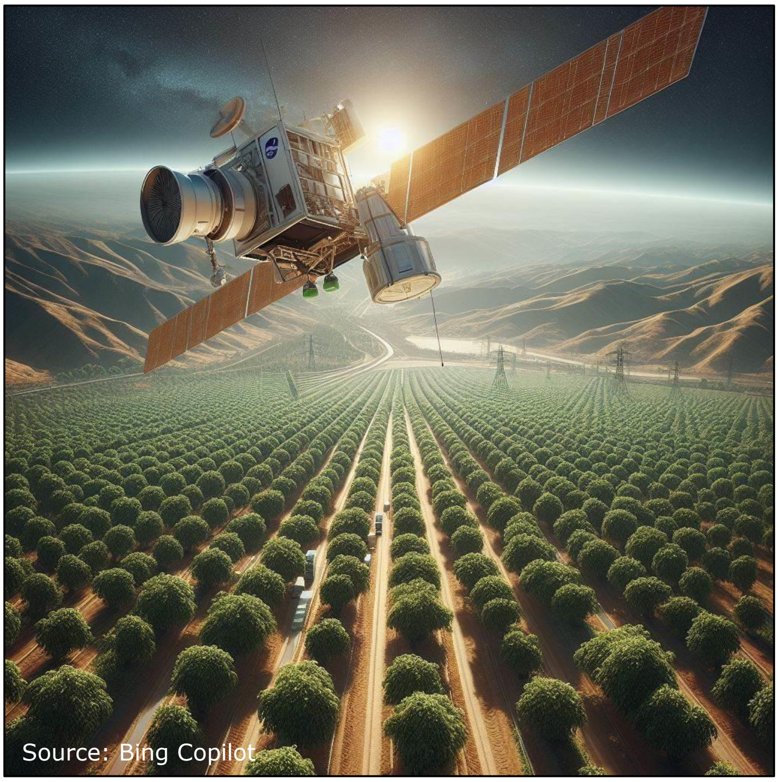
Geo-information Science and Remote Sensing
Thesis Report GIRS-2024-12

Texture analysis for detecting avocado orchards in Uruapan, Mexico

Dennis Mallens

Supervisors: Frida Ruiz Mendoza & Aldo Bergsma



Source: Bing Copilot

GRS-80436

Wageningen University & Research

Laboratory of Geo-Information Science and Remote Sensing

Texture analysis for detecting avocado orchards in Uruapan, Mexico

Dennis Mallens

Registration number: 1038987

Supervisors:

Frida Ruiz Mendoza

Aldo Bergsma

A thesis submitted in partial fulfilment of the degree of Master of Science

at Wageningen University and Research,

The Netherlands.

11-03-2024

Wageningen, The Netherlands

Thesis code number: GRS-80436
Thesis Report: GIRS-2024 -12
Wageningen University and Research
Laboratory of Geo-Information Science and Remote Sensing

Acknowledgements

While finishing this paper was not always as simple as I sometimes thought it to be, I still am grateful for all the people around me that motivated me to keep going. Firstly, I want to thank my supervisors Frida Ruiz Mendoza and Aldo Bergsma for all their help and advice in completing the MSc thesis. They came with a great subject and gave me the freedom to approach it my way and doing so, I learned a lot on the process of GRS research.

Secondly, I want to thank the friends I had in the thesis room. The time we had helping, laughing and drinking infinite amounts of coffee helped me stay motivated and made, the for some people hard and difficult period, a time I can look back to and be happy that I made the right decisions in time.

During writing this thesis I learned that the sector of GIS and remote sensing is continuously expanding and new discoveries are made every day. Undoubtedly, I am not done learning about the world of GRS, this was just the beginning.

Abstract

This study focuses on addressing ecological challenges located in Michoacán's Trans-Mexican Volcanic Belt due to the expansion of avocado orchards, which is a risk to local wildlife, water resources, and forest fragmentation, but in the meantime has economic benefits to the area.

Applying machine learning models for land use classification, the research explores the potential improvement through texture analysis, specifically applying Gray-Level Co-occurrence Matrix (GLCM) within the municipality of Uruapan, Michoacán, known for significant avocado cultivation and biodiversity. The investigation encompasses varying resolutions, from high-resolution Planetscope imagery of 3 meters to lower-resolution Sentinel-2 and Landsat 8 data varying from 10 to 60 meters.

In the first research question, Random Forest (RF) and Support Vector Machine (SVM) models are compared for avocado orchard localization using Planetscope data. The RF model has the greater accuracy, achieving an F-score of 0.9545 compared to SVM's 0.9481. Following the application of GLCM texture features provides minimal improvement in RF predictions with a F-score of 0.9551 but adversely affects SVM in the spatial prediction, even with an F-score of 0.9453. Lastly, the study applies the Sentinel-2 and Landsat 8 data with coherent GLCM textures to the machine learning models which shows a decreased F-score and spatial predictions.

Following these results, several recommendations are discussed for alternative accuracy metrics, feature extraction methods, and the application of deep learning with pre-trained models. To finally conclude that with the current results the GLCM texture features did not improve the overall accuracy. However, can be of potential use when accounting for the recommended model improvements.

Table of Contents

Introduction

There is a significant problem in the Michoacán region of the Trans-Mexican Volcanic Belt, where a high rate of forest land is being transformed into avocado orchards. (Mas et al., 2017) This poses an additional threat to the local wildlife in the temperate forests. As a result of this conversion process, the temperate forests are shrinking, and the remaining parts are becoming isolated from each other. This widespread occurrence primarily affects the wetter slopes of the Trans-Mexican Volcanic Belt region, situated at elevations between 1,500 and 2,000 meters. As of 2005, more than 22,270 hectares of forest had already been converted into avocado orchards. (Monterrubio-Rico et al., 2019) The transformed orchards are mostly Hass avocado trees. This use of homogeneity in avocado production brings along an increase in the use of pesticides, finally the increasing amounts of orchards puts pressure on the water reserves which increasingly being used. (De La Vega-Rivera & Merino-Pérez, 2021)

Nonetheless, the avocado orchards do stimulate the economy and boost technological innovations to efficiently improve the cultivation of avocados. (Vargas-Canales et al., 2020) These two contradictory visions on the cultivation of avocados, makes it therefore important to localize with high accuracy these type of orchards as the exact locations are currently unkown.

Land use classification in this region can help understand the impact size of avocado orchards. Multiple studies have already been conducted to detect land use change in the Michoacán region. The study of Latorre-Cardenas et al. (2023) used supervised classification on Sentinel-2 imagery to identify land use within the avocado belt to help determine temperate forest fragmentation. With an accuracy of 68.4%. Guerrero et al. (2008) use a multi-layered land use model called GEOMOD, from physical to social layers, to determine change in land use. GEOMOD is a model for land-use and land-cover change that operates on a grid system and can simulate the spatial progression or regression of land changes over time (Pontius, 2006) and determine driver for deforestation. Denvir (2023) uses the EGO 5 model to assess forest loss due to avocado orchards under various climate change scenarios. As well as the research of Arima et al. (2022) where a probit model is used to determine future deforestation due to an increasing amount of avocado orchards and how the spatial distribution of avocado cultivation is changing in response to both present and future climate conditions. Applying machine learning (ML) models to land use classification is done regularly and give more promising results than others. Talukdar et al. (2020) stated in a literature review that Support Vector Machine (SVM), Random Forest (RF) models gave the best results in machine learning land use classification. The SVM aims to determine the optimal class separation or optimal hyperplane to separate the training data in pre-defined classes. It does this by identifying the support vectors, which is training data that is closest to the optimal decision boundary. Whereas, a RF model uses a collection of classifier models that work together to form a classification system. This collection of classifiers are used as a voting system for data to be placed in a pre-defined class by the majority vote. (Sheykhmousa et al., 2020)

However, can these models be further improved? For this the thesis will go into the texture domain. In general, a texture can be informally defined as a set of texture elements which occurs in regular or repeated pattern. (Hung, C et al., 2019) Texture features describe the characteristics of image textures which is expressed as a function of the spatial variability of pixel intensities that are measured in grayscale. (Tüceryan & Jain, 1993) Examples of these grayscale patterns are seen in figure 1 of several land uses within Uruapan where clear distinction of texture is seen. Texture operators such as grey level co-occurrence matrix (GLCM), local binary pattern (LBP) and texture spectrum (TS) are used to extract the texture features which contain the texture information. (Hung, C et al., 2019) Therefore, Erener and Duzgun, (2009) already suggest the application of texture layers on aerial bands to possibly come up with an improved model. Additionally, in the study of Puissant et al., (2005) is concluded that the spatial resolution does effect the accuracy of texture analysis on land use classification. They state, as expected, that higher resolution improve the land use classification using texture analysis. However, they used resolutions of 1, 2.5 and 5-10 meters which was analysed in 2006. In this research, a current image in combination with ML algorithms using the spectral information of the RS images is needed to assess the avocado orchard impact. Therefore, Planetscope is used with a resolution of 3 meters, but also open-source datasets like Sentinel-2 and Landsat 8 with a resolution of 10 and 30 meters respectively to assess if texture analysis on lower resolution images is still viable.

In short, this study is going to assess if additional texture analysis can assist in improving the overall land use classification accuracy on multiple resolutions and thereby ameliorate avocado orchard detection.

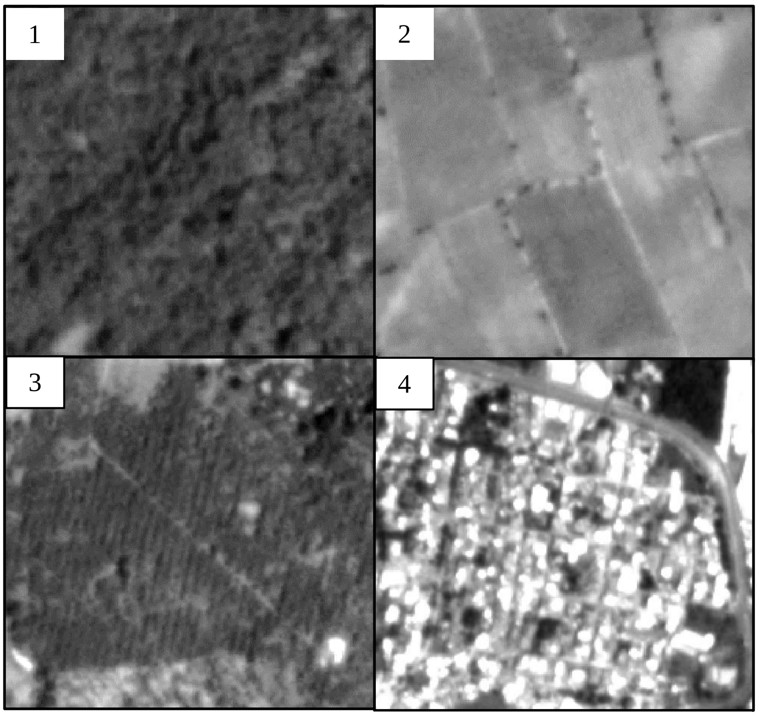


Figure 1. Grayscale textures of 1. forest, 2. farmland, 3. avocado orchard and 4. urban area within the municipality of Uruapan.

1. Research needs

The application of texture analysis on land use classification gave promising results for in the paper of Erener en Duzgun, (2009) They engage in the classification of historical black and white aerial photograph records, making them suitable for use in change detection studies. The use of only GLCM on black and white aerial images with an land use classification accuracy of 86% (Erener & Duzgun, 2009). Ciriza et al., (2017) used GLCM texture analysis on detecting uprooted orchards with an accuracy of 88%, suggesting that the identification technique could also be applied on other permanent crops due to their plantation pattern.

