MSc thesis in Geomatics

Integrating radar and multi-spectral data to detect cocoa crops: a deep learning approach

Adele Therias

2023

MSc thesis in Geomatics

Integrating radar and multi-spectral data to detect cocoa crops: a deep learning approach

Adele Therias

June 2023

A thesis submitted to the Delft University of Technology in partial fulfillment of the requirements for the degree of Master of Science in Geomatics

Adele Therias: *Integrating radar and multi-spectral data to detect cocoa crops: a deep learning approach* (2023)

Cover image sources
Satellite data: Google Earth accessed via QGIS XYZ tiles layer connection
Polygon data: Meridia Land B.V

Code access
The data preparation and model training code can be accessed at the following link:
Cocoa Detection Thesis Github repository

The work in this thesis was carried out in the:

GDCM Lab
Delft University of Technology

in collaboration with:

MERIDIA

Meridia Land B. V.

Supervisors:	Dr. Azarakhsh Rafiee
	Dr. Stef Lhermitte
	Philip van der Lugt
	Thomas Vaassen
Co-reader:	Dr. Roderik Lindenbergh

Abstract

The production of cocoa beans contributes to 7.5% of European Union (EU) driven deforestation. For this reason, the recent European Union Deforestation-free Regulation (EUDR) requires producers to perform comprehensive tracking of cocoa farm extents. However, cocoa crops present unique detection challenges due to their complex canopy structure, spectral similarity to forest, variable farming methods, and location in frequently cloudy regions. Previous work employs Multispectral Imagery (MSI) and/or Synthetic Aperture Radar (SAR) for pixel-based classification of satellite images. Convolutional Neural Network (CNN)s offer a promising approach to semantic segmentation of cocoa parcels that considers both spectral and spatial characteristics.

This thesis aims to evaluate the impact of combining SAR and MSI data in the training of a CNN for cocoa detection, in order to demonstrate the importance of texture, moisture and canopy characteristics in identifying cocoa canopies. A U-NET is employed to evaluate how prediction results are impacted by the stacking of MSI datasets with different SAR polarizations, seasons and temporality. The results show that the addition of single-day and temporal SAR to a single-day MSI image can improve the predictions, reaching an F1 score of 86.62%. This research demonstrates the influence of SAR measurement season and polarization, and ground truth classes, on the semantic segmentation of cocoa.

Keywords: cocoa farms, SAR, MSI, CNN, U-NET, semantic segmentation

Acknowledgements

I am blessed to have been surrounded by supportive colleagues, friends and family throughout my entire MSc Geomatics and thesis journey.

Thank you to my supervisor and advisor power-team including Dr. Azarakhsh Rafiee, Dr. Stef Lhermitte, Philip van der Lugt, Dr. Liangliang Nan, and Dr. Roderik Lindenbergh for your ongoing enthusiasm and insightful questions along the way.

Thank you to Jonathan Seipl for suggesting the initial connection with Meridia that made this collaboration possible, and to Thomas Vaassen for being generous with Meridia resources that truly enriched my thesis experience.

I am hugely grateful for the extra technical discussions, inspiration and guidance from Nikolai Kalischek, Dr. David Tax, Ioanna Panagiotidou, Simon Pena Pereira, Tim Straubinger, Naill Ibrahimli and Imke Lansky. Special thank you to Elina Sakketou for in-house design consultations.

Mom, Papa and Emeline: I couldn't have done it without you! Grandma and Grampa, I couldn't ask for better cheerleaders.

Last, but certainly not least, a big thank you to the rest of my lovely family and friends around the world for your support through the ups and downs, your patience as I became a hermit for the last few months, and your confidence that I could, in fact, do this!

Contents

1. Introduction **1**
- 1.1. Motivation . . . 1
- 1.2. Objective and scope . . . 2
- 1.3. Research questions . . . 3
- 1.4. Thesis outline . . . 3

2. Theoretical foundations **5**
- 2.1. Deep Learning . . . 5
  - 2.1.1. Neural Networks . . . 5
- 2.2. Convolutional Neural Networks . . . 6
  - 2.2.1. Convolution layers . . . 7
    - Activation function . . . 8
  - 2.2.2. Pooling layers . . . 8
  - 2.2.3. Training and predictions . . . 8
    - Loss functions . . . 9
- 2.3. Semantic Segmentation . . . 9
  - 2.3.1. U-NET architecture . . . 10
  - 2.3.2. Training labels . . . 11
    - Class imbalance . . . 11
    - Label uncertainty . . . 12
- 2.4. Remotely sensed data . . . 12
  - 2.4.1. Multispectral Imagery . . . 12
  - 2.4.2. Synthetic Aperture Radar . . . 13
    - Backscatter . . . 14
    - Polarization . . . 15

3. Related work **17**
- 3.1. Pixel-based cocoa detection . . . 17
- 3.2. Object-based cocoa detection . . . 19
- 3.3. Semantic segmentation of cocoa parcels . . . 20
- 3.4. SAR for vegetation detection . . . 22

4. Background and context **25**
- 4.1. Study area . . . 25
- 4.2. Cocoa crop characteristics . . . 25
- 4.3. Cocoa detection challenges . . . 26
- 4.4. Ground truth datasets . . . 27
- 4.5. Hypotheses . . . 30

5. Methodology **33**
- 5.1. Satellite data preparation . . . 33
  - 5.1.1. Multispectral Imagery . . . 33

5.1.2. Synthetic Aperture Radar . . . 34
5.1.3. Stacking images . . . 34
5.1.4. Data normalization . . . 34
5.2. Ground truth labels . . . 35
5.3. Data split . . . 35
5.4. U-NET Segmentation . . . 35
5.4.1. Weighted loss function . . . 36
5.4.2. Probability and prediction . . . 36
5.5. Evaluation . . . 37

6. Implementation **41**
6.1. Data preparation . . . 41
6.1.1. Satellite rasters . . . 41
6.1.2. Ground truth polygons . . . 42
6.2. Creating patches . . . 42
6.3. Data split . . . 43
6.4. UNET . . . 43

7. Results and analysis **45**
7.1. Reference training with MSI . . . 45
7.2. Reference training with SAR . . . 47
7.2.1. Single day SAR . . . 47
7.2.2. Multi-day SAR . . . 48
7.3. Experiment results . . . 50
7.3.1. Single day MSI and SAR . . . 50
Single day MSI and single-day SAR . . . 50
Single day MSI and multi-day SAR . . . 53
7.3.2. Multi-day MSI and SAR . . . 55
Multi-day MSI and single day SAR . . . 55
Multi-day MSI and multi-day SAR . . . 57
7.3.3. Overview of Set 1 experiments . . . 59
7.4. Intercrop detection . . . 60
7.5. Set 2 experiments: Impact of labelling . . . 62
7.6. Set 3 experiments: Impact of clouds . . . 65
7.7. Overview of all experiments . . . 65

8. Discussion, conclusion and future work **69**
8.1. Discussion . . . 69
8.2. Limitations . . . 71
8.3. Conclusion . . . 72
8.4. Contribution . . . 74
8.5. Future work . . . 74

A. Metrics **77**

B. Plots and graphs **79**

C. Reproducibility self-assessment **81**
C.1. Marks for each of the criteria . . . 81
C.2. Reflection on reproducibility . . . 82

List of Figures

1.1. Cocoa typologies . . . 2
1.2. Comparison of cocoa prediction rasters from three different methodologies. . . 2

2.1. Neural network . . . 5
2.2. Biological neuron and its mathematical model . . . 6
2.3. CNN basic architecture . . . 7
2.4. Convolution Layer . . . 7
2.5. Graph of ReLU function . . . 8
2.6. Pooling layer example . . . 9
2.7. Convolutional Neural Network . . . 10
2.8. U-NET architecture . . . 11
2.9. Spectral signatures from a cocoa-producing region in Brazil . . . 13
2.10. S1 acquisition modes . . . 14
2.11. SAR types of scattering . . . 15

3.1. Classifier using SAR and MSI datasets . . . 17
3.2. Pixel-based cocoa detection . . . 18
3.3. Visual Comparison of pixel-based and object-based classification results . . . 20
3.4. U-NET for cocoa detection . . . 21
3.5. Visualization of validation batch for agroforest cocoa prediction . . . 22
3.6. Influence of undersampling on segmentation . . . 22
3.7. Related work cocoa prediction map . . . 23
3.8. Evaluation of cocoa segmentation . . . 24
3.9. Importance of SAR polarization in canopy gap predictions [1, p. 23] . . . 24

4.1. Study area map . . . 26
4.2. Cocoa tree characteristics . . . 27
4.3. Low- and Elevated-canopy shade trees . . . 28
4.4. Cocoa parcel size, Meridia ground truth dataset . . . 28
4.5. Satellite imagery showing (1) Monocrop cocoa, (2) Intercrop cocoa and (3) Forest 29
4.6. Examples of overlap between cocoa and forest ground truth. . . . 29
4.7. Satellite imagery showing non-cocoa crops . . . 30
4.8. SAR imagery in different polarizations over study area . . . 31
4.9. SAR backscatter variability . . . 31

5.1. Methodology flowchart . . . 38
5.2. Possible stacking combinations of MSI and SAR datasets. . . . 39

6.1. Seasons of southern Ghana and distribution of datasets over one year. . . . 41
6.2. Imagery included in the MSI temporal stack . . . 41

7.1. Loss curves for MSI-only reference experiments . . . 45

7.2. MSI-only metrics . . . 46
7.3. MSI-only maps . . . 46
7.4. Loss curves for single-day SAR-only reference experiments (May) . . . 47
7.5. Single-day SAR maps . . . 47
7.6. Single-day SAR metrics . . . 48
7.7. Loss curves for multi-day SAR-only reference experiments (multi-season) . . . 49
7.8. Multi-day SAR maps . . . 49
7.9. Multi-day SAR metrics . . . 50
7.10. Loss curves for December MSI + May SAR . . . 51
7.11. Single-day MSI + single-day SAR maps . . . 51
7.12. Single-day MSI + single-day SAR metrics . . . 52
7.13. Single-day MSI + single-day SAR loss box plot . . . 52
7.14. Loss curves for December MSI + multi-season SAR . . . 53
7.15. Single-day MSI + Multi-day SAR metrics . . . 53
7.16. Single-day MSI + Multi-day SAR maps . . . 54
7.17. Single-day MSI + SAR maps . . . 55
7.18. Loss curves for multi-day MSI + May SAR . . . 55
7.19. Multi-day MSI + Single-day SAR metrics . . . 56
7.20. Multi-day MSI + Single-day SAR maps . . . 56
7.21. Loss curve for multi-day MSI + multi-season SAR VV . . . 57
7.22. Multi-day MSI + SAR metrics . . . 57
7.23. Multi-day MSI + SAR maps . . . 58
7.24. Comparison of IoU and Loss for best results from each experiment. . . . 59
7.25. Box plot comparison of loss metric values from original 10-fold experiments. . . 60
7.26. Multi-day MSI + SAR intercrop maps . . . 61
7.27. Single-day MSI + SAR intercrop maps . . . 62
7.28. Comparison of loss from 10-fold experiments with (1) original labels (Cocoa and Forest) and (2) modified labels (Cocoa and Not Cocoa) . . . 63
7.29. Single-day MSI + Multi-day SAR . . . 64
7.30. Comparison of loss from 10-fold experiments with (1) original labels (Cocoa and Forest) and (2) original labels without clouded areas . . . 66
7.31. Comparison of metrics for experiments from Sets 1, 2 (varied non-cocoa class), and 3 (cloudless) . . . 67

B.1. Loss curves for single-day SAR-only reference experiments (January) . . . 79
B.2. Loss curves for multi-day SAR-only reference experiments (dry season) . . . 80
B.3. Loss curves for December MSI + January SAR . . . 80
B.4. Loss curve for single-day MSI-trained model . . . 80

C.1. Reproducibility criteria. . . . 81

List of Tables

5.1. MSI bands used in this research 33

6.1. List of experiments indicating dataset stacks and seasonality 44

8.1. Comparison with related work metrics 71

A.1. Set 1 experiment results: average metrics across all folds 77
A.2. Set 2 experiment results: modified labels with more varied "Not cocoa" class . 78
A.3. Set 3 experiment results: original labels with clouds removed 78

C.1. Evaluation of reproducibility criteria 81

Acronyms

ANN Artificial Neural Network . . . 5
CNN Convolutional Neural Network . . . v
CR Cross Ratio . . . 75
CRIG Cocoa Research Institute of Ghana . . . 26
DL Deep Learning . . . 3
EU European Union . . . v
EUDR European Union Deforestation-free Regulation . . . v
GLCM Grey-Level Co-occurence Matrix . . . 17
GRD Ground Range Detected . . . 22
IoU Intersection over Union . . . 20
IW Interferometric Wide . . . 13
LiDAR Light Detection and Ranging . . . 21
ML Machine Learning . . . 1
MSI Multispectral Imagery . . . v
NDVI Normalized Difference Vegetation Index . . . 18
NIR Near Infrared . . . 12
OA Overall Accuracy . . . 17
ReLU Rectified Linear Unit . . . 8
RMSE Root Mean Squared Error . . . 17
S1 Sentinel 1 . . . 18
S2 Sentinel 2 . . . 17
SAR Synthetic Aperture Radar . . . v
SLC Single Look Complex . . . 72
TCA Tasseled Cap Analysis . . . 75
VH Vertical-Horizontal . . . 3
VV Vertical-Vertical . . . 3

1. Introduction

Forests play a key role in the functioning of ecosystems at local and global scales, providing important services including habitat for over 80% of terrestrial biodiversity and the sequestration of around 289 Gt of carbon [2]. However, forests are also under significant threat due to human activities such as commodity extraction, urbanization, and agricultural intensification, with severe and long-term impacts on ecological and human well-being including rising temperatures, habitat destruction, climate change, pollution and soil degradation [2].

One of the major contributing factors to deforestation is agricultural production for export to the EU [3]. In December 2022, the EU approved the EUDR, which aims to reduce the impact of EU consumption on global deforestation by banning the import of products that are issued from deforested areas, with a particular focus on cattle, wood, palm oil, soy, cocoa and coffee [3]. Due to the due diligence requirements for companies that produce such commodities and derived products [3], the enforcement of this regulation will require highly accurate and timely tracking of farm extents using geodata and advanced geospatial analysis.

1.1. Motivation

Around 16% of the world's forests are located in Africa [4, p. 14], and this continent faces particularly high rates of deforestation. Between 2015 and 2020, Africa had the highest rate of deforestation globally, with a total of 4.41 million deforested hectares of land including 1.90 million hectares in Central and Western Africa [4]. The focus of this research is the detection of cocoa crops, of which West Africa is one of the main producing regions [5] and which are estimated to cause 7.54 % of EU-driven deforestation [3, p. 27].

While many crops can be detected via the automated classification of MSI [5], cocoa presents unique challenges. First, West Africa has frequent cloud cover due to its Monsoon climate, which limits the availability of cloud-free MSI datasets and the temporal resolution of those datasets [6]. Second, agroforestry land cover, a common practice which integrates shade trees and other crops to improve growing conditions, has a spectral signature and canopy structure similar to nearby forest [6, p. 2], and the canopy structure of cocoa can vary widely, as shown in Figure 1.1.

Researchers have aimed to address these challenges by using machine learning Machine Learning (ML) algorithms trained with SAR) and/or MSI datasets to identify cocoa crops. While many of these implementations use pixel-based classifications that do not consider spatial context, recent work has applied a CNN trained with MSI data and shows promising results in Ghana and Cote d'Ivoire [8].

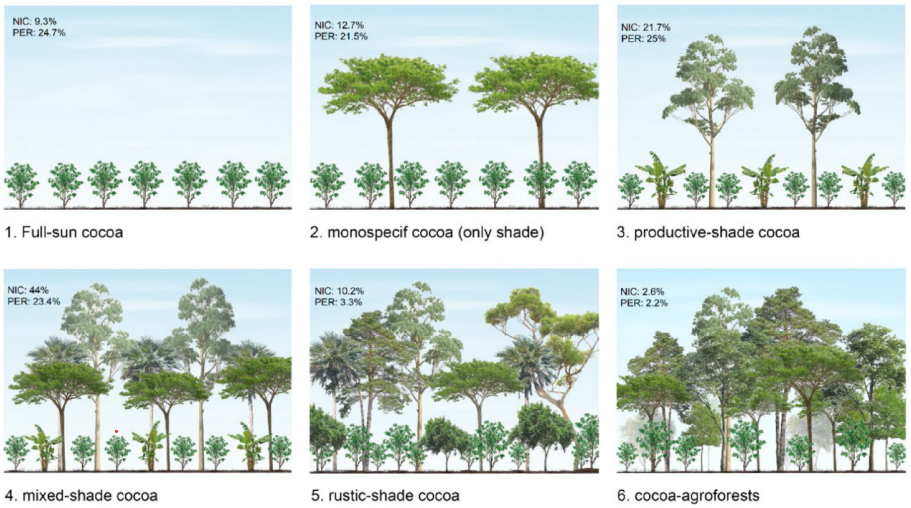


Figure 1.1.: Cocoa typologies

Source: [7, p. 5]

1.2. Objective and scope

The topic of this thesis emerged from a need identified by Meridia Land B.V., a company that works to improve data transparency and traceability in smallholder supply chains [9]. Among other datasets, Meridia makes use of cocoa prediction maps in order to evaluate clients' compliance with the EUDR. In Ghana, "over a quarter of agricultural conversion stems from cocoa expansion" [10, p. 1], and cocoa prediction maps can assist in identifying discrepancies between client farm data and the reality on the ground. However, the existing maps from previous studies often do not align in their predictions, as shown in Figure 1.2. This discrepancy raises questions regarding various maps' reliability and the level of confidence with which they can be used for compliance tracking.

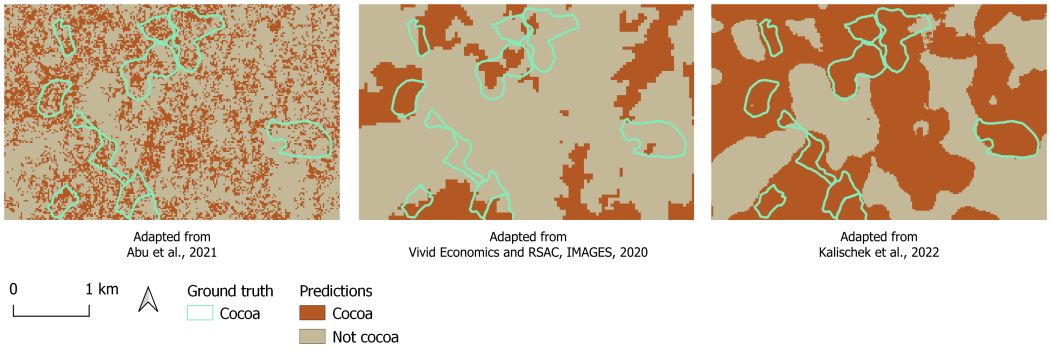


Figure 1.2.: Comparison of cocoa prediction rasters from three different methodologies.

Sources: [5] [11] [8]

Meridia is interested in generating a better understanding of the origin of and differences between existing cocoa maps. It is also in the interest of organizations implicated in the cocoa industry to work towards more accurate detection of cocoa farm extents for the implementation of the EUDR. The objective of this thesis is to build on existing deep learning approaches for cocoa detection by using SAR data in the training of a CNN in order to better understand the impact of inputs on the results of automated cocoa detection. Given the complexity of detecting cocoa crops and the time-sensitive need for reliable maps for the enforcement of the EUDR, this thesis aims to generate insight into the relevance of open-source satellite datasets in order to inform the future development of cocoa detection models in academia and industry.

1.3. Research questions

The research conducted in this MSc thesis builds upon existing work on cocoa detection via ML, with a focus on advancing Deep Learning (DL) approaches. The overarching research question is as follows:
To what extent can a CNN trained with multispectral *and* SAR datasets enable the automated detection of cocoa crops in Ghana?

The sub-questions guiding this research include:

1. How does the combination of MSI and SAR data affect the results of cocoa parcel segmentation trained with data from a single day?
2. How does the combination of MSI and SAR data affect the results of cocoa parcel segmentation trained with temporal datasets?
3. Why does the use of different polarizations (i.e. Vertical-Vertical (VV) or Vertical-Horizontal (VH)) affect the influence of SAR datasets on the cocoa segmentation results?
4. What is the impact of SAR and MSI training data on the detection of intercrop cocoa?

1.4. Thesis outline

The thesis is composed of eight chapters. In Chapter 2, the theoretical foundations for the research are explained including key concepts related to ML, semantic segmentation, MSI and SAR data. Chapter 3 outlines existing research and findings related to ML for cocoa detection. Chapter 4 provides geographical and practical context for the research and hypotheses based on the known study area characteristics. The methodology, including data preparation and experimental set-up, is described in Chapter 5. The implementation tools and procedures are explained in chapter 6 and the experiment results are shown and described in chapter 7. Finally, a discussion of the results, future work and conclusions can be found in Chapter 8.

2. Theoretical foundations

2.1. Deep Learning

ML is a field of research that involves training computers to complete specific tasks autonomously and improve performance through learning and responding to patterns in large datasets [12]. These patterns are learned via networks that enable the computer to make predictions, measure their accuracy and improve its performance iteratively [12]. In many ML applications, humans play an important role in selecting and experimenting with features of the dataset that they expect to be relevant to the learning process. DL is a subset of ML that involves a more complex network with a greater number of hidden layers within the network [12]. Most importantly, humans do not play a direct role in selecting and testing features of the dataset, and instead rely on architecture and hyperparameter modifications in order to improve training results. In the field of computer vision, DL has been enabling unprecedented accuracy in computers' ability to derive information from images [13].

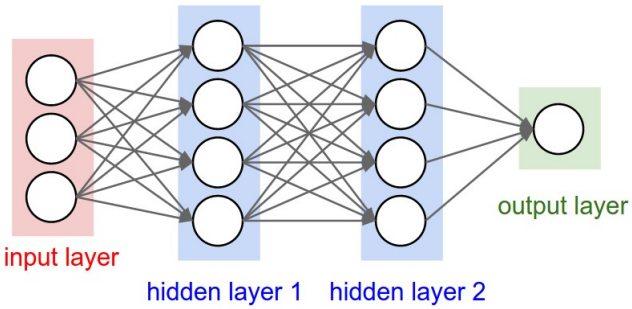


Figure 2.1.: Neural network

Source: [14]

2.1.1. Neural Networks

The foundation of DL is the Artificial Neural Network (ANN), which has been modelled after the human brain. Just as a brain contains neurons that receive information and output signals with varying degrees of influence which inform a person's understanding or action, the ANN contains nodes which receive data, activate, and send the signal on to the next layer of nodes [14]. The complex combination of signals between nodes and layers leads to the performance of a task. For the purpose of machine vision, an ANN takes an image as input, with its pixels fully connected to a series of hidden layers. All neurons in the input layer are connected to all layers in the next layer, and so on. As shown in Figure 2.2, each neuron applies an activation function to the input x and outputs a value y that is carried forward

to the next layer. The final layer contains neurons indicating the probability of an image belonging to each possible class.

Each connection between neurons (as seen in Figure 2.1) represents a weight, which indicates the influence of one node's activation on the next node. For example, nodes connected to pixels in a specific region of the image may have more influence on the next set of neurons than other regions due the magnitude of their weights. The network is initialized with random weights, and is trained by inputting data and its corresponding class label [15]. The network aims to optimize the weights and activations across all layers in order to have its prediction come as close as possible to the true labelled class. By detecting patterns in a large number of images and generating a series of weights and activations for the neurons in the hidden layers, the network can then take a new input image, detect the patterns in its pixels, and output a probability of it belonging to each of the potential classes [15].

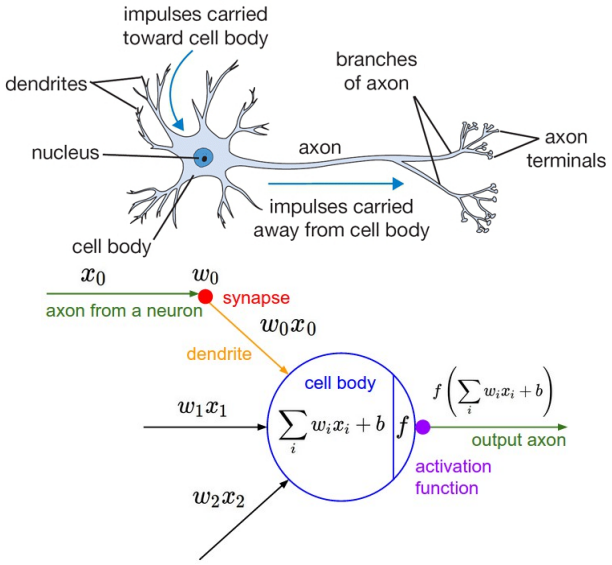


Figure 2.2.: Biological neuron and its mathematical model

Source: [14]

2.2. Convolutional Neural Networks

The challenge with a fully connected ANN is that the number of parameters that need to be computed grows quickly when inputting image datasets, especially when using input images that have multiple bands (e.g. R,G,B, other satellite bands). To avoid the high computation cost, a more appropriate model for image processing is the CNN [16]. A CNN enables the detection of patterns by connecting nodes in one layer with a subset of the nodes in the previous layer, as shown in Figure 2.7. An additional benefit of CNNs is that the 2D kernels representing the weights maintain the spatial relationships between pixels.

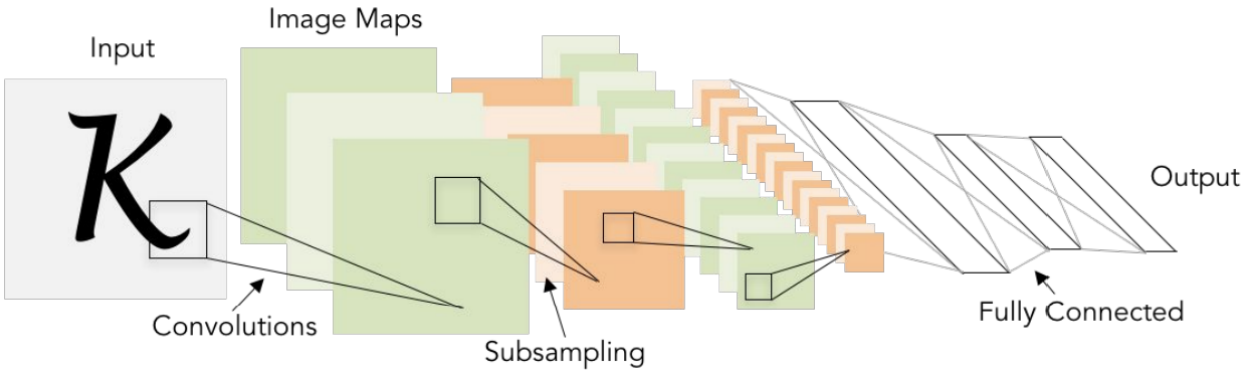


Figure 2.3.: CNN basic architecture

Source: [16]

2.2.1. Convolution layers

A CNN uses a number of filters (matrices containing weights) which represent the relationship between an input and subsequent layers [17]. A convolutional layer involves the application of filters to the input image, each filter leading to an activation map for the image. The activation maps are stacked to form a multi-channel output as shown in Figure 2.4. The output image may change in size depending on the dimensions, padding and stride of the filters. A padded filter enables convolutions to maintain the same input image size, whereas unpadded filters will progressively make the image smaller as convolutions are applied. As the convolutional layers are applied, the features and patterns detected become increasingly abstract. This allows the network to learn both high-level and low-level patterns which inform the classification of an image [17].

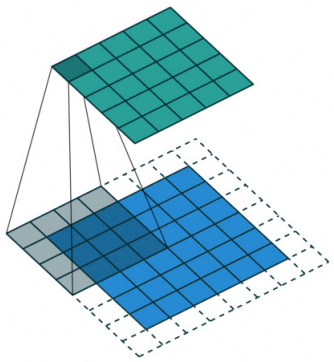


Figure 2.4.: Convolution Layer

Source: [18]

Activation function

In order to detect characteristics in an input dataset, an ANN applies activation functions which activate nodes that meet certain criteria, and de-activate those that do not. This enables the network to focus on the nodes that are deemed "useful" for the task at hand and these nodes are used for further learning in deeper layers of the network. In a CNN, the activation function is generally applied after a convolution layer: activating certain pixels means that those parts of the image will influence the next convolution layer and this helps to reduce the computation cost by discarding the values of irrelevant pixels. There are several possible activation functions for use in CNNs; one of the most common is the Rectified Linear Unit (ReLU) for its computational efficiency and fast convergence [19]. When ReLU is applied, pixels with negative values are assigned a value of zero [20], and the next convolutions focus on the pixels with positive values.

