



**MANIPAL INSTITUTE OF TECHNOLOGY**  
**MANIPAL**  
*A Constituent Institute of MAHE, Manipal*

*SEMINAR REPORT ON*

**Change Detection of Deforestation in the Brazilian Amazon Using  
Landsat Data and Convolutional Neural Networks**

*SUBMITTED TO*

**Department of Computer Science & Engineering**

*by*

**Abhijeet Sahdev**

**170905316**

**Roll No. 42**

**Semester 7 Section A**

*Guide*

**Dr.Prema K.V**

**Professor, Dept. of CSE**

*Evaluated by*

**Dr. R. Vijaya Arjunan & Dr. Manjunath K N**

**August 30, 2020**

# Contents

<b>1</b>	<b>INTRODUCTION</b>	<b>2</b>
1.1	Abstract . . . . .	2
1.2	Motivation . . . . .	2
<b>2</b>	<b>LITERATURE REVIEW</b>	<b>3</b>
<b>3</b>	<b>METHODOLOGY</b>	<b>4</b>
3.1	Train and Test Sites . . . . .	4
3.2	Deep Learning Models . . . . .	6
3.3	Data Structure . . . . .	7
3.4	Ground Truth . . . . .	8
3.5	Hyperparameters . . . . .	9
3.6	Modelling Approach . . . . .	10
3.7	Accuracy Assessment . . . . .	10
<b>4</b>	<b>RESULTS</b>	<b>11</b>
4.1	Quantitative Analysis . . . . .	11
4.2	Visual Analysis . . . . .	13
4.3	Deforested Area . . . . .	17
4.4	Processing Time . . . . .	18
<b>5</b>	<b>DISCUSSION &amp; FUTURE ENHANCEMENTS</b>	<b>18</b>

# 1 INTRODUCTION

## 1.1 Abstract

When it comes to managing rainforests, analyzing the pattern of deforestation is vital. It enables us to monitor both legal and illegal deforestation and its implications, such as climate change through greenhouse gas emissions. Given that there is ample room for improvements when it comes to mapping deforestation using satellite imagery, in this study, Pablo Pozzobon de Bem , Osmar Abílio de Carvalho Junior \* , Renato Fontes Guimarães and Roberto Arnaldo Trancoso Gomes aimed to test and evaluate algorithms belonging to the fields of deep learning (DL), particularly convolutional neural networks (CNNs), to this end. They mapped the deforestation between images approximately one year apart, 2017 and 2018 and between 2018 and 2019. Three CNN architectures —SharpMask, U-Net, and ResUnet—were used to classify the changes and were then compared to machine learning (ML) algorithms—random forest (RF) and multilayer perceptron (MLP). They found that the DL models were better in most performance metrics, with the ResUnet model achieving the best overall results. Visually, the DL models also provided classifications with better defined deforestation patches and did not need any sort of post-processing to remove noise, unlike the ML models, which needed some noise removal to improve results.

## 1.2 Motivation

Deforestation is one of the primary sources of climate change as it is one of the largest sources of greenhouse gas emissions in the world. Within the region of the Brazilian Amazon, it has been observed that deforestation, along with forest fires, can make up to 48% of the total emissions. Locally, estimates have also shown that unchecked deforestation could lead to reductions in seasonal rainfall and into the savannization of the environment [1].The Brazilian National Institute for Space Research (INPE) releases annual deforestation and land use information derived from satellite imagery data through their Program for Deforestation Monitoring (PRODES) and TerraClass projects. This information has been widely used for monitoring, research, and policymaking. Carbon emission estimates from deforestation are also dependent on land use and land-use change data [2] . However, they are likely to be underestimated due to the omission of illegal logging data, i.e, the activity of felling trees and preparing timber.[3]

Coming to remote sensing, i.e., acquiring information about an object (Earth) via satellite imagery, was used for change detection analysis;the process of analyzing and quantifying the state of an object or phenomenon at different times. The changes present in the images can be semantic (of the object under analysis) or noisy (variations in lighting, shadows, among others). Thus, the challenge here is to use a method that establishes features to minimize noisy changes and emphasize the semantic changes that are intertwined. The final result is a binary classification that contains ‘no-change’ and ‘changed’ regions.

Change detection methods based on machine learning algorithms typically use direct classification [4], which takes a set of stacked temporal images as input and uses complex nonlinear functions to model and determine changes. However, it’s not necessary to use

pre-classification techniques that seek to define the best measures to detect changes. On the other hand, deep learning has attracted attention from remote sensing researchers because of its ability to automatically extract features from the image data-set, high-level semantic segmentation, nonlinear problem modeling, and mapping in complex environments [5]. The capacity for pattern recognition in the three dimensions of the image (spatial, spectral, and temporal) makes DL algorithms especially effective when used to change detection with common and recurring patterns [6]. Among the DL algorithms, convolutional neural networks (CNN) are one of the leading types of architectures [7]. CNNs, and deep neural networks in general, are so revolutionary because they take the task of feature extraction out of the hands of human beings. Prior to using CNNs, researchers would often have to manually decide which characteristics of the image were most important for detecting a cat. However, neural networks can build up these feature representations automatically, determining for themselves which parts of the image are the most meaningful.[8]

The objective of this study was to investigate the use of CNNs for the detection of deforestation within the Brazilian Amazon to verify the hypotheses that DL algorithms are a viable and possibly better alternative in comparison to classic ML algorithms when it comes to mapping deforestation. Three different CNN architectures were used to classify deforested areas yearly and were compared to two classical machine learning algorithms.

## 2 LITERATURE REVIEW

The advancements in Computer Vision with Deep Learning has been constructed and perfected with time, primarily over one particular algorithm — a Convolutional Neural Network. A Convolutional Neural Network (ConvNet/CNN) is a Deep Learning algorithm which can take in an input image, assign importance (learnable weights and biases) to various aspects/objects in the image and be able to differentiate one from the other. The architecture of a ConvNet is analogous to that of the connectivity pattern of Neurons in the Human Brain and was inspired by the organization of the Visual Cortex. This network is capable of successfully capturing the Spatial and Temporal dependencies in an image through the application of relevant filters[9].

Every subsequent winning architecture uses more layers in a deep neural network to reduce the error rate [10]. This works for less number of layers, but when we increase the number of layers, there is a common problem in deep learning associated with that called Vanishing/Exploding gradient. This causes the gradient to become 0 or too large. Thus when we increases number of layers, the training and test error rate also increases which seems absurd. ResNet, which was proposed in 2015 by researchers at Microsoft Research introduced a new architecture called Residual Network in order to tackle this. The approach behind this network is instead of layers learning the underlying mapping, we allow network to fit the residual mapping, using a technique called skip connections . The skip connection skips training from a few layers and connects directly to the output. The advantage of this skip connection is that if any layer hurt the performance of architecture then it will be skipped by regularization. So, this results in training very deep neural network without the problems caused by vanishing/exploding gradient.

Fully Convolutional Networks (FCNs) owe their name to their architecture, which is built only from locally connected layers, such as convolution, pooling and upsampling. Note that no dense layer is used in this kind of architecture. This reduces the number of parameters and computation time. Also, the network can work regardless of the original image size, without requiring any fixed number of units at any stage, given that all connections are local [11]. The U-Net architecture is built upon the Fully Convolutional Network and modified in a way that it yields better segmentation in medical imaging. Compared to FCN-8, the two main differences are that U-net is symmetric and that the skip connections between the downsampling path and the upsampling path apply a concatenation operator instead of a sum. These skip connections intend to provide local information to the global information while upsampling. Because of its symmetry, the network has a large number of feature maps in the upsampling path, which allows to transfer information. To obtain a segmentation map (output), segmentation networks usually have two parts :

- Downsampling path is used to extract and interpret the context.
- Upsampling path is used to recover spatial information.

The skip connection here is used to transfer local information by concatenating or summing feature maps from the downsampling path with feature maps from the upsampling path. Merging features from various resolution levels helps combining context information with spatial information [12].

