

RESEARCH ARTICLE

Using deep convolutional neural networks to forecast spatial patterns of Amazonian deforestation

James G. C. Ball¹  | Katerina Petrova² | David A. Coomes¹  | Seth Flaxman³ ¹Department of Plant Sciences and Conservation Research Institute, University of Cambridge, Cambridge, UK²Department of Mathematics, Imperial College London, London, UK³Department of Computer Science, University of Oxford, Oxford, UK

Correspondence

James G. C. Ball

Email: ball.jgc@gmail.com

Funding information

Engineering and Physical Sciences Research Council, Grant/Award Number: EP/V002910/1; Natural Environment Research Council, Grant/Award Number: PDAG/501

Handling Editor: Hooman Latifi

Abstract

1. Tropical forests are subject to diverse deforestation pressures while their conservation is essential to achieve global climate goals. Predicting the location of deforestation is challenging due to the complexity of the natural and human systems involved but accurate and timely forecasts could enable effective planning and on-the-ground enforcement practices to curb deforestation rates. New computer vision technologies based on deep learning can be applied to the increasing volume of Earth observation data to generate novel insights and make predictions with unprecedented accuracy.
2. Here, we demonstrate the ability of deep convolutional neural networks (CNNs) to learn spatiotemporal patterns of deforestation from a limited set of freely available global data layers, including multispectral satellite imagery, the Hansen maps of annual forest change (2001–2020) and the ALOS PALSAR digital surface model, to forecast deforestation (2021). We designed four model architectures, based on 2D CNNs, 3D CNNs, and Convolutional Long Short-Term Memory (ConvLSTM) Recurrent Neural Networks (RNNs), to produce spatial maps that indicate the risk to each forested pixel (~30m) in the landscape of becoming deforested within the next year. They were trained and tested on data from two ~80,000 km² tropical forest regions in the Southern Peruvian Amazon.
3. The networks could predict the location of future forest loss to a high degree of accuracy ($F_1 = 0.58$ – 0.71). Our best performing model (3D CNN) had the highest pixel-wise accuracy ($F_1 = 0.71$) when validated on 2020 forest loss (2014–2019 training). Visual interpretation of the mapped forecasts indicated that the network could automatically discern the drivers of forest loss from the input data. For example, pixels around new access routes (e.g. roads) were assigned high risk, whereas this was not the case for recent, concentrated natural loss events (e.g. remote landslides).
4. Convolutional neural networks can harness limited time-series data to predict near-future deforestation patterns, an important step in harnessing the growing volume of satellite remote sensing data to curb global deforestation. The

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2022 The Authors. *Methods in Ecology and Evolution* published by John Wiley & Sons Ltd on behalf of British Ecological Society.

modelling framework can be readily applied to any tropical forest location and used by governments and conservation organisations to prevent deforestation and plan protected areas.

KEYWORDS

Amazon, artificial intelligence, convolutional neural networks, deep learning, deforestation forecasting, machine learning, spatial forecasting, tropical forests

1 | INTRODUCTION

To achieve the pledge made by world leaders at COP26 to end deforestation by 2030, innovative approaches to forest protection are urgently required. Previous efforts to curb tropical primary forest loss have failed to have a net positive impact (Potapov et al., 2017) despite countries and companies making substantial commitments (e.g. the New York Declaration of Forests in 2014). In regions of the Amazon deforestation is surging (Beuchle et al., 2021) which risk inducing the biome to undergo a critical transition with profound consequences for climate and biodiversity (Lovejoy & Nobre, 2019). Although the ultimate causes of deforestation need to be addressed—notably global demand for agricultural and wood products—interventions to tackle proximate causes, such as illegal logging and mining, are also necessary (Asner & Tupayachi, 2016; Finer et al., 2014). One measure identified as effective in reducing deforestation is ‘targeting protected areas to regions where forests face higher threat’ (Busch & Ferretti-Gallon, 2017). However, knowing how to optimally allocate limited resources to tackle diverse threats across vast, difficult to access tropical forest landscapes is challenging. To enable effective, on-the-ground prevention measures, up-to-date information on the location, relative severity and likely evolution of threats is required.

Governmental and NGO commitments to make timely interventions have led to products that give near-real-time alerts on the location of deforestation events (Hansen et al., 2016; Reiche et al., 2021), but access to these alerts has had little material benefit in terms of curbing deforestation (Moffette et al., 2021). One issue has been that responses based on near-real-time mapping can only ever be reactionary. As a result, there have been calls to develop early warning systems that inform decision-makers of the location of near-term deforestation risk (e.g. WWF, 2020). By developing innovative solutions that harness the growing volume of remote sensing (RS) data, interventions stand a greater chance of preventing deforestation. In the longer term, cost effective conservation plans should not only consider the spatial distribution of conservation features (e.g. species, carbon stocks) and the financial costs, but also account for how threats are likely to evolve (Wilson et al., 2007).

Deforestation is difficult to predict as it results from complex interactions within human-ecological systems but characteristic drivers—such as agricultural expansion and ease of access—have been identified (Geist & Lambin, 2002). Drivers can be mapped to spatially resolved geographical, economic, social and biological variables (e.g. agricultural land value, distance to roads). Within a statistical or machine learning (ML) framework, these can be correlated to

the likelihood of deforestation at a given location typically, by linking the spatial predictor layers to changes in forest extent over a single time step of several years (see e.g. Cushman et al., 2017; Mayfield et al., 2017; Saha et al., 2020 for intercomparisons of different ML frameworks). However, Rosa et al. (2014)'s review of such approaches concluded that they had a limited or poorly defined ability to forecast locations of deforestation, a key reason being that spatially explicit datasets on the drivers of deforestation are often incomplete, out-of-date or unreliable in tropical forest regions. For example, proximity of roads is widely accepted to be a strong predictor of future deforestation (Barber et al., 2014) but roads in tropical forest landscapes are often unofficial, illegal or not comprehensively mapped (Perz et al., 2007), and can be highly dynamic; appearing, expanding and changing course to access new resources (Ahmed et al., 2014). Most studies that have included roads as predictors used static road maps which is inappropriate in fast changing contexts, while others have used highly speculative methods to predict the development of roads in tropical forest landscapes (Ahmed et al., 2014; Mena et al., 2017). With many of the decisions that lead to road development in tropical forest regions likely to remain hidden from public view, a fundamentally new approach to making predictions of the spread of deforestation is required. Analogous issues exist for other landscape features that help predict deforestation including the location/spread of human settlements and agricultural land.

The ML frameworks used to date have been limited in how they represent local context. Without the ability to retain the spatial structures of the data, pixels are represented in isolation or local context is represented as averaged neighbourhood metrics. Contrasting approaches that identify threatened regions based on dynamic data (i.e. resolved over numerous time steps) have tended to focus simply on spatial and temporal trends in intensity of deforestation (e.g. the emerging hot spot approach of Harris et al., 2017). Although these approaches are easy to implement and update, they are naïve to the drivers of the observed forest loss and therefore unable to reliably predict whether a deforestation front is likely to spread. This risks inefficient targeting of conservation resources. A forecasting approach that could automatically learn and detect the contextual features of the landscape that signal deforestation risk from satellite imagery, and update its predictions accordingly, would bypass the limitations of both existing ML and simple trend-based approaches.

Through implementing modern computer vision approaches, it may be possible for an artificial neural network to automatically learn what features of a landscape are indicative of future deforestation from historic events and thereby avoid the need to rely on problematic

data layers. Deforestation often exhibits striking spatiotemporal patterns when viewed from space (e.g. fishbone pattern; de Filho & Metzger, 2006). The emergence of these patterns suggests a degree of predictability while their characteristic spatial and spectral structures can give clues as to the drivers. Over recent years, the quantity of freely available, high-resolution imagery has been increasing at a rate that has exceeded our ability to harness it to support decision-making. However, deep convolutional neural networks (CNNs), that can retain spatial structure and be trained to automatically extract features of images, have revolutionised the field of computer vision (Voulodimos et al., 2018). They have led to radical advances in our ability to make predictions in countless fields and have recently been taken up by RS researchers (Zhu et al., 2017) and ecologists working with Earth observation data (Brodrick et al., 2019) to generate novel insights. This includes work on detecting logging trails (Abdi et al., 2022), human settlements (Corbane et al., 2021), deforestation (Torres et al., 2021), forest disturbance (Kislov et al., 2021) and quantifying the properties of terrestrial vegetation (Kattenborn et al., 2021). Importantly, however, these studies have focused on assessing the current state of systems rather than predicting future states.

