

# Detecting the Attack of the Fall Armyworm (Spodoptera Frugiperda) in Cotton Plants with Machine Learning and Spectral Measurements

Felipe David Georges Gomes<sup>1</sup>, Mayara Maezano Faita Pinheiro<sup>1</sup>, Danielle Elis Garcia Furuya<sup>1</sup>, Wesley Nunes Gonçalves<sup>2</sup>, José Marcato Junior<sup>2</sup>, Mirian Fernandes Furtado Michereff<sup>3</sup>, Maria Carolina Blassioli-Moraes<sup>3</sup>, Miguel Borges<sup>3</sup>, Raúl Alberto Alaumann<sup>3</sup>, Veraldo Liesenberg<sup>4</sup>, Lúcio André de Castro Jorge<sup>5</sup>, Ana Paula Marques Ramos<sup>1</sup>, and Lucas Prado Osco<sup>6</sup>✉

<sup>1</sup>Post-Graduate Program of Environment and Regional Development, University of Western São Paulo (UNOESTE), Rodovia Raposo Tavares, km 572-Limoeiro, 19067-175, Presidente Prudente, São Paulo, Brazil

<sup>2</sup>Faculty of Engineering, Architecture, and Urbanism and Geography, Federal University of Mato Grosso do Sul (UFMS), Avenida Costa e Silva, 79070-900, Campo Grande, Mato Grosso do Sul, Brazil

<sup>3</sup>National Research Center of Development of Genetic Research and Biotechnology, Brazilian Agricultural Research Agency (EMBRAPA), Parque Estação Biológica W5 Asa Norte, 70770-917, Brasília, DF, Brazil

<sup>4</sup>Forest Engineering Department, Santa Catarina State University (UDESC), 88520-000, Av. Luiz de Camões, 2090, Conta Dinheiro, Lages - SC, Brazil

<sup>5</sup>National Research Center of Development of Agricultural Instrumentation, Brazilian Agricultural Research Agency (EMBRAPA), Rua XV de Novembro, 1452, 13560-970, São Carlos, São Paulo, Brazil

<sup>6</sup>Faculty of Engineering and Architecture and Urbanism, University of Western São Paulo (UNOESTE), Rodovia Raposo Tavares, km 572-Limoeiro, 19067-175, Presidente Prudente, São Paulo, Brazil

In cotton cultivars, an insect that causes irreversible damage is the *Spodoptera frugiperda*, known as the fall armyworm. Since the visual detection of plants is a burdensome task for human inspection, the spectral information related to plant damage, registered on a spectral scale, can be useful. These measurements, associated with machine learning techniques, produce useful information for a rapid and non-invasive inspection method development. To contribute to this gap fulfillment, this paper proposes a machine learning framework to model the spectral response of cotton plants under the attack of the fall armyworm. Additionally, a theoretical model is presented, built from the results of the machine learning analysis, to infer this damage with up-to-date orbital sensors. The data was composed of the reflectance measurements collected at a cotton field with control plants and plants submitted to *Spodoptera frugiperda* damage. Their spectral response was recorded with a hand-held spectroradiometer ranging from 350 to 2,500 nm, for eight consecutive days. Different machine learning models were evaluated and the overall best model was defined by accuracies comparisons on a testing-set. A ranking approach was adopted based on the model accuracy, returning the most contributive wavelengths for the classification. Sequentially, an unsupervised neural network (Self-Organizing Map - SOM) was implemented to reduce data-dimensionality and assist in the definition of important spectral regions. The regions were associated with the spectral bands of the two sensors (OLI and MSI) and a theoretical model using a band simulation process with the overall best machine learning model was proposed. The results indicated that the Random Forest (RF) algorithm is the most suitable to predict cotton-plants damaged by insects and that the last day of analysis (8th day) was better to separate it, with F-measure equals 0.912. The ranking approach combined with the SOM method indicated the spectral regions at the red to near-infrared (650 to 1,350 nm) and shortwave infrared (1,570 to 1,640 nm) as the most important regions to the analysis. The proposed theoretical model simulated with the OLI and MSI sensor-bands returned an F-Measure of 0.865 and 0.886, respectively. In conclusion, this framework can be used to map cotton-plants under insect-attack. The theoretical model presents high accuracy to

infer the insect-damaged on cotton plants based on multispectral bands from other sensors, being a useful tool for future research that intends to evaluate it in other areas and at different field scales.

machine learning | insect-damage | spectral data | theoretical model

Correspondence: [lucasosco@unoeste.br](mailto:lucasosco@unoeste.br)

## Introduction

Agronomic crops around the world are susceptible to diseases and insect damage. Attacks of pests in plants are some of the issues that most impact crop yield rates (1, 2). Diseases in plants implicate some form of physiological modification that disturbs their normal processes for healthy development (2). Because agriculture is the basis for many worldwide economies, a requirement for all producer countries is developing effective strategies to monitor and manage their crops. In Brazil, the agriculture sector is responsible for an important part of the country's gross domestic product (GDP). According to (3), GDP reached 1.6 trillion in 2019, representing a growth of 1.1% from 2018. This value is related to an increased agriculture sector, which was 1.3% in later years, with the following crops presenting the most annual outgrowth: cotton (39.8%), orange (5.6%), and bean (2.2%) (3).

Brazil is one of the main cotton fiber producers and exporters in the world (4). The estimated planted area with cotton cultivar in 2019/2020 in this country is approximately 1,665 mil ha, and the last harvest reached a record production with more than 7,372 thousand Mg of seed cotton, which implies a yield rate of 1,774 kg.ha<sup>-1</sup>. The estimate is an export of 1.9 million tons in 2020, an amount 18% higher than the 1.6 million tons shipped in 2019, which was a record. Mato Grosso (MT) and Bahia (BA) are the two main producers states in Brazil, farming more than 88% of the cotton area of the current harvest (4).

To maintain, or even to increase the cultivar production rates, by performing the correct management of crop-fields, farmers are adopting precision agriculture practices associated with remote sensing approaches. Remote sensing data has already been used to investigate several agronomic-related problems for a long time. Among them, it is possible to highlight crop yield prediction for different cultivars, such as sugar beet (5) and maize (6); nutrient deficiencies estimative, like nitrogen and phosphorus in winter wheat (7), and nitrogen in citrus (8); disease detection (9, 10); and diseases and insect pests management (11, 12). In recent years, a type of puissant processing technique that has been gaining attention in both remote sensing and precision agriculture communities, including to monitor crop diseases and pests, is machine learning analysis.