## 3 METHODOLOGY

### 3.1 Train and Test Sites

In this study, the authors had selected three regions within the Brazilian Amazon as study sites. These scenes encompass major deforestation centers that have developed along the “Trans-Amazon” (BR-230) [13],[14] and “Cuiabá–Santarem” (BR-163) highways (Figure 1) [15].

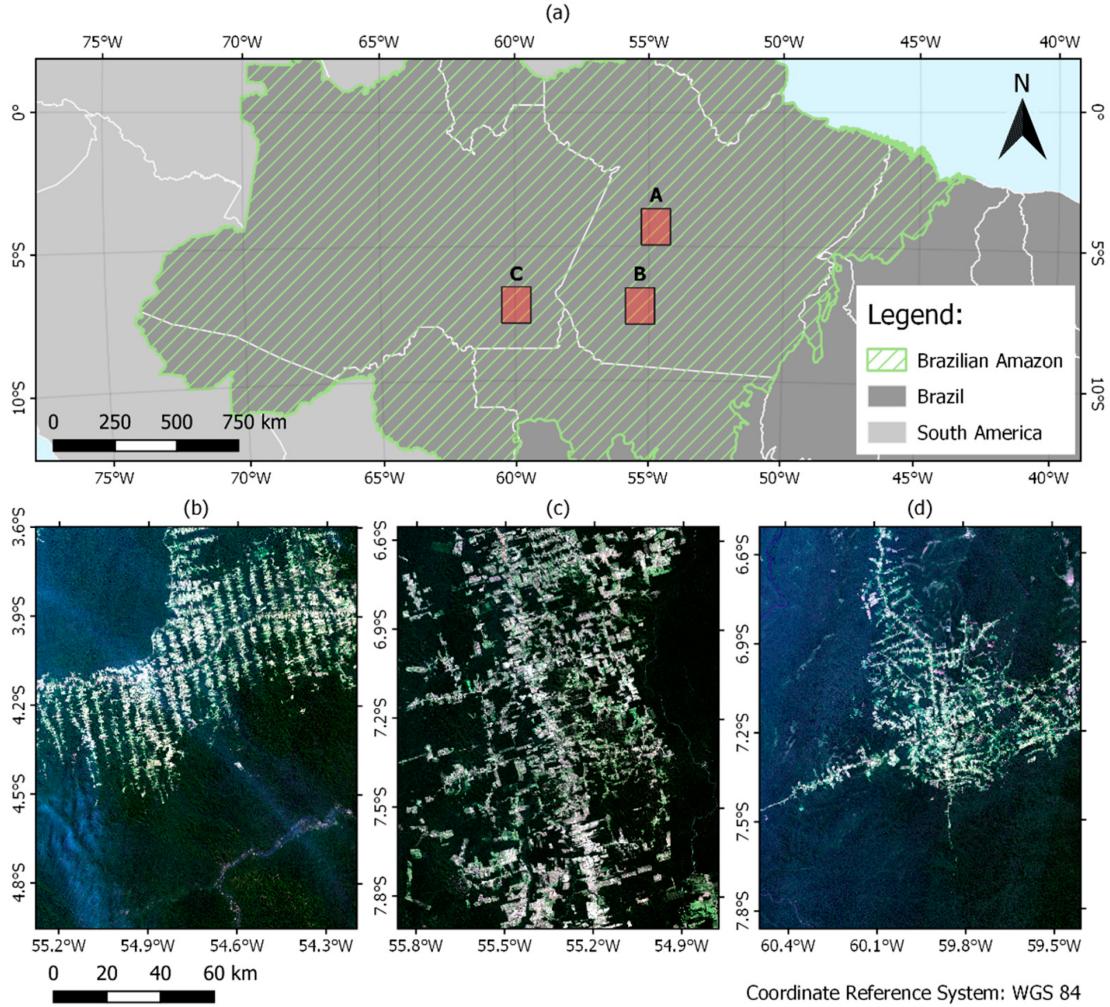


Figure 1: (a) Location of the study sites within the Amazon region with the (b,c) train sites A and B and (d) test site C as true color composite Landsat images taken from June and July 2018.

Construction of roads is the driving force for the spatial distribution of deforestation in the Amazon, with a majority of it around the neighborhood of the main highway [3] as it favors the establishment of settlements, attracts migrants, facilitates the extraction of resources, increases the profitability of livestock and agriculture, and provides access to wood. Landsat 8 is a Landsat satellite that carries the Operational Land Imager (OLI) and the Thermal Infrared Sensor (TIRS) instruments [16]. The training used two scenes (Sites A and B), and validation utilized the remaining scene (Site C). Then, they defined a bi-temporal approach for modeling and obtained Landsat 8/OLI imagery for each site for the years of 2017, 2018, and 2019, with approximately one year between each observation. Multi-temporal images from similar periods of the year reduce variations in the phenology and sun-terrain-sensor geometry. Moreover, images acquired were from the dry season to minimize cloud cover and reduce noise (Table 1).

Site	Landsat Scene	2017	2018	2019 \$
A	227_63	July 18	July 21	July 24
B	227_65	July 18	July 21	July 24
C	223_65	June 21	June 24	July 13

Table 1: Acquisition dates for each site and corresponding Landsat scenes.

### 3.2 Deep Learning Models

U-Net, SharpMask and ResUnet DL architectures have been used in this study. To point out, U-Net and SharpMask algorithms were not developed for classification with remote sensing data in mind, although studies have shown that they offered results that have a leading edge. [17] These algorithms are based on the auto-encoders architecture with the addition of bridges or residual connections. Autoencoders down-sample the feature maps generated through convolutional filters while incrementally increasing their number to learn low-level features compactly, and then up-sample them back to the original input shape for inference. This process is further enhanced using connections bridging the down-sampling and up-sampling steps (Figure 2) to propagate information.

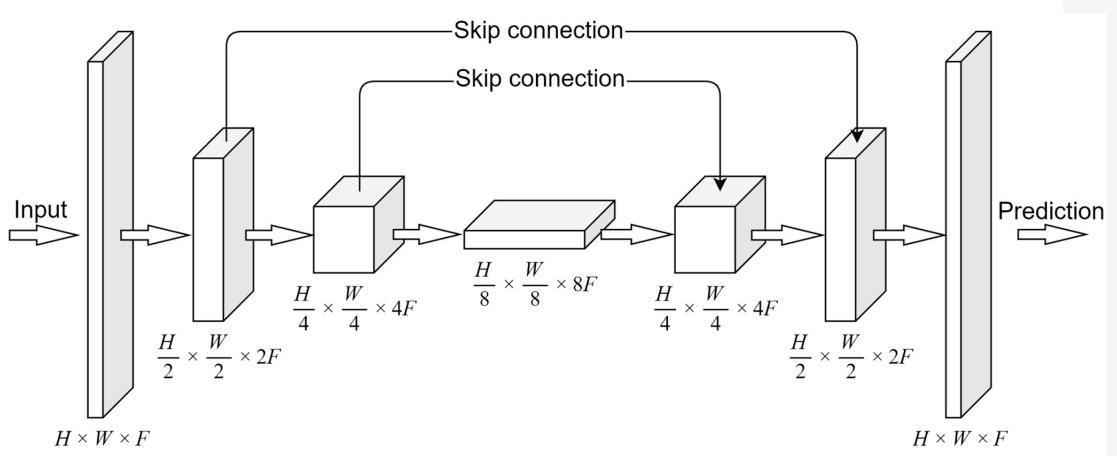


Figure 2: Simple representation of an autoencoder architecture with the addition of skip connections. H, W and F represent the height, width, and number of filtered feature maps, respectively. In this study, H and W are both 200 pixels, while F depends on the specific model architecture.

These connections help speed up training and reduce the degradation of data by combining both low-level detail and high-level contextual information. Low-level spatial detail is essential for change detection and land cover classifications, and that is the main reason behind the choice of this specific type of architecture for this study.

However, these three architectures differ in depth and complexity (Table 2).

