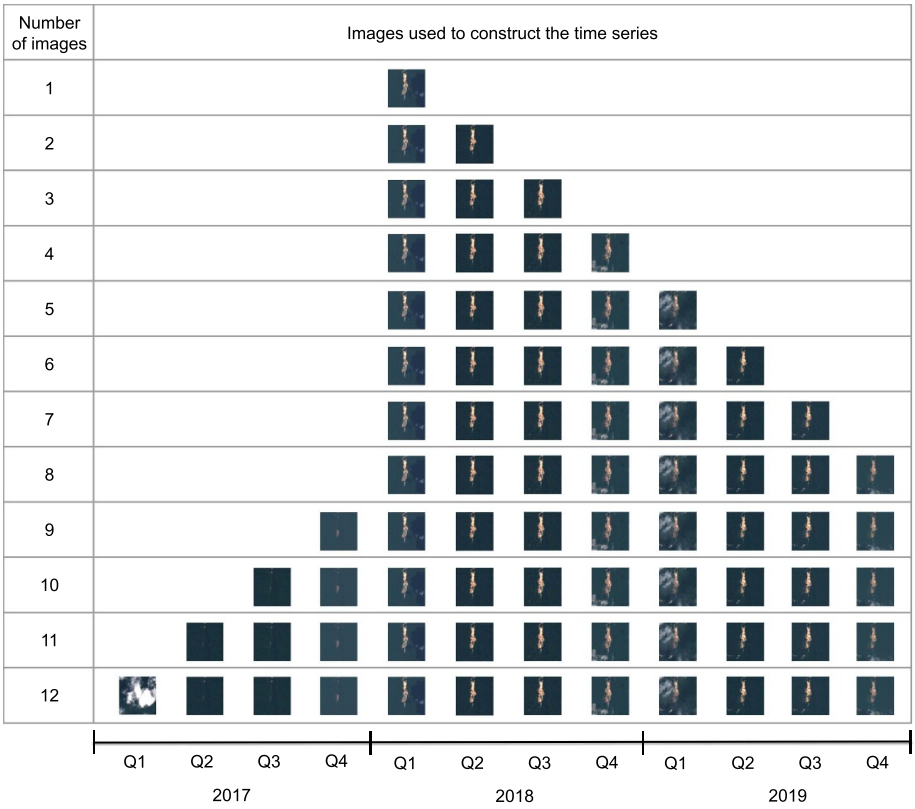


Figure C1. Time series sampling for an example forest loss site with estimated forest loss year 2018; images from one year before and one year after the forest loss year are acquired; when constructing time series of different lengths, images from the forest loss year are used first, followed by the year after it and lastly including images from the year before the forest loss event.

image, we used images from the first quarter of the estimated forest loss year. For constructing the time series, we continued to add images from the next quarter to a maximum of 8 images covering the forest loss in that year and the year after. For constructing longer time series, we added images from the year *before* the forest loss, starting with the latest image (i.e. sampled in the last quarter of the year).

Appendix D. Experimental setup

We used ResNet34 pre-trained on ImageNet [43]. The pre-trained weights only have three input channels while we used four. Therefore, we augmented the first convolutional layer by duplicating the weights corresponding to the first channel, resulting in a shape matching our input data. When stacking multiple images together for the multi-temporal baseline, we duplicated the weights to match the input data shape. We have also tried to use weights pre-trained on satellite images but did not obtain satisfactory results.

All experiments were executed three times with varying seed values and the results were averaged across each set of three experiments. All models were trained with the cross entropy loss function and Adam optimizer. We used early stopping to stop training after the validation loss did not improve for 20 epochs. During training, we randomly flipped images as a regularization technique to reduce overfitting. We used the F1 score as an evaluation metric as it better captures the model's performance when working with imbalanced datasets.