Machine learning is inherently a multidisciplinary field that draws on results, among others, from artificial intelligence, probability and statistics, and mathematical and computational theory, that can be applied to model different types of data (13). Algorithms of machine learning are generally classified as supervised and unsupervised, and their main characteristic is extracting patterns in a dataset based on the learning attributes of a given sample (14). Machine learning methods are capable of analyzing hierarchical and nonlinear relationships between the independent variables and the dependent variable, usually resulting in better performance compared with conventional linear regression models (15, 16).

Several recent studies evaluated the contribution of machine learning to model remote sensing data regarding plant analysis (1, 2, 10, 17–21). Machine learning algorithms play an important role in detecting plant diseases and pests (1). For example, (22) applied several machine learning algorithms to detect caterpillar (*Spodoptera littoralis*) attacks in cotton plants implemented at a greenhouse experiment and achieved a prediction accuracy of 84%. Also related, other studies modeled the spectral response of cotton plants under different grades of aphid infestation, at leaf level (23), as well as in attacks associated with mealybug infestation (24, 25).

As demonstrated by the literature (26, 27), in the context of remote sensing, plants infected by diseases and pests present differences in their spectral response compared to healthy plants. This is because diseases modify both the photosynthetic process and generate changes in the internal structures of the leaves (27). These differences in reflectance response of damage and healthy plants may be mapped particularly using the visible, near-infrared, and medium-infrared regions of the electromagnetic spectrum (1, 28). In this sense, an important contribution is indicating the most related spectral regions to predict a particular kind of insect in plants. For cotton crops, the herbivory-insect attacks represent a serious threat. Herbivore-insects can affect cotton in different phenological stages, such as in the vegetative phase, consuming the leaves (Pascua and Pascua 2002), or in the reproductive phase, feeding on flower buds, flowers, and bolls (29). It has been reported that they can affect the amino acid composition, water content, and the oxidative status of the cotton (30–32).

As aforementioned, few studies evaluated the insect-herbivore damage in cotton plants with machine learning and spectral reflectance measurements. Moreover, the identification of the most related spectral regions to differentiate damage and healthy plants can benefit from these algorithms. Another possibility to large scaling mapping is the proposal of a theoretical model to evaluate this type of insect-damage effect with other sensors, at different scales, like those embedded in aerial and orbital platforms. In particular, two up-to-date orbital sensors, known to be frequently used for agriculture mapping, are the OLI, embedded in Landsat 8, and the MSI, embedded in Sentinel 2. The effectiveness of these sensors to explore agronomic-related variables has been already revised (1, 33, 34), and simulation processes initially conducted at a terrestrial scale and later converged at aerial or orbital is also something validated by previous studies in other issues (8, 35). However, for the agricultural context, this type of proposal is, still, underexplored.

The main idea behind this approach is to offer a rapid and unbiased method to detect insect-damage in plantations and assist precision agriculture practices in crop management. In cotton plants, an insect that causes irreversible damage is the *Spodoptera frugiperda*, known as the fall armyworm. Thus, this paper proposes a framework for modeling the spectral response of insect-herbivore damage in cotton plants based on a machine learning approach. Moreover, as an additional contribution, theoretical models were developed for both Landsat-8 /OLI and Sentinel-2/MSI systems, capable of simulating the spectral behavior of a cotton plant under the fall armyworm attack in these satellite-images. These models were implemented adopting the band simulation process (8) using results of a ranking-SOM (Self-Organizing Map) approach. This framework may be useful for future studies that intends to implement a proximal sensing survey in insect-damaged plant analysis, as well as novel research that aims to evaluate it with multispectral sensors embedded in aerial or orbital platforms.

## Materials and Method

The proposed framework was divided into the following phases: 1) data acquisition and organization; where the experiment was conducted, and the spectral wavelengths were registered and converted into reflectance factor values, divided it into training/validation and testing subsets, and submitted to different learning models; 2) machine learning evaluation; where the performances of multiple models were compared by measuring their robustness in relation to a day-to-day analysis, and applying the ranking approach combined with a clustering method to identify the most contributive spectral wavelengths and regions; 3) theoretical model application, where a band simulation process was used to indicate the expected spectral behavior of a cotton plant affected by the fall armyworm based on the information extracted within the ranking and clustering approach. The summarization of the proposed method is displayed in Figure 1.