Architectures	Layers	Parameters
U-Net	69	19,33,866
SharpMask	114	2,21,386
ResUnet	93	20,68,554

Table 2: Acquisition dates for each site and corresponding Landsat scenes.

Other differences between them :

- U-Net and SharpMask algorithms down-sample the feature maps through a pooling operation, whereas the ResUnet architecture down-samples by using a stride of two between convolutional filter windows.
- U-Net and SharpMask use exclusively long connections (linking the down-sampling and up-sampling sides of the architecture) while ResUnet makes use of long and short connections (between convolutional blocks).

### 3.3 Data Structure

Landsat 8 Operational Land Imager (OLI) and Thermal Infrared Sensor (TIRS) images consist of nine spectral bands with a spatial resolution of 30 meters for Bands 1 to 7 and 9. For this study, dataset consisted only of bands 1 through 7, as they share the same spatial resolution and contain most of the spectral information. Our initial training data was a bi-temporal cube stacking the base image and next year's image, constituting 14 bands. This data structure was maintained for the RF and MLP algorithms, where each pixel is an observation, and each band is a variable. Data sets had to be restructured for the DL algorithms due to the inner workings of the CNNs and hardware memory constraints. To build and train the models in this study, Keras, a high-level wrapper for the well-known Tensorflow library was used. When working with three-dimensional image data, Keras accepts inputs in the form of a four-dimensional array with shape (samples, sample rows, sample columns, channels). To convert our images to the correct format, patches were extracted through 200x200 pixel windows with a 10-pixel overlap on each side (Figure 3). This process generated a total of 844 samples per site per time sequence, with a total of 3376 training samples and 1688 test samples. beginfigure[H]

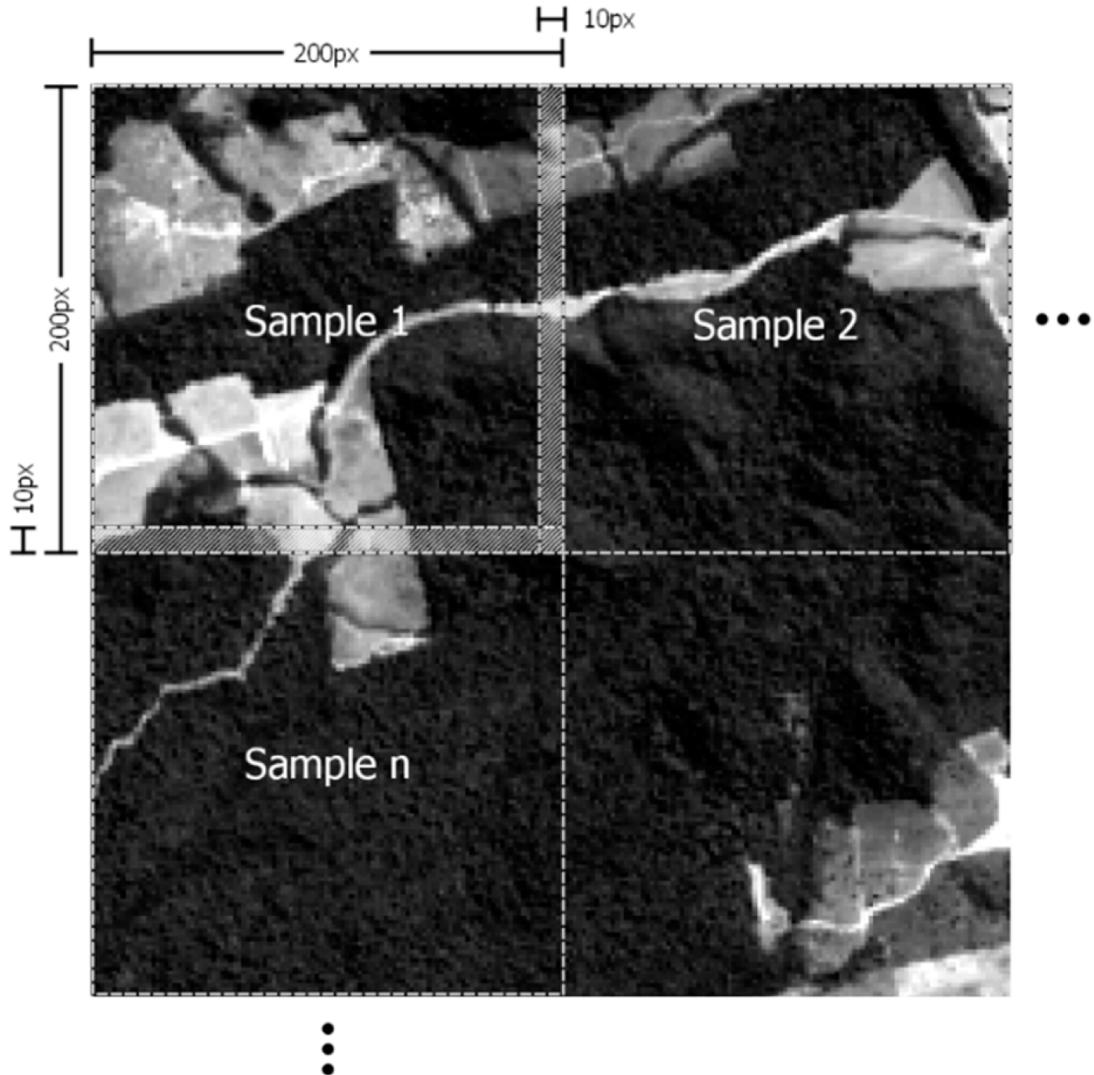


Figure 3: Simple representation of an autoencoder architecture with the addition of skip connections. H, W and F represent the height, width, and number of filtered feature maps, respectively. In this study, H and W are both 200 pixels, while F depends on the specific model architecture.

### 3.4 Ground Truth

To create our ground truth masks, the Brazilian Institute of Space Research's Project for Deforestation Mapping (INPE's PRODES) data [7] for the years of 2018 and 2019 served as a visual guide. It was further refined by remapping the deforestation polygons on a smaller scale. The changes were mapped using digitizing tools from the QGIS software [87] at 1:30,000 scale and subsequently transformed into binary raster files with 0 and 1 as absence–presence codes, respectively. Changes exclusive to the natural forest, regardless of the land cover type in the following year were only mapped (Figure 4).