Appendix E. Attention score distribution

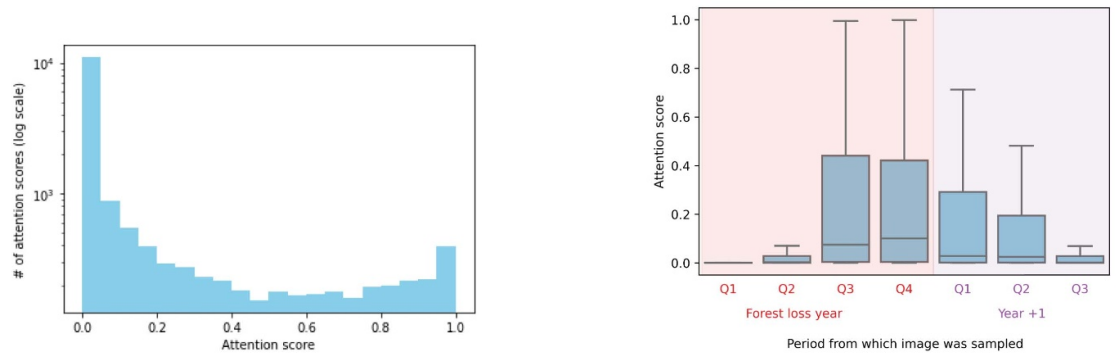
The overall distribution of the attention scores over the test set is shown in figure E1(a). The large majority of images are not used to make a prediction (i.e. most attention scores are 0). We believe this is because there is often relatively little change between multiple image acquisitions, resulting in redundant data.

E.1. Time series entropy of attention scores

To better understand how many images from a time series the model typically uses and how that differs between classes, we calculated the Shannon entropy [44] for every set of attention scores corresponding to a time series:

$$H = \sum (\mathbf{a} * \log(\mathbf{a})), \tag{E.1}$$

where **a** corresponds to the set of attention scores the model produced for a time series (equation (A.1)). If the model entirely attends to a single image only, this corresponds to entropy of 0. The maximum entropy is reached if the model attends to all images to the same



(a) Histogram of attention scores (b) Temporal attention scores

Figure E1. Temporal distribution of attention scores.