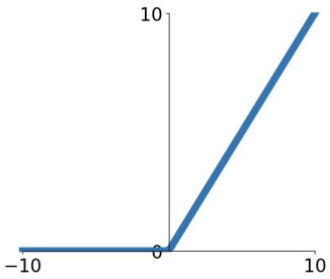


Figure 2.5.: Graph of ReLU function

Source:[16]

2.2.2. Pooling layers

Following convolution layers, pooling layers reduce the dimension of the input image while keeping the important features resulting from the previous activation. Max pooling layers are used to downsample the image and increase computation efficiency due to dimension reduction [16]. The size and stride of the pooling filters influences the size of the output. For example, with a 2x2 max pooling filter, each downsampling step would involve reducing the spatial dimension of the image by a factor of 2.

2.2.3. Training and predictions

The process of training a CNN involves initializing its weights, providing an input image for it to classify (forward pass), comparing the prediction to the ground truth, and updating the weights of the model through backpropagation [21]. This process is repeated for all images in the input training dataset for a designated number of repetitions called "epochs." The predictions at the output of a CNN are numerical values that quantify the probability of belonging to the various classes in the classification. In order to normalize these values between 0-1 and classify the output, another activation function is applied to the output. In

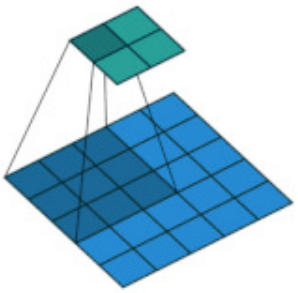


Figure 2.6.: Pooling layer example

Source: [18]

the case of multi-class classifications, one commonly used function is Softmax, which takes the prediction values and forces them into a vector of values between zero and one, which add up to one [22].

$$f_j(z) = -\frac{e^{z_j}}{\sum_k e^{z_k}} \tag{2.1}$$

where z is the input vector containing k real numbers.

In order to assign a specific class prediction based on these probabilities, an Argmax operation is typically applied which identifies the class with the highest probability and assigns the prediction to that class label.

Loss functions

Once a prediction has been made, a loss function is used to compare the network prediction to the ground truth in a quantitative way and guide the learning process. The loss function evaluates the correlation between the model results and the ground truth, and the goal is to minimize the loss value in order to have the predictions be as close to reality as possible [22]. Furthermore, loss functions can be designed or adapted to penalize or encourage certain characteristics of interest. In the case of multi-class classification, categorical cross-entropy loss is a common loss function that focuses on improving the model's predictions by optimizing per-pixel accuracy [23] as cited in [24].

2.3. Semantic Segmentation

In many cases, it is equally as important to identify the location of a feature in an image, not only its class [25]. This technique is called segmentation: it involves pixel-wise classification and simultaneous detection of object instances. Garcia-Garcia et al. (2017) provide an overview of existing datasets and methods which apply DL for semantic segmentation [13]. Most state-of-the-art semantic segmentation is based on the Fully Convolutional Network,

an architecture developed by Long et al. (2014) which replaces fully connected layers with convolutional layers [26] and outputs "spatial maps" indicating per-pixel class labels [13, p. 9]. CNNs designed with segmentation in mind involve upsampling the spatial maps back to "dense per-pixel labeled outputs" [13, p. 9]. This upsampling can take the form of a kind of "reverse max pooling" or "deconvolution filters."

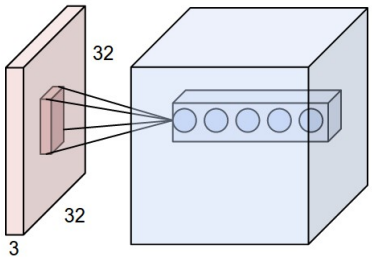


Figure 2.7.: Convolutional Neural Network

Source: [14]

2.3.1. U-NET architecture

One popular CNN used for image segmentation is the U-NET, which is named based on the shape of its architecture as seen in Figure 2.8 [25]. The U-NET is an effective network well suited to land classification, and has been shown to yield improved results compared to the popular **rf!** (rf!) algorithm [27]. Due to its symmetrical shape and use of skip-connections to maintain spatial detail after each convolution, the U-NET outputs segmentation maps at the same resolution as its input images. In a comparison of oil palm segmentation performance, the U-NET was shown to lead to accuracy values over 10% greater than that of AlexNet [27]. Furthermore, the U-NET is a popular architecture due to its ability to reach a high accuracy with a relatively small number of training images [25].

The U-NET takes raw image data as its input, which it carries through two main components: the encoder ("contracting path") and the decoder ("expansive path") [25]. The encoder is responsible for detecting the high and low level patterns in the image as described for the regular CNN above, therefore progressively convolving the image to a higher abstraction. Each encoder block is made up of two convolutional + ReLU activation layers followed by a max pooling layer to downsample the image and increase computation efficiency [25]. Prior to max pooling, the convolution output is set aside and saved for future use as it contains important contextual information that will be used in the decoding process. Each downsampling step involves reducing the spatial dimension of the image by a factor of 2, and a doubling of the feature channels [25].

The final level of the encoder (the bottleneck) includes only two convolution and activation layers and no max pooling. The decoder takes as input the output of the bottleneck and gradually up-convolves the image while bringing back the spatial context that had been saved in each level of the encoder. Each expansive step upsamples the image via a 2x2 convolution and halves the feature channels, concatenates the feature map to the results of the corresponding encoder level, then performs two convolutions [25]. The addition of spatial context allows for the features in the image to be detected and grouped into

segments of a class. The final level of the decoder applies 1x1 convolutions for each class being detected, therefore producing a segmentation map for each class.

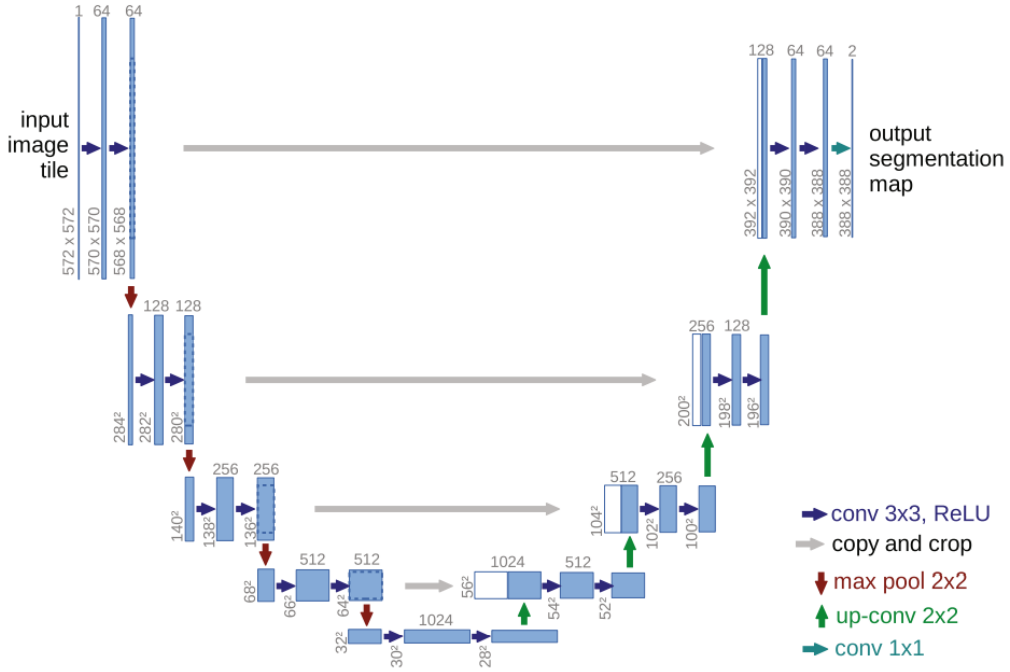


Figure 2.8.: U-NET architecture

Source: [25, p. 235]

2.3.2. Training labels

In order to train a network to predict classes, the training dataset must be accompanied with ground truth labels. The choices and method of preparing ground truth labels have a crucial part to play in the machine learning process. These labels are created to accompany imagery and inform the model of the location of various classes in order for the network to learn the patterns that are most likely to determine the location of a parcel.

Class imbalance

In the case of classification in which classes are not represented equally in the dataset, there is a risk that the minority class cannot be generalized by the model, leading to poor classification results and an overprediction of the majority class. There are several ways to address this imbalance, including undersampling the majority class, oversampling the minority class via data augmentation or modifying the loss function used by the network [28].

Label uncertainty

Another common challenge in creating ground truth datasets is uncertainty. In classification problems, the ground truth is the optimal class allocation which the network is aiming for through the training process, and which serves to train and evaluate the model via the computation of error matrices [29, p. 51]. However, if the reference data cannot be fully relied upon, this risks negatively influencing the effectiveness of the model.

In some cases, there is uncertainty within the labels as a result of errors in ground truth data collection, or as a result of the raster representation of objects, which generally has a lower level of certainty in edge pixels compared to core pixels of an object. In these cases, adding an uncertainty weight to the loss function (computed per class and/or based on pixel location) can help reduce the influence of the more uncertain pixels [28].

In other cases, the uncertainty occurs outside the labelled areas, as partially labelled data can lead to sparse ground truth labels which miss labels for certain classes [30]. This heterogeneity can occur due to incomplete datasets that do not contain labels for all objects in an image: for example, a dataset of cocoa polygons may not contain ground truth for all cocoa farms in a region, therefore they remain unlabelled in the ground truth dataset. Another scenario would be the combination of multiple specialized datasets (e.g. one dataset of labelled forest areas and a separate dataset of labelled cocoa areas) for a multi-class segmentation problem. The cocoa parcels would be missing from the forest dataset, and vice-versa.

One way to address this challenge is to exclude uncertain regions from the training, and assign them to an ”unknown” or ”void” class. While this limits the amount of training data available, it avoids the use of training data that may confuse the model [30].

2.4. Remotely sensed data

CNNs are trained with images, which can be regular RGB photographs, or remotely-sensed images with additional characteristics beyond the visible light spectrum. The following section describes the characteristics of two types of remotely sensed satellite data that are used in this thesis:MSI and SAR data.

2.4.1. Multispectral Imagery

MSI imagery is obtained via passive remote sensing from the energy that is reflected from the Earth’s surface due to the properties of surface objects [31]. The bands of interest are typically detected in the visible, Near- and Shortwave-infrared regions of the electromagnetic spectrum[32]. Different objects reflect the sun’s energy in different ways, which makes it possible to differentiate between them; for instance, the chlorophyll in vegetation causes it to reflect visible light in the green part of the spectrum which leads vegetation to appear green. This phenomenon also occurs in non-visible parts of the spectrum: healthy plants will reflect more Near Infrared (NIR) radiation than their unhealthy counterparts. Combined, these surface characteristics give land cover types unique ”spectral signatures” which refers to amount of radiation they reflect in different parts of the electromagnetic spectrum (see Figure 2.9) [32]. Different vegetation types will often have different spectral signatures, enabling, for instance, the identification of different crops. This becomes more challenging

when nearby land covers have similar signatures, as is the case with cocoa and surrounding jungle [32].

Considering that MSI relies on reflectance from the Earth's surface, reflected radiation must be able to be detected by the sensor. However, in the case of cloud or other atmospheric characteristics, the radiation may be intercepted and may therefore reduce the amount of data that is available from the land surface [33]. Some pre-processing steps can reduce the impact of clouds, such as interpolation to fill data gaps, but this is overall a very challenging aspect of MSI. To address this effect, it is common to rely on temporal data that will increase the probability of collecting data on a cloud-free day.

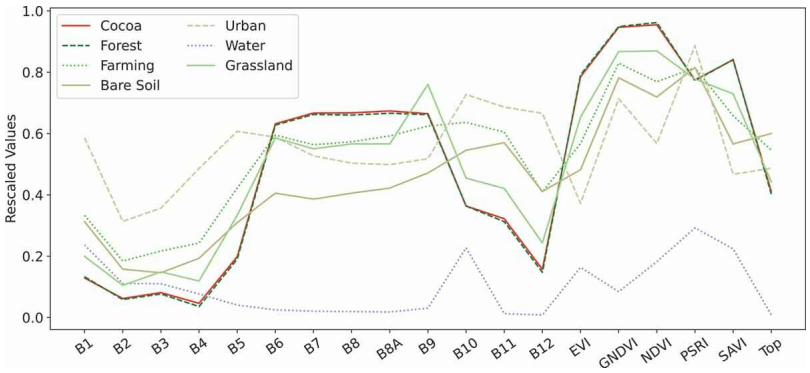


Figure 2.9.: Spectral signatures from a cocoa-producing region in Brazil

Source: [32]

2.4.2. Synthetic Aperture Radar

SAR is an active remote sensing technology that involves sending microwave pulses from the satellite to the surface of the Earth, and receiving back a signal (backscatter) with a certain phase and amplitude. The phase provides information on the distance of the surface objects from the sensor (i.e. height), whereas, as seen in Figure 2.11, the amplitude indicates the intensity of the signal that is returned, varying based on geometry, surface roughness and water content [34]. Considering that it captures data related to surface texture, geometry and moisture, SAR can provide additional information that may further differentiate between vegetation types when optical spectral signatures are not sufficient. Furthermore, it can offer a more consistent source of data that is not affected by atmospheric changes [35] as cited in [33].

Unlike MSI which is captured directly below the satellite, SAR is captured at an angle as the satellite's sensor sends pulses to one side as it orbits the Earth. The most widely used acquisition mode for Sentinel 1 is the Interferometric Wide (IW) mode, which collects data in 250 km swaths at 5 m by 20 m resolution [36], see Figure 2.10. This characteristic of SAR means that the direction of orbit should be considered when comparing SAR datasets [37].

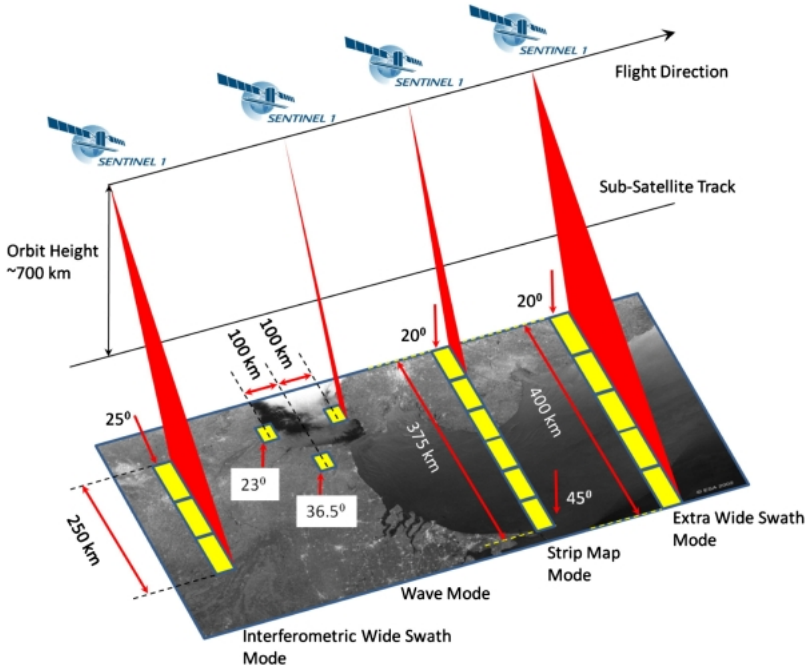


Figure 2.10.: S1 acquisition modes

Source: [36]

Backscatter

Backscatter refers to the amount of energy that is returned to the sensor following the initial pulse and after having interacted with the Earth's surface [34]. There are four main types of mechanisms that affect the intensity of the backscatter:

1. Surface roughness: as the pulse arrives at the surface of the Earth, it interacts with the texture of the ground. A smoother surface will contribute to more specular reflection, reflecting the energy away from the sensor. A rougher surface will cause the pulse to be reflected in different directions, and a portion of the initial pulse energy will be reflected back toward the sensor (see Figure 2.11) [34].

2. Volume scattering: the pulse encountering a vegetated surface will bounce off the objects that it encounters before a portion of the initial energy is returned to the sensor. A less dense vegetation area will cause less volume scattering, meaning that the overall intensity of the backscatter will be higher. A more dense vegetated area, such as a forest, will lead to significant volume scattering with the pulse bouncing off many different branches and leaves before reflecting back to the sensor. This will lead to a lower intensity backscatter [34].

3. Double bounce: when objects on the surface have a structure that is perpendicular to the ground, such as buildings or tree trunks, pulses moving towards such objects and landing on the ground at at certain angles may reflect in a specular way, then bounce on the perpendicular object before returning to the sensor. This double-bounce effect

leads to a very high backscatter value as little energy is lost before the pulse is reflected to the sensor [34].

4. Moisture: combined with the characteristics above, the level of moisture in surface materials affects the backscatter intensity due to the dielectic constants of surface materials. The dielectic constant expresses the ability of a material to store electrical energy. Water has a dielectic constant of around 80, which is much higher than that of a dry soil (around 4) and therefore materials containing water will absorb more energy from the pulse, leading to a weaker backscatter [38].

Furthermore, the size of the wavelength determines the types of surface features that will affect the backscatter: microwaves will penetrate objects smaller than their wavelength, and reflect off objects of a similar size [39].

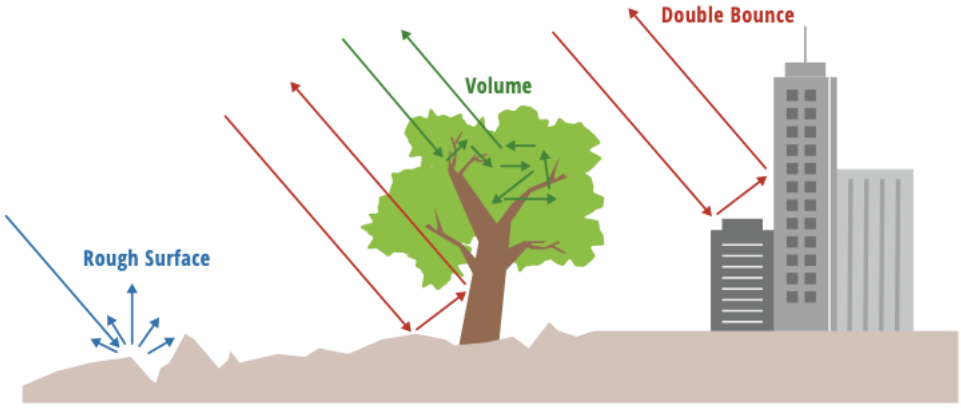


Figure 2.11.: SAR types of scattering

Source: [34]

Polarization

The polarization of a radar pulse refers to the direction in which the wave oscillates in relation to the surface it is imaging (horizontal or vertical) [34]. As shown in Figure **??**, microwave pulses can be sent with Horizontal or Vertical polarization, and the polarization of the pulse that is transmitted and that which is measured upon return can be the same (co-polarization, such as vertically transmitted and vertical received [VV]) or opposite (cross-polarization, such as vertical transmitted and horizontally received [VH]) [34]. The polarization will have an influence on the types of surface characteristics that are detected by the sensor based on how the pulse interacts with the surface. For instance, VV will have the strongest backscatter when interacting with a rough and dry soil surface, HH will have the highest double-bounce backscatter intensity, and cross-polarized (VH or HV) will have the greatest intensity as a result of volume scattering [34].

3. Related work

This section highlights related academic work that employs SAR and/or MSI data and ML to detect cocoa crops and other vegetation types.

3.1. Pixel-based cocoa detection

The majority of existing cocoa detection studies focus on pixel-based classification by implementing ML algorithms and different combinations of datasets. Some researchers have considered only MSI (such as Landsat or Sentinel 2 (S2) data), and applied classification algorithms including Maximum Likelihood Algorithm (Overall Accuracy (OA) = 82.6 %) [40], Random Forest (OA = 89.8 %) [10] and XGBoost, a type of boosted Random Forest (OA = 95.17 %) [32].

Some researchers have focused on classification methods that are based on SAR data only. SAR-based classification has been implemented using Supervised Maximum-likelihood Classifier (OA = 89 %) [33], Random Forest combined with Grey-Level Co-occurence Matrix (GLCM) (OA = 88.1 %) [39], and Multi-Layer Perceptron Neural Networks Regression (Root Mean Squared Error (RMSE) = 7.18 %) [1].

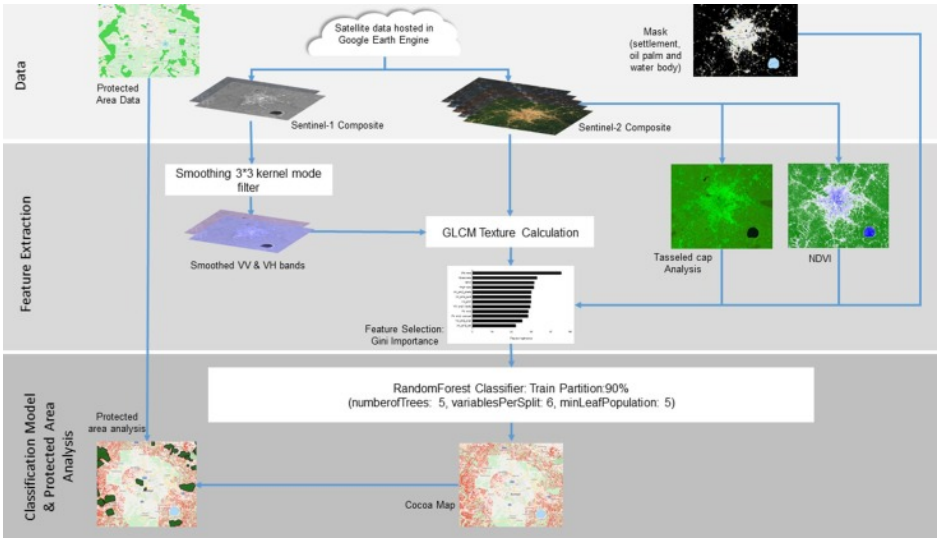


Figure 3.1.: Classifier using SAR and MSI datasets

Source: [5]

Two recent articles perform classification on a combination of SAR and MSI data. In a 2021 study to measure the encroachment of cocoa farms on protected forest reserves, Abu et al. created composite images of each Sentinel 1 (S1) (SAR) and S2 (MSI) datasets, and performed textural (GLCM) and spectral (Normalized Difference Vegetation Index (NDVI), Tasseled Cap index) analyses before applying a multi-feature Random Forest classifier [5], see Figure 3.1. The producer's accuracy and user's accuracy were respectively evaluated at 82.9% and 62.2% [5].

In a 2022 paper, Tamga et al. explore the spatial distribution of classification errors with a focus on cocoa mapping in Ghana and Côte d'Ivoire [41]. The authors use S1 datasets acquired with IW swath mode and with VV and VH polarisations as well as red, green, blue, near-infrared and red-edge bands from S2 multispectral data to perform textural (GLCM) and spectral analyses before applying a multi-feature Random Forest classifier [41]. The main differences between the two studies is that Tamga et al. applied seven different vegetation indices for spectral analysis and they calculated Shannon entropy per pixel to remove pixels with a high probability of error [41]. The higher producer's accuracy (88%) and user's accuracy (91%) compared to the work by Abu et al. is attributed to the fact that this study area is considerably smaller and focused only on a cocoa producing region [41].

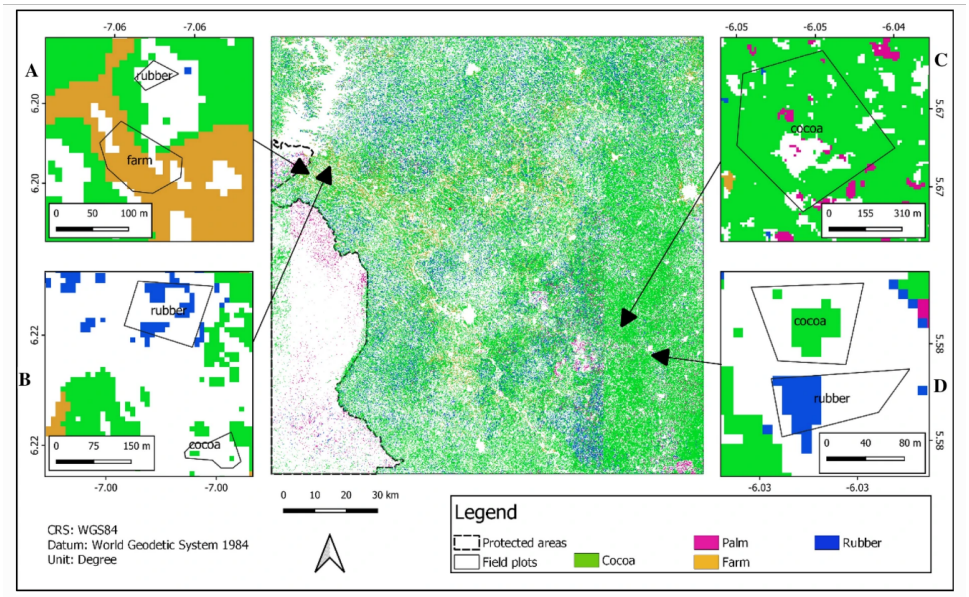


Figure 3.2.: Pixel-based cocoa detection

Source: [41, p. 9]

As can be seen in Figure 3.2, the results of pixel-based classifications can lead to a "salt and pepper" effect, even when smoothing filters are applied [41]. In both articles, the authors conclude that the classification output could be improved with the use of deep learning, with one paper specifically suggesting the use of "semantic image segmentation" [5].

3.2. Object-based cocoa detection

A landscape is composed of class patches, not only pixels [6] which can vary greatly in appearance [42] as cited in [6, p. 3]. Therefore, when classifying land cover, it is important to consider not only the spectral characteristics of pixels, but spatial pattern characteristics within the landscape. This approach can be implemented via algorithms that first detect image objects and then classify them [6].

One early example of such an approach combines optical and dual-polarimetric radar satellite data to map rice and cocoa parcels in Indonesia [43]. The authors detect image objects from a panchromatic dataset, then use a time series of co- and cross-polarized SAR datasets at a resolution of 15 meters and MSI datasets at a resolution of 30 meters for classification of each object [43]. This study makes use of an object-based nearest neighbour classifier and applies it to different combinations of datasets, and the highest OA (89%) is obtained by using MSI *and* cross-polarized SAR data [43].

Another more recent example of object-based classification is a study from 2020 which aims to detect and differentiate between open forest and agroforestry cocoa [6]. The authors first detect image objects from combined SAR and MSI datasets using the Multiresolution Segmentation algorithm, and then apply Random Forest classification to three experimental datasets: MSI only (OA = 79.02%), MSI and SAR (OA = 80.49%), and finally MSI, SAR and image objects (OA = 89.76) [6].

As shown in Figure 3.3, not only is the accuracy of the object-based classification higher, but the visual output is also a more realistic representation of the spatial characteristics of cocoa parcels on the ground. One challenge of object detection is the importance of selecting an object scale that is relevant to the dataset; furthermore, incorrectly defined image objects will cause the mis-classification of all pixels in that object [6, p. 11].

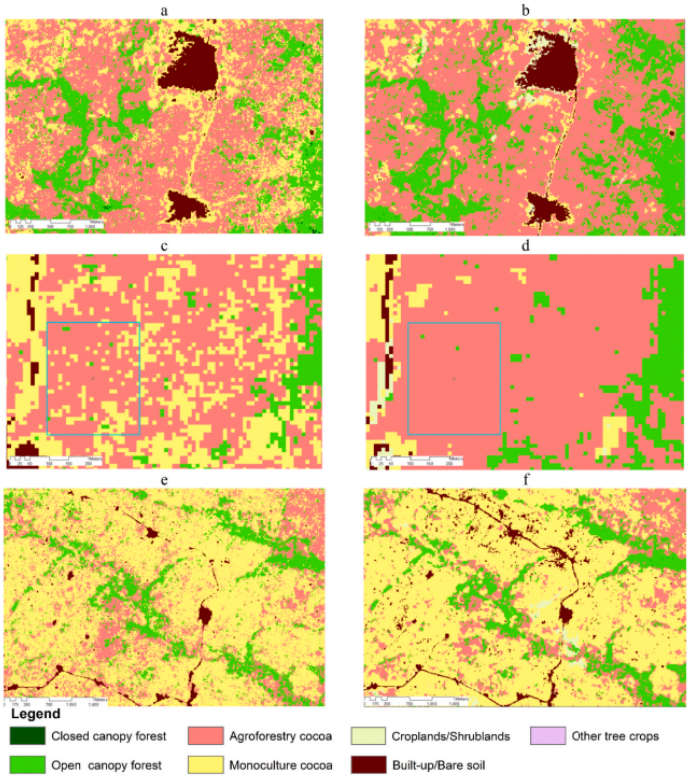


Figure 3.3.: Visual Comparison of pixel-based and object-based classification results

Source: [6, p. 10]

3.3. Semantic segmentation of cocoa parcels

CNNs offer the possibility to detect and classify cocoa parcels by considering spectral and spatial characteristics in a deep learning architecture by automating feature selection [8]. Filella (2018) adapts the U-NET architecture to detect cocoa based on S2 imagery (using bands 2-8, 11 and 12) in Ghana and Ecuador in order to measure the impact of different inputs on the detection of full-sun and agroforestry cocoa [44]. By training the U-NET in Figure 3.4 with MSI images and an Ecuadorean ground-truth polygon dataset (full-sun cocoa), the final recall is 93%, the final precision is 98% , and the IoU is 92% .