Forecasting deforestation is a temporal extension of the land-cover classification task. 2D CNNs (designed for working with image data; Kussul et al., 2017) and networks that incorporate an additional dimension (e.g. time) implicitly in their characterisation of data, including 3D CNNs (Li et al., 2017) and recurrent convolutional neural networks (ConvRNNs; Interdonato et al., 2019), have been shown to have state-of-the-art accuracy in detecting and classifying landcover change. Similarly, Convolutional Long Short-Term Memory (ConvLSTM)-based RNNs have been shown to improve on the performance of ConvRNN for RS classification tasks (Rußwurm & Körner, 2018). These promising developments suggest that 2D CNNs, 3D CNNs and ConvRNNs architecture types, if trained on large volumes of spatially resolved historic deforestation data, could

characterise and learn the features of the imagery associated with changing forest cover and be used to predict the future location of deforestation (something that has not previously been attempted). Critically, they preserve and work with the spatial (and temporal) structures of the input datasets and therefore make predictions based on the local context of the scenes they are presented with.

This article describes a range of deep CNN model architectures (MAs) that we designed to predict the risk of forest loss at a given pixel (~30m) when presented with a multilayered scene (of data from freely available, global datasets) centred on that pixel, with the aim of automatically identifying areas threatened by deforestation. To test these networks, we trained and evaluated their performance on two regions in the Peruvian Amazon, one of the most biodiverse regions on the planet and exceptionally rich in endemic amphibians, birds, fishes, bats, and trees (Bass et al., 2010; Jenkins et al., 2013) but facing a range of threats including deforestation driven by copper and gold mining, logging, agriculture, cattle ranching and crude oil extraction (Finer et al., 2014, 2015; Piotrowski, 2019). We trained and tested the models with cloud free satellite data for 2014–2020 and spatially resolved annual deforestation data for 2001–2020 (Hansen et al., 2013). The best performing model was used to forecast deforestation in across regions in 2021, predicting the risk of deforestation for every forested pixel in the landscape.

2 | MATERIALS AND METHODS

2.1 | Study regions

This study focused on two regions in the south of Peru (Figure 1). The size of the study regions were suitable as they provided many millions of training points while allowing processing to remain within the computing resources available to us for reasonable experimentation.

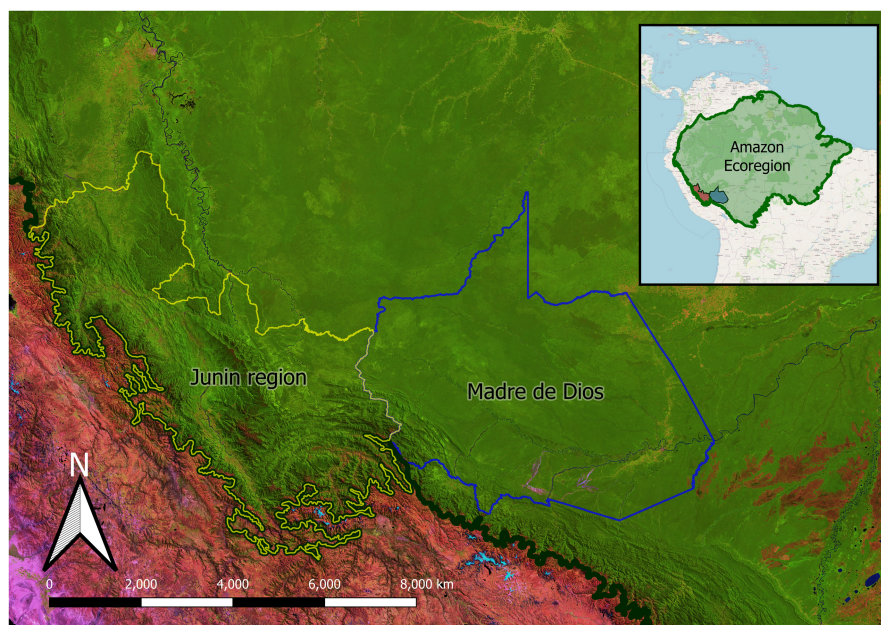


FIGURE 1 Southern Peruvian study regions. Base image: Landsat composite (bands 5,4,3) 2019 (Hansen/UMD/Google/USGS/NASA).

A preliminary broad intercomparison of a broad set of models focused on the Madre de Dios (MdD) department of Peru. MdD has approximately 97.4% primary forest cover and primary forest loss has steadily increased since 2000 (Hansen et al., 2013; Turubanova et al., 2018). It is almost entirely intact Southwest Amazon moist forest (Olson et al., 2001; Potapov et al., 2017). The Interoceanic Highway—a 2575-km road from the coast of Peru to Brazil—was completed in 2011, enabling access to portions of MdD previously protected from human activity. This increased access supported immigration that has increased pressure on the forest. Legal and illegal gold mining is prevalent, seeding remote deforestation frontiers in otherwise intact forest (Nicolau et al., 2019). The region exports a range of agricultural products including cotton, coffee, sugarcane, cocoa beans, Brazil nuts and palm oil. Large logging concessions and illegal selective logging (particularly for mahogany) place additional pressure on the forest there (Chirinos & Ruiz, 2003). The dynamic and diverse pressures made it a compelling trail region.

The most promising models were further developed and additionally applied to a second region. It centred on the Junín department, included surrounding departments and was cropped by the limit of the Amazon Ecoregion. It has approximately 88% forest cover, primarily split between Southwest Amazon moist forest and Peruvian Yungas (subtropical cloud forest) with some Ucayali moist forest (lowland to premontane moist forests; Olson et al., 2001). While most of its forest is considered primary there are some highly fragmented regions (Potapov et al., 2017; Turubanova et al., 2018). Drivers of deforestation are similar to MdD but with less mining and more hydrocarbon exploration (Finer et al., 2014, 2015). It gave the opportunity to trail the networks on a similarly sized area but one with greater environmental heterogeneity.

2.2 | Datasets

We limited ourselves to global, freely available datasets so the approach could be replicated and scaled to any forest location. The

forest state labels used in this study were taken from the Global Forest Change dataset (Hansen et al., 2013). It tracks the location of forest loss globally and is collated on an annual basis (latest version: 2001–2020) at ~30m resolution. Loss events are based on a pixel's forest cover (here >30% to ~0%) and are neutral to the cause of transition. Management activities, unauthorised resource extraction and loss due to other biotic or abiotic factors are not differentiated, thus observed patterns of loss relate to all possible causes.

The dataset includes a data mask describing areas of land or permanent water which we used to filter valid pixels. It also includes percentage tree cover observed in 2000 which we included as a predictor layer (for information to on the forest density) and to filter out pixels with low cover (<30%). From the same source, cloud-free composite Landsat imagery of four spectral bands was available for 2014 to 2020. Near infrared (NIR) and short-wave infrared (SWIR) bands are informative for the RS of vegetation. These were included allow the computer vision networks to learn features from the spectral signals. We included the Japan Aerospace Exploration Agency's (JAXA) 30m resolution ALOS Global DSM as a layer to give another dimension for the networks to learn from and exploit the features of the topology. It is an L-band SAR derived product which has better canopy penetration than C-band products and so gives a clearer signal of topology. Historic forest change labels were processed into four one-hot encoded layers, which assigned pixels to classes based on how recently (from current time t) they lost forest cover: (i) 0–1 years from t , (ii) 2–4, years from t , (iii) 5–8 years from t or (iv) more than 8 years from t . A pixel that does not register in these layers was free from loss since 2000. These layers encode the proximity, in time and space, of recent forest loss and thereby assist the models in representing contagion effects. All layers (see Table 1) were normalised to take values between 0 and 1 before being fed into the models.

Although additional layers (e.g. distance to roads, settlements) could have been included and may have improved overall accuracy, we wanted to test the ability of the deep CNNs to automatically extract useful features from a set of regularly updated, scalable RS

Variable	Description	Value
datamask	Land surface (0) or permanent water body (1)	0, 1
treecover2000	Percentage tree cover observed in 2000	0–100
elevation ^a	Height above sea level in meters	0 - max(h)
recent_loss1(t)	If forest to non-forest in years [t , $t - 2$)	0, 1
recent_loss2(t)	If forest to non-forest in years [$t - 2$, $t - 5$)	0, 1
recent_loss3(t)	If forest to non-forests [$t - 5$, $t - 8$)	0, 1
recent_loss4(t)	If forest to non-forests [$t - 8$, 2000)	0, 1
last_b30(t) ^b	Normalised TOA reflectance Landsat 7 band 3 (red)	0–255
last_b40(t) ^b	Normalised TAO reflectance Landsat 7 band 4 (NIR)	0–255
last_b50(t) ^b	Normalised TAO reflectance Landsat 7 band 5 (SWIR)	0–255
last_b70(t) ^b	Normalised TAO reflectance Landsat 7 band 7 (SWIR)	0–255

^aIncluded as an optional data layer but not included in final models.

