

Detection of Lime Crops Using Sentinel-2 Satellite Images and Machine Learning

César Villaseñor-Sotomayor¹, Ramona E. Chávez-Valdez¹, Patricia E. Figueroa-Millán¹,
Fernando Pech-May², Santiago Arceo-Díaz¹, Ismael Amezcua-Valdovinos³

¹ Tecnológico Nacional de México, Campus Colima,
Mexico

² Tecnológico Nacional de México, Campus de los Ríos,
Mexico

³ Universidad de Colima, Facultad de Telemática,
Mexico

villasarcesar@gmail.com, {echavez, patricia.figueroa santiago.arceo }@colima.tecnm.mx,
fernando.pech@cinvestav.mx, ismaelamezcua@ucol.mx

Abstract. Lime is a widely consumed agricultural product in Mexico and is predominantly produced in the state of Colima, making it a key component of the region's economy. Lime trees are commonly affected by Huanglongbing, a disease present in Colima since 2010. Its spread, combined with the lack of control measures, has significantly reduced cultivated areas and production, resulting in substantial economic losses and negatively impacting local growers. To address this issue, mapping cultivated areas is essential for monitoring plant health and conducting accurate, up-to-date, and reliable lime crop censuses. Recent studies have shown that remote sensing and machine learning are effective tools for crop mapping and monitoring in specific regions. Many of these studies leverage optical satellite imagery and phenological analysis to assess plant health. This study proposes a strategy for lime crop detection and mapping using the Google Earth Engine (GEE) code editor, Sentinel-2 imagery, and machine learning algorithms: Random Forest (RF), Support Vector Machine (SVM), Classification and Regression Trees (CART), and Gradient Tree Boosting (GTB). The study area covers the municipalities of Tecoman and Armeria, Colima. The models were trained using 3,932 sampling points. Model performance evaluation indicates that GTB achieved the highest accuracy in lime crop detection, supporting informed decision-making, efficient management of agricultural resources, and the implementation of strategic policies.

Keywords. Sentinel-2, Machine Learning, RF, SVM, CART, GTB, NDVI, GNDVI.

1 Introduction

Mexico is considered one of the main producers of lime, the second most important citrus fruit in Mexico. Colima and Michoacán are the states that stand out in the production of Mexican lime with 89% of the production [1]. According to the Agri-Food and Fisheries Information Service (SIAP, by its initials in Spanish) [2], in 2008 there were around 29 thousand hectares (ha) planted with lime and a production of approximately 657 thousand metric tons (mt). This fruit is not exempt from pathogens; the disease that most affects this citrus is Huanglongbing (HLB), considered one of the most serious and destructive. In Mexico, it has caused serious damage, such as what occurred in 2010 in the state of Colima [3], severely affecting production; planting was reduced to 27% (21 thousand hectares) and production to 42% (381 thousand metric tons). These multimillion-dollar losses led to the implementation of actions to control the disease and recover the planted areas.

This is essential for environmental management and precision agriculture. For this purpose, emerging tools and technologies have been used, such as remote sensing, classification algorithms based on machine learning and deep learning, which, in combination with metrics like spectral indices, have facilitated the identification

of agricultural study areas. In response to this situation, the use of Sentinel-2 satellite images and machine learning is proposed for the detection of lime crops, thus providing a viable option for its applicability in various scenarios such as agricultural censuses, deforestation monitoring, or forest mapping.

In this context, an agricultural census is addressed as a case study for the identification of lime crops in the state of Colima, Mexico; which arises from the need to have censuses that report accurate data on the planted area and serve as tools to establish strategic policies for the crop, as well as to manage resources that support the activity.

Having updated censuses is essential for decision-making and data analysis. However, conducting them in situ is a complex task that involves a large amount of human and economic resources [4]. In addition, a very large number of surveyors and supervisors is required to ensure the reliability of the data collected properly, and finally, extensive processing of this data is necessary, which can lead to delays in the results. Therefore, the application of crop detection through satellite images helps mitigate the aforementioned issues and allows for obtaining results similar to those of a traditional census. In addition, this contributes to the development of strategies for the detection of Land Use and Land Cover (LULC). These approaches have been fundamental for environmental management, precision agriculture, and the mapping of vegetation and crops.

1.1 Theoretical Framework

Satellite images are obtained through sensors installed on satellite platforms, mainly aimed at Earth observation. These images are composed of multiple bands, and the sensors used can be classified depending on their information acquisition mechanism (active or passive) [5].

Active sensors operate with their own light source; they send a signal to the Earth and then check the response, that is, the amount received. In contrast, passive sensors rely on the natural energy from the sun that bounces off the Earth [6].

These images have proven to be highly useful in areas such as agriculture, mining, the

environment, hydrology, natural disaster management, among others.

There are different types of programs that have launched various satellites into orbit, designed to fulfill specific functions in different fields of knowledge and technology, such as Copernicus, Galileo, EGNOS, and Meteosat.

One way to categorize satellite images is based on the number and nature of the spectral bands they record. The main types are: multispectral, hyperspectral, panchromatic, and Synthetic Aperture Radar (SAR) images.

A multispectral image corresponds to a set of spectral bands acquired from the same scene, with each band capturing information within a specific range of wavelengths in the electromagnetic spectrum. They typically consist of between 4 and 15 bands [7, 8].

A panchromatic image, on the other hand, is generated from a single spectral band that integrates information from the red, green, and blue bands. The resulting image does not provide specific spectral information, as it is composed of a single broadband. The captured variable corresponds to the brightness intensity reflected by the objects in the scene, which are represented in a grayscale format [9, 10]. Likewise, the acquisition of hyperspectral images is a process that allows capturing high-resolution spectral information through the decomposition of light. This type of image is composed of a large set of bands, usually exceeding 100 [11, 12].