Results are less reliable for agroforest cocoa, shown in Figure 3.5, for which the U-NET is trained using 12 farm polygons and a time-series of 4 multispectral images from December 2017 to January 2018. In order to measure the impacts of under-sampling and temporal data on segmentation outputs, the U-NET is trained and validated separately with temporal and non-temporal data, and different levels of undersampling. Non-temporal data provides the worst results across all metrics, with undersampling leading to an Intersection over Union (IoU) of 47.2%. The use of temporal data performs better: the highest IoU of 58.2% is achieved with a batch size of 64 and a minimum of 100 cocoa pixels. Attempting to reduce batch size to increase the proportion of cocoa is not effective: the model predicts forested areas rather than cocoa crops, which is attributed to a lack of background ("non-

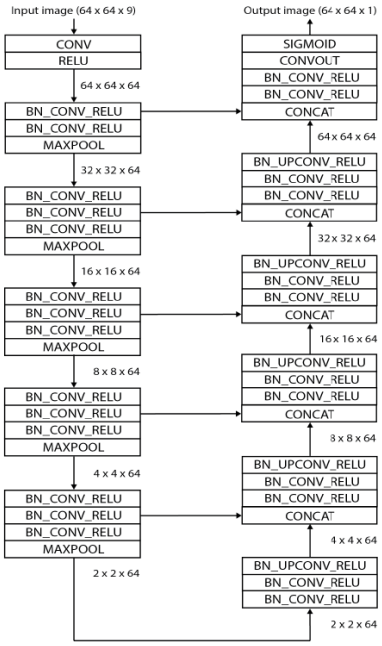


Figure 13: Modified U-Net architecture

Figure 3.4.: U-NET for cocoa detection

Source: [44]

cocoa") labels. While the author concludes that their model is effective for full-sun cocoa segmentation, the study is inconclusive regarding agroforestry cocoa. Some of the key areas of future research mentioned in this paper include training the network with a larger dataset and a longer time series, improving the labelling of non-cocoa data, detecting and processing clouds separately, and implementing a more powerful network.

In another paper by Kalischek et al. (2022), a CNN is trained using 100,000 cocoa farm polygons, 10,000 non-cocoa polygons and a time-series of 10 S2 images collected from each 6-month time period between October 2018 and December 2021. As part of the data pre-processing, cloud-covered samples are marked as "nodata" and the authors choose input patches that are at least 10% labelled. In order to save time and provide additional data, the authors train a deep learning network with S2 and canopy height Light Detection and Ranging (LiDAR) data, then feed the predicted height values into the CNN. The output of this model is a probability map that indicates, for each pixel, the probability of that pixel containing cocoa (between 0 and 1). This probability map can be converted to binary map based on desired level of confidence; in Figure 3.7, 65% probability is used as the minimum threshold for symbolizing a pixel as "cocoa."

The authors create 10 replicas of the network using the same data but with different random initializations. The results of the segmentation are shown in Figure 3.8; compared to the pixel-based classification implemented to map cocoa at a similarly large scale [5], this DL approach improved "precision and recall by more than 26% and 4% respectively" [8, p. 2]. While the implementation of this CNN had good results, it requires the use of time-series

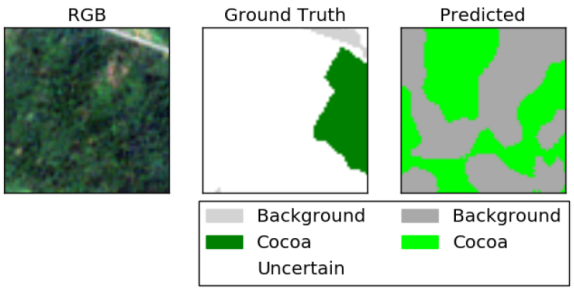


Figure 3.5.: Visualization of validation batch for agroforest cocoa prediction

Source: [44, p. 46]

Minimal amount	Percentage of cocoa	Accuracy	Recall	Precision	Intersection over Union
0	2%	53.2%	0.0%	/	0.0%
100	5%	75.7%	61.1%	99.7%	58.2%
200	7%	49.8%	44.2%	99.6%	44.1%

Figure 3.6.: Influence of undersampling on segmentation

Source: [44, p. 42]

datasets due to the limitations of MSI data when dealing with cloud cover. Therefore, the authors suggest that integrating the use of SAR datasets could allow for monthly or even weekly cocoa mapping updates [8].

3.4. SAR for vegetation detection

SAR has been used in a range of studies related to vegetation monitoring, such as differentiating between forest types [45] and estimating biomass [46]. A higher amount of biomass leads to more volume scattering [1], causing a higher intensity of cross-polarized backscatter [45]. One team of researchers developed a platform that issues deforestation alerts based on S1 VV and VH Ground Range Detected (GRD) datasets: forest disturbances are identified based on a backscatter image from a single day, and are gradually validated or rejected via imagery from subsequent days. This enables the near-real-time detection of forest disturbances of 0.2 ha or larger with a high level of certainty [47].

Beyond a certain amount of biomass, this difference is no longer detectable [1] and spatial distribution of vegetation becomes more relevant. Specifically, SAR is able to detect the presence of gaps in the canopy, where backscatter over bare ground has a high contrast with the surrounding vegetation [1], and tree trunks at the edges of the gaps can cause the double-bounce scatter [1, p. 21]. Numbisi, F. N., & Van Coillie, F. (2020) evaluate the influence of SAR datasets on the prediction of canopy gaps in cocoa agroforestry systems in Cameroon [1]. As shown in Figure 3.9, the importance of different datasets in the Random Forest regression

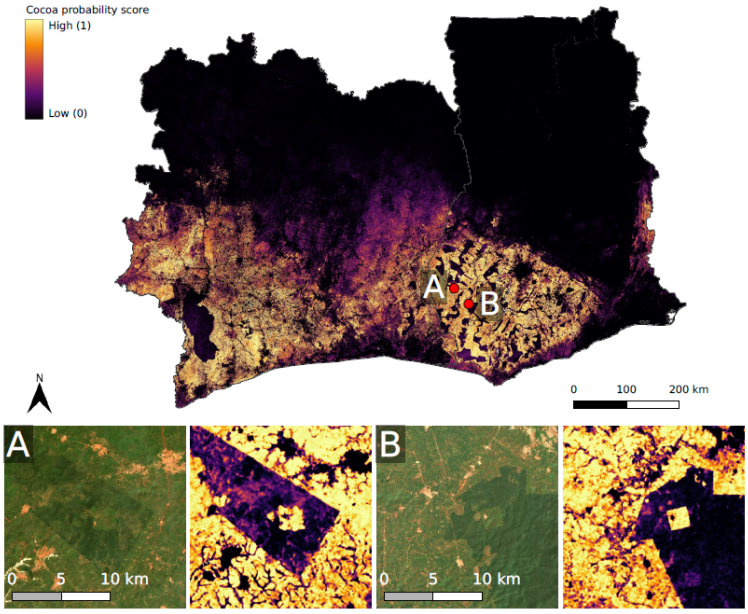


Figure 3.7.: Related work cocoa prediction map

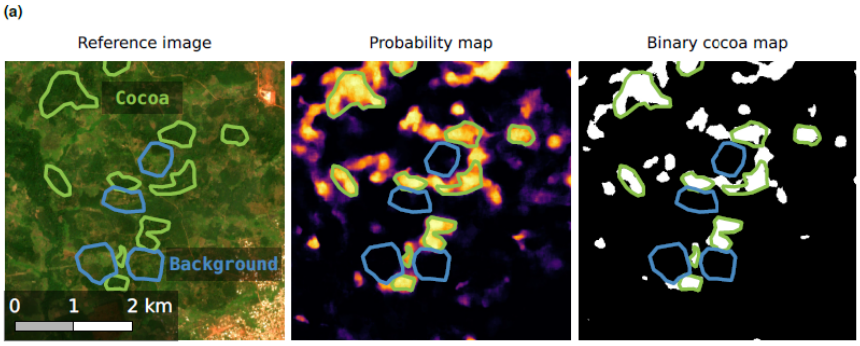
Source: [8, p. 3]

are measured using the Gini importance coefficient, and the VV polarization is found to have the greatest influence on predictions [1, p. 23].

SAR can capture the temporal variation in vegetated land cover and for different vegetation and crop types [46]. More detailed and frequent time series are likely to provide a more precise delineation of crop types, however a land classification with >85% accuracy can be achieved with the use of two polarizations from four days over the course of one year [48]. Numbisi et al. (2018) compare the effectiveness of a Random Forest model trained with temporal SAR and MSI in detecting agroforestry cocoa, Overall, the GLCM derived from SAR from multiple seasons are found to more reliably classify vegetated areas, whereas the MSI datasets led to a higher classification of non-vegetated areas. The authors conclude that temporal SAR provides sufficient information to distinguish between agroforests and transition forests [49], and that GLCM derived from co- and cross-polarization SAR leads to a better classification compared to the use of MSI data with or without indices.

It is important to note that the detection and classification of vegetation depends on the sensor's wavelength, polarization, spatial resolution and incidence angle [1]. While longer wavelength SAR bands, such as L-band (24 cm) are able to penetrate into the understorey and soil of a forested area, C-band has a shorter wavelength of 5 cmand therefore only penetrates into the upper canopy before scattering [1], therefore it is not as commonly used for vegetation monitoring such as tropical deforestation [47]. However, the high temporal resolution of some satellite measurements can compensate for their limited spatial resolution for the tracking of forest disturbances [46].

To conclude this section, existing research has demonstrated the potential for CNNs to detect the extent of cocoa farms based on MSI data; however, neither of the CNN-based cocoa de-



(b)

Class	Precision (%)	Recall (%)	F1 (%)	Accuracy (%)
Cocoa	88.5	87.2	87.3	85.9
Non-cocoa	83.9	84.3	84.1	

Figure 3.8.: Evaluation of cocoa segmentation

Source: [8, p.4]

tection studies found in the literature integrate SAR datasets as input to their deep learning network. Therefore, this research builds on opportunities for further research that integrates a larger ground-truthing dataset in the implementation of a U-NET architecture and evaluates the impact of SAR data on the segmentation of cocoa parcels.

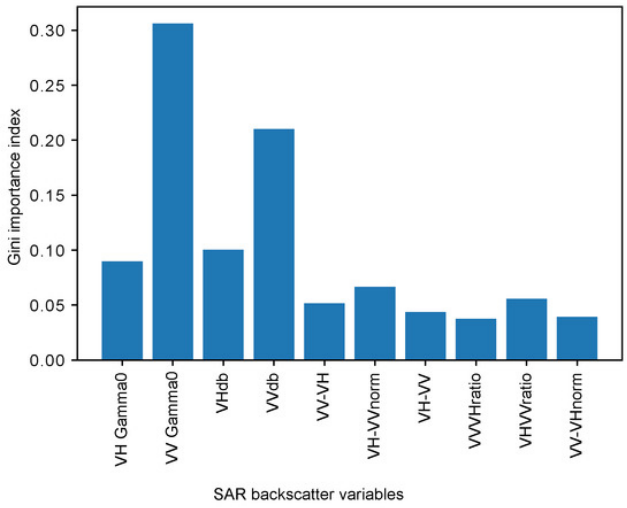


Figure 3.9.: Importance of SAR polarization in canopy gap predictions [1, p. 23]

4. Background and context

The following section provides context on the geography of the study area and cocoa crop characteristics in order to formulate hypotheses that respond to the research questions.

4.1. Study area

This research focuses on detecting cocoa crops in Ghana. The study area is delineated by a 110x110 km S2 tile located in central south Ghana shown in Figure4.1. The tile was selected for having a diversity of land cover, including forest reserves, monocrop cocoa and intercrop cocoa polygons. The region has a semi-equatorial climate consisting of dry and rainy seasons, including a major rainy season (from April – July), and a minor rainy season (from September–October) [50, p. 2]. Considering that cocoa requires a high amount of precipitation in order to grow, these rainy seasons are respectively referred to "main crop" and "light crop" cocoa seasons [50, p. 2]. During the main crop season, the high levels of rainfall, high relative humidity and low temperatures lead to a higher production of cocoa. During the light crop season, the semi-drought conditions lead to a lower productivity and a change in canopy structures as deciduous trees lose their foliage [50].

4.2. Cocoa crop characteristics

After Cote D'Ivoire, Ghana has become the second largest producer of cocoa globally [51]. The cocoa plant is a small evergreen tree which produces pods containing cocoa beans which are dried and processed. It grows best in temperatures between 18–30 °C, above which its photosynthetic abilities are impeded [52]. Unlike many seasonal crops that are fully harvested and replaced from one year to the next, cocoa trees grow and produce over several years, ranging in size from around 1m to a maximum height of around 8m [8] as shown in Figure4.2 [53]. Overall, the plant requires a shaded environment: ranging from 70% shade for trees 0-3 years old, and 30–40% shade for older trees [54] as cited in [52, p. 1]. Cocoa farms are 1-5 hectares on average [52], and considering the lifespan of the plant, a farm can contain trees with a wide range of ages, heights and levels of productivity.

Most commonly, cocoa is grown as a homogeneous crop which integrates large trees that provide shade for the plants, although in recent years there has been a shift towards full-sun cultures [52]. Both shaded and full-sun cultures are referred to as "monocrop cocoa" in this thesis. Alternatively, cocoa can be grown alongside other crops, ranging from distinct plots on a farm, to small plants (e.g. tam, plantain) below the cocoa trees [52] to complex agroforestry: "deliberate integration of regenerated or planted forest or fruit tree species in cocoa farms for ecological and socio-economic benefits" [50, p. 2]. These varieties of mixed-crop practices is referred to as "intercrop cocoa" in this thesis.

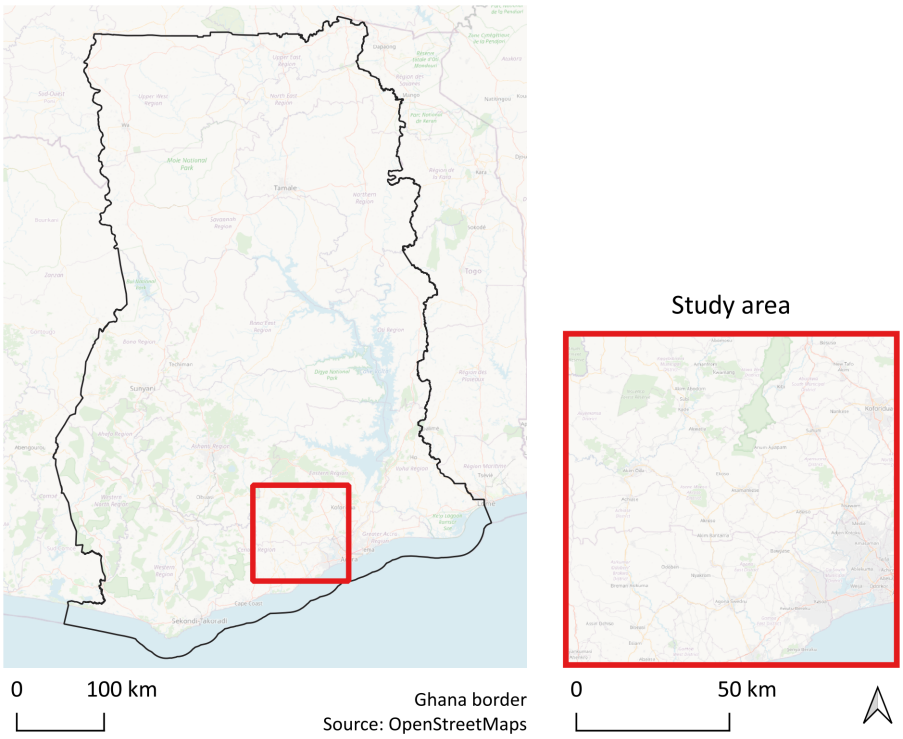


Figure 4.1.: Study area map

Agroforestry offers many benefits for the cocoa crop productivity and the local ecosystem, including enhanced soil fertility, pest and disease management, biodiversity and carbon sequestration [50, p. 2]. The traditional method of agroforestry establishment is to thin the understory of the forest and replace it with cocoa plants [55], but it can also be established by planting diverse and complementary trees and other plants which will grow into an agroforest in the longer term [56].

4.3. Cocoa detection challenges

One of the challenging aspects about detecting and tracking cocoa crop extents is the diversity of practices described above, which limits the definition of cohesive, homogeneous and strictly defined cocoa land cover classes. For example, even within monocrop cultures, the shade tree species used across different farms will vary widely based on the farmers' preferences and the pre-existing ecology of the site, leading to canopies with different tree sizes, morphologies and spectral characteristics [50]. Furthermore, while organizations such as the Cocoa Research Institute of Ghana (CRIG) issue recommendations and best practices, these are not followed unanimously. For example, the CRIG advises farmers to maintain 16–18 shade trees per hectare spaced around 24 m x 24 m apart and having a height above 12 m in order to provide permanent shade to 30–40% of the crown cover [56, p. 2]; however, it is estimated that only 25–42% of the 1.6 million hectares of cocoa crop in Ghana is grown

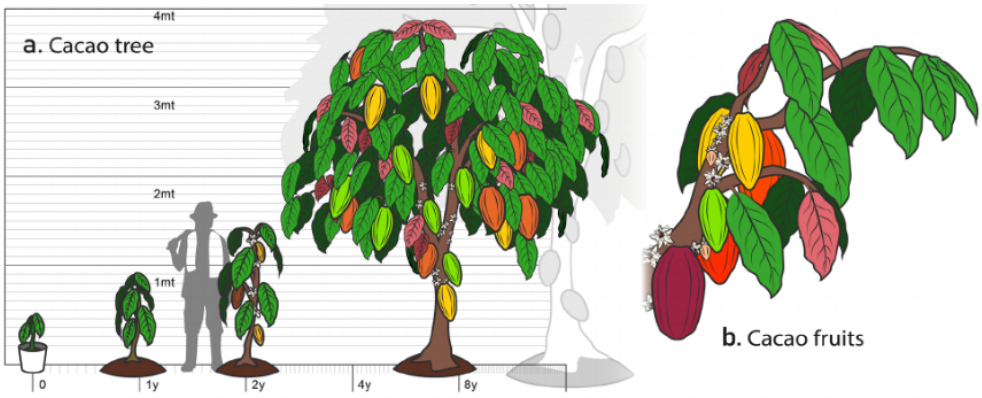


Figure 4.2.: Cocoa tree characteristics

adapted from [53, p. 4]

under sufficient shade [52]. Moreover, official sources recommend maintaining 1730 cocoa plants per hectare, however cocoa plant density in Ghana ranges from 1000 to 2500 trees per hectare [55].

Another challenging aspect is the detection of cocoa in the more complex intercrop cultivation environments, such as agroforestry farms. These farms are generally composed of a variety of shade trees above 5 m tall that make up a stratified canopy that provides at least 10% shade cover [1]. While the practice offers numerous ecological benefits mentioned above, the characteristics are much more challenging to track with the use of MSI data and can be mistaken for transition forests [57] as cited in [49, p. 339].

However, there are some subtle differences that may enable the differentiation between agroforestry cocoa and forest. First of all, forest ecosystems usually have a more dense and intact tree canopy, whereas agroforests generally have a lower density of trees and are more likely to have canopy gaps [1, p. 339]. Another key difference is that while forest canopy remains rather stable over time due to the persistence of evergreen trees in the natural ecosystem, agroforest canopy characteristics exhibit seasonal changes such as the loss of shade tree foliage during the dry season [1, p. 340].

Depending on the farm characteristics, vegetation height can provide additional insight into the nature of different land cover types. In monocrop farms, shade trees can vary in crown shape and height, with some species' canopies partially within the cocoa foliage, and others having crowns that are much higher and distinct from the cocoa trees as shown in Figure4.3 [58]. In the case of agroforestry that is established below pre-existing tropical forest, the height difference between vegetation types can be significant: cocoa grows to a maximum of 8m tall [8], in contrast with tropical forest trees in Ghana which can grow from 5m to over 50m with a mean height of around 20m [59].

4.4. Ground truth datasets

As a result of the collaboration with Meridia, three cocoa polygon datasets were available for training and testing purposes, totalling 88,975 polygons. The data was collected by

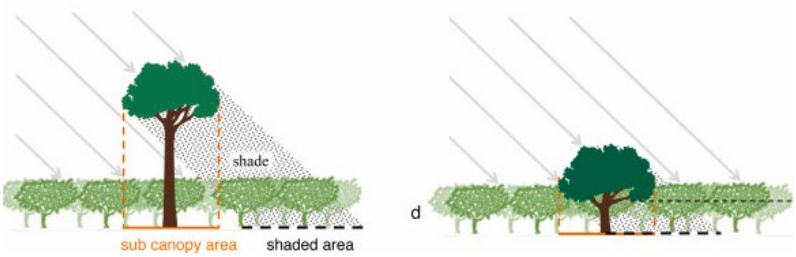


Figure 4.3.: Low- and Elevated-canopy shade trees

adapted from [58, p. 2]

Meridia's field agents in Ghana, who mapped the farms alongside farmers, farm owners or their representatives with an average accuracy of 2m between 2017-2022. The polygons range in size from <1 hectare to >10 hectares, although the vast majority of polygons have an area under 3 hectares as shown in Figure 4.4.

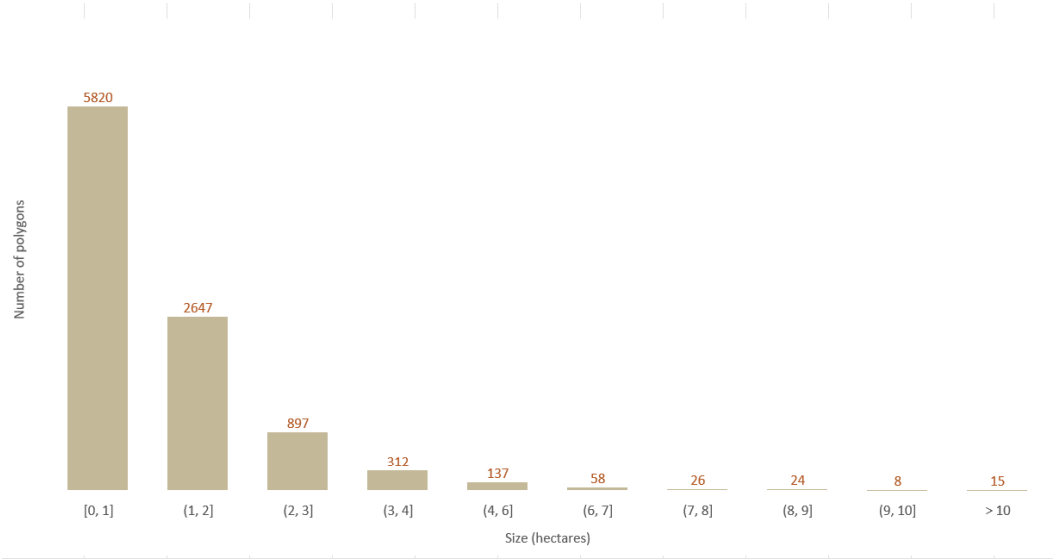


Figure 4.4.: Cocoa parcel size, Meridia ground truth dataset

Each polygon dataset contains attributes that were collected during specific field visits based on the best estimates of farmers or their representatives, such as cultivation type, tree age, density, and farm productivity. The meaning of different attributes was clarified by one of Meridia's data engineers, however there remained some ambiguity in terms of definitions, and some contradictions (e.g. some polygons were indicated as "entire" cocoa cover while listing other crop types grown on the farm). The most consistent and clear attribute across all datasets was that of cultivation type: monocrop, intercrop, sparse, both, or unknown. Farmers could indicate whether their farms were "Mostly mono-cropped cocoa" or "Mostly intercropped cocoa." This wording suggests that the categories are fuzzy, e.g. that a mostly cocoa-producing farm may contain a small section of other crop(s) as shown in Figure 4.7, therefore it is important to note that there is a degree of labelling uncertainty in the dataset.

For the purposes of this research, only cocoa polygons labelled as monocrop or intercrop and located within the study area were considered for training purposes.

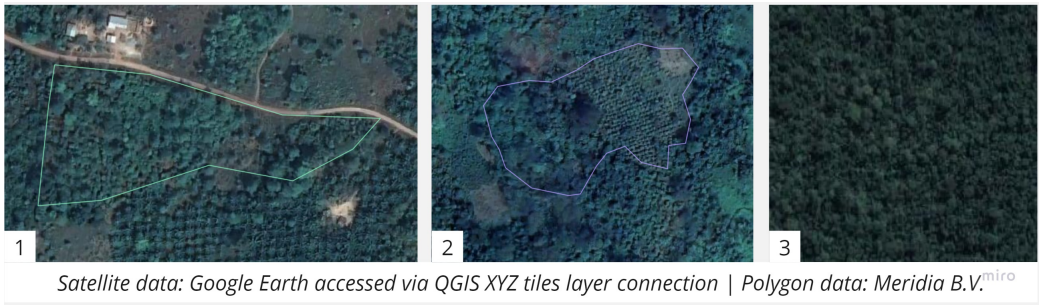


Figure 4.5.: Satellite imagery showing (1) Monocrop cocoa, (2) Intercrop cocoa and (3) Forest

Meridia also provided access to a 2019 forest reserve dataset issued from the Ghana Forestry Commission, as well non-cocoa crop polygons. While forest reserves are technically protected under national, regional or international agreements, a reserve designation does not guarantee that such areas are undisturbed forest as this is dependent on the enforcement of the protected area. Cocoa producers are increasingly encroaching on protected areas due to environmental and social pressures [5], [8]. As can be seen in Figure 4.6, numerous cocoa polygons in the Meridia datasets overlap with protected area polygons. This is an important indication of the limitation of using forest reserves as forest ground truth labels. However, considering the importance of distinguishing between forest and cocoa and the limited time to create more accurate forest labels, the labels were deemed sufficient for this thesis research.

Figure 4.6.: Examples of overlap between cocoa and forest ground truth.

The non-cocoa crop polygons in the study are varied, including tree crops (cashew, oil palm, mango) and smaller crops (maize, cassava and beans). Some of these crops are can be recognized relatively easily through satellite observation as shown in Figure 4.7, which enables a qualitative analysis of the labelling data and the model predictions.

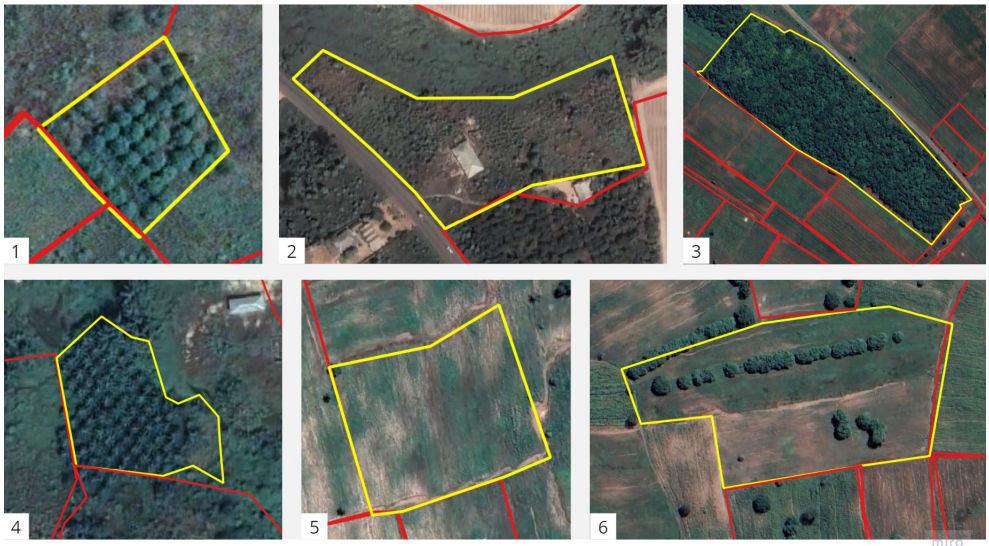


Figure 4.7.: Satellite imagery showing non-cocoa crops

(1) Mango, (2) Yam cassava, (3) Cashew, (4) Oil palm, (5) Beans and (6) Beans + mango

4.5. Hypotheses

Overall, the use of SAR is expected to provide more clear differentiation between land types and across seasons, as shown in Figure 4.9. Examples of the GRD S1 SAR imagery are shown in Figure 4.8. Considering the characteristics of cocoa, results from related work and the features of S1 SAR data:

1. The combination of SAR and MSI is expected to lead to improvement in monocrop cocoa predictions compared to training without the use of SAR.
2. The addition of SAR temporal stacks covering the wet and dry seasons is expected to improve the results in comparison with MSI + single day SAR training.
3. The addition of SAR is expected to enable the detection of cocoa in intercrop polygons where MSI-training would predict forest.
4. The addition of SAR is expected to improve predictions in cloudy images because of SAR's ability to penetrate cloud cover.
5. Combining both polarizations of SAR is expected to yield better results than a single polarization as it will capture the effects of volume scattering, surface scattering and double bounce.