^bLatest available cloud-free observation at pixel.

TABLE 1 Predictor layers. The first three layers are static and the rest change year-on-year (t). All are ~30m. abbreviations: Top-of-atmosphere (TOA), near infrared (NIR) and short-wave infrared (SWIR)

data layers. This tests their utility in poorly mapped or highly dynamic regions. Additional details of the datasets are given in S2.

2.3 | Modelling process

The aim was to train a classifier that, when presented with a scene centred on a single forested pixel, could predict whether the target central pixel would remain forested (0) or transition to a non-forest state (1) in the year ahead. The model training and testing used 2014 to 2020 data, and forecasts were produced for 2021 (a year for which data were not yet unavailable). This section overviews the model training and testing process. Additional details are given in S4 and corresponding scripts are available at <https://github.com/PatBall1/DeepForestcast>.

At a time t (years since 2000), the set of valid pixels (J_t) were those that had forest cover. When a pixel transitioned from a forest state to a non-forest state, this change was taken to be permanent. Specifically, a datapoint with a specific spatial location (central pixel), j , and point in time, t , consisted of data from a square $(2r + 1) \times (2r + 1)$ pixel scene, where r is the number of pixels the scene has in vertical and horizontal directions from the target central pixel, of the layers given in Table 1 (stacked bands). This allowed the network to observe what was happening in the neighbourhood of the pixel being considered (i.e. make decisions based on the field of view), which contrasts with typical ML approaches that use information on the pixel in isolation or with averaged neighbourhood metrics that lose the spatial structure information of the data. Each datapoint was associated with a non-forest (0) or forest (1) label from the year ahead ($t + 1$; see Figure 2).

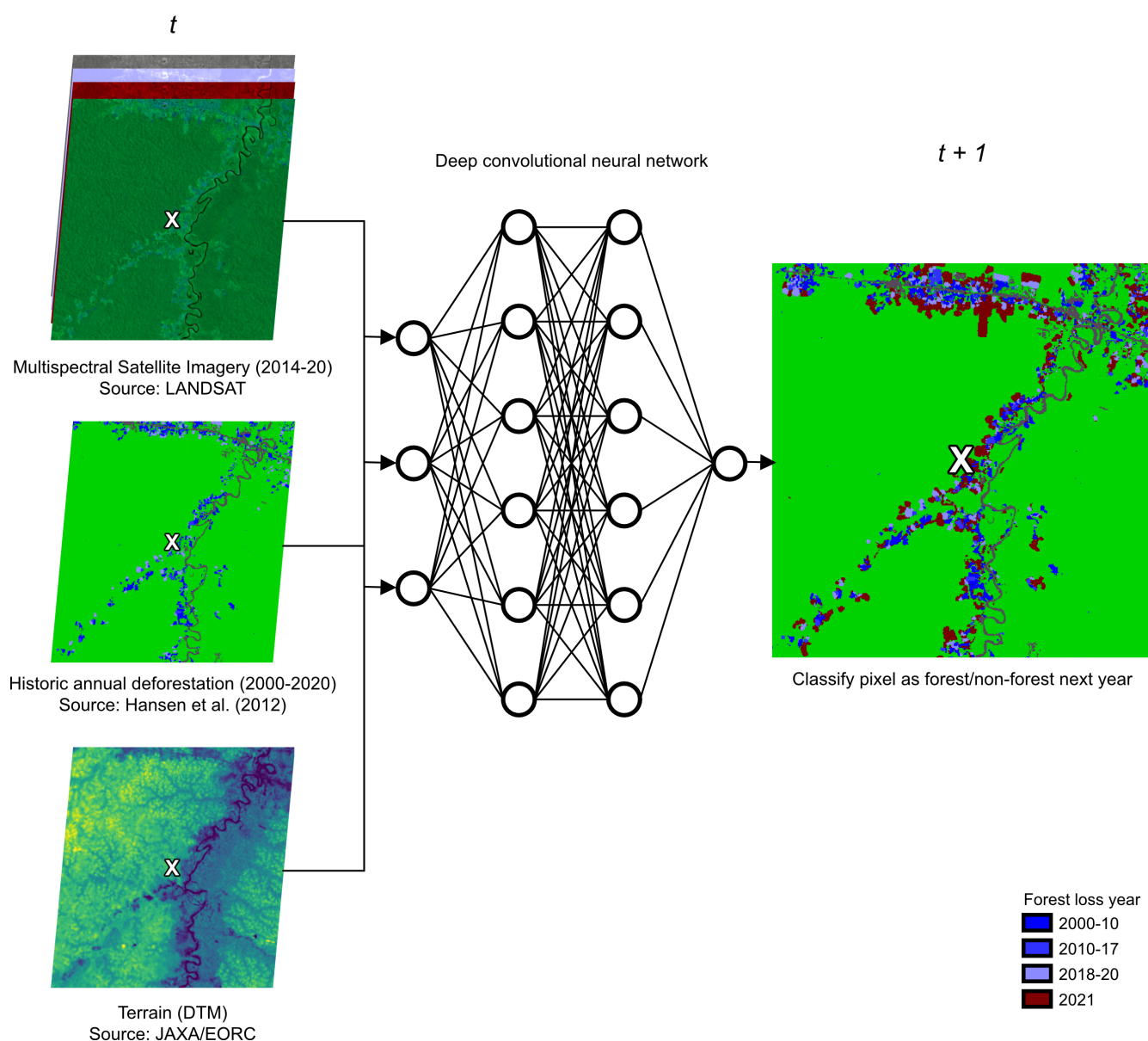


FIGURE 2 For a given forested pixel at the Centre of a scene, the convolutional neural networks learned to predict the state of the pixel (forest/non-forest) in the year ahead based on multispectral satellite imagery, historic loss patterns and the terrain.

For each datapoint, two types of 3D tensors were inputted into the models. The first, S^j , contained the stacked static layers (see Table 1) and was of the shape $S^j \in \mathbb{R}^{s \times (2r+1) \times (2r+1)}$, where s is the number of static layers (here 2 or 3 depending on whether the DSM layer was included). The second type, X_t^j , contained the stacked dynamic layers (see Table 1) and was of the shape $X_t^j \in \mathbb{R}^{d \times (2r+1) \times (2r+1)}$ where d is the number of dynamic layers (here 8). The label associated with the datapoint was defined as $Y_{t+1}^j \in \{0, 1\}$, and took a value of 1 only if the central target pixel was labelled as non-forest in year $t + 1$. By allowing the models to learn the features (and interactions) of the data layers in tensors that are associated with the year-ahead forest/non-forest labels, the model would then be able to make predictions of the year-ahead label (\hat{Y}_{t+1}^j) when presented with new input tensors. Output \hat{Y}_{t+1}^j could vary continuously between 0 (no chance of forest loss) and 1 (certain transition) with a threshold applied to determine the binary classification (what we refer to here as deforestation 'risk').

The input data for each year, with associated year-ahead ($t + 1$) labels, was partitioned, with stratified random sampling, into training (80%) and test (20%) sets. The training set was further split with stratified random sampling into five even splits so that fivefold cross validation could be performed while tracking model accuracy during the training process. Stratification was based on the proportion of class labels. Less than 0.5% of each study area transitioned from forest to non-forest in any given year which meant the number of pixels that remained as forest the next year (0) vastly outweighed the number that transitioned to non-forest (1). This extreme class imbalance, if not addressed, would likely have led to very low predictive accuracy for the infrequent class. As suggested by Buda et al. (2018), we modified our training data by under-sampling the over-represented class so that there was at least one positive case (1) to four negative cases (0). Fortunately, there remained several million valid datapoints available in each study region each year.

A preliminary, broad intercomparison of all model classes was first carried out on the MdD study region using 2014–2018 training and testing data. Using a grid search of hyperparameters, a broad set of models (different architectures and hyperparameters) was trained (with fivefold cross validation) on 2014–2017 input data based on labels up to 2018. The accuracy of these models was tested on a (within training period) withheld test set (2014–2017 input data, 2018 labels). Model selection was based on AUC (see below). This process was used to evaluate the most effective network depths for each model type.

Training deep convolutional neural networks is resource intensive so, from the initial intercomparison, the most promising MAs, in terms of model accuracy and training efficiency, were selected to be taken forward to be further refined and tested on the Junín region as well as MdD. Using Bayesian hyperparameter tuning (Snoek et al., 2012; implemented on wandb.ai) and retraining on the 2014–2018 inputs (2019 labels), optimised models were produced. The trails and performances achieved during the hyperparameter sweep is available to inspect at <https://wandb.ai/patball/forecasting/sweeps/df5v36lz>.