Finally, Synthetic Aperture Radar (SAR) images use electromagnetic waves in the radar band to generate detailed representations of the Earth's surface. The system emits energy pulses that, upon striking the surface, are backscattered toward the sensor. SAR images are constructed based on the analysis of the characteristics of this backscattering, which depend on the physical properties of the terrain [13, 14].

Copernicus is a multi-satellite project developed by the European Space Agency (ESA), dedicated to Earth and environmental observation. It includes various satellites, among which the Sentinel family stands out. In addition, it is supported by additional satellites, both commercial and governmental [15, 16]. The Sentinel family is composed of [17, 18]:

1. Own satellites, Sentinel-1, -2, -3, and -6, designed to meet the specific requirements of the Copernicus program:
 - Sentinel-1 captures radar images day and night, under any conditions, for land and ocean monitoring.
 - Sentinel-2 provides high-resolution optical images that allow the observation of vegetation, soil, water coverage, among other landscape elements.
 - Sentinel-3 provides valuable optical, radar, and altimetric data essential for marine and land monitoring.
 - Sentinel-6 provides precise altimetric measurements to monitor the height of the global sea surface.
2. Sentinel-4 and -5 serve as instruments onboard meteorological satellites managed by EUMETSAT:
 - Sentinel-4 collects data on the atmosphere in order to observe the presence of trace gases and aerosols over Europe, assessing air quality.
 - Sentinel-5, like Sentinel-4, monitors the composition of the atmosphere by providing precise data on pollutant compounds such as ozone, various oxides, and particulate matter present in the atmosphere.

Sentinel-2 [19] operates with multispectral images of wide spatial coverage and high resolution. Its main purpose is to support operational applications aimed at monitoring terrestrial variables such as vegetation, soil, and bodies of water. The platform is equipped with an optical instrument that captures information in 13 spectral bands.

1.2 Related Work

In the literature, there are various proposals that address LULC management through the use of geospatial images and machine learning techniques. During the thorough analysis, some related works stand out.

Wang et al. [20] presented a review of 107 LULC satellite products with the aim of providing

a guide to facilitate the selection of the most suitable product according to specific needs. The review aims to highlight products that cover various categories, not only those focused on forests, which are usually the most common. The authors point out that limitations in accuracy and validation still persist, the use of technologies such as deep learning and cloud computing is driving improvements in this area.

Kanja et al. [21] carried out a mapping of the Miombo woodlands in Zambia into three specific classes: reference forest, degraded forest, and regenerating forest. Using RF through GEE for LULC classification and a U-Net convolutional neural network for the different types of classes. As a result, RF achieved an overall accuracy of 93% and an F1 score ranging from 93% to 96%, depending on the seasons. Meanwhile, the U-Net neural network obtained an F1 score of 85% for reference forests, 73% for degraded forests, and 72% for regenerating forests.

Shah et al. [22] estimated LULC and classified land cover in Kolhapur, India, by implementing Sentinel-2 along with RF and k-nearest neighbor (k-NN). Five classes were estimated: water, urbanization, agriculture, forests, and barren land, with the following results highlighted: RF achieved an overall accuracy of 94.24%, and k-NN 84.12%.

Ouchra et al. [23], Alciaturi et al. [24], and Tan et al. [25] present a strategy for land use classification using the Google Earth Engine (GEE) platform. For the land use analysis, they used machine learning algorithms along with spectral indices to improve the accuracy of the results.

Ouchra et al. [23] evaluated six classification algorithms: Support Vector Machine (SVM), Random Forest (RF), Classification and Regression Trees (CART), Minimum Distance (MD), Decision Tree (DT), and Gradient Tree Boosting (GTB). The MD algorithm achieved the highest accuracy, with 93.85% without indices and 95.76% with them, while SVM showed the poorest performance (74.56% and 81.35%, respectively). In general, the inclusion of spectral indices improved the accuracy across all classifiers. Similarly, Alciaturi et al. [24] analyzed land use and crop types during the summer and winter seasons.

For the evaluation, they used RF, SVM, and GTB with Sentinel-1 and Sentinel-2 images, in addition to using elevation data. In the summer season, they identified rice crops, grasslands, forests, and urban areas, while in winter they added post-agriculture fields and winter crops. RF and GTB achieved the highest accuracy in summer (94% accuracy, Kappa = 0.92), and GTB in winter (94%, 0.92), followed by RF and SVM (89%, 0.87). On the other hand, Tan et al. [25], using Sentinel-2A images, compared pixel-based (PBIA) and object-based (OBIA) classification approaches with RF and SVM. The results showed that RF with OBIA was the most accurate (with an average of 99% accuracy, Kappa 0.98), followed by RF with PBIA (93%, Kappa 0.89). SVM with OBIA (85%, Kappa 0.79) and with PBIA (85%, Kappa 0.78).

Sheykhoumousa et al. [26] conducted a systematic review of 251 articles on the performance of SVM and RF classifiers in remote sensing image classification. Of these, 42% reported implementations using RF and 68% using SVM. The review results indicated that, in the crop classification category, SVM achieved an average accuracy of 70%, while RF reached 90%.

Ashfaq et al. [27] predicted wheat yield in Multan, Pakistan, based on seasonal climate, satellite imagery, and soil properties using RF, SVM, and Least Absolute Shrinkage and Selection Operator (LASSO) in GEE and ArcGIS. RF was the most accurate, with 97% accuracy and an R^2 of 0.88, followed by SVM with 93%, while Lasso Regression achieved 85%.

Zhang et al. [28] identified topsoil loss in Heilongjiang, China, using Sentinel-2 imagery, topographic data, and machine learning. They applied RF and SVM optimized with Particle Swarm Optimization (PSO) to tune their hyperparameters.

RF with PSO achieved the highest accuracy (97% accuracy, Kappa 0.94), followed by SVM (94% and 0.90). Without optimization, RF and SVM achieved only 72% and 68% accuracy, respectively.