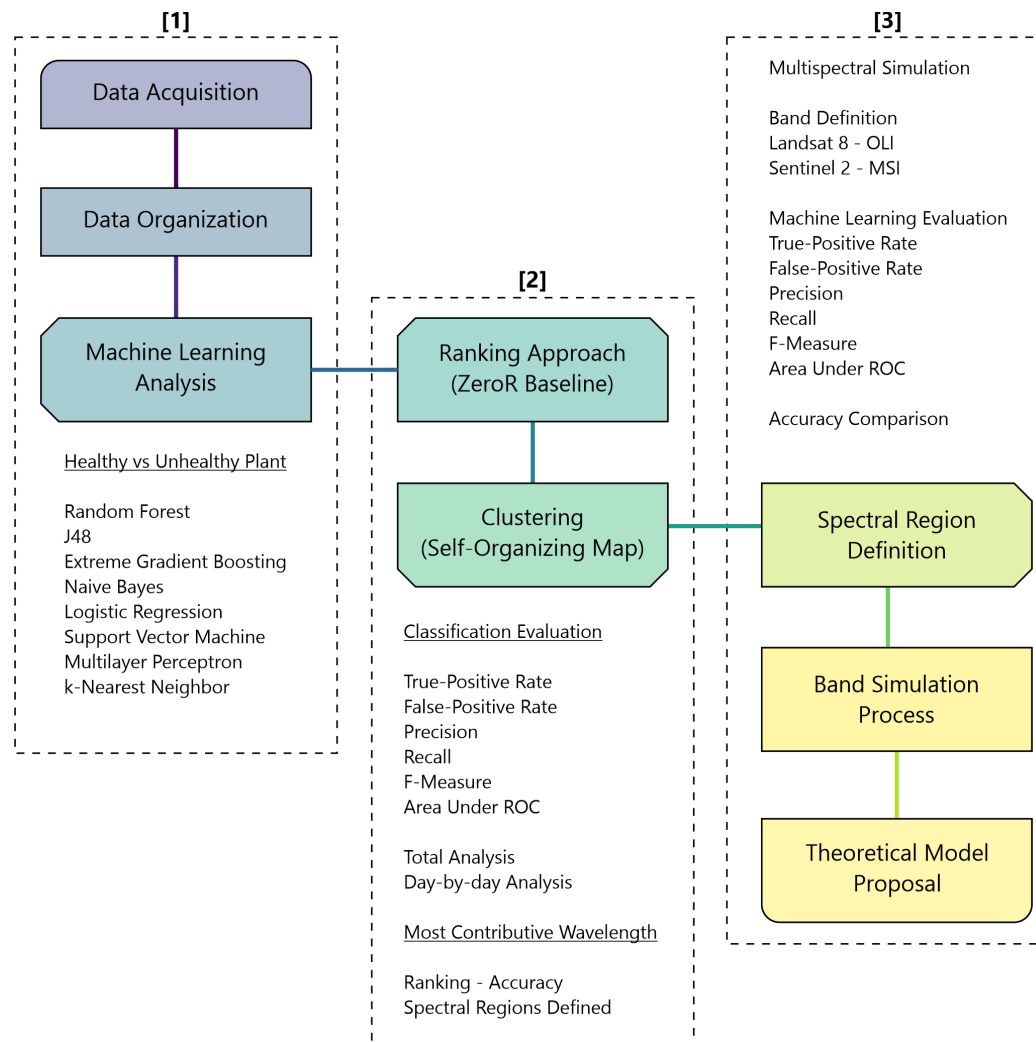


Fig. 1. The workflow implemented in this study.

**Spectral Data Acquisition and Organization.** Plants and insects were grown and reared at the Embrapa Recursos Genéticos e Biotecnologia station, in Brasília, DF, Brazil. Plants were grown in experimental plots and the fall armyworms were reared in rooms of an insectary at  $25 \pm 1$  Celsius degrees,  $60 \pm 10\%$  RH, and 14:10 L:D. The cotton (var. Delta Opal) plants used in the experiments were 6-week-old at the vegetative stage (with up to 6 expanded leaves and about 30 cm high). The insect reared in the experiment was the *Spodoptera frugiperda* larvae (fall armyworm). They were reared in plastic containers on an artificial diet based on beans (*Phaseolus vulgaris*), and the adults were maintained in 3 L plastic containers with the walls lined with paper for oviposition. Adults were fed with a sugar solution prepared from 1 L of water, 50 g of honey, 50 g of sugar, 1 mg of Nipagin, and 1 mg of ascorbic acid, following the instructions proposed at (31)

For the spectral reading of the plants, the insects were removed and measures of both damaged and undamaged (healthy/control) plants were performed, with a direct exposition of the plants to the sunlight. The spectral measurements were conducted with a handheld spectroradiometer

ASD FieldSpec 3 high-resolution (Analytical Spectral Devices Inc., Boulder, USA). The calibration of the instrument was performed using a Spectralon white reference panel, immediately before the first measurement. The procedure was periodically repeated during the measurement process. This equipment operates within the spectral range of 350–2,500 nm (VIS-SWIR) at the spectral resolution of 1.4 nm in between 350 and 1,000 nm and 2 nm in between 1,000 and 2,500 nm. The apparatus was connected to a portable microcomputer, where the data were stored after they were obtained.

The reflectance spectrum ( $\rho$ ) was computed by dividing the amount of radiance reflected from the target (LT) of interest (i.e. cotton plant), in the specific wavelength interval, by the amount of radiance reflected in this same wavelength interval from a reference sample (Lr), multiplying this result by a correction factor (K) related to the ratio of solar irradiance to reference plate exitance (27). The equation (1) was used to perform the signal convention is presented below:

$$\rho_T = \frac{L_T}{L_r} \times K \quad (1)$$

After that, each spectral reflectance measurement was orga-

nized to be used as input features over the machine learning models. An additional label corresponding with the ground-truth (health or damaged cotton plant) was also created. It should be emphasized that the two main affected spectral regions were removed from the spectra dataset due to the absorption of water vapor on the atmosphere (27), as well as the final spectral wavelengths related to low-to-signal noise of the spectroradiometer. These regions corresponded to, respectively: 1,350-1,410 nm and 1,820-1,940 nm, and 2,460-2,500 nm, resulting in 1,934 variables (spectral wavelengths) to be used as input in machine learning models. As spectral measurements were conducted within different hours from each other (associated with another purpose from this paper), only those collected beyond 09:00 and before 15:00 were evaluated. This interval is recommended to collect spectral measurements considering a more regular solar irradiance upon plants (27). The experiment was monitored for eight consecutive days and the spectral measurements were extracted according to each day of analysis.

**Data Processing and Machine Learning Evaluation.** For modeling the spectral response of caterpillar-damage in the cotton plants, an analysis based on multiple machine learning methods was implemented. The algorithms used in this classification problem were: Random Forest (RF) (36); Decision Tree J48 (37); Extreme Gradient Boosting (XGB) (38); Naive Bayes (NB) (39); Logistic Regression (LoR) (40); Support Vector Machine (SVM) (41); Multilayer Perceptron (MLP) (42) and; k-Nearest Neighbor (kNN) (43). The algorithms were implemented through the open-source Weka 3.9.4, with the integration of libraries from R and Scikit-Learn. Additionally, meta classifiers were evaluated and different search algorithms were also tested into an experimental phase, like Random Search, Multi-Search, and Grid-Search. Nevertheless, the accuracy gain per performance was significantly small, increasing the computational cost of each training model. Because of that, the baseline values of the algorithms were maintained, using their respective library default parameters.

Decision tree-based algorithms like RF and J48 were set with similar construction characteristics in relation to the number of nodes and leaves at each tree. The J48 was configured to generate a pruned tree with the C4.5 pruning method, and the RF was configured to compute 100 decision trees and a bagging-size equal to 100. The LoR model was implemented with a Ridge log-likelihood of  $10e-8$ . The SVM was set as a C-SVC type, with the Kernel radial basis function, and epsilon and gamma values of, respectively, 0.001 and  $\exp(-\gamma \cdot \|u-v\|^2)$ . The kNN algorithm was configured to adopt a Euclidean distance metric and a number of neighbors equal to 1. The MLP used one hidden layer with 200 neurons, a learning rate and momentum equal to 0.05 and 0.1, respectively, the Adam solver and a SoftMax function. The XGB and NB learners were executed with their respective default values.

A repetitive evaluation of the training performance of each model was performed. For this, the training subset was evaluated with a cross-validation approach in 10 folders. The

cross-validation method was repeated 10 times per model, which means that 100 validations were constructed from the repetitions. This was necessary to ensure that the final model, built with the training subset, was not biased and could produce accurate results in relation to the testing subset. With the 100 validations, the interquartile values were calculated, which were later used for the mentioned comparison with the testing subset, and were also able to generate an averaged accuracy of the training process. This type of evaluation was performed considering firstly the dataset with data accumulated from all of the measured days. The same method was applied for the dataset considering the days individually.