The models were tested on the withheld 2014–2018 inputs (2019 labels) test set before being tested on the year beyond the training period test set (2015–19 inputs; 2020 labels). It was important to test the models' ability to predict the labels for a year outside of its training period to understand whether the models are transferable through time and therefore able to produce reliable forecasts. The optimised models (with hyperparameters and weights retained) were then updated by continued training on 2015–2019 input data and 2020 labels. The final models were then used to produce deforestation forecasts across the entirety of each study area for 2021 (available for inspection as a GeoTiff at <https://doi.org/10.5061/dryad.hdr7sqvjz>). Figure S13 summarises the arrangement of modelling processes.

Fresh models, with the same architecture and parameterization as the optimised models, were trained only on the MdD data. To test spatial transferability, the accuracy of this model was evaluated first on MdD and then the Junín region. This model was then used to forecast 2021 deforestation for MdD (available for inspection as a GeoTiff at <https://doi.org/10.5061/dryad.hdr7sqvjz>). Additional models, with the same parameterization as the optimised models but sequentially trained on data on MdD data then Junín region data were assessed to help understand the effect of expanding the geographic diversity of training cases.

The metrics we used to evaluate the models were area under receiver operating characteristic curve (AUC), precision, recall and F_1 score. AUC integrates model performance over all classification thresholds. As the threshold could be adjusted to suit the user's relative preference towards a good true-positive rate (recall or sensitivity) or false-positive rate, AUC was insightful to assess overall model performance. As F_1 score (harmonic mean of the precision and recall) can communicate accuracy while accounting for variable class imbalance it was selected as a metric for model intercomparison over simple prediction accuracy (which can be misleading in cases of class imbalance).

We used the Adam optimizer with weighted cross entropy loss to train the models through stochastic gradient descent (Kingma & Ba, 2014). The penalty for missing a loss event was set to be double that of incorrectly identifying a loss event at a pixel that remained forested as we felt it preferable for the network to be tuned to avoid missing the relatively rare loss events. To avoid overfitting, regularisation was implemented through a dropout layer (Srivastava et al., 2014) and early stopping criteria. The models were trained with repeated exposure to the full sampled training set (i.e. multiple epochs).

We used the PyTorch ML framework to develop the models (code available at <https://github.com/PatBall1/DeepForestcast>). We used GPUs (4 × Tesla P100-PCIE-16GB) and CPUs (12 × 5980MB) with a 12-hour time limit on a high-performance computing cluster to train and test the models and make forecasts. Details on the computational resources used, including hardware and software, are given in S5. The trials conducted as part of the training process, including hyperparameters used, accuracies attained, and computing resources used is available at <https://wandb.ai/patball/forecasting/>.

2.4 | Model architectures

A variety of MAs were designed and implemented, and described in brief below. Technical details and developed rationales behind the choices of the network architectures can be found in S3. A deep neural network has a set of weights which define the strength of connections between the nodes and the predictions it makes from input data; these are adjusted as the network learns from exposure to training data. A single MA has a corresponding set of variable hyperparameters. Hyperparameters define aspects of how a model handles data, builds interactions, and how it adjusts its weights to learn from the data. A unique set of hyperparameters as well as the architecture defines what we refer to as a *model*. By varying the hyperparameters and comparing the resultant model accuracies, the most effective set of hyperparameters can be identified (hyperparameter tuning; see Section 2.3). A spatial pyramid pooling layer in each network allowed the spatial input window size to be varied and optimised as a hyperparameter (He et al., 2015).

2.4.1 | Model architecture 1: 2D convolutional neural network

The first and simplest MA we employed (MA1) was a type of 2D convolutional neural network (2D CNN). In a 2D CNN, filters slide ('convolve') in two spatial dimensions across a 3D input tensor (with a spatial extent and a depth that corresponds to the number of bands) to extract spatial features that can be used by the network to make predictions. Inputs were simply the dynamic tensor static tensor stacked at a point in space. Potential temporal features of the data were not implicitly characterised by this type of network. Temporal change was simply captured by pairing the input data for 1 year (at time t) with the forest/non-forest labels from a year ahead (at $t + 1$). See Figure S10 for a visual representation of the MA.

2.4.2 | Model architecture 2: 3D convolutional neural network

The second MA (MA2) was a type of 3D convolutional neural network (3D CNN). Li et al. (2017) demonstrated that 3D CNNs applied to hyperspectral data could achieve state-of-the-art land use classification accuracy. We thought that, by substituting the spectral dimension with the temporal dimension, it would be possible to draw out the spatiotemporal features of forest change. This type of network slides ('convolves') filters in three dimensions (two spatial and one temporal). This meant, for a given central target pixel, the series of dynamic tensors (see Section 2.4.3) could be stacked temporally along the channel axis to form a 4D input tensor and 3D convolutions used to characterise the spatio-temporal features of the input data. The static tensor was passed to a separate 2D convolutional branch. See Figure S11 for a visual representation of this MA.

2.4.3 | Model architectures 3 and 4: Convolutional long short-term memory recurrent neural network

MAs 3 and 4 were also designed to implicitly handle the temporal characteristics of the input data. Instead of using 3D convolutions, they use a Convolutional Long Short Term Memory (ConvLSTM)-based Recurrent Neural Network (RNN) type architecture. Rußwurm and Körner (2018) demonstrated the state-of-the-art performance of this network type for crop cover classification from multispectral satellite data. At a given time step, RNNs can *remember* and make use of information from previous time steps; they are designed to characterise and predict sequential data. LSTM is a specialised RNN design that allows for long time dependencies to be learnt. By integrating spatial convolutional elements within the LSTM component, a ConvLSTM can characterise and predict the sequential evolution of spatial patterns. MA3 had a single ConvLSTM cell, whereas MA4 was 'deeper' as it had a stack of ConvLSTM cells (see S3.4 for details). We compared these two architectures as stacking ConvLSTM cells can improve prediction accuracy in some situations (e.g. Kim et al., 2017). These MAs take the same 4D input tensors as the 3D CNN and pass the static tensor to a 2D convolutional branch. These MAs, while able to achieve comparable accuracies to the first two, were dropped after a broad intercomparison as their training could not be parallelised and so it took far longer to achieve these accuracies. However, their model classes are available for experimentation from <https://github.com/PatBall1/DeepForestcast> as their implementation may become more effective as the amount of historic data and available computing resources increases.

2.4.4 | Baseline: Random forest

To assess the relative benefit of using deep CNN-type methods, we trained, tuned and tested random forest classifiers to perform the same task. The data structure was equivalent to that of the 2D CNN, that is, a spatial window around a focal pixel was still used as an input and forest state labels were offset by 1 year. Spatial structure was not preserved—the 3D tensors were necessarily flattened into a 1D feature vector—but the model was free to learn which pixels (closer or farther from the central pixel) were of greatest importance. Fivefold cross-validation with a randomised search over the hyperparameters and window size was performed to tune the model.

3 | RESULTS

3.1 | Model performances

3.1.1 | Broad intercomparison

The best models from each model class (i.e. selected after hyperparameter tuning) predicted 2018 MdD test set deforestation with similar accuracies (AUC = 0.935–0.944). Precision and recall were

approximately equal across MA classes. The best performing MA after hyperparameter tuning was the 3D CNN (MA3; AUC = 0.944). This was followed by the deep ConvLSTM (MA4; AUC = 0.938), the ConvLSTM (MA3; AUC = 0.937) and the 2D CNN architecture (MA1; AUC = 0.935). This suggested that models that could implicitly handle and characterise the temporal with the spatial structure of the input data may be better suited to predicting future deforestation. In other words, those models that could retain and work with the spatiotemporal patterns of the data (how the spatial patterns evolved) rather than just use spatial patterns tended to perform better. MA4 and MA3, while achieving comparable accuracies, were uncompetitive with respect to training time due to their constraints on parallelisation (a well-known problem with RNNs). Therefore, MA1 and MA2 were selected to be taken forward for further development and testing.

The spatial dimensions of the input tensors determined the size of the scene that models were able to view and had an influence on model performance. Model AUC plateaued around an input frame of $\sim 35 \times 35$ pixels ($\sim 1 \text{ km} \times 1 \text{ km}$) and began to decrease for input frames larger than $\sim 50 \times 50$ pixels ($\sim 1.5 \text{ km} \times 1.5 \text{ km}$). Between these sizes seemed to be the sweet spot where the networks could learn from local context while also remaining focused on the pixel of interest. Full details of the model performances, including hyperparameters and receiver operating characteristic (ROC) curves, are given in S7.