As suggested in the introduction, higher classification accuracy could help visualize avocado orchards impact. The study of Cho, et al. (2021) and Latorre-Cardenas et al. (2023) are the only papers that use Machine Learning to locate possible avocado orchards within the Michoacán region. Cho, et al. (2021) only uses six multi-layered 30 meter resolution Landsat images to get an avocado orchard location prediction, while Latorre-Cardenas et al. (2023) used nine bands with a resolution of 10 and 20 meters from Sentinel-2 imagery. These two studies have in common that they only use bands provided by the satellite of choice with resolutions between the 10, 20 and 30 meters. (U.S. Geological survey, 2018)(Sentinel Online, n.d.) Meanwhile, Machine Learning models do not have a limit in input variables (Nikparvar & Thill, 2021) and no research has been conducted on the localization of avocado orchards in Michoacán combining these with texture analysis. Which for this paper is the aim to overcome this research gap.

2. Research questions

2.1. Main objective

The main objective of this thesis is to find out if GLCM texture analysis in addition with satellite band combinations implemented in a machine learning model can improve the accuracy of avocado orchard localization in Uruapan, Mexico with varied satellite imagery resolutions.

2.2. Main research question

How can texture analysis improve machine learning models in detecting avocado orchards in Uruapan, Mexico?

2.3. Sub research questions

RQ1: Which machine learning algorithm, either Random Forest (RF) or Support Vector Machines (SVM), yields the most accurate results in detecting avocado orchards when applied to Planetscope's high-resolution images?

RQ2: What is the contribution of the GLCM texture analysis method for the accurate identification of avocado orchards in Uruapan, Mexico, using the machine learning models that have been previously employed in RQ1?

RQ3: Can GLCM maintain its accuracy and reliability while working with lower-resolution (Sentinel-2 & Landsat 8/9) images as input, and how might its performance be influenced by the change in spatial resolution?

3. Methodology

This section will cover the method and materials used to investigate every research question. The research will build up on each other per research question. In the first research question a focus was laid upon the current machine learning models and their accuracies, whilst SVM and RF models are created and assessed. For the second research question the machine learning models created are extended with additional GLCM layers where the accuracy of the different models of RQ1 and RQ2 are assessed. For the final research question, the models of the second research question are reused with Landsat 8 and Sentinel-2 satellite imagery to determine if the GLCM with lower spatial resolution is also effective in detecting avocado orchards. Finally, the combination of results gives us a clear understanding of the main research question.

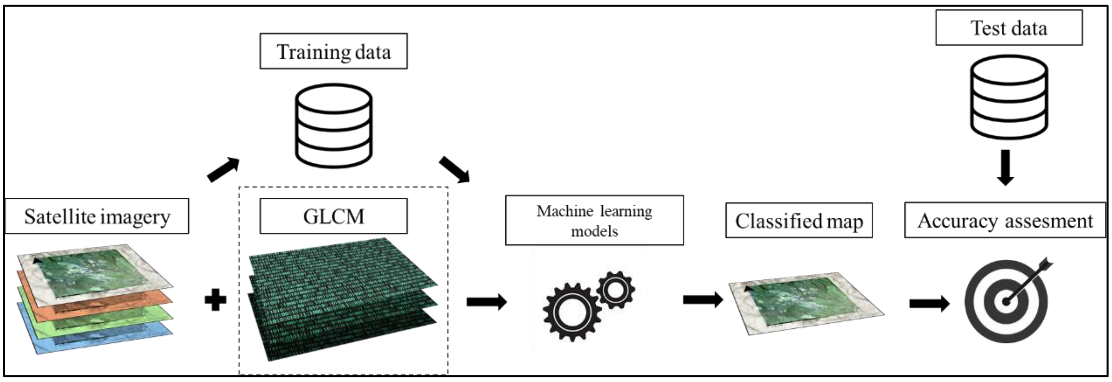


Figure 2. Flowchart of the thesis methodology.

3.1. *Study area*

Planetscope data has a limit of 5000 km^2 per month of high resolution satellite imagery when used for research purposes. Therefore, within the 'avocado belt' I selected the area between the Pico de Tancítaro and the city of Uruapan. The study area ranges from 19.33 N to 19.52 N latitude and 102.02 W to 102.32 W longitude (figure 3). The interest of this area is due to the fact that 20% (16,588 hectares) of the total area of the municipality of Uruapan cultivate avocados since 2014 due to its ideal climatic conditions and provides plenty of job opportunities (Vidales & Ortíz, 2014) and 63% is covered by forest in 2010 (Global Forest Watch, 2022). Meanwhile, the Pico de Tancítaro is a 'Región Terrestre Prioritaria' (Priority Land Region) due to its rich biodiversity and ecosystem. (Medina-García et al., 2020) (Conabio, n.d.)

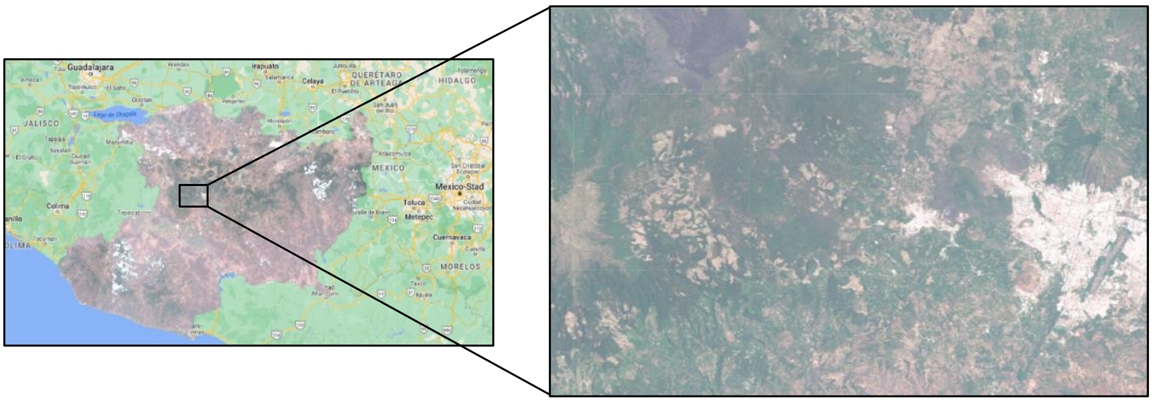


Figure 3. Left: satellite image of Michoacan, Mexico. Right: satellite image of the study area, with on the left Pico de Tancítaro and on the right the city of Uruapan. (Google Earth Engine)

3.2. *Data collection and pre-processing*

To cover the study area, satellite imagery from Planetscope's DOVE cubesat was used. Planetscope is paid service or gives out academic licenses that provides 5000 km^2/month, and has a better resolution than open source imagery from Sentinel-2 and Landsat 8, which will be used in the final research question to determine their relevancy in this study. An overview of the classifications of the satellites used can be found in table 1.

Table 1. Satellite types given resolution, bands and library.

Satellite type	Resolution	Bands	Image library	Source
Planetscope (DOVE)	3 meters	Coastal Blue: 431 - 452 nm Blue: 465 – 515 nm Green I: 513 - 549 nm Green: 547 – 583 nm Yellow: 600 - 620 nm Red: 650 – 680 nm Red Edge: 697 – 713 nm NIR: 845 – 885 nm	Planet Explorer	European Space Agency (2022)
Sentinel-2	10 meters 20 meters* 60 meters**	Coastal Blue**: 443 nm Blue: 490 nm Green: 560 nm Red: 665 nm Red edge 1*: 705 nm Red edge 2*: 740 nm Red edge 3*: 783 nm NIR: 842 nm	Copernicus	ESA (n.d.)
Landsat 8	30 meters	Coastal Blue: 430-450 nm Blue: 450 – 510 nm Green: 530 – 590 nm Red: 640 – 670 nm NIR: 850 – 880 nm SWIR 1: 1570 – 1650 nm SWIR 2: 2110 – 2290 nm	USGS Earth Explorer	NASA (2021)

The satellite imagery have requirements to give the best accuracy and analysis usage: The satellite imagery that is used has little to no cloud coverage (<5%), which can be setup within the satellite portal. The dates used are as recent as possible and the difference between the images from the three different platforms are within a range of 3 months. The imagery is obtained from their satellite image library, which can be found in table 1. To acquire the study area, a bounding box of the area is created, whereafter the exact study area is clipped in ArcGIS to the municipality of Uruapan.

For the supervised machine learning models, which will be discussed in the next section, training data is needed for the machine learning models to run and test data to determine the accuracy of the final classification. Obtaining this data is done by creating random selected areas of interest (AOI) within the satellite imagery and assigning a class to it via visual observation. Visual observation however can create human error and most importantly is a tedious task to do. (Abdi, 2019) The AOI is categorised as: 'orchard' and 'other'. It is important that a sufficient amount of training- and test data is obtained, since this will have an impact on the overall accuracy of the predicted model. (Ouma et al., 2023) Therefore, a total of 122 AOI's were obtained and classed, then a 70%-30% training-test data ratio be held into account and will be randomly selected. These AOI are in Arcgis used to extract the feature layers to be used in the machine learning models.

Finally, for the accuracy assessment, the data from INEGI is used. INEGI is autonomous public organization that focuses on the regulation and coordinating statistical and geographical information of Mexico. (De Estadística Y Geografía Inegi, n.d.) The percentage of avocado cultivation sown per hectare within the municipality of Uruapan is obtained from their database. In addition to this, land use and vegetation vector data set from 2016 of INEGI (Inegi, 2016) was used (figure 4.) to give an indication of potential avocado orchard locations. Since avocado trees are classed as permanent crop (*Glossary:Permanent Crops*, 2023), the INEGI data represents it as permanent cycle rainfed agriculture and permanent annual cycle rainfed agriculture.

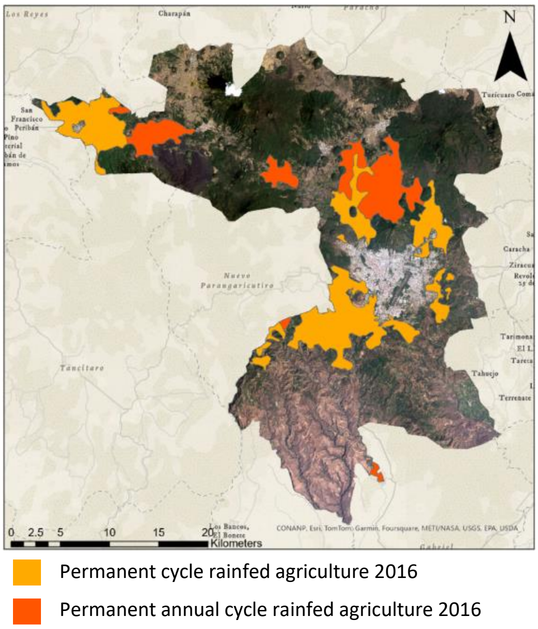


Figure 4. Possible avocado orchard locations in the municipality of Uruapan.

3.3. Machine learning models

As already mentioned, there are several studies that cover land use classification using machine learning. In this research several of these studies will be analysed in context of model usage, classification and accuracy assessment. Finally, some of these machine learning models will be applied on the study area. In this thesis two supervised machine learning models will be used: random forest model (RF) and the support vector machine (SVM). Both ML models will be created in Python.

3.3.1. Random Forest

The first ML model is the 'Random forest'. The RF model uses a collection of regression and classification trees. RF is a supervised model, thus needs training data to train the model otherwise called 'in-bag samples' (Belgiu & Drăguţ, 2016) and can have as many predictor variables to use as possible input.