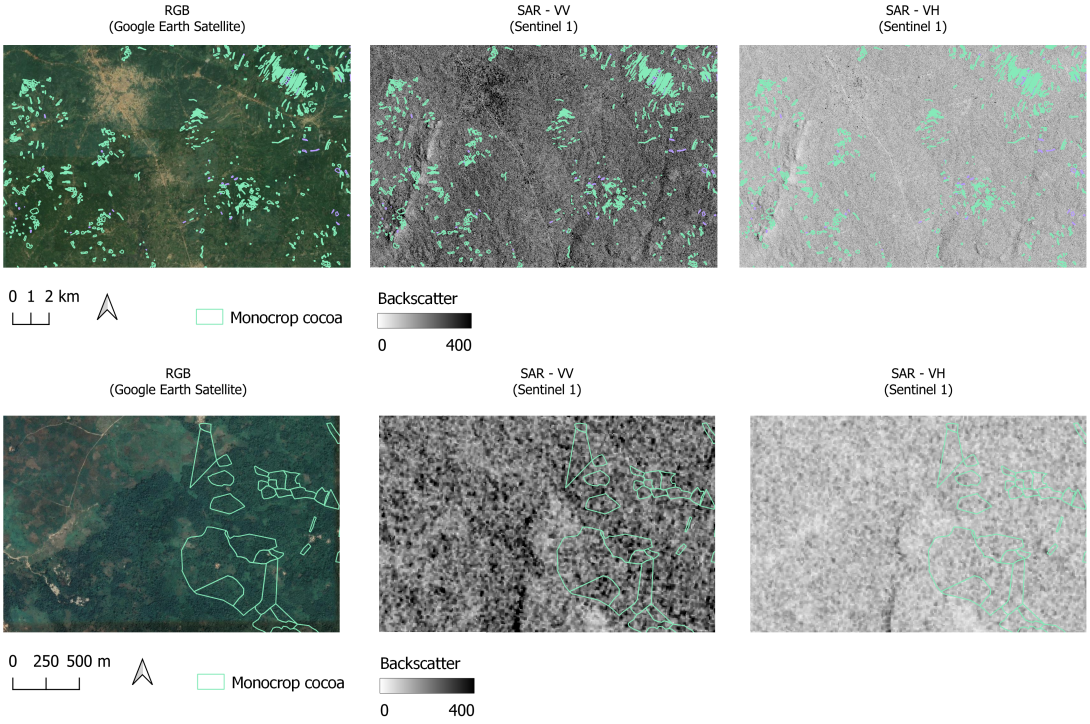


Figure 4.8.: SAR imagery in different polarizations over study area

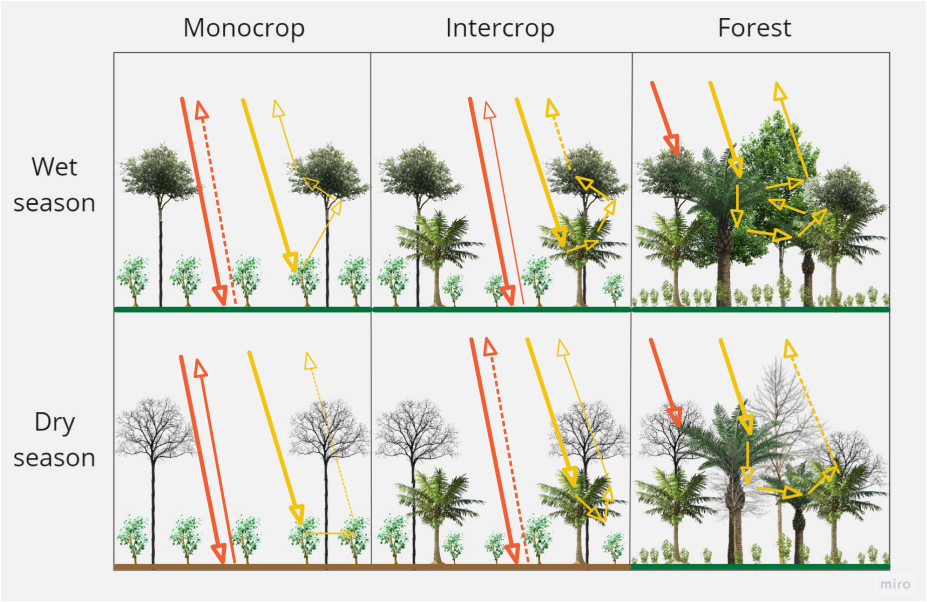


Figure 4.9.: SAR backscatter variability

Dark colour arrow = VV, Light colour arrow = VH

5. Methodology

This section provides an overview of the steps involved in the research for this thesis, as well as their theoretical and practical underpinnings. The steps are illustrated in the flowchart presented in Figure 5.1.

5.1. Satellite data preparation

After defining the study area, the first step is to retrieve satellite data for the area. The Sentinel datasets are frequently used for their open-source availability and wide coverage, both spatially, spectrally and temporally.

5.1.1. Multispectral Imagery

Considering that this research focuses on the addition of SAR to MSI datasets for semantic segmentation of cocoa parcels, the MSI dataset plays an important role as "reference" dataset. S2 captures MSI imagery in 13 bands, ranging from 442 nm (visible light) to 2202 nm (Short-wave Infrared) [60]. Considering the typical size of cocoa parcels, the lowest resolution S2 datasets (60m) are likely to generalize details along the borders of parcels and potentially decrease the effectiveness of the segmentation. Therefore, in accordance with the methodology of Kalichek et al. (2022), the S2 bands of 10m and 20m resolution MSI are used for this analysis, which include those listed in Table 5.1.1 [8]. The 20m resolution images are resampled to 10m using bilinear interpolation in order to be able to stack images using a consistent resolution [6].

Band	Resolution	Wavelength	Description
B2	10 m	490 nm	Blue
B3	10 m	560 nm	Green
B4	10 m	665 nm	Red
B5	20 m	705 nm	Visible and Near Infrared (VNIR)
B6	20 m	740 nm	Visible and Near Infrared (VNIR)
B7	20 m	783 nm	Visible and Near Infrared (VNIR)
B8	10 m	842 nm	Visible and Near Infrared (VNIR)
B8a	20 m	865 nm	Visible and Near Infrared (VNIR)
B11	20 m	1610 nm	Short Wave Infrared (SWIR)
B12	20 m	2190 nm	Short Wave Infrared (SWIR)

Table 5.1.: MSI bands used in this research

While most studies described in Chapter 3 correct or exclude clouded areas via pre-processing, this research is interested in the impact of SAR data on the presence of clouds. Therefore, MSI data is retrieved at level 1-C, which means that the imagery contains "orthorectified Top-Of-Atmosphere (TOA) reflectance, with sub-pixel multispectral registration" [60] and limiting the cloud cover threshold to 15%.

5.1.2. Synthetic Aperture Radar

SAR data is collected differently than MSI data, therefore its raw form is not compatible and stackable with MSI. This study makes use of Level 1 GRD S1 (C-band) datasets from ascending orbit that have already been processed, multi-looked and projected using the WGS84 coordinate reference system [36] with a pixel resolution of 10 m and a spatial resolution of 20 m × 22 m [61].

As seen in [47], the SAR values used in this research are gamma0 backscatter in decibels, in order to calibrate the values to the logarithmic scale which normalizes the values and increases contrast between features on the ground [1].

5.1.3. Stacking images

In order to input combined datasets into the U-NET, the rasters must be aligned and stacked into a 3-dimensional matrix in which each of the image pixels has a vector of reflectance and/or backscatter values from each of the dataset bands, as described in [8] and [6]. The bands used differ depending on the experiment; some possible combinations are illustrated in Figure 5.2.

5.1.4. Data normalization

Once all bands have been stacked, a normalization process is applied to each band independently in order to increase the speed of model convergence [62]. Typically, imagery for semantic segmentation training is normalized in order for the mean to be centered at 0 and to have a standard deviation of 1 [62]. Although this works when distribution of values is normal, this is not the case for satellite reflectance values: distributions for reflectance are long tailed and 0-bounded, and typical normalization risks being influenced by outlier reflectance values. Therefore, normalization is instead applied using the following equation for linear scaling [62].

$$val_{out} = (val_{in} - c)(\frac{b-a}{d-c}) + a \tag{5.1}$$

where val_{out} is the normalized pixel value, val_{in} the original pixel index, a and b are the lower and upper values of the desired range, and c and d are the lower and upper values of the original range. c and d can be set to the minimum and maximum values of the original range, or to the first and last quartile in order to further prevent the influence of outlier values. For this research, a and b are set to 0 and 1, and c and d are set to the minimum and maximum of each band.

5.2. Ground truth labels

The majority of the studies described in Chapter 3 use the following labels: cocoa, non-cocoa and unknown/void. Since forest and cocoa have many similarities as described in Chapter 4, forest polygons are used as "non-cocoa" regions in order to encourage their differentiation by the model.

The ground truth and satellite must be converted to a format that can be accepted by the U-NET: small image patches with a uniform size. While the U-NET can process square images of different sizes, the dimensions must be divisible by 8 as they undergo 5 convolutions (each applying a 2x2 filter and therefore dividing the original dimensions by 2^2), and they must be consistent across the entire dataset.

As the unknown class is largely overrepresented in this dataset, an undersampling method is applied by setting the minimum of labelled pixels (cocoa or forest) to 10%, as seen in [44]. In order to address the research questions, two labelled datasets are created:

1. Monocrop dataset (for training, validation and testing): at least 10% of the pixels are labelled as monocrop cocoa or forest
2. Intercrop dataset (for testing only): contains labelled intercrop cocoa and no labelled monocrop or forest pixels

5.3. Data split

In order to test the effectiveness of a classification model, a high-quality reference map must be generated against which the predictions will be compared [63]. 10% of the training dataset is set aside to be used for an independent testing of the predictions for images that the model has not previously seen.

In order to get a more robust estimate of the model's performance and determine whether it is consistent across different splits of the data, a 10-fold cross-validation approach is used. This method divides the dataset into ten folds, and trains the network ten times, each time using a different fold as the validation dataset.

5.4. U-NET Segmentation

The U-NET employed in this research is adapted from the code described in [64] based on the original U-NET architecture. It is composed of five encoder blocks and five decoder blocks using ReLU as its activation function and "same" padding. The random weights are initialized using the He normal distribution, the Adam optimizer and the Sparse Categorical Cross Entropy loss function are used for training, and a 0.3 dropout probability is applied to prevent overfitting. In addition, following the work of Filella (2018), the mini-batch size is set to 32, the learning rate to 10^{-5} and an L2 regularization rate of 10^{-2} [44]. For each experiment, the model with the best validation loss value is saved using checkpoints and used for testing. In order to prevent overfitting, early stopping is used to automatically stop the training when there has not been validation loss improvement for 10 epochs.

To generate reference results, the U-NET is trained with single day and multi-day stacks of MSI data. It is also trained separately with single- and multi-day SAR datasets. Next, experiments are conducted which involve stacking SAR data (single- and multi-day rasters) onto MSI datasets in order to observe the impacts on evaluation metrics and visual output. A summary of all experimental datasets is depicted in Table 6.1 of Chapter 6.

5.4.1. Weighted loss function

In order to avoid relying on undersampling forest labels to balance the ground truth classes, the class imbalance is rectified via a weighted loss function. The weights are determined based on the "inverse proportion of class frequencies" [65]: the proportion of labels from each class is calculated in order to increase the influence of the minority class (i.e. cocoa) on the loss computation, as depicted in the following equation:

$$w_c = \frac{1}{\frac{n_c}{N}} \tag{5.2}$$

where c is a given class, n is the number of labels in this class and N is the total number of labels in the dataset. Any pixel without a ground truth label is assigned a weight of zero, therefore not having any influence on the training. These weights are integrated into the categorical cross entropy function as shown below.

$$L_{cross-entropy}(\hat{y}, y) = -\frac{1}{N}\sum_{j}^{N}\sum_{c}^{M} w_c y_{c,j} \ln(\hat{y}_{c,j}) \tag{5.3}$$

where N is the number of labeled samples, and M is the number of classes, $y_{c,j}$ is a binary indicator (0 or 1) of a pixel belonging to a certain class c and $\hat{y}_{c,j})$ is the predicted probability of pixel belonging to a certain class c.

5.4.2. Probability and prediction

The output of the U-NET is a n-dimensional matrix in which n is the number of classes and each channel contains the probability that a pixel belongs to the given class. The Softmax activation function is applied in order to squeeze these vectors to have values between 0 and 1 [44] [24].. This output enables a visualization of the per-pixel cocoa probability map. This continuous raster enables a choice in visualization as the user can establish a probability threshold that is most suiTable to their needs, such as visualizing as "cocoa" areas that are predicted with a probability above 65% as shown in [8]. The U-NET also outputs a prediction mask which is generated by the Argmax function, which identifies for each pixel the class with the highest probability. This raster contains discrete labels, which enable the computation of evaluation metrics described in the following section.

5.5. Evaluation

In order to evaluate the segmentation outputs for each of the experiments, satellite images from the test datasets are provided to the U-NET and the predicted masks are compared to the ground truth labels, excluding the "unknown" regions. The following metrics are observed to evaluate the effectiveness of the model:

1. Loss: measures the difference between the ground truth of each pixel and the probabilities predicted for each class. The loss values are averaged for all labelled pixels in all test images.
2. IoU: measures the degree of similarity between the predicted mask and ground truth, and is computed only for the Cocoa class [66, p. 8].

$$IOU = \frac{Areaof Overlap}{Areaof Union} = \frac{TP}{TP + FP + FN} \tag{5.4}$$

3. Recall: calculates the proportion of correctly predicted ground truth cocoa, also referred to as Producer's Accuracy [66, p. 10]

$$PA = \frac{TP}{TP + FN} \tag{5.5}$$

4. Precision: calculates the proportion of true cocoa predictions, also referred to as User's accuracy [66, p. 10]

$$UA = \frac{TP}{TP + FP} \tag{5.6}$$

5. F1: the harmonic mean of Precision and Recall [66, p. 10]

$$F1 = \frac{2 * UA * PA}{UA + PA} \tag{5.7}$$

The metric means across all experiment folds are plotted in order to observe the differences between dataset trainings and identify any significant trends. In the case that the means appear very similar, or that a particular comparison is deemed crucial for addressing the research questions, additional box plots are created in order to observe the spread of the data and any statistically significant differences between datasets.

Two challenges arise from the exclusion of large unknown areas. One challenge is that a model which overpredicts cocoa will have very good metrics, as the ground truth polygons will contain high probability cocoa predictions, but so will other regions in the image. Another challenge is that there is missing information from the majority of most images where the prediction accuracy is not quantified or estimated. These two challenges can be addressed via visual observation of the predicted masks and probabilities. By viewing the predicted maps, it can be identified whether cocoa is predicted across most of the image. Furthermore, by comparing the predictions to satellite imagery, it is possible to estimate qualitatively the likelihood that certain predictions are correct (e.g. by visually identifying crops, urban areas, water and other features that are visible to the human eye).

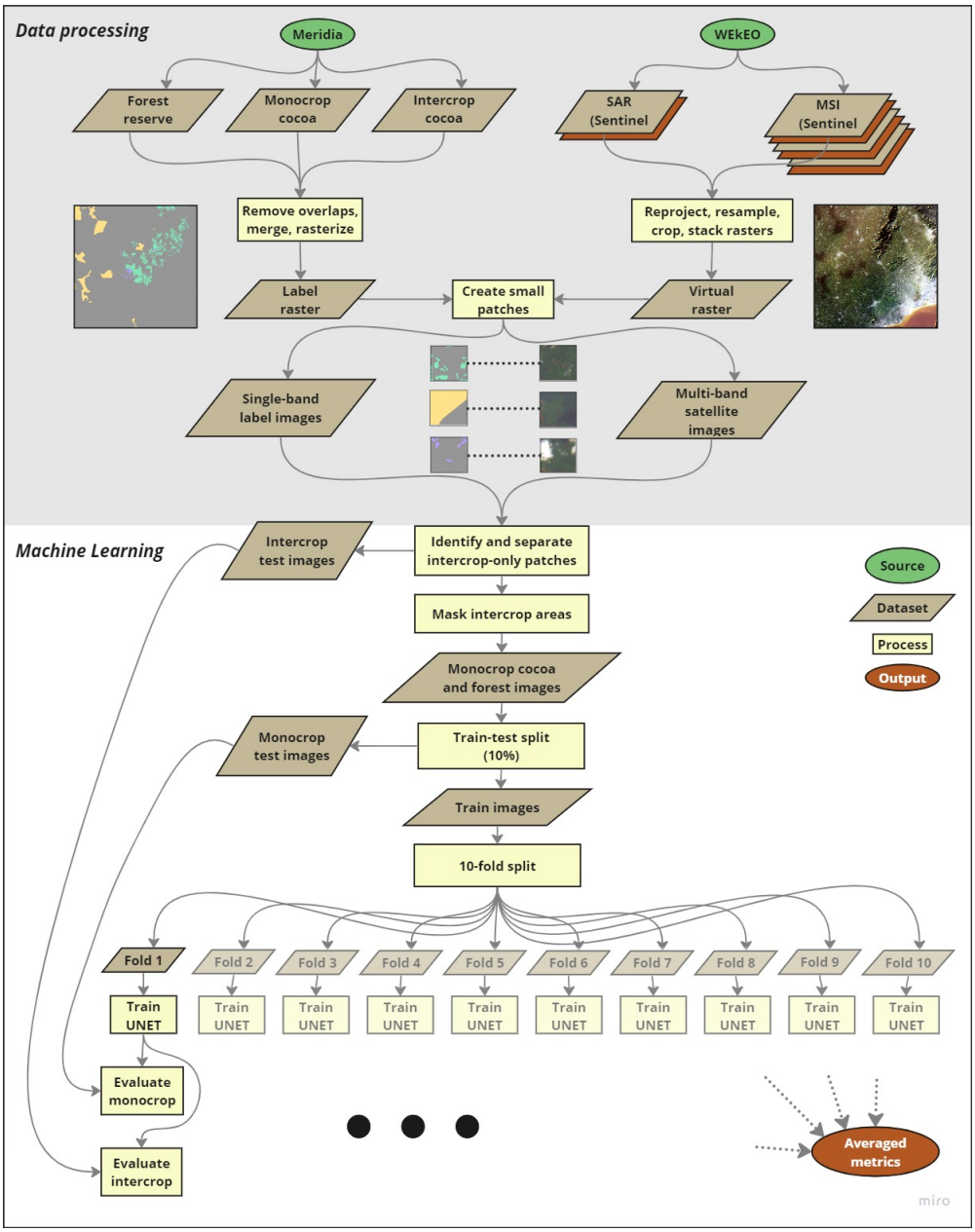


Figure 5.1.: Methodology flowchart

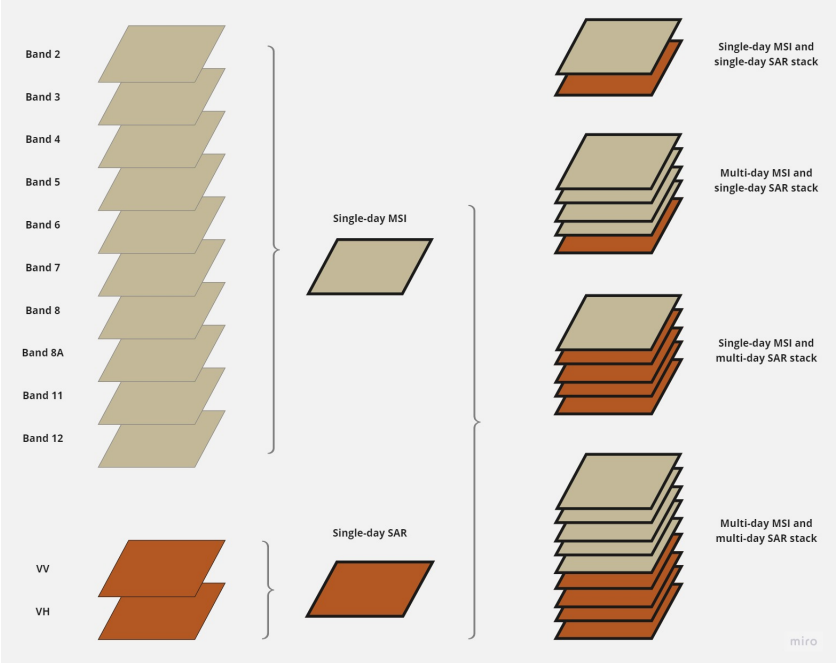


Figure 5.2.: Possible stacking combinations of MSI and SAR datasets.

6. Implementation

6.1. Data preparation

6.1.1. Satellite rasters

S1 and S2 datasets are downloaded via the WEkEO JupyterHub (Earth Observation Server) using the WEkEO Harmonized Data Access API [67]. The MSI datasets (Level 2A) are filtered to contain less than 15% cloud cover and efforts are made to select imagery distributed across the wet and dry seasons (see figure 6.1). The year 2020-2021 is selected because it contains datasets that match as closely as possible to an even distribution across the seasons, and because this timeline overlaps with the ground truth dataset collection dates. The SAR datasets (Level 1C) are filtered to IW mode and GRD products only. For the dry season stack, the images are selected as close as possible to the dates of MSI data.

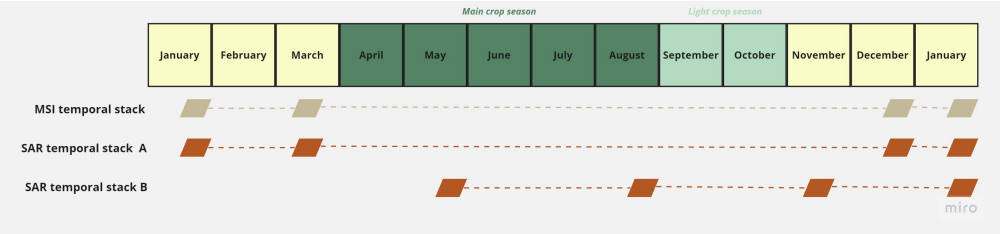
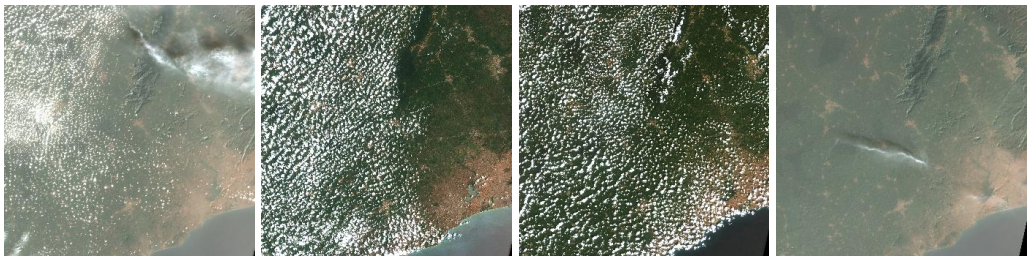


Figure 6.1.: Seasons of southern Ghana and distribution of datasets over one year.

After downloading the datasets, the following steps are implemented using QGIS:

1. Project MSI data to UTM zone 30N and re-sample all 20 m resolution bands to 10 m resolution using bi-linear interpolation



(a) January 2020 (b) March 2020 (c) December 2020 (d) January 2021

Figure 6.2.: Imagery included in the MSI temporal stack

2. Project SAR to UTM zone 30N, re-sample to 10 m resolution and clip to study area
3. Stack MSI and SAR in various virtual rasters with each file saved as a separate band.

6.1.2. Ground truth polygons

After collecting the ground truth polygons from Meridia, the following steps are conducted using QGIS:

1. Quality-control ground truth polygons: conduct visual checks on cocoa ground truth to check for possible misclassification, such as overlaps between cocoa and non-cocoa ground truth
2. Cocoa polygons: remove any polygons classified as ”sparse,” ”unknown” or ”both” to keep only ”intercrop” and ”monocrop” cultivation types. Remove any intercrop polygons intersecting with monocrop polygons and vice-versa.
3. Forest polygons: remove any polygons that intersect with cocoa polygons and apply a 500m internal buffer to exclude the forest reserve edges.
4. Merge cocoa and forest polygons into one layer, then rasterize with label values (0 = unknown, 1 = monocrop cocoa, 2 = forest, 3 = intercrop cocoa)

Note: the labelling process described above is repeated to create a separate dataset composed of a more varied ”Not cocoa” class which also contains non-cocoa crop polygons and is used in some of the experiments described in chapter 7.

6.2. Creating patches

Using the label and satellite rasters described in the previous section, the following steps are implemented using a custom Python script with the help of the following libraries: rasterio, geopandas, shapely, pandas, numpy, and fiona.

1. Initialize moving window at top left of study area (i.e. 128 x 128 pixels)
2. Move window across entire tile with a stride that is equal to the window size, in order to ensure no duplication of data between train and test datasets. For each location, check if labels satisfy the monocrop or intercrop dataset requirements.
3. If so, crop and save a copy of the label raster, save a corresponding cropped portion of the masked virtual raster.

When following the above steps for the cocoa and forest label dataset, the study area generates 56 images with intercrop-only labels and 496 images containing a minimum of 10% monocrop cocoa and/or forest labels. The monocrop dataset labelled pixels consist of 13.38% cocoa and 86.62% forest. When removing clouded areas from this dataset, the study area generates 56 images with intercrop-only labels and 537 images containing a minimum of 10% monocrop cocoa, forest and/or non-cocoa crop. The monocrop dataset labelled pixels consist of 20.30% cocoa and 79.70% forest.

When following the above steps for the cocoa and non-cocoa (forest+other crop) label dataset, the study area generates 57 images with intercrop-only labels and 556 images containing a

minimum of 10% monocrop cocoa, forest and/or non-cocoa crop. The monocrop dataset labelled pixels consist of 18.44% cocoa and 82.46% non-cocoa.

6.3. Data split

The patches are uploaded to a AWS Studio Lab workspace to be further processed via a custom Studio notebook that is scripted in Python [68]. The datasets are loaded and the 10% test split is applied using the sci-kit learn library (train test split function). Considering that the cocoa class is the primary class of interest, and that it is underrepresented compared to the forest class, the aim is to increase the proportion of cocoa in the test dataset. Therefore, 10 different randomizations are applied and the one with the greatest proportion of cocoa (21.58%) is kept as the test dataset. The 10-fold validation and training data split is implemented using the sci-kit learn k-fold function.

6.4. UNET

The U-NET architecture is adapted from the cocoa segmentation work of [44] and the Jupyter notebook U-NET implementation by [64] with the following parameters:

- Input: 128x128x[number of bands]
- Number of filters: 32
- Number of classes: 3
- Encoder: 5 blocks of two Conv Layers (3x3 filters, 'same' padding) with relu activation and HeNormal initialization, max pooling
- Decoder: 4 blocks of transpose convolution, concatenate with skip connection from encoder, two Conv layers (3x3 filters, 'same' padding)
- Model output: one Conv layer (3x3 filters, 'same' padding) followed by one 1x1 convolution layer to get image to same size as input.

The following table summarizes the experiments carried out, and the dataset combinations used for each:

Set	#	MSI Data	SAR Data	Season
1	1	December	N/A	Dry
	2	March	N/A	Dry
	3	Jan, March, December, Jan	N/A	Dry
	4	N/A	May - VV	Wet
	5	N/A	May - VH	Wet
	6	N/A	May - VV VH	Wet
	7	N/A	Jan - VV	Dry
	8	N/A	Jan - VH	Dry
	9	N/A	Jan - VV VH	Dry
	10	N/A	Jan, March, December, Jan - VV	Dry
	11	N/A	Jan, March, December, Jan - VH	Dry
	12	N/A	Jan, March, December, Jan - VV VH	Dry
	13	N/A	May, August, November, Jan - VV	Multi
	14	N/A	May, August, November, Jan - VH	Multi
	15	N/A	May, August, November, Jan - VV VH	Multi
	16	December	Jan - VV	Dry
	17	December	Jan - VH	Dry
	18	December	Jan - VV VH	Dry
	19	December	May - VV	Multi
	20	December	May - VH	Multi
	21	December	May - VV VH	Multi
	22	December	May, August, November, Jan - VV	Multi
	23	December	May, August, November, Jan - VH	Multi
	24	December	May, August, November, Jan - VV VH	Multi
	25	Jan, March, December, Jan	May - VV	Multi
	26	Jan, March, December, Jan	May - VH	Multi
	27	Jan, March, December, Jan	May - VV VH	Multi
	28	Jan, March, December, Jan	May, August, November, Jan - VV	Multi
2	29	December (label change)	N/A	Dry
	30	December	May	Multi
	31	December	May, August, November, Jan	Multi
3	32	December (cloudless)	N/A	Dry
	33	December	May	Multi
	34	December	May, August, November, Jan	Multi

Table 6.1.: List of experiments indicating dataset stacks and seasonality

7. Results and analysis

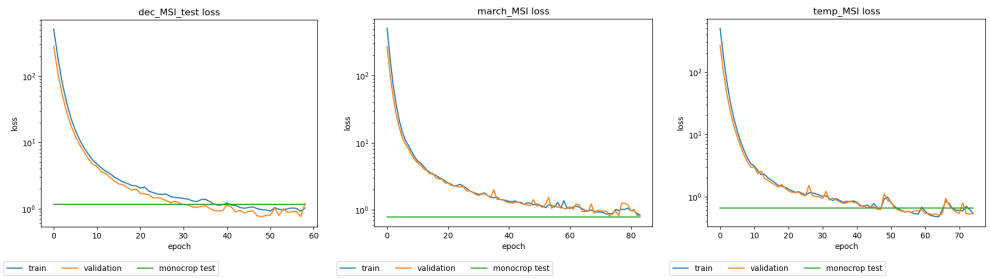
The following section first describes metrics from reference experiments that involve the use of only one kind of data (i.e. SAR or MSI), followed by experiments stacking different combinations of datasets. These initial experiments are referred to as "Set 1."

