Zander, Mats
Kieler Straße 100, 22769 Hamburg
m.zander@tu-braunschweig.de
Environmental Sciences
4991054

Tree Species Segmentation and Classification for Tree Health Assessment

Master Thesis
for the degree of
M.Sc. in Environmental Sciences

Technological University Brunswick

Department of Architecture, Civil Engineering and Environmental Sciences
Institute of Geoecology (IGÖ)

Reviewers:
Prof. Harald Biester
PhD. Matthias Beyer

Mentor:
Malkin Gerchow (PhD Student)

Declaration of Authorship

I hereby declare that the thesis submitted is my own unaided work. All direct or indirect sources used are acknowledged as references.

I am aware that the thesis in digital form can be examined for the use of unauthorized aid and in order to determine whether the thesis as a whole or parts incorporated in it may be deemed as plagiarism. For the comparison of my work with existing sources I agree that it shall be entered in a database where it shall also remain after examination, to enable comparison with future theses submitted.

This paper was not previously presented to another examination board and has not been published.

Mats Zander, 19.04.2022, Hamburg

Table of Contents

Table of Figures

Table of Tables

1. Introduction

Forests span nearly 30% of the world's land area and are present in nearly every climate zone. As biomes forests are the world's biggest host to biodiversity, regulate micro climates as well as impacting the world's climate as carbon storage (Huete 2012, Trumbore et al. 2015). Forests have always faced human-related destruction as they are a source of energy, food and building materials. In later history humans are imposing new indirect stressors in form of climate change, air pollution and invasive species on forests (Trumbore et al. 2015).

The introduction of these indirect stressors has impacts on forest diversity as species resilience to environmental factors differs. The loss of biodiversity is therefore inevitable if species are stressed beyond their resilience. If stress exceeds resilience levels of trees the disappearance of whole ecosystems as experienced in parts of Germany with the bark beetle infestation may follow (Zimmermann & Hoffmann 2020).

With the necessity to preserve forests not only for human consumption but also as biodiversity hotspots and for carbon storage timely and accurate information on tree health is important to understand and better protect forests (Huete 2012, Trumbore et al. 2015).

Environmental impact on tree stress levels can first be measured in the canopy. This is due to leaves being highly sensible and able to adapt faster than other tree parts to nutrient availability, water ability, radiation, temperature or pest presence (Trumbore et al. 2015). Gathering forest canopy information can therefore give insights into current tree health status and stress levels.

Remote sensing as discipline and research field aims at the assessment of ecosystems by measuring its radiation reflectance without physical contact (Huete 2012). It enables the forest health status evaluation by capturing the spectral reflectance of leaves and compiling Vegetation Indices (VIs).

Using VIs as a measure of plant wellbeing, performance or stress from aerial footage is common practice in remote sensing (Candiago et al. 2015, Fawcett et al. 2020). Accurate interpretation of VIs as health status proxies requires the context of observed species. Precision agriculture uses VIs to evaluate and optimize crop performance, with the distinct advantage of having same crop fields and therefore a reference within the plot for interpretation.

When using VIs to evaluate plant performance and stress in form of deficiencies or pests in heterogenous vegetation, interpretation becomes critical. As every species has different physiological strategies and VIs that would indicate severe plant stress or death are explicable and expected for other species. With that in mind, computer vision approaches are able to segment and classify different species in heterogenous vegetation and could improve the feasibility of VIs for species-based performance evaluation.

The connection of remotely sensed spectral information and computer vision enables the monitoring of heterogenous vegetation and could significantly increase the knowledge about the prevailing environmental conditions as well as specific plant strategies within studied ecosystems.

Continuous improvements in the quality of aerial imagery further increase the abilities remote sensing. This includes the improvements of Unmanned Aerial Vehicles (UAVs) to increase weight lifting capacities and flight time as well as improvements of multispectral sensors which are getting lighter in weight and easier to use. The simultaneous emergence of neural networks in form of Convolutional Neural Networks (CNNs) or Fully Convolutional Neural Networks (FCNs) as well as the increase in computing power improve the use of computer vision. The combined improvements strengthen the potential of remote sensing by increasing its use for environmental monitoring and conservation tasks.

Referenced training data is needed to enable computer vision approaches (models) to differentiate trees in the canopy. Models train on dedicated training data to learn features which allow for successful tree crown classification. Classification is the process of labeling segmented individual tree crowns with a dedicated class label and segmentation is the extraction of position and shape of tree crowns. More high quality training data often amounts in better tree crown predictions (Ghamisi et al. 2017, Hartling et al. 2021, Onishi & Ise 2021).

1.1. Study area

The experiment was conducted with data acquired at the Estacion Experimental Forestal Horizontes, which is part of the Area de Conservaci on Guanacaste, located in the northwest of Costa Rica (Figure 1). Precipitation in the study site (COSTA study site) shows pronounced seasonal differences and falls almost exclusively between May and November. It is further influenced by the El Niño Southern Oscillation resulting in heavy interannual variability with

annual sums ranging between 880 and 3030 mm. The site, previously agriculturally used land, has not been managed for more 30 years. This prolonged period of no human interference resulted in a now regrown tropical rainforest (Kühnhammer et al. 2022).

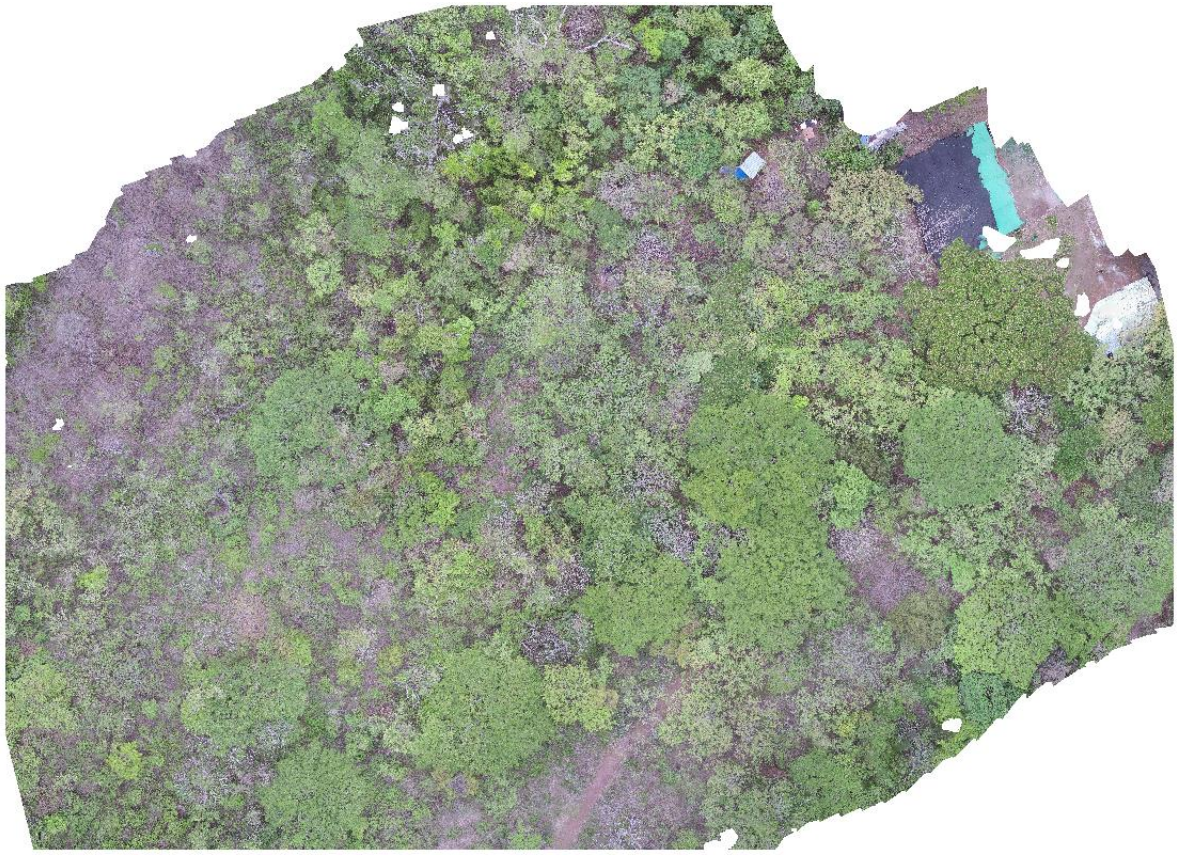


Figure 1: RGB-Image of the COSTA study site (Estacion Experimental Forestal Horizontes, Area de Conservaci on Guanacaste, Costa Rica), the image spans around 3 ha (196.7m west to east and 144.6m north to south).

Five species (classes) were chosen for this study Caoba (*Sweitenia macrophylla*), Guacimo *(Guazuma ulmifolia),* Guapinol (*Hymenaea courbaril*), RonRon (*Astronium graveolens*) and Tempisque (*Sideroxylon capiri*). As addition the class Other was established to account for segmented trees that do not belong to the species classes (Figure 2). Classes Caoba and Tempisque (Kühnhammer et al. 2022) represent species that do not shed leaves during the dry season.

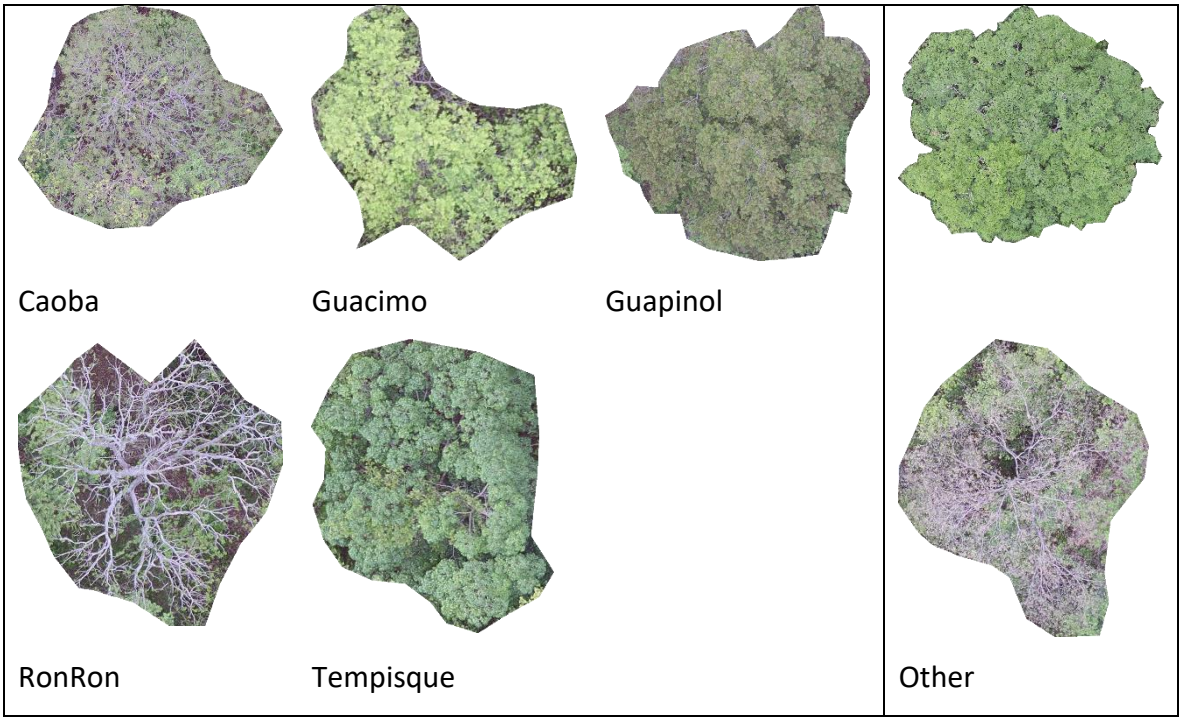


Figure 2: Detailed overview of the classes.

1.2. Research problem

The challenge with the COSTA dataset is the high tree species diversity on the comparatively small area. For comparison, the COSTA study site covers less than 3 ha while other studies used datasets covering 11 ha (Onishi & Ise 2021) and 51 ha (Schiefer et al. 2020) to name just two. Compared studies divided their datasets into 7 (Onishi & Ise 2021) and 14 (Schiefer et al. 2020) classes respectively. This is in contrast to the use of 6 classes in the COSTA experiment, including the species conglomeration in form of class Other. The ratio of dataset size to classes makes the COSTA dataset an interesting research object.

The comparatively small study site in combination with the highly diverse, dense and unmanaged forest impose a challenge for computer vision approaches and ultimately raise the following research questions:

> *"Are CNNs and FCNs able to correctly predict tree species at the COSTA study site?"*
>
> *"Are classification predictions via CNNs or semantic segmentation predictions via FCNs sufficient information to asses COSTA dataset tree species health status with VIs?"*

Given this, we can formalize our Hypothesis as follows: If VIs formed with neural network model predictions display the same VI values as the COSTA reference data, then model predictions are sufficient information to predict COSTA dataset tree health status.

The first objective of this study is to test and compare different approaches to classification and semantic segmentation in their usability to predict tree species from UAV derived COSTA canopy images. Classification experiments are conducted by using CNNs that are fine-tuning on differently constructed datasets. Trained classification models are tested and compared to SVMs. The semantic segmentation experiment on the other hand uses FCNs with U-Net architecture and different encoders the FCNs are fine-tuned and tested on the COSTA semantic segmentation dataset. Best performing models are used to create species predictions for the COSTA datasets. The second objective is the VI creation from reference and prediction data and their comparison. The comparison is the evaluation basis for the capability assessment of tested models as a tool to predict species specific health status.

1.3. Document Outline

The document is structured as follows; the chapter Literature Review provides a thorough review of research conducted in remote sensing relevant to this work. The review includes an introduction to tree crown segmentation, neural networks as well as providing a brief overview on chosen VIs used in this study. Differences in studied neural network applications used on literature for tree classification (CNNs) and forest mapping (FCNs) are discussed.

Chapter three discusses the chosen methods to answer the questions introduced in the introduction. Means of dataset creation are explained including information used in the datasets as well as applied data augmentation. The chapter further explains the process of CNN and FCN training as well as model performance evaluation. Chapter Methods finishes with a brief summary of the VI creation process from spatial reference and prediction data.

The chapter Results presents the COSTA datasets as a result of the tree crown segmentation. Model performance results of the classification training and prediction are shown. Semantic segmentation training and prediction results are displayed and explained. Lastly VIs constructed from combined COSTA datasets are presented and compared.

Model training as well as prediction results are given context and their informative value is explained in the Discussion chapter. Further the VI values from reference and prediction data

are discussed. Explanations for observed differences within classes as well as between classes are given. The suitability of model predictions on the COSTA dataset for tree health status analysis is discussed. Chapter five finishes with limitations and possible adjustments to improve future results.

The work concludes with the chapter Conclusion which answers the questions posed in the introduction.

2. Literature review

Aim of the literature review is to give an overview on the scientific state of the art in tree segmentation and classification approaches as well as a brief overview on different VIs that can detect plant wellbeing.

2.1. Individual Tree Crown Segmentation

Individual tree detection requires information on position and shape of trees. Unmanned Aerial Vehicles (UAVs) capture canopy imagery that is compiled to orthomosaics. Orthomosaics are mosaics of images stitched together, overlap of images allows for the creation of different height models. The vertical position of each reflection captured by the camera is composed to the Digital Surface Model (DSM), the underlying ground is referred to as Digital Terrain Model (DTM). The subtraction result is called Canopy Height Model (CHM) and stores height information of the canopy or structures on the ground (Candiago et al. 2015, Chenari et al. 2017, Schiefer et al. 2020).

$$CHM = DSM - DTM$$

Different procedures have been proposed to acquire position and shape of individual tree crowns from the orthomosaic (Natesan et al. 2019, Onishi & Ise 2021, Schiefer et al. 2020).

Watershed segmentation is inspired by geographical watershed pouring, in which ridges represent high regions and valleys low regions as in topographic landscapes (Hartling et al. 2021). This approach uses the height information stored in the CHM. A Gaussian filter is used to blur the CHM prior to finding local maxima as tree tops to avoid false positives (Hartling et al. 2021, Natesan et al. 2019). The CHM is then flooded from the tree top markers resulting in shapes corresponding with the tree outline. Natesan et al. (2019) pointed out that this procedure may produce incorrect tree crown outlines in areas with high tree density.

Another segmentation procedure is Multiresolution segmentation, which several studies used as part of the eCognition Developer software (Chenari et al. 2017, Onishi & Ise 2021). This approach uses primarily the visible (Red-Green-Blue (RGB)) spectral information. Additional information, such as a slope model, was used to improve segmentation results (Onishi & Ise 2021). The slope model contains the maximum elevation change between neighboring pixels and is derived from the DSM. Parameter values for the Multiresolution segmentation were

adjusted via trial and error and resulting masks were checked for fragmentation or merger artifacts (Onishi & Ise 2021). Chenari et al. (2017) successfully tested the same procedure for object segmentation in sparsely populated wild pistachio and wild almond woodlands (Iran).

Manual segmentation, segmentation by hand, of individual tree crowns was used by Schiefer et al. (2020) to create tree crown reference data.

2.2. Support Vector Machines

Species classification of individual tree crowns is a broad discipline in remote sensing (Hartling et al. 2021, Li et al. 2021, Mäyrä et al. 2021, Natesan et al. 2019, Onishi & Ise 2021, Sothe et al. 2020, Sothe et al. 2019). Support Vector Machines (SVMs) are a prominent machine learning approach that has been used in recent history to classify individual tree crowns (Hartling et al. 2021, Onishi & Ise 2021, Sothe et al. 2019).