3.3.2. Support Vector Machine

The second ML is the 'Support Vector Machine'. Just like the RF, the SVM is a supervised model that aims to find a distinction via the pre-defined classes (which are set via the training data) via a hyperplane. The optimal hyperplane refers to the separation of classes with the least amount of misclassifications. (Mountrakis et al., 2011) Scaling is performed on the feature training- and test set to standardize the features by removing the mean and scaling to unit variance. This preprocessing step make certain that the features have comparable scales which can help to improve performance and computation speed. (Testas, 2023)

3.3.3. Parameter selection

Obtaining the fit for the machine learning models is via a five-fold cross validation to optimize the models hyperparameters to acquire the optimal predication accuracy. For the RF model the following parameters are used:

- n_estimators: 50, 100, 200
- max_depth: ‘None’, 10, 20
- min_samples_split: 2, 5, 10
- min_samples_leaf: 1, 2, 4

For the SVM model the following parameters are used:

- C: 0.1, 1, 10
- gamma: 'scale', 'auto', 0.1, 0.01
- kernel: 'linear', 'rbf' (Radial Basis Function), 'poly' (polynomial)
- degree: 2, 3, 4 (Only in case of a polynomial kernel)

After analysing, the best hyperparameters are used to fit the model.

3.3.4. Confusion matrix

To assess the accuracy of the classified maps by the machine learning models, a confusion matrix is used. A confusion matrix provides a summarized overview of the prediction distribution across all classes. Within the confusion matrix, the predicted pixel of the test set will be divided in the True-Positive (TP), True-Negative (TN), False-Positive (FP) and False-Negative (FN) which will be used to calculate the precision, recall and F-score metrics: (Heydarian et al., 2022)

$$Precision = \frac{TP}{TP+FP} \qquad (1)$$

$$Recall = \frac{TP}{TP+FN} \qquad (2)$$

$$Fscore = 2 \cdot \frac{precision \cdot recall}{precision + recall} \qquad (3)$$

The F-score returns high values if the precision and recall give high values as well, a harmonic mean. (Kozák et al., 2022) Thus, the F-score will give a proper insight into the accuracy of the classified map. The confusion matrix metrics will be applied to every machine learning model created in this thesis. The results of the above mentioned metrics will be compared within every sub research question, giving us with improved or reduced accuracies the conclusion for the accuracy assessment.

Since only the test set of the model is used to determine the accuracy of the model, it is also important to know if the total predicted avocado orchards area in the whole study extent is accurate. To do so, the area sown for avocado production of Uruapan are collected via INEGI and the percentage over the whole study area are determined. By dividing the total sown avocado orchards over the total Uruapan municipality area, the percentage of the area meant for avocado orchards is obtained. However, since this data is gathered from a table, no spatial accuracy can be drawn from this. Then, by obtaining the ML prediction, the pixels that are classed as avocado orchard are divided by the total area in pixels. This gives the percentage of predicted avocado orchards and so by comparing this with the actual avocado orchards percentage an accuracy conclusion can be drawn of the machine learning model.

3.4. GLCM texture analysis

The second research question is about the usage of grey level co-occurrence matrix and defining the contribution of the GLCM texture analysis for classifying avocado orchards. For this process the GLCM statistical measures are applied on the Planetscope, Sentinel-2 and Landsat 8 satellite imagery.

GLCM functions analyse the texture of a grayscale image by determining the frequency of occurrence of pixel pairs with specific values and in defined spatial arrangements within the image. This process involves generating a GLCM and subsequently deriving statistical measures from the resulting matrix. (MathWorks, n.d.) The texture operator GLCM has a variety of statistical measures that are performed within the defined window size, which is due to computational power set on 3, and an angle of 45 degrees (Nizalapur & Vyas, 2020):

1. Homogeneity of the pixels within the window size
2. Contrast is the rate of the variation between pixels within the window size
3. Dissimilarity is defined by the difference between the absolute values of the grayscale
4. Entropy is defined by the irregularity between the pixels within the window size
5. Angular Second Moment (ASM) signifies how the grey level in the image show the regularity of distribution
6. Correlation quantifies the probability of occurrence for the specified pairs of pixels simultaneously.

(O'Byrne et al., 2012) (Gaudêncio et al., 2022)

Ultimately, the importance per statistical measure is obtained. This process will be accomplished by feature importance. Since the RF- and SVM model are used, I need two separate methods to get the optimal feature set. The RF model uses in this case the 'Mean Decrease of Impurity' (MDI). MDI calculates the cumulative reduction in loss or impurity resulting from all splits associated with a specific feature based on their Gini-index value. (Xiao et al., 2019) For the SVM model, the permutation-based feature importance is used. The permutation importance assesses the importance of a used feature, using F and systematically shuffling the values. The permutation disrupts the link between the F and the feature and thereby altering the predictive performance of the model. A notable impact on the prediction performance occurs if F holds a substantial importance. The degree of change in the prediction accuracy helps to identify the features that contribute the most in the model. (Oh, 2022) Comparing both optimal selected features will give us an answer of the most useful GLCM statistical measure.

3.5. Resolution Assessment

For the final sub-research question, the ML model, previously made for the GLCM analysis, will be used to evaluate the practicability of GLCM on various open-source satellite imagery. As already mentioned, for this research Sentinel-2 and Landsat 8 will be used in combination with the GLCM features of these satellites. The model input will be the band combination showed in table 1 respectively to their resolution. Finally, a confusion matrix will be created for each satellite usage, where the prediction, recall and F-score metrics are calculated from. These metrics are compared with each other, giving a conclusion to the final sub-research question.

4. Results

What machine learning model for land use classification, when applied to Planetscope's high-resolution images, yields the most accurate results in detecting avocado orchards?

First, the satellite data from Planetscope is obtained with the date 13-03-2023, of a series of images put into a mosaic. The imagery put into the RF and SVM models gave the following results.

- *Random forest model*

From the parameter selection the following parameters were selected: 'max_depth': None, 'min_samples_leaf': 1, 'min_samples_split': 2 and 'n_estimators': 200. The Random Forest model gave with the usage of the MDI a ranking where the NIR band has the highest importance of approximately 0.37 to predict the avocado orchards. Followed by green with a feature importance of around 0.12. After that the sequence follows down with blue, red edge, red, coastal blue, yellow and green 1 which have a feature importance close to each other between 0.6 and 0.11.

The prediction given in the 30% test set of the study area created an confusion matrix with the accessory precision and recall variables. TP is 79,042, FP is 3,161 and FN is 4,372 (figure 5) giving the following accuracy metrics:

$$Precision = \frac{79042}{79042 + 3161} = 0.962$$

$$Recall = \frac{79042}{79042 + 4372} = 0.948$$

$$Fscore = \ 2 \cdot \frac{0.962 \cdot 0.946}{0.962 + 0.946} = 0.9545$$

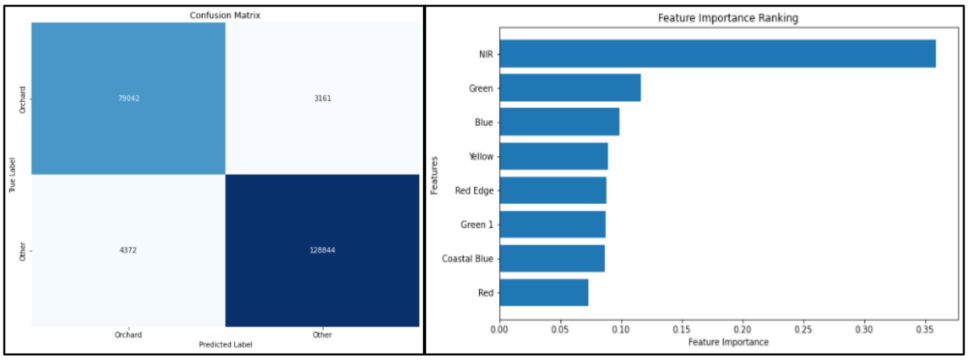


Figure 5. Confusion matrix and Feature importance of Planetscope RF model.

- *Support vector machine*

From the parameter selection the following parameters were selected: 'C': 10, 'gamma': scale and 'kernel': 'rbf'. The Support Vector Machine has after feature permutation a more gradually decrease in feature importance where the yellow feature has the most importance, followed by NIR, blue, green, red, green 1, coastal blue and red edge. The prediction given of the test set gives an TP of 78,829, FP of 3,374 and a FN of 5,248 as seen in the confusion matrix. (figure 6) Giving the accuracy metrics:

$$Precision = \frac{78829}{78829 + 3374} = 0.959$$

$$Recall = \frac{78829}{78829 + 5248} = 0.938$$

$$Fscore = \ 2 \cdot \frac{0.959 \cdot 0.938}{0.959 + 0.938} = 0.9481$$

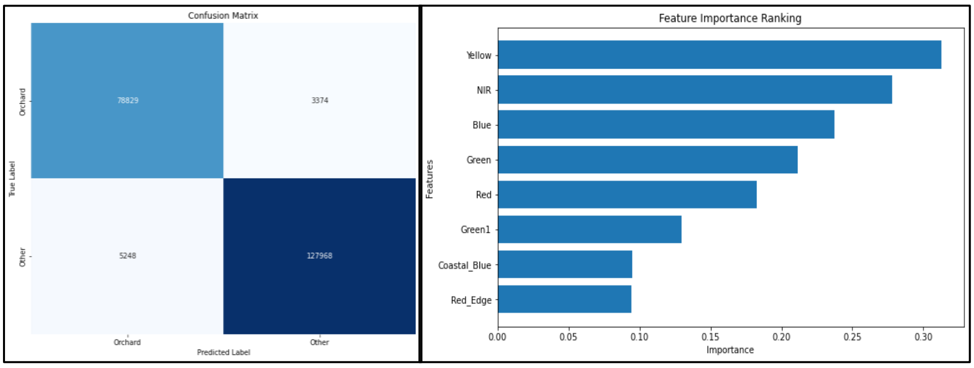


Figure 6. Confusion matrix and Feature importance of Planetscope SVM model.

The total agricultural area is determined to be $\frac{23604.77}{101290.00} \cdot 100 = 23\%$ (Inegi, 2020). Where 23,604.77 represents the hectares sown and 101,290.00 the total area of Uruapan in hectares. Meanwhile, only $\frac{17640}{101290.00} \cdot 100 = 17\%$ is avocado orchard in 2022. (Inegi, 2020) For these two models the total avocado orchard pixels in the RF and SVM models are 24,148,485 and 22,237,289 pixels respectively. Given the total pixel area of 112,687,097, this tells us that 21.4% for RF and 19.7% for SVM of the total area is predicted as avocado orchard. The spatial distribution from the RF and SVM give similar spatial disribution patterns that are coherent with the permanent cycle agriculture from 2016. (figure 7) The main faults that are observed are located in forests North of the study area. Meanwhile, some predictions are not within or close to the permanent cycle agricultural boundaries however, are observed to be orchards.