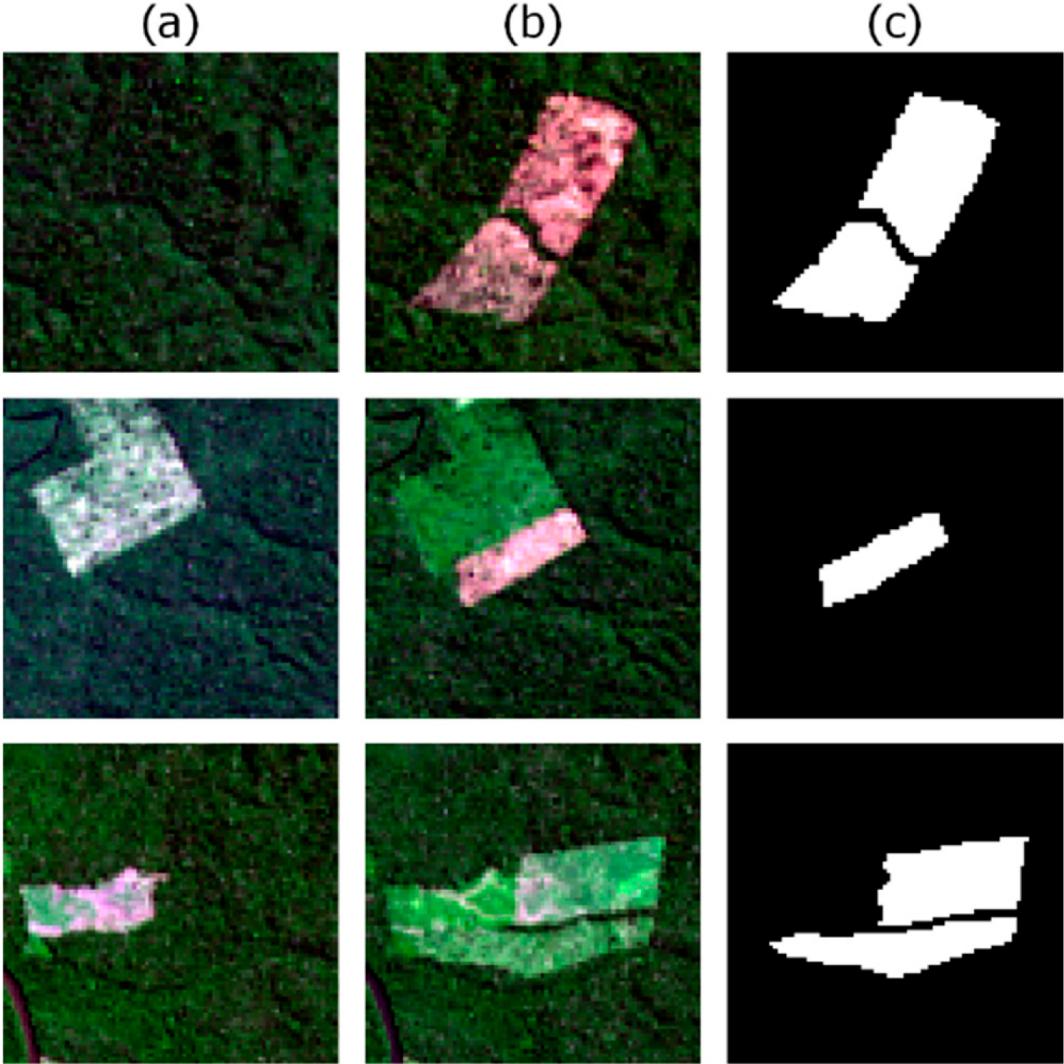


Figure 4: Example of the change mapping in three locations between (a) 2017 and (b) 2018 and the respective (c) rasterized deforestation mask.

### 3.5 Hyperparameters

RF model needed two hyperparameters set, the number of trees to build (`ntree`) and the number of variables randomly sampled as candidates at each split (`mtry`). These were set to 500 trees, and three variables, respectively. The structure of the MLP algorithm consisted of a simple 3-layer network containing an input layer, a hidden layer with 256 nodes, and an output layer. The DL algorithms and MLP shared the same hyperparameters for training. Focal loss was used as the loss function as it excels in classification problems with an uneven number of observations in each class. For gradient descent optimization, adaptive moment estimation (ADAM) algorithm [18] with incorporated Nesterov Momentum (NADAM) with a learning rate of 2e-3,  $\beta_1$  0.9 and  $\beta_2$  0.999 was used. The number of epochs was set to 250 and the batch size to 16 to fit the training process into memory.

### 3.6 Modelling Approach

Given the context of the main methodological steps described in the previous sections, a top-down view of the modeling approach is described in Figure 5.

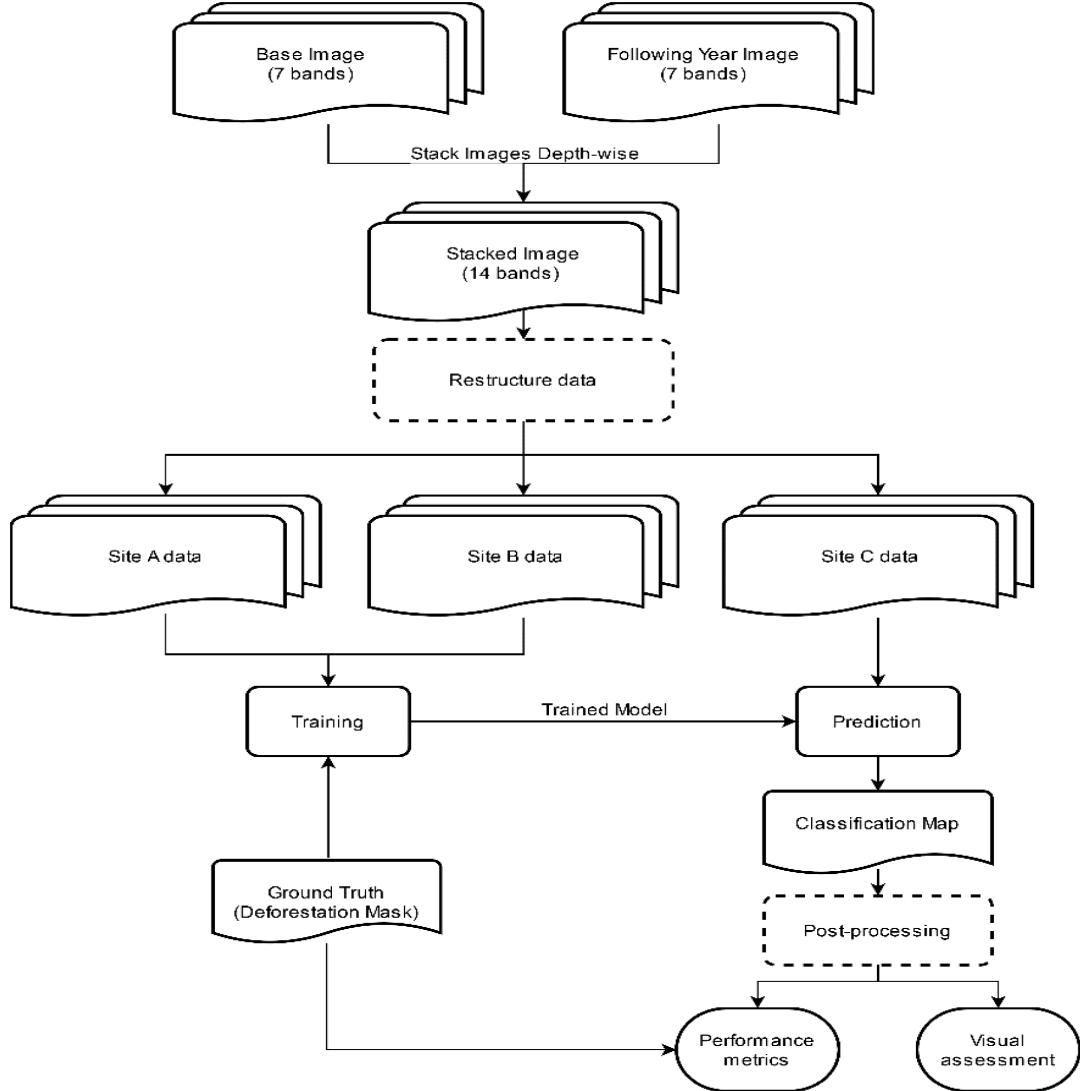


Figure 5: Flowchart of the modeling approach taken in this study.

### 3.7 Accuracy Assessment

The accuracy metrics were calculated using the test site data exclusively to avoid overfitting. To calculate the accuracy measures, classification results were compared to the ground truth mask for the test site. Given that deforestation related change is typically a rare phenomenon, the change–no-change ratio is highly imbalanced [19]. Therefore, change detection research usually shows a predominance of invariant areas, causing a bias in some accuracy metrics. For example, overall accuracy is relatively high on most change maps [20]. The Precision and Recall measures (Equations (1) and 2) were used to offer more insight in the distribution of

errors in the classifications, along with three other measures besides accuracy: F1 score (also, Dice coefficient), Kappa index, and mean intersection over union (mIoU) measure (Equations 3–5, respectively).

$$Precision = \frac{True\ Positives}{True\ Positives + False\ Positives} \quad (1)$$

$$Recall = \frac{True\ Positives}{True\ Positives + False\ Negatives} \quad (2)$$

$$F1 = \frac{Precision \times Recall}{Precision + Recall} * 2 \quad (3)$$

$$Kappa = \frac{p_o - p_e}{1 - p_e} \quad (4)$$

where  $p_o$  is the rate of agreement between the ground truth and the classification, and  $p_e$  is the expected rate of agreement due to chance.

$$mIoU = \frac{IoU_1 + IoU_2 + \dots + IoU_n}{n} \quad (5)$$

where  $IoU$  is the area of intersection divided by the area of union between the classification and ground truth for a class and  $n$  is the total number of classes. Finally, we used McNemar[21] to evaluate the statistical significance of differences between the classifications.