Several experiments are repeated with an extended "Not cocoa" class ("Set 2"), and with the removal of clouded areas ("Set 3"). Finally, observations on the detection of intercrop cocoa are explained. All metrics are recorded in Appendix A. **Note:** a low Loss value indicates a more effective prediction model.

7.1. Reference training with MSI

When comparing the results of training with single-day MSI from March (end of the Dry season) and December (beginning of the dry season), March imagery leads to better results, with a higher IoU and lower loss value, as shown in Figure 7.2. This may be due to the fact that over the course of the Dry season, vegetation cover undergoes more significant changes in agricultural areas compared to natural forest, leading to more distinguishable spectral differences by the end of the Dry season. For the single-day MSI experiments described later in this chapter, the December imagery is used as the reference in order to evaluate whether the addition of SAR may improve its results.

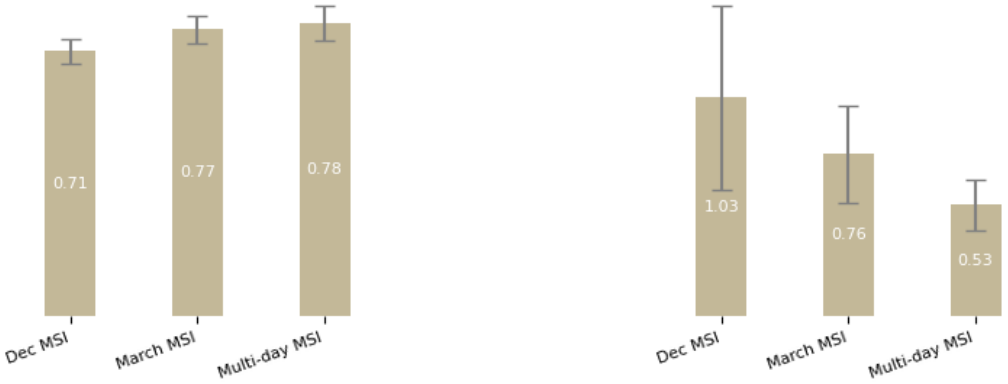
Training the U-NET with a temporal stack of MSI imagery from four days across the Dry season leads to improved results compared to both of the single-day datasets. The use of temporal images enables the changes in vegetation to inform the training of the model. Furthermore, visual comparisons show that the December- and March-trained models are predicting clouded areas as forest. On the other hand, training with the temporal imagery



(a) December MSI (b) March MSI (c) Multi-season MSI

Figure 7.1.: Loss curves for MSI-only reference experiments

increases the amount of False Positives, which could suggest that the model is instead associating clouded areas with cocoa fields. This is possibly due to the fact that clouds cover different fields on different days in the temporal stack (see Figure 7.3).



(a) IoU

(b) Loss

Figure 7.2.: MSI-only metrics

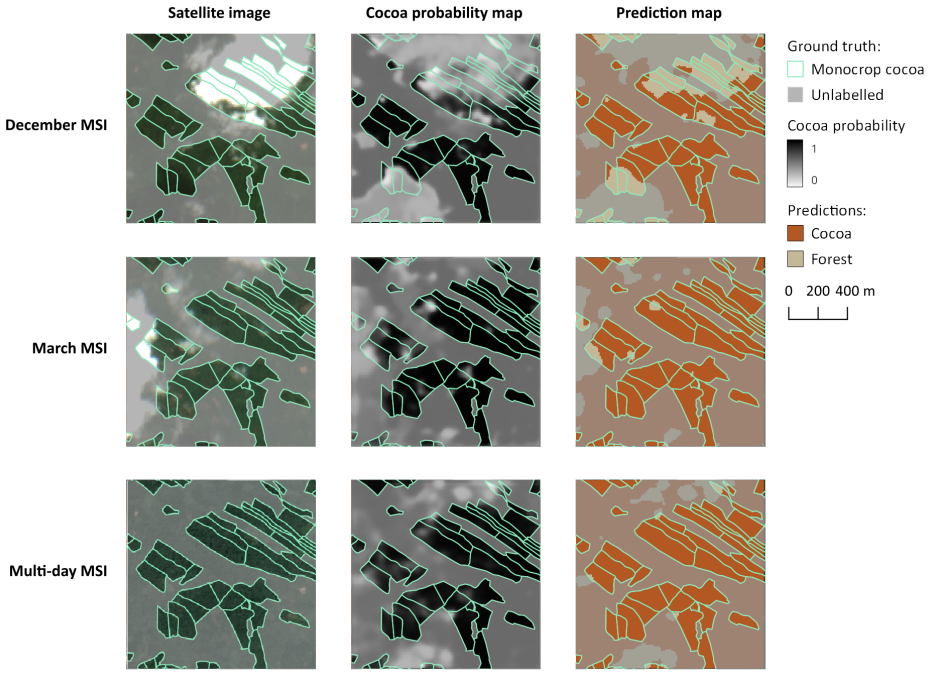
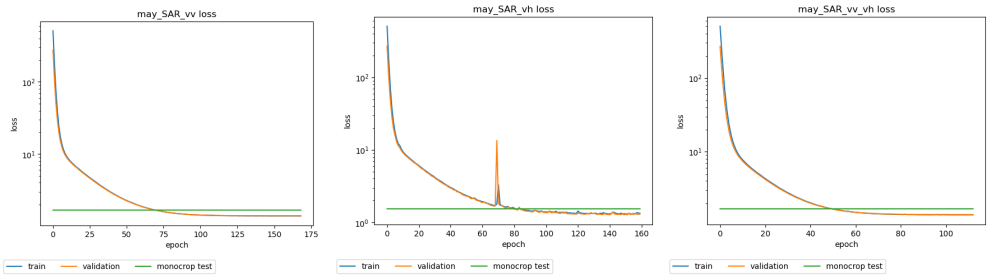


Figure 7.3.: MSI-only maps

7.2. Reference training with SAR

7.2.1. Single day SAR

Overall, the results from the January dataset (Dry season) are better than those trained with May data (Wet season) (see Figure 7.6). This may be attributed to the fact that during the Dry season, differences in the canopy structure and moisture are likely more significant between the classes compared to the Wet season.



(a) May VV (b) May VH (c) May VV VH

Figure 7.4.: Loss curves for single-day SAR-only reference experiments (May)

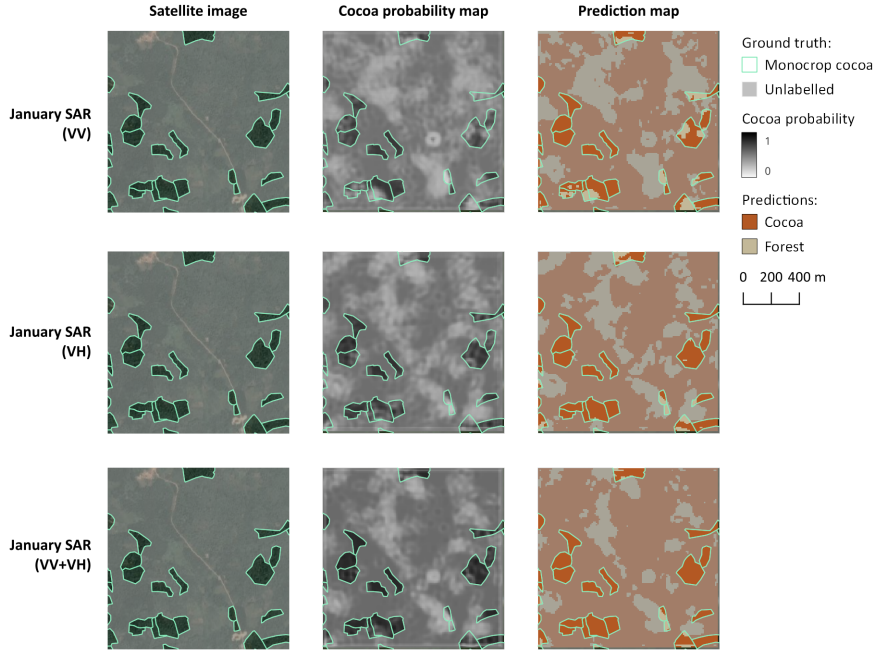
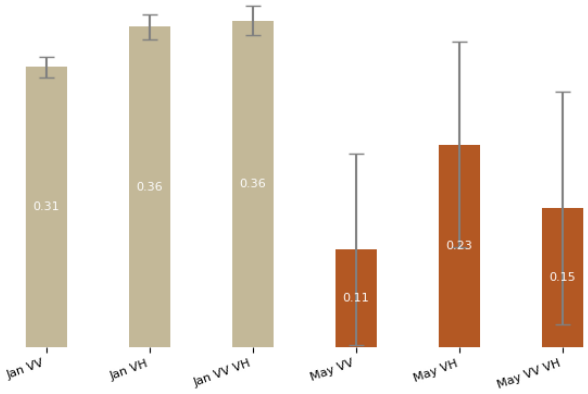


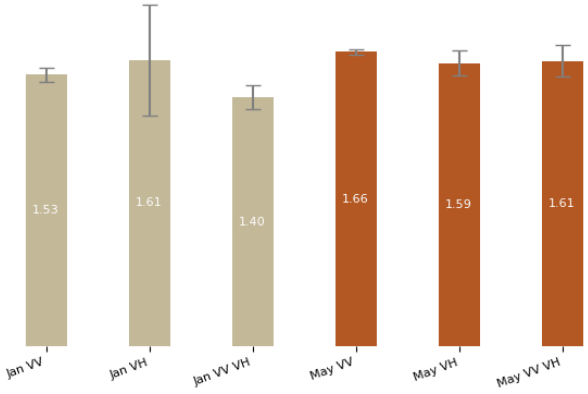
Figure 7.5.: Single-day SAR maps

When comparing the results from different polarizations of data captured on the same day, VH offers the best results, followed by combined VV and VH, with VV leading to the poorest metrics. When comparing the prediction maps for different polarizations in January, VV is

more likely to predict forest in cocoa polygons (see Figure 7.5). This may be influenced by the density of vegetation in some areas of cocoa fields which prevent co-polarized ground backscatter.



(a) IoU



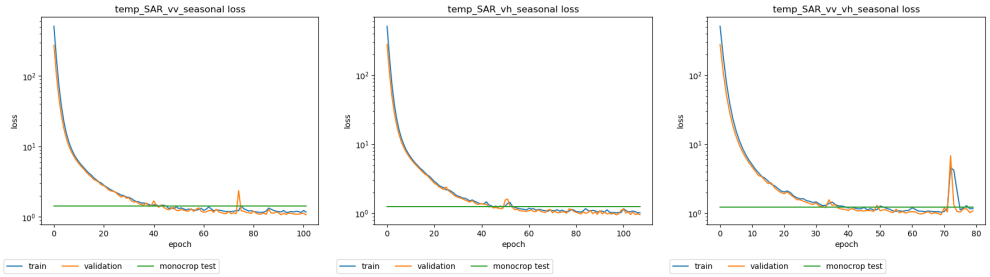
(b) Loss

Figure 7.6.: Single-day SAR metrics

7.2.2. Multi-day SAR

The reference multi-day SAR experiments compare the training of the UNET using a stack of SAR datasets from across the Dry season, and another that is composed of four images distributed over the Wet and Dry seasons (referred to as "multi-season").

Overall, the training results for the multi-season stacked are better than Dry season only (see Figure 7.9). This is likely due to the additional insight into seasonal changes that is provided by the Wet season imagery, such as the loss of foliage of cocoa shade trees compared to the more seasonally stable forest canopies.



(a) Multi-season VV (b) Multi-season VH (c) Multi-season VV VH

Figure 7.7.: Loss curves for multi-day SAR-only reference experiments (multi-season)

When comparing polarization metrics, the use of VV leads to significantly worse results in both temporal stacks compared to VH or VV+ VH which have similar results. It may be due to the fact that seasonal changes in ground backscatter alone is not sufficient to distinguish between different vegetation types. A comparison of prediction maps of single day SAR stacks for January and May to multi-season SAR in the same polarization (VH), January and multi-season have visually similar results, predicting more cocoa than the May trained model, which predicts forest in most of the cocoa polygons (see Figure 7.8).

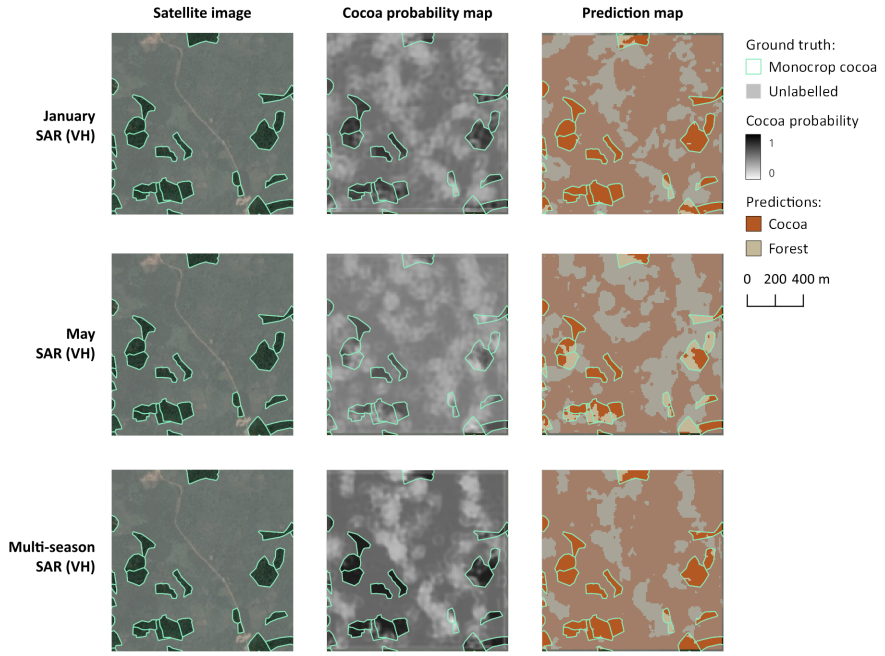
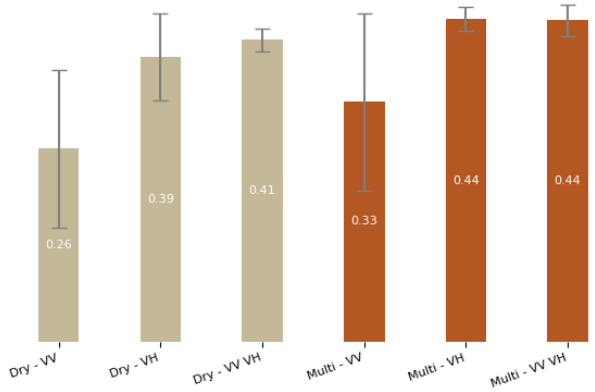
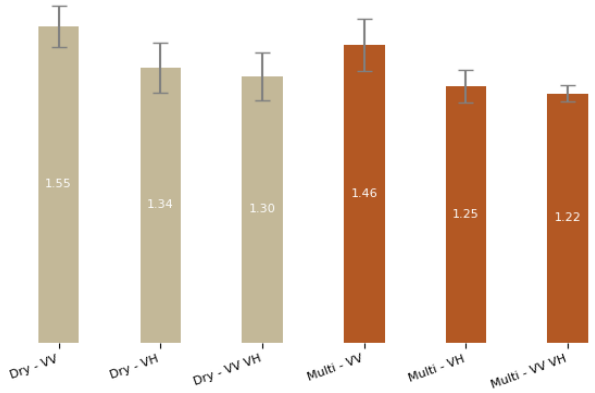


Figure 7.8.: Multi-day SAR maps



(a) IoU



(b) Loss

Figure 7.9.: Multi-day SAR metrics

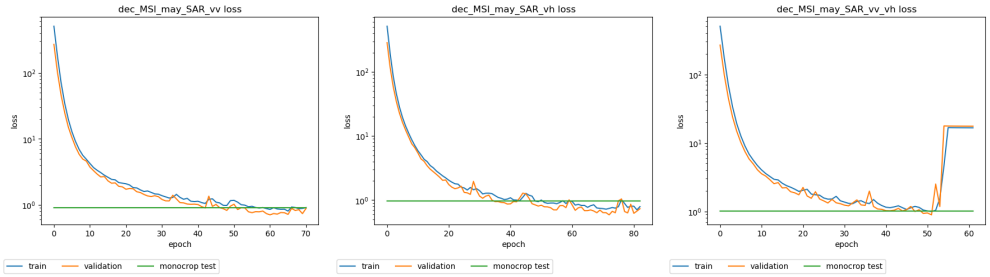
7.3. Experiment results

7.3.1. Single day MSI and SAR

This subsection investigates whether the stacking of SAR on the single-day December MSI can improve training metrics, analyzing the influence of season(s) and polarization.

Single day MSI and single-day SAR

The seasonality of datasets affects the effectiveness of different SAR polarizations. When stacking VV data with MSI, imagery from the Dry season (the same season as the MSI imagery) leads to better metrics (see Figure 7.12). This may be due to the fact that during the Dry season, cocoa is likely to have more canopy gaps and therefore a higher contrast in ground backscatter compared to forest.



(a) VV (b) VH (c) VV VH

Figure 7.10.: Loss curves for December MSI + May SAR

In contrast, stacking VH data leads to better results when working with Wet season SAR data. A comparison of the loss box plots (Figure 7.13) shows that the difference is statistically significant. This may be attributable to the fact that combining a Dry season MSI image with a Wet season SAR provides temporal information about the canopy that is not available when both datasets originate from the same season. By comparing prediction maps in Figure 7.11, it can be observed that the VH May dataset predicts the least amount of cocoa in the forest compared to the two other polarization stacks. When combining VV and VH, the result is similar across seasons. Overall, stacking December MSI with May VH SAR leads to the lowest loss value.

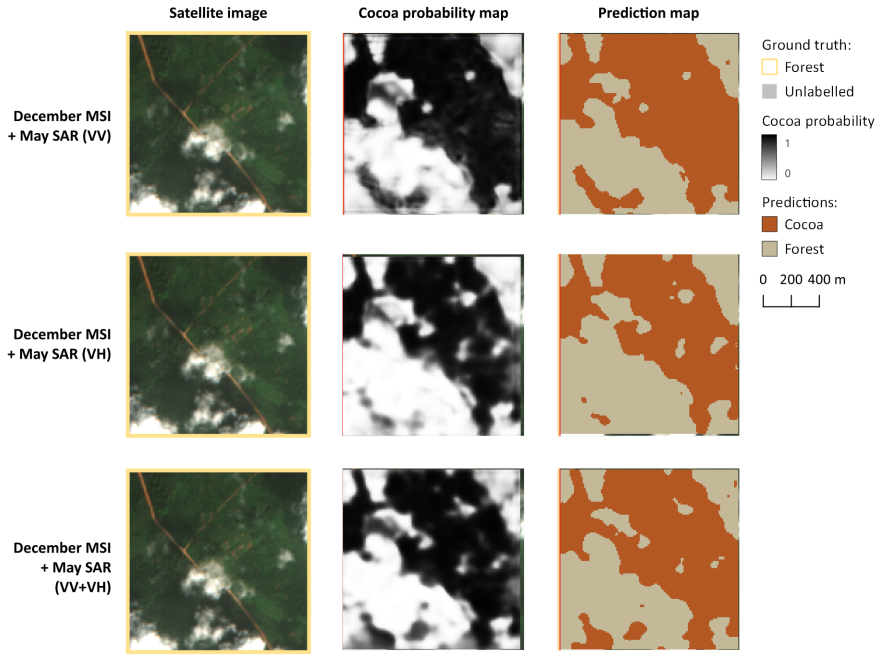
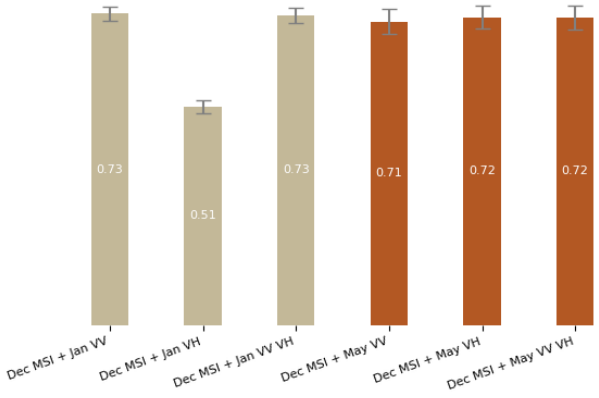
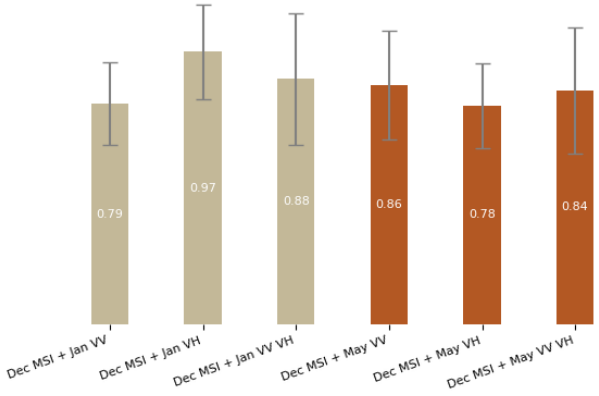


Figure 7.11.: Single-day MSI + single-day SAR maps



(a) IoU



(b) Loss

Figure 7.12.: Single-day MSI + single-day SAR metrics

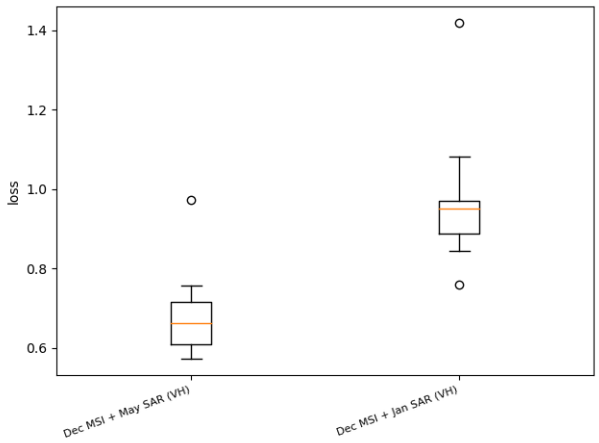
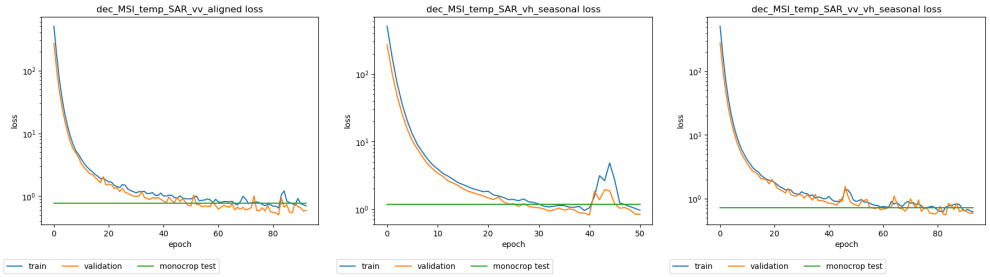


Figure 7.13.: Single-day MSI + single-day SAR loss box plot

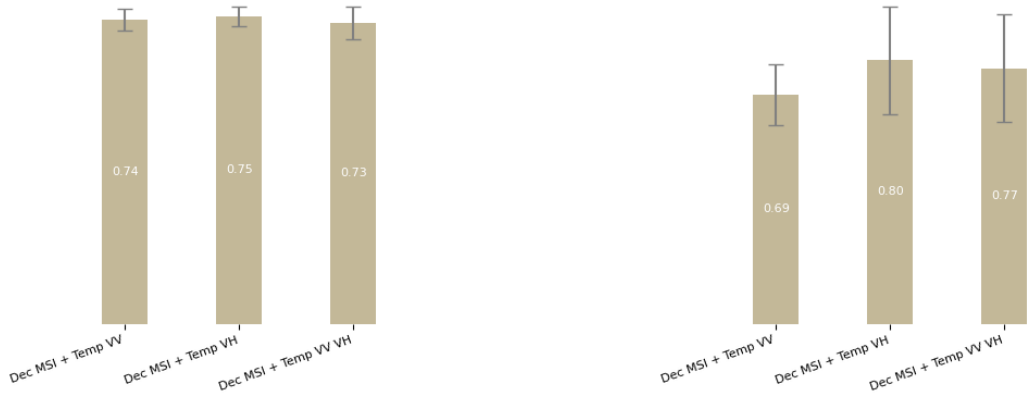
Single day MSI and multi-day SAR



(a) VV (b) VH (c) VV VH

Figure 7.14.: Loss curves for December MSI + multi-season SAR

In these experiments, the December MSI imagery is stacked with the multi-seasonal SAR data in order to compare the use of different polarizations. As shown in Figure 7.15, the IoUs for all three experiments are nearly identical, and visual observation of the prediction maps indicate that all three have a large amounts of False Positives. However, when comparing the Loss values, VV has the best result, followed by VV+VH, then VH. In Figure 7.16, it can be observed that there is patch of possibly clear-cut or agricultural land; this area is most effectively differentiated from the forest by the VV stack. While cocoa is not the correct class for this patch, it does provide insight into the potential for temporal VV combined with single-day MSI to differentiate between vegetated areas with differing amounts of canopy cover.



(a) IoU (b) Loss

Figure 7.15.: Single-day MSI + Multi-day SAR metrics

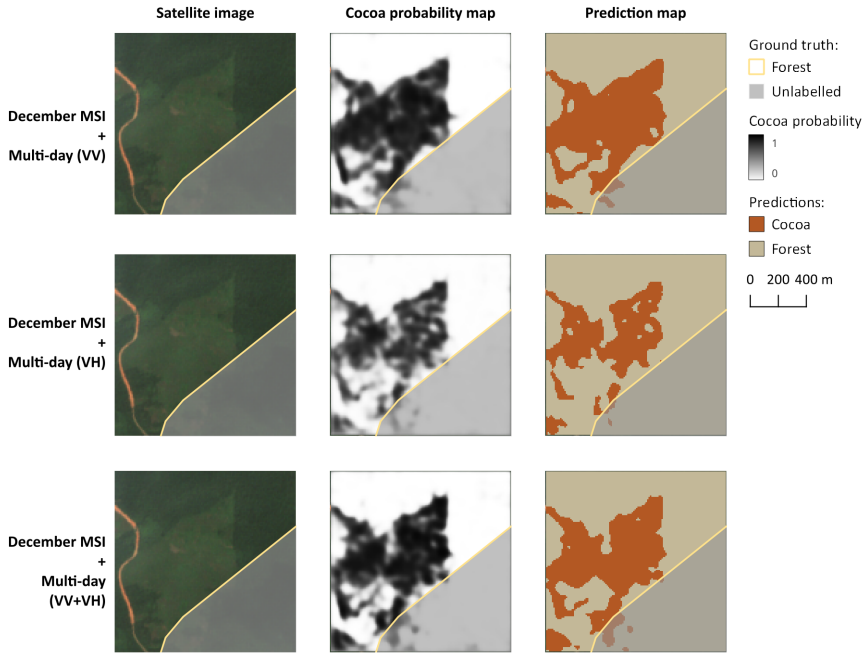


Figure 7.16.: Single-day MSI + Multi-day SAR maps

Figure 7.17 compares the predictions for the best results from single-day MSI, single-day MSI + single-day SAR and single-day MSI + multi-day SAR by focusing on an easily-identifiable crop: oil palm. Looking outside of the labelled cocoa polygons, it is visually evident where palm oil crops are located because they have a distinct texture. When comparing the probability maps, the single-day MSI + single-day SAR experiment has the best results as cocoa has the lowest probability in the palm field. One possible reason is that palm and cocoa may behave more similarly across the seasons in contrast with forest (e.g. less dense canopy, more exposed ground) which makes the co-polarized backscatter less effective at differentiating them, but palm and forest have similarly dense canopies in the wet season, and therefore similar levels of cross-polarized backscatter.