The SVM defines an optimal separating class boundary, called hyperplane, in a multidimensional feature space that differentiates classes (Ghamisi et al. 2017). To achieve this only samples in the training data are used which are close to the class boundary, these samples are referred to as support vectors (Ghamisi et al. 2017). This concept however was initially designed to distinguish in binary classification problems. To account for this, two strategies can be used one-against-one as well as one-against-rest (Ghamisi et al. 2017). SVMs have the upside of being easy to use and a swift training stage.

Studies using SVMs proved their capability of multi class individual tree crown classification (Hartling et al. 2021, Onishi & Ise 2021, Sothe et al. 2019). They however revealed some accuracy deficiencies if dealing with small datasets (Hartling et al. 2021, Sothe et al. 2019).

2.3. Convolutional Neural Networks

Another option is the use of neural networks for image classification. Neural networks are commonly used in the form of Convolutional Neural Networks (CNNs) for individual tree crown classification (Fricker et al. 2019, Li et al. 2021, Mäyrä et al. 2021, Natesan et al. 2019, Onishi & Ise 2021, Sothe et al. 2020).

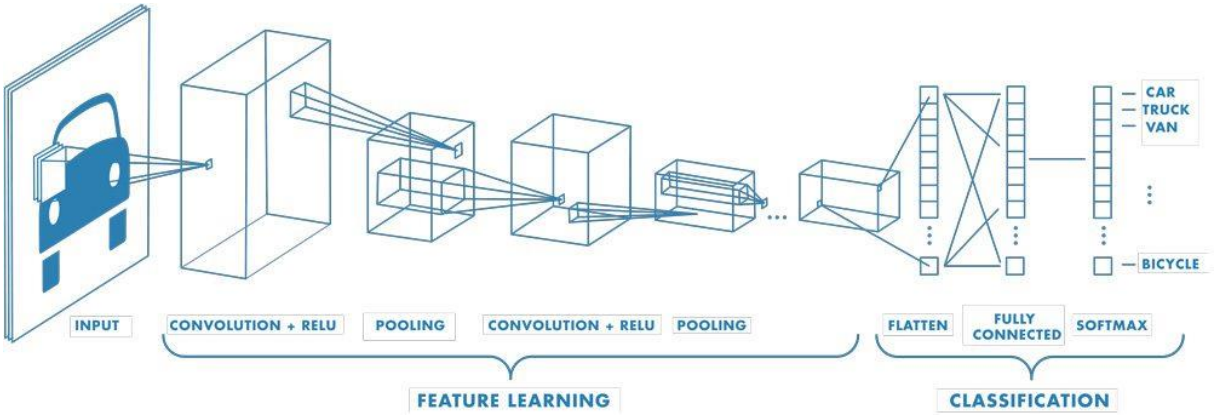


Figure 3: Illustration of a CNN. (The MathWorks, Inc. 2021)

CNNs, are deep learning networks that are inspired by neuro science. Deep learning refers to the network having three or more layers. Each layer consists of a convolution layer, nonlinear operation and a pooling layer (Figure 3). The convolution produces a feature map on which the nonlinear operation is used, the result is passed on to the pooling layer that creates a new feature map. The resulting feature map of this stage is of lesser resolution and gets passed on to the next layer, who does a similar process further convoluting the image. This chaining of convolutions and reprocessing is able to extract invariant and robust features (Ghamisi et al. 2017). The condensed features then allow for the classification of the image. CNN training consists of several complete considerations of the training data, each loop over the training data is called epoch. During this training weights in the nonlinear operation are trained and modified to minimize an error function. The training time required for CNNs is considerably higher compared to SVMs which is caused by the repeated consideration of the training dataset. To avoid the observed overfitting of CNNs regularization methods are required (Ghamisi et al. 2017). Overfitting describes models achieving good results on training datasets but with less accuracy on the validation and testing datasets.

The most commonly used CNN for individual tree crown classification is Residual Networks (ResNet) (Fricker et al. 2019, Onishi & Ise 2021, Volpi & Tuia 2017). This Network was the winning entry to the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) in 2015. The ImageNet dataset is a research effort to provide large amounts of natural images to researchers. ImageNet consists of more than 15 million images in more than 22,000 classes. The challenge winning ResNet was 152 layers deep (He et al. 2015).

The use of deep networks is especially useful for large datasets (He et al. 2015). To increase accuracies in other scientific applications, such as individual tree crown classification, a

method called fine-tuning is applied (Rusu et al. 2016, Yosinski et al. 2014). Features and information in the first network layers are not task specific. The information acquired on the ImageNet dataset can therefore be used for classification of different target images (Yosinski et al. 2014). This is widely applied in remote sensing to reduce the negative impacts of smaller datasets (Onishi & Ise 2021, Schiefer et al. 2020).

Several studies have been conducted using CNNs as well as SVMs to classify individual tree crowns. Onishi and Ise (2021) used four different neural networks with ImageNet weights, namely AlexNet, VGG16, ResNet18, and ResNet152, all of which were entries to ILSVRCs. The aim of their research was to show the ability of CNNs to classify tree crowns in leaf as well as fall seasons in a heterogenous forest and to compare the classification results to SVMs accuracies on the same dataset. The networks were fine-tuned on the dataset created for this study. To increase the amount of training data, each tree crown was augmented eight times by horizontal flipping and rotating 90 degrees. Tree crown images were uniformed in size as CNNs require even image input sizes. To account for this, tree crown images were center cropped to an edge size of 224 pixels after resizing the images to 256x256 pixels (Onishi & Ise 2021). The dataset was divided into 50% training data, 25% for validation and the remaining 25% is preserved for testing. While the ratios remained the same, data is selected at random for model training. Onishi and Ise (2021) succeeded in showing the superiority of deep neural networks in form of CNNs if compared to SVMs in individual tree crown classification tasks. They observed that the fine-tuned ResNet152 is outperforming SVMs in leaf season as well as fall season. Accuracies achieved by CNNs were 93.3% for leaf season and 97.6% in fall season, the SVM achieved 80.3% and 91.8% respectively.

A similar study was conducted by Natesan et al. (2019). Their study site was dominated by pine trees, and was imaged three times in three years, two times during leaf season and once in fall season. The tree crown images were augmented, including flipping and rotating, then rescaled to an edge size of 224 pixels and used in fine-tuning the ResNet50. The achieved overall accuracy did not surpass 80%.

2.4. Fully Convolutional Neural Networks

Another use of CNNs is studied by Schiefer et al. (2020). Their study was conducted to show the applicability of a less labor-intensive way of mapping tree species in a heterogenous forest.

The tested approach is called semantic segmentation. The difference to classification is that the network is classifying each pixel of the image instead of assigning a class to the whole image (Figure 4). Networks capable of pixelwise classification are referred to as Fully Convolutional Neural Networks (FCNs) (Lobo Torres et al. 2020, Schiefer et al. 2020). The advantage of this end-to-end approach is the eradication of prior, costly tree crown segmentation in the application of the network after completed training (Lobo Torres et al. 2020, Schiefer et al. 2020, Volpi & Tuia 2017). Schiefer et al. (2020) further reasoned prior tree crown segmentation or feature engineering are limiting study transferability and increase computational load.

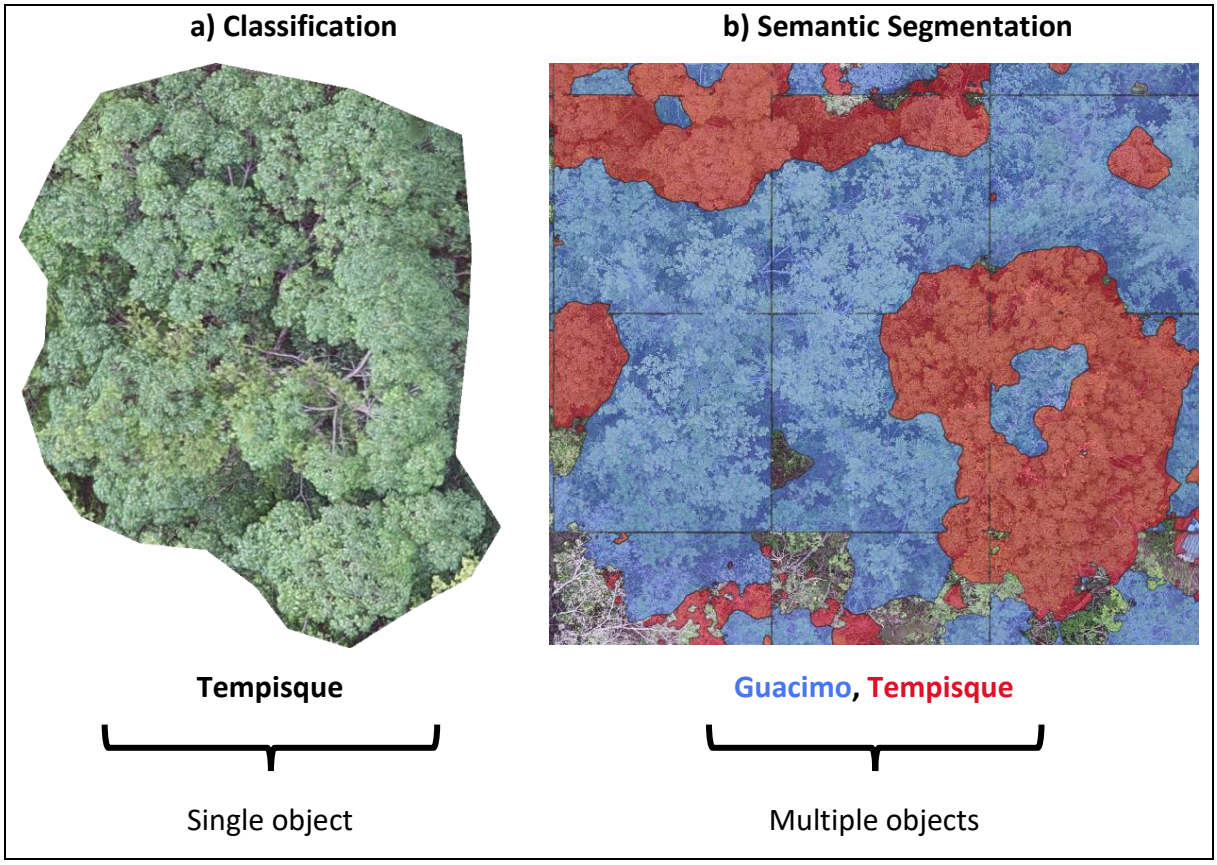


Figure 4: Different classification approaches tested in this study: classification (a) and semantic segmentation (b)

The study used the U-Net architecture (Figure 5). U-Net consists of an encoder and a decoder. FCNs follow the convoluting procedures of CNNs, and are referred to as encoder (Ronneberger et al. 2015). The decoder then symmetrically expands the encoding results to the original image resolution in order to map the classification (Lobo Torres et al. 2020, Schiefer et al. 2020). U-Net further possess skip connectors, which allow for concatenation of convolution and corresponding deconvolution of the decoder. This further improves the class mapping on the input image spatial resolution (Lobo Torres et al. 2020).

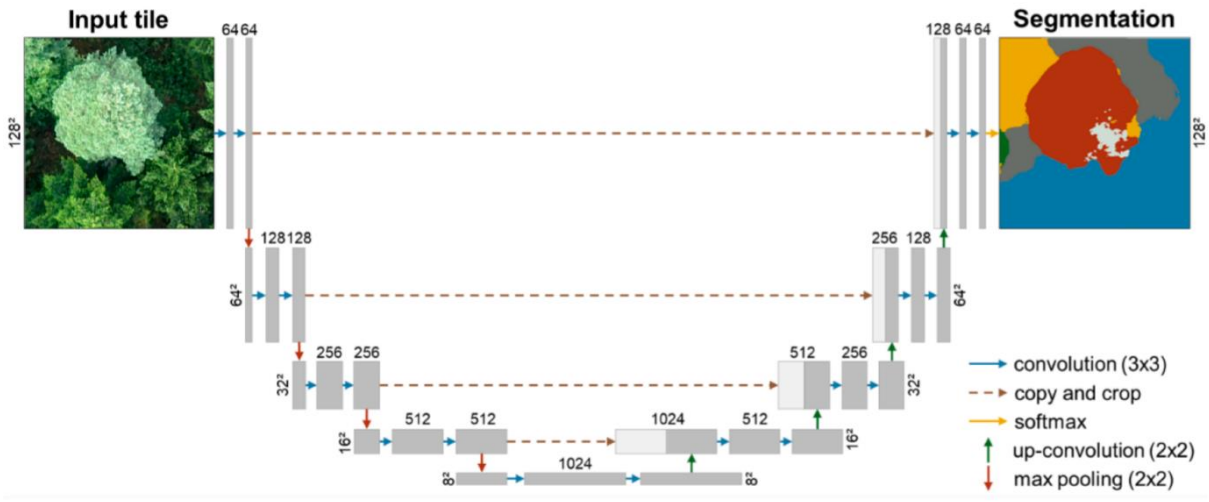


Figure 5: U-net FCN-Architecture for tree crown segmentation. Analyzation scheme of tiles with pixel size 128x128 (Schiefer et al. 2020).

The FCNs analyzed by Schiefer et al. (2020) was classifying 14 different tree species in two forest sites in Germany. The study site images were divided into tiles of equal size (128x128, 256x256 and 512x512 pixels respectively). 10% of the dataset was used as test data, 75% of the remaining dataset was used for training while the rest (18%) was used as validation dataset. Validation was done for each epoch to prevent overfitting. The study of Schiefer et al. (2020) achieved Overall Accuracies (OAs) of 89% with the lowest tile size and a Pseudo-RGB dataset. The Pseudo-RGB consisted of RGB tree crown images with CHM height information as additional pixel information. Their findings showed differences in OA of 3% between the six model setups (3 tile sizes x 2 datasets). An additional observation was the negative impact of increasing tile sizes for the detection rate of underrepresented species (Schiefer et al. 2020).

2.5. Vegetation Indices

Detecting plant health status with the use of Vegetation Indices (VIs) is an important research area in remote sensing (Candiago et al. 2015, Fawcett et al. 2020, Saura et al. 2019). Vegetation indices measure canopy greenness, a composite property of canopy structure, leaf area, and canopy chlorophyll content and therefore can predict plant health status (Huete et al. 2006). The use of UAVs improved precision and accessibility of means to detect water deficiencies, pathogenic pressure or other nutrient deficiencies (Candiago et al. 2015, Saura et al. 2019).

The creation of VIs however requires the use of multispectral cameras, as additional wave lengths to the visible spectral bands red, green and blue are needed (Huete 2012, Saura et al.

2019). Spectral VIs reveal information about leaf and canopy structure as well as biogeochemical composition (Figure 6). This information is a result of spectral reflectance signatures by different parts of the leaves. Water absorbs shortwave-infrared (SWIR) in wavelengths between 1300 and 2100 nm whereas leaf pigments absorb most of the visible spectral wavelengths 400 to 700 nm another relevant spectral region is the near infrared region (NIR, 700-1300) as reflectance in this region varies with leaf structure and amount. VIs are a normalization of 2 or more wavelengths with one reflective spectral bandwidth and one absorbed wavelengths (Huete 2012).

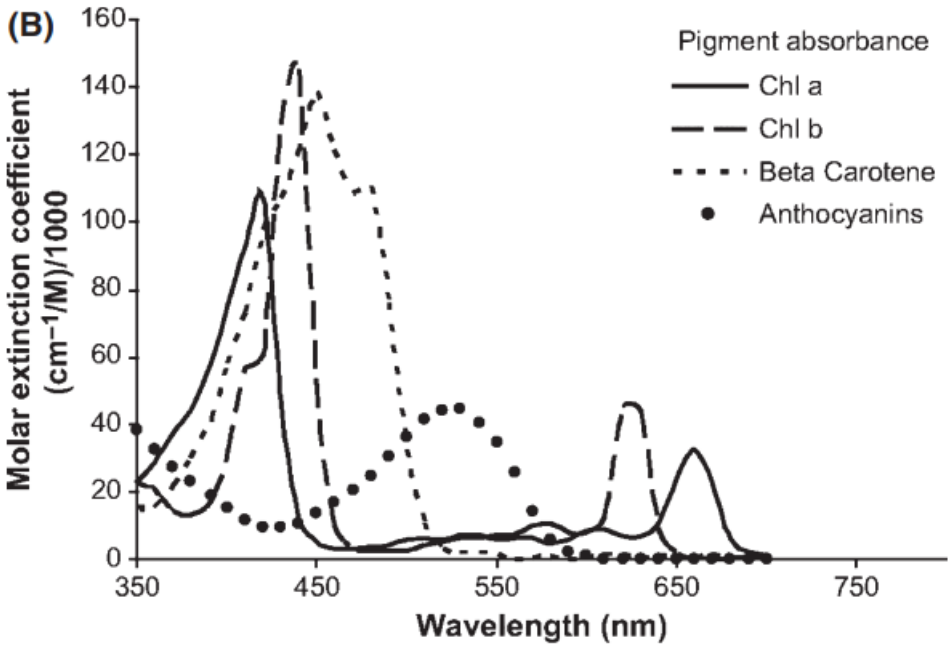


Figure 6: Biochemical absorption spectra of common leaf pigments (Huete 2012)

A fundamental VI is the Normalized Difference Vegetation index (NDVI), this index is constructed with the emissions of the canopy-absorbing red band (600-700 nm) and the NIR (Huete 2012, Tucker 1979, Xue & Su 2017).

$$NDVI = \frac{\rho_{NIR} - \rho_{RED}}{\rho_{NIR} + \rho_{RED}}$$

The NDVI is an Index which gives information on vegetation coverage, which ranges between -1 and 1 whereas higher values represent healthy plant coverage. It is widely used in remote sensing to detect areas with high vegetation (Huete 2012, Saura et al. 2019, Xue & Su 2017). A limitation of NDVI is its saturation for Leaf Area Index (LAI) values of 2-3 and above (Huete 2012, Xie et al. 2018). The LAI is a ratio of leaf area to per unit ground surface area, hence

dimensionless. The Index can be related to photosynthesis, evaporation and transpiration as well as carbon flux. It is therefore a strong index to go from for understanding the processes in the studied ecosystem (Zheng & Moskal 2009). This is explained by the high absorbance of the red band by upper most foliage layers, no information on lower leaves can be gained from using of the visual red spectral band (Huete 2012).