## 4 RESULTS

### 4.1 Quantitative Analysis

Quantitatively, the DL models showed a clear advantage over RF and MLP (Table 4.1). The ResUnet model had the best results with regard to every measure with the exception of Precision in the 2017–2018 time frame. In comparison, the RF model showed the worst results in most measures, although the performance measures still indicated a good classification. Moreover, the RF and MLP classifications exhibited a considerable amount of impulse noise (“salt-and-pepper” type), and a majority filter was applied to reduce the noise and improve the classification both visually and quantitatively. On the other hand, DL models did not require any post-processing steps as they produced classifications with virtually no noise. All models showed very high overall accuracy, but, as explained previously, the ratio between the change and no-change classes is highly imbalanced and is mostly explained by the larger, no-change class. McNemar’s test results indicate that despite the seemingly similar results, the model classifications were all significantly different from each other (Table 4.1).

**Table 3.** Performance measures for the model validation results for the 2017–2018 and 2018–2019 sequences. Best results in the column in bold text.

Model	2017–2018						2017–2019					
	F1	Kappa	mIoU	Precision	Recall	Overall Accuracy	F1	Kappa	mIoU	Precision	Recall	Overall Accuracy
RF	0.8014	0.8003	0.8332	<b>0.9414</b>	0.6976	0.9979	0.8902	0.8892	0.9000	0.8877	0.8928	0.9979
MLP	0.8926	0.8920	0.9024	0.9282	0.8597	0.9987	0.9101	0.9093	0.9167	0.9314	0.8898	0.9983
Resunet	<b>0.9432</b>	<b>0.9428</b>	<b>0.9459</b>	0.9252	<b>0.9619</b>	<b>0.9993</b>	<b>0.9465</b>	<b>0.9460</b>	<b>0.9487</b>	<b>0.9358</b>	<b>0.9574</b>	<b>0.9990</b>
Unet	0.9112	0.9106	0.9179	0.9223	0.9003	0.9989	0.9339	0.9332	0.9373	0.9175	0.9508	0.9987
Sharpmask	0.9223	0.9218	0.9274	0.9173	0.9274	0.9990	0.9337	0.9331	0.9372	0.9218	0.9460	0.9987

**Table 4.** McNemar’s test p-values between model classifications. Values below  $p = 0.05$  indicate the differences between classifications are statistically significant.

	2017–2018					2018–2019				
	MLP	ResUnet	RF	SharpMask	U-Net	MLP	ResUnet	RF	SharpMask	U-Net
MLP										
ResUnet	<0.001					<0.001				
RF	<0.001	<0.001				<0.001	<0.001			
SharpMask	<0.001	<0.001	<0.001			<0.001	<0.001	<0.001		
U-Net	<0.001	<0.001	<0.001	<0.001		<0.001	<0.001	<0.001	<0.001	

Other quantitative observations are as follows :

- RF’s higher precision in the 2017–2018 frame can be understood by the low number of false positives produced. Conversely, it produced a very high number of false negatives within the same frame (Figure 6).
- The ResUnet model had the lowest number of misclassified pixels in both time sequences. It also produced the least number of false negatives out of all the models.
- Considering the number of false-positive cases, the DL algorithms did not show a large difference over the ML models.
- Considering the number of false-negatives, DL algorithms showed a clear advantage. This is a significant advantage given that underestimating the extent of deforestation is a less desirable outcome than its overestimation.

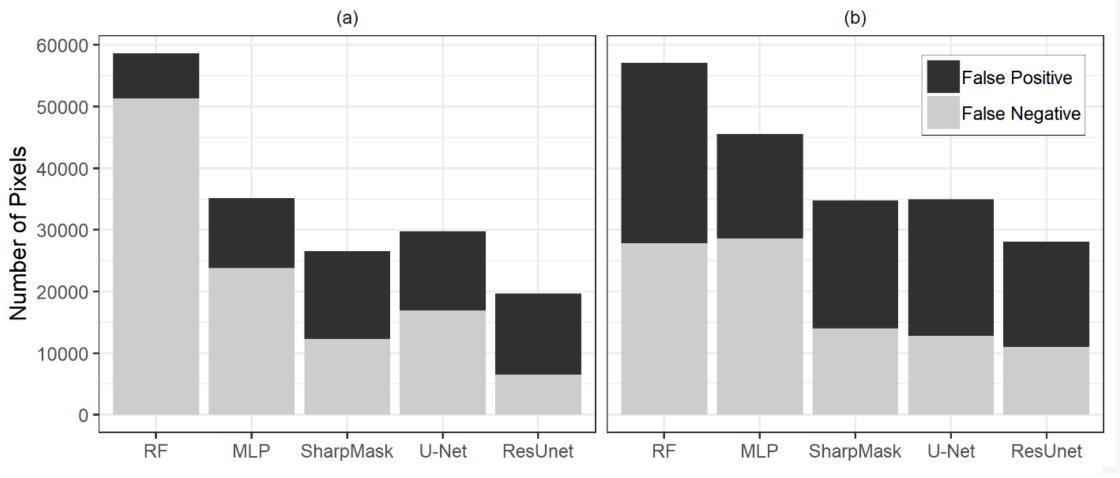
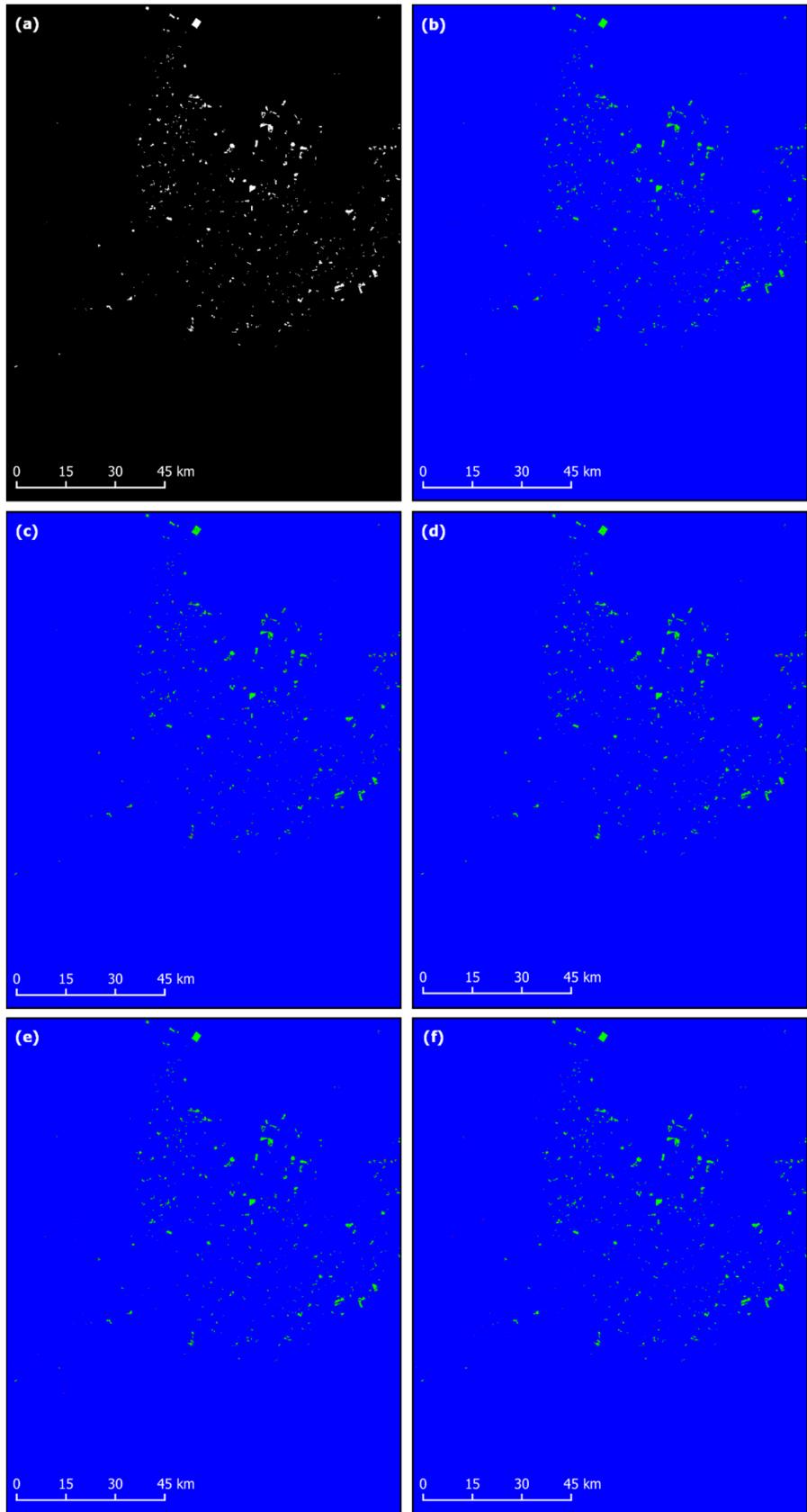