3.1.2 | Model development

Further tuning and extended training improved within-year accuracy of the 2D CNN and 3D CNN model classes which performed better than the baseline models. Across the board, the models predicted less accurately on the Junín region than MdD. It suggests that the Junín region is inherently more difficult to predict on (it has a more cloud cover, diversity of forest types and fragmentation). Models that included the topological layer as an input predictor layer were no more accurate than those without suggesting it was uninformative to the model. To reduce redundancy, this layer was dropped.

There was a drop in model prediction accuracy from the within training period test set and the year ahead test set (see Table 2). This was to be expected but it could also indicate some change in the patterns of deforestation or a change in the nature of the input data (it is difficult to produce satellite data that is exactly consistent across dates due to e.g. atmospheric conditions). The greater the complexity of the model, the greater the drop off in accuracy from nowcast to forecast predictions indicating a degree of overfitting. The model that was trained on both regions sequentially was only marginally better at predicting year ahead deforestation than the model that was trained on just the region to be predicted (for 3D CNN). The highest year ahead AUC from the 3D CNN for the Junín region was returned by the model trained exclusively on MdD data. This suggests that data quality is at least as important as volume and spatial range of training data. It also suggests that the spatio-temporal patterns of forest loss between the two regions are somewhat similar and that the models are spatially transferable.

TABLE 2 Within year and year ahead pixel wise classification accuracy of the 3D CNN, 2D CNN and random forest baseline model classes. The reported accuracies are all for unseen test sets. The 2019 labels are within the training period and 2020 are beyond (to test forecast accuracy). The ratio of 0 to 1 labels in the test sets were 4:1

Model class	Training period (training labels)	Training region	Prediction region	Test set 2019				Test set 2020			
				AUC	Precision	Recall	F_1	AUC	Precision	Recall	F_1
3D CNN	2014–2018 (2018–2019)	Junín	Junín	0.990	0.903	0.903	0.903	0.839	0.580	0.580	0.580
3D CNN	2014–2018 (2018–2019)	Madre de Dios	Junín	0.879	0.663	0.663	0.663	0.861	0.593	0.593	0.593
3D CNN	2014–2018 (2018–2019)	Madre de Dios; Junín	Junín	0.997	0.904	0.904	0.904	0.849	0.589	0.589	0.589
3D CNN	2014–2018 (2018–2019)	Junín	Madre de Dios	0.961	0.902	0.903	0.903	0.863	0.667	0.667	0.667
3D CNN	2014–2018 (2018–2019)	Madre de Dios	Madre de Dios	0.997	0.964	0.964	0.964	0.890	0.715	0.715	0.715
2D CNN	2014–2018 (2015–2019)	Junín	Junín	0.984	0.865	0.865	0.865	0.876	0.623	0.623	0.623
2D CNN	2014–2018 (2015–2019)	Madre de Dios	Madre de Dios	0.882	0.673	0.673	0.673	0.868	0.654	0.654	0.654
Random forest (7×7)	2014–2018 (2015–2019)	Junín	Junín	0.806	0.688	0.434	0.532	0.803	0.679	0.430	0.527
Random forest (7×7)	2014–2018 (2015–2019)	Madre de Dios	Madre de Dios	0.816	0.750	0.492	0.594	0.814	0.750	0.489	0.592

Bold shows the best performing models for forecasting at each site.

While the 3D CNN forecasted most accurately on the MdD region, the 2D CNN gave the best forecast accuracy for the Junín region. The 3D CNN has the advantage of being able to convolve over spatial and temporal dimensions together (and so learn spatio-temporal features) but the 2D CNN has more datapoints available to learn from as the inputs are for a single year input data (with an offset label) and not stacked across time like with the other model types. More details of the model performances over different hyperparameters is available at <https://wandb.ai/patball/forecasting2d/> and <https://wandb.ai/patball/forecasting/> for the 2D CNN and 3D CNN, respectively.

3.2 | Forecasts

The near-future forecasts exhibited some interesting spatial features. To illustrate the 3D CNN's assessment of deforestation risk, we have focused on four recent forest loss events in the two study regions. The 2021 risk forecast was overlaid over Planet/NICFI biannual composite satellite imagery for 2020 (see Figure 3). Figure 3a shows a rapidly growing commercial agricultural area well serviced by a river and roads in Junín. The network signals a very high risk that agricultural clearance will continue to expand into the remaining forest areas. Figure 3b shows a new, remote unauthorised gold mine in MdD. The network anticipates an immediate expansion of the mining operations into surrounding areas. Figure 3c shows a newly laid

access route connecting newly established agricultural settlements to a large river. The network highlights a risk of clustered conversion along parts of its route. Figure 3d shows a large recent forest loss event caused by a landslide. Despite the scale of the event, the network deems its surroundings as relatively low risk from further change of state compared with the anthropogenic causes of loss. A remote landslide is unlikely to result in a progressive loss of forest. The contrasting predictions (in combination with the high reported accuracy) suggest a level of discrimination (or *intelligence*) in the model's evaluation of transition likelihood. Forecasts are available for inspection in GeoTiff format at <https://doi.org/10.5061/dryad.hdr7sqvjz/https://datadryad.org/stash/share/QeVIFyTmNbbcQEoOmoOqmWg9Ql4UaxvCB7lkTTZumyc>.

4 | DISCUSSION

4.1 | Improving spatial forecasts

For the first time, deep CNNs were used to forecast the spatial development of forest loss. From a limited set of freely available, global predictor layers, we predicted year-ahead forest loss at 30 m resolution to a good degree of accuracy (best models $F_1 = 0.62$ – 0.71) relative to the baseline models ($F_1 = 0.52$ – 0.60) for two ~80,000 km² regions in the Southern Peruvian Amazon while also demonstrating a degree of spatial and temporal transferability of the models.

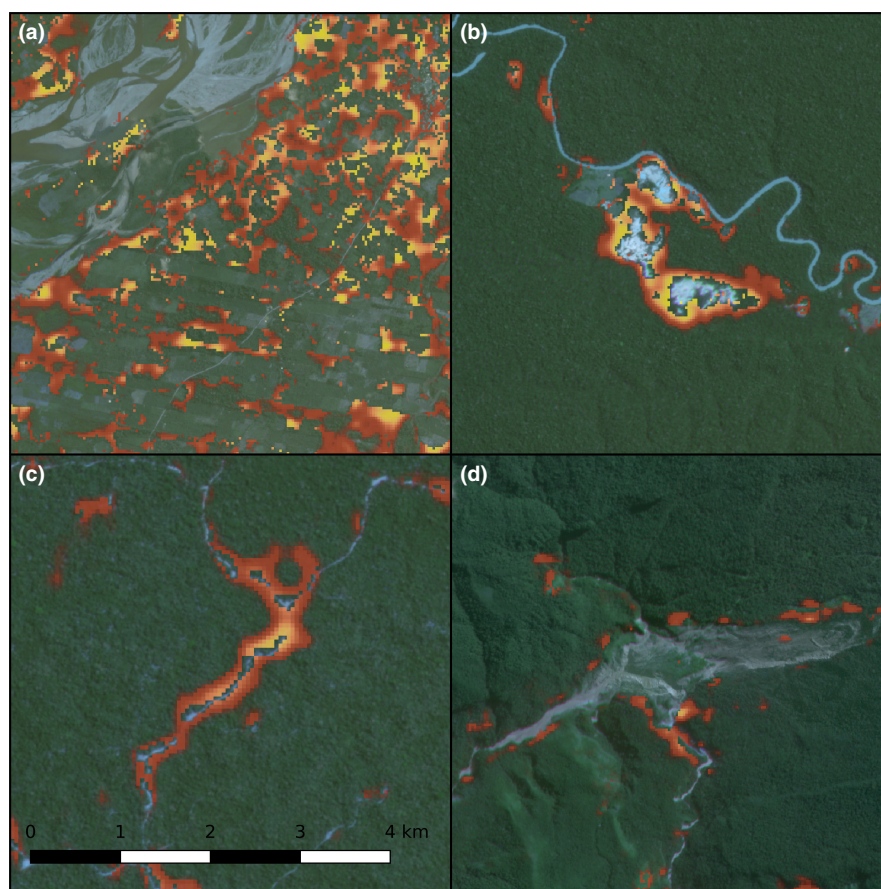


FIGURE 3 Examples from the 2021 deforestation risk forecasts where yellow indicates the areas most likely to transition to non-forest states fading to red as areas are less likely to transition: (a) expanding commercial agriculture; (b) a newly established, unauthorised gold mine; (c) a new access route connecting undeveloped forest to a river and (d) a recent remote landslide. Source of satellite imagery: Planet/NICFI.