Later, the model's prediction accuracy was evaluated in the testing subset samples. The testing sample consisted of 10% of the original dataset and was composed of data unknown to the algorithms, i.e. data not used neither for training nor validation. For both training and testing evaluation, the following classification metrics were used (14): True-Positive Rate (TP); False-Positive Rate (FP); Precision (P); Recall (Re) and F-measure (F). Since F-measure consists of the harmonic mean between Recall and Precision, it is a reliable metric to indicate the overall robustness of the classification. In this regard, the F-measure values were used for the comparison between the algorithms' performance, returned by the multiple repetitive training/validation process. Initially, a Shapiro-Wilk test was performed, followed by the Mann-Whitney pairwise comparison. These statistical tests were conducted at a confidence interval of 95%.

Once defined the overall best learner and measured-day for the classification task, a ranking approach followed by a clustering algorithm was applied. The ranking approach is based on a Metric score returned from a direct comparison between the accuracy returned by the model against the accuracy obtained with a baseline algorithm. This comparison consists of the difference between each other. The baseline algorithm used for this comparison is the ZeroR, since it uses the average value of the target variables and applies it to the prediction. The implementation of the ZeroR as a baseline into the ranking approach, considering the accuracy values, is based on a previous study where it evaluated the impact of variable reduction on the models' performance (20). Here, however, this approach is considered after the models' application to identify the most important spectral wavelengths to model the insects' attacks on cotton crops.

After identifying the most contributive wavelengths into the overall best model, a clustering approach based on the Self-Organizing Map was used to reduce visualization in data-dimensionality and help to organize the spectral wavelengths evaluated into groups. As the SOM method is used for reducing and, thus, representing highly dimensional data into a lower dimension, as well as preserving the respective topological characteristics of the dataset (44, 45), in this case, the wavelengths, it can create sequential clusters considering both wavelengths orders and their related Metric score. In this sense, this method was used to help in isolating spectral regions of interest associated with the most contributive wavelengths in the model. This approach was necessary to



improve the detection of spectral regions to be used by the theoretical model proposed.

**Band Simulation and Theoretical Model Proposal.** With the most contributive spectral regions defined in the previous step, the spectral regions of two up-to-date multispectral sensors were used as an example for the theoretical model proposal. In this aspect, spectral sensitivity filters of both Landsat-8/OLI and Sentinel-2/MSI sensors (Figure 2) were used in this simulation. The simulation process consists of a transformation from a known reflectance value, related to the target of interest, at different wavelengths, as if they were recorded at the same intervals of another sensor. To simulate the value of each band, the equation bellow was applied:

$$R_{Sim}(\lambda_i) = \frac{\sum \lambda HCRF(\lambda) * SRF(\lambda)}{\sum \lambda SRF(\lambda)} \quad (2)$$

where:  $HCRF(\lambda)$  corresponds to the reflectance value obtained with the  $\rho T$  (equation 1) for each wavelength, and the  $SRF(\lambda)$  is the spectral response function obtained for each sensor.

To determine the most proximal value of the in-field spectral measurement into the same interval as the multispectral sensors, the recommendations of (46) (OLI sensor), and (47) (MSI sensor) were followed. Additionally, both NASA (National Aeronautics and Space Administration) and ESA (European Space Agency) released spreadsheets containing information from their respective spectral bands at their respective webpages: NASA<sup>1</sup> and ESA<sup>2</sup>. These spreadsheets contain information related to the per-wavelength, band-averaged, and in-band relative spectral response. Both sensors' spectral wavelengths response function was sampled at 1nm intervals, and information related to their respective standard deviation of each measurement is specified. The simulation process was then conducted similarly as reported in previous studies (8, 35).

As aforementioned, the band simulation process occurred after the determination of the most contributive spectral regions obtained by the combination between the ranking approach and the SOM method. Therefore, not all spectral bands of the OLI and MSI sensors were simulated here, but only the bands related to the predefined regions. Once these bands were simulated, the spectral response of every measured target was calculated as if they were measured by the multispectral sensor itself. This assured that these measures maintained the same quantity of samples used in the previous experiments. With these new variables, the overall best learner, which was defined in the previous step, was adopted. Here, the same configuration set using the training and validation approach with 10 consecutive runs was used, and the real performance in the testing set was calculated. In this step, the same classification metrics used for the spectral modeling phase were calculated, and the ROC curve and its metrics were evaluated for both classes: healthy cotton plants and insect-herbivore damaged cotton plants.

## Results and Discussion

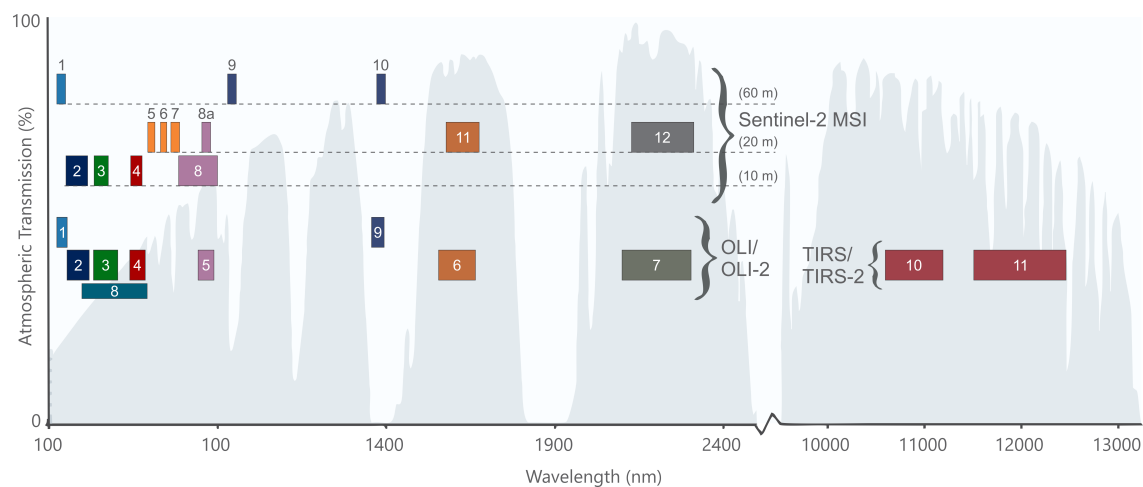
The results of the proposed framework are organizing into two segments: The first section presents the results returned by the spectral response modeling with the machine learning algorithms, and; the second section it is demonstrated the outcomes of the proposed approach to select spectral regions and to suggest a theoretical model for the two selected multispectral sensors.