Figure 6: Error distributions in the (a) 2017–2018 and (b) 2018–2019 time sequences in total pixel numbers.

## 4.2 Visual Analysis

The models detected roughly the same deforestation sites at the validation site across both time sequences (Figure 7 and Figure 8). However, the DL models provided more detailed classifications within smaller scales, particularly around feature edges. Moreover, all models were able to classify “easy” deforestation patches with less complex spectral mixtures (Figure 9), but the classification of the ML algorithms degraded as the spectral signatures within the patches increased in complexity (Figure 10).



Legend:

- True Negative
- False Positive
- False Negative
- True Positive

Figure 7: Deforestation masks according to the (a) ground truth and classifications produced by the (b) random forest (RF), (c) multilayer perceptron (MLP), (d) SharpMask, (e) U-Net, and (f) ResUnet models in the 2017–2018 sequence.  
14

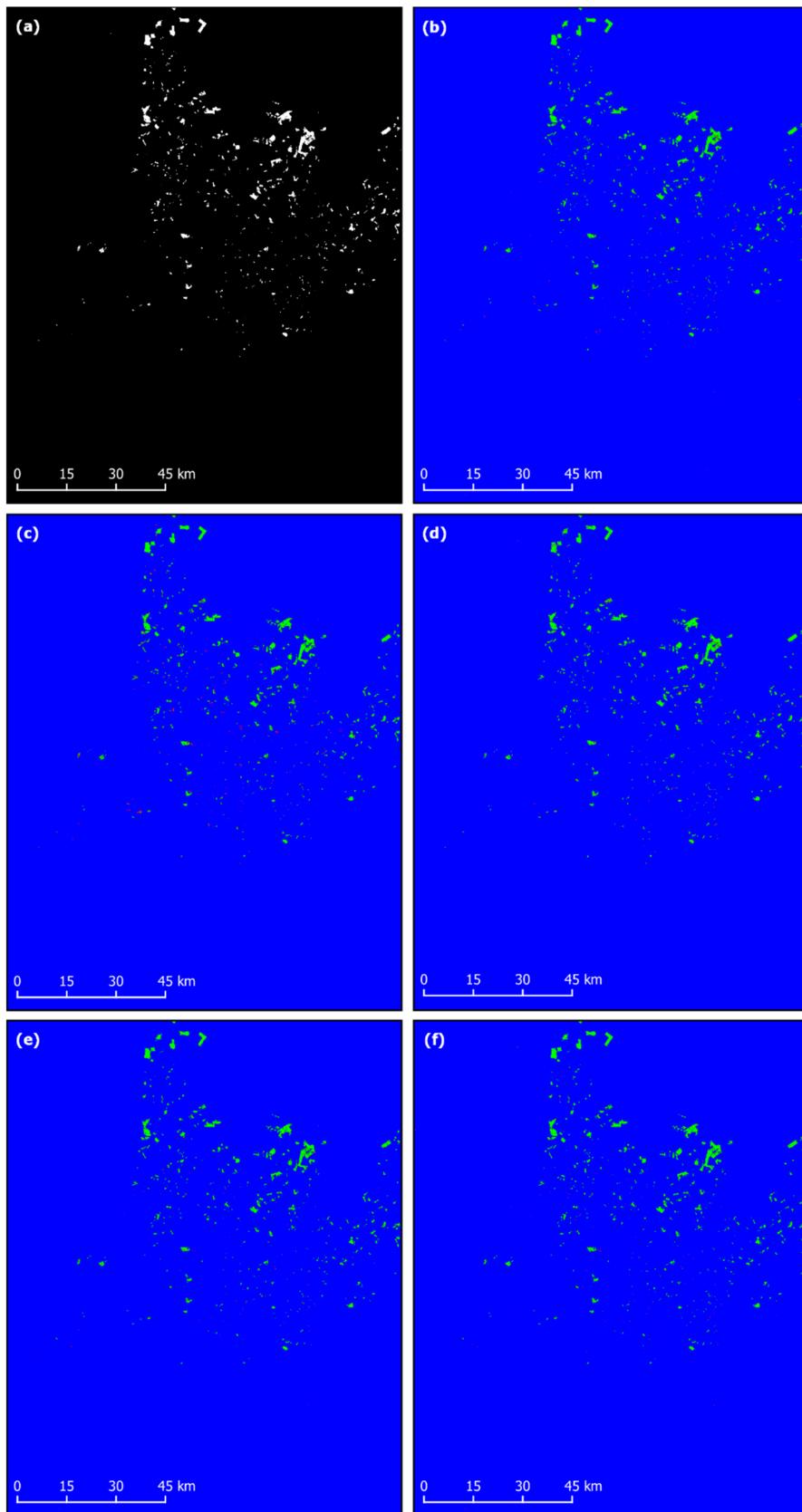


Figure 8: Deforestation masks according to the (a) ground truth and classifications produced by the (b) RF, (c) MLP, (d) SharpMask, (e) U-Net, and (f) ResUnet models in the 2018–2019 sequence.