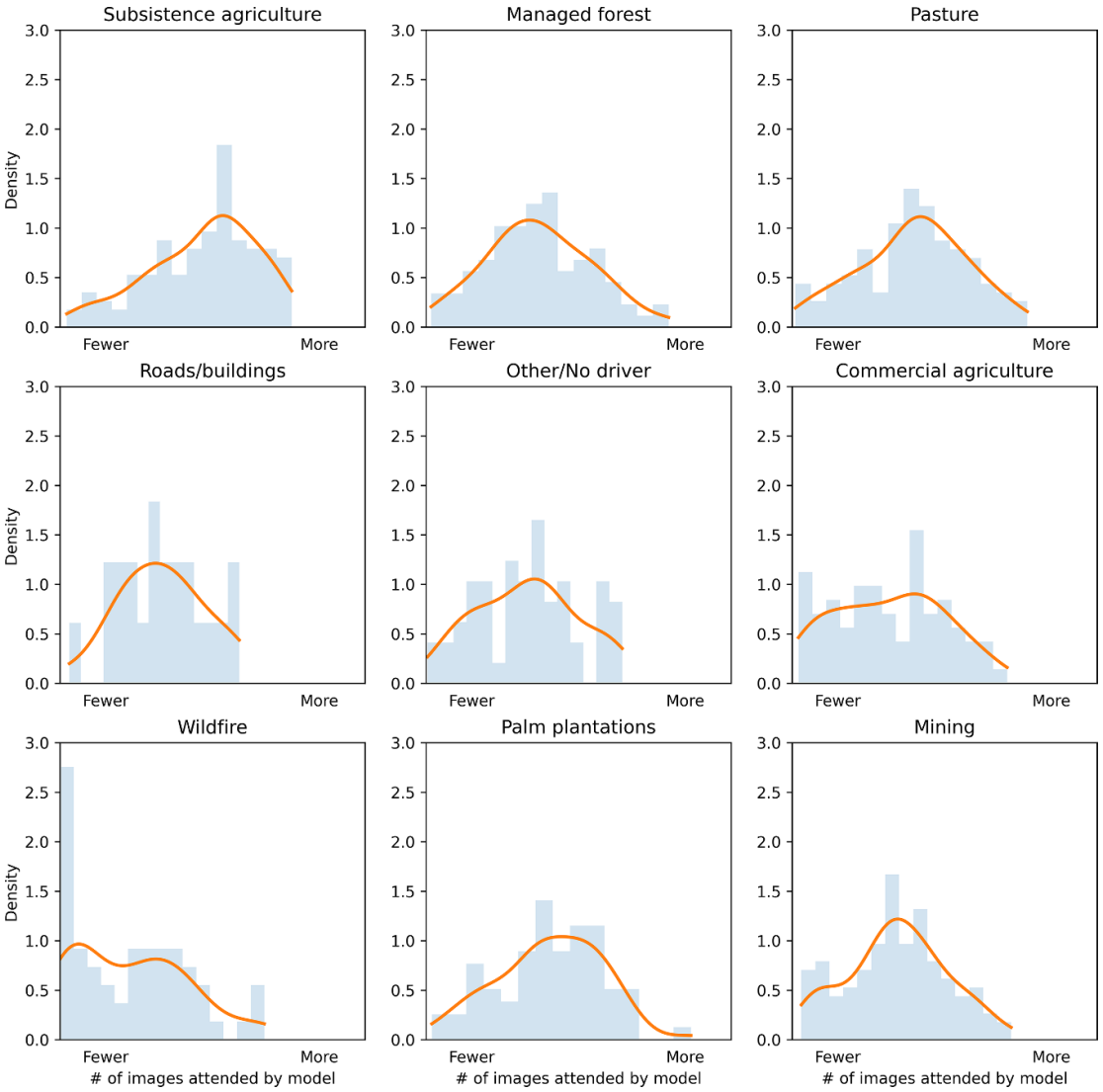


Figure E2. Entropy per class with a fitted KDE; lower entropy values correspond to the model attending to fewer images from the input time series; higher entropy values signal more images used.

extent (i.e. all attention scores are equal). We calculated the entropy for each time series over the test set and fitted a kernel density estimator (KDE) to the resulting density for easier interpretation.

The results are shown in figure E2. For most classes, the distribution follows a skewed normal distribution. This means that most of the time, the model attends to multiple, but not all, images in the