**Spectral Response Modeling.** To assist in understanding the implications of the subsequent results, the averaged spectral wavelengths of both groups, healthy (control) and unhealthy (damaged) cotton plants, were plotted (Figure 3). When considering both average and standard deviation of the entire measured spectral response, it is possible to observe that spectral wavelength located at the near-infrared regions (between. 700 and 1,350) nm present the highest differences in amplitude. This is followed by wavelengths at the short-wave infrared region (at 1,550 to 1,700 nm). The regions situated at the visible range and the second shortwave infrared region (from 2,000 to 2,450 nm) did not return high amplitudes as differences. This potentially indicates which regions are prominent to separate a healthy cotton plant from an insect-herbivore-damaged cotton plant. Still, since the standard deviation suggests that some spectral measurements from a class overlay the other, this may propose a hindrance to direct or linear methods.

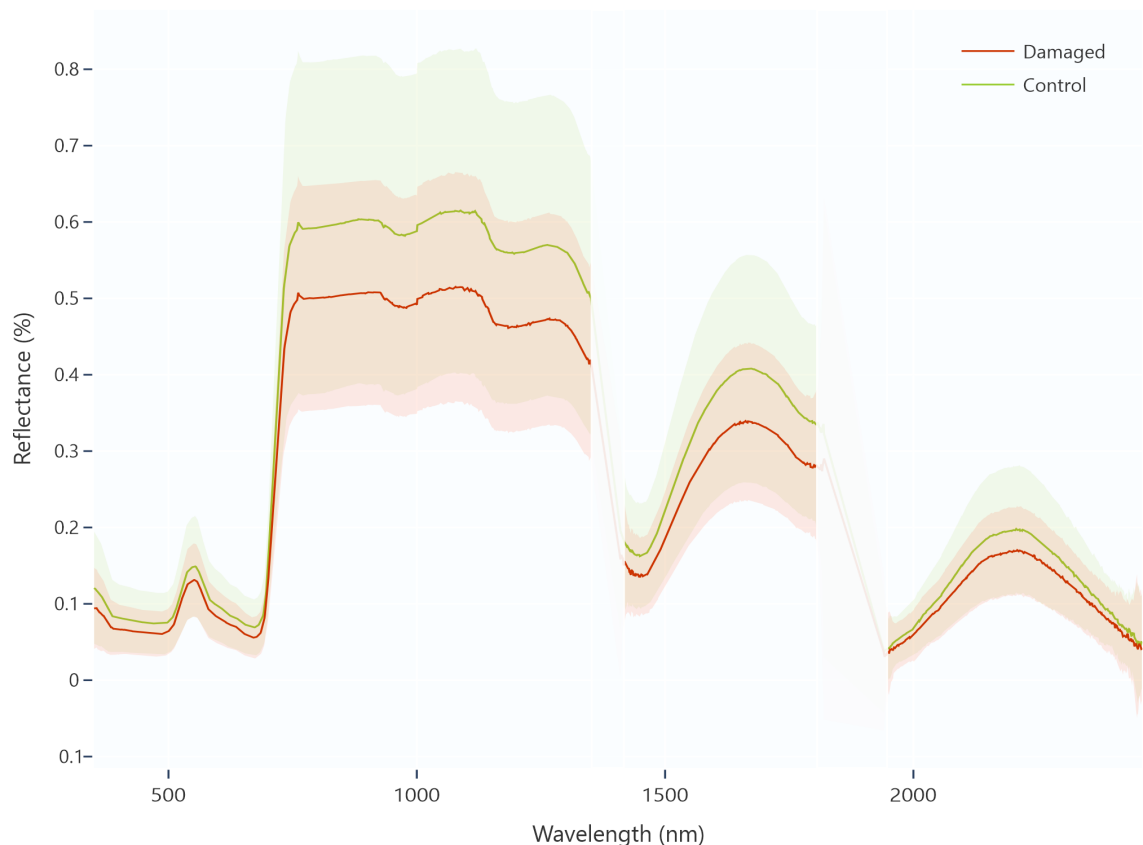
The spectral behavior of the evaluated cotton plants, under the fall armyworm attack, suggests that an overall reduction in reflectance occurred. This behavior was also observed in similar research that evaluated, with spectral vegetation indices, the spectral divergence between cotton plants under different grades of aphid infestation at leaf level (23). Cotton plants also displayed this kind of behavior in another type of insect-attack, provoked by mealybug infestation (24, 25). The mentioned investigations were able to ascertain the relationship between spectral data and the damage that occurred on the plants. However, in this paper, the proposed approach was conducted on a multiple-day analysis, and for that, more robust techniques than the traditional statistical methods, based on machine learning, were necessary.

The overall evaluation of the machine learning algorithms indicates that the RF and kNN learners are the most robust methods to model the spectral behavior of cotton plants under the armyworm attack (Figure 4). They were able to separate them with satisfactory performance in relation to other models in both training and validation sequences (Figure 4) and testing (Table 1). At the training phase, the kNN algorithm returned higher F-measures values than RF, and this also occurred at the testing-set, which was considered as the final evaluation since it incorporates spectral wavelengths not used neither for training nor validation of the models (Table 1). In this sense, the kNN can be considered the most appropriate algorithm to model the insect-damage attack in the cotton plants when considering data from different consecutive days of analysis.

Considering the entire dataset, both NB and MLP returned



**Fig. 2.** The spectral bands of Sentinel 2 - MSI and Landsat 8 - OLI sensors (adapted from: NASA, 2020). Averaged spectral wavelengths considering all the health and insect-damaged cotton plants