Liu et al. [29] studied vegetation recovery after forest fires in Yunnan Province, China, using GEE and machine learning techniques along with spectral indices to analyze post-fire recovery.

They compared four classification models: RF, SVM, CART, and Naive Bayes (NB). The results showed that RF achieved the highest accuracy (90%), followed by SVM (88%), CART (85%), and NB (78%).

Yu et al. [30] identified and monitored irrigated areas (2019 to 2023) in arid regions of Xinjiang, China, using Sentinel-2 imagery, spectral indices, and machine learning algorithms. They compared RF, CART, and SVM, determining that RF was the most accurate model, with a Kappa value ranging from 0.82 to 0.92 across different years and overall accuracy between 88% and 93%. CART followed (Kappa between 0.67 and 0.75, overall accuracy between 84% and 87%), and finally SVM (Kappa between 0.45 and 0.75, and overall accuracy between 71% and 86%).

Morell-Monzó et al. [31] developed a classification framework to identify abandoned citrus crops based on machine learning, spectral indices extracted from Sentinel-2 imagery, and the Random Forest classifier to categorize the crops into three states: non-productive, productive, and abandoned. When evaluating the model on three datasets with temporal and spatial variability, they achieved an overall accuracy ranging from 89% to 92%.

Among the algorithms presented in the cited studies, the most commonly used were SVM and RF. In most cases, RF stood out by delivering the best results, although MD and GTB also demonstrated notable performance in some studies. The use of these algorithms and their results in vegetation and agricultural crop detection support their relevance for identifying lime crops through the use of satellite imagery.

1.3 Machine Learning

There are various machine learning techniques that can be used for classification tasks. The most prominent algorithms are RF, SVM, CART, and GTB.

First, RF [32] is proposed as a classification method based on an ensemble of decision trees, where each tree is generated from randomly sampled variables that are independently and identically distributed across all trees in the ensemble.

The generalization ability of this type of model depends both on the individual accuracy of the trees that compose it and on the degree of correlation between them.

In turn, the concept of SVM [33] was introduced as a binary classification method. This approach constructs a decision function based on an optimal hyperplane, whose purpose is to separate instances belonging to two different classes with the maximum possible margin between them, while simultaneously minimizing the classification error rate.

Similarly, the CART algorithm [34] is a predictive modeling technique that builds a binary decision model designed to predict a target variable, based on learning simple decision rules derived from the features present in the data.

Finally, GTB [35] consists of an ensemble of binary decision trees, in which each tree represents a sequence of binary evaluations leading to two possible outcomes. This ensemble is built sequentially, progressively incorporating new trees. The first tree is initially trained, and each subsequent tree is fitted with the goal of reducing the error made by the previous ensemble.

It is important to note that satellite images require processing. In this regard, Google's platform (GEE) enables large-scale geospatial data processing [36], providing a suitable tool for implementing the classification models mentioned above. Additionally, it provides a solution to the limitations that arise when working with large volumes of training points and processing tasks directly within the GEE code editor.

1.4 Radiometric Indices

Spectral indices are values derived through mathematical operations between broadband spectral bands. These indices are able to highlight specific features that may be of interest, such as vegetation, water bodies, soil moisture, among others [37]. The most commonly used in land studies are:

- Normalized Difference Vegetation Index (NDVI). It is obtained from a normalized difference between the red and near-infrared

(NIR) spectral bands, taking advantage of the contrast between chlorophyll pigment absorption in the red band and the high reflectance characteristic of vegetation in the NIR band. As a result, NDVI allows for the estimation of vegetation vigor and canopy density [38]:

$$NDVI = \frac{NIR - RED}{NIR + RED}. \quad (1)$$

- Green Normalized Difference Vegetation Index (GNDVI). Similar to NDVI, but instead of using the red band, it uses the green band. This allows for the estimation of photosynthetic activity and the assessment of water and nitrogen uptake in vegetation [39]:

$$GNDVI = \frac{NIR - GREEN}{NIR + GREEN}. \quad (2)$$

- Normalized Multiband Drought Index (NMDI). It uses a calculation similar to the previous indices, but instead of using the red or green band, it relies on the difference between two shortwave infrared (SWIR) bands to determine soil and vegetation moisture. It is based on NIR and SWIR bands—the first reflects the overall moisture level, while the second captures absorption related to water content [40,41]:

$$NMDI = \frac{NIR - (SWIR 1 - SWIR 2)}{NIR + (SWIR 1 - SWIR 2)}. \quad (3)$$

1.5 Evaluation Metrics

To analyze the results of each algorithm, various evaluation metrics are used, such as the Kappa index, F1-Score, and Intersection over Union (IoU), these metrics help assess the quality of the performance obtained, determining the level of agreement, effectiveness, and the similarity between the classified data and the actual data.

The kappa index (κ) measures the relationship between two evaluation indicators (classified data

Table 1. Kappa Scale by Landis and Koch

Kappa (κ)	Degree of Agreement
< 0.00	No agreement
0.00 - 0.20	Slight
0.21 - 0.40	Fair
0.41 - 0.60	Moderate
0.61 - 0.80	Substantial
0.81 - 1.00	Almost perfect

and actual data); it calculates the agreement between the obtained results, excluding those expected by random. A value of 1 indicates perfect agreement, while a value of 0 represents a random result [42]:

$$\kappa = \frac{P_o - P_e}{1 - P_e}, \quad (4)$$

P_o = the proportion of observed agreements

P_e = the proportion of expected agreements

As mentioned in [42], Landis and Koch [43] proposed the kappa index rating scale, shown in Table 1. This scale is widely used to assess the reliability of classifications.