Other spectral products are therefore necessary to determine and differentiate plant performance and limitations in highly performing ecosystems e.g., dry forests in the wet season. Such a product is the Normalized Difference Red Edge Index (NDRE), it is constructed similar to the NDVI by replacing the red the Red-Edge spectral band (715 nm) (Xie et al. 2018). The Red-Edge reflection is less prone to saturation and is able to gather reflective information of deeper canopy layers thereby revealing additional information on lower hanging foliage.

$$NDRE = \frac{\rho_{NIR} - \rho_{REDEGDE}}{\rho_{NIR} + \rho_{REDEDGE}}$$

Another VI that maintains sensitivity even in high LAI canopies by relying on near-infrared canopy reflectance is the Enhanced Vegetation Index (EVI). The EVI is a proxy for canopy photosynthetic capacity and gross primary production (GPP) (Huete et al. 2006). The VI was developed to minimize the effects of varying background soil reflectance and atmospheric influences in measuring vegetation status (Tanaka et al. 2015). EVI has shown to correlate with GPP for North American biome types (Huete et al. 2006, Hunt et al. 2013).

$$EVI = G * \frac{\rho_{NIR} - \rho_{RED}}{\rho_{NIR} + C1 * \rho_{RED} - C2 * \rho_{BLUE} + L}$$

The Red-Edge Chlorophyll Index (CI) is used to estimate leaf chlorophyll content. CI values are sensitive to small variations in the chlorophyll content and consistent across most species (Hunt et al. 2013). Leaf chlorophyll levels are correlated to leaf nitrogen levels (Hunt et al. 2013). It is used as an indicator for leaf nutrient levels as well as water stress and photosynthetic capacity.

$$CI_{Red-edge} = \frac{\rho_{NIR}}{\rho_{REDEDGE}} - 1$$

Huete (2012) states that even though most VIs are highly correlated, the use of multiple gives a much better overview on the ecosystem performance.

2.6. Summary, Limitations and Gaps of Literature

A detailed literature review on the application of computer vision approaches in tree species segmentation and classification as well as the use of vegetation indices to determine plant health status was made. Literature exploring various aspects of individual tree crown classification, its applications and the use of VIs, was studied as part of solving the problem of predicting species specific tree health status in the COSTA dataset. The review pointed out different options for reference dataset creation in form of tree crown segmentation. It continued by showing the challenges and opportunities of tree crown classification via the use of SVMs and CNNs as well as the option of semantic segmentation with FCNs. Lastly it explained the creation and use of VIs as health status proxies.

To the best knowledge of the author, there are no research papers which use predicted tree species for the estimation and evaluation of species-specific heath status.

Limitations and research gaps identified by the literature review are addressed by the research questions posed in the Introduction chapter:

> *"Are CNNs and FCNNs able to correctly predict tree species at the COSTA study site?"*
>
> *"Are classification predictions via CNNs or semantic segmentation predictions via FCNs sufficient information to asses tree species health status with vegetation indices?"*

3. Methods

This chapter explains gives an overview of the methods chosen to answer the questions posed in the introduction.

All programming was done using Python 3.9 and the machine learning packages PyTorch (Paszke et al. 2019) and Scikit-learn (Buitinck et al. 2013).

Model training and testing used the GPU hardware acceleration of *Google Colaboratory*[1].

3.1. Data acquisition

The COSTA dataset was acquired using the *DJI Matrice 210* (UAV) carrying the *DJI Zenmuse XT2* for RGB imagery as well as the *RedEdge MX Dual Camera Imaging System* consisting of *RedEdge MX* and *RedEdge MXblue* by *MicaSense.* The system captures 10 different spectral bands which are necessary for the VI creation. Several imagery acquisition campaigns were conducted at the study site. This study used data acquired on April 30th 2021. As well as RGB imagery acquired on May 1st and May 3rd 2021. During flights, the UAV maintained an above ground height of 50 meters with flight speed of 2 m/s and an image overlap of 90%. Agisoft (version 1.7.4) was used to create georeferenced orthomosaics. This procedure resulted in very high image resolutions, specifically a resolution 1 cm for RGB data and 2.5 cm for multispectral data.

3.2. Datasets

The reference tree crown data is manually segmented in *QGIS* (version 3.20.1 "Odense"). The segmented tree crowns were assigned to the six classes RonRon, Tempispue, Caoba, Guacimo, Guapinol and Other. This was achieved by comparing tree crowns of trees with confirmed class assignment. The tree crown class assignment results were cross-validated by an expert.

To create the classification dataset, extraction of the tree crowns into single images was necessary. The polygon layer with the manually created tree crown shapes was split into single polygons using the native *QGIS*-function "*split polygon layer*". Single crown shapes were then clipped to the raster resulting in the desired individual tree crown image.

To test and evaluate the impact of dataset augmentation three different datasets were constructed from the RGB imagery, which are referred to as standard, augmented and 3-flight

[1] Free to use as of 01.04.2022, (https://colab.research.google.com/notebooks/intro.ipynb)

dataset. Additionally, a dataset with 5-band multispectral images of the tree crowns was used. All tree crown images in the classification datasets were structured into folders for each class. The folders are then randomly split into training, validation and test data using the *split-folders* package (Ver. 0.4.3) in Python. 20 percent of the data was used as test data, while 20 percent of the remaining data were used as validation data and the remaining images were used as training data. All RGB datasets used the same test data, image augmentation or dataset expansion was only conducted on training and validation data.

The standard dataset consisted of unmanipulated tree crown images. The 3-flight dataset used unmanipulated tree crown images from three flight dates. Therefore, three different images of the same tree crown were used in this dataset. The augmented dataset used standard dataset images which were manipulated by turning the images three times 90 degrees and flipping them horizontally. This resulted in the augmented dataset with eight times the size of the standard dataset.

As part of the classification experiment a comparison with networks training on a multispectral dataset was conducted. For this dataset the spectral bands blue (475 nm), green (560 nm), red (668 nm), red edge (717 nm), and NIR (842 nm) were combined to create a multispectral image. The dataset structure and distribution of training, validation and test data was done same as for the RGB datasets.

The segmentation dataset used six classes. In contrast to the classification dataset a Background class was used to describe all pixels that do not represent the five species classes, class Other was not part of this dataset. The dataset was structured in two parts. Part one was the image data while the second part was the label. The label is an array with class values for each pixel of the image data. The label file was created by extracting the location of every class from the reference data into layers containing one class and adding up the layers to form the label layer containing one value for each class. The single file RGB image map was split due to computing restrictions and model performance. The split was done by creating tiles from the image map with pixel size of 500 x 500. Resulting tiles which did not span the required pixel size were expanded with black to the required tile size, this was done to account for tiles at the map edge. Label tiles were created with the same procedure from the label map, and if necessary, expanded with the background value. Training, validation and test data were divided using the function *test_train_split* (Buitinck et al. 2013). As image and target data need

to be concatenated the "*random_state*" parameter was set identical. 10 percent of the dataset was used as test data while 15 percent of the remaining data was used for validation and the resulting rest used as training data.

The shape of the COSTA study site resulted in image tiles which displayed more than 50 percent white pixels. These more half empty image tiles were excluded from the COSTA semantic segmentation dataset, corresponding target tiles were removed accordingly.

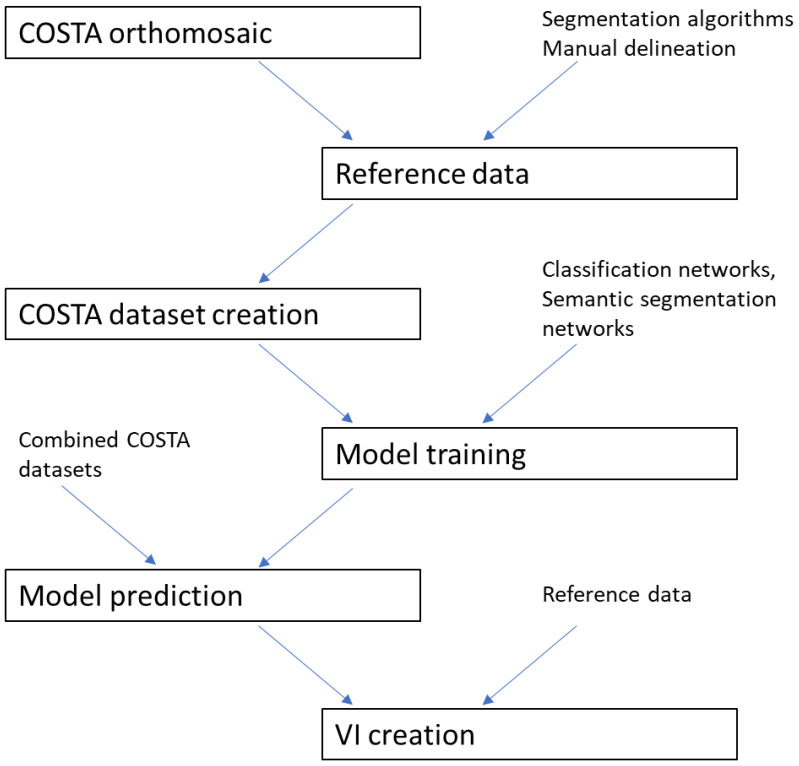


Figure 7: Workflow for tree health assessment using neural networks at the COSTA study site.

3.3. Classification

The classification experiments were conducted using CNNs with ResNet architecture and SVMs as comparison. Different layer depths were used in combination with different datasets to evaluate the impact of different model training setups on model performance. Tested layer depths in this experiment were 34, 50 and 101 referred to as ResNet34, ResNet50 and ResNet101. All considered networks were downloaded with pretrained ImageNet weights and fine-tuned on COSTA datasets created in this study.

During model training data was loaded, resized, augmented and normalized. All individual tree crowns were resized to edge length of 224 pixel. Image data from standard, 3-Flight and 5-

band multispectral datasets was augmented by randomly flipping and rotating during model training. As an additional measure to increase model performance and training speed all datasets were normalized. Normalization was achieved by providing dataset mean and standard deviation values of the spectral bands to the network.

The training consisted of fitting the model to the training data and validation on a distinct validation dataset using the cross-entropy loss function of *PyTorch* (Paszke et al. 2019). Models were training on RGB datasets for 60 epochs while 5-band multispectral training was conducted for 150 epochs. The network weights of the epoch in which the lowest error score on the validation data by the loss function was achieved were saved and used as best model configuration.

Training on the 5-band MS dataset required modifications to the networks. The ImageNet dataset consists of RGB imagery, which consist of three channels. To enable fine-tuning on five channel data two additional channels were added. Channel weights of the red channel were copied into the new channels. Network training remained the same procedure previously explained for RGB data.

To evaluate the general suitability of CNNs to accurately classify the individual tree crowns of COSTA classification test data, results were compared to SVM test data predictions. SVM training used CNN training data predicts corresponding test data. This procedure ignored the validation data used in the CNN training process. Best SVM prediction results on COSTA classification test data were achieved using a *C* value of “1000”, “*rbf”* as kernel option and *gamma* value “1”.

The best fitting CNN was used to create the prediction. Validation and test data were combined for the prediction. This dataset is referred to as combined COSTA classification dataset. Prediction results of the classification were matched to individual tree crowns via the tree ID for further tree health assessment (Figure 7).

3.4. Semantic segmentation

To test the performance of semantic segmentation via FCNs the COSTA semantic segmentation dataset was used. The dataset consisted of image and target tiles with size of 500 x 500 pixel. Prior to model selection and training tiles containing solely background and their corresponding image tiles were excluded from model training, validation and testing.

This was done to increase model training speed and improve general model performance. Excluded tiles were however later used in model predictions of the COSTA study site.

The semantic segmentation experiment used the FCN *U-Net* architecture with different pretrained encoders. Six different encoders were tested and evaluated, namely resnet50, resnet101, se_resnet50, se_resnet101, densenet121 and densenet169. All encoders were pretrained, using ImageNet weights as starting weights.

Unbalanced class distribution was accounted for by changing loss function weights to represent class distribution in the dataset. Cross entropy loss was chosen as the loss function for model training and validation. All models trained for 40 epochs and the network weights of the epoch with the lowest error score on the validation data was saved for further performance evaluation.

The network with the highest performance on the semantic segmentation test data was used for tree species prediction at the COSTA site. The combined COSTA semantic segmentation dataset consisting of all image tiles was used for the prediction. This included training, test, validation data as well as tiles depicting background only. The predicted tiles were aligned to create a prediction map with the size of the original input image map. This enabled georeferencing of the semantic segmentation prediction results for future VI analysis.

3.5. Performance Evaluation

The evaluation of models requires performance metrics for comparison. The afore mentioned error score produced by the loss function is one of these performance scores. However, to compare between networks other metrics were used. This study used test accuracy referred to as OA and the F1-Score which was constructed to compare model performance between individual classes. These scores were calculated for each class in the COSTA dataset by looking at the number of correct classifications against misclassifications.

OA is the number of correct classifications divided by the number of all classifications. True positives (TP) and true negatives (TN) are the correct classifications while false positive (FP) and false negative (FN) denote misclassifications. The OA can be misleading in cases of highly imbalanced dataset such as the COSTA dataset.

$$OA = \frac{TP + TN}{TP + FP + TN + FN}$$

Therefore, performance metrics are included that give insight in the class wise performance. Such a metric is the F1-Score is the weighted mean of precision and recall. The Score ranges from 0 to 1, 1 being the best score possible. The F1-Score is better suited to evaluate models with unevenly distributed classes.

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$F1 = \frac{2 * Precision * Recall}{Precision + Recall}$$

Semantic segmentation uses an additional metric to validate model performance, mean Intersection over Union (mIOU). This metric is used to evaluate the degree of similarity between the reference data and the model prediction. If the shape of the prediction matches exactly the reference data the IOU is 1. The mIOU is used in the evaluation of multiclass semantic segmentation as this value reflects the mean value of IOUs achieved for all classes.

$$IOU = \frac{Reference \cap Prediction}{Referenc \cup Prediction}$$

3.6. Vegetation status

This study compares four VIs, which were chosen as they enable a comprehensive overview on the observed ecosystem status. Chosen VIs were NDVI, NDRE, EVI and CI. VI calculation was implemented as a custom Python script. This script used the multispectral information captured on April 30th 2021 of the study site to create georeferenced VI maps.

Two different approaches were pursued. The classification prediction enabled the creation of VI mean values for each individual tree crown, whereas the semantic segmentation prediction was unsuitable for individual tree crown means. Hence, VI means values for each species were created for semantic segmentation reference and prediction data.

VI values for classification reference tree shapes were created by using the *Zonal statics* tool in *QGIS*. The values were matched to tree IDs which allowed for the use of the same data as VI prediction mean values. In order to create species specific VIs with semantic segmentation

maps the *QGIS* tool *Raster layer zonal statistics* was used. This was done for the reference map as well as semantic segmentation prediction map.

The VI means were qualitatively compared to further evaluate model success in correctly predicting species in the COSTA dataset as well as the feasibility of model predictions as a mean of species health status assessment.

4. Results

This chapter presents the results of the research experiments. It describes the findings of the data preparation as well as giving an overview on the datasets. It further explains the model training results and model performance on unseen data. Lastly results of the use of whole dataset model predictions for vegetation status detection are presented. The chapter is divided into the following subsections.

4.1. Tree crown segmentation and datasets

The segmentation of tree crowns was the most important step in the COSTA dataset creation. For this, two algorithms were tested and compared to the manual tree crown segmentation. The first algorithm tested was *watershed segmentation* (Figure 8). Secondly the *multiresolution segmentation algorithm* by *eCognition* was tested but its results were not further used as exporting from required a license which was not available.

a) Reference data b) Watershed segmentation

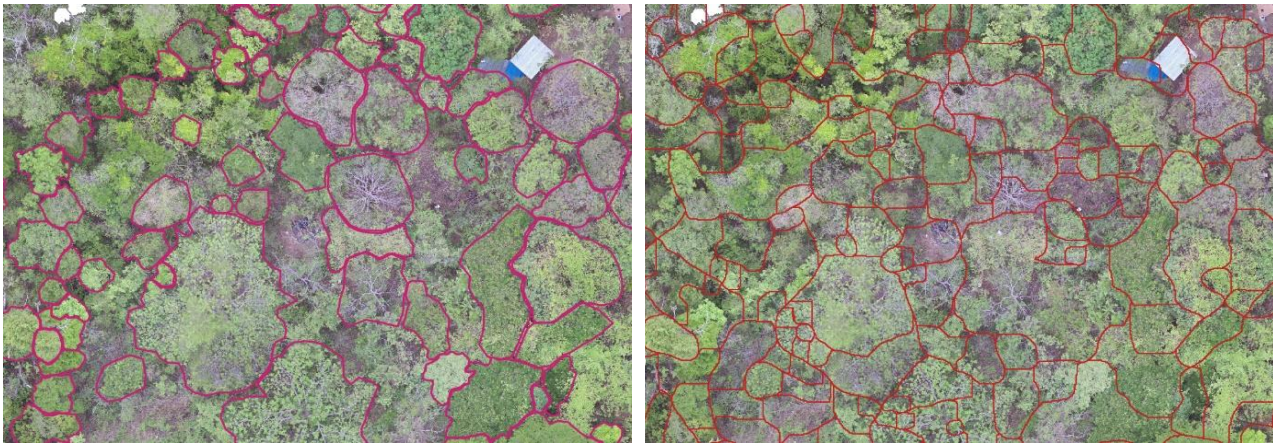


Figure 8: Manual delineation (left) and watershed segmented (right) tree crowns.