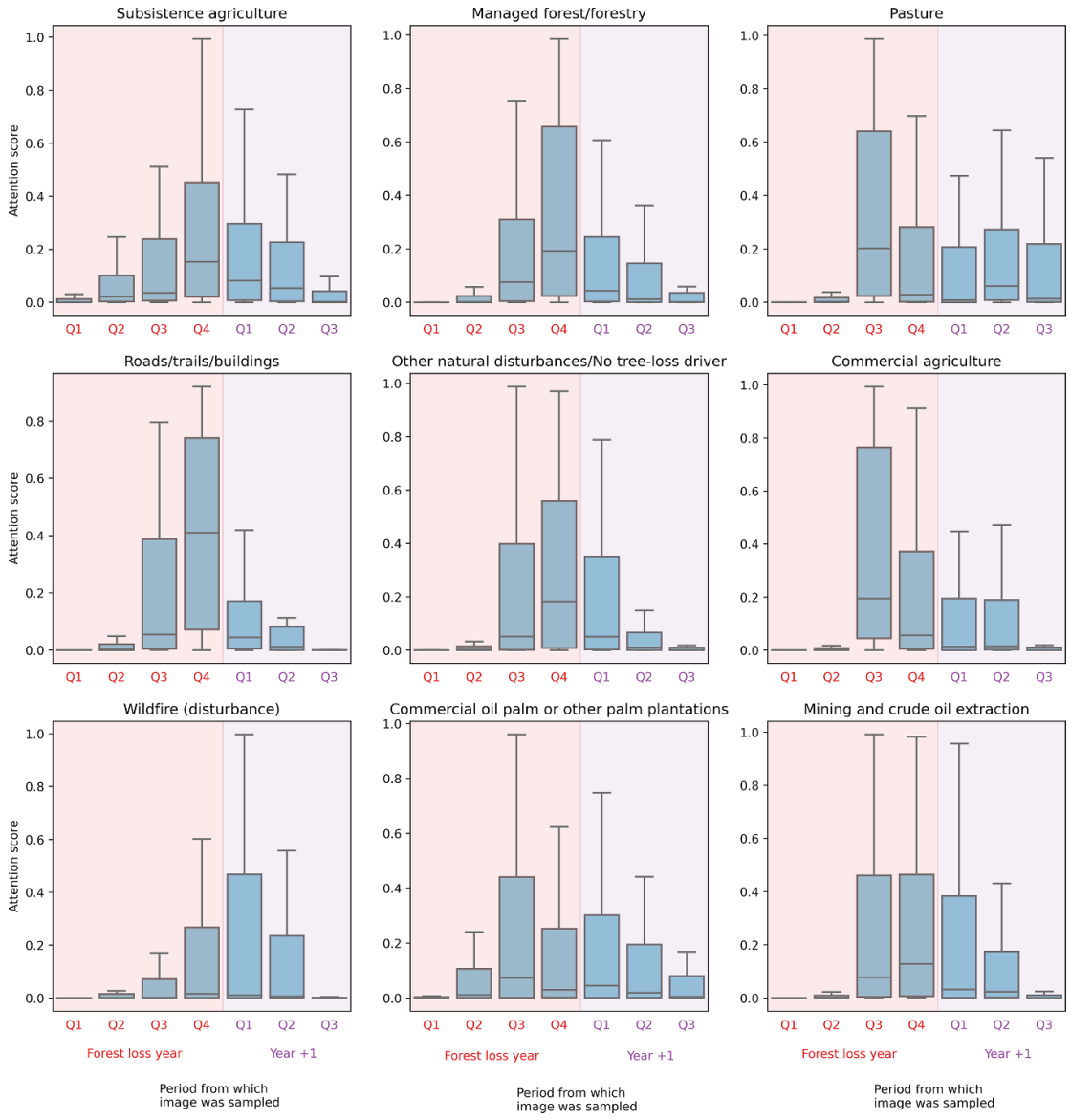


Figure E3. Temporal attention scores per class.

time series. The classes related to agriculture tend to be more negatively skewed, i.e. it is more common for the model to attend to multiple images. We expect this to be caused by seasonal patterns associated with these classes.

There is a distinctly different shape associated with the class *wildfire*. Here, the most common case is a very low entropy value, corresponding to the model attending mostly to a single image to make a prediction. This is consistent with the example in figure 4(b), where the model recognizes wildfire almost entirely based on a single image that shows distinct burned scars. We believe this may explain the model's relatively low performance with respect to this class. The model learns to extract features that are most useful for the overall performance. If the features required for recognizing *wildfire* are very different from all other classes, the model might not learn them well. Also, if the distinct features are present on a single or a few images, the detection of *wildfire* may be more sensitive to cloud cover. It is worth noting that there is a wide range of techniques for detecting wildfire from remotely sensed imagery, such as using the Normalized Burn Ratio index. While such methods may have improved the results of our model for this particular class, we did not use it because the aim of this work is to train a model in a data-driven way, without hand-crafting features for individual drivers.

E.1.1. Attention score distribution in time

Figure E1(b) shows the distribution of attention scores over time for the best-performing model which was trained on time series of 7 images. The model