Also, by combining two parameters, precision and recall, the F1-Score metric emerges. Precision represents the number of positive predictions that were correctly assigned, while recall indicates the number of actual positive cases that the model correctly identified. The F1-Score is calculated as the harmonic mean of both metrics, thus providing a balanced indicator of the model's performance [44]:

$$F1 = \frac{2TP}{2TP + FP + FN}, \quad (5)$$

where TP = number of true positives,

FN = number of false negatives,

FP = number of false positives.

In turn, the IoU evaluates the similarity between the ground truth data and the classification predicted by the model. It is calculated as the ratio between the intersection area, that is, the samples classified in the same category in both the prediction and the ground truth, and the union area, which includes all samples classified in either category, whether correctly or incorrectly [45].

$$IoU = \frac{B_1 \cap B_2}{B_1 \cup B_2}, \quad (6)$$

where B_1 = predicted area,

B_2 = actual area.

These concepts provide a solid foundation for conducting a comparative analysis in the specific case of the study area, facilitating the appropriate selection of the algorithm through an objective evaluation of performance. In this way, supporting the methodology and justifying the use of the selected tools.

2 Methodology

This section describes the case study, the data source, the dataset used, the preprocessing, the parameters of each classifier, and the workflow to achieve lime crop detection using satellite imagery.

2.1 Workflow

To select the appropriate algorithm, the workflow shown in Figure 3 was used, which allowed training and evaluating the different algorithms under the same conditions. First, through the GEE code editor, Sentinel-2 images were accessed and preprocessed to obtain a clean image of the study area. Then, a visual selection of training points was carried out, and these data were exported to the GEE cloud in order to be processed in Python, as shown in Figure 3.

To select the appropriate algorithm, the workflow shown on the right side of Figure 3 was applied, which allowed the different algorithms to be trained and evaluated under the same conditions.

Using the GEE API in Python, the Sentinel-2 images and the previously selected training points were imported. A preprocessing step was performed to clean the resulting image, to which spectral indices were applied. Subsequently, each algorithm was trained, and finally, a comparison was conducted to evaluate their performance.

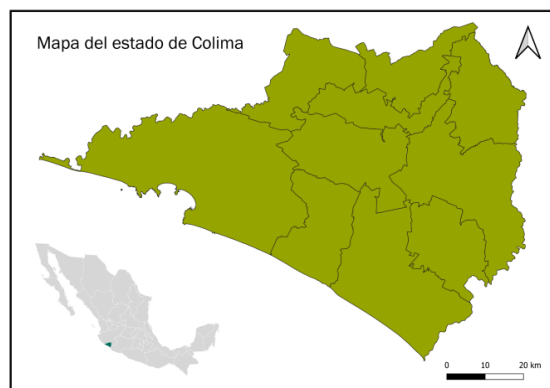


Fig. 1. Map of the State of Colima

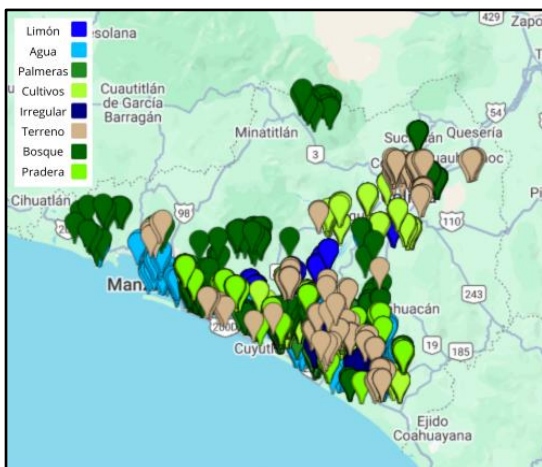


Fig. 2. Points assigned for the classification

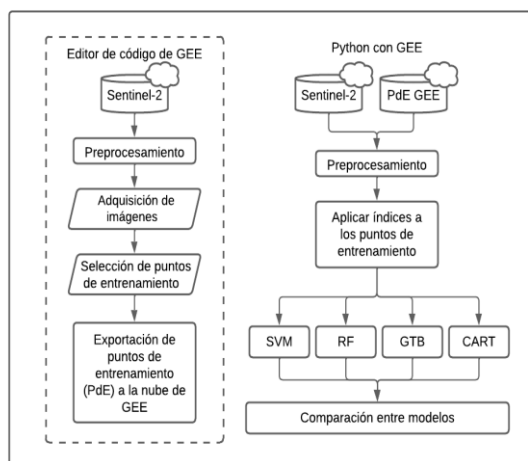


Fig. 3. Workflow

2.2 Study Area

The study was conducted in the region of Colima, Mexico, as shown in Figure 1. The geographic coordinates are: 19°30'45" north latitude at the northernmost point, 18°20'20" at the southernmost point, 103°29'11" west longitude in the east, and 114°45'34" in the west, covering an area of 5,627 km², it comprises ten municipalities: Armería, Colima, Comala, Coquimatlán, Cuauhtémoc, Ixtlahuacán, Manzanillo, Minatitlán, Tecomán, and Villa de Álvarez.

The predominant climate in most of the region is warm sub-humid with rainfall during the summer.

However, in some areas such as Armería and Tecomán, the climate varies to very warm and warm semi-dry, also with rainfall during the summer [46].

2.3 Data Source

These points have been identified through images from the Sentinel-2 mission, which consists of two satellites aimed at monitoring changes in land surface conditions to support various applications such as vegetation, soil, and water cover monitoring, providing data that is systematically and freely available to the public [47].

- Download platform. The dataset provider used was Copernicus, provided by the European Union, together with the cloud processing platform GEE.
- Image period. A date range was used, from January 1st to May 1st, 2024.
- Cloud mask. To ensure the image undergoes preprocessing, a 'mask clouds' function is first created, which performs cloud and cirrus masking in order to obtain the most complete and clean image possible. Only images without the presence of these elements are selected, resulting in a clear sky to observe the case study. Subsequently, in another function called 'image with indices', spectral indices are calculated, in this case, NDVI, GNDVI, a combination of both indices ('Combined Index'), and NDMI, returning an image with the computation of these four parameters.