Trial-and-error testing of parameters for the watershed segmentation did not yield the required results to justify its use as a delineation base for the creation of individual tree crowns

images for the COSTA dataset. The manual delineated reference data was therefore used to create the datasets.

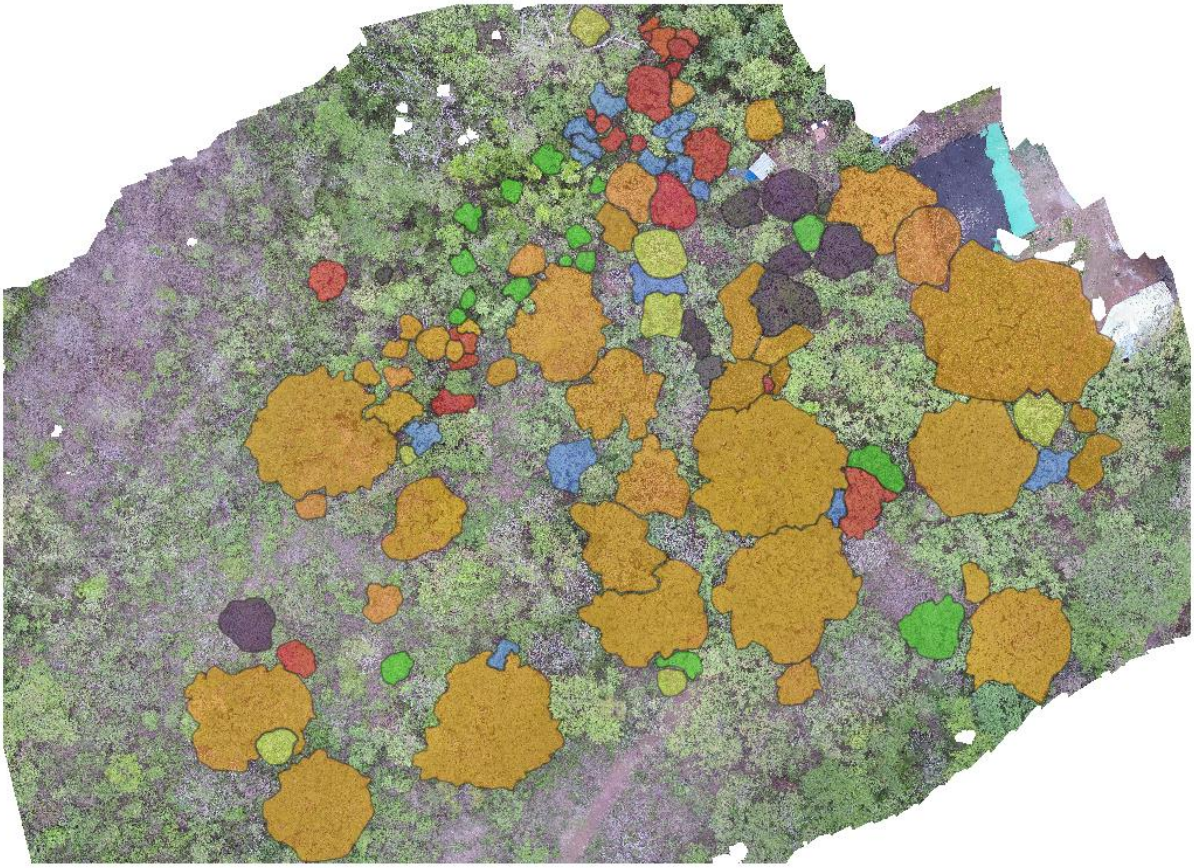


Figure 9: COSTA Reference dataset. The individual tree crowns of classes Caoba (Black), Guacimo (Blue), Guapinol (Green), Tempisque (Red), RonRon (Yellow) and Other (Orange) are shown. Class Other is not used for semantic segmentation.

Classification dataset

The delineation resulted in 110 individual tree crowns (Figure 9) of which 65 are used for model training, 20 for validation and 25 for testing (Table 1). The classes Caoba and RonRon were underrepresented in the COSTA dataset (Table 1). This led to only one RonRon individual in the validation data und two in the test data. The class Other (41) displayed the highest amount of individual tree crown.

Table 1: Classification dataset distribution (individual tree crowns)

Classes	Total occurences	Training	Validation	Test
Caoba	10	6	2	2
Guacimo	15	9	3	3
Guapinol	16	9	3	4
Other	41	25	7	9
RonRon	7	4	1	2
Tempisque	21	12	4	5
Total	**110**	**65**	**20**	**25**

Semantic segmentation dataset

The dataset for the semantic segmentation consisted of 923 tiles with edge size of 500 pixels. Tiles that showed none of the five classes of interest are excluded from the training, validation and testing process. This procedure resulted in 144 tiles used for training, 26 used during the validation and further 19 used to evaluate the model performance during testing. The semantic segmentation dataset showed great differences in area-related shares between the classes (Table 2). The shares were constructed without the use of background only tiles. All 923 tiles were used to create the COSTA combined dataset on which the model prediction of the whole study site was executed.

Table 2: Area-related share distribution of classes in the semantic segmentation dataset.

Classes	Area-related share (Total)	Area-related share (Training)	Area-related share (Validation)	Area-related share (Test)
Background	67.83%	67.14%	72.76%	66.31%
Guapinol	5.01%	4.96%	6.12%	3.88%
Guacimo	6.36%	6.80%	5.00%	4.86%
Caoba	10.11%	10.34%	4.37%	16.21%
RonRon	3.72%	3.66%	4.26%	3.47%
Tempisque	6.97%	7.10%	7.50%	5.28%

The largest species class in the semantic segmentation dataset was the Caoba class (Table 2) this in contrast to the low individual count observed in the in the classification dataset (Table 1). Area-related shares of classes in the data varied due to the nature of random selecting tiles for testing, validation and training. This led to the class Caoba having lower area-related shares than the classes Guapinol, Guacimo and Tempisque in the validation data.

4.2. Model performance

This research tested different approaches to predicting individuals tree crowns and their segmentation. It was therefore necessary to evaluate the models to create a comparison. The evaluation was done using OA and F1-score, the segmentation evaluation additionally used the mIOU.

The models with the highest performance were used to predict the whole dataset, the results of which were used to assess the ability of model prediction the tree health status for the classes in the COSTA dataset.

4.2.1. Classification

The classification experiments have shown performance differences between datasets as well as between models in the same dataset. The best performing dataset was the augmented dataset. Within this dataset the ResNet50 and ResNet101 model performed best with OAs of 76% (Table 3). ResNet50 on augmented data (ResNet50-A) outperformed other classification models regarding F1-Score (0.78) (Table 4). The network ResNet34, fine-tuned on the standard dataset, showed the lowest OA (48%) of CNNs training on RGB datasets. All models training on the standard, 3-Flight or Augmented dataset trained for 60 epochs and used 5 as batch size. SVMs performed overall worse in OA as well as F1-Score than any CNNs training on the same datasets. SVMs produced OA between 32% and 36%. The SVM training on the Augmented dataset showed the highest OA as well as F1-Score of all SMVs in this experiment (Table 3).

Table 3: Model accuracies and F1-Scores of the classification experiment

	Mean F1-Score	OA
Standard dataset		
Resnet34	0.49	0.48
Resnet50	0.62	0.64
Resnet101	0.67	0.68
SVM	0.31	0.32
3-Flight dataset		
Resnet34	0.75	0.72
Resnet50	0.58	0.60
Resnet101	0.67	0.68
SVM	0.31	0.32
Augmented dataset		
Resnet34	0.68	0.72
Resnet50	0.78	0.76
Resnet101	0.74	0.76
SVM	0.33	0.36
Multispectral dataset		
Resnet34	0.47	0.50
Resnet50	0.58	0.58
Resnet101	0.53	0.54
SVM	0.32	0.33

ResNet50-A achieved an F1-Score of 0.78. This network achieved F1-scores of 0.8 or higher in all but two classes. Lowest F1-Scores were achieved in the classes Caoba (0.5) and Guapinol (0.67). The classes Guacimo, Tempisque and RonRon performed equal with F1-Scores of 0.8 (Table 4).

Table 4: ResNet50 trained on the Augmented dataset prediction results of unseen tree crown images.

Classes	F1-Score	Occurrences
Caoba	0.5	2
Guacimo	0.8	3
Guapinol	0.67	4
Other	0.88	9
RonRon	0.8	2
Tempisque	0.8	5
Weighted avg	0.78	
Overall accuracy	0.76	

The confusion matrix (Figure 10) gives a detailed impression on ResNet50-A's test data prediction. The columns show the model prediction while the rows show the designated label. Brighter cell colors indicate higher prediction counts for the specific combination. White cells have the highest number of ground truth to prediction count while black has no entry for the combination of ground truth and model prediction. While ResNet50-A correctly classified most of the individual tree crowns overfitting in the Caoba class was observed (Figure 10).

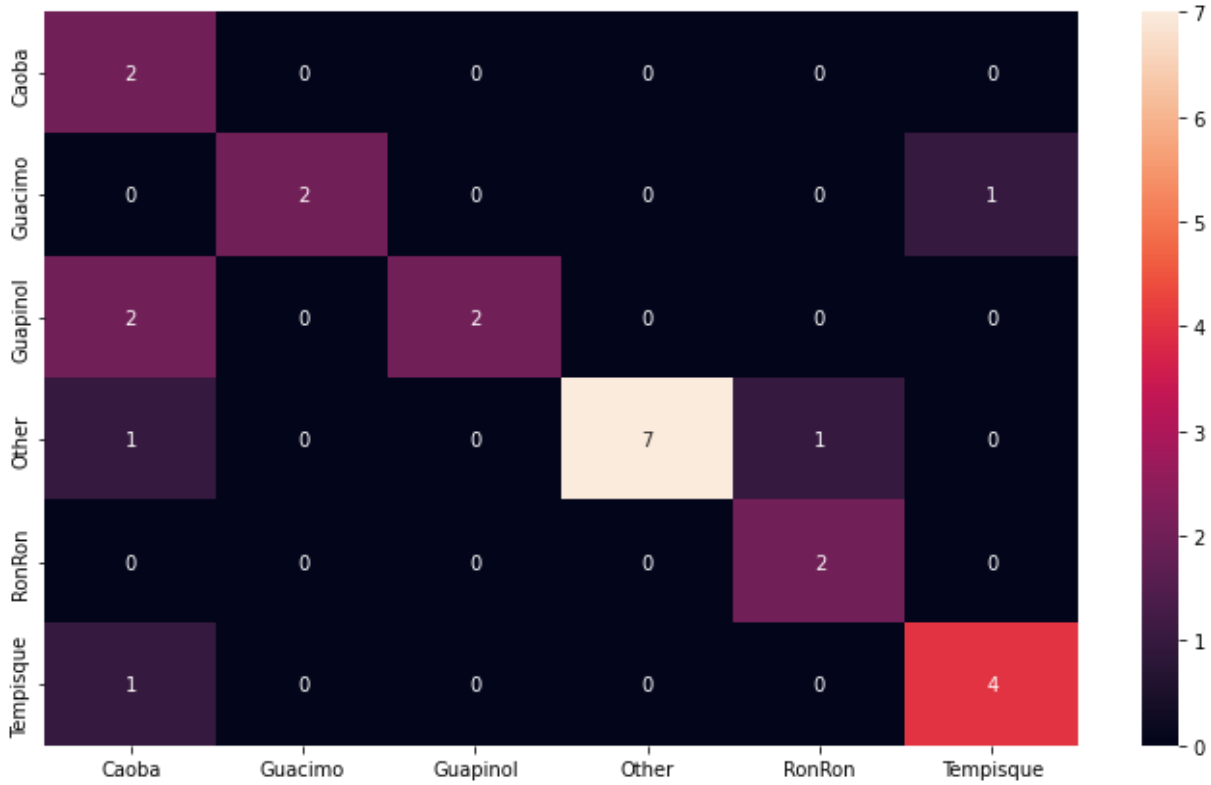


Figure 10: Confusion matrix of the test data prediction by ResNet50-A

In contrast to this the best performing SVM showed overfitting in the class "Other" (Figure 11). SVM did not surpass OAs of 36% (Table 3). The same accuracy was achieved by placing all

tree images in the class other. The per class F1-Scores were considerably lower compared to CNNs (Sup. Table 3). SVMs were unable to correctly identify the classes Caoba, Guapinol and RonRon in the COSTA classification dataset.

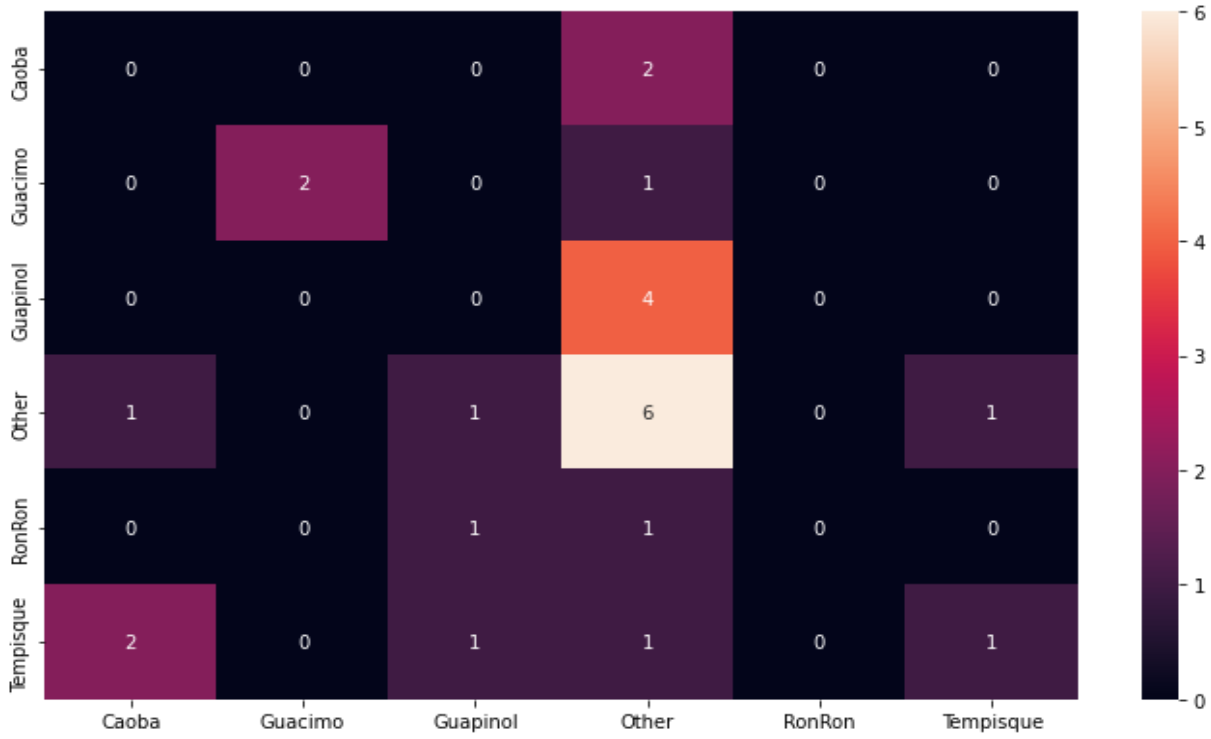


Figure 11: SVM prediction results unseen data.

Networks fine-tuned on the 5-band multispectral dataset achieved OAs of up to 58% (Table 5). They were outperformed by all networks training on RGB data with the exception ResNet34 fine-tuned on the standard dataset (Sup. Table 2). Multispectral models required more epochs to learn as can be seen by the loss values reaching a plateau after 100+ epochs (Sup. Figure 4). Learning was therefore conducted for 150 epochs. The model performing best on the MS dataset was ResNet50. This model was not able to correctly predict the Caoba class in the test data (Table 5). The model reached a mean F1-Score of 0.58 and had highest success in predicting the Guacimo class in the test data.

Table 5: F1-Score and OA of ResNet50. Trained on 5-Band MS dataset and predicting unseen data.

Classes	F1-Score	Occurrences
Caoba	0.00	2
Guacimo	0.80	3
Guapinol	0.67	4
Other	0.67	9
RonRon	0.50	2
Tempisque	0.44	4
Weighted avg	0.58	
Overall accuracy	0.58	

A non-pretrained model was tested to compare the results against the fine-tuned models (Figure 12). This comparison showed inferior performance of the non-pretrained model on the 5-band multispectral dataset compared to models fine-tuned on the COSTA multispectral dataset.