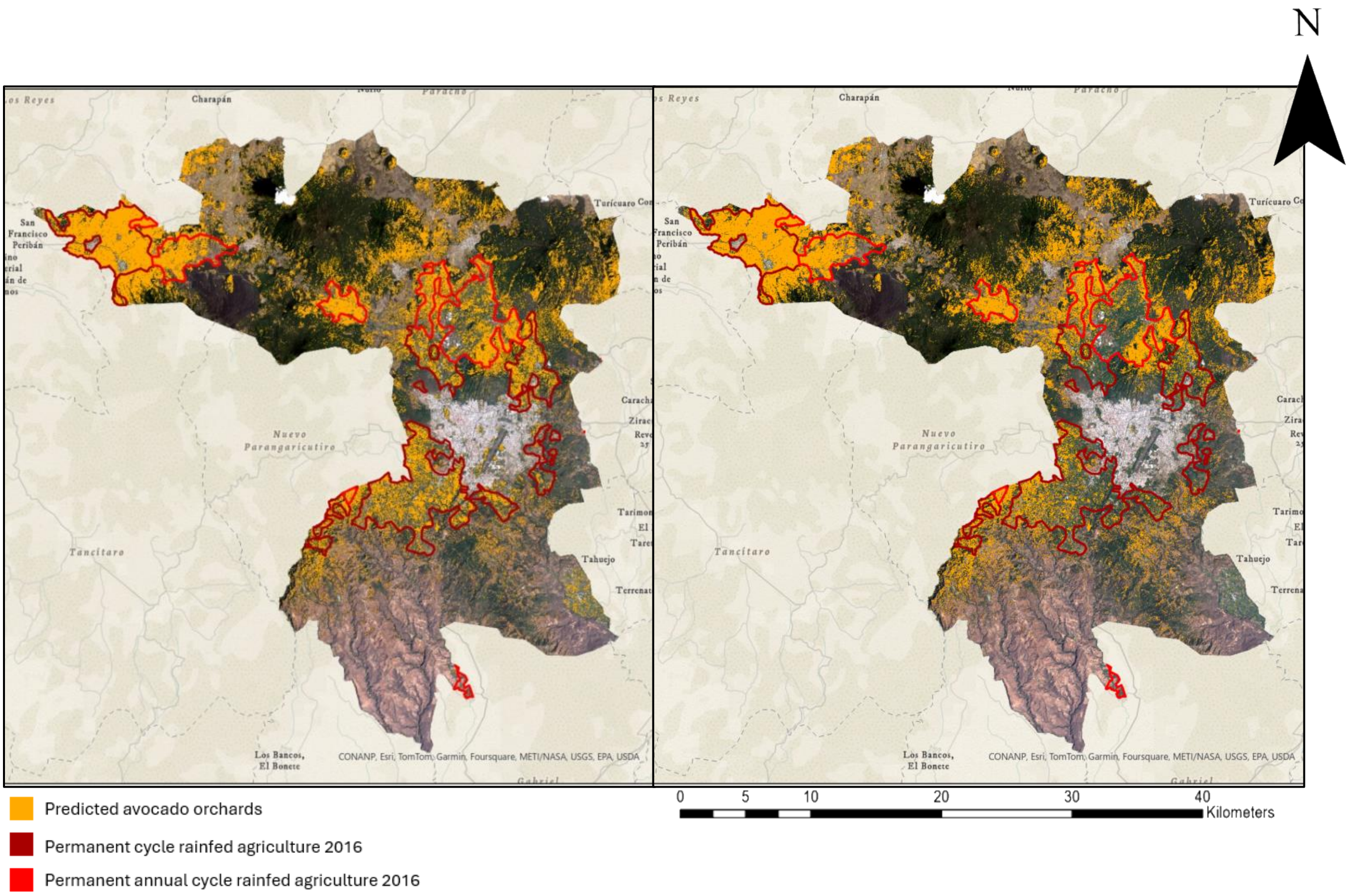


Figure 7. Planetscope avocado orchard prediction RF (left) and SVM (right).

What is the optimal spatial texture analysis method within GLCM for the accurate identification of avocado orchards in Uruapan, Mexico, using the same machine learning models that have been previously employed?

For this research question I added the GLCM textures to the feature layers and repeated the models. Where the results for the RF and SVM are drawn.

- *Random forest*

From the parameter selection the following parameters were selected: 'max_depth': None, 'min_samples_leaf': 1, 'min_samples_split': 2 and 'n_estimators': 200. The RF model with these parameters shows a similar pattern where the NIR band has the most influence with an importance of 0.33, whereas the other spectrum features retaliate between 0.07 and 0.11. The six GLCM textures do not have a higher importance than 0.01 The confusion matrix with a TP of 79,118, FP of 3,374 and a FN of 5,248 (figure 8) give the following accuracy assessment:

$$Precision = \frac{79118}{79118 + 3085} = 0.962$$

$$Recall = \frac{79118}{79118 + 4344} = 0.948$$

$$Fscore = \ 2 \cdot \frac{0.962 \cdot 0.948}{0.962 + 0.948} = 0.9551$$

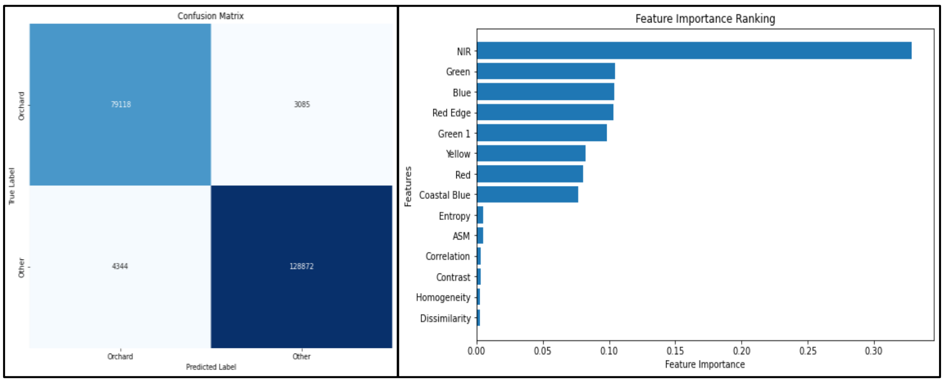


Figure 8. Confusion matrix and Feature importance of Planetscope + GLCM texture RF model.

- *Support Vector Machine*

From the parameter selection the following parameters were selected: 'C': 10, 'gamma': scale and 'kernel': 'rbf'. This SVM model tells us that also in this case the yellow and NIR feature have the highest importance with 0.27. Followed by the other Planetscope spectral band ranging between 0.09 and 0.21. Finally, the GLCM features are listed between 0.03 and 0.04, with ‘dissimilarity’ close to zero. The confusion matrix with a TP of 78,857, FP of 3,346 and a FN of 5,759 (figure 9) give the following accuracy assessment:

$$Precision = \frac{78857}{78857 + 3346} = 0.959$$

$$Recall = \frac{78857}{78857 + 5759} = 0.932$$

$$Fscore = \ 2 \cdot \frac{0.959 \cdot 0.932}{0.959 + 0.932} = 0.9453$$

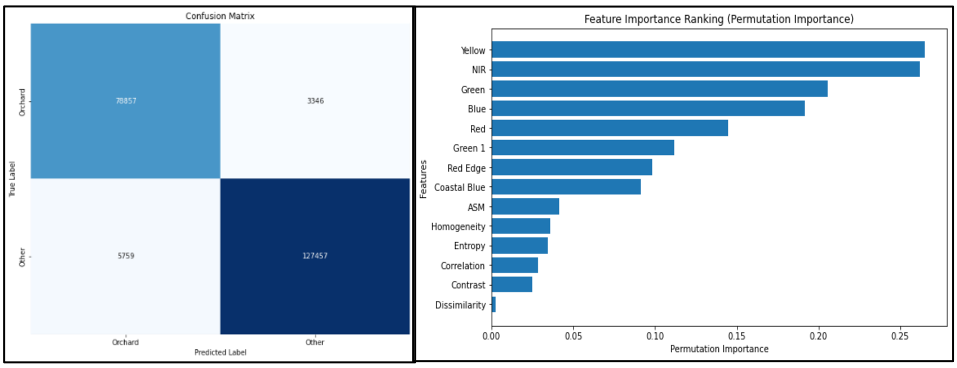


Figure 9. Confusion matrix and Feature importance of Planetscope + GLCM texture SVM model.

For the two models in combination with GLCM texture the total avocado orchard pixels in the RF and SVM models are 23,479,219 and 4,531,641 pixels respectively. Since the image size remained the same, by dividing the predicted pixels by the total area pixels which gives us 20.8% for RF and 4,0% for SVM of the total area that is predicted as avocado orchard. The locations of the predictions for the RF model show similar results as in figure 7, where they mostly are predicted in the expected areas in combination with some predictions in forests. The SVM model however shows an unusual low predicted area. (figure 10) Looking at the image reveals that the SVM models creates its predictions also greatly based on some of the GLCM features and thus the contours of the avocado orchards.

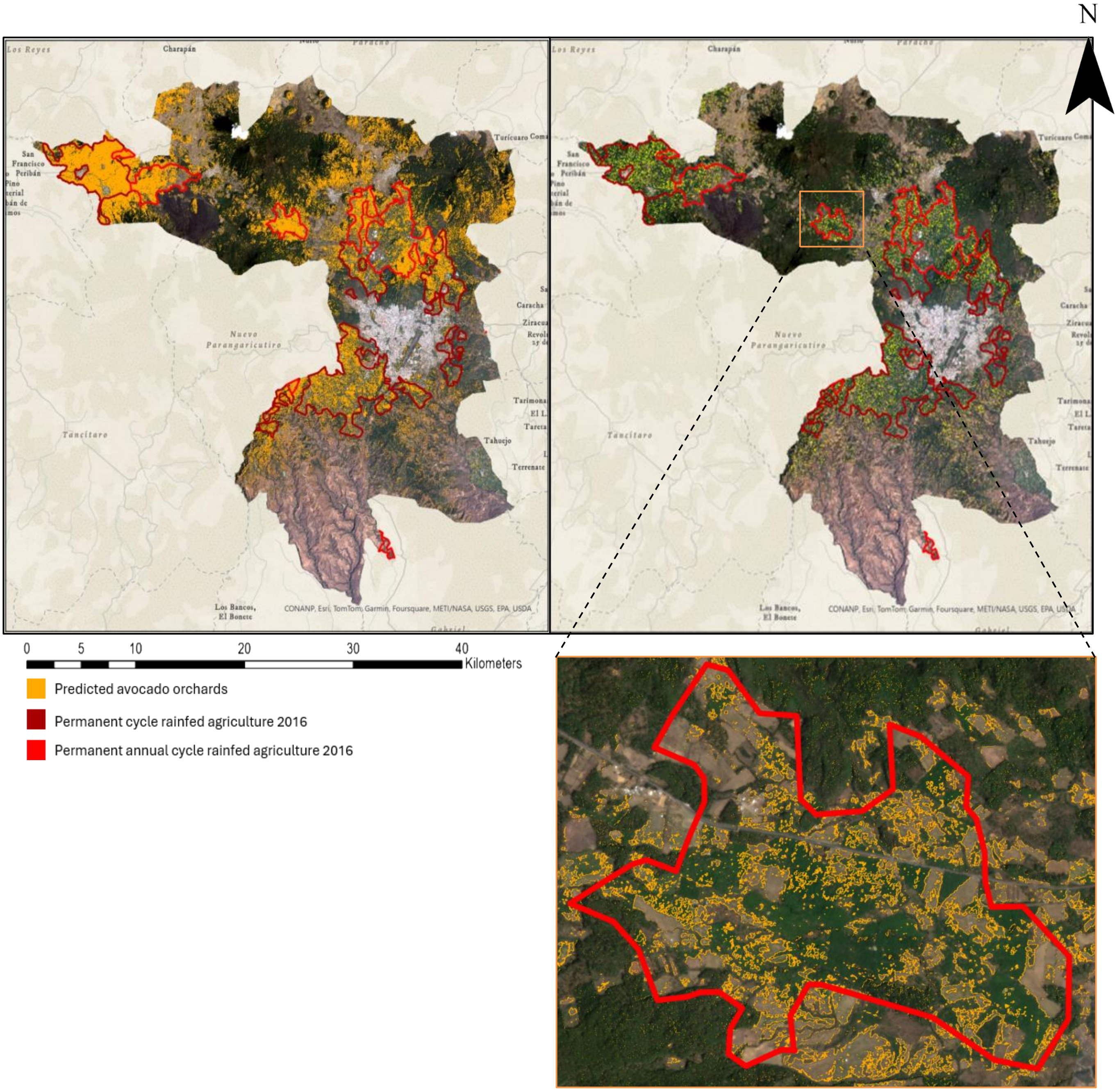


Figure 10. Planetscope + GLCM texture avocado orchard prediction RF (left) and SVM (right-up) with a zoomed in location (right-bottom).