The apparent ability of the networks to automatically extract landscape features relating to different drivers of forest loss and infer whether loss was likely to continue from a source was encouraging. It highlights the potential for deep learning computer vision methods to process large volumes of RS data to manage risks and make predictions in complex human-natural systems. Previous attempts at forecasting deforestation have often had limited predictive capacity due to a lack of reliable input data on dynamic landscape features. From satellite data and other complementary spatial datasets, our model appeared able to discriminate between one-off loss events (e.g. landslide) from frontiers that are likely to progress over time (e.g. from a newly built road, an unauthorised gold mine).

The dimensionality of convolutions in the networks—spatial in the 2D CNN case and spatiotemporal in the 3D CNN case—involved a trade-off. The ability of the 3D CNN to exploit spatiotemporal features allowed it to perform most accurately in the case of MdD, the region with less topological variation and ecosystem variety. Indeed, the ability of 3D CNNs to view a pixels spatial and temporal context together has been shown to improve classification accuracy from RS data (Ji et al., 2018; Xu et al., 2018). This implicit handling of the temporal dimension with the spatial dimensions would seem especially important in the case of forecasting. However, stacking the inputs across the time axis reduced the number of individual datapoints available to the network to learn from. The 2D CNN, with a greater volume and temporal variety of data available for it to learn from, performed better on the Junín region. The Junín region has persistent cloud cover and so stacking the data over time may propagate noise and degrade training data. The individual year inputs of the 2D CNN are more resistant to this as potential noise or gaps in the data would not contaminate informative datapoints. The greater volume of data may have helped the network deal with the environmental variability.

Deep learning models have many hyperparameters that need tuning. We demonstrated a degree of spatial transferability of our models, but users should be aware that additional tuning and retraining may be necessary to achieve optimal performance in a new region. The optimal window width to view the scene was 1–1.5 km. For larger windows, the marginal benefit from the additional information provided less benefit than the cost of handling this information. In other words, there was a point at which adding further away pixels to a scene began to distract the network from the area of focus.

We stress that, while the reported accuracy is encouraging, there are reasons to receive these metrics cautiously. The GFC (Hansen et al., 2013) dataset provided all data labels and so the reported accuracies specifically relate to how well the networks predict future outputs of this classifier and not necessarily the ground truth. The GFC dataset was generated by a bagged decision tree that classifies filtered Landsat data. Our models were likely to have approximated and incorporated behaviours of this classifier and accuracy may be inflated (relative to on-the-ground change) from learned correlations.

Any ML approach is necessarily limited by the fact it is based on historic data and the assumption that observed trends will continue, and therefore does not implicitly account for regime changes in the

system. This is relevant here, as political changes have substantial impacts on Amazonian deforestation (Pereira et al., 2020). A changing climate, volatile international economic conditions and advances in technology further challenge the assumption of stationarity. However, as the forecasting system can be dynamically updated it is able to learn emerging patterns which is an improvement on previous approaches. Additionally, the models output a continuous measure of risk across all pixels and, by varying the threshold at which the binary classification is made, it is possible to adjust the overall allocation of forest loss in predictions. It would be possible to couple the spatial evaluation of risk from the CNN system (accounting for landscape scale changes) to a model of overall forest loss based on external drivers to generate forecasts that reflect global trends. The pixelwise risks can be integrated to assess relative risk across a region.

4.2 | Data sources

We limited our set of data layers to test the networks' ability to automatically extract dynamic landscape features. However, including additional RS data layers as predictors could help to inform the networks and improve accuracy. Of particular interest is night lights, which provide a dynamic signal for human presence and economic activity. At current, the resolutions of available night light products are too coarse to integrate into our system, but this is likely to change in the near future (Levin et al., 2020). The ~30m resolution of Landsat derived products is likely to fail to pick up on selective logging. Higher resolution imagery (e.g. Planet NICFI, Sentinel-2) may make that feasible but there is not currently a matching dataset for forest cover labels. The GFC dataset only provides annual information on forest loss and not forest gain which means the forecasts presented here cannot account for reforestation. The Copernicus Global Land Cover collection (Buchhorn et al., 2020) could provide labels for more types of landcover transitions but is only available for 2015–2019 and at 100m resolution. As this data source develops more intricate predictions that support forest regeneration analyses may be possible. Furthermore, rather than simply predicting transitions between forest and non-forest states it may also be valuable to make predictions on gradual declines in relation to disturbance events and forest resilience. In this context, it would be possible to train networks similar to those presented here to predict the evolution of greenness metrics such as NDVI (Boulton et al., 2022). The ALOS JAXA elevation layer was found to not improve model performance, which was somewhat surprising in Junín, which has inaccessible mountainous areas. It may be that processing the elevation into a relative slope layer would allow the networks to better comprehend the features of the local topology.

4.3 | Probing the black box

Deep neural networks are considered 'black box' models so it is difficult to say how the networks were making decisions and to what

degree they were recognising different potential drivers of loss. There has been work on interpreting deep CNNs to explore their internal reasoning (Li et al., 2021). By applying interpretation algorithms, it may be possible to identify which parts of a scene are triggering the networks' decision. It may even be possible to generate groups of scenes (e.g. by driver) that typically signal forest loss (see e.g. the Activation Atlas of Carter et al., 2019). Automatic classification of drivers has been demonstrated at coarse resolutions (10km) with simple decision tree models applied to derived RS products (Curtis et al., 2018). Given the relative sophistication of the CNN technology available, it is conceivable that a network could be trained to classify and map drivers directly from satellite imagery at high resolution, particularly as the amount of high-resolution imagery is rapidly increasing (e.g. Planet NICFI, Sentinel-2). However, a comprehensive training set of labelled deforestation drivers at this resolution does not exist and would require a considerable investment to produce.

Fieldwork to potential transition regions may be another way to better understand the networks' predictions. Of particular interest would be to investigate the cases where the models are failing to make accurate predictions. Identifying specific drivers that lead to misclassification would allow for more focused model training. Active learning algorithms (Settles, 2010) could guide expeditions by identifying regions that would be particularly informative to gather extra data on.

4.4 | Technological advances

The MAs were inspired by models that have performed well on land cover classification and adapted to fit the specific forecasting task. However, recent architectural innovations could further help to improve performance. Transformer networks (originally from natural language processing) have proven themselves to be the state-of-the-art in handling sequential data (Vaswani et al., 2017) and recently in some computer vision tasks (Dai et al., 2021). Using self-attention, they avoid the difficulties in parallelising RNNs, and so can be trained more efficiently. Integrating transformer components into the networks may prove to be an effective way of handling the spatiotemporal features of the data (especially as the available time series grows). This could help to extend the networks forecasting horizon. Further improvements could also be gained by refining the training process to focus it on contentious datapoints. The under sampling of the non-change class was random and many training points would be uninformative as they are in very remote locations. A method that directs the training to cases that sit close the decision threshold between the classification states would address this. We have provided open-source code and benchmark results, and hope further improvements can be integrated through collaborations.

4.5 | Applications

The forecasting approach presented here was designed to be flexible, useful to a range of potential users and scalable to forests

globally. All pixels in a landscape can be ranked according to the likelihood of loss allowing for spatial prioritisation. An organisation may have resources to police a certain proportion of a region and the forecasts would allow them to monitor the most threatened areas. The forecasts could also direct interventions to emerging frontiers of loss in otherwise isolated and intact landscapes. On the other hand, when planning a permanent protected area, it may be preferable to avoid areas that are most imminently threatened but address areas that may undergo transition if left unprotected. In either case, it is beneficial to understand how threats are likely to evolve.

Governmental and non-governmental organisations can use the tools presented here to understand how the landscapes in which they operate will change into the future and refine their protocols for managing and responding to deforestation risk. The effectiveness of deep CNNs in the current context suggests that they should also be applied to other ecological forecasting tasks based on RS data such as forest fire forecasting.

AUTHOR CONTRIBUTIONS

James G. C. Ball, Seth Flaxman and Katerina Petrova conceived of the ideas and designed methodology presented; Katerina Petrova, James G. C. Ball and Seth Flaxman coded models and routines; James G. C. Ball led the writing of the manuscript; Seth Flaxman, David A. Coomes and Katerina Petrova reviewed and edited the manuscript.

ACKNOWLEDGEMENTS

Thanks to Mark Burgman for initiating the collaboration between the authors. Thanks to Cool Earth for their early support. J.G.C.B. is supported by the Natural Environment Research Council. S.F. is supported by the Engineering and Physical Sciences Research Council.