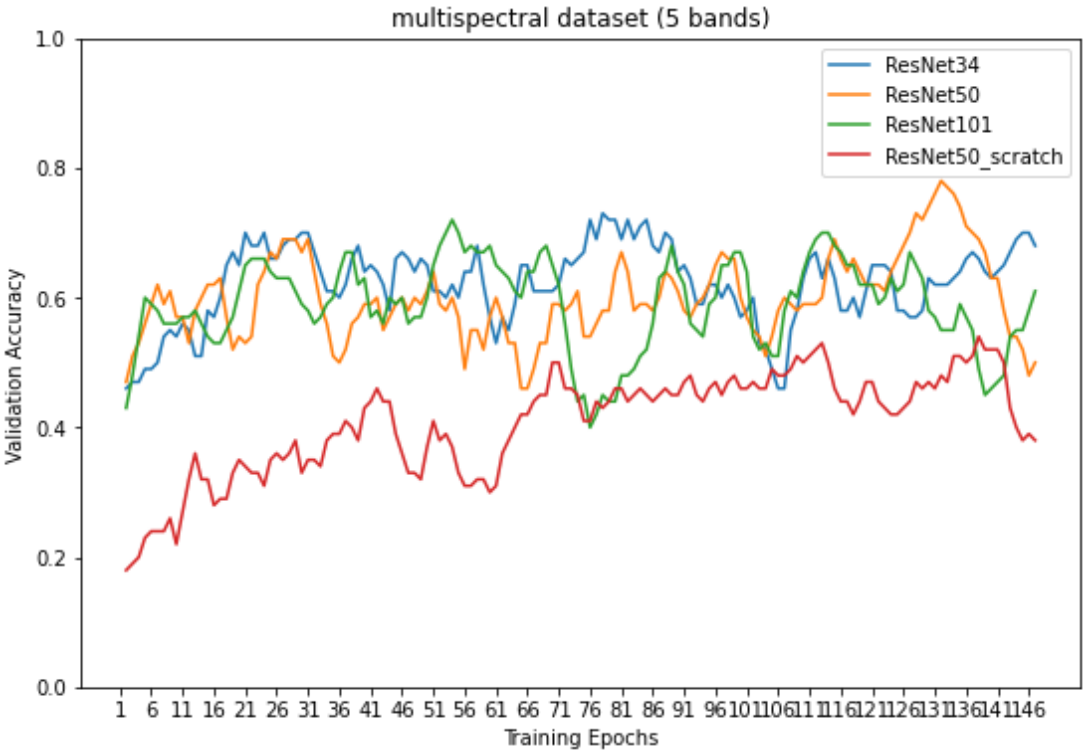


Figure 12: Comparison of validation accuracies achieved by CNNs fine-tuned on the multispectral COSTA dataset with a non-pretrained CNN (ResNet50_scratch).

Model prediction results on the combined COSTA dataset are shown in Figure 13. The model prediction on the combined COSTA dataset reached 76% OA. The results of the combined prediction are used for VI comparison with the refernce data.

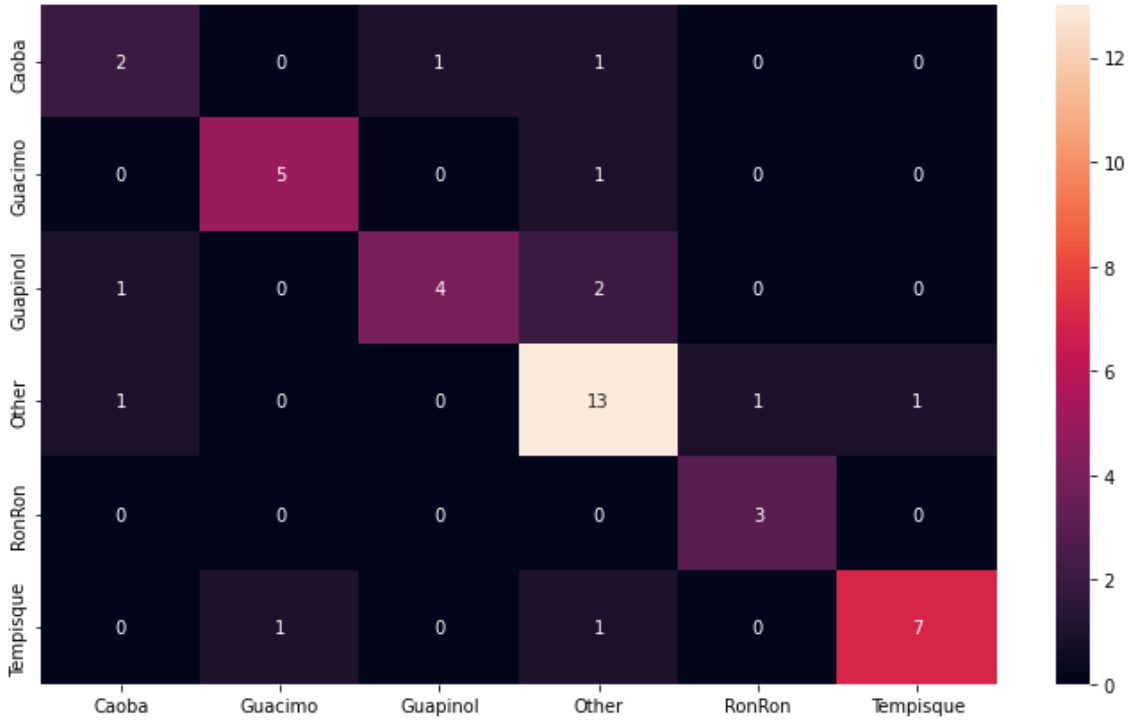


Figure 13: Combined COSTA dataset prediction.

4.2.2. Semantic segmentation

All semantic segmentation networks reached their minimal validation loss withing 40 epochs (Figure 14). The validation loss of all models in this test increased after reaching their respective validation loss minimum. The error score increase indicates overfitting and therefore hints at a finished training process on the dataset at hand.

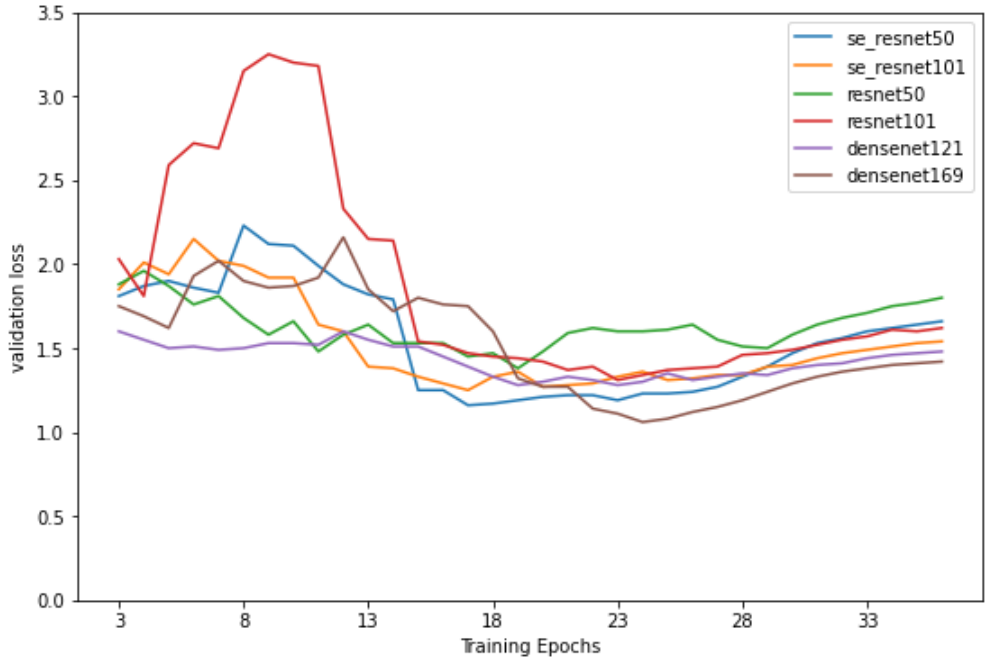


Figure 14: Validation loss of semantic segmentation training.

Model performance results for predicting semantic segmentation test data revealed OAs between 46% and 63% (Table 6). Both Se_ResNet layer depths performed similar in F1-Score, OA and mIOU. This was in contrast to Densenet and ResNet layer depths which display performance variability. The single best model performance was achieved by ResNet101 with an OA of 63%, F1-Score of 0.64 and reached mIOU 0.37. The single worst mIOU was produced by ResNet50 (0.22). Densnet169 achieved a similar mIOU (0.33) compared to Densenet121 (0.32) while performing inferior in OA and F1-Score.

Table 6: Overview on Semantic Segmentation model performance

	Mean F1-Score	Overall Accuracy	mIOU
ResNet50	0.52	0.51	0.22
ResNet101	0.64	0.63	0.34
Se_ResNet50	0.63	0.61	0.35
Se_ResNet101	0.62	0.6	0.37
Densenet121	0.64	0.61	0.32
Densenet169	0.45	0.46	0.33

Class-specific results are shown for model Se_ResNet101 which had achieved a F1-Score for Caoba of 0.63 which was the highest for any of the tree species classes (Table 7). This class, except background, was best detected throughout all models in the semantic segmentation experiment. Caoba holds the highest area-related share of any tree species within the dataset. All models scored the lowest F1-Scores for RonRon which held the smallest area-related share in the test dataset. Class-specific F1-Scores for other classes however varied strongly between model predictions (Sup. Table 5).

Table 7: Detailed results of Se_ResNet101 test data prediction.

Classes	F1-Score	Area-related share
Background	0.65	66.31%
Guapinol	0.50	3.88%
Guacimo	0.41	4.86%
Caoba	0.63	16.21%
RonRon	0.38	3.47%
Tempisque	0.62	5.28%
Weighted avg	0.62	
Overall accuracy	0.60	

The combined semantic segmentation COSTA dataset used data already used in training and validation. Semantic segmentation prediction results show increased area-related shares for

all species (Table 8). Highest area-related shares gains were achieved by the Guacimo class with an increase of 8% and a 16% increase for the Caoba class (Table 8). The pronounced edge effect was a result of splitting the study area into image tiles with edge length of 500 pixels (Figure 15).

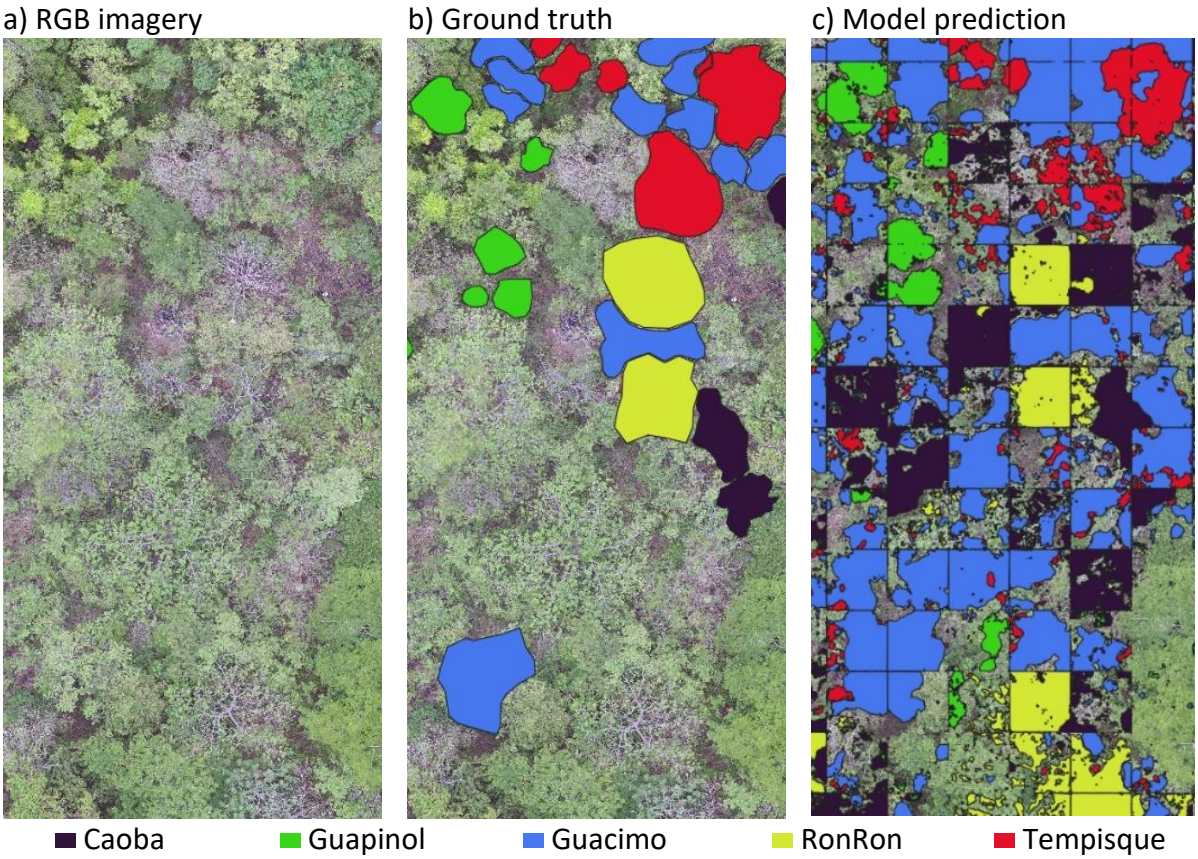


Figure 15: Excerpt of semantic segmentation prediction results. The edge effect of the tiles is clearly visible.

Area-related shares (Table 8) for reference data differed from aera-related shares in the semantic segmentation dataset (Table 2) as the combined COSTA semantic segmentation dataset included background tiles, which were excluded from the semantation dataset.

Table 8: Species specific area-related shares in reference and prediction data.

Class	Reference	Prediction
Guapinol	1.02%	2.53%
Guacimo	1.30%	9.91%
Caoba	2.06%	18.34%
RonRon	0.76%	3.03%
Tempisque	1.42%	2.48%

4.3. Vegetation status evaluation

VI results for the classification datasets revealed differences in VI means between classes (Table 9). Highest mean values for NDVI (0.83), EVI (0.61) and CI (0.84) in the reference data were reached by the Tempsique class. The highest NDRE mean value (0.28) was shared by classes Guapinol and Tempisque. The same distribution was observed for the VI means on predicted classes. Class RonRon displayed the lowest VI mean values of all classes.

Table 9: Class specific VI means in the COSTA combined classification dataset. CI values are calculated with Red-Edge spectral information. Means are presented with standard deviance in brackets. Highest means for each VI are highlighted.

Species	NDVI		NDRE		CI		EVI	
	Reference	Prediction	Reference	Prediction	Reference	Prediction	Reference	Prediction
Caoba	0.68 (0.06)	0.67 (0.10)	0.21 (0.02)	0.20 (0.01)	0.55 (0.06)	0.51 (0.03)	0.39 (0.07)	0.37 (0.10)
Guacimo	0.73 (0.05)	0.72 (0.04)	0.22 (0.02)	0.21 (0.03)	0.59 (0.06)	0.55 (0.10)	0.50 (0.06)	0.50 (0.05)
Guapinol	0.79 (0.05)	0.77 (0.06)	**0.28** (0.06)	0.29 (0.06)	0.83 (0.22)	0.84 (0.21)	0.50 (0.07)	0.48 (0.06)
Other	0.76 (0.10)	0.77 (0.08)	0.27 (0.04)	0.27 (0.04)	0.75 (0.17)	0.77 (0.17)	0.51 (0.13)	0.51 (0.10)
RonRon	0.56 (0.09)	0.58 (0.09)	0.21 (0.04)	0.22 (0.03)	0.55 (0.12)	0.57 (0.11)	0.33 (0.04)	0.35 (0.05)
Tempisque	**0.83** (0.05)	**0.85** (0.02)	**0.28** (0.07)	**0.30** (0.05)	**0.84** (0.26)	**0.91** (0.20)	**0.61** (0.09)	**0.66** (0.04)

NDRE and NDVI class differences (Figure 16) as a measure of canopy greenness and development showed similar reference values in NDVI for the classes Guacimo, Guapinol and Other. NDRE results on the other hand showed near equal mean values for classes Caoba, Guacimo and RonRon with Guapinol and Other achieving higher means.

VI creation from predicted tree crown classification revealed highest mean values in all VIs for class Tempisque. The calculations of VIs from CNN predictions on the COSTA combined dataset were able to reproduce the observed trends in VI mean values derived from reference data.

Figure 16: Box-Plots of reference (left) and prediction (right) derived VIs. Normalized Difference Red Edge Index (NDRE), Red-Edge Chlorophyll Index (CI), Normalized Difference Vegetation Index (NDVI) and Enhanced Vegetation Index (EVI) are displayed.

VI values derived from the reference semantic segmentation data showed the same distribution of VI mean values as observed for the classification dataset reference. Semantic segmentation prediction VIs were consistent for the classes Background, Guacimo, RonRon and Tempisque (Table 10). However, prediction results differed from reference data for classes Guapinol and Caoba.

Table 10: Class specific VI means in the COSTA combined semantic segmentation dataset. CI values are calculated with Red-Edge spectral information. Highest means for each VI are highlighted. Values in brackets is the difference to the reference data.