Can GLCM maintain its accuracy and reliability while working with lower-resolution (Sentinel-2 & Landsat 8/9) images as input, and how might its performance be influenced by the change in resolution?

For the final research question four predictions were made. RF and SVM models for Sentinel 2 and Landsat 8 in combination with the GLCM textures. The satellite imagery was taken for Sentinel-2 is 29-04-2023 and Landsat 8 is 08-05-2023.

- *Random forest*
    - *Sentinel 2*

Firstly, the RF model for Sentinel 2 is created. From the parameter selection the following parameters were selected: 'max_depth': 20, 'min_samples_leaf': 2, 'min_samples_split': 2 and 'n_estimators': 200. This gave the feature importance and confusion matrix of the RF model for Sentinel 2. The ranking gives us the result that ‘Red edge 3’ has the most influence on the prediction with an importance of 0.175. Followed by other Sentinel 2 spectral bands between 0.13 and 0.07. Finally, the texture features are listed with an importance between 0.03 and 0.015. Next, the confusion matrix gave as results TP of 1,084, FP of 292 and a FN of 314 (figure 11) which give the following accuracy metrics:

$$Precision = \frac{1084}{1084 + 292} = 0.788$$

$$Recall = \frac{1084}{1084 + 314} = 0.775$$

$$Fscore = 2 \cdot \frac{0.788 \cdot 0.775}{0.788 + 0.775} = 0.781$$

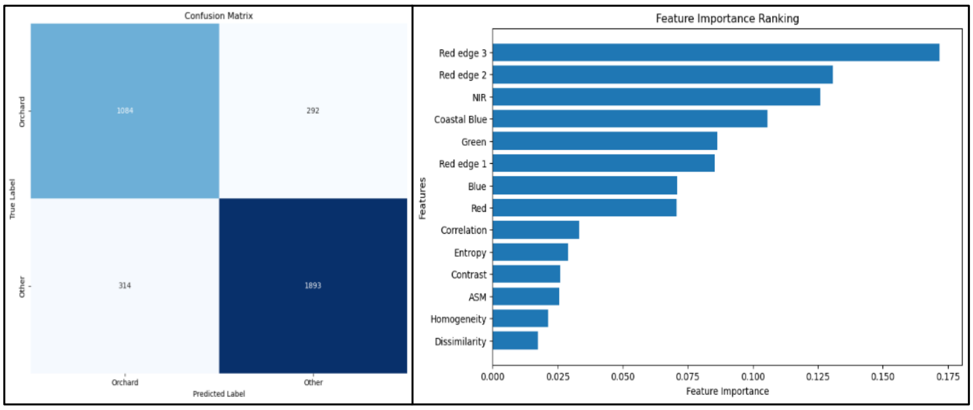


Figure 11. Confusion matrix and Feature importance of Sentinel-2 RF model.

-
    - *Landsat 8*

Secondly, the RF model for Landsat was formed. The parameter selection gave us the following result: 'max_depth': None, 'min_samples_leaf': 1, 'min_samples_split': 5 and 'n_estimators': 100. This gave us the feature importance and confusion matrix. Herein, the NIR feature has the highest importance with 0.34, followed by the other Landsat 8 spectral bands ranging from 0.06 to 0.15. At the bottom of the ranking the GLCM texture features are to be found ranging from 0.015 to 0.03. The confusion matrix has the TP of 732, FP of 111 and the FN of 103 (figure 12), giving the accuracy metrics:

$$Precision = \frac{732}{732 + 111} = 0.868$$

$$Recall = \frac{732}{732 + 103} = 0.877$$

$$Fscore = 2 \cdot \frac{0.868 \cdot 0.877}{0.868 + 0.877} = 0.872$$

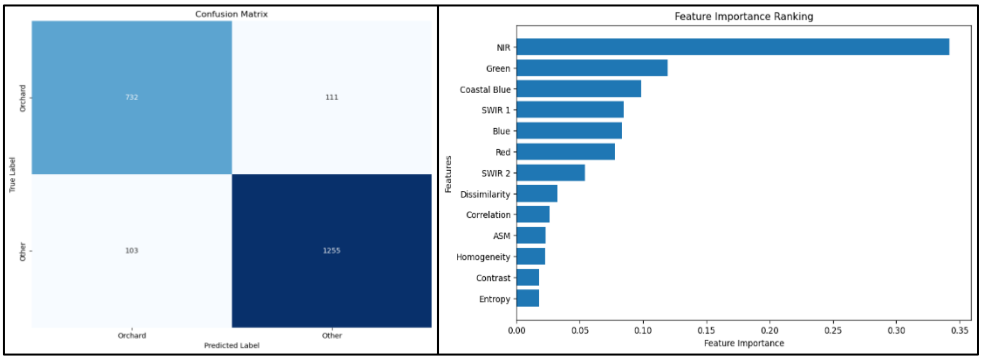


Figure 12. Confusion matrix and Feature importance of Landsat 8 RF model.

- *Support vector machine*
  - *Sentinel 2*

The second part of this research question is the SVM model. For the Sentinel 2 the parameter selection was 'C': 10, ‘degree’: 2, 'gamma': 0.1 and 'kernel': 'rbf'. This gave the feature importance with the green feature as most important with a value of 0.2. Followed by a more gradually importance decline with the Sentinel 2 spectral bands ranging between 0.06 to 0.175. After the spectral features, the texture features are listed with an importance between 0.015 and 0.05. The confusion matrix gave a result with a TP of 1,022, FP of 354 and a FN of 302 (figure 13), which give the following accuracy metrics:

$$Precision = \frac{1022}{1022 + 354} = 0.743$$

$$Recall = \frac{1022}{1022 + 302} = 0.772$$

$$Fscore = 2 \cdot \frac{0.743 \cdot 0.772}{0.743 + 0.772} = 0.757$$

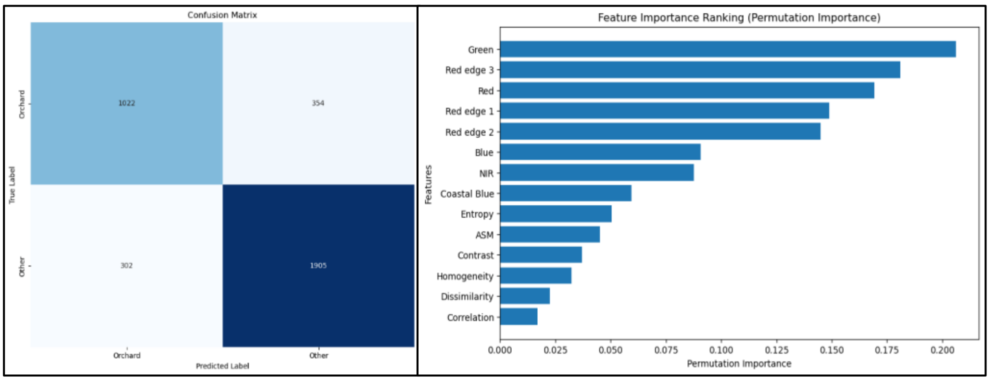


Figure 13. Confusion matrix and Feature importance of Sentinel-2 SVM model.

  - *Landsat 8*

Finally, the SVM model for the Landsat 8 is created and gave a result using the parameter selection of 'C': 10, ‘degree’: 2, 'gamma': 0.1 and 'kernel': 'rbf'. The feature importance of this model was given and gave as most important feature the NIR band with an importance of 0.3. After that, the other spectral features were presented, except blue, with an importance between 0.05 and 0.16. The blue feature was valued as 0.02. The texture layers are also in this case together between 0.01 and 0.04. The confusion matrix had a TP of 747, a FP of 96 and a FN of 102 (figure 14), which give the models accurate assessment:

$$Precision = \frac{747}{747 + 96} = 0.886$$

$$Recall = \frac{747}{747 + 102} = 0.880$$

$$Fscore = 2 \cdot \frac{0.886 \cdot 0.880}{0.886 + 0.880} = 0.883$$

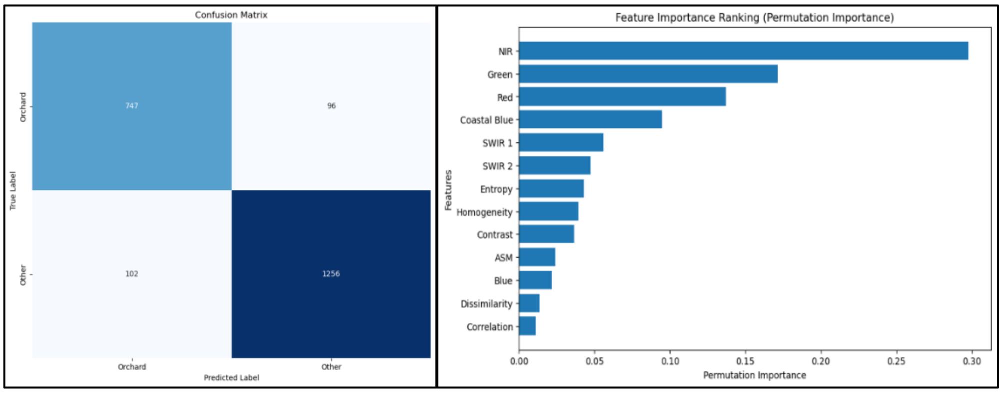


Figure 14. Confusion matrix and Feature importance of Landsat 8 SVM model.