CONFLICT OF INTEREST

None declared.

PEER REVIEW

The peer review history for this article is available at <https://publons.com/publon/10.1111/2041-210X.13953>.

DATA AVAILABILITY STATEMENT

All codes are freely available at the project Github repository <https://doi.org/10.5281/zenodo.6858022> (Ball et al., 2022a). Records of model runs and hyperparameter tuning is available at <https://wandb.ai/patball/>. All input data are freely available online. Processed, model ready inputs, model weights and model specifications are available from the Dryad Digital Repository <https://doi.org/10.5061/dryad.hdr7sqvjz> (Ball et al., 2022b). Additional forecasts can be produced by the authors on request.

ORCID

James G. C. Ball  <https://orcid.org/0000-0002-0165-5290>

David A. Coomes  <https://orcid.org/0000-0002-8261-2582>

Seth Flaxman  <https://orcid.org/0000-0002-2477-4217>

REFERENCES

- Abdi, O., Uusitalo, J., & Kivinen, V.-P. (2022). Logging Trail segmentation via a novel U-net convolutional neural network and high-density laser scanning data. *Remote Sensing*, 14, 349. <https://doi.org/10.3390/rs14020349>
- Ahmed, S. E., Ewers, R. M., & Smith, M. J. (2014). Large scale Spatio-temporal patterns of road development in the Amazon rainforest. *Environmental Conservation*, 41(3), 253–264.
- Asner, G. P., & Tupayachi, R. (2016). Accelerated losses of protected forests from gold mining in the Peruvian Amazon. *Environmental Research Letters*, 12, 094004. <https://doi.org/10.1088/1748-9326/aa7dab>
- Ball, J., Petrova, K., Coomes, D., & Flaxman, S. (2022a). v1.0.0 PatBall1/DeepForestcast: Publication. *Github Repository*. <https://doi.org/10.5281/zenodo6858022>
- Ball, J., Petrova, K., Coomes, D., & Flaxman, S. (2022b). Data from: Using deep convolutional neural networks to forecast spatial patterns of Amazonian deforestation: Supporting data and outputs. *Dryad Digital Repository*. <https://doi.org/10.5061/dryad.hdr7sqvjz>
- Barber, C. P., Cochrane, M. A., Souza, C. M., & Laurance, W. F. (2014). Roads, deforestation, and the mitigating effect of protected areas in the Amazon. *Biological Conservation*, 177(September), 203–209.
- Bass, M. S., Finer, M., Jenkins, C. N., Kreft, H., Cisneros-Heredia, D. F., McCracken, S. F., Pitman, N. C. A., English, P. H., Swing, K., Villa, G., di Fiore, A., Voigt, C. C., & Kunz, T. H. (2010). Global conservation significance of Ecuador's Yasuní National Park. *PLoS ONE*, 5(1), e8767.
- Beuchle, R., Achard, F., Bourgoïn, C., Vancutsem, C., Eva, H., & Follador, M. (2021). *Deforestation and Forest degradation in the Amazon*. European Union. <https://doi.org/10.2760/61682>
- Boulton, C. A., Lenton, T. M., & Boers, N. (2022). Pronounced loss of Amazon rainforest resilience since the early 2000s. *Nature Climate Change*, 12(3), 271–278. <https://doi.org/10.1038/s41558-022-01287-8>
- Brodrick, P. G., Davies, A. B., & Asner, G. P. (2019). Uncovering ecological patterns with convolutional neural networks. *Trends in Ecology & Evolution*, 34(8), 734–745. <https://doi.org/10.1016/j.tree.2019.03.006>
- Buchhorn, M., Lesiv, M., Tsendbazar, N.-E., Herold, M., Bertels, L., & Smets, B. (2020). Copernicus global land cover layers—Collection 2. *Remote Sensing*, 12(108), 1044. <https://doi.org/10.3390/rs12061044>
- Buda, M., Maki, A., & Mazurowski, M. A. (2018). A systematic study of the class imbalance problem in convolutional neural networks. *Neural Networks: The Official Journal of the International Neural Network Society*, 106(October), 249–259.
- Busch, J., & Ferretti-Gallon, K. (2017). What drives deforestation and what stops it? A meta-analysis. *Review of Environmental Economics and Policy*, 11(1), 3–23.
- Carter, S., Armstrong, Z., Schubert, L., Johnson, I., & Olah, C. (2019). Activation atlas. *Distill*, 4(3). <https://doi.org/10.23915/distill.00015>
- Chirinos, C., & Ruiz, M. (2003). *Informe final del estudio: Desarrollo e implementación de lineamientos de control de la extracción ilegal para un manejo forestal sostenible en el Perú*. ITTO, Ciudad de Panamá.
- Corbane, C., Syrris, V., Sabo, F., Politis, P., Melchiorri, M., Pesaresi, M., Soille, P., & Kemper, T. (2021). Convolutional neural networks for global human settlements mapping from Sentinel-2 satellite imagery. *Neural Computing and Applications*, 33, 6697–6720. <https://doi.org/10.1007/s00521-020-05449-7>
- Curtis, P. G., Slay, C. M., Harris, N. L., Tyukavina, A., & Hansen, M. C. (2018). Classifying drivers of global Forest loss. *Science*, 361(6407), 1108–1111.
- Cushman, S. A., Macdonald, E. A., Landguth, E. L., Malhi, Y., & Macdonald, D. W. (2017). Multiple-scale prediction of Forest loss risk across Borneo. *Landscape Ecology*, 32(8), 1581–1598.
- Dai, Z., Liu, H., Le, Q. V., & Tan, M. (2021). Coatnet: Marrying convolution and attention for all data sizes. *Advances in Neural Information Processing Systems*, 34, 3965–3977.
- de Filho, F. J. B. O., & Metzger, J. P. (2006). Thresholds in landscape structure for three common deforestation patterns in the Brazilian Amazon. *Landscape Ecology*, 21, 1061–1073. <https://doi.org/10.1007/s10980-006-6913-0>
- Finer, M., Babbitt, B., Novoa, S., Ferrarese, F., Pappalardo, S. E., De Marchi, M., Saucedo, M., & Kumar, A. (2015). Future of oil and gas development in the western Amazon. *Environmental Research Letters*, 10(2), 024003. <https://doi.org/10.1088/1748-9326/10/2/024003>
- Finer, M., Jenkins, C. N., Blue Sky, M. A., & Pine, J. (2014). Logging concessions enable illegal logging crisis in the Peruvian Amazon. *Scientific Reports*, 4, 4719. <https://doi.org/10.1038/srep04719>
- Geist, H. J., & Lambin, E. F. (2002). Proximate causes and underlying driving forces of tropical Deforestation Tropical forests are disappearing as the result of many pressures, both local and regional, acting in various combinations in different geographical locations. *Bioscience*, 52(2), 143–150.
- 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. G. (2013). High-resolution global maps of 21st-century Forest cover change. *Science*, 342(6160), 850–853.
- Hansen, M. C., Krylov, A., Tyukavina, A., Potapov, P. V., Turubanova, S., Zutta, B., Ifo, S., Margono, B., Stolle, F., & Moore, R. (2016). Humid tropical Forest disturbance alerts using Landsat data. *Environmental Research Letters: ERL [Web Site]*, 11(3), 034008.
- Harris, N. L., Goldman, E., Gabris, C., Nordling, J., Minnemeyer, S., Ansari, S., Lippmann, M., Bennett, L., Raad, M., Hansen, M., & Potapov, P. (2017). Using spatial statistics to identify emerging hot spots of Forest loss. *Environmental Research Letters: ERL [Web Site]*, 12(2), 024012.
- He, K., Zhang, X., Ren, S., & Sun, J. (2015). Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(9), 1904–1916. <https://doi.org/10.1109/TPAMI.2015.2389824>
- Interdonato, R., Ienco, D., Gaetano, R., & Ose, K. (2019). DuPLO: A DuAl view point deep learning architecture for time series classification. *ISPRS Journal of Photogrammetry and Remote Sensing: Official Publication of the International Society for Photogrammetry and Remote Sensing*, 149(March), 91–104.
- Jenkins, C. N., Pimm, S. L., & Joppa, L. N. (2013). Global patterns of terrestrial vertebrate diversity and conservation. *Proceedings of the National Academy of Sciences of the United States of America*, 110(28), E2602–E2610.
- Ji, S., Zhang, C., Anjian, X., Shi, Y., & Duan, Y. (2018). 3D convolutional neural networks for crop classification with multi-temporal remote sensing images. *Remote Sensing*, 10(1), 75. <https://doi.org/10.3390/rs10010075>
- Kattenborn, T., Leitloff, J., Schiefer, F., & Hinz, S. (2021). Review on Convolutional Neural Networks (CNN) in Vegetation Remote Sensing. *ISPRS Journal of Photogrammetry and Remote Sensing: Official Publication of the International Society for Photogrammetry and Remote Sensing*, 173(March), 24–49.
- Kim, S., Hong, S., Joh, M., & Song, S.-K. (2017). Deepbrain: ConvLstm network for precipitation prediction using multichannel radar data. *arXiv [cs.LG]* arXiv:1711.02316. <https://doi.org/10.48550/arXiv.1711.02316>
- Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv [cs.LG]* arXiv. <http://arxiv.org/abs/1412.6980>
- Kislov, D. E., Korznikov, K. A., Altman, J., Vozmishcheva, A. S., & Krestov, P. V. (2021). Extending deep learning approaches for forest disturbance segmentation on very high-resolution satellite images. *Remote Sensing in Ecology and Conservation*, 7, 355–368. <https://doi.org/10.1002/rse2.194>
- Kussul, N., Lavreniuk, M., Skakun, S., & Shelestov, A. (2017). Deep learning classification of land cover and crop types using remote sensing data. *IEEE Geoscience and Remote Sensing Letters*, 14(5), 778–782.