Table 2. Number of points for each category

Category	Lime	Water	Palm Tree	Crops	Bare Soil	Forest	Grassland	Irregular
Points	794	200	397	733	207	730	671	200

Table 3. Parameters of Each Classification Mode.

Algorithms	Parameters
RF	numberOfTrees = 100, maxNodes = None, minLeafPopulation = 1, bagFraction = 0.8, variablesPerSplit = None
GTB	numberOfTrees = 200, maxNodes = 100, shrinkage = 0.1, samplingRate = 0.7
CART	maxNodes = 100, minLeafPopulation = 5
SVM	decisionProcedure = 'Voting', svmType = 'C_SVC', kernelType = 'RBF', shrinking = True, gamma = 1, cost = 1000

- Sentinel-2. The satellite images correspond to the Sentinel-2 MultiSpectral Instrument (MSI) sensor, processed at Level-2A for surface reflectance (SR).

Next, the image collection is loaded by applying the previously mentioned processes, first, the images that intersect the study area are filtered, for training, the previously mentioned date range is used, then, filters are applied to ensure clear skies, reducing the number of processed images to only those with less than ten percent cloud cover, subsequently, the 'mask clouds' function is applied to ensure clean-sky images.

In the next step, the 'image with indices' function is applied, the mean of the images is calculated to obtain a single image, and finally, the image is clipped to the study area.

2.4 Training Data

For the case study, eight categories were selected that significantly represent land use and land cover: Lime, Water, Palm Trees, Crops (agricultural areas excluding lime or palm plantations), Bare Soil, Forest, Grassland, and Irregular Pattern (areas with difficult-to-interpret patterns that did not fit into other categories).

These categories were defined with the purpose of adapting the classification to the specific characteristics of the region, thus ensuring better model performance.

Training points were assigned to each category, as shown in Figure 2, the category of interest, which is Lime, was assigned 794 training points. The remaining categories were distributed as shown in Table 2.

80% of the data was used for training and the remaining 20% for validation testing, a widely used approach in this type of context, as presented by authors Ouchra et al. and Liu et al. [23, 29]. To perform random sampling, the function 'randomColumn()' is used. It generates a column with random values between 0 and 1, resulting in an approximately uniform random distribution, that is, training data is selected with random values less than or equal to 0.80, approximately 80%, and validation data corresponds to the remaining 0.20, resulting in roughly 20%.

To analyze the performance of each classifier in the case study, a comparison was carried out among the most commonly used ML models for this type of task. The four models evaluated were: RF, SVM, CART, and GTB, implemented in Python using the GEE library.

Table 4. Comparison between Classification Models

Models	RF	CART	SVM	GTB
Kappa Index for Lime	0.91	0.68	0.64	0.94
Overall Kappa Index	0.88	0.70	0.75	0.88
F1-score	0.90	0.75	0.80	0.90
Intersection IoU	0.82	0.61	0.68	0.82

Table 5. Comparison between the Number of Classified Hectares and Those Reported by SIAP

Year	SIAP	RF	CART	SVM	GTB
2023	21832.04	16439.81 (-24.69%)	35719.28 (+63.60%)	31900.87 (+46.11%)	19163.72 (-12.22%)
2022	21796.81	13301.17 (-38.97%)	32453.41 (+48.89%)	30546.00 (+40.13%)	17491.69 (-19.75%)
2021	21107.45	12925.06 (-38.76%)	36430.08 (+72.59%)	29618.07 (+40.32%)	16142.33 (-23.52%)
2020	19269.10	14265.15 (-25.96%)	61715.81 (+220.28%)	23886.40 (+23.96%)	16001.19 (-16.95%)
2019	19244.85	13097.12 (-31.94%)	39769.96 (+106.65%)	24718.46 (+28.44%)	15540.13 (-19.25%)

2.5 Parameters of Each Classifier

In RF, a total of 100 decision trees were used, as shown in Table 3, each tree was trained with 80% of the available data, to perform the splits within a tree, a number of variables per split was selected. Since the parameter "ss" was set to None, GEE used the square root of the number of features, each tree node used at least one sample, and since the maximum number of leaf nodes per tree was not specified, there was no limit to how much the trees could grow.

On the other hand, to train the SVM in terms of making classification decisions, the voting method was used, a penalized support vector machine was implemented, For the kernel type, a radial basis function was used, the shrinking algorithm was enabled, a gamma value of 1 was added to the kernel function, and finally, a cost

parameter value of 1000 was used, as shown in Table 3.

In CART, as shown in Table 3, the maximum number of leaf nodes per tree was limited to 100, and the minimum size of each leaf node was set to 5 points.

GTB was configured to create 200 decision trees. The shrinkage parameter was set to 0.1, a sampling rate of 70% was used for each tree, and the maximum number of leaf nodes per tree was limited to 100, as detailed in Table 3.

2.6 Postprocessing Performed

Once the classification was executed, postprocessing was carried out to improve the quality of the results, first, a focal filter was implemented using 'ee.Image.focalMode', subsequently, other filter denominated