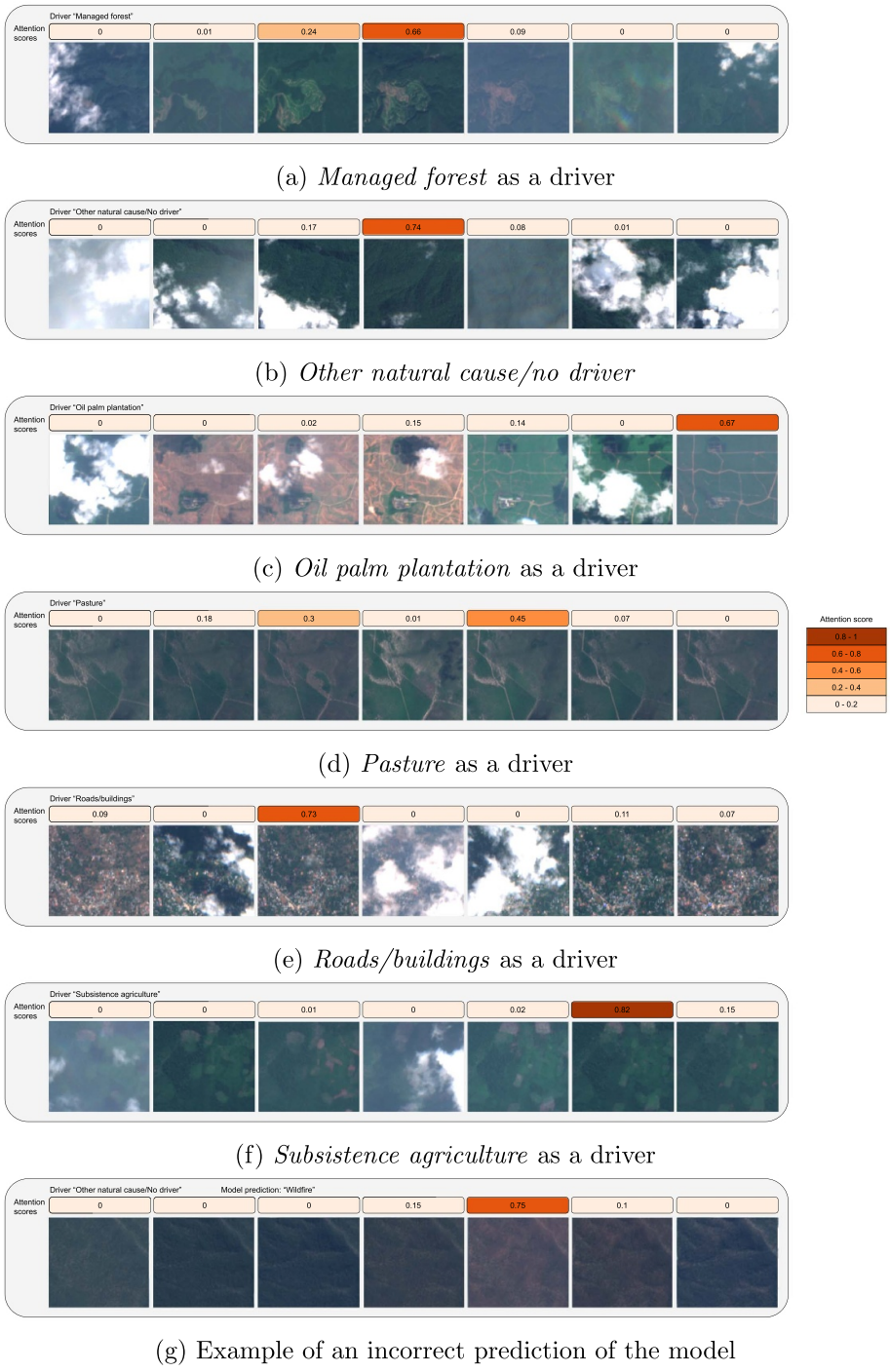


Figure E4. Examples of attention scores the model produced for different time series, complementing those reported in figure 4.

learns to mostly attend to images in the second half of the estimated forest loss year, followed by the first half of the subsequent year. This mirrors the evolution of the F1 score per increasing number of images and confirms that the model attends mostly to the period immediately after the event.

Figure E3 shows this disaggregated by class. We can see that most classes follow the overall trend and the model attends most to images in that period. In contrast, to recognize *pasture*, the model attends to later images as well. We believe this is because shortly after the forest loss event, agriculture-related classes may have a similar appearance to *pasture*. Later, *pasture* may become recognizable as the cleared area does not show any signs of crops or trees being planted.

Figure E4 contains more qualitative examples of input time series and the attention scores produced by the model.