- Levin, N., Kyba, C. C. M., Zhang, Q., Sánchez de Miguel, A., Román, M. O., Li, X., Portnov, B. A., Molthan, A. L., Jechow, A., Miller, S. D., Wang, Z., Shrestha, R. M., & Elvidge, C. D. (2020). Remote sensing of night lights: A review and an outlook for the future. *Remote Sensing of Environment*, 237, 111443. <https://doi.org/10.1016/j.rse.2019.111443>
- Li, X., Xiong, H., Li, X., Wu, X., Zhang, X., Liu, J., Bian, J., & Dou, D. (2021). Interpretable deep learning: Interpretation, interpretability, trustworthiness, and beyond. *arXiv [cs.LG]*. arXiv. <http://arxiv.org/abs/2103.10689>
- Li, Y., Zhang, H., & Shen, Q. (2017). Spectral-spatial classification of hyperspectral imagery with 3D convolutional neural network. *Remote Sensing*, 9(1), 67.
- Lovejoy, T. E., & Nobre, C. (2019). Amazon Tipping Point: Last Chance for Action. *Science Advances*, 5(12), eaba2949.
- Mayfield, H., Smith, C., Gallagher, M., & Hockings, M. (2017). Use of freely available datasets and machine learning methods in predicting deforestation. *Environmental Modelling & Software*, 87(January), 17–28.
- Mena, C. F., Laso, F., Martinez, P., & Sampedro, C. (2017). Modeling road building, deforestation and carbon emissions due deforestation in the Ecuadorian Amazon: The potential impact of oil frontier growth. *Journal of Land Use Science*, 12(6), 477–492.
- Moffette, F., Alix-Garcia, J., Shea, K., & Pickens, A. H. (2021). The impact of near-real-time deforestation alerts across the tropics. *Nature Climate Change*, 11(2), 172–178.
- Nicolau, A. P., Herndon, K., Flores-Anderson, A., & Griffin, R. (2019). A spatial pattern analysis of forest loss in the Madre de Dios region, Peru. *Environmental Research Letters*, 14(12), 124045. <https://doi.org/10.1088/1748-9326/ab57c3>
- Olson, D. M., Dinerstein, E., Wikramanayake, E. D., Burgess, N. D., Powell, G. V. N., Underwood, E. C., D'Amico, J. A., Itoua, I., Strand, H. E., Morrison, J. C., Loucks, C. J., Allnutt, T. F., Ricketts, T. H., Kura, Y., Lamoreux, J. F., Wettengel, W. W., Hedao, P., & Kassem, K. R. (2001). Terrestrial ecoregions of the world: A new map of life on earth. *Bioscience*, 51(11), 933–938.
- Pereira, E. J., de Area, L., de Santana Ribeiro, L. C., da Silva Freitas, L. F., & de Barros Pereira, H. B. (2020). Brazilian policy and agribusiness damage the Amazon rainforest. *Land Use Policy*, 92(March), 104491.
- Perz, S. G., Caldas, M. M., Arima, E., & Walker, R. J. (2007). Unofficial road building in the Amazon: Socioeconomic and biophysical explanations. *Development and Change*, 38(3), 529–551.
- Piotrowski, M. (2019). *Nearing the tipping point: Drivers of deforestation in the Amazon region*. Inter-American Dialogue.
- Potapov, P., Hansen, M. C., Laestadius, L., Turubanova, S., Yaroshenko, A., Thies, C., Smith, W., Zhuravleva, I., Komarova, A., Minnemeyer, S., & Esipova, E. (2017). The last Frontiers of wilderness: Tracking loss of intact Forest landscapes from 2000 to 2013. *Science Advances*, 3(1), e1600821.
- 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. *Environmental Research Letters: ERL [Web Site]*, 16(2), 024005.
- Rosa, I. M. D., Ahmed, S. E., & Ewers, R. M. (2014). The transparency, reliability and utility of tropical rainforest land-use and land-cover change models. *Global Change Biology*, 20(6), 1707–1722.
- Rußwurm, M., & Körner, M. (2018). Multi-temporal land cover classification with sequential recurrent encoders. *ISPRS International Journal of Geo-Information*, 7(4), 129.
- Saha, S., Saha, M., Mukherjee, K., Arabameri, A., Ngo, P. T. T., & Paul, G. C. (2020). Predicting the deforestation probability using the binary logistic regression, random Forest, ensemble rotational Forest, REPTree: A case study at the Gumani River basin, India. *The Science of the Total Environment*, 730(August), 139197.
- Settles, B. (2010). *Active learning literature survey*. Computer Sciences Technical Report 1648. University of Wisconsin.
- Snoek, J., Larochelle, H., & Adams, R. P. (2012). Practical Bayesian optimization of machine learning algorithms. In F. Pereira, C. J. C. Burges, L. Bottou, & K. Q. Weinberger (Eds.), *Advances in Neural Information Processing Systems* (Vol. 25). Curran Associates, Inc. <https://proceedings.neurips.cc/paper/2012/file/05311655a15b75fab86956663e1819cd-Paper.pdf>
- Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research: JMLR*, 15(56), 1929–1958.
- Torres, D. L., Turnes, J. N., Vega, P. J. S., Feitosa, R. Q., Silva, D. E., Junior, J. M., & Almeida, C. (2021). Deforestation detection with fully convolutional networks in the Amazon Forest from Landsat-8 and Sentinel-2 images. *Remote Sensing*, 12, 5084. <https://doi.org/10.3390/rs13245084>
- Turubanova, S., Potapov, P. V., Tyukavina, A., & Hansen, M. C. (2018). Ongoing primary forest loss in Brazil, Democratic Republic of the Congo, and Indonesia. *Environmental Research Letters*, 13(7), 074028.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, 30. <https://papers.nips.cc/paper/2017>
- Voulodimos, A., Doulamis, N., Doulamis, A., & Protopapadakis, E. (2018). Deep learning for computer vision: A brief review. *Computational Intelligence and Neuroscience*, 2018(February), 7068349.
- Wilson, K. A., Underwood, E. C., Morrison, S. A., Klausmeyer, K. R., Murdoch, W. W., Meyers, B., Wardell-Johnson, G., Marquet, P. A., Rundel, P. W., McBride, M. F., Pressey, R. L., Bode, M., Hoekstra, J. M., Andelman, S., Looker, M., Rondinini, C., Kareiva, P., Shaw, M. R., & Possingham, H. P. (2007). Conserving biodiversity efficiently: What to do, where, and when. *PLoS Biology*, 5(9), e223.
- WWF. (2020). Predicting deforestation with an early warning system. WWF Netherlands, 2020. <https://www.wwf.nl/wat-we-doen/aanpak/internationale/lokale-bevolking/conservation-for-and-with-people/predicting-deforestation>
- 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 Journal of Photogrammetry and Remote Sensing*, 144(2018), 423–434.
- Zhu, X. X., Tuia, D., Mou, L., Xia, G.-S., Zhang, L., Feng, X., & Fraundorfer, F. (2017). Deep learning in remote sensing: A comprehensive review and list of resources. *IEEE Geoscience and Remote Sensing Magazine*, 5(4), 8–36.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Ball, J. G. C., Petrova, K., Coomes, D. A., & Flaxman, S. (2022). Using deep convolutional neural networks to forecast spatial patterns of Amazonian deforestation. *Methods in Ecology and Evolution*, 00, 1–13. <https://doi.org/10.1111/2041-210X.13953>