At last the land coverage was
calculated for all four models. The total area for Sentinel 2 is 2,137,289 pixels and for Landsat 8 it is 1,127,467 pixels. The predicted pixels for every model are for Sentinel 2 RF: 647,021 and SVM: 590,530 and for Landsat 8 RF: 421,993 and SVM: 591,906. Using the same method as in the previous questions the total area predicted is 30.3%, 37,4%, 27,6% and 52,5% respectively. The spatial distribution for all predictions well placed within the permanent cycle agricultural borders. However, both satellite data have their own flaws when an prediction is made. The Sentinel-2 models have the city of Uruapan classified as avocado orchards, while the Landsat 8 models classify excessive amounts of land as avocado orchards in the South of the study area. (figure 15)

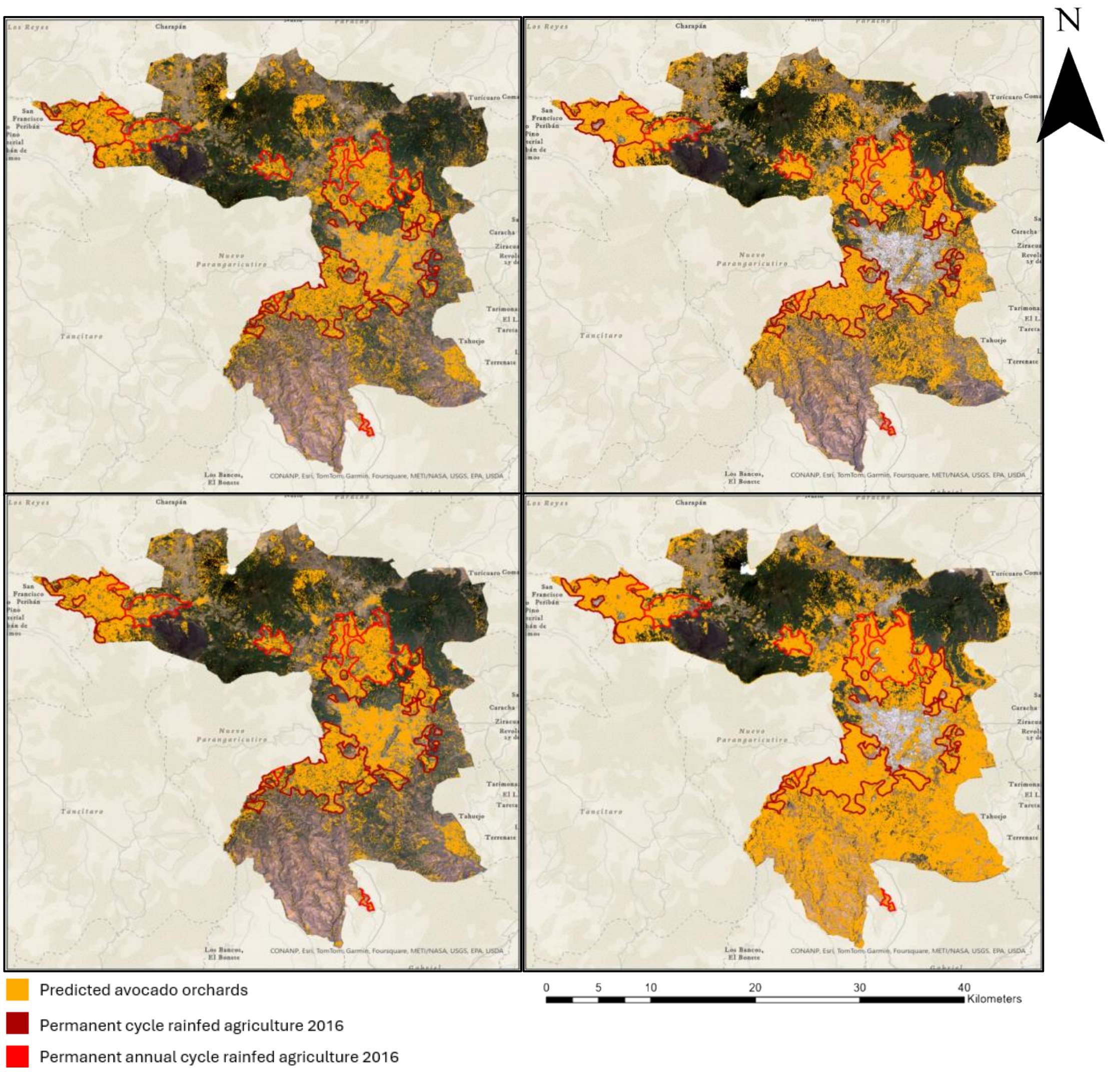


Figure 15.Avocado orchard predictions of a) Sentinel-2 + GLCM texture RF, b) Sentinel-2 + GLCM texture SVM, c) Landsat 8 + GLCM texture RF and d) Landsat 8 + GLCM texture SVM.

5. Discussion

What machine learning model for land use classification, when applied to Planetscope's high-resolution images, yields the most accurate results in detecting avocado orchards?

From the first results between the RF and SVM model for the Planetscope data, assessing the first accuracy metrics comes with a F-score of 0.9545 and 0.9481 respectively. So, this tells us that the RF model in this case has the upper hand in predicting the avocado orchards in the study area. The study of Talukdar et al. (2020) created a comparison between multiple ML models on land use classifications where they came to the conclusion that the RF model gives the highest accuracy. Additionally, in a more recent paper, the accuracy assessment for RF and SVM for land use classification is made and came out that RF is the superior model in this case. (Adugna et al., 2022) This is in line with the findings of the sub research question based on the F-score of the confusion matrices. Next to that the accuracy was also tested on the complete study area. Based on the statistical and geographical data of INEGI, 17% of the Uruapan municipality is used for avocado orchards. Meanwhile, RF predicts 21.4% and SVM predicts 19.7%. Lastly, the spatial predictions give the last insights in the performance of the models. From the prediction maps, the RF model shows more predictions within the INEGI vector, but have to take into account that this dataset is from 2016 and land use could be different in the satellite data used in this paper. Based on all these results the RF model would in this case yield the most accurate predictive results.

What is the optimal spatial texture analysis method within GLCM for the accurate identification of avocado orchards in Uruapan, Mexico, using the same machine learning models that have been previously employed?

After adding the GLCM texture feature layers it can be concluded that the RF model has the same F-score of 0.9551 as the model without the GLCM texture features. Surprisingly, when applying the models fit to predicting the percentage of avocado orchards of the entire study area, an improvement is seen where the new model predicts 20.8% of the total area as avocado orchards. Even though the feature importance on GLCM features was almost non-relevant it did improve the RF model. Ciriza et al., (2017) found that the total accuracy increases by 3.3% when adding GLCM features to satellite bands features. However, that is with an feature amount of 56 and after selection to only five features it increases by 9.2%. Additionally, the GLCM features justified in this study are correlation, contrast and entropy. Meanwhile, the RF feature importance shows that entropy as top contributor for the texture features. However, due to the limited importance and increase in F-score this does not say enough about the prediction improvement of the entire RF model. The SVM model showed a similar pattern with the GLCM features on the bottom with as top contributor ASM. The justified feature entropy is only on the third spot of the GLCM features. However, in the case of the SVM model, a slight decrease within the F-score is seen to 0.9453, compared to the SVM model of RQ1. Additionally, the spatial prediction showed a different predict method where is mainly focused its prediction on some of the GLCM features which explains the 4,0% of the total prediction. Showing only the contours of the fields makes it difficult to visually assess the accuracy in combination with the percentage. This makes its difficult also in future classifications to assess the prediction quality if only contours are sown. This only, seems to be an issue for the SVM model, but looking at the feature importances, it takes the GLCM more into account than the RF model. This says that the GLCM features negatively influence the localization of avocado orchards with these models. Nonetheless, both models exhibit comparable spatial predictions obtained from the land use data provided by INEGI.

Can GLCM maintain its accuracy and reliability while working with lower-resolution (Sentinel-2 & Landsat 8) images as input, and how might its performance be influenced by the change in resolution?

Finally, Sentinel-2 and Landsat 8 bands and GLCM textures were used as replacement for the Planetscope wavelets and textures. The results gave a RF F-score of 0.781 and 0.872 for Sentinel-2 and Landsat 8 respectively and a SVM F-score of 0.757 and 0.883 for Sentinel-2 and Landsat 8 respectively. Which concludes that the accuracy assessment using the F metrics says that open source data like Sentinel-2 and Landsat 8 does not maintain its accuracy when predicting avocado orchards when using GLCM features and thus perform worse than Planetscope's satellite imagery. This also counts for the area coverage prediction where the avocado orchards prediction is more than 10% to 35% higher than with Planetscope's imagery. As already mentioned in the introduction, Puissant et al., (2005) tells that higher resolution have a higher accuracy as is seen in this accuracy assessment. However, even though the F-score is a pleasantly high, the reliability of this prediction is doubtful. When looking at the maps in figure 15, the maps of the Sentinel-2 and Landsat 8 predictions show predictions that are definitely false. For example, Sentinel-2 predictions show that also the city of Uruapan is viewed as avocado orchards for both RF and SVM. Meanwhile, the Landsat 8 models are overpredicting with high amounts of area that is not avocado orchards. This is especially visible in the SVM model of Landsat 8. This brings me to the practical implications and limitations of this study.

Practical implications

Computational time

When comparing the RF and SVM models, I noticed during running the models that the computation time of SVM is way higher than that of the RF model. Example of this is predicting the whole Planetscope data around 100 hours, which was reduced to 12 hours using multiple cores from the high performance computer from the GRS faculty. This was also found in the research of Adugna et al., (2022). To find a trade-off between accuracy and time-efficiency, a radial basis function would be most optimal. (Jozdani et al., 2019) In the parameter selection this came forward in the lower resolution that 'rbf' was also the selected kernel to use. Due to time restrictions, the 'rfb' kernel is also used to obtain the SVM Planetscope fit. Reason for this is that the change in resolution causes the amount of pixels per feature layer to be 50 and 100 times smaller for Sentinel-2 and Landsat 8 respectively.

Regrettably, this was not enough for the model to be finished within a reasonable timeframe and therefore split the study area up into four quarters. The upper right quarter was used since this still had the most information to come up with a reasonable prediction as can be seen in figure 8a and c. This is less the case in the lower resolution predictions, perhaps due to the reduction of training pixels in an already limited amount of training pixels compared to Planetscope's model. The high amount of dimensions of the model in combination with a decrease/low amount of training pixels can lead to the curse of dimensionality. Ultimately, this can decrease the accuracy of the ML model. (Salimi et al., 2018)

Methodological reflection

Accuracy assessment

Reading the land use and land cover papers like Adugna et al., (2022), Erener, A., & Duzgun, S. (2009) and from the paper reviews of Talukdar et al. (2020), the kappa coefficient was mainly used as a accuracy metric. It is also seen as a standard, but statistically robust metric to test the accuracy in classification. It considers outcomes that may have occurred by random chance. (Ben-David, 2008) In a more recent review study, they looked at differences between Matthews Correlation Coefficient (MCC), Cohen's Kappa Coefficient and Brier Score. Herein, the MCC is seen as more informative than the Kappa coefficient and Brier Score, but only in binary classifications such as in this paper. (Chicco et al., 2021) Therefore, for future classifications, a more widely accuracy assessment metrics could be used to assess the models performance.

Feature extraction

To counter the high amount of dimensions/curse of dimensionality and computation time, a dimensionality reduction method could be applied to also improve accuracy and test if texture features are still reasonable the take into the model. Methods that could be applied are Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA), Independent Component Analysis (ICA) and Minimum Noise Fraction (MNF). (Moharram & Meena, 2022) In this paper the RF model was still reasonable to use. Nevertheless, for future research using high dimensionality and time consuming models, such as SVM, a feature extraction method could be used.

Texture analysis

In this thesis the use of machine learning models is utilized. Nonetheless, with the upcoming of deep-learning models, an option could be the use of conventional neural networks (CNN) with the use of pre-trained models such as ResNet50. (Dewangkoro & Arymurthy, 2021) Or with the use of Discrete Cosine Transforms (DCT) to enhance the fine-grained multi-scale land cover object identification for high resolution imagery, which could be used to detect land cover change (Zhu et al., 2023). These methods could be more modern practices and perhaps promising alternatives in detection of avocado orchards.

Significance of findings

Overall, I believe that the sub-research questions results give a swift overview of the influence of GLCM texture analysis, however more could be done to strengthen the main objective as explained in the methodological reflection. As for the predicted avocado orchard, only 76 pixels for RF and pixels 28 for SVM added to the TP when adding the GLCM texture features. Despite the limited change in the confusion matrix, this change reflects to 0.6% or 608 ha of improvement to the total predicted study area, despite not being spatially explicit. However, more research has to be done to say more about the significance of GLCM texture in discovering avocado orchards.

6. Conclusion

The primary aim of this thesis was to assess if GLCM texture analysis in combination with various satellite imagery can improve the detection of avocado orchards in Uruapan, Mexico. The study was structured around three sub-research questions to evaluate the possibility of GLCM texture analysis.