ORCID iD

Linda See https://orcid.org/0000-0002-2665-7065

References

[1] Gibson L *et al* 2011 Primary forests are irreplaceable for sustaining tropical biodiversity *Nature* **478** 378–81
[2] Goodman R and Herold M 2014 Why maintaining tropical forests is essential and urgent for a stable climate *Center for Global Development Working paper No. 385*
[3] Wright S J 2010 The future of tropical forests *Ann. New York Acad. Sci.* **1195** 1–27
[4] Curtis P G, Slay C M, Harris N L, Tyukavina A and Hansen M C 2018 Classifying drivers of global forest loss *Science* **361** 1108–11
[5] Seymour F and Harris N L 2019 Reducing tropical deforestation *Science* **365** 756–7
[6] Goldman E, Weisse M, Harris N, and Schneider M 2020 Estimating the role of seven commodities in agriculture-linked deforestation: Oil palm, soy, cattle, wood fiber, cocoa, coffee, and rubber
[7] Heilmayr R, Rausch L L, Munger J and Gibbs H K 2020 Brazil's Amazon Soy Moratorium reduced deforestation *Nat. Food* **1** 801–10
[8] Hansen M C *et al* 2013 High-resolution global maps of 21st-century forest cover change *Science* **342** 850–3
[9] Vancutsem C, Achard F, Pekel J F, Vieilledent G, Carboni S, Simonetti D, Gallego J, Aragao L E and Nasi R 2021 Long-term (1990–2019) monitoring of forest cover changes in the humid tropics *Sci. Adv.* **7** eabe1603
[10] Hosonuma N, Herold M, De Sy V, De Fries R S, Brockhaus M, Verchot L, Angelsen A and Romijn E 2012 An assessment of deforestation and forest degradation drivers in developing countries *Environ. Res. Lett.* **7** 044009
[11] Pendrill F, Persson U M, Godar J, Kastner T, Moran D, Schmidt S and Wood R 2019 Agricultural and forestry trade drives large share of tropical deforestation emissions *Glob. Environ. Change* **56** 1–10
[12] Pendrill F *et al* 2022 Disentangling the numbers behind agriculture-driven tropical deforestation *Science* **377** eabm9267
[13] De Sy V, Herold M, Achard F, Beuchle R, Clevers J, Lindquist E and Verchot L 2015 Land use patterns and related carbon losses following deforestation in south america *Environ. Res. Lett.* **10** 124004
[14] Wijaya A, Sugardiman Budiharto R, Tosiani A, Murdiyarso D and Verchot L V 2015 Assessment of large scale land cover change classifications and drivers of deforestation in indonesia *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **40** 557–62
[15] Austin K G, Schwantes A, Gu Y and Kasibhatla P S 2019 What causes deforestation in Indonesia? *Environ. Res. Lett.* **14** 024007
[16] Meijaard E and Gaveau D 2021 Global oil palm map
[17] Masolele R N, De Sy V, Herold M, Marcos Gonzalez D, Verbesselt J, Gieseke F, Mullissa A G and Martius C 2021 Spatial and temporal deep learning methods for deriving land-use following deforestation: a pan-tropical case study using Landsat time series *Remote Sens. Environ.* **264** 112600
[18] Irvin J A, Sheng H, Ramachandran N, Johnson-Yu S, Zhou S, Story K, Rustowicz R, Elsworth C, Austin K and Ng A 2020 Forestnet: Classifying drivers of deforestation in indonesia using deep learning on satellite imagery *NeurIPS Workshop on Tackling Climate Change With Machine Learning*
[19] Mitton J and Murray-Smith R 2021 Rotation equivariant deforestation segmentation and driver classification *NeurIPS Workshop on Tackling Climate Change With Machine Learning*
[20] Kolesnikov A *et al* 2021 An image is worth 16 × 16 words: transformers for image recognition at scale *Int. Conf. on Learning Representations*
[21] Kaselimi M, Voulodimos A, Daskalopoulos I, Doulamis N and Doulamis A 2022 A vision transformer model for convolution-free multilabel classification of satellite imagery in deforestation monitoring *IEEE Trans. on Neural Networks and Learning Systems*