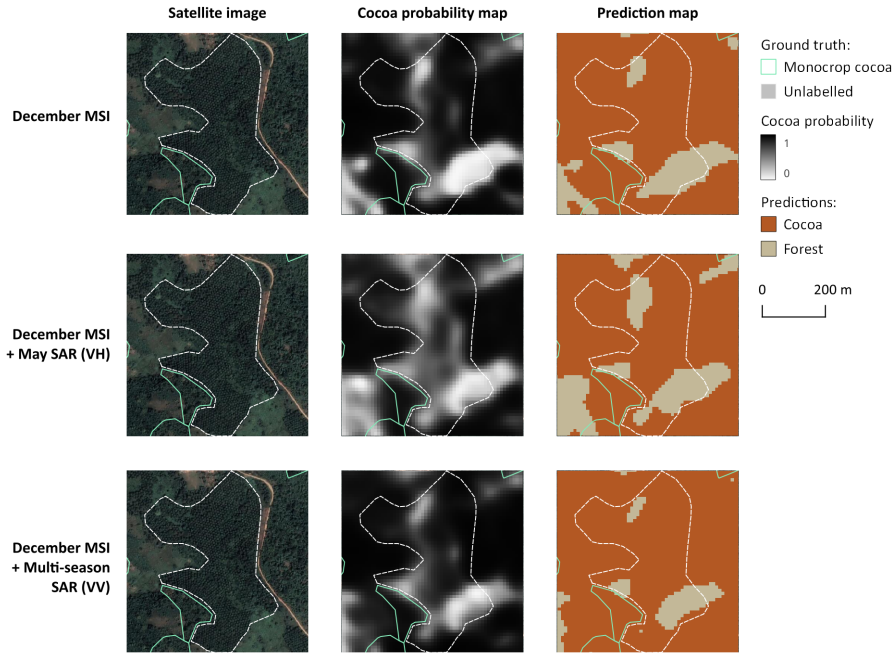
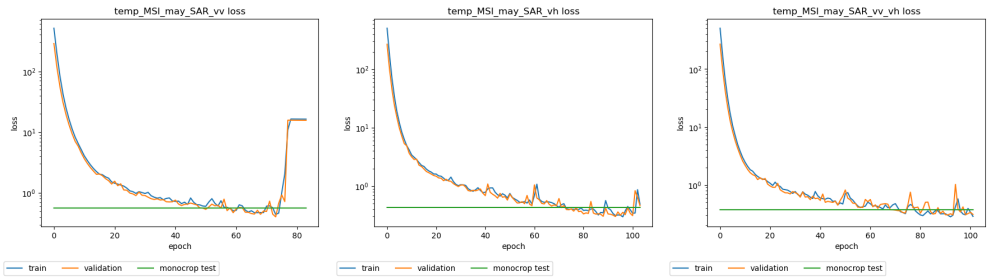


Figure 7.17.: Single-day MSI + SAR maps

7.3.2. Multi-day MSI and SAR

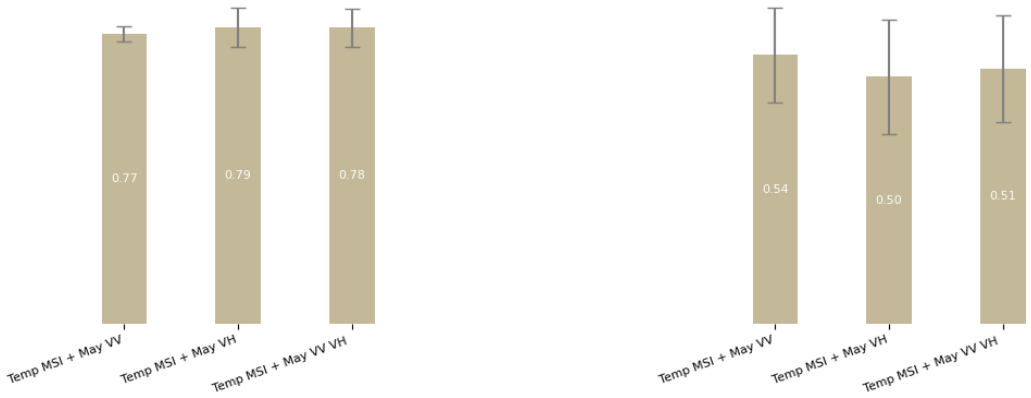
Multi-day MSI and single day SAR

When stacking May SAR with the temporal MSI stack, the resulting IoUs are plotted in Figure 7.19 showing very slight differences. In both Loss and IoU metrics, VH has a slightly better result, followed closely by VV+VH, then VV. The prediction maps in Figure 7.20 indicate cocoa in many parts of the images, including a high number of False Positives, but there some alignments between boundaries of the ground truth polygons and surrounding forest predictions in the probability and prediction maps.



(a) VV (b) VH (c) VV VH

Figure 7.18.: Loss curves for multi-day MSI + May SAR



(a) IoU (b) Loss

Figure 7.19.: Multi-day MSI + Single-day SAR metrics

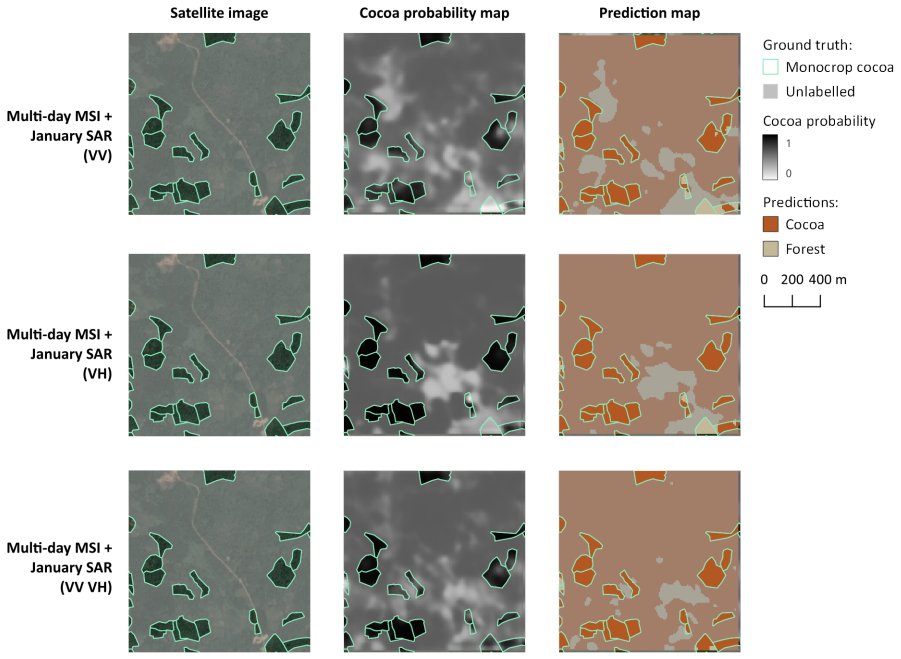


Figure 7.20.: Multi-day MSI + Single-day SAR maps

Multi-day MSI and multi-day SAR

The U-NET was trained with one dataset of multi-day MSI stacked with a multi-day SAR dataset in VV polarization. Co-polarization is selected for this experiment considering that the single-day MSI + multi-day SAR yields the best results in VV.

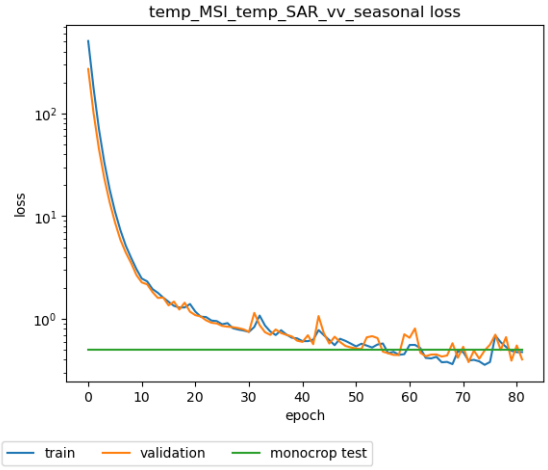
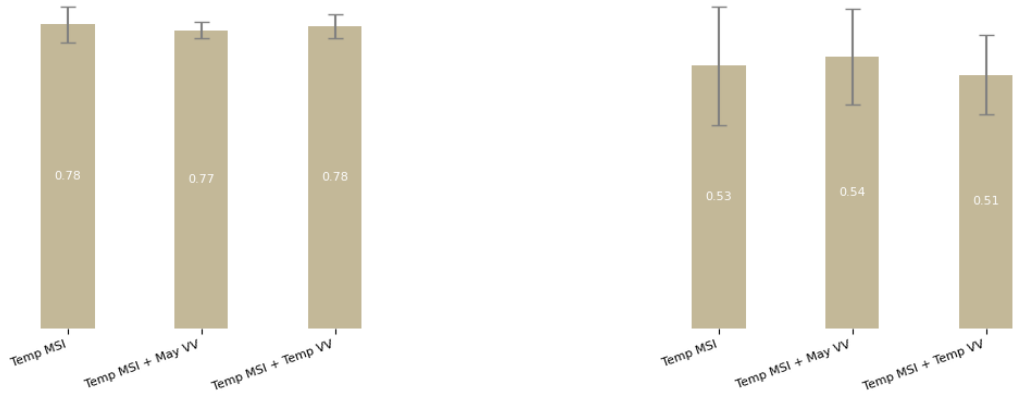


Figure 7.21.: Loss curve for multi-day MSI + multi-season SAR VV

When comparing these results with multi-day MSI alone and multi-day MSI stacked with single-day SAR, the IoUs are nearly identical (see Figure 7.22). However the loss score shows a slightly worse result when stacking a single day of SAR, and a slightly improved result when stacking a multi-day SAR dataset. By comparing the predictions (Figure 7.23), it can be observed that the use of temporal SAR excludes the urban area from the cocoa class, whereas single day SAR and MSI label this area as cocoa.



(a) IoU

(b) Loss

Figure 7.22.: Multi-day MSI + SAR metrics

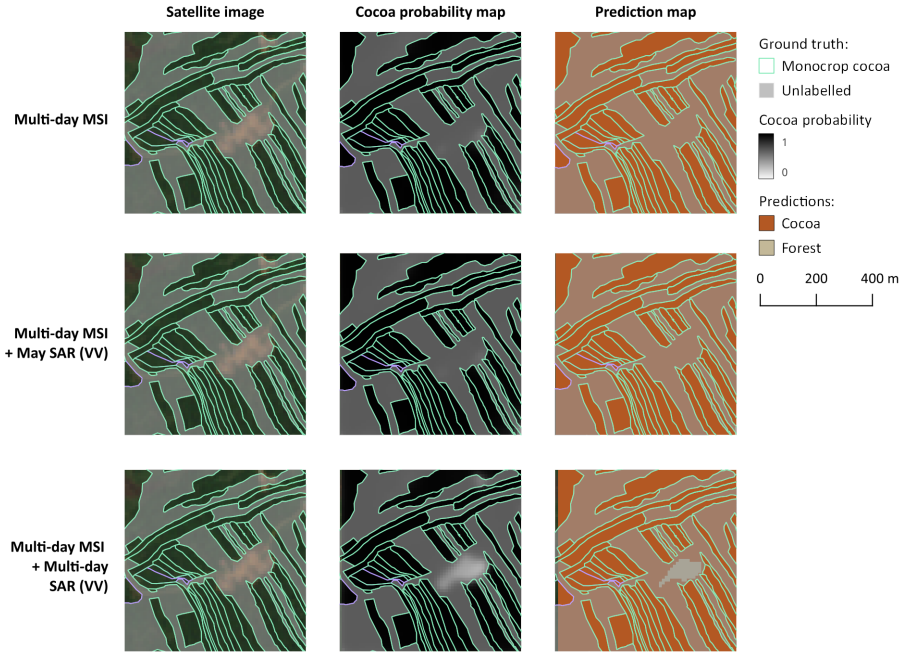
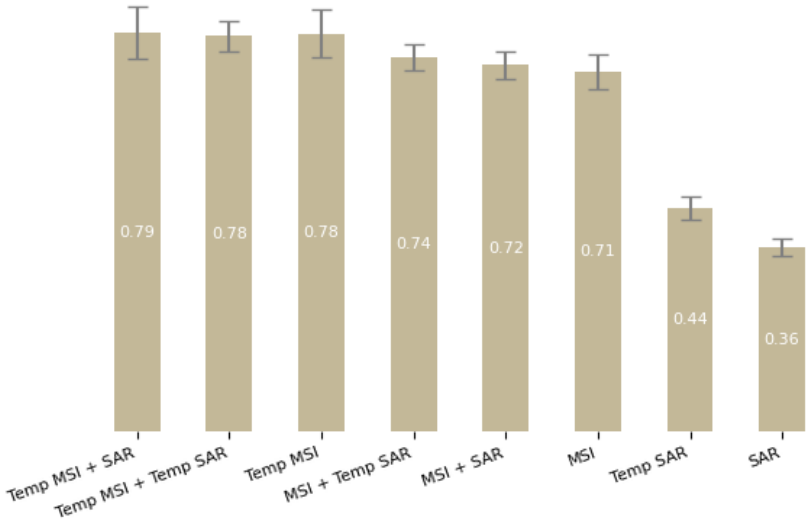


Figure 7.23.: Multi-day MSI + SAR maps

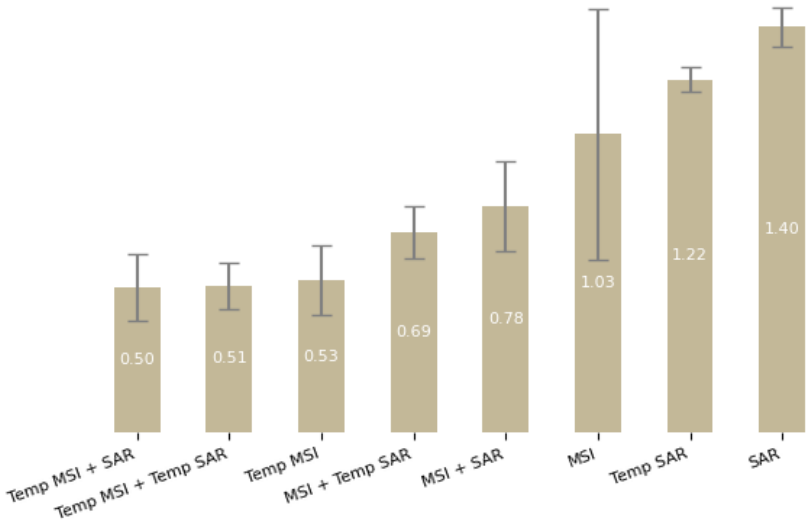
Comparison of predictions from multi-day MSI stacked with single-day and multi-day SAR, showing the detection of an urban area.

7.3.3. Overview of Set 1 experiments

The metrics of the best results from each experiment in Set 1 are plotted in Figure 7.24. The multi-day MSI datasets have the best results, and the SAR-only datasets have the worst results. This is not surprising, as the combination of temporal and multi-band data can provide significantly more insight into the differences between land cover type compared to backscatter only.



(a) IoU



(b) Loss

Figure 7.24.: Comparison of IoU and Loss for best results from each experiment.

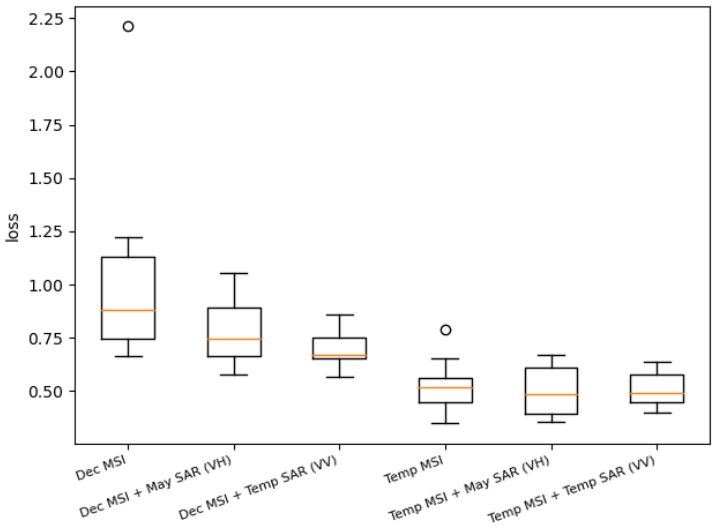


Figure 7.25.: Box plot comparison of loss metric values from original 10-fold experiments.

Based on Figure 7.24, it appears that the addition of single-day and multi-season SAR both improve the results for multi-day MSI data. This is also the case when stacking single-day MSI with temporal SAR. However, adding single-day SAR to single-day MSI appears to lead to slightly worse results. By looking closer at the distribution of the loss values across the 10-fold experiments, it becomes clear that an outlier is affecting the loss average for the December MSI and Multi-day MSI experiments.

Therefore, comparing the box plots provides a more accurate assessment of the experiment results as shown in Figure 7.25. Single-day MSI results are improved slightly with the addition of single day SAR and also reduce the spread of the data, suggesting that the model is less sensitive to the split of the training and validation data and is therefore more robust. This spread is further reduced with the use of multi-season SAR, and the median loss value is improved in a statistically significant way. In the case of multi-day MSI, the results are less conclusive. While they are indeed better than any single-day experiment, the addition of SAR from a single day or from across the seasons does not significantly improve the loss results.

7.4. Intercrop detection

Considering the uncertainty of intercrop labels, the use of quantitative metrics does not provide meaningful insight into the effectiveness of the models. Efforts were made to observe visually whether certain experiments led to more accurate intercrop detection.

In the case of the original training labels (cocoa and forest), a comparison of predictions overlaid on an intercrop polygon does offer some insight (see Figure 7.26. The stacked multi-day MSI and single-day SAR, which has the highest loss metric (least effective monocrop cocoa prediction) predicts cocoa in the entire polygon with a high certainty and uniformity. This prediction is unlikely considering that the defining characteristic of ”intercrop” is the

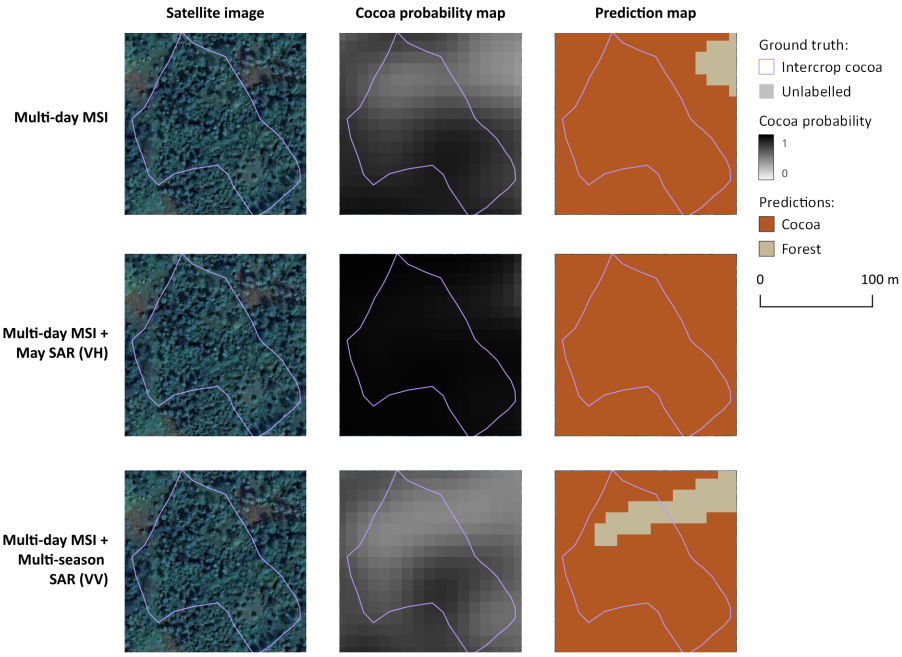


Figure 7.26.: Multi-day MSI + SAR intercrop maps

combination of different crop types. In contrast, the multi-day MSI stack predicts a lower cocoa probability in the top half of the polygon where a more dense tree canopy is visible on the satellite image. Finally, the multi-day MSI and multi-day SAR stack has a lower cocoa probability in the top half and lower right corner of the polygon, where a line of large trees is visible. While it is not possible to determine the exact location of cocoa in the imagery at this resolution, it can be observed that the addition of multi-seasonal SAR enables the model to be sensitive to some vegetation differences that are not detected by MSI only.

Another observation from the intercrop imagery is the influence of shadow on cocoa detection, in addition to cloud cover. In Figure 7.27, the majority of the image is under the shadow of a cloud in the bottom right of the image. The MSI-trained model predicts forest in the clouded and shaded region covering nearly the entire intercrop polygons, likely affected by the lower reflectance in this portion of the image. However, the addition of SAR from a single day and multi-day reduces the amount of forest predicted in the image and does not follow the shadow outline, suggesting that additional data supports the model in overcoming the atmospheric interference with the spectral signatures of vegetation.

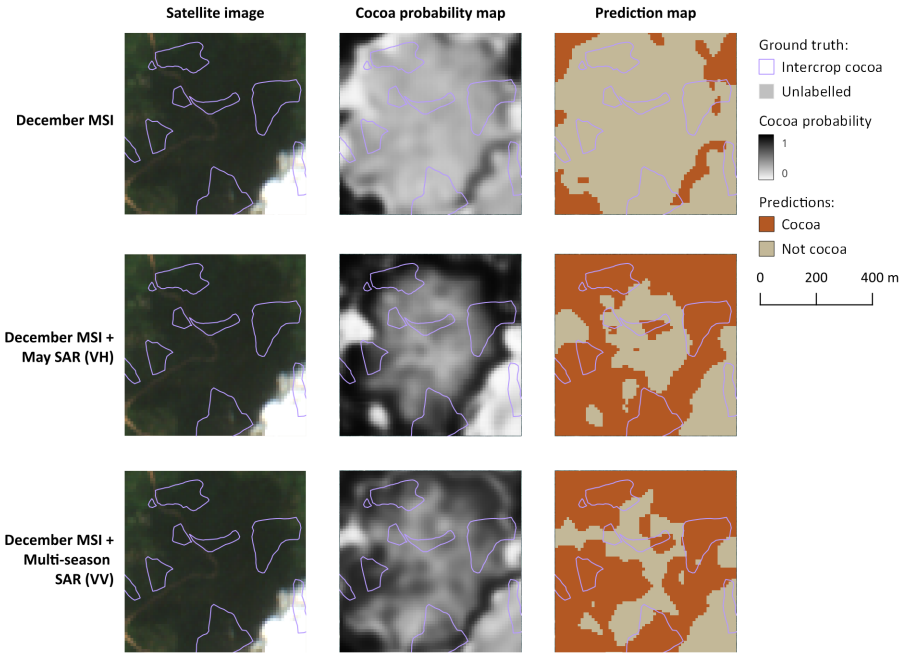
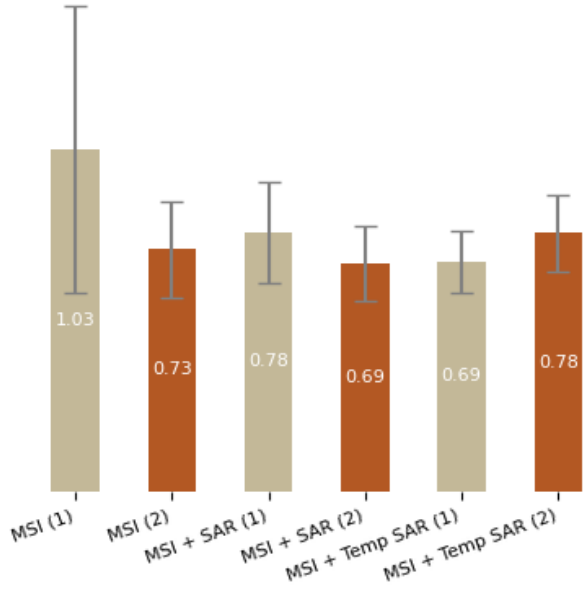


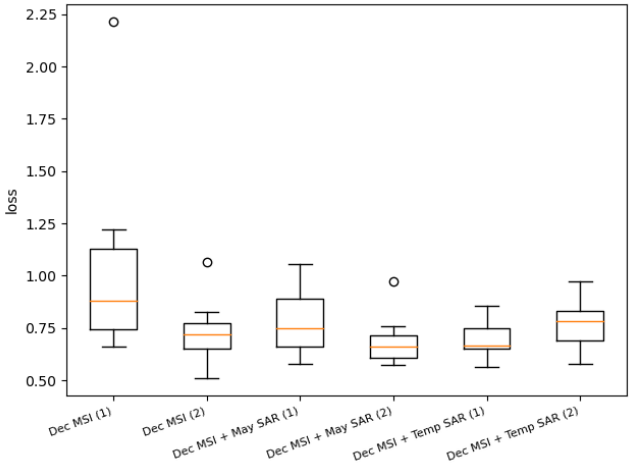
Figure 7.27.: Single-day MSI + SAR intercrop maps

7.5. Set 2 experiments: Impact of labelling

Several experiments are repeated with a more comprehensive dataset. The cocoa class is not changed, but the "non-cocoa" class that previously included only forest reserves, is supplemented with non-cocoa crops. These experiments lead to some different results shown in Figure 7.28, and all metrics are reported in Table A.2.



(a) Loss averages



(b) Loss box plot

Figure 7.28.: Comparison of loss from 10-fold experiments with (1) original labels (Cocoa and Forest) and (2) modified labels (Cocoa and Not Cocoa)

In the case of single-day MSI data, the loss value is significantly lower, and slightly lower in the case of single-day MSI stacked with single-day VH SAR, therefore indicating improved results in both cases. The loss metric for the single-day MSI stacked with multi-season VV SAR is higher, indicating less effective predictions. Unlike the original experiment, the experiment with modified labels does have an outlier in the single day MSI + single day SAR plot. This can be explained by the fact that the non-cocoa crops made up a relatively small

portion of the total ground truth data, therefore in the case that the random split excludes any of these areas from training, the predictions may be negatively affected.

By observing the predicted maps, it can indeed be observed that the single-day MSI + single-day SAR is frequently labelling oil palm crops as cocoa (see for example Figure 7.29), whereas the single-day MSI + multi-day SAR identifies these areas as "not cocoa." These results are reversed in comparison with training using only cocoa and forest labels described in the previous section (illustrated in Figure 7.17).

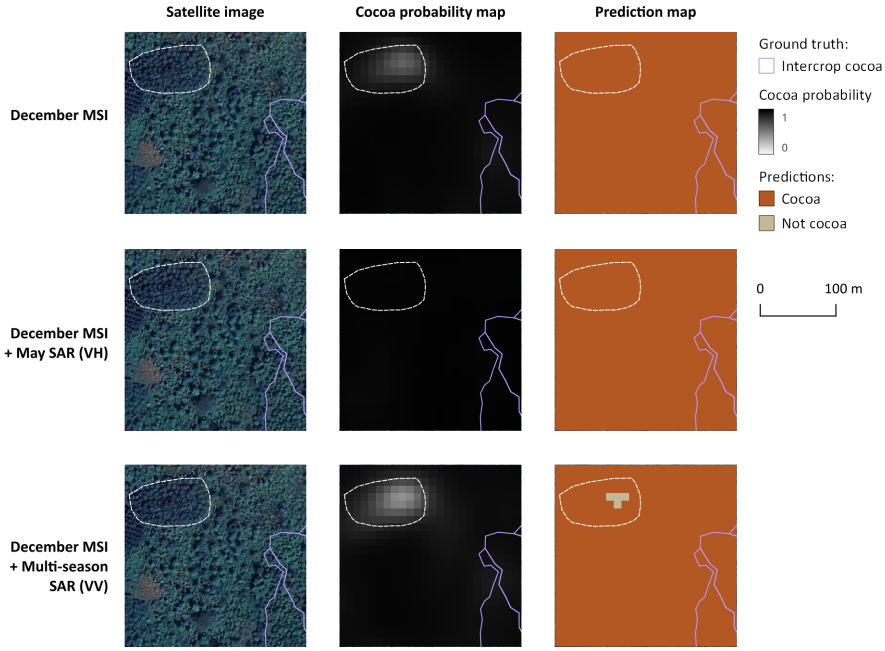


Figure 7.29.: Single-day MSI + Multi-day SAR

Single-day MSI (December, dry season) stacked with single and multi-season SAR datasets showing impact on palm crop differentiation when training with non-cocoa crop labels (white dotted polygon).

A possible explanation for these differences is the characteristics of the palm vegetation. Palm and cocoa are both evergreen, however cocoa has deciduous shade trees above the canopy, which palm does not have. Both are likely to have some ground scattering; during the Dry season, cocoa will have less shade foliage therefore a higher intensity of VV, whereas palm will exhibit a small increase in VV due to a dryer soil and forest will have a consistently low intensity of VV backscatter. Both vegetation types have a rather high amount of volume scattering, although palm appears to have a more lush canopy with fewer gaps, therefore the VH intensity of palm is likely more similar to forest, and higher than that of cocoa.