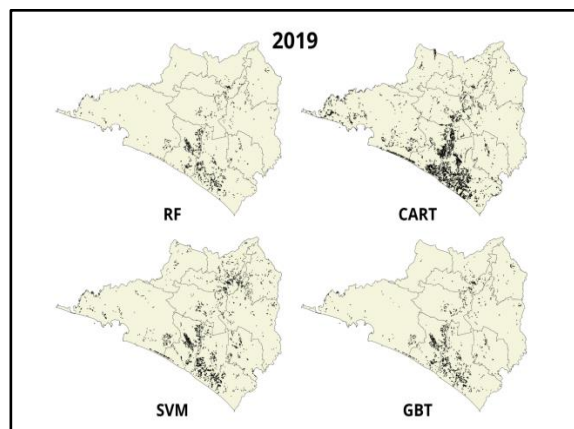


Fig. 4. Lime Classification Comparison in Colima for the Year 2019

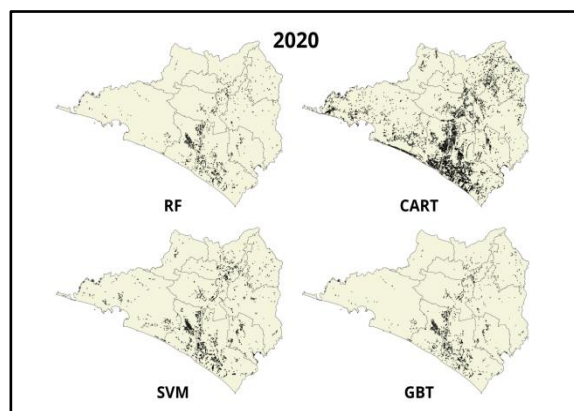


Fig. 5. Lime Classification Comparison in Colima for the Year 2020

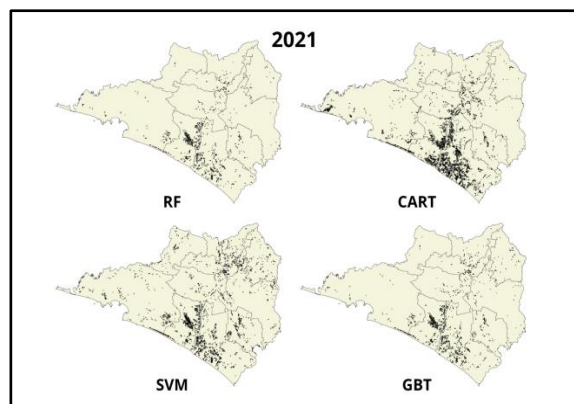


Fig. 6. Lime Classification Comparison in Colima for the Year 2021

'ee.Image.connectedPixelCount' method was used to calculate how many connected pixels of the same value surround each pixel, limited to 256 pixels. Then, a filter was created to identify pixels considered as noise, those with fewer than 150 surrounding pixels, making them transparent or null, resulting in a classification without 'noise', finally, from all category pixels, only those classified as lime were extracted.

3 Results

The models used for classification enabled the detection of lime crops through Sentinel-2 satellite images and machine learning. Using the Kappa index, two main values were measured: the first was the Kappa value for the lime category, and the second was a global Kappa value that considers all eight categories. This approach provides a more balanced view of the algorithm's performance and for identifying potential imbalances among categories. In addition, global F1-score values were also evaluated to analyze performance, as well as Intersection over Union (IoU) to assess similarity, as shown in Table 4.

The GTB model achieved the best result in the Kappa index for the lime category, with a value of 0.94, indicating an almost perfect level of agreement along with RF, although the latter had a slightly lower result of 0.91. The GTB and RF models achieved the same values in the following evaluation metrics: global Kappa index, F1-score, and Intersection over Union (IoU). In the global Kappa index, which reflects the average agreement across all categories, they obtained an almost perfect agreement with a value of 0.88. For the F1-score, they showed a performance of 0.90, and in the IoU metric, they demonstrated a similarity of 0.82.

CART and SVM did not reach the expected performance; both models obtained Kappa coefficients in the range of 0.64 to 0.75, indicating a substantial level of agreement, their F1-score performances were 0.75 and 0.80, respectively, and their Intersection over Union (IoU) similarity values were below 0.70.

The statistical yearbook of agricultural production from SIAP was consulted as an

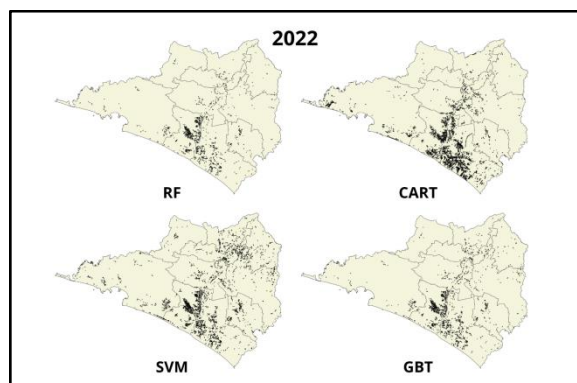


Fig. 7. Lime Classification Comparison in Colima for the Year 2022

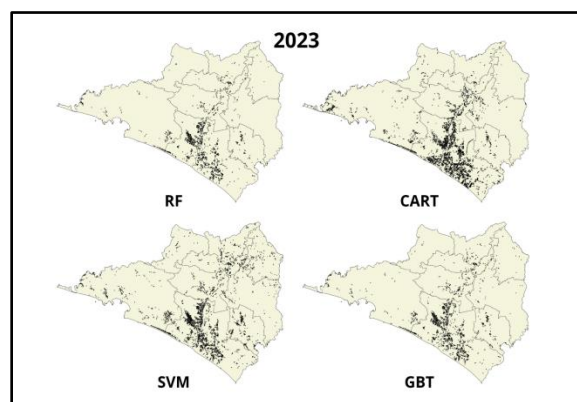


Fig. 8. Lime Classification Comparison in Colima for the Year 2023

external reference to validate the results, obtaining the area planted with lime (in hectares) in the state of Colima over the past five years starting from 2023, as shown in Table 5.

These values were compared with the hectares identified by each model during the same years, using an ImageCollection with the aforementioned preprocessing and with dates ranging from January 1st to December 31st of each year.

In relation to the performance shown in Table 4, Table 5 reveals that the GBT model achieves better classification of lime hectares compared to other models, with an average discrepancy of 18.33% relative to the hectares reported by SIAP, and showing its best performance in 2023 with a discrepancy of 12.22%. On the other hand, CART

shows the worst performance, with an average discrepancy of 102.40%, and its best reappraisal occurring in 2022 with a discrepancy of 44.89%.