[22] Li Y, Zhang H and Shen Q 2017 Spectral–spatial classification of hyperspectral imagery with 3D convolutional neural network *Remote Sens.* **9** 67
[23] Pelletier C, Webb G I and Petitjean F 2019 Temporal convolutional neural network for the classification of satellite image time series *Remote Sens.* **11** 523
[24] Shi X, Chen Z, Wang H, Yeung D Y, Wong W K and Woo W c 2015 Convolutional lstm network: a machine learning approach for precipitation nowcasting *Advances in Neural Information Processing Systems* vol 28
[25] Benedetti P, Ienco D, Gaetano R, Ose K, Pensa R G and Dupuy S 2018 M3f usion: a deep learning architecture for multiscale multimodal multitemporal satellite data fusion *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **11** 4939–49
[26] Interdonato R, Ienco D, Gaetano R and Ose K 2019 Duplo: A dual view point deep learning architecture for time series classification *ISPRS J. Photogramm. Remote Sens.* **149** 91–104
[27] Bahdanau D, Cho K and Bengio Y 2014 Neural machine translation by jointly learning to align and translate *Int. Conf. on Learning Representations*
[28] Rußwurm M and Körner M 2020 Self-attention for raw optical satellite time series classification *ISPRS J. Photogramm. Remote Sens.* **169** 421–35
[29] Garnot V S F, Landrieu L, Giordano S and Chehata N 2020 Satellite image time series classification with pixel-set encoders and temporal self-attention *IEEE Conf. on Computer Vision and Pattern Recognition*
[30] Goldenberg B *et al* 2017 Planet: understanding the amazon from space
[31] De Sy V, Herold M, Achard F, Avitabile V, Baccini A, Carter S, Clevers J G, Lindquist E, Pereira M and Verchot L 2019 Tropical deforestation drivers and associated carbon emission factors derived from remote sensing data *Environ. Res. Lett.* **14** 094022
[32] Fritz S *et al* 2022 A continental assessment of the drivers of tropical deforestation with a focus on protected areas *Front. Conserv. Sci.* **3** 830248
[33] Song X P *et al* 2021 Massive soybean expansion in south america since 2000 and implications for conservation *Nat. Sustain.* **4** 784–92
[34] Tyukavina A *et al* 2022 Global trends of forest loss due to fire from 2001 to 2019 *Front. Remote Sens.* **3** 825190
[35] Hochreiter S and Schmidhuber J 1997 Long short-term memory *Neural Comput.* **9** 1735–80
[36] He K, Zhang X, Ren S and Sun J 2016 Deep residual learning for image recognition *IEEE Conf. on Computer Vision and Pattern Recognition* pp 770–8
[37] Pisl J, Hughes L H, Rußwurm M and Tuia D 2023 Classification of tropical deforestation drivers with machine learning and satellite image time series *IEEE Int. Geoscience and Remote Sensing Symp.*
[38] Graves A and Schmidhuber J 2005 Framewise phoneme classification with bidirectional lstm and other neural network architectures *Neural Netw.* **18** 602–10
[39] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez A N, Kaiser Ł and Polosukhin I 2017 Attention is all you need *Advances in Neural Information Processing Systems* vol **30**
[40] Laso Bayas J C *et al* 2022 Drivers of tropical forest loss between 2008 and 2019 *Sci. Data* **9** 146
[41] Geist H J and Lambin E F 2002 Proximate causes and underlying driving forces of tropical deforestation *BioScience* **52** 143
[42] Maus V, Silva D d, Gutschlhofer J, Rosa R D, Giljum S, Gass S L B, Luckeneder S, Lieber M and McCallum I 2022 Global-scale mining polygons (version 2) *Creative Commons Attribution-Sharealike 4.0 International*
[43] Deng J, Dong W, Socher R, Li L J, Li K and Fei-Fei L 2009 Imagenet: a large-scale hierarchical image database *IEEE Conf. on Computer Vision and Pattern Recognition* pp 248–255
[44] Shannon C E, 1948 A mathematical theory of communication *Bell Syst. Tech. J.* **27** 379–423