When training the model with only forest and cocoa labels, with palm included in the "unknown" class, the seasonal variation in palm ground backscatter may appear more similar to that of cocoa, causing better results with the temporal VV stack. In contrast, the single day VH training will group together forest and palm because they both have higher amounts of volume backscatter due to canopy structure.

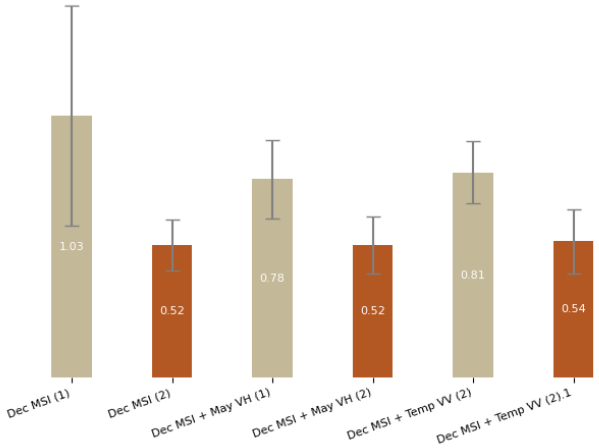
When training the model with cocoa and non-cocoa classes (the latter of which contains forest and non-cocoa crops including palm), the seasonal variation in palm is less likely to be confused for cocoa, and instead the model will learn to better differentiate between their seasonal patterns. However, the wide range of crop types and canopies in the non-cocoa class may create a less clear distinction between cocoa canopy and other canopies, therefore explaining why the single day VH predicts most vegetated areas as cocoa.

7.6. Set 3 experiments: Impact of clouds

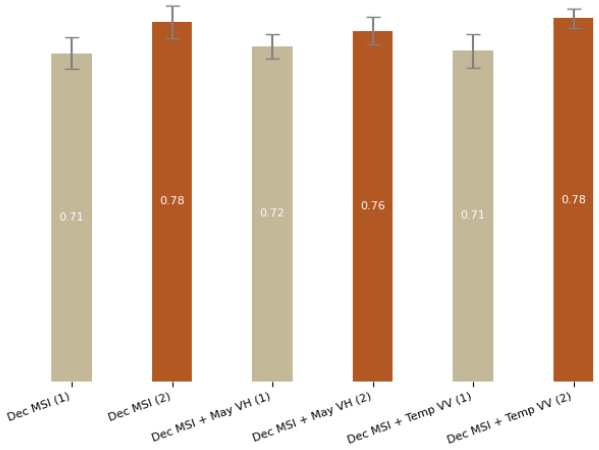
The results described above are generated using training data that is partially cloudy. In the case of the Cocoa/Forest experiment, 12% of the labelled data is covered by cloud. The single-day MSI + SAR stack experiments are repeated with the exclusion of clouded areas, and all metrics are reported in Table A.3. The removal of cloudy training data leads to overall lower loss values and higher IoUs as shown in Figure 7.30. These results suggests that the relevance of SAR is more important in the case of partially cloudy MSI.

7.7. Overview of all experiments

Figure 7.31 depicts the metrics from each of the experiment sets.

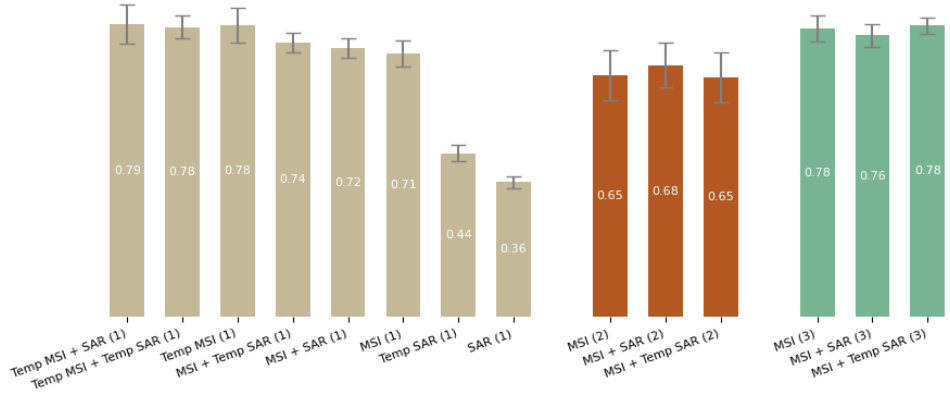


(a) Loss averages

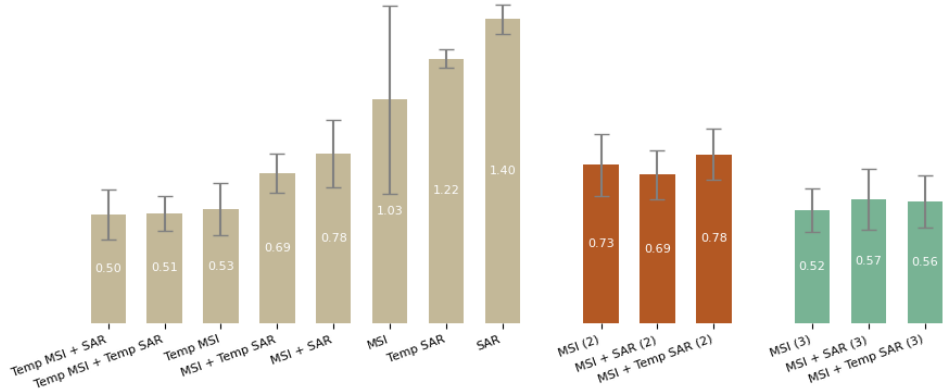


(b) IoU averages

Figure 7.30.: Comparison of loss from 10-fold experiments with (1) original labels (Cocoa and Forest) and (2) original labels without clouded areas



(a) IoU



(b) Loss

Figure 7.31.: Comparison of metrics for experiments from Sets 1, 2 (varied non-cocoa class), and 3 (cloudless)

8. Discussion, conclusion and future work

8.1. Discussion

In this section, each of the hypotheses are addressed in the context of the related work described in Chapter 3 and the research questions outlined in 1.

Firstly, the combination of SAR and MSI was expected to improve the model predictions for monocrop cocoa compared to MSI-only training. The results show that the addition of SAR to *single-day* MSI can improve the accuracy of the model, depending on the seasonality and polarization of the dataset introduced. In the case of a *single-day* image from December (Dry season), the addition of *single-day* VH Wet season imagery slightly improved the loss metrics, and it was found that less cocoa was predicted in forested areas and in visibly identifiable oil palm fields, suggesting an improved distinction between vegetation canopies. Repeating the training with modified labels (Cocoa and Not Cocoa, containing non-cocoa crops) led to improved metrics but a less effective differentiation between cocoa and oil palm, perhaps attributable to the similarities between these crop canopies. This insight suggests that careful labelling and fine-tuning of the model could potentially lay the foundation for cocoa predictions at a higher temporal frequency than is possible with multi-temporal dataset analyses.

Furthermore, the addition of temporal stacks of SAR data covering the wet and dry seasons was expected to lead to better results in comparison with adding single-day SAR. The addition of *multi-seasonal* VV data to *single-day* MSI was indeed found to cause the most significant improvement, decreasing the average loss value from 0.76 to 0.69, as well as reducing the spread of the 10-fold loss values and the vulnerability of the model to imbalanced data splits. This improvement is in line with the findings presented in [1] which identify multi-temporal VV SAR as having the greatest importance in cocoa detection via Random Forest Regression. Nevertheless, closer observation of the predicted images outside of the labelled polygons indicated that the model was identifying some non-cocoa crops (such as oil palm) as cocoa, a mistake that is not captured in the metrics. Repeating the training with modified labels (Cocoa and Not Cocoa, containing non-cocoa crops) led to a higher loss value but a more effective differentiation between cocoa and oil palm.

Regarding the combination of SAR with multi-day MSI datasets, the experiments suggest slightly improved results but the box plot analysis indicates that these differences are not statistically significant. While the metrics are inconclusive, it was observed that some land cover types outside of the ground truth polygons (e.g. urban areas) were detected by the model trained by stacked multi-day MSI and multi-day SAR which was not the case for other models. Further investigation using a more extensive "Not cocoa" class would provide more insight into this phenomenon. Therefore, it can be concluded that the addition of single- and multi-day SAR can improve the differentiation between cocoa and other vegetation types, but is very sensitive to the underlying MSI dataset, the classes used in the training of the model, as well as the SAR seasonality and polarization.

Secondly, the addition of SAR to MSI was expected to enable the detection of cocoa on intercrop polygons where MSI-training would predict only forest. While the nature of the intercrop cocoa labels do not lend themselves to the use of the metrics applied for monocrop cocoa polygons, visual observation of prediction masks does suggest some improvement in the model's effectiveness as there is more differentiation between vegetation types within an intercrop polygon. Due to the lack of attribute data specifying the location of agroforestry cocoa, it is not clear whether the addition of SAR would improve the detection of this type of cocoa crop.

The addition of SAR was expected to improve predictions in images with cloud cover on a single day measurement where MSI reflectance is inhibited. While this was not observed equally across all samples, seasons and polarizations, it was observed that the stacking of May SAR in combined VV and VH polarization lessened the influence of clouds on the resulting predictions. It is likely that a larger training dataset with a lower proportion of cloud cover could improve such results as the model would have additional positive samples from which to learn combined reflectance and backscatter characteristics.

Combining both polarizations of SAR was expected to yield better results than a single polarization as it was expected to capture the effects of volume scattering, surface scattering and double bounce, all of which provide insight into the type of land cover. In the case of SAR only trainings (both single-day and multi-day), the combined polarizations did indeed generate the highest cocoa F1 score and the lowest loss value, indicating the best results. This is in line with the results presented in [49] and [39] which indicate that the use of GLCM derived from combined co- and cross-polarized SAR data led to the best classification of land cover types. However, there is not a clear tendency for the combined polarizations to improve predictions when combined with MSI. In the case of single-day MSI + single-day SAR, single-day MSI + temporal SAR, and temporal MSI + temporal SAR, the loss values of combined polarizations were consistently between that of the best result and the worst result. The reason for this difference in results is not immediately clear, but it is possible that the strength of the reflectance patterns and high volume of MSI data combined with a higher volume of SAR data causes more noisy datasets and impedes the effectiveness of the model.

It is challenging to compare the results of the present research with previous research on the detection of cocoa using ML considering the wide range of methods used, the limited time and attention devoted to hyperparameter tuning in this thesis and the limited temporal and spatial scope of datasets used for the model training. Furthermore, as discussed throughout the results section, the quantitative metrics do not fully capture the effectiveness of the different training datasets. However, it can be helpful to observe trends in the metrics of different studies, therefore the present research results are compared to those of the two previous semantic segmentation studies for cocoa detection presented in [8] and [44]. The experiment yielding the best metrics in this thesis was the one trained with multi-day MSI stacked with May (wet season) VH SAR data and is compared with other studies in table 8.1.

The comparison shows that, in all three cases, the recall score is lower than the precision score, suggesting that it remains challenging to detect all of the cocoa that is present in an area, likely due to complexities such as canopy cover which often mask the presence of cocoa below. Furthermore, the precision values reported by [44] and the current research are both extremely high, which may suggest a tendency in both cases for the model to over-predict cocoa across vast areas of all images. Through visual observation of predicted maps, it appears that this is likely to generate a high number of False Positives outside of the ground

Source	Accuracy	Precision*	Recall*	F1*	IoU*
Filella, 2018 [44]	75.7	99.7	61.1	75.8	58.2
Kalischek et al., 2022 [8]	85.9	88.5	87.2	87.3	n/a
Current study	95.1	98.1	83.2	90.0	78.7

Table 8.1.: Comparison with related work metrics

*cocoa only

truth areas. However, since the metrics are computed only on labelled ground truth, this is not immediately evident in the metrics.

Considering that [44] also acknowledges a low number of non-cocoa ground truth polygons, neither a very high precision value, nor the F1 score that is derived from it, is sufficient to confirm the reliability of either model. With these considerations in mind, it is important to acknowledge the limitations of quantitative metrics in evaluating cocoa detection strategies and emphasize the importance of qualitative evaluation and ground-truth verification of such models if they are to be used for future enforcement of the EUDR.

8.2. Limitations

A number of limitations were encountered in this thesis, regarding the labels used for network training, the satellite imagery involved in the research and the processing steps applied in the methodology.

Label classes. For the purpose of clarity, two classes are used (forest and cocoa), however this excludes a very large portion of the land cover types in the study area such as other crops, water bodies, urban areas etc. As a result, the network is limited in its ability to differentiate between other classes and can therefore not be immediately applied to predict cocoa in regions that are not known cocoa or forest parcels. The additional experiments undergone with the Cocoa and Not cocoa classes (containing non-cocoa crops) do improve the predictions, but a significant number of classes are still missing from the ground truth, such as urban areas, water, and other vegetation types.

Label uncertainty. The forest labels are reserves, which does not guarantee that the polygons contain pure forest without encroaching cocoa crops. Therefore, the network may be mistakenly trained with illegal cocoa or other crops. Furthermore, there was some uncertainty in the label data provided by Meridia: within intercrop farms, the exact location of cocoa is not known and can therefore not be compared to predictions with a high level of certainty. The cocoa dataset does not contain specific attributes related to the type of intercrop farms, which limits conclusions related to agroforest cocoa detection. It would be helpful for further analysis to identify intercrop type during the field visit survey. Finally, due to the use of forest reserve polygons and restricting the study to one S2 tile, there are no masks used for training or testing purposes that contain both forest and cocoa. Considering that patterns along the borders of such land types are complex and likely difficult to distinguish, it would be helpful to collect additional polygons of smaller forest patches adjacent to cocoa parcels during field visits.

Metrics. The labelling challenges described above limit the relevance of the metrics used to evaluate cocoa crop predictions. Furthermore, the metrics are computed after the application of the Argmax function, which means that predictions with cocoa probabilities above 50% are considered "cocoa" predictions, which does not reflect the difference in levels of certainty between different experiments. Computing metrics from predictions above a certain threshold (e.g. 80%) would offer more meaningful comparisons between experiments.

Satellite data. The 10m resolution of S1 and S2 datasets is likely not sufficient to detect many important details that differentiate between intercrop cocoa and forest. For instance, a lower resolution can increase the uncertainty along boundaries between forested and non-forested areas [69] as cited in [61, p. 5]. Another limitation is that the datasets are pre-processed by the data provider, reducing the amount of control over the steps taken in pre-processing and potential loss of information (e.g. when converting S1 data from Single Look Complex (SLC) to GRD).

Temporal resolution. The temporal stacks were limited to four images per year; a higher temporal resolution would increase the amount of information available to inform the training. Furthermore, the temporal resolution of the MSI is limited to the dry season in order to avoid the use of imagery with >15% cloud cover. This limits the effectiveness of training with MSI temporal stack and could be improved by spreading out the data evenly across the year while masking cloudy areas.

Variety of data. Finally, additional steps in preparing the data for training could maximize the use of the datasets, such as rotating and translating the imagery in order to increase the ability of the U-NET to identify land cover classes from different angles. Increasing the study area would also allow for a more diverse representation of cocoa and forest from different regions.

8.3. Conclusion

This section addresses each of the research sub-questions and concludes by drawing conclusions on the main research question for this thesis.

- How does the combination of MSI and SAR data affect the results of cocoa parcel segmentation trained with MSI data from a single day?

The results of Chapter 7 demonstrate how using different combinations of datasets for training can influence the effectiveness of the same U-NET in detecting cocoa crops. Compared to a reference MSI-only experiment, the combination of wet season single-day SAR in VH polarization and dry season single-day MSI is shown to produce slightly improved F1 score and loss metrics, and demonstrates an improvement in the ability for the U-NET to differentiate between forest and cocoa. Furthermore, the combination of multi-seasonal SAR in VV polarization and dry season single-day MSI led to an even greater improvement in the model's abilities to differentiate cocoa from forest. Additional experiments including non-cocoa ground truth led these results to be reversed. However, it was observed that the ability of a specific polarization and season of SAR to distinguish cocoa from palm oil crop was dependent on the labels included in the dataset.

- How does the combination of temporal MSI and SAR data affect the results of cocoa parcel segmentation?

The experimental results suggest that, while there was a slight improvement in the average loss value for stacked multi-day MSI and single or multi-day SAR, this difference is not statistically significant. The spread of the loss values across experimental folds is slightly reduced, suggesting that the model may be more robust in the face of different data splits due to the additional characterization of crops based on texture and moisture. Furthermore, based on observations of the prediction maps, the combination of multi-day MSI and multi-day SAR has potential to improve differentiation with other class types that were not included in this research.

- Why does the use of different polarizations (i.e. VV or VH) affect the influence of SAR datasets on the cocoa segmentation results?

The use of different polarizations of SAR enables the model to learn from different characteristics of surface features. Considering that different types of vegetation will lead to different intensities of co- and cross-polarized backscatter, and that these intensities will vary differently over time, it is important to use data that is relevant to the specific land cover types being considered. The importance of different polarizations varies depending on whether they are stacked with MSI or not. This is likely due to the fact that MSI already contains a significant amount of data related to different spectral signatures, and the addition of SAR should be selected to complement this data. When stacking a single-day dry season MSI imagery, SAR data from the Wet season is has the most positive impact on results when considering VH polarization as it will provide the most insight into canopy structure, and therefore volume backscatter, during the productive season. Temporal SAR added to MSI offers the best results when using the VV polarization, likely because it offers more insight into canopy gaps and ground texture and moisture, which change in more distinct ways between forest and cocoa compared to biomass. The different results observed between experiments also indicate that different vegetation types have more or less similarity in co- and/or cross polarized data, therefore the labels used for training must be carefully selected as they can influence the effectiveness of the model.

- What is the impact of SAR andMSI training data on the detection of intercrop cocoa?

As described in Chapter 7, the effect of SAR data on intercrop cocoa detection is challenging to quantify due to labelling limitations, and difficult to observe visually due to the challenge of identifying intercrop cocoa from satellite imagery. However, by comparing map predictions, it could be observed that intercrop areas are not homogeneously predicted as either forest or cocoa, which suggests some improvement in the sensitivity of the model. Therefore, more detailed labels indicating the exact location of cocoa would need to be used to evaluate the effectiveness of stacked MSI and SAR in more depth. Furthermore, given the complexity of agroforestry cocoa and the spatial resolution of GRD SAR at 20 x 22 m, this kind of dataset is not sufficient for the detection of cocoa within complex and dense canopies with a high level of certainty.

Conclusion on the main research question:

- To what extent can a Convolutional Neural Network trained with multispectral *and* SAR datasets enable the automated detection of cocoa crops in Ghana?

This thesis aims to evaluate the influence of introducing SAR datasets into MSI on the ability for a CNN to detect cocoa. The research carried out with a U-NET combines different stacks of data and compares the resulting predictions and metrics. Quantitative and qualitative evaluations of the results suggest that the addition of SAR can improve predictions when

that additional data is selected carefully to provide additional information (i.e. by making use of relevant seasons and polarizations). Furthermore, the research has demonstrated the need for high quality labels and a careful choice of classes used in the classification in order to generate results that are relevant to the purpose of the cocoa map. Whereas this research offers insight into the effect of SAR in differentiating between cocoa, forest and other crops, it is expected that the conclusions drawn in this thesis are relevant to cocoa classification with a greater number of "non-cocoa" ground truth classes. Finally, whereas evaluating the potential for combined dataset to better detect intercrop cocoa was an important motivation for this thesis, the lack of precise ground truth polygons makes it difficult to draw clear conclusions on this subject. Visual analysis suggests that SAR does improve the distinction between vegetation types. However, the lower resolution of SAR datasets and the complexity of agroforestry land cover suggest that it may not be suitable for detecting cocoa plants within the agroforest canopy.

8.4. Contribution

This thesis contributes to a more in-depth understanding of the potential, limitations and complexity of combining S1 and S2 datasets for cocoa crop detection. In particular, the findings address the following topics:

Temporal data: the research provides an exploration into the the importance of temporal datasets that capture changes in reflectance, moisture and biomass of different land cover types over the seasons. It suggests ways to combine temporal SAR data with single day MSI (and vice-versa) in order to improve predictions and potentially reduce the amount of data needed to identify cocoa.

Polarization: the research demonstrates the importance of carefully choosing SAR polarizations when stacking with MSI data by showing the outcome of different dataset stacks depending on the scenario (e.g. SAR alone, temporal SAR and MSI, single-day SAR and MSI).

Labelling: in this work, the complexity of labelling decisions and uncertainty has been described in detail and contributes an improved understanding of the necessity for high accuracy cocoa and forest ground truth with detailed attributes.

8.5. Future work

As the reporting requirements for cocoa producers continue to grow in response to the threat of deforestation, future work on the use of ML and remotely sensed data will play an important role in producing more reliable cocoa maps. There are several directions for future research to be considered.

First, considering the high cost of commercial remotely sensed data, it is important to further explore the potential for detecting cocoa using DL and open-source S1 and S2 data as has been done in this thesis. The first step to deepen this research would be to apply hyperparameter tuning in order to optimize the results using the current architecture and datasets. Another way to build on the current research would be to increase the temporal resolution of the dataset stacks in order to improve the models ability to detect land cover changes over the seasons, and during a longer timespan. Considering the improvements introduced by a more

varied ”Not Cocoa” ground truth class, it would also be beneficial to conduct additional experiments with a more extensive ”Not Cocoa” class in order to observe the impacts on trainings with different datasets.

Feature extraction and selection can reduce the amount of data that must be processed by the model, and can sometimes improve classification results [5]. Furthermore, some previous ML cocoa detection algorithms have made use of vegetation indices (e.g. NDVI) and texture metrics (e.g. GLCM). Considering that texture is considered in the U-NET kernels, it may be useful to employ similar analyses prior to the CNN in order to guide the network towards identifying meaningful patterns. For example, the SAR datasets could be pre-processed by applying the GLCM [39]. MSI can be further processed using vegetation indices which can identify unique characteristics of different vegetation types [70]. One commonly used vegetation index is the NDVI, and a previous study has applied the Tasseled Cap Analysis (TCA) which uses more bands than NDVI and combines characteristics of brightness, greenness and wetness [5]. One solution to reduce the effect of the soil moisture is the Cross Ratio (CR), the ratio between VH and VV, which increases alongside biomass [46, p. 13]. Future research could explore the impact of using vegetation indices and metrics on the training speed and effectiveness of the U-NET.

Considering that cocoa trees and tropical forest trees and shade trees can have a significant difference in height [8] and additional collected or inferred vegetation height data could be integrated with MSI and SAR to further improve model predictions.

Another way to develop this research would be to refine the training process by employing a more appropriate loss function, such as Lovasz loss function, which optimizes the IoU rather than focusing on individual pixels [24, p. 1463], or the Dice loss, which aims to optimize the F1 score [8]. It is possible that the U-NET, although one of the most popular networks for semantic segmentation, is not the best network for this use case. Other CNN architectures may provide more adapted convolutional layers [8]. Therefore, carrying out similar experiments with a more complex architecture may yield improved results.

Considering the limitations of S1 and S2 datasets described previously, including the spatial resolution and rather short wavelength of C-band SAR data, future research could explore the potential of semantic segmentation of cocoa using more precise remotely sensed data. One possible avenue of research would be exploring the use of longer wavelengths (such as L- and P-bands) which can penetrate deeper into the canopy in combination with S2 MSI. Another avenue would be the exploration of higher resolution MSI and hyperspectral imagery in combination with SAR in order to explore the potential for such datasets to enable deep learning to detect cocoa in complex ecosystems such as agroforestry crops. Finally, it would be beneficial to analyze the role of spatial resolution in the segmentation outputs for different forms of cocoa crops.

A. Metrics

DS = Dry Season
MS = Multi Season
*cocoa metric

#	label	time (s)	loss	accuracy	IoU*	precision*	recall*	F1*
1	Dec MSI	624.89	1.0304	0.9249	0.7086	0.8923	0.7975	0.8423
2	March MSI	797.97	0.7622	0.9384	0.7670	0.9496	0.8115	0.8752
3	Temp MSI	735.07	0.5254	0.9501	0.7842	0.9803	0.8320	0.9001
4	May VV	1216.9988	1.6587	0.5063	0.1087	0.4891	0.2069	0.2908
5	May VH	1148.9945	1.5931	0.6019	0.2251	0.5923	0.3105	0.4074
6	May VV VH	957.91	1.6066	0.5599	0.1543	0.5263	0.2427	0.3322
7	Jan VV	1038.0268	1.5291	0.6744	0.3106	0.6597	0.3823	0.4840
8	Jan VH	938.9624	1.6129	0.7213	0.6819	0.6819	0.4369	0.5325
9	Jan VV VH	893.8322	1.4008	0.7195	0.3630	0.7074	0.4357	0.5393
10	Temp VV (DS)	689.24	1.5482	0.5284	0.2645	0.7401	0.3231	0.4498
11	Temp VH (DS)	774.93	1.3449	0.7095	0.3908	0.7592	0.4581	0.5714
12	Temp VV VH (DS)	757.40	1.3028	0.7654	0.4138	0.7336	0.4946	0.5908
13	Temp VV (MS)	779.66	1.4572	0.6932	0.3291	0.6745	0.4081	0.5085
14	Temp VH (MS)	802.28	1.2549	0.7840	0.4422	0.7695	0.5158	0.6176
15	Temp VV VH (MS)	666.53	1.2191	0.7776	0.4405	0.7819	0.5072	0.6153
16	Dec MSI + Jan VV	841.14	0.7875	0.9310	0.7319	0.9217	0.8008	0.8570
17	Dec MSI + Jan VH	820.53	0.9741	0.7928	0.5137	0.9315	0.5292	0.6749
18	Dec MSI + Jan VV VH	749.13	0.8758	0.9287	0.7279	0.9334	0.7876	0.8543
19	Dec MSI + May VV	738.84	0.8556	0.9223	0.7130	0.9318	0.7721	0.8445
20	Dec MSI + May VH	793.41	0.7809	0.9263	0.7230	0.9275	0.7862	0.8511
21	Dec MSI + May VV VH	761.57	0.8352	0.9256	0.7226	0.9401	0.7775	0.8511
22	Dec MSI + Temp VV	800.59	0.6919	0.9323	0.7379	0.9331	0.7997	0.8613
23	Dec MSI + Temp VH	724.16	0.7971	0.9359	0.7473	0.9172	0.8205	0.8662
24	Dec MSI + Temp VV VH	756.44	0.7725	0.9293	0.7320	0.9185	0.7998	0.8550
25	Temp MSI + May VV	723.67	0.5432	0.9464	0.7697	0.9767	0.8200	0.8915
26	Temp MSI + May VH	771.41	0.4989	0.9506	0.7869	0.9814	0.8320	0.9005
27	Temp MSI + May VV VH	743.04	0.5147	0.9503	0.7849	0.9791	0.8331	0.9002
28	Temp MSI + Temp VV	731.36	0.5066	0.9492	0.7796	0.9793	0.8278	0.8972

Table A.1.: Set 1 experiment results: average metrics across all folds

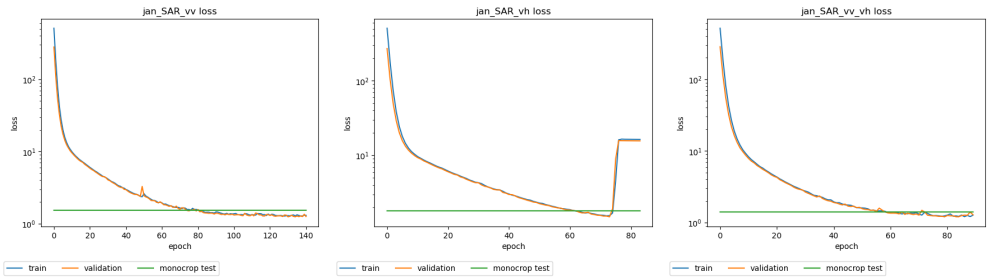
#	label	time (s)	loss	accuracy	IoU*	precision*	recall*	F1*
29	Dec MSI	671.23	0.7296	0.9298	0.6499	0.9355	0.6816	0.7887
30	Dec MSI + May VH	712.4836	0.6857	0.9379	0.6771	0.9440	0.7024	0.7885
31	Dec MSI + Temp VV	502.6102	0.7783	0.9312	0.6450	0.8799	0.7163	0.7444

Table A.2.: Set 2 experiment results: modified labels with more varied ”Not cocoa” class

#	label	time (s)	loss	accuracy	IoU*	precision*	recall*	F1*
32	Dec MSI	875.7600	0.5202	0.9520	0.7761	0.9744	0.7933	0.8640
33	Dec MSI + May VH	705.9654	0.5712	0.9426	0.7412	0.9640	0.7644	0.8380
34	Dec MSI + Temp VV	688.5537	0.5600	0.9527	0.7737	0.9534	0.8055	0.8542