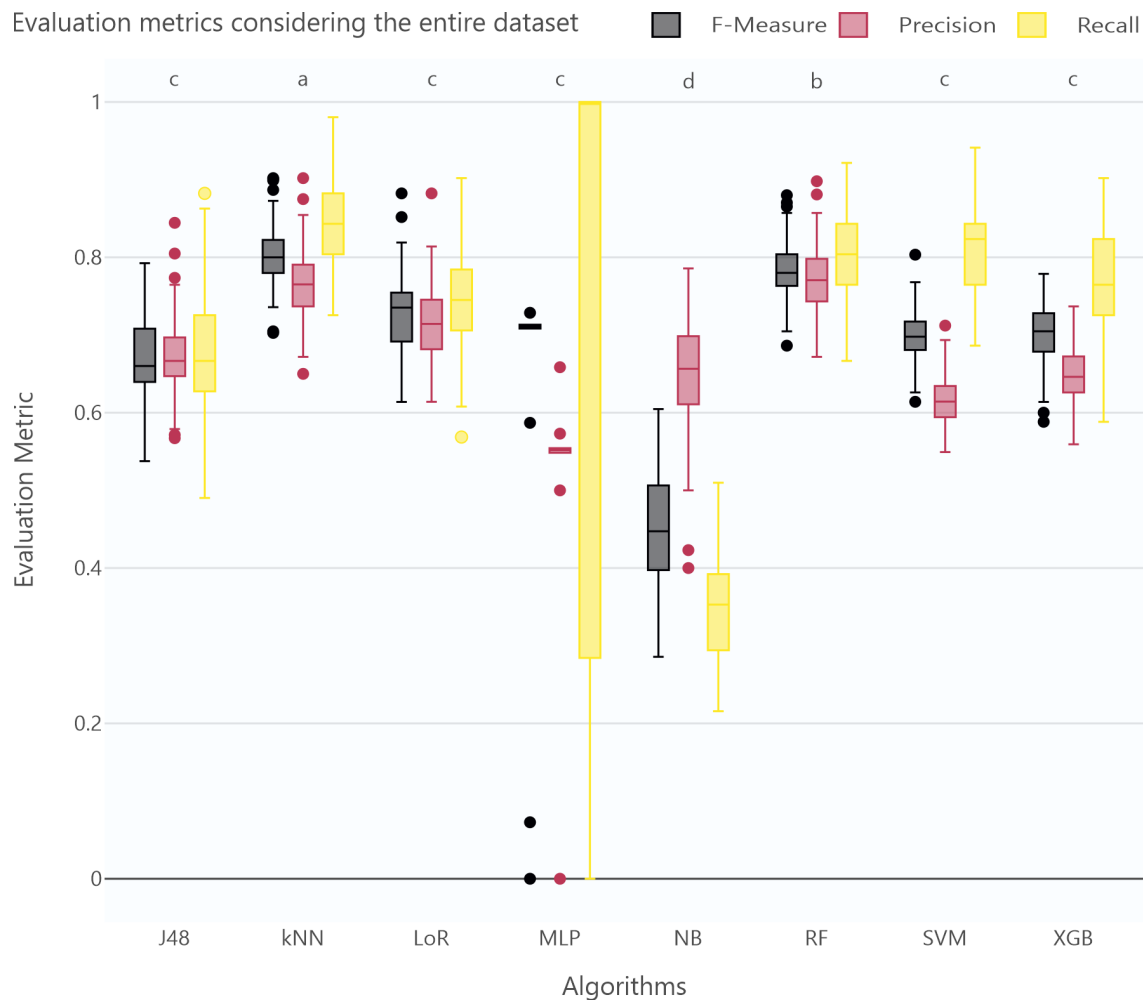


**Fig. 3.** Averaged spectral wavelengths for both the healthy cotton plants and the caterpillar attacked plants.

the worst performances in predicting plants’ health and unhealthy. While NB presented a more “appropriate” response, the MLP resulted in an unexpected behavior by producing low Precision and highly variegated Recall values. This may be due to an imbalanced classification in relation to one of the classes, returning high prediction values for one type of class and low or zero values for the other class. This may be a characteristic associated with the dataset itself, as a similar MLP was used for other previous classification tasks in agricultural-related problems (18) and returned satisfactory results. The remaining algorithms (J48, XGB, LoR, and

SVM), although resulted in interesting performances (Table 1), were not able to withstand the RF and kNN algorithms. These algorithms, especially RF, presented high accuracies in other situations regarding spectral and multispectral data (48–50), which indicates the robustness of these learners for similar tasks.

To determine the predictions according to the day of analysis, the classification task was performed with the same algorithms at the different days, from the initial moment (day 1), where few infestations were reported, to the last day (day 8), where the insect was already speeded throughout the exper-



**Fig. 4.** Training performance of the algorithms when considering the entire dataset composed of spectral wavelengths acquired during eight consecutive days. Letters above the box-plots of each algorithms' result relate to a statistical difference between the F-Measure metric.

imental plantations' fields. This analysis confirmed an expected result, in which, for most predictions, the final day of analysis was the optimal moment to separate the insect-damaged from the healthy cotton plant (Figure 5). Similar behavior to the entire dataset analysis (Figure 4) was noticed in this analysis in relation to the performances of the algorithms, where the kNN and RF returned the best overall predictions. LoR also presented high F-measures, followed by XGB, JB48, and, at a lower level, NB. MLP was still the worst model generated for this task. The F value from each algorithm was compared. The Mann-Whitney pairwise com-

parison indicated that no statistical difference at p-value under 0.05 occurred for the kNN, RF, and LoR algorithms.

During the testing phase, the RF algorithm presented the overall best metrics for the classification task (Table 2), which differentiates from the complete dataset analysis (Table 1) where kNN was better. This may be an indicator of how the day of analysis, and therefore the severity of the insect-herbivory damage on the cotton plant, impacts the performance of the algorithms. An increase in performance was evident since the F-measure values went from 0.775 (day 1 - RF) to 0.950 (day 8 - RF) (Figure 5). Although the 8th day can be considered the most adequate moment from this analysis (and thus, being incorporated into the next step of the framework for this very reason), it is important to consider that from day 5 the performance of the algorithms went from lower predictions (around 0.7 and 0.8) to higher values (0.9) for most algorithms. In real field conditions, where the earliest possible estimations of insect-herbivore damage are required to assist farmers in control practices, the combined information of proximal sensing and machine learning methods can ensure rapid and accurate data to be incorporated into decision-making.

Since the RF learner returned the highest evaluation metric

**Table 1.** Testing performance of the algorithms when considering the complete dataset.

Algorithm	TP	FP	Precision	Recall	F-Measure
RF	0.746	0.254	0.748	0.746	0.746
J48	0.617	0.383	0.618	0.617	0.617
XGB	0.634	0.393	0.631	0.634	0.627
NB	0.544	0.413	0.587	0.544	0.525
LoR	0.718	0.292	0.717	0.718	0.717
SVM	0.616	0.429	0.617	0.616	0.596
MLP	0.502	0.498	0.507	0.502	0.503
kNN	0.773	0.246	0.775	0.773	0.770