Initially, Planetscope's data was utilized to determine which machine learning model, RF or SVM, gave the superior accuracy in estimating avocado orchards. From the F-score, the RF model had a score of 0.9545 versus a score of 0.948 of that SVM. The spatial prediction gave 21.4% for RF and 19.7% for SVM, which is closer to the 17% of avocado orchards, but visually the RF model gave more promising results. In combination with the quicker running time of RF, to conclude that RF model has the upper hand of the two models. Afterwards, these models were used in combination with GLCM features: homogeneity, dissimilarity, ASM, entropy, contrast and correlation, to check the influence on the predictions. The F-score for RF is 0.9551 and for SVM 0.945. This also gave a limited improvement in the spatial prediction where the GLCM features negatively influenced the SVM model where the total area prediction is 4,0% in combination with the visual prediction of only contours of land. Ultimately, the feature importance tells us that entropy came highest from the GLCM features, also including the Sentinel-2 and Landsat 8 results. Nevertheless, the GLCM features are barely significant to the model with the current model set up. Finally, the models were used on open source data from Sentinel-2 and Landsat 8. The F-score came logically back as less accurate with RF 0.781 and 0.872 respectively and for SVM 0.757 and 0.883 respectively. Also with the addition of the area prediction which is more than 10% off, it is not reliably enough compared to Planetscope's data, but has potential if more is done in creating the ML model when assessing the prediction placement in the study area.

From these results and some literature reviewing, extra or different accuracy assessments could be used such as MCC or Kappa coefficient for the use of binary classification and to research the impact of GLCM texture analysis and reduce the computation time, a feature extraction like PCA, LDA, ICA or MNF could help realise this.

With the sub-research question covered and the discussion handled, the main research question can be reviewed: "How can texture analysis improve machine learning models in detecting avocado orchards in Uruapan, Mexico?" While the inclusion of GLCM features did not significantly improve the overall accuracy of the machine learning models for avocado orchard detection in Planetscope data, the analysis minimally suggests that texture information can potentially aid in orchard identification and more and/or other research has to be done to discover the competence of GLCM texture analysis with the recommendations given.

7. References

1. Abdi, A. M. (2019). Land cover and land use classification performance of machine learning algorithms in a boreal landscape using Sentinel-2 data. Giscience & Remote Sensing, 57(1), 1–20. https://doi.org/10.1080/15481603.2019.1650447
2. Adugna, T., Xu, W., & Fan, J. (2022). Comparison of Random Forest and Support Vector Machine Classifiers for Regional Land Cover Mapping Using Coarse Resolution FY-3C Images. Remote Sensing, 14(3), 574. https://doi.org/10.3390/rs14030574
3. Aksoy, S., Yalniz, I. Z., & Tasdemir, K. (2012). Automatic detection and segmentation of orchards using very high resolution imagery. IEEE Transactions on Geoscience and Remote Sensing, 50(8), 3117–3131. https://doi.org/10.1109/tgrs.2011.2180912
4. Arima, E., Denvir, A., Young, K. R., González-Rodríguez, A., & García-Oliva, F. (2022). Modelling avocado-driven deforestation in Michoacán, Mexico. Environmental Research Letters, 17(3), 034015. https://doi.org/10.1088/1748-9326/ac5419
5. Belgiu, M., & Drăguţ, L. (2016). Random Forest in Remote Sensing: A review of applications and future directions. Isprs Journal of Photogrammetry and Remote Sensing, 114, 24–31. https://doi.org/10.1016/j.isprsjprs.2016.01.011
6. Ben-David, A. (2008). Comparison of classification accuracy using Cohen's Weighted Kappa. Expert Systems With Applications, 34(2), 825–832. https://doi.org/10.1016/j.eswa.2006.10.022
7. Chicco, D., Warrens, M. J., & Jurman, G. (2021). The Matthews Correlation Coefficient (MCC) is More Informative Than Cohen's Kappa and Brier Score in Binary Classification Assessment. IEEE Access, 9, 78368–78381. https://doi.org/10.1109/access.2021.3084050
8. Cho, K., Goldstein, B., Gounaridis, D., & Newell, J. P. (2021). Where does your guacamole come from? Detecting deforestation associated with the export of avocados from Mexico to the United States. Journal of Environmental Management, 278, 111482. https://doi.org/10.1016/j.jenvman.2020.111482
9. Ciriza, R., Sola, I., Albizua, L., Álvarez-Mozos, J., & González-Audícana, M. (2017). Automatic detection of uprooted orchards based on orthophoto texture analysis. Remote Sensing, 9(5), 492. https://doi.org/10.3390/rs9050492
10. Conabio. (n.d.). Regiones terrestres prioritarias de México | Biodiversidad mexicana. Biodiversidad Mexicana. https://www.biodiversidad.gob.mx/pais/regiones-terrestres-prioritarias-de-mexico
11. De Estadística Y Geografía Inegi, I. N. (n.d.). Acerca del INEGI. https://www.inegi.org.mx/inegi/acercade.html
12. De La Vega-Rivera, A., & Merino-Pérez, L. (2021). Socio-Environmental impacts of the avocado boom in the Meseta Purépecha, Michoacán, Mexico. Sustainability, 13(13), 7247. https://doi.org/10.3390/su13137247
13. Denvir, A. (2023). Avocado expansion and the threat of forest loss in Michoacán, Mexico under climate change scenarios. Applied Geography, 151, 102856. https://doi.org/10.1016/j.apgeog.2022.102856
14. Dewangkoro, H. I., & Arymurthy, A. M. (2021). Land Use and Land Cover Classification Using CNN, SVM, and Channel Squeeze & Spatial Excitation Block. IOP Conference Series: Earth And Environmental Science, 704(1), 012048. https://doi.org/10.1088/1755-1315/704/1/012048

15. Erener, A., & Duzgun, S. (2009). A methodology for land use change detection of high resolution PAN images based on texture analysis. *Rivista italiana di telerilevamento*, 47–59. https://ocw.metu.edu.tr/mod/resource/view.php?id=2043
16. ESA. (n.d.). Spatial resolution. Sentinel Online. https://sentinels.copernicus.eu/web/sentinel/user-guides/sentinel-2-msi/resolutions/spatial
17. European Space Agency. (2022). PlanetScope. Earth Online. https://earth.esa.int/eogateway/missions/planetscope
18. Gaudêncio, A. S. F., Hilal, M., Cardoso, J. M. R., Humeau-Heurtier, A., & Vaz, P. G. (2022). Texture analysis using two-dimensional permutation entropy and amplitude-aware permutation entropy. Pattern Recognition Letters, 159, 150–156. https://doi.org/10.1016/j.patrec.2022.05.017
19. Global Forest Watch. (2022). Forest change in Uruapan, Michoacán, Mexico. Consulted on 4 October 2023, van https://www.globalforestwatch.org/dashboards/country/MEX/16/102/
20. Glossary:Permanent crops. (2023, 9 August). Eurostat. Consulted on 10 March 2024, of https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Glossary:Permanent_crops
21. Guerrero, G., Masera, O., & Mas, J. (2008). Land use / Land cover change dynamics in the Mexican Highlands: Current situation and long term scenarios. In Environmental science and engineering. https://doi.org/10.1007/978-3-540-68498-5_2
22. Heydarian, M., Doyle, T. E., & Samavi, R. (2022). MLCM: Multi-Label Confusion Matrix. IEEE Access, 10, 19083–19095. https://doi.org/10.1109/access.2022.3151048
23. Hung, C., Song, E., & Lan, Y. (2019). Image Texture, Texture Features, and Image Texture Classification and Segmentation. In Image Texture Analysis. https://doi.org/10.1007/978-3-030-13773-1_1
24. Inegi. (2016, 15 December). Publicaciones y mapas. Consulted on 1 March 2024, of https://www.inegi.org.mx/app/biblioteca/ficha.html?upc=889463173359
25. Jozdani, S., Johnson, B. A., & Chen, D. (2019). Comparing Deep Neural Networks, Ensemble Classifiers, and Support Vector Machine Algorithms for Object-Based Urban Land Use/Land Cover Classification. Remote Sensing, 11(14), 1713. https://doi.org/10.3390/rs11141713
26. Kozák, J., Probierz, B., Kania, K., & Juszczuk, P. (2022). Preference-Driven classification measure. Entropy, 24(4), 531. https://doi.org/10.3390/e24040531
27. Latorre-Cardenas, M. C., González-Rodríguez, A., Godínez-Gómez, O., Arima, E., Young, K. R., Denvir, A., García-Oliva, F., & Ghilardi, A. (2023). Estimating fragmentation and connectivity patterns of the temperate forest in an Avocado-Dominated landscape to propose conservation strategies. Land, 12(3), 631. https://doi.org/10.3390/land12030631
28. Mas, J., Lemoine-Rodríguez, R., González, R., López-Sánchez, J. G., Piña-Garduño, A., & Herrera-Flores, E. (2017). Evaluación de las tasas de deforestación en Michoacán a escala detallada mediante un método híbrido de clasificación de imágenes SPOT. Madera Y Bosques, 23(2), 119–132. https://doi.org/10.21829/myb.2017.2321472
29. MathWorks. (n.d.). Texture Analysis Using the Gray-Level Co-Occurrence Matrix (GLCM). https://nl.mathworks.com/help/images/texture-analysis-using-the-gray-level-co-occurrence-matrix-glcm.html
30. Medina-García, C., De Azcárate, J. G., & Montes, A. V. (2020). Las comunidades vegetales del bosque de Coníferas Altimontano en el macizo del Tancítaro (Michoacán, México). Acta Botanica Mexicana, 127. https://doi.org/10.21829/abm127.2020.1584

31. Moharram, M. A., & Meena, S. D. (2022). Dimensionality reduction strategies for land use land cover classification based on airborne hyperspectral imagery: a survey. Environmental Science And Pollution Research, 30(3), 5580–5602. https://doi.org/10.1007/s11356-022-24202-2
32. Monterrubio-Rico, T. C., Charre-Medellín, J. F., & López-Ortiz, E. I. (2019). Wild felids in temperate forest remnants in an avocado plantation landscape in Michoacán, Mexico. Southwestern Naturalist, 63(2), 137. https://doi.org/10.1894/0038-4909-63-2-137
33. Mountrakis, G., Im, J., & Ogole, C. (2011). Support vector Machines in Remote Sensing: a review. Isprs Journal of Photogrammetry and Remote Sensing, 66(3), 247–259. https://doi.org/10.1016/j.isprsjprs.2010.11.001
34. NASA. (2021, 3 December). Landsat 8 Bands | Landsat Science. Landsat Science | A joint NASA/USGS Earth observation program. https://landsat.gsfc.nasa.gov/satellites/landsat-8/landsat-8-bands/
35. Nikparvar, B., & Thill, J. (2021). Machine learning of spatial data. ISPRS international journal of geo-information, 10(9), 600. https://doi.org/10.3390/ijgi10090600
36. Nizalapur, V., & Vyas, A. (2020). TEXTURE ANALYSIS FOR LAND USE LAND COVER (LULC) CLASSIFICATION IN PARTS OF AHMEDABAD, GUJARAT. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLIII-B3-2020, 275–279. https://doi.org/10.5194/isprs-archives-xliii-b3-2020-275-2020
37. O'Byrne, M. L., Ghosh, B., Pakrashi, V., & Schoefs, F. (2012). Texture Analysis based Detection and Classification of Surface Features on Ageing Infrastructure Elements. HAL (Le Centre Pour La Communication Scientifique Directe). https://hal.science/hal-01009012
38. Oh, S. (2022). Predictive case-based feature importance and interaction. Information Sciences, 593, 155–176. https://doi.org/10.1016/j.ins.2022.02.003
39. Ouma, Y. O., Keitsile, A., Nkwae, B., Odirile, P. T., Moalafhi, D. B., & Qi, J. (2023). Urban land-use classification using Machine learning classifiers: comparative evaluation and post-classification Multi-feature fusion approach. European Journal of Remote Sensing, 56(1). https://doi.org/10.1080/22797254.2023.2173659
40. Pontius, R. G. (2006). GEOMOD Modelling. Clark University.
41. Puissant, A., Hirsch, J., & Weber, C. (2005). The utility of texture analysis to improve per-pixel classification for high to very high spatial resolution imagery. International Journal of Remote Sensing, 26(4), 733–745. https://doi.org/10.1080/01431160512331316838
42. Rustam, Z., & Kharis, S. A. A. (2020). Comparison of Support Vector Machine recursive feature elimination and kernel function as feature selection using Support Vector Machine for lung cancer classification. Journal of physics, 1442(1), 012027. https://doi.org/10.1088/1742-6596/1442/1/012027
43. Salimi, A., Ziaii, M., Amiri, A., Zadeh, M. H., Karimpouli, S., & Moradkhani, M. (2018). Using a Feature Subset Selection method and Support Vector Machine to address curse of dimensionality and redundancy in Hyperion hyperspectral data classification. The Egyptian Journal Of Remote Sensing And Space Science, 21(1), 27–36. https://doi.org/10.1016/j.ejrs.2017.02.003
44. Sheykhmousa, M., Mahdianpari, M., Ghanbari, H., Mohammadimanesh, F., Ghamisi, P., & Homayouni, S. (2020). Support Vector Machine Versus Random Forest for Remote Sensing Image Classification: A Meta-Analysis and Systematic Review. IEEE Journal Of Selected