Species	NDVI		NDRE		CI		EVI	
	Reference	Prediction	Reference	Prediction	Reference	Prediction	Reference	Prediction
Background	0.63	0.64 (-0.01)	0.22	0.22 (0.00)	0.60	0.61 (-0.01)	0.37	0.37 (0.00)
Guapinol	0.79	0.74 (0.05)	**0.31**	0.26 (0.05)	**0.92**	0.73 (0.20)	0.48	0.45 (0.03)
Guacimo	0.74	0.75 (-0.01)	0.24	0.24 (0.00)	0.65	0.66 (-0.01)	0.47	0.49 (-0.02)
Caoba	0.71	0.53 (0.18)	0.22	0.19 (0.03)	0.58	0.50 (0.08)	0.43	0.29 (0.14)
RonRon	0.63	0.63 (0.00)	0.24	0.23 (0.01)	0.66	0.64 (0.03)	0.36	0.37 (-0.02)
Tempisque	**0.81**	**0.80** (0.00)	0.27	**0.27** (0.00)	0.78	**0.79** (-0.01)	**0.58**	**0.59** (-0.01)

Observed differences were most pronounced in NDVI (-0.18) (Figure 18) and EVI (-0.14) (Figure 17) for the Caoba class. Guapinol showed the biggest difference (-0.05) of all classes between reference and prediction data in the NDRE (Figure 17). The class further depicted the biggest decline in CI (- 0.19) between reference and prediction (Figure 18,Table 10).

Guapinol VIs constructed on the prediction dataset and using Red-Edge in the calculation showed a decrease of 16.4% for NDRE and 21.2% for CI compared to reference data. This was in contrast to VIs constructed from visual spectral information, in which declines of 6.9% for EVI and 6.4% for NDVI were observed. The polar opposite was observed for the class Caoba. NDVI values decreased by 25.3%, EVI by 32.2% while CI values declined by 13.6 and NDRE by 13.4%.

The creation of VIs from the prediction of semantically segmented tree species via FCNs on the COSTA combined dataset matched the observed reference VIs for three of five studied species in this experiment. Predicted VIs for Caoba and Guapinol were lower than in the reference data.

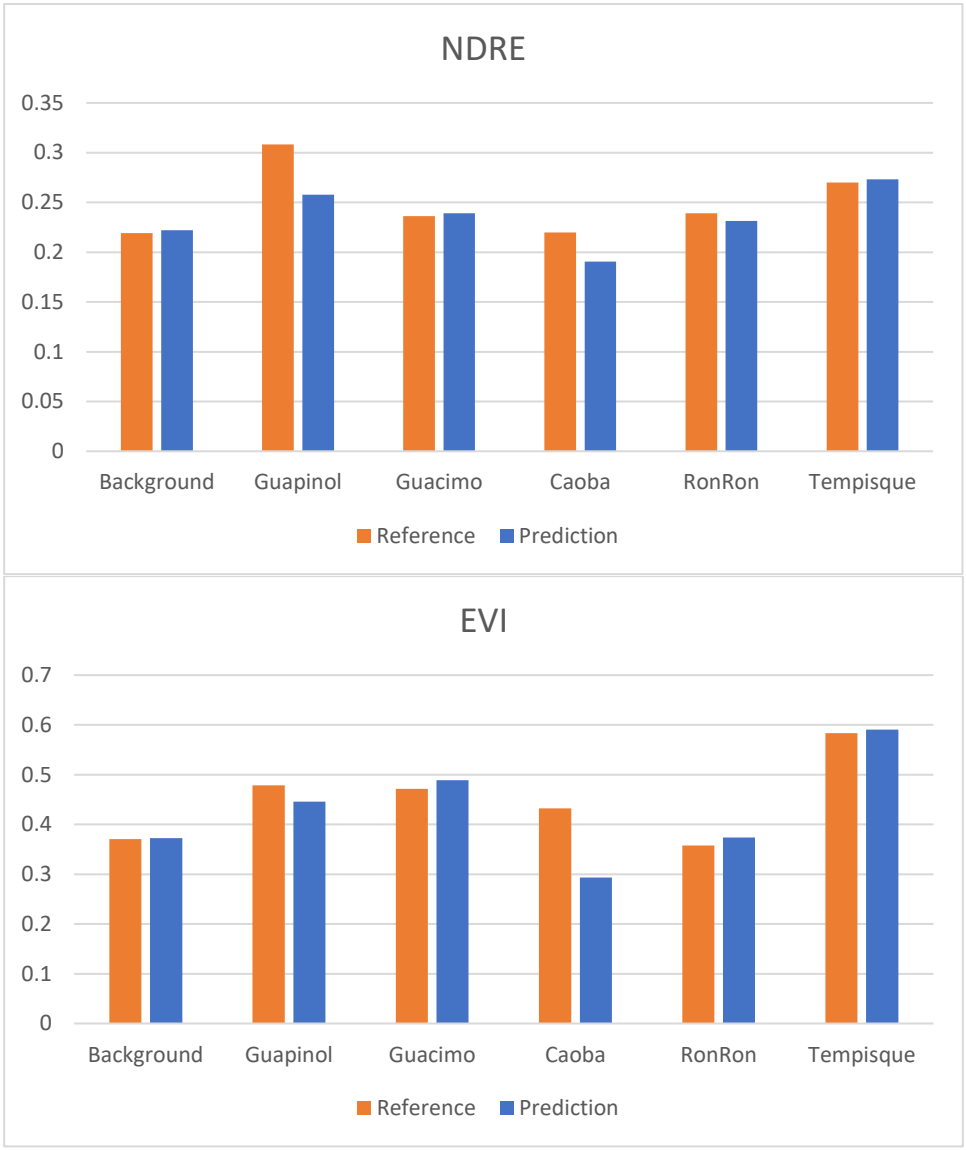


Figure 17: Plot of NDRE and EVI results for the semantic segmentation combined COSTA dataset. Values displayed are class means.

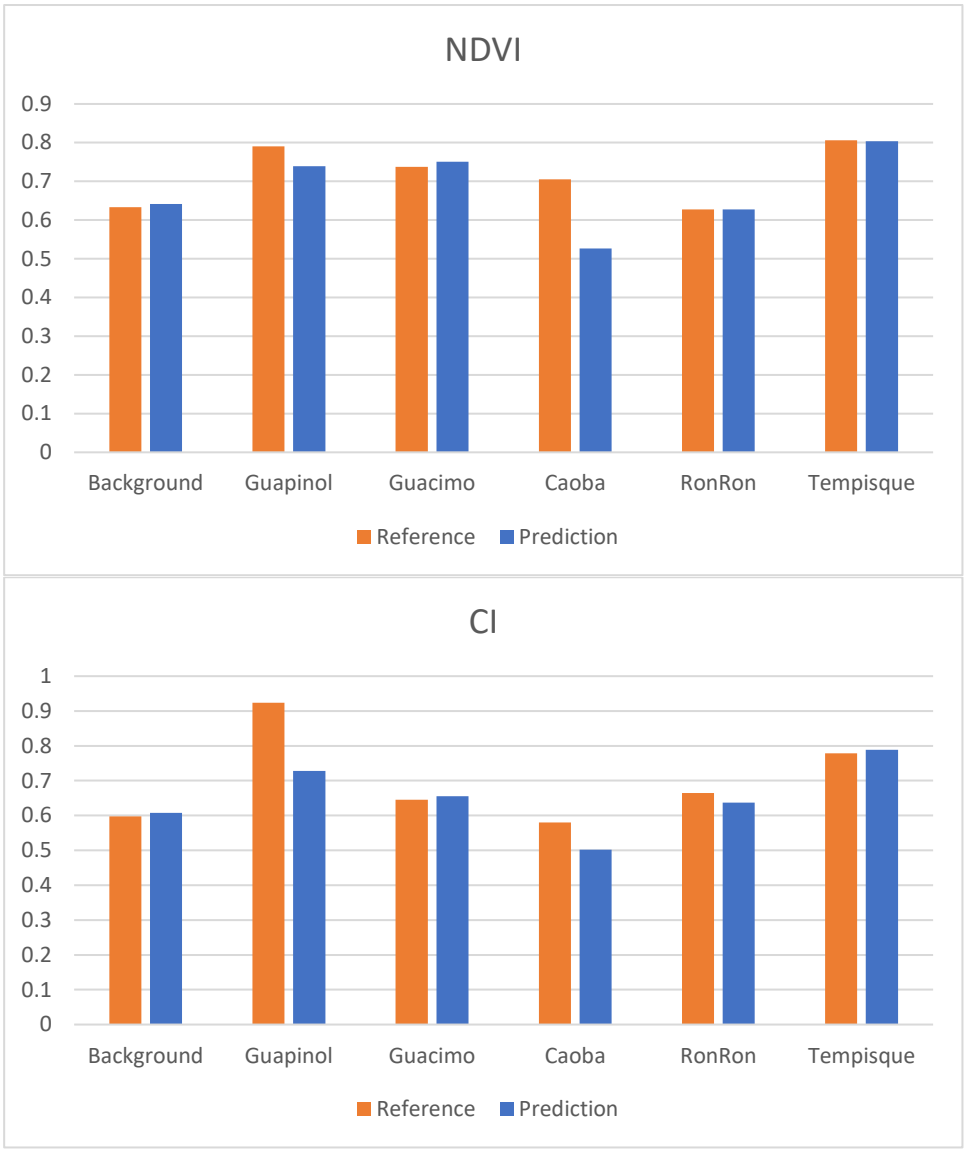


Figure 18: Plot of NDVI and CI_Red-Edge (CI) results for the semantic segmentation combined COSTA dataset. Values displayed are class means.

5. Discussion

The labor-intensive manual delineation proofed to be the only valid solution to create the classification and semantic segmentation datasets. As the delineation results of watershed segmentation (Figure 8) need manual correction to a degree in which complete manual delineation was less time intensive. The inaccuracy of watershed segmentation for dense trees is documented in literature (Hartling et al. 2021, Natesan et al. 2019) and was further confirmed in this study. Visual algorithms such as the multiresolution algorithm have proven to successfully delineate densely populated woodlands (Onishi & Ise 2021), but a required license was unavailable for this study.

Model performance

CNNs are well known for their ability to classify individual tree crowns (Mäyrä et al. 2021, Natesan et al. 2019, Onishi & Ise 2021). This study confirmed this ability on the comparatively small COSTA dataset as the tested CNNs were able to classify unseen individual tree crown images with OAs of up to 76% (Table 3). This shows that CNNs are able to learn features in the dataset and can be used to correctly classify tree crowns. This is in line with earlier observations (Natesan et al. 2019, Onishi & Ise 2021). The achieved accuracies in this study are comparable with other studies using only RGB information of tree crowns (Natesan et al. 2019). Lack however compared to Pseudo-RGBs (Onishi & Ise 2021).

While the models were able to classify the individual tree crown images reasonably well, they had difficulties correctly classifying the Coaba class (F1-Score = 0.5). This is in all likely hood due to the heterogeny of the foliage cover during the image acquisition and the presence of liana in the canopy altering the characteristics the model trains on (Figure 19). In class heterogeny also proofed to lower classification accuracies in literature (2019, Onishi & Ise 2021).

Figure 19: Heterogeny in appearance within the class Caoba.

Testing classification of individual tree crowns using the 5-band multispectral dataset did not increase OAs. This is in contrast to other studies that found higher overall accuracies if hyperspectral data is used is model training (Fricker et al. 2019). This contradiction could be explained by the increased complexity of the model in combination and the small dataset being insufficient for the model to successfully train. Larger datasets could however proof the superiority of models trained on multispectral data in classification results. This study was not able to show these results.

The inferior performance of the 5-band multispectral classification networks could be explained with the lesser resolution of the multispectral data. Its resolution was 2.5 cm compared to RGB data which had a 1 cm resolution.

SVMs showed to be less accurate in correctly predicting unseen data. This is in line with screened literature (Onishi & Ise 2021). The SVMs used in this study did not exceed OAs of 36%. Other studies using 5-band multispectral and equally sized datasets achieved comparable OAs (Hartling et al. 2021). 36% OA in the COSTA dataset is however significant as classifying the whole test dataset as “Other” would yield the same results. SVMs proofed insufficient as a mean to predict classes of individual tree crown images in the COSTA dataset.

Classification results of model prediction on the combined COSTA dataset reach an OA of 76%. The prediction was done on the combined COSTA classification dataset which consists of the standard test and validation data. There are two possible reasons for the combined dataset to might have compromised result quality. Firstly, the model is fitted to the validation data, thus the model performing best on the validation data (lowest error score) is chosen as model for the prediction. Secondly, while test data was the same for all models, validation data was randomly selected from the remaining data. As the best performing model trained on the

Augmented dataset it could have used parts of the standard validation set in model training. However, considering the size of the COSTA classification dataset as well as class heterogeny, the results show that CNNs are able correctly predict individual tree crowns in the highly divers combined COSTA dataset.

The accuracies achieved by FCNs in this study did not reach the reported OAs in other literature (Schiefer et al. 2020). There are two likely reasons for this observation. Firstly, the combination of high class amount to comparatively small dataset (Schiefer et al. 2020). Schiefer et al. (2020) classified 14 classes on 51 ha while the COSTA semantic segmentation dataset consist of 6 classes on less than 3 ha combined COSTA segmentation dataset, of which 3,600 m² are used as training data. Secondly, the previously mentioned high in class heterogeny (Schiefer et al. 2020). The results of the model testing are, in regards to dataset size and class heterogeny, impressive. The models are able to correctly predict some classes. Overfitting caused liberate classification of the classes Caoba and Guacimo (Figure 20, Table 8). Both of which show in class heterogeny. The missing foliage of RonRon tree crowns is likely to have caused the miss classification of empty branches as RonRon (Figure 20).

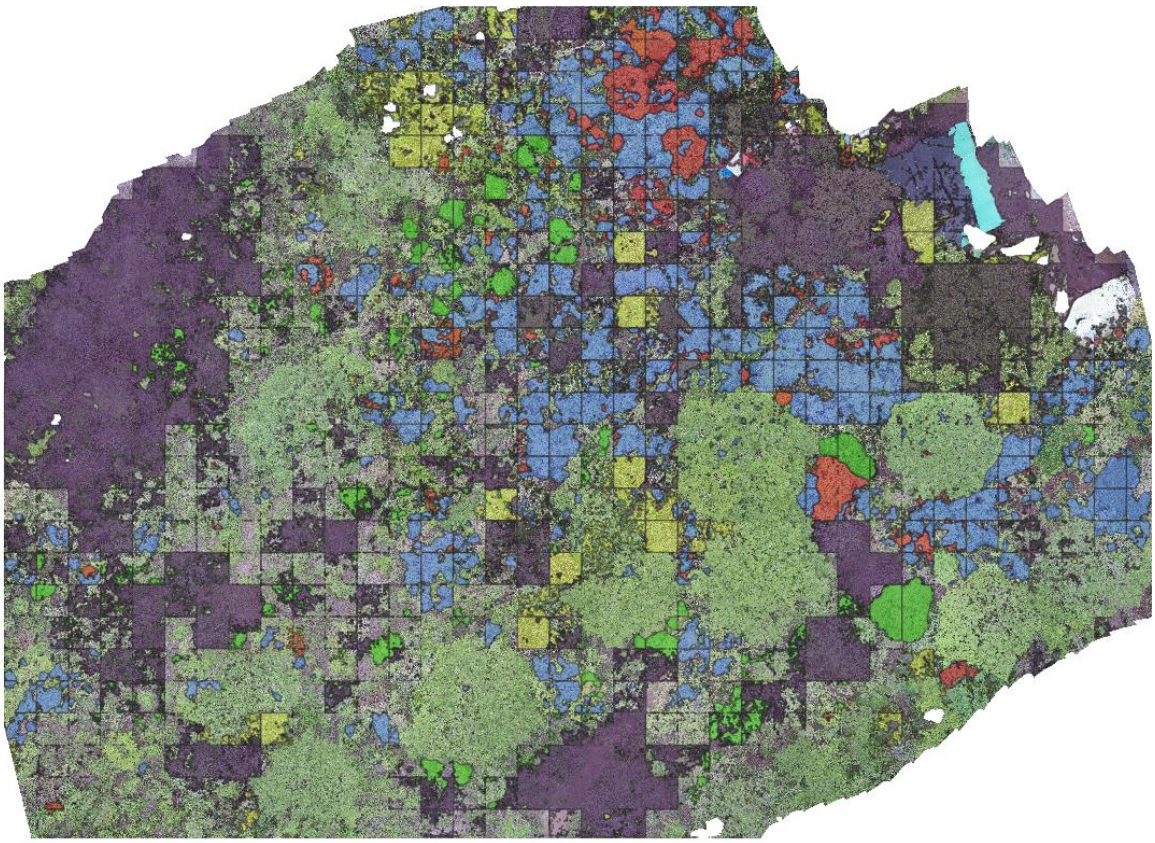


Figure 20: Semantic segmentation prediction of the whole COSTA dataset. The prediction of classes Caoba (Black), Guacimo (Blue), Guapinol (Green), Tempisque (Red) and RonRon (Yellow) is shown.

Due to the variety of CNN approaches, CNN architectures, forest types, and tree species studied make a more detailed comparison of classification and semantic segmentation results with existing literature increasingly difficult.

Vegetation Indices

Results of the reference VIs reveal differences in plant health status for the observed tree species. The low VI means in NDVI, NDRE and EVI are expected for the class RonRon, as the class did possess no to little foliage cover at the time of data acquisition. While displaying similar EVI means as Guacimo, Guapinols NDRE mean value is considerably higher (Table 9). The observed disconnect hints to higher leave chlorophyll contents as well as higher amounts canopy leaf layers for Guapinol (Boiarskii 2019, Huete 2012). This is further amplified by the high CI values observed in the Guapinol class. The combination of VIs suggests that Tempisque followed by Guapinol are the healthiest tree species analyzed in the COSTA dataset. This observation is in line with non-leaf shedding species at the end of dry season (April). Interestingly, Guapinol and Tempisque outperform Caoba in VI values which, as the two other species, does not shed leaves during dry season (Kühnhammer et al. 2022).