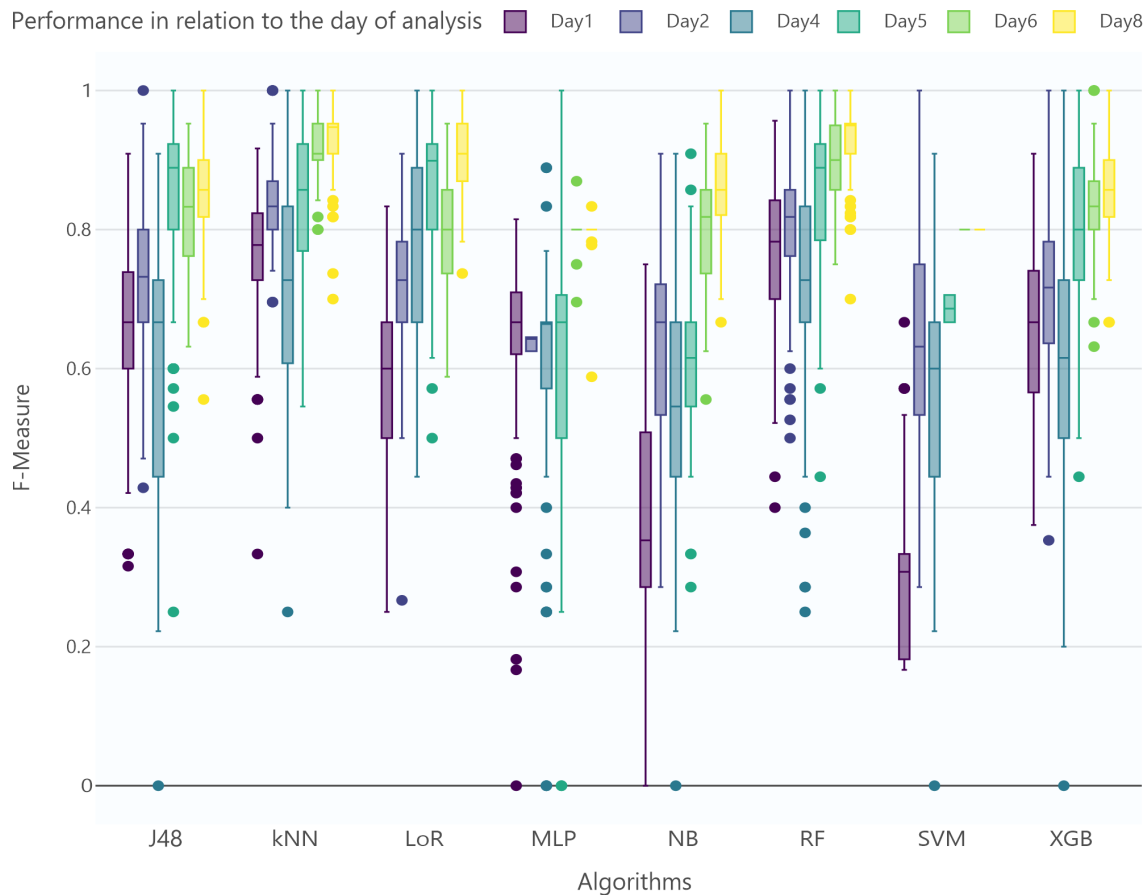


Fig. 5. Training performance of the algorithms in relation to the day of analysis.

values in the day-by-day analysis, this algorithm was chosen to be incorporated into the subsequent steps of the proposed framework and create the theoretical model used to distinguish healthy cotton plants from plants damaged by the insect-herbivore attack. As related, RF is a robust algorithm that already produced satisfactory performances in other remote sensing tasks related to agronomic and vegetation variables estimations (18, 20) and considered a state-of-the-art predictor at machine learning benchmarks and competitions on Kaggle.com (2020). Therefore, a sequential process that incorporates the information from the previous step into the next one was adopted. The spectral modeling suggested that the RF associated with the 8th day of analysis was the most suitable to separate the spectral wavelengths of both classes. Because of that, this configuration-set was used to perform

the subsequent processing.

Using the RF in comparison against the baseline algorithm ZeroR, a ranking approach that considered the difference in accuracy obtained from individual wavelengths into the model's performance was performed. This ensured that a metric score was attributed to each spectral wavelength, corresponding to the increase in the overall accuracy of the RF model. The highest scores correspond with the most contributive wavelengths used in the classification task. This type of approach is important to precise plant or leaf spectral modeling and can assist in the creation of mathematical operations like spectral indices or highly specific sensors. Here, however, a theoretical model is proposed considering just multispectral bands. Therefore, instead of identifying only the individual wavelengths, this framework proposes to separate the most contributive spectral regions of the model. For this, a reductive data-dimensionality approach based on the Self-Organizing Map (SOM) method was used.

The information returned by the ranking approach in conjunction with the SOM results indicates the spectral regions to be considered in the theoretical model (Figure 6). The SOM method, as previously described, consists of a clustering approach based on an unsupervised artificial neural network. This method used the value of the metric score while preserving the topological information of the spectral wavelengths. In this sense, it was able to return almost sequential clustering's that presented similar metric results. By consid-

Table 2. Testing evaluation metrics considering the performance of each algorithm during the 8th day of analysis.

Algorithm	TP	FP	Precision	Recall	F-Measure
RF	0.913	0.143	0.914	0.913	0.912
J48	0.780	0.290	0.777	0.780	0.778
XGB	0.767	0.307	0.764	0.767	0.765
NB	0.813	0.253	0.810	0.813	0.811
LoR	0.900	0.170	0.902	0.900	0.897
SVM	0.667	0.667	0.500	0.667	0.583
MLP	0.667	0.667	0.500	0.667	0.583
kNN	0.893	0.173	0.894	0.893	0.891



ering the cluster with the highest metric score (cluster 3 in Figure 6), two spectral regions can be highlighted, one situated between red and near-infrared (650 to 1,350 nm) and the other at the shortwave infrared (1,570 to 1,640 nm). The score in these regions resulted in mean and standard deviation values of, respectively,  $0.4365 \pm 0.0207$ , being higher than the other clusters:  $0.3127 \pm 0.0177$  for cluster 2;  $0.3717 \pm 0.0177$  for cluster 1, and;  $0.2495 \pm 0.0259$  for cluster 0.

**Multispectral Regions Simulation.** Between the predefined spectral regions, the chosen multispectral sensors evaluated in this proposal possess a total of 10 spectral bands, of which 3 are from the OLI sensor (Landsat 8) and 7 from the MSI sensor (Sentinel 2). For the simulation process, the response functions of each sensor were implemented, and the wavelengths situated between their respective minimal and maximum ranges were used. For the OLI sensor, the red (c.w.: 655 nm), near-infrared (c.w.: 865 nm), and shortwave infrared 1 (c.w.: 1,610 nm) bands were simulated, and for the MSI sensors, the red (c.w.: 655 nm), red-edge 1 (c.w.: 705 nm), red-edge 2 (c.w.: 740 nm), red-edge 3 (c.w.: 783 nm), near-infrared 1 (c.w.: 842 nm), near-infrared 2 (c.w.: 865 nm) and, shortwave infrared (c.w.: 1,610 nm) were also simulated (Figure 2). As all samples of the 8th day were simulated, the same number of samples used at the previous step was obtained with this practice, with the only alteration being the number of attributes (i.e. input variables) into the model. The histograms indicating the distribution of the simulated reflectance data at each spectral band helped to highlight the spectral behavior of the registered cotton plants according to their class (Figure 7).