Topics in Applied Earth Observations And Remote Sensing, 13, 6308–6325. https://doi.org/10.1109/jstars.2020.3026724
45. Talukdar, S., Singha, P., Mahato, S., Shahfahad, Liou, Y. A., & Rahman, A. (2020). Land-Use Land-Cover Classification by Machine Learning Classifiers for Satellite Observations—A review. Remote Sensing, 12(7), 1135. https://doi.org/10.3390/rs12071135
46. Testas, A. (2023). Support Vector Machine Classification with Pandas, Scikit-Learn, and PySpark. In Apress eBooks (pp. 259–280). https://doi.org/10.1007/978-1-4842-9751-3_10
47. Tüceryan, M., & Jain, A. K. (1993). TEXTURE ANALYSIS. In WORLD SCIENTIFIC eBooks (pp. 235–276). https://doi.org/10.1142/9789814343138_0010
48. Vargas-Canales, J. M., Carbajal-Flores, G., Lara, T. I. B., Vera, J. H. C., Fresnedo-Ramírez, J., Palacios-Rangel, M. I., & Rodríguez-Haros, B. (2020). Impact of the market on the specialization and competitiveness of avocado production in Mexico. International Journal of Fruit Science, 20(sup3), S1942–S1958. https://doi.org/10.1080/15538362.2020.1837711
49. Vidales, K. B. V., & Ortíz, D. A. A. (2014). Responsabilidad social de las empresas agrícolas y agroindustriales Aguacateras de Uruapan, Michoacán, y sus implicaciones en la competitividad. Contaduría y Administración, 59(4), 223–251. https://doi.org/10.1016/s0186-1042(14)70161-5
50. Xiao, L., Wang, Y., Basu, S., Kumbier, K., & Yu, B. (2019). A debiased MDI feature importance measure for random forests. arXiv (Cornell University), 32, 8047–8057. https://arxiv.org/pdf/1906.10845.pdf
51. Zhu, Y., Fan, L., Li, Q., & Jing, C. (2023). Multi-Scale Discrete Cosine Transform Network for Building Change Detection in Very-High-Resolution Remote Sensing Images. Remote Sensing, 15(21), 5243. https://doi.org/10.3390/rs15215243

Satellite data

1. Landsat-8 image (2022) courtesy of the U.S. Geological Survey
2. Planet Team (2022). Planet Application Program Interface: In Space for Life on Earth. San Francisco, CA. https://api.planet.com
3. Sentinel-2 image (2022) courtesy of the Copernicus SciHub

8. Appendix

Table 2. INEGI dataset of crop production in Uruapan, Mexico.

Year	Id status	Name province	Idddr	Name ddr	Idcader	Nomcader	Id municipality	Gemeente	Idciclo	Productive name	Id modality	Name modality	Measure Id unity	Nomination	Id	Name cultivation	Sown (ha)	Harvested (ha)	Damaged (ha)	Production volume	Yield	Price medio rural (pesos)	Production value (pesos)
2022	16	Michoacán	86	Uruapan	1	Parangaricutiro	102	Uruapan	1	Otoño-Invierno	1	Riego	200201	Tonelada	5740000	Calabacita	10	10	0	195.5	19.55	7749.92	1515109.36
2022	16	Michoacán	86	Uruapan	1	Parangaricutiro	102	Uruapan	1	Otoño-Invierno	1	Riego	200201	Tonelada	6120000	Chile verde	12	12	0	73.2	6.1	8707.45	637385.34
2022	16	Michoacán	86	Uruapan	1	Parangaricutiro	102	Uruapan	1	Otoño-Invierno	1	Riego	200201	Tonelada	7490000	Maíz grano	155	155	0	550.25	3.55	6200	3411550
2022	16	Michoacán	86	Uruapan	1	Parangaricutiro	102	Uruapan	1	Otoño-Invierno	1	Riego	200201	Tonelada	8970000	Tomate rojo (jitomate)	12.5	12.5	0	176.25	14.1	10109.36	1781774.7
2022	16	Michoacán	86	Uruapan	1	Parangaricutiro	102	Uruapan	1	Otoño-Invierno	1	Riego	200201	Tonelada	8980000	Tomate verde	8	8	0	70.7	8.84	9500	671650
2022	16	Michoacán	86	Uruapan	1	Parangaricutiro	102	Uruapan	2	Primavera-Verano	2	Temporal	200201	Tonelada	5690000	Cacahuate	12	12	0	21.6	1.8	13447.96	290475.94
2022	16	Michoacán	86	Uruapan	1	Parangaricutiro	102	Uruapan	2	Primavera-Verano	2	Temporal	200201	Tonelada	5740000	Calabacita	15	15	0	196.5	13.1	9223.06	1812331.29
2022	16	Michoacán	86	Uruapan	1	Parangaricutiro	102	Uruapan	2	Primavera-Verano	2	Temporal	200201	Tonelada	7490000	Maíz grano	4730	4730	0	10926.3	2.31	6607.06	72190719.68
2022	16	Michoacán	86	Uruapan	1	Parangaricutiro	102	Uruapan	2	Primavera-Verano	2	Temporal	200201	Tonelada	8970000	Tomate rojo (jitomate)	25	25	0	352.5	14.1	9636.6	3396901.5
2022	16	Michoacán	86	Uruapan	1	Parangaricutiro	102	Uruapan	3	Perennes	1	Riego	200201	Tonelada	5060000	Aguacate	11260	11110	0	121572	10.94	25925.11	3151768051
2022	16	Michoacán	86	Uruapan	1	Parangaricutiro	102	Uruapan	3	Perennes	1	Riego	200201	Tonelada	5460000	Ave del paraíso	6	6	0	1751	291.83	306	535806
2022	16	Michoacán	86	Uruapan	1	Parangaricutiro	102	Uruapan	3	Perennes	1	Riego	200201	Tonelada	5820000	Caña de azúcar	171.27	171.27	0	13504.64	78.85	980	13234547.2
2022	16	Michoacán	86	Uruapan	1	Parangaricutiro	102	Uruapan	3	Perennes	1	Riego	200201	Tonelada	5830000	Semilla de caña de azúcar	2	2	0	206	103	945	194670
2022	16	Michoacán	86	Uruapan	1	Parangaricutiro	102	Uruapan	3	Perennes	1	Riego	200201	Tonelada	6530000	Durazno	30	30	0	264	8.8	10974.28	2897209.92
2022	16	Michoacán	86	Uruapan	1	Parangaricutiro	102	Uruapan	3	Perennes	1	Riego	200201	Tonelada	7060000	Guayaba	222	222	0	2222	10.01	9823.87	21828639.14
2022	16	Michoacán	86	Uruapan	1	Parangaricutiro	102	Uruapan	3	Perennes	1	Riego	200201	Tonelada	7560000	Mango	15	15	0	226.5	15.1	7812.52	1769535.78
2022	16	Michoacán	86	Uruapan	1	Parangaricutiro	102	Uruapan	3	Perennes	1	Riego	200201	Tonelada	7830000	Nanche	11	11	0	60.39	5.49	7750.84	468073.23
2022	16	Michoacán	86	Uruapan	1	Parangaricutiro	102	Uruapan	3	Perennes	1	Riego	200201	Tonelada	7920000	Nopalitos	31	31	0	1653.01	53.32	3540.13	5851870.29
2022	16	Michoacán	86	Uruapan	1	Parangaricutiro	102	Uruapan	3	Perennes	1	Riego	200201	Tonelada	9310000	Zarzamora	300	180	0	4500	25	20045.44	90204480
2022	16	Michoacán	86	Uruapan	1	Parangaricutiro	102	Uruapan	3	Perennes	2	Temporal	200201	Tonelada	5060000	Aguacate	6380	6380	0	64507	10.11	26565.03	1713630206
2022	16	Michoacán	86	Uruapan	1	Parangaricutiro	102	Uruapan	3	Perennes	2	Temporal	200201	Tonelada	5460000	Ave del paraíso	12	12	0	5214.1	434.51	402	2096068.2
2022	16	Michoacán	86	Uruapan	1	Parangaricutiro	102	Uruapan	3	Perennes	2	Temporal	200201	Tonelada	5830000	Semilla de caña de azúcar	5	5	0	422.14	84.43	810	341933.4
2022	16	Michoacán	86	Uruapan	1	Parangaricutiro	102	Uruapan	3	Perennes	2	Temporal	200201	Tonelada	6530000	Durazno	14	14	0	119	8.5	11168.92	1329101.48
2022	16	Michoacán	86	Uruapan	1	Parangaricutiro	102	Uruapan	3	Perennes	2	Temporal	200201	Tonelada	7830000	Nanche	14	14	0	73.21	5.23	7875.6	576572.68
2022	16	Michoacán	86	Uruapan	1	Parangaricutiro	102	Uruapan	3	Perennes	2	Temporal	200201	Tonelada	7920000	Nopalitos	42	42	0	2523.99	60.1	3708.82	9361024.59
2022	16	Michoacán	86	Uruapan	1	Parangaricutiro	102	Uruapan	3	Perennes	2	Temporal	200201	Tonelada	15050000	Pastos y praderas	110	110	0	2812.11	25.56	435.28	1224055.24
Total																	23604.77	23334.77	0	234193.89	1314.23	220250.25	5103019742

Table 3. List of deliverables.

List of research data	
Type	Data
Final report	word, PDF
Midterm presentation	pptx
Final presentation	pptx
Planetscope satellite imagery	TIFF
Sentinel-2 satellite imagery	TIFF
Landsat 8 satellite imagery	TIFF
Mexico administrative levels	Shapefile
INEGI land use and land vegetation	Shapefile
INEGI crop production 2022	Excel
Random Forest model	Python Source File
Support Vector Machine model	Python Source File
GLCM texture	R Source File
Grayscale images	Figure
Study area	Maps
Avocado prediction maps	Maps
Confusion matrices and feature importances	Figure
Flowchart methodology	Figure
Satellite data	Table