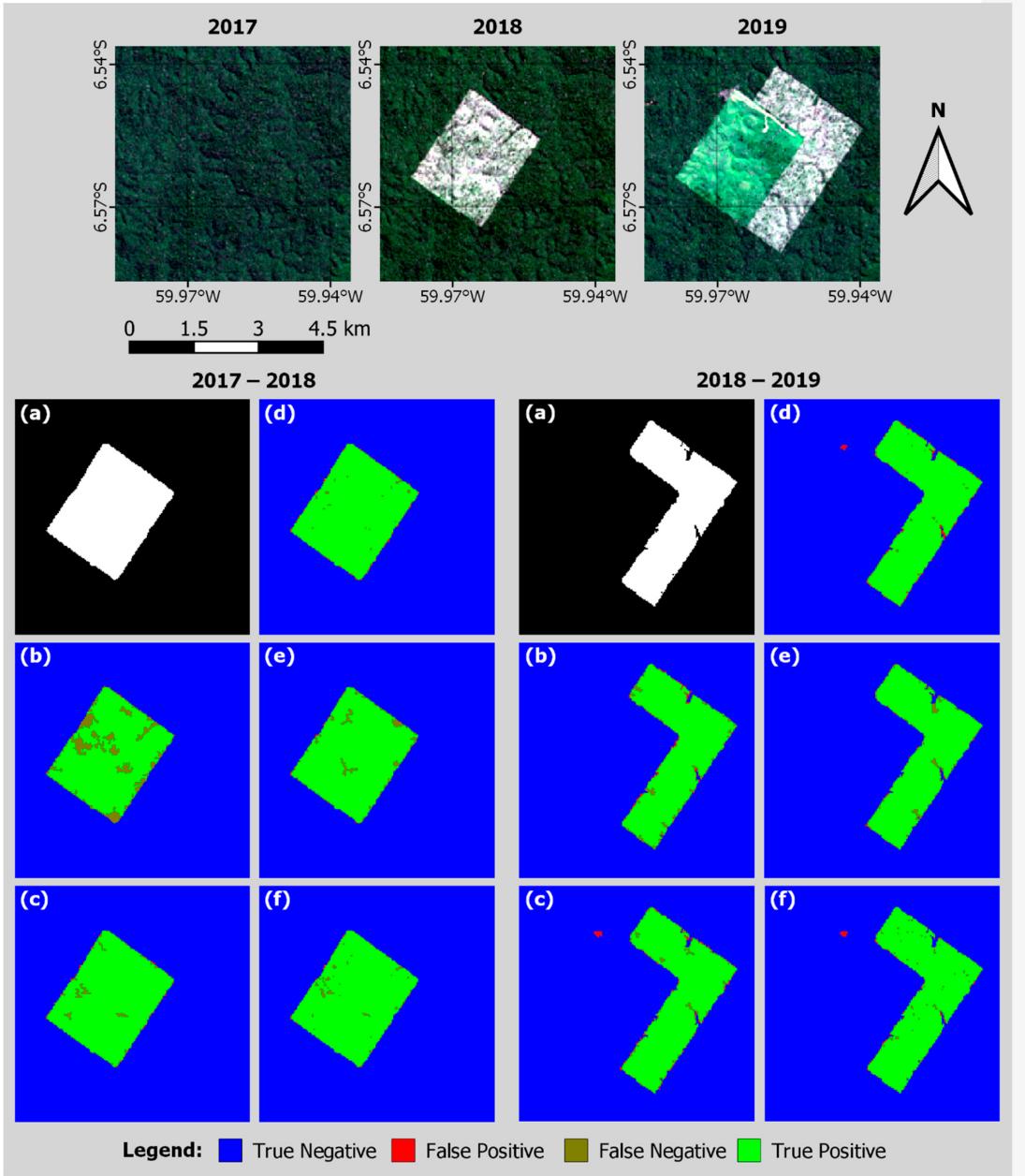


Figure 9: First example location within the test site with the (a) ground truth and classifications made by the (b) RF, (c) MLP, (d) SharpMask, (e) U-Net, and (f) ResUnet models in each time sequence..

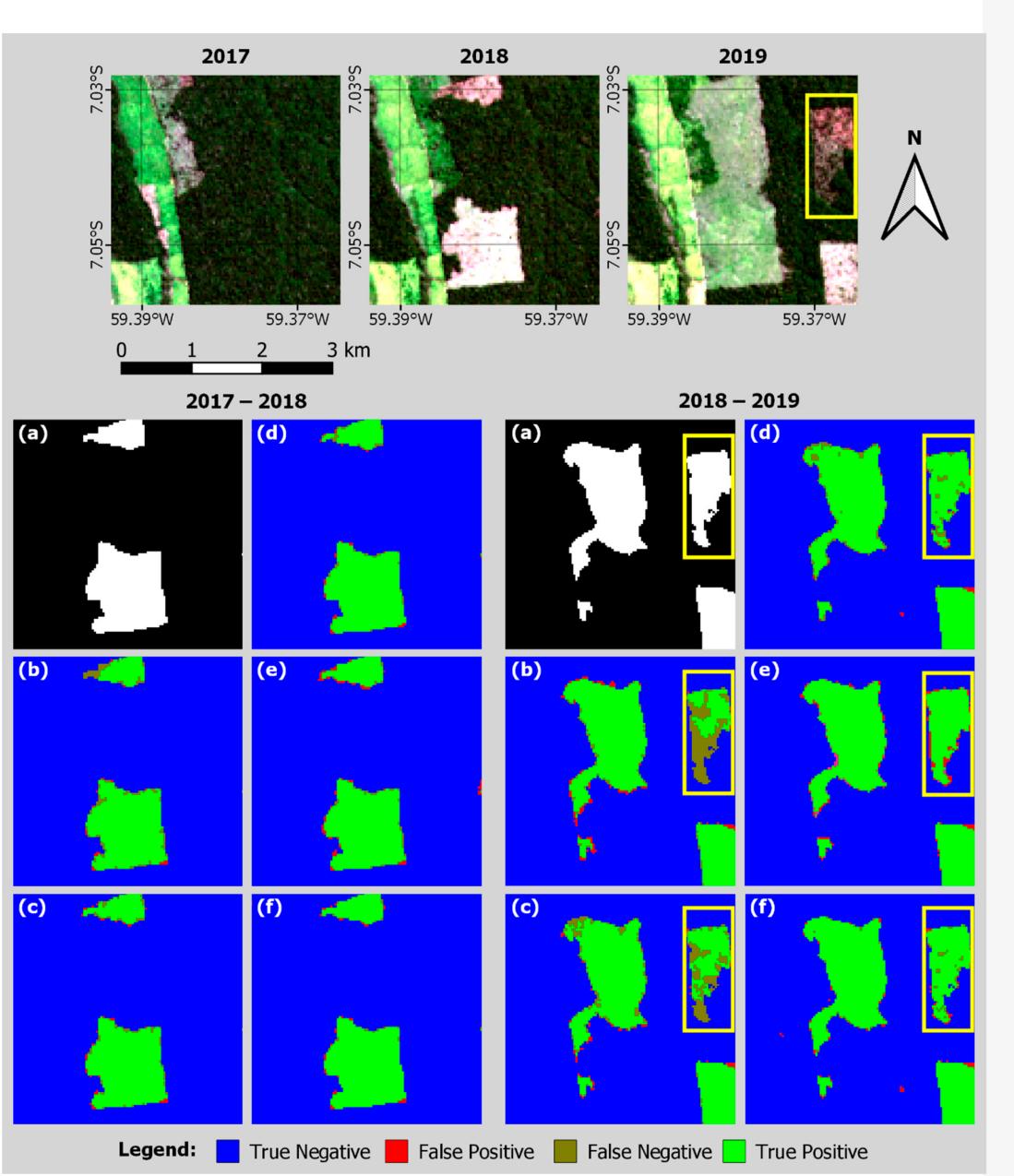


Figure 10: Second example location within the test site with the (a) ground truth and classifications made by the (b) RF, (c) MLP, (d) SharpMask, (e) U-Net, and (f) ResUnet models in each time sequence. The yellow rectangle highlights an example of a “hard-to-classify” deforestation patch.

Overall, RF showed a higher tendency to produce false-negatives both visually and quantitatively.

### 4.3 Deforested Area

The total deforested area was higher than the ground truth in the SharpMask and ResUnet predictions in both time sequences (Fig 11), whereas the opposite was true for the MLP prediction,

as it slightly underestimated the total area in both time spans. RF model underestimated the total deforested area by a very large portion (almost  $40\text{km}^2$  or a 26% decrease in area) in the 2017–2018 sequence due to a large number of false negative predictions).

**Table 5.** Total deforested area according to the ground truth and each model’s prediction.

Reference	Deforested Area ( $\text{km}^2$ )		Difference from Ground Truth (%)	
	2017–2018	2018–2019	2017–2018	2018–2019
Ground Truth	152.73	233.44	—	—
Random Forest	113.17	234.79	-25.90	+0.58
MultiLayer Perceptron	141.45	223.01	-7.39	-4.47
SharpMask	154.40	239.56	+1.10	+2.62
U-Net	149.10	241.90	-2.38	+3.62
ResUnet	158.78	238.84	+3.96	+2.31

Figure 11: Total deforested area according to the ground truth and each model’s prediction.