The VI values derived from the individual tree crown prediction with CNNs yielded comparable results if compared to the reference VI values. Predicted spatial information by CNN classification is therefore sufficient to assess species specific tree health status in the COSTA dataset.

VIs created with the use of the semantic segmentation data where comparable for most classes. However, Caoba and Guapinol classes showed considerable differences in VIs (Table 10). Guapinol NDRE(Figure 17) and CI (Figure 18)mean value predictions displayed a rise, whereas prediction of the class Caoba was impacted increases in NDVI (Figure 18) and EVI (Figure 17) mean values.

This observation could again be explained by the heterogeny in the Caoba class and resulting misclassifications in RGB imagery. The smaller decrease in VIs constructed from visual spectral information suggests misclassification of trees which possess canopy but not as developed as Guapinol individuals reference data. Reasons for the misclassifications could be threefold. Firstly, bias in the creation of the reference data. Secondly, dataset size restrictions and high in-class variability in reflectance of the Red-Edge spectrum, which results in observed high

variability in NDRE and CI results (Figure 16). Lastly, the limit of RGB input data for semantic segmentation prediction of Guapinol, this however is less likely than reasons one and two.

VIs derived from the semantic segmentation prediction suggest that Tempisque is the healthiest species as it achieves highest means in all analyzed VIs in the combined COSTA semantic segmentation dataset. Second best VI mean values were achieved by Guapinol, matching the trend observed in the classification data. This is in line with reference VIs.

The combined COSTA semantic segmentation dataset included training and validation data. This approach was chosen to increase the dataset size for VI interpretation. The approach further limits the information value gained form the VI comparison of reference and model prediction. As model training fits the network to the training data. The prediction of data used in training is therefore more accurate than accuracies of unseen data. This can produce VI values which are not representative for the model performance.

Observed differences in reference VI values (Caoba, Guapinol) compared to semantic segmentation prediction VIs caused by general model performance show the limitations of the COSTA dataset. The FCN approach tested in this study was not able to reach accuracies in species prediction on the COSTA segmentation dataset to create reliable species specific VI values. Tree health estimations on the basis of the predictions could therefore differ significantly from actual species wellbeing.

5.1. Limitations and improvements

The most likely reason for the low observed accuracies compared to literature is the small size of the COSTA dataset. Several options to increase the dataset and with that the OAs are listed in literature. (Natesan et al. 2019) were able to significantly increase model performance by adding same site imagery of three years to the dataset. This study used same site imagery of three consecutive days and with that was unable to show the reported increase in model performance. A different mean to create a bigger dataset is to increase the study area (Onishi & Ise 2021, Schiefer et al. 2020). Higher counts of class members also helps the generalization of the model and would decrease the observed problems in dealing with in class heterogeny (Natesan et al. 2019, Onishi & Ise 2021, Schiefer et al. 2020). Another option to increase model performance discussed in literature, even though with contradicting results, is the addition of height information as model feature (Onishi & Ise 2021, Schiefer et al. 2020, Sothe et al. 2020).

The delineation and classification of reference data solely via the use of aerial imagery in dense woodlands such as the COSTA data can lead to miss classification in the Training data. In combination with the small dataset this may lead to decreased model performance.

The creation of VIs from semantic segmentation prediction accuracy could be increased by using multispectral data. This is hinted by the observed differences in Guapinol VI mean values constructed from Red-Edge spectral information, which is not included in RGB data.

A different classification approach using FCNs is called instance segmentation. Instance segmentation as well as semantic segmentation are able to classify each pixel of the input the additionally instance segmentation is able to segment individual objects. The output of the model is threefold; object class label, object bounding box and object mask (He et al. 2015, Lobo Torres et al. 2020). It would therefore greatly increase usability and transferability of FCNs, for use in forest mapping and health assessment. However, multiclass instance segmentation studies have not yet been published in scientific literature.

6. Conclusion

This research was done to answer two questions. Firstly, *Are CNNs and FCNNs able to correctly predict tree species at the COSTA study site?* While neither of the approaches tested achieved accuracies reported in literature, classification predictions with the use of CNNs showed considerable success in predicting the six classes of the COSTA dataset (OA = 76%). FCNs used to create semantic segmentation predictions reached OAs of 58% on COSTA test data. The prediction success of the chosen semantic segmenstation FCN approach is inconclusive.

The second question answered by this study was; *Are classification predictions via CNNs or semantic segmentation predictions via FCNs sufficient information to assess tree species health status with vegetation indices?* VI values derived from CNN classification predictions of the COSTA combined dataset displayed considerable similarity to the reference. Thus, CNN prediction results are considered sufficient to assess COSTA tree species heath status. Semantic segmentation prediction results were able to recreate VI values for some species but failed for Caoba and Guapinol. This study therefore concludes that the chosen FCN approach is not producing prediction with sufficient accuracy to predict tree species health status at the COSTA site.

However, considering the limitations COSTA dataset in form of size and heterogeny, model performance of CNNs and FCNs on the COSTA dataset is remarkable. The study showed the impact computer vision could have by predicting tree species in small, divers study sites and datasets. This makes computer vision in form of neural networks a valuable asset in predicting tree health status. If used on well understood species this remotely sensed tree health could enable assessment of below ground processes of ecosystems, such as ground water availability, nutrient availability or possibly root depth.

Bibliography

Boiarskii B. 2019. Comparison of NDVI and NDRE Indices to Detect Differences in Vegetation and Chlorophyll Content. *JMCMS* spl1 (4)

Buitinck L, Louppe G, Blondel M, Pedregosa F, Mueller A, et al. 2013. API design for machine learning software: experiences from the scikit-learn project. http://arxiv.org/pdf/1309.0238v1

Candiago S, Remondino F, Giglio M de, Dubbini M, Gattelli M. 2015. Evaluating Multispectral Images and Vegetation Indices for Precision Farming Applications from UAV Images. *Remote Sens* 7 (4):4026–47

Chenari A, Erfanifard Y, Dehghani M, Pourghasemi HR. 2017. Woodland mapping at single-tree levels using object-oriented classification of unmanned aerial vehicle (UAV) images. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.* XLII-4/W4:43–49

Fawcett D, Panigada C, Tagliabue G, Boschetti M, Celesti M, et al. 2020. Multi-Scale Evaluation of Drone-Based Multispectral Surface Reflectance and Vegetation Indices in Operational Conditions. *Remote Sens* 12 (3):514

Fricker GA, Ventura JD, Wolf JA, North MP, Davis FW, Franklin J. 2019. A Convolutional Neural Network Classifier Identifies Tree Species in Mixed-Conifer Forest from Hyperspectral Imagery. *Remote Sens* 11 (19):2326

Ghamisi P, Plaza J, Chen Y, Li J, Plaza AJ. 2017. Advanced Spectral Classifiers for Hyperspectral Images: A review. *IEEE Geosci. Remote Sens. Mag.* 5 (1):8–32

Hartling S, Sagan V, Maimaitijiang M. 2021. Urban tree species classification using UAV-based multi-sensor data fusion and machine learning. *GIScience & Remote Sensing* 58 (8):1250–75

He K, Zhang X, Ren S, Sun J. 2015. *Deep Residual Learning for Image Recognition*

Huete AR. 2012. Vegetation Indices, Remote Sensing and Forest Monitoring. *Geography Compass* 6 (9):513–32

Huete AR, Didan K, Shimabukuro YE, Ratana P, Saleska SR, et al. 2006. Amazon rainforests green-up with sunlight in dry season. *Geophys. Res. Lett.* 33 (6)

Hunt ER, Doraiswamy PC, McMurtrey JE, Daughtry CS, Perry EM, Akhmedov B. 2013. A visible band index for remote sensing leaf chlorophyll content at the canopy scale. *International Journal of Applied Earth Observation and Geoinformation* 21:103–12

Kühnhammer K, Dahlmann A, Iraheta A, Gerchow M, Birkel C, Marshall JD, Beyer M. 2022. Continuous in situ measurements of water stable isotopes in soils, tree trunk and root xylem: Field approval. *Rapid communications in mass spectrometry : RCM* 36 (5):e9232

Li H, Hu B, Li Q, Jing L. 2021. CNN-Based Individual Tree Species Classification Using High-Resolution Satellite Imagery and Airborne LiDAR Data. *Forests* 12 (12):1697

Lobo Torres D, Queiroz Feitosa R, Nigri Happ P, La Elena Cué Rosa L, Marcato Junior J, et al. 2020. Applying fully convolutional architectures for semantic segmentation of a single tree species in urban environment on high resolution UAV optical imagery. *Sensors (Basel, Switzerland)* 20 (2)

Mäyrä J, Keski-Saari S, Kivinen S, Tanhuanpää T, Hurskainen P, et al. 2021. Tree species classification from airborne hyperspectral and LiDAR data using 3D convolutional neural networks. *Remote Sens. Environ.* 256:112322

Natesan S, Armenakis C, Vepakomma U. 2019. ResNet-based tree species classification using UAV images. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.* XLII-2/W13:475–81

Onishi M, Ise T. 2021. Explainable identification and mapping of trees using UAV RGB image and deep learning. *Scientific reports* 11 (1):903

Paszke A, Gross S, Massa F, Lerer A, Bradbury J, et al. 2019. *PyTorch: An Imperative Style, High-Performance Deep Learning Library*

Ronneberger O, Fischer P, Brox T. 2015. U-Net: Convolutional Networks for Biomedical Image Segmentation. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, ed. N Navab, J Hornegger, WM Wells, AF Frangi, pp. 234–41. Cham: Springer International Publishing

Rusu AA, Rabinowitz NC, Desjardins G, Soyer H, Kirkpatrick J, et al. 2016. *Progressive Neural Networks*

Saura JR, Reyes-Menendez A, Palos-Sanchez P. 2019. Mapping multispectral Digital Images using a Cloud Computing software: applications from UAV images. *Heliyon* 5 (2):e01277

Schiefer F, Kattenborn T, Frick A, Frey J, Schall P, Koch B, Schmidtlein S. 2020. Mapping forest tree species in high resolution UAV-based RGB-imagery by means of convolutional neural networks. *ISPRS J. Photogramm. Remote Sens.* 170:205–15

Sothe C, Almeida CM de, Schimalski MB, La Rosa LEC, Castro JDB, et al. 2020. Comparative performance of convolutional neural network, weighted and conventional support vector machine and random forest for classifying tree species using hyperspectral and photogrammetric data. *GIScience & Remote Sensing* 57 (3):369–94

Sothe C, Dalponte M, Almeida CM de, Schimalski MB, Lima CL, et al. 2019. Tree Species Classification in a Highly Diverse Subtropical Forest Integrating UAV-Based Photogrammetric Point Cloud and Hyperspectral Data. *Remote Sens* 11 (11):1338

Tanaka S, Kawamura K, Maki M, Muramoto Y, Yoshida K, Akiyama T. 2015. Spectral Index for Quantifying Leaf Area Index of Winter Wheat by Field Hyperspectral Measurements: A Case Study in Gifu Prefecture, Central Japan. *Remote Sens* 7 (5):5329–46

The MathWorks, Inc., ed. 2021. *Introducing Deep Learning with MATLAB*

Trumbore S, Brando P, Hartmann H. 2015. Forest health and global change. *Science (New York, N.Y.)* 349 (6250):814–18

Tucker CJ. 1979. Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sens. Environ.* 8 (2):127–50

Volpi M, Tuia D. 2017. Dense Semantic Labeling of Subdecimeter Resolution Images With Convolutional Neural Networks. *IEEE Trans. Geosci. Remote Sensing* 55 (2):881–93

Xie Q, Dash J, Huang W, Peng D, Qin Q, et al. 2018. Vegetation Indices Combining the Red and Red-Edge Spectral Information for Leaf Area Index Retrieval. *IEEE J. Sel. Top. Appl. Earth Observations Remote Sensing* 11 (5):1482–93

Xue J, Su B. 2017. Significant Remote Sensing Vegetation Indices: A Review of Developments and Applications. *J. Sens.* 2017:1–17

Yosinski J, Clune J, Bengio Y, Lipson H. 2014. How transferable are features in deep neural networks?

Zheng G, Moskal LM. 2009. Retrieving Leaf Area Index (LAI) Using Remote Sensing: Theories, Methods and Sensors. *Sensors (Basel, Switzerland)* 9 (4):2719–45

Zimmermann S, Hoffmann K. 2020. Evaluating the capabilities of Sentinel-2 data for large-area detection of bark beetle infestation in the Central German Uplands. *J. Appl. Rem. Sens.* 14 (02):1

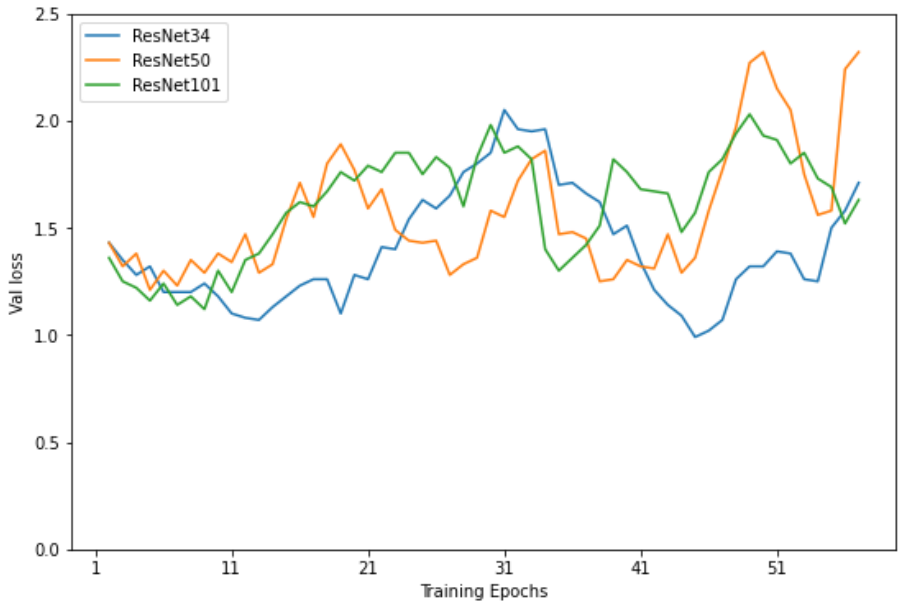
Appendix

3-flight dataset

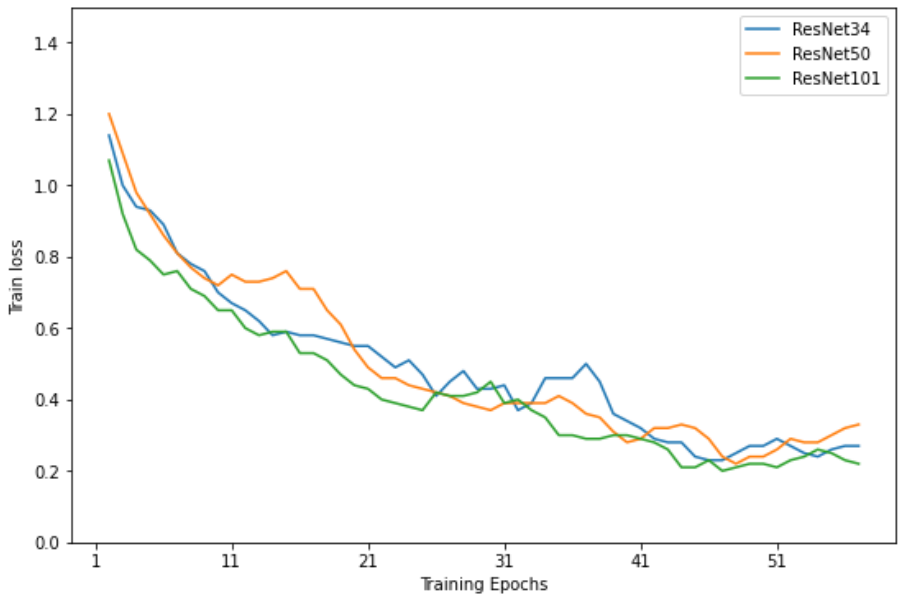
Sup. Table 1: Performance evaluation metrics of 3-flight CNNs and SVM on the classification test data.