The maps generated from the lime classification across different years show a certain similarity in the presence of this crop in the regions of Armería and Tecmán, as illustrated in Figures 4, 5, 6, 7, and 8.

4 Discussion

In the previously mentioned studies, RF stands out with better results compared to other models such as SVM and CART. On the other hand, some studies also highlight the use of GTB, showing very good performance.

These trends are reflected in the results obtained in this research, as RF and especially GTB demonstrate better performance compared to SVM and CART, which did not show outstanding results.

Although the GTB model achieved favorable results in the evaluation metrics, a deviation ranging from 24% to 12% was observed compared to the expected values, which affects the model's accuracy. For this study, an expected error margin below 10% was considered desirable in order to approach official census data.

5 Conclusion

Based on the analysis of the performance of various lime classification models, the following main results were derived regarding the detection of lime crops using Sentinel-2 satellite imagery and machine learning.

The RF and GTB models achieved the best results, with the only difference being that the GTB model stood out by achieving a Kappa value for the lime category that was 3% higher than RF, and by producing values closer to reality when compared to the hectare data reported by SIAP.

Although several classification models integrated in GEE were evaluated in this study, due to the limitations encountered, it is recommended that future research explore different models such as Naive Bayes, Minimum

Distance, and the k-NN classifier, or consider the use of Deep Learning with other tools.

This research represents a relevant contribution to improving the accuracy of results, bringing them closer to actual values and enhancing their reliability for future applications beyond this case study. It also shows potential for broader use in areas such as deforestation monitoring, forest mapping, and other related fields, it also has potential for use in areas such as deforestation monitoring, forest mapping, among others. Promoting preventive analysis and collaboration among agro-environmental stakeholders through accurate information

References

1. **Secretaría de Agricultura y Desarrollo Rural. (2020).** Limón mexicano, único en el mundo. Gobierno de México.
2. **Servicio de Información Agroalimentaria y Pesquera (SIAP) (2025).** Anuario estadístico de la producción agrícola. Servicio de Información Agroalimentaria y Pesquera.
3. **Robles-González, M.M., Orozco-Santos, M., Manzanilla-Ramírez, M.Á., Velázquez-Monreal, J.J., & Carrillo-Medrano, S.H. (2017).** Efecto del HLB sobre el rendimiento de limón mexicano en Colima, México. *Revista Mexicana de Ciencias Agrícolas*, Vol. 8, No. 5, pp. 1101–1111. DOI: 10.29312/remexca.v8i5.111.
4. **Organización de las Naciones Unidas para la Alimentación y la Agricultura (FAO) (2019).** Programa mundial del censo agropecuario 2020. Vol. 2, Directrices operativas. Organización de las Naciones Unidas para la Alimentación y la Agricultura.
5. **Environmental Systems Research Institute (Esri). (s.f.).** Satellite imagery. GIS Dictionary. Esri GIS Dictionary: Satellite Imagery
6. **EOS Data Analytics (2025).** Teledetección: tipos y aplicaciones de los sensores remotos. EOS Data Analytics. Teledetección: tipos y aplicaciones de los sensores remotos
7. **Nicolis, O., Gonzalez, C. (2021).** Wavelet-based fractal and multifractal analysis for detecting mineral deposits using multispectral images taken by drones. In *Methods and Applications in Petroleum and Mineral Exploration and Engineering*, pp. 295–307. DOI: 10.1016/B978-0-323-85617-1.00017-5.
8. **Environmental Systems Research Institute (Esri) (2025).** Multispectral image. GIS Dictionary.
9. **EOS Data Analytics (2023).** Panchromatic imagery and its band combinations in use. EOS Data Analytics. Panchromatic and Pansharpened Satellite Imagery.
10. **Environmental Systems Research Institute (Esri) (2025).** Panchromatic image. GIS Dictionary.
11. **eoPortal Directory (2025).** Hyperspectral imaging. eoPortal Directory.
12. **Environmental Systems Research Institute (Esri) (2025).** Hyperspectral imagery. GIS Dictionary.
13. **NASA Earthdata (2025).** Synthetic Aperture Radar (SAR). NASA Earthdata.
14. **Hajkova, Z. (2024).** What is SAR imagery? Introduction to Synthetic Aperture Radar. European Space Imaging.
15. **Comisión Europea (2025).** Sobre Copernicus. Copernicus.
16. **Comisión Europea (2025).** Copernicus en pocas palabras. Copernicus.
17. **Comisión Europea (2025).** Infraestructura. Copernicus.
18. **Comisión Europea (2025).** Estos son nuestros satélites. Copernicus.
19. **Copernicus Data Space Ecosystem (2025).** Sentinel-2. Copernicus Data Space Ecosystem.
20. **Wang, Y., Sun, Y., Cao, X., Cheng, X., Wang, Y., Zhang, W. (2023).** A review of regional and global scale land use/land cover (LULC) mapping products generated from satellite remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 206, pp. 311–334. DOI: 10.1016/j.isprsjprs.2023.11.014.
21. **Kanja, K., Zhang, C., Lippe, M., Atkinson, P. M. (2025).** Mapping tropical dry Miombo woodlands into functional forest classes