#### 4.4 Processing Time

MLP and DL models offered faster training and prediction times than RF, due to the fact that the Tensorflow framework uses the computer’s graphical processing unit (GPU) for parallel processing instead of the central processing unit (CPU), that is used for ML. Using an NVIDIA GTX 1070 GPU and a batch size of 16, the total training time ranged from approximately 40 minutes for the simpler MLP model (around 10 seconds per epoch) to almost three hours for the more complex ResUnet model (approximately 40 seconds per epoch). Given the size of the datasets, RF took approximately six hours to train using parallel processing with an Intel Core i5-4690k processor.

The difference in processing times was particularly considerable when using the models to classify the images after training. The DL models and MLP classified the test scene within seconds, whereas RF took almost an hour to complete the task.

### 5 DISCUSSION & FUTURE ENHANCEMENTS

The CNN architectures used in this study showed a clear advantage to the classic ML algorithms, both quantitatively and visually, regarding deforestation mapping. Similarly, a comparative study of methods was developed by [22] for wetland mapping. Although, here it was found that deep learning methods obtained better accuracy than RF and support vector machine. The authors discovered that CNN may produce inferior performance when the training sample size is small.

The difference in performance between the DL and traditional ML methods stems from the former’s capability to understand both the spatial and spectral context, whereas the latter inherently only sees the spectral information (as mentioned in previous sections). While choice and development of architectures for certain targets is a relevant topic for future research,

authors have found that autoencoder networks with residual connections seem to be a good starting point for classifications in remote sensing imagery as they can take advantage of spatial and spectral information in a very efficient manner.

Despite their advantages, DL algorithms are still not as accessible or as easy to use as classic ML models. Besides needing specific hardware for training, they require a relatively large quantity of samples, and developing ground truth masks for specific targets can be challenging (1:1 ratio is desired) and time-consuming in large extents as both spatial and spectral context are strictly needed, whereas the traditional ML algorithms work with simpler sampling schemes and can produce reasonably good results with a much smaller sample size.

Therefore, the process of building a model for broader use (i.e., country-wide monitoring) is complicated. However, these models have another advantage in the fact that they can be incrementally trained, meaning they could be gradually provided with new samples to update the model weights and improve their classifications with time.

This “black box” nature of these networks can make them undesirable for those who might wish to disclose their internal workings such as public and governmental entities. Authors have pointed out that while the models had showed good capability for generalization within their field of study, it cannot be asserted that those models would achieve the same results in different areas where deforestation is a common occurrence. A broader reaching model would necessarily require samples from different regions to account for the possible spatial and spectral variability from one region to another.

Further research should be carried out to study the applicability of the models for similar targets in different areas. In addition, while Landsat data are enough for annual deforestation mapping between dry seasons, more frequent monitoring is virtually impossible as clouds are present above the forest canopy during most of the year, and the ground reflectance cannot reach the satellite’s optical sensors. One solution would be the use of radar data to be able to cross the cloud cover. As such, authors also recommend the investigation of the use of radar data and DL algorithms to detect deforestation within a shorter time-frame.

## References

- [1] J. P. Boisier, P. Ciais, A. Ducharne, and M. Guimberteau, “Projected strengthening of amazonian dry season by constrained climate model simulations,” *Nature Climate Change*, vol. 5, no. 7, pp. 656–660, 2015.
- [2] C. Quéré, R. Andrew, P. Friedlingstein, S. Sitch, J. Hauck, J. Pongratz, P. Pickers, J. Ivar Korsbakken, G. Peters, J. Canadell *et al.*, “Global carbon budget 2018,” *Earth System Science Data*, vol. 10, no. 4, pp. 2141–2194, 2018.
- [3] T. R. Pearson, S. Brown, L. Murray, and G. Sidman, “Greenhouse gas emissions from tropical forest degradation: an underestimated source,” *Carbon balance and management*, vol. 12, no. 1, p. 3, 2017.
- [4] A. P. Tewkesbury, A. J. Comber, N. J. Tate, A. Lamb, and P. F. Fisher, “A critical synthesis of remotely sensed optical image change detection techniques,” *Remote Sensing of Environment*, vol. 160, pp. 1–14, 2015.
- [5] L. Zhang, L. Zhang, and B. Du, “Deep learning for remote sensing data: A technical tutorial on the state of the art,” *IEEE Geoscience and Remote Sensing Magazine*, vol. 4, no. 2, pp. 22–40, 2016.
- [6] L. Mou, L. Bruzzone, and X. X. Zhu, “Learning spectral-spatial-temporal features via a recurrent convolutional neural network for change detection in multispectral imagery,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 2, pp. 924–935, 2018.
- [7] L. Ma, Y. Liu, X. Zhang, Y. Ye, G. Yin, and B. A. Johnson, “Deep learning in remote sensing applications: A meta-analysis and review,” *ISPRS journal of photogrammetry and remote sensing*, vol. 152, pp. 166–177, 2019.
- [8] “Machine learning vs neural networks: Why it’s not one or the other.” [Online]. Available: <https://www.verypossible.com/insights/machine-learning-vs.-neural-networks>
- [9] S. Saha, “A comprehensive guide to convolutional neural networks-the eli5 way,” Dec 2018. [Online]. Available: <https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53>
- [10] pawangfgCheck out this Author’s contributed articles., Pawangfg, and C. out this Author’s contributed articles., “Residual networks (resnet) - deep learning,” Jun 2020. [Online]. Available: <https://www.geeksforgeeks.org/residual-networks-resnet-deep-learning/>
- [11] “U-net¶.” [Online]. Available: <http://deeplearning.net/tutorial/unet.html>
- [12] “Fully convolutional networks (fcn) for 2d segmentation¶.” [Online]. Available: [http://deeplearning.net/tutorial/fcn\\_2D\\_segm.html](http://deeplearning.net/tutorial/fcn_2D_segm.html)
- [13] J. Godar, E. J. Tizado, and B. Pokorny, “Who is responsible for deforestation in the amazon? a spatially explicit analysis along the transamazon highway in brazil,” *Forest Ecology and Management*, vol. 267, pp. 58–73, 2012.

- [14] G. C. Carrero and P. M. Fearnside, “Forest clearing dynamics and the expansion of landholdings in apuí, a deforestation hotspot on brazil’s transamazon highway,” *Ecology and Society*, vol. 16, no. 2, 2011.
- [15] H. Müller, P. Griffiths, and P. Hostert, “Long-term deforestation dynamics in the brazilian amazon—uncovering historic frontier development along the cuiabá–santarém highway,” *International Journal of Applied Earth Observation and Geoinformation*, vol. 44, pp. 61–69, 2016.
- [16] “Landsat missions.” [Online]. Available: <https://www.usgs.gov/land-resources/nli/landsat/landsat-8>
- [17] R. Kemker, C. Salvaggio, and C. Kanan, “Algorithms for semantic segmentation of multi-spectral remote sensing imagery using deep learning,” *ISPRS journal of photogrammetry and remote sensing*, vol. 145, pp. 60–77, 2018.
- [18] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” *arXiv preprint arXiv:1412.6980*, 2014.
- [19] S. V. Stehman, “Sampling designs for accuracy assessment of land cover,” *International Journal of Remote Sensing*, vol. 30, no. 20, pp. 5243–5272, 2009.
- [20] S. Stehman, “Comparing estimators of gross change derived from complete coverage mapping versus statistical sampling of remotely sensed data,” *Remote Sensing of Environment*, vol. 96, no. 3-4, pp. 466–474, 2005.
- [21] Q. McNemar, “Note on the sampling error of the difference between correlated proportions or percentages,” *Psychometrika*, vol. 12, no. 2, pp. 153–157, 1947.
- [22] T. Liu, A. Abd-Elrahman, J. Morton, and V. L. Wilhelm, “Comparing fully convolutional networks, random forest, support vector machine, and patch-based deep convolutional neural networks for object-based wetland mapping using images from small unmanned aircraft system,” *GIScience & remote sensing*, vol. 55, no. 2, pp. 243–264, 2018.