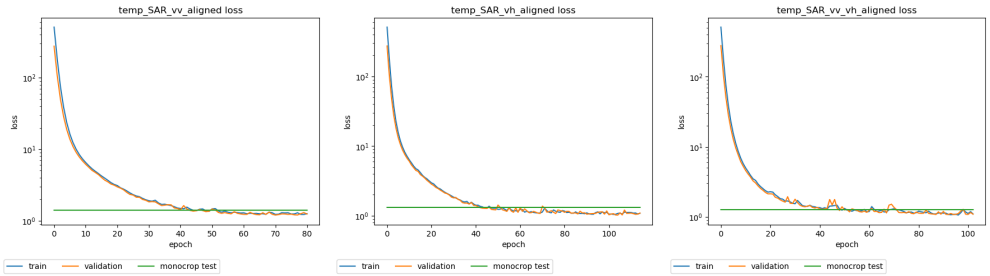
Table A.3.: Set 3 experiment results: original labels with clouds removed

B. Plots and graphs



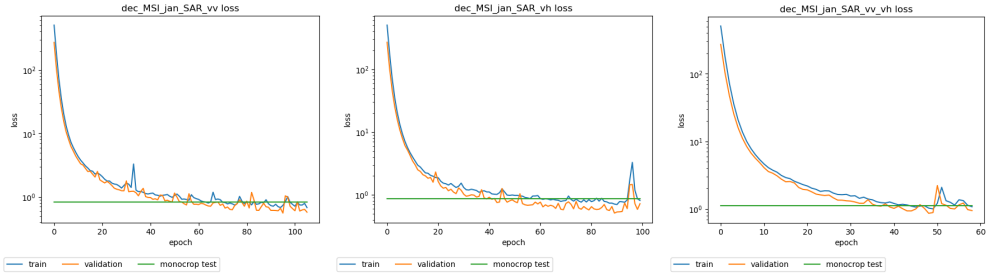
(a) Jan VV (b) Jan VH (c) Jan VV VH

Figure B.1.: Loss curves for single-day SAR-only reference experiments (January)



(a) Multi-day VV (b) Multi-day VH (c) Multi-day VV VH

Figure B.2.: Loss curves for multi-day SAR-only reference experiments (dry season)



(a) Dec MSI + Jan VV (b) Dec MSI + Jan VH (c) Dec MSI + Jan VV VH

Figure B.3.: Loss curves for December MSI + January SAR

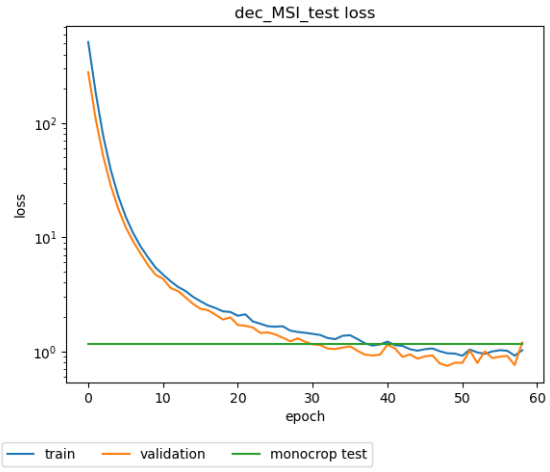


Figure B.4.: Loss curve for single-day MSI-trained model

C. Reproducibility self-assessment

C.1. Marks for each of the criteria

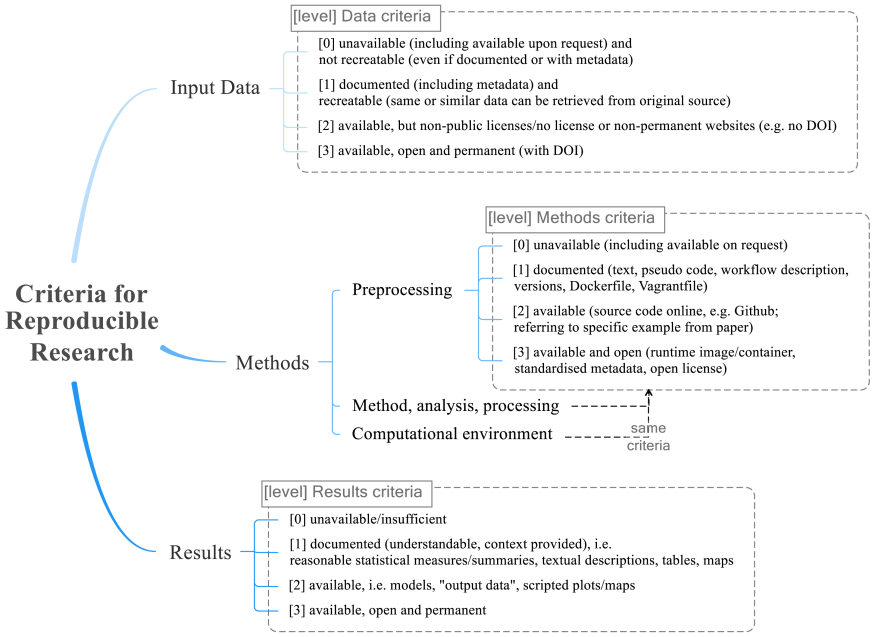


Figure C.1.: Reproducibility criteria.

category	criteria	grade
1. Input data	Satellite imagery	3
	Ground truth polygons	0
2. Methods	Pre-processing	2
	Analysis & processing	2
	Computational environment	3
3. Results		1

Table C.1.: Evaluation of reproducibility criteria

C.2. Reflection on reproducibility

Input data
The S1 and S2 datasets are openly available via the WEkEO platform in the pre-processed formats. It can also be obtained directly via the European Space Agency's Copernicus Data Hub and processed via the European Space Agency's Sentinel Application Platform (SNAP).

The ground truth polygons obtained from Meridia for this thesis are confidential and proprietary data collected from or about the company's clients and cannot be made available. Furthermore, cocoa ground truth is a commercial resource that is not openly available and is challenging to collect.

Methods
The implementation of this research is described in detail in Chapter 5 and Chapter 6. All tools for pre-processing are open source and available.

The pre-processing and post-processing scripts were implemented in Visual Studio Code and are available via Github. QGIS is an open-source GIS software that was used for visualization and pre-processing purposes.

The CNN implementation was carried out via AWS Studio Lab, a free version of AWS Sagemaker which makes use of Jupyter notebooks and cloud-based storage. Studio Lab provides access to a GPU runtimes for 8 hours per day (each runtime lasts a maximum of 4 hours) and 15 GB of persistent storage. These resources were sufficient to carry out 1-2 experiments in one 24 hour cycle. The Jupyter notebook used in this research (including the U-NET architecture, data splits and experiment folds, and visualization) are available via Github.

Results
The results are documented and described in Chapter 7 in the form of maps, tables and graphs. A summary of all experiment metric averages can be found in Appendix A, and additional plots and maps are included in Appendix B and **??**.

Bibliography

[1] Frederick N. Numbisi and Frieke Van Coillie. Does sentinel-1a backscatter capture the spatial variability in canopy gaps of tropical agroforests? a proof-of-concept in cocoa landscapes in cameroon. *Remote Sensing*, 12(24), 2020. ISSN 2072-4292. doi: 10.3390/rs12244163. URL https://www.mdpi.com/2072-4292/12/24/4163.

[2] Rahul Kumar, Amit Kumar, and Purabi Saikia. *Deforestation and Forests Degradation Impacts on the Environment*, chapter Deforestation and Forests Degradation Impacts on the Environment, pages 19–39. Springer International Publishing, Cham, 2022. ISBN 978-3-030-95542-7. doi: 10.1007/978-3-030-95542-7_2. URL https://doi.org/10.1007/978-3-030-95542-7_2.

[3] European Commission. Proposal for a regulation of the european parliament and of the council on the making available on the union market as well as export from the union of certaincommodities and products associated with deforestation and forest degradation andrepealing regulation (eu) no 995/2010, 2021. URL https://environment.ec.europa.eu/system/files/2021-11/COM_2021_706_1_EN_ACT_part1_v6.pdf.

[4] Global forest resources assessment. *Food and Agriculture Organization of the United Nations*, 2020. doi: https://doi.org/10.4060/ca9825en. URL https://www.fao.org/3/ca9825en/ca9825en.pdf.

[5] Itohan-Osa Abu, Zoltan Szantoi, Andreas Brink, Marine Robuchon, and Michael Thiel. Detecting cocoa plantations in côte d'ivoire and ghana and their implications on protected areas. *Ecological Indicators*, 129:107863, 2021. ISSN 1470-160X. doi: https://doi.org/10.1016/j.ecolind.2021.107863. URL https://www.sciencedirect.com/science/article/pii/S1470160X21005288.

[6] George Ashiagbor, Eric K. Forkuo, Winston A. Asante, Emmanuel Acheampong, Jonathan A. Quaye-Ballard, Prince Boamah, Yakubu Mohammed, and Ernest Foli. Pixel-based and object-oriented approaches in segregating cocoa from forest in the juabeso-bia landscape of ghana. *Remote Sensing Applications: Society and Environment*, 19:100349, 2020. ISSN 2352-9385. doi: https://doi.org/10.1016/j.rsase.2020.100349. URL https://www.sciencedirect.com/science/article/pii/S2352938520301178.

[7] Luis Orozco-Aguilar, Arlene Lopez-Sampson, Mariela E. Leandro-Munoz, Valentina Robiglio, Martin Reyes, Melanie Bordeaux, Norvin Sepulveda, and Eduardo Somarriba. Elucidating pathways and discourses linking cocoa cultivation to deforestation, reforestation, and tree cover change in nicaragua and peru. *Frontiers in Sustainable Food Systems*, 5, 2021. ISSN 2571-581X. doi: 10.3389/fsufs.2021.635779. URL https://www.frontiersin.org/articles/10.3389/fsufs.2021.635779.

[8] Nikolai Kalischek, Nico Lang, Cécile Renier, Rodrigo Daudt, Thomas Addoah, William Thompson, Wilma Blaser-Hart, Rachael Garrett, and Jan Wegner. Satellite-based high-resolution maps of cocoa planted area for côte d'ivoire and ghana. 06 2022.

[9] Meridia. Field data solutions for smallholder supply chains, n.d. URL https://www.meridia.land/.

[10] George Ashiagbor, Winston Adams Asante, Eric Kwabena Forkuo, Emmanuel Acheampong, and Ernest Foli. Monitoring cocoa-driven deforestation: The contexts of encroachment and land use policy implications for deforestation free cocoa supply chains in ghana. *Applied Geography*, 147:102788, 2022. ISSN 0143-6228. doi: https://doi.org/10.1016/j.apgeog.2022.102788. URL https://www.sciencedirect.com/science/article/pii/S014362282200159X.

[11] vivideconomics. State and trends of deforestion in cote d'ivoire (2019-2020)., July 2020. URL https://www.vivideconomics.com/wp-content/uploads/2020/07/State-and-Trends-of-Deforestation-in-CdI-1.pdf. Accessed via IMAGES.

[12] M. I. Jordan and T. M. Mitchell. Machine learning: Trends, perspectives, and prospects. *Science*, 349(6245):255–260, 2015. doi: 10.1126/science.aaa8415. URL https://www.science.org/doi/abs/10.1126/science.aaa8415.

[13] A. Garcia-Garcia, S. Orts-Escolano, S.O. Oprea, V. Villena-Martinez, and J. Garcia-Rodriguez. A review on deep learning techniquesapplied to semantic segmentation. *arXiv preprint*, 2017. URL https://arxiv.org/pdf/1704.06857.pdf.

[14] A. Karpathy. Cs231n: Convolutional neural networks for visual recognition. neural networks. Website, 2018. URL https://cs231n.github.io/neural-networks-1/.

[15] Grant Sanderson. But what is a neural network? — chapter 1, deep learning. Video, October 2017. URL https://www.youtube.com/watch?v=aircAruvnKk&list=PLZHQObOWTQDNU6R1_67000Dx_ZCJB-3pi&ab_channel=3Blue1Brown.

[16] Fei-Fei Li, Yunzhu Li, and Ruohan Gao. Image classification with cnns, April 2023. URL http://cs231n.stanford.edu/slides/2023/lecture_5.pdf.

[17] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998. doi: 10.1109/5.726791.

[18] Vincent Dumoulin and Francesco Visin. A guide to convolution arithmetic for deep learning. *ArXiv e-prints*, mar 2016.

[19] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. Imagenet classification with deep convolutional neural networks. *Commun. ACM*, 60(6):84–90, may 2017. ISSN 0001-0782. doi: 10.1145/3065386. URL https://doi.org/10.1145/3065386.

[20] Udeme Udofia. Basic overview of convolutional neural network (cnn). *Medium*, 2018. URL https://medium.com/dataseries/basic-overview-of-convolutional-neural-network-cnn-4fcc7dbb4f17.

[21] Fei-Fei Li, Justin Johnson, and Serena Yeung. Training neural networks, part i, April 2017. URL http://cs231n.stanford.edu/slides/2017/cs231n_2017_lecture6.pdf.

[22] Fei-Fei Li, Justin Johnson, and Serena Yeung. Loss functions and optimization, April 2017. URL http://cs231n.stanford.edu/slides/2017/cs231n_2017_lecture3.pdf.

[23] Pieter-Tjerk de Boer, Dirk P. Kroese, Shie Mannor, and Reuven Y. Rubinstein. A tutorial on the cross-entropy method. *Annals of operations research*, 134(1):19–67, January 2005. ISSN 0254-5330. doi: 10.1007/s10479-005-5724-z.

[24] Iñigo Alonso, Matan Yuval, Gal Eyal, Tali Treibitz, and Ana C. Murillo. Coralseg: Learning coral segmentation from sparse annotations. *Journal of Field Robotics*, 36(8): 1456–1477, 2019. doi: https://doi.org/10.1002/rob.21915. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/rob.21915.

[25] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pages 234–241, 2015. URL https://link.springer.com/content/pdf/10.1007/978-3-319-24574-4.pdf?pdf=button.

[26] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. *CoRR*, abs/1411.4038, 2014. URL http://arxiv.org/abs/1411.4038.

[27] Andrew Clark, Stuart Phinn, and Peter Scarth. Optimised u-net for land use–land cover classification using aerial photography. *PFG–Journal of Photogrammetry, Remote Sensing and Geoinformation Science*, pages 1–23, 2023.

[28] Patrik Olã Bressan, José Marcato Junior, José Augusto Correa Martins, Maximilian Jaderson de Melo, Diogo Nunes Gonçalves, Daniel Matte Freitas, Ana Paula Marques Ramos, Michelle Taís Garcia Furuya, Lucas Prado Osco, Jonathan de Andrade Silva, Zhipeng Luo, Raymundo Cordero Garcia, Lingfei Ma, Jonathan Li, and Wesley Nunes Gonçalves. Semantic segmentation with labeling uncertainty and class imbalance applied to vegetation mapping. *International Journal of Applied Earth Observation and Geoinformation*, 108:102690, 2022. ISSN 1569-8432. doi: https://doi.org/10.1016/j.jag.2022.102690. URL https://www.sciencedirect.com/science/article/pii/S0303243422000162.

[29] Pontus Olofsson, Giles M. Foody, Martin Herold, Stephen V. Stehman, Curtis E. Woodcock, and Michael A. Wulder. Good practices for estimating area and assessing accuracy of land change. *Remote Sensing of Environment*, 148:42–57, 2014. ISSN 0034-4257. doi: https://doi.org/10.1016/j.rse.2014.02.015. URL https://www.sciencedirect.com/science/article/pii/S0034425714000704.

[30] Mark Schutera, Luca Rettenberger, Christian Pylatiuk, and Markus Reischl. Methods for the frugal labeler: Multi-class semantic segmentation on heterogeneous labels. *PLOS ONE*, 17(2):1–14, 02 2022. doi: 10.1371/journal.pone.0263656. URL https://doi.org/10.1371/journal.pone.0263656.

[31] Spectral signatures, . URL https://www.esa.int/SPECIALS/Eduspace_EN/SEMPNQ3Z2OF_0.html.

[32] João E. Batista, Nuno M. Rodrigues, Ana I. R. Cabral, Maria J. P. Vasconcelos, Adriano Venturieri, Luiz G. T. Silva, and Sara Silva. Optical time series for the separation of land cover types with similar spectral signatures: cocoa agroforest and forest. *International Journal of Remote Sensing*, 43(9):3298–3319, 2022. doi: 10.1080/01431161.2022.2089540. URL https://doi.org/10.1080/01431161.2022.2089540.

[33] S. Saatchi, D. Agosti, K. Alger, J. Delabie, and J. Musinsky. Examining fragmentation and loss of primary forest in the southern bahian atlantic forest of brazil with radar imagery. *Conservation Biology*, 15(4):867–875, 2001. doi: https://doi.org/10.1046/j.1523-1739.2001.015004867.x. URL https://conbio.onlinelibrary.wiley.com/doi/abs/10.1046/j.1523-1739.2001.015004867.x.

[34] Alaska satellite facility. introduction to sar. URL https://hyp3-docs.asf.alaska.edu/guides/introduction_to_sar/.

[35] D. L. Evans, C. Elachi, E. R. Stofan, B. Holt, J. B. Way, M. Kobrick, M. Vogt, S. Wall, J. van Zyl, M. Schier, H. Öttl, and P. Pampaloni. The shuttle imaging radar-c and x-sar mission. *Eos, Transactions American Geophysical Union*, 74(13):145–158, 1993. doi: https://doi.org/10.1029/93EO00097. URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/93EO00097.

[36] *Sentinel-1 SAR User Guide*. European Space Agency. URL https://sentinels.copernicus.eu/web/sentinel/user-guides/sentinel-1-sar.

[37] Sahel Mahdavi, Meisam Amani, and Yasser Maghsoudi. The effects of orbit type on synthetic aperture radar (sar) backscatter. *Remote Sensing Letters*, 10(2):120–128, 2019. doi: 10.1080/2150704X.2018.1530481. URL https://doi.org/10.1080/2150704X.2018.1530481.

[38] Hari Shanker Srivastava, Parul Patel, Yamini Sharma, and Ranganath R. Navalgund. Large-area soil moisture estimation using multi-incidence-angle radarsat-1 sar data. *IEEE Transactions on Geoscience and Remote Sensing*, 47(8):2528–2535, 2009. doi: 10.1109/TGRS.2009.2018448.

[39] Frederick N. Numbisi, Frieke M. B. Van Coillie, and Robert De Wulf. Delineation of cocoa agroforests using multiseason sentinel-1 sar images: A low grey level range reduces uncertainties in glcm texture-based mapping. *ISPRS International Journal of Geo-Information*, 8(4), 2019. ISSN 2220-9964. doi: 10.3390/ijgi8040179. URL https://www.mdpi.com/2220-9964/8/4/179.

[40] Daniel Tutu Benefoh, Grace B. Villamor, Meine van Noordwijk, Christian Borgemeister, Winston A. Asante, and Kwabena O. Asubonteng. Assessing land-use typologies and change intensities in a structurally complex ghanaian cocoa landscape. *Applied Geography*, 99:109–119, 2018. ISSN 0143-6228. doi: https://doi.org/10.1016/j.apgeog.2018.07.027. URL https://www.sciencedirect.com/science/article/pii/S0143622817310470.

[41] Dan Kanmegne Tamga, Hooman Latifi, Tobias Ullmann, Roland Baumhauer, Michael Thiel, and Jules Bayala. Modelling the spatial distribution of the classification error of remote sensing data in cocoa agroforestry systems. *Agroforestry Systems*, 97(1):109–119, oct 2022. doi: https://doi.org/10.1007/s10457-022-00791-2.

[42] Thomas Blaschke and Josef Strobl. What's wrong with pixels? some recent developments interfacing remote sensing and gis. *GIS – Zeitschrift für Geoinformationssysteme*, 14:12 – 17, 06 2001.

[43] S. Erasmi and A. Twele. Regional land cover mapping in the humid tropics using combined optical and sar satellite data—a case study from central sulawesi, indonesia. *International Journal of Remote Sensing*, 30(10):2465–2478, 2009. doi: 10.1080/01431160802552728. URL https://doi.org/10.1080/01431160802552728.

[44] Guillem Bonet Filella. Cocoa segmentation in satellite images with deep learning. Master's thesis, ETH Zurich, 2018. URL https://ethz.ch/content/dam/ethz/special-interest/baug/igp/photogrammetry-remote-sensing-dam/documents/pdf/Student_Theses/BA_BonetFilella.pdf.

[45] Marius Rüetschi, Michael E. Schaepman, and David Small. Using multitemporal sentinel-1 c-band backscatter to monitor phenology and classify deciduous and coniferous forests in northern switzerland. *Remote Sensing*, 10(1), 2018. ISSN 2072-4292. doi: 10.3390/rs10010055. URL https://www.mdpi.com/2072-4292/10/1/55.

[46] Mariette Vreugdenhil, Wolfgang Wagner, Bernhard Bauer-Marschallinger, Isabella Pfeil, Irene Teubner, Christoph Rüdiger, and Peter Strauss. Sensitivity of sentinel-1 backscatter to vegetation dynamics: An austrian case study. *Remote Sensing*, 10(9), 2018. ISSN 2072-4292. doi: 10.3390/rs10091396. URL https://www.mdpi.com/2072-4292/10/9/1396.

[47] Johannes Reiche, Eliakim Hamunyela, Jan Verbesselt, Dirk Hoekman, and Martin Herold. Improving near-real time deforestation monitoring in tropical dry forests by combining dense sentinel-1 time series with landsat and alos-2 palsar-2. *Remote Sensing of Environment*, 204:147–161, 2018. ISSN 0034-4257. doi: https://doi.org/10.1016/j.rse.2017.10.034. URL https://www.sciencedirect.com/science/article/pii/S0034425717304959.

[48] Carolin Thiel, Oliver Cartus, Robert Eckardt, Nicole Richter, Christian Thiel, and Christiane Schmullius. Analysis of multi-temporal land observation at c-band. In *2009 IEEE International Geoscience and Remote Sensing Symposium*, volume 3, pages III–318–III–321, 2009. doi: 10.1109/IGARSS.2009.5417764.

[49] F. N. Numbisi, F. Van Coillie, and R. De Wulf. Multi-date sentinel1 sar image textures discriminate perennial agroforests in a tropical forest-savannah transition landscape. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-1:339–346, 2018. doi: 10.5194/isprs-archives-XLII-1-339-2018. URL https://isprs-archives.copernicus.org/articles/XLII-1/339/2018/.

[50] Bismark Kwesi Asitoakor, Philippe Vaast, Anders Ræbild, Hans Peter Ravn, Vincent Yao Eziah, Kwadwo Owusu, Eric Opoku Mensah, and Richard Asare. Selected shade tree species improved cocoa yields in low-input agroforestry systems in ghana. *Agricultural Systems*, 202:103476, 2022. ISSN 0308-521X. doi: https://doi.org/10.1016/j.agsy.2022.103476. URL https://www.sciencedirect.com/science/article/pii/S0308521X22001123.

[51] Kenneth Peprah. Cocoa plant, people and profit in ghana. In Peter Osobase Aikpokpodion, editor, *Theobroma Cacao*, chapter 8. IntechOpen, Rijeka, 2019. doi: 10.5772/intechopen.81991. URL https://doi.org/10.5772/intechopen.81991.

[52] James S. Kaba, Alexander Otu-Nyanteh, and Akwasi A. Abunyewa. The role of shade trees in influencing farmers' adoption of cocoa agroforestry systems: Insight from semi-deciduous rain forest agroecological zone of ghana. *NJAS - Wageningen Journal of Life Sciences*, 92:100332, 2020. ISSN 1573-5214. doi: https://doi.org/10.1016/j.njas.2020.100332. URL https://www.sciencedirect.com/science/article/pii/S1573521420300336.

[53] Myriam Rojas, Arne Hommes, Hero Heeres, and Farid Chejne Janna. Physicochemical phenomena in the roasting of cocoa (theobroma cacao l.). *Food Engineering Reviews*, 14, 01 2022. doi: 10.1007/s12393-021-09301-z.

[54] Richard Asare, Sonii David, et al. *Planting, replanting and tree diversification in cocoa systems-learning about sustainable cocoa production: a guide for participatory farmer training*. Number 13. Forest & Landscape Denmark, 2010.

[55] Denis J Sonwa, Stephan F Weise, Goetz Schroth, Marc JJ Janssens, and Howard-Yana Shapiro. Structure of cocoa farming systems in west and central africa: a review. *Agroforestry Systems*, 93:2009–2025, 2019.

[56] Winston Adams Asante, Gabriel Ahoma, Benjamin Apraku Gyampoh, Boateng Kyereh, and Richard Asare. Upper canopy tree crown architecture and its implications for shade in cocoa agroforestry systems in the western region of ghana. *Trees, Forests and People*, 5:100100, 2021. ISSN 2666-7193. doi: https://doi.org/10.1016/j.tfp.2021.100100. URL https://www.sciencedirect.com/science/article/pii/S266671932100039X.

[57] Denis Sonwa, Stephan Weise, Bernard Nkongmeneck, Mathurin Tchatat, and Marc Janssens. Structure and composition of cocoa agroforests in the humid forest zone of southern cameroon. *Agroforestry Systems*, 91, 06 2017. doi: 10.1007/s10457-016-9942-y.

[58] W.J. Blaser-Hart, S.P. Hart, J. Oppong, D. Kyereh, E. Yeboah, and J. Six. The effectiveness of cocoa agroforests depends on shade-tree canopy height. *Agriculture, Ecosystems Environment*, 322:107676, 2021. ISSN 0167-8809. doi: https://doi.org/10.1016/j.agee.2021.107676. URL https://www.sciencedirect.com/science/article/pii/S0167880921003807.

[59] Gaia Vaglio Laurin, Jianqi Ding, Mathias Disney, Harm Bartholomeus, Martin Herold, Dario Papale, and Riccardo Valentini. Tree height in tropical forest as measured by different ground, proximal, and remote sensing instruments, and impacts on above ground biomass estimates. *International Journal of Applied Earth Observation and Geoinformation*, 82:101899, 2019. ISSN 1569-8432. doi: https://doi.org/10.1016/j.jag.2019.101899. URL https://www.sciencedirect.com/science/article/pii/S0303243419300844.

[60] *Sentinel-2 MSI User Guide*, . URL https://sentinels.copernicus.eu/web/sentinel/user-guides/sentinel-2-msi.

[61] Johannes Reiche, Adugna Mullissa, Bart Slagter, Yaqing Gou, Nandin-Erdene Tsendbazar, Christelle Odongo-Braun, Andreas Vollrath, Mikaela J Weisse, Fred Stolle, Amy Pickens, et al. Forest disturbance alerts for the congo basin using sentinel-1. *Environmental Research Letters*, 16(2):024005, 2021.

[62] Nika Oman Kadunc. How to normalize satellite images for deep learning. *Medium*, 2022. URL https://medium.com/sentinel-hub/how-to-normalize-satellite-images-for-deep-learning-d5b668c885af.

[63] Stephen Stehman. Sampling design for accuracy assessment of land cover. *International Journal of Remote Sensing - INT J REMOTE SENS*, 30:5243–5272, 09 2009. doi: 10.1080/01431160903131000.

[64] Vidushi Bhatia. U-net implementation from scratch using tensorflow. *Medium*, July 2021. URL https://medium.com/geekculture/u-net-implementation-from-scratch-using-tensorflow-b4342266e406.

[65] Balanced weights for imbalanced classification. *Medium*, 2022. URL https://medium.com/grabngoinfo/balanced-weights-for-imbalanced-classification-465f0e13c5ad#:~:text=Theweightsarecalculatedusing,theminorityclassis97.919.

[66] Oscar D. Pedrayes, Darío G. Lema, Daniel F. García, Rubén Usamentiaga, and Ángela Alonso. Evaluation of semantic segmentation methods for land use with spectral imaging using sentinel-2 and pnoa imagery. *Remote Sensing*, 13(12), 2021. ISSN 2072-4292. doi: 10.3390/rs13122292. URL https://www.mdpi.com/2072-4292/13/12/2292.

[67] Wekeo harmonized data access. URL https://help.wekeo.eu/en/collections/3530725-wekeo-harmonized-data-access.

[68] Aws sagemaker studio lab. URL https://studiolab.sagemaker.aws/.

[69] Matthew C Hansen, Alexander Krylov, Alexandra Tyukavina, Peter V Potapov, Svetlana Turubanova, Bryan Zutta, Suspense Ifo, Belinda Margono, Fred Stolle, and Rebecca Moore. Humid tropical forest disturbance alerts using landsat data. *Environmental Research Letters*, 11(3):034008, mar 2016. doi: 10.1088/1748-9326/11/3/034008. URL https://dx.doi.org/10.1088/1748-9326/11/3/034008.

[70] Susan Barati, Behzad Rayegani, Mehdi Saati, Alireza Sharifi, and Masoud Nasri. Comparison the accuracies of different spectral indices for estimation of vegetation cover fraction in sparse vegetated areas. *The Egyptian Journal of Remote Sensing and Space Science*, 14(1):49–56, 2011. ISSN 1110-9823. doi: https://doi.org/10.1016/j.ejrs.2011.06.001. URL https://www.sciencedirect.com/science/article/pii/S1110982311000147.

Colophon

This document was typeset using LaTeX, using the KOMA-Script class `scrbook`. The main font is Palatino.