- using Sentinel-1 and Sentinel-2 imagery and machine learning. *Remote Sensing Applications: Society and Environment*, pp. 101615. DOI: 10.1016/j.rsase.2025.101615.
22. **Shah, S. V., Gaikwad, S. V., Vibhute, A. D. (2025).** Land use land cover mapping of Kolhapur District using machine learning and Sentinel-2 satellite imagery. *Procedia Computer Science*, Vol. 257, pp. 249–258. DOI: 10.1016/j.procs. 2025.03.034.
 23. **Ouchra, H., Belangour, A., Erraissi, A. (2023).** Machine learning algorithms for satellite image classification using Google Earth Engine and Landsat satellite data: Morocco case study. *IEEE Access*, Vol. 11, pp. 71127–71142. DOI: 10.1109/ACCESS.2023.3293828.
 24. **Alciaturi, G., Wdowinski, S., García-Rodríguez, M.P., Fernández, V. (2025).** Seasonal land use and land cover mapping in South American agricultural watersheds using multisource remote sensing: The case of Cuenca Laguna Merín, Uruguay. *Sensors*, Vol. 25, No. 1, pp. 228. DOI: 10.3390/s25010228.
 25. **Tan, Y.C., Duarte, L., Teodoro, A.C. (2024).** Comparative study of Random Forest and Support Vector Machine for land cover classification and post-wildfire change detection. *Land*, Vol. 13, No. 11, pp. 1878. DOI: 10.3390/land13111878.
 26. **Sheykhou, 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*, Vol. 13, pp. 6308–6325. DOI: 10.1109/JSTARS .2020.3026724.
 27. **Ashfaq, M., Khan, I., Alzahrani, A., Tariq, M. U., Khan, H., Ghani, A. (2024).** Accurate wheat yield prediction using machine learning and climate-NDVI data fusion. *IEEE Access*, Vol. 12, pp. 40947–40961. DOI: 10.1109/ACCESS.2024. 3376735.
 28. **Zhang, X., Li, Y., Wang, J., Chen, H., Liu, Z., Zhao, M. (2025).** Study on the extraction of topsoil-loss areas of cultivated land based on multi-source remote sensing data. *Remote Sensing*, Vol. 17, No. 3, pp. 547. DOI: 10.3390/rs17030547.
 29. **Liu, P., Zhuang, W., Kou, W., Wang, L., Wang, Q., Deng, Z. (2025).** Fire severity outperforms remote sensing indices in exploring post-fire vegetation recovery dynamics in complex plateau mountainous regions. *Forests*, Vol. 16, No. 2, pp. 263. DOI: 10.3390/f16020263.
 30. **Yu, L., Zhang, J., Li, X., Wang, Y., Zhao, Y., Liu, B. (2024).** Identification and monitoring of irrigated areas in arid areas based on Sentinel-2 time-series data and a machine learning algorithm. *Agriculture*, Vol. 14, No. 10, pp. 1693. DOI: 10.3390/agriculture 14101693.
 31. **Morell-Monzó, S., Sebastiá-Frasquet, M.-T., Estornell, J., Moltó, E. (2023).** Detecting abandoned citrus crops using Sentinel-2 time series: A case study in the Comunitat Valenciana Region. *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 201, pp. 54–66. DOI: 10.1016/j.isprsjprs. 2023.05.003.
 32. **Breiman, L. (2001).** Random forests. *Machine Learning*, Vol. 45, No. 1, pp. 5–32. DOI: 10.1023/A:1010933404324.
 33. **Cortes, C., Vapnik, V. (1995).** Support-vector networks. *Machine Learning*, Vol. 20, No. 3, pp. 273–297. DOI: 10.1007/BF00 994018.
 34. **Romanelli, A., Palmese, S., De Vita, S., Calicchio, A., Gammaldi, R. (2024).** Stratifying mortality risk in intensive care: A comprehensive analysis using cluster analysis and classification and regression tree algorithms. *Intensive Care Research*, Vol. 4, pp. 116–128. DOI: 10.1007/s44231- 024-00064-9.
 35. **Krueger, K., Mueller, F., Gebhardt, P., Weissler, B., Schug, D., Schulz, V. (2023).** High-throughput FPGA-based inference of gradient tree boosting models for position estimation in PET detectors. *IEEE Transactions on Radiation and Plasma Medical Sciences*, Vol. 7, No. 3, pp. 253–262. DOI: 10.1109/TRPMS.2023.3238904.

36. **Google Developers (2024).** Google Earth Engine. Google Developers.
37. **The MathWorks, Inc. (2025).** Spectral indices. MathWorks. Spectral Indices – MATLAB & Simulink.
38. **Environmental Systems Research Institute (Esri) (2025).** NDVI. ArcGIS Pro.
39. **Environmental Systems Research Institute (Esri) (2025).** GNDVI. ArcGIS Pro.
40. **Wang, L., Qu, J.J. (2007).** NMDI: A normalized multi-band drought index for monitoring soil and vegetation moisture with satellite remote sensing. *Geophysical Research Letters*, Vol. 34, No. 20. DOI: 10.1029/2007GL031021.
41. **Salas, J., Sepúlveda, R., Vera, P. (2025).** Water clarity assessment through satellite imagery and machine learning. *Water*, Vol. 17, No. 2, pp. 253. DOI: 10.3390/w17020253.
42. **Abraira, V. (2001).** El índice kappa. *Medicina de Familia. SEMERGEN*, Vol. 27, No. 5, pp. 247–249.
43. **Landis, J.R., Koch, G.G. (1977).** The measurement of observer agreement for categorical data. *Biometrics*, Vol. 33, No. 1, pp. 159–174. DOI: 10.2307/2529310.
44. **Microsoft. (2023).** Resultados de los modelos de aprendizaje automático. Microsoft Learn. Evaluación de los resultados de los experimentos de aprendizaje automático automatizado.
45. **Murel, J., Kavlakoglu, E. (2024).** ¿Qué es la detección de objetos? IBM.
46. **Instituto Nacional de Estadística y Geografía (INEGI) (2025).** México en cifras. Instituto Nacional de Estadística y Geografía.
47. **Copernicus Data Space Ecosystem (2024).** Sentinel-2. Copernicus Data Space Ecosystem.

Article received on 06/12/2025; accepted on 16/01/2026.

**Corresponding author is César Villaseñor-Sotomayor.*