Classes	F1-Score				Occurrences
	ResNet34	ResNet50	ResNet101	SVM	
Caoba	0.29	0.00	0.00	0.00	2
Guacimo	0.86	0.75	0.86	0.40	3
Guapinol	0.86	0.44	0.33	0.25	4
Other	0.78	0.60	0.70	0.42	9
RonRon	0.67	0.67	1.00	0.00	2
Tempisque	0.75	0.75	0.89	0.36	5
Weighted avg	0.72	0.58	0.67	0.31	
Overall accuracy	0.75	0.60	0.68	0.32	

a)



b)



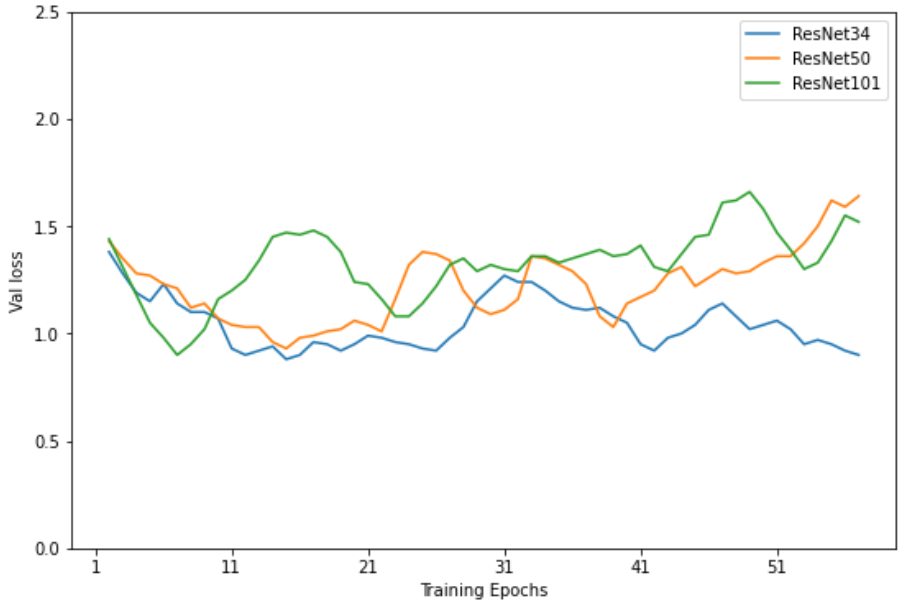
Sup. Figure 1: Smoothed plot of validation loss (a) and training loss (b) of CNNs fine-tuning on the 3-flight dataset

Standard dataset

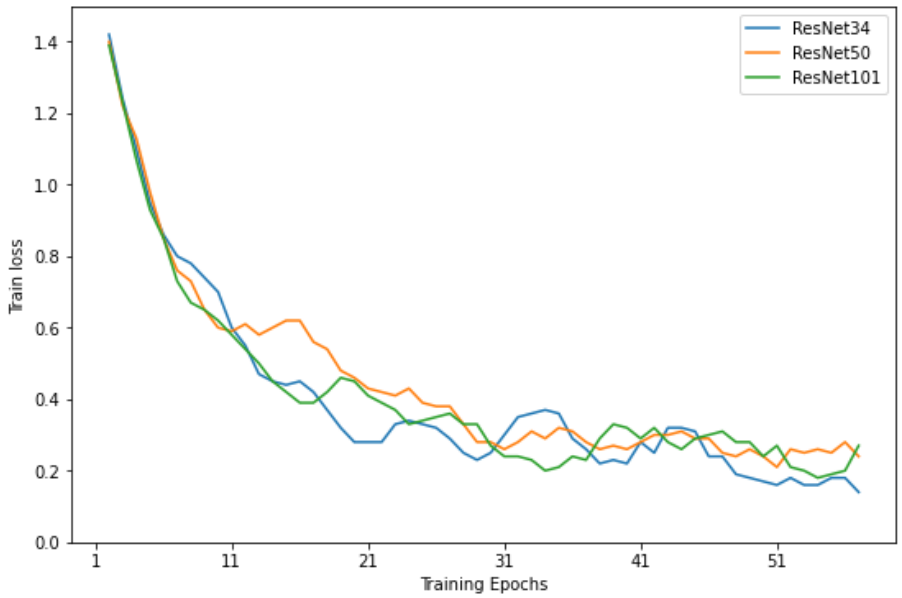
Sup. Table 2: Performance evaluation metrics of standard CNNs and SVM on the classification test data.

Classes	F1-Score				Occurrences
	ResNet34	ResNet50	ResNet101	SVM	
Caoba	0.00	0.00	0.00	0.00	2
Guacimo	0.50	0.75	1.00	0.40	3
Guapinol	0.44	0.40	0.67	0.25	4
Other	0.44	0.63	0.63	0.42	9
RonRon	0.67	1.00	1.00	0.00	2
Tempisque	0.73	0.80	0.71	0.36	5
Weighted avg	0.48	0.62	0.67	0.31	
Overall accuracy	0.49	0.64	0.68	0.32	

a)



b)



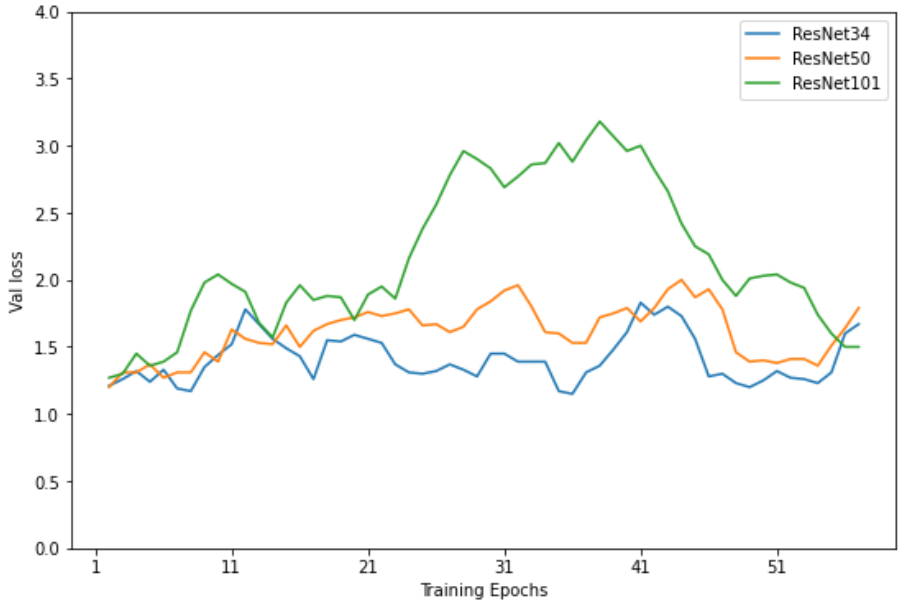
Sup. Figure 2: Smoothed plot of validation loss (a) and training loss (b) of CNNs fine-tuning on the standard dataset

Augmented dataset

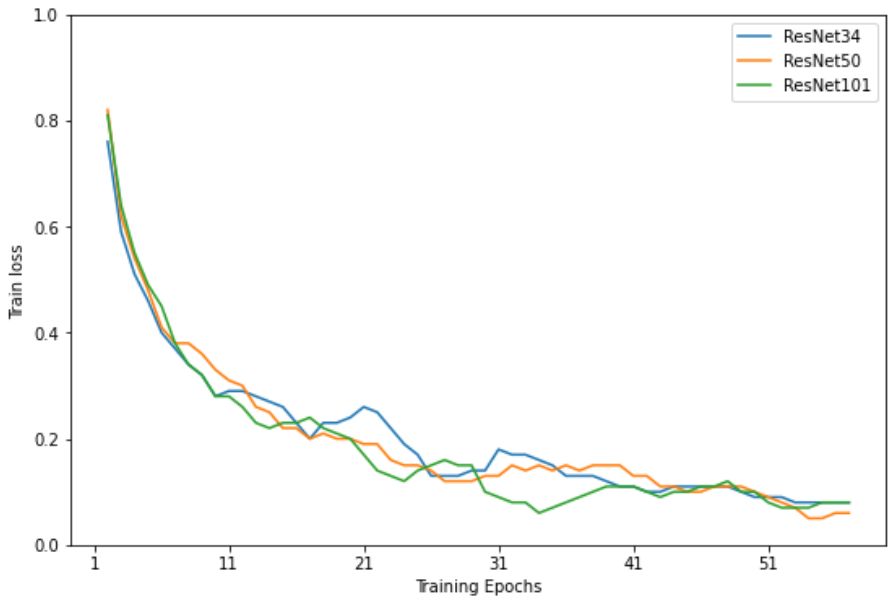
Sup. Table 3: Performance evaluation metrics of augmented CNNs and SVM on the classification test data.

Classes	F1-Score				Occurrences
	ResNet34	ResNet50	ResNet101	SVM	
Caoba	0.00	0.5	0.50	0.00	2
Guacimo	0.86	0.8	1.00	0.80	3
Guapinol	0.80	0.67	0.55	0.00	4
Other	0.71	0.88	0.95	0.50	9
RonRon	1.00	0.8	0.00	0.00	2
Tempisque	0.60	0.8	0.75	0.29	5
Weighted avg	0.68	0.78	0.74	0.33	
Overall accuracy	0.72	0.76	0.76	0.36	

a)



b)

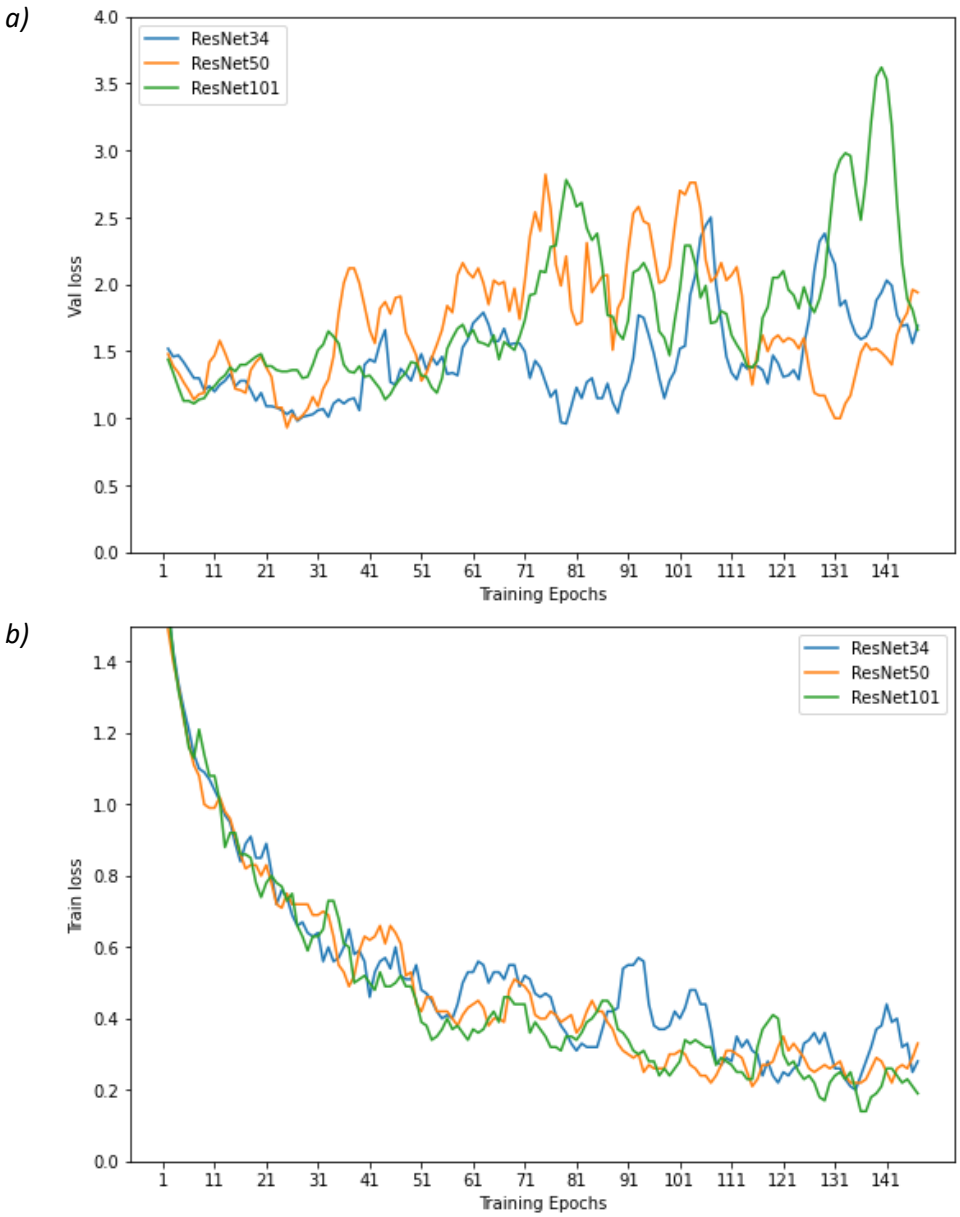


Sup. Figure 3: Smoothed plot of validation loss (a) and training loss (b) of CNNs fine-tuning on the augmented dataset

5-band multispectral dataset

Sup. Table 4: Performance evaluation metrics of multispectral CNNs and SVM on the classification multispectral test data. The non-pretrained (NP) ResNet50 is listed for comparison.

Classes	F1-Score					Occurrences
	ResNet34	ResNet50	ResNet101	ResNet50 (NP)	SVM	
Caoba	0.00	0.00	0.00	0.00	0.00	2
Guacimo	1.00	0.80	0.86	0.86	0.33	3
Guapinol	0.44	0.67	0.57	0.27	0.40	4
Other	0.60	0.67	0.56	0.14	0.45	9
RonRon	0.00	0.50	0.50	0.50	0.00	2
Tempisque	0.29	0.44	0.44	0.00	0.25	4
Weighted avg	0.47	0.58	0.49	0.25	0.32	
Overall accuracy	0.50	0.58	0.54	0.29	0.33	

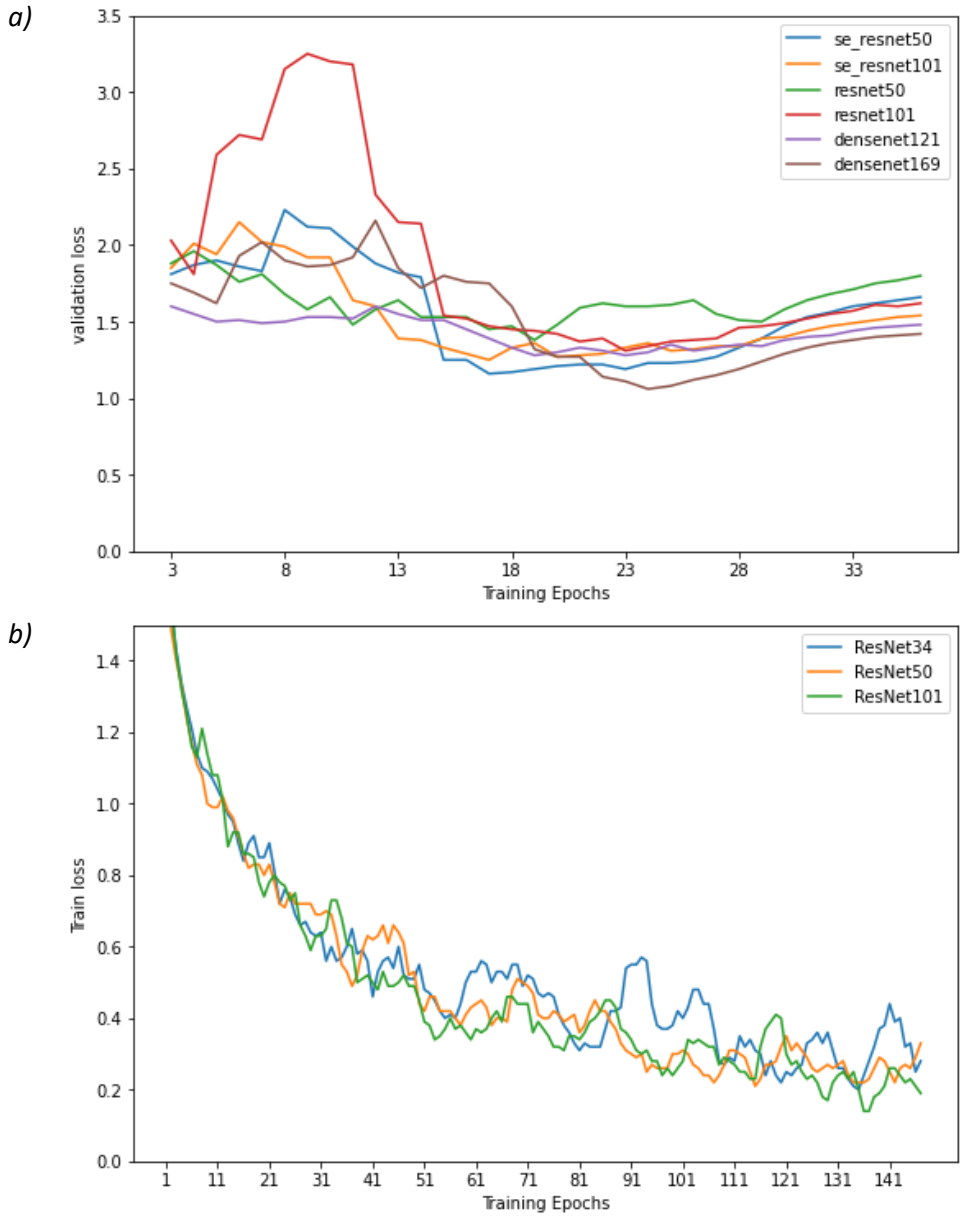


Sup. Figure 4: Smoothed plot of validation loss (a) and training loss (b) of CNNs fine-tuning on the 5-band multispectral dataset

Semantic Segmentation

Sup. Table 5: Performance evaluation metrics of FCNs with U-Net architecture on the semantic segmentation test data. Encoders are specified.

Classes	F1-Score						Area-related share
	Densenet121	Densenet169	Se_ResNet50	Se_ResNet101	ResNet50	ResNet101	
Background	0.72	0.41	0.68	0.65	0.70	0.70	66.31%
Guapinol	0.44	0.38	0.48	0.50	0.34	0.54	3.88%
Guacimo	0.29	0.36	0.56	0.41	0.37	0.38	4.86%
Caoba	0.58	0.57	0.64	0.63	0.04	0.65	16.21%
RonRon	0.27	0.59	0.19	0.38	0.21	0.26	3.47%
Tempisque	0.54	0.60	0.48	0.62	0.39	0.39	5.28%
Weighted avg	0.64	0.49	0.63	0.62	0.53	0.64	
OA	0.61	0.46	0.61	0.60	0.52	0.63	



Sup. Figure 5: Smoothed plot of validation loss (a) and training loss (b) of FCNs fine-tuning on the semantic segmentation dataset.