The band simulation process was able to return the spectral response of the measured in-field plants as if they were registered by another sensor. It is important to highlight that the simulation process may be used for different sensors aside from the ones investigated in this proposal (OLI and MSI). They were chosen mainly because they are up-to-date sensors and provide medium to high spatial resolution imagery to be freely available to the entire world. However, the proposed framework could be adapted into other sensors, some of which are also related to commercial cameras equipped at UAV-based platforms. Regardless, it is important to consider that the spectral measurements were conducted at a plant level and that the spatial resolution of the orbital sensors may offer a potential hindrance for the prediction since it can produce a spectral mixture of other objects at the terrestrial surface, like, for example, soil brightness. Even so, the band simulation process is considered a robust representation of the spectral response of the target at the simulated sensor (8, 35) and may be used for a theoretical model. Using the simulated bands of both sensors (OLI and MSI), two configuration-sets that were implemented in a computational environment were used, and the RF learner performance considering the same consecutive validation approach executed at the spectral experiment was evaluated. At the training phase (Figures 8 and 9), the results indicated that the RF model was able to predict, with median values above 0.9, the overall classification metrics (Precision, Re-

call, F-Measure, and Area Under ROC) satisfactorily. In this situation considering the training set, no statistical differences pending the Mann-Whitney pairwise comparison test were obtained. Also, when considering the relation of True-Positive and False-Positive rates at the Receiver Operating Characteristics (ROC) curves, a slightly better adjustment for the classes with the MSI sensor. Nonetheless, both OLI and MSI returned high estimates at the two classes, which indicates that the theoretical model based on the simulated bands indicated by the combined ranking and SOM method may assist in predicting this type of insect-herbivore damage effect in cotton plants with different sensors.

At the testing phase (Tables 3 and 4), the RF learner was able to return approximate evaluation result metrics to the ones acquired during the spectral classification (Table 2). However, even if a slight decrease in the performance was obtained when adopting the OLI and MSI spectral bands, with F-measure values from 0.912 to 0.886 and 0.865, these results can be considered important outcomes for the RF learner, as it achieved these performances using only 3 (OLI) and 7 (MSI) spectral bands, in comparison against the 1,934 wavelengths implemented with the spectral dataset. Another important observation to be made is that, when evaluating the separated classes, the model performed better at identifying healthy cotton plants in comparison against the damaged plants in both sensors. Nonetheless, the damaged class received lower values of False-Positive (FP) rates, which indicates that the model can predict the cotton plants under insect-herbivore attack with more guarantee than it can predict healthy plants.

Table 3. Testing results of the RF model into the OLI (Landsat 8) sensor dataset.

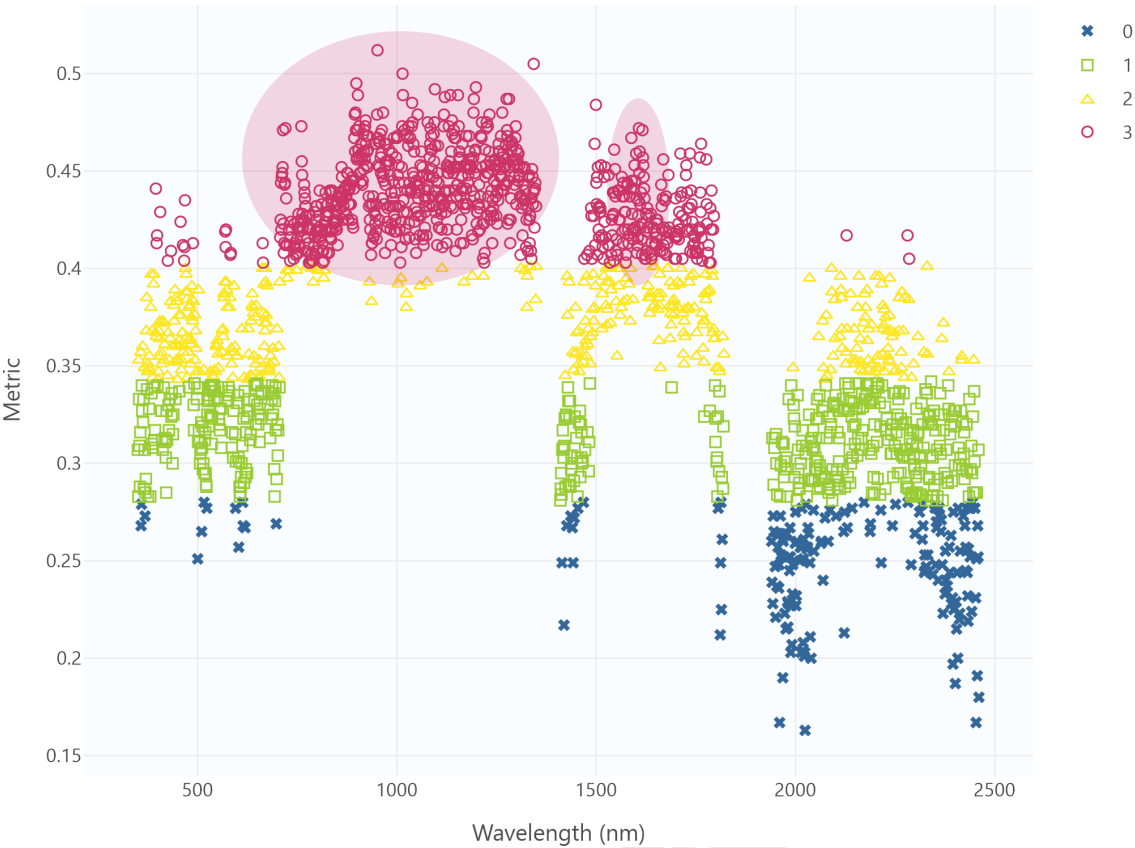
Class	TP	FP	Precision	Recall	F-Measure
Control	0.920	0.240	0.885	0.920	0.902
Damaged	0.760	0.080	0.826	0.760	0.792
Average	0.867	0.187	0.865	0.867	0.865

Table 4. Testing results of the RF model into the MSI (Sentinel 2) sensor dataset.

Class	TP	FP	Precision	Recall	F-Measure
Control	0.930	0.200	0.903	0.930	0.916
Damaged	0.800	0.070	0.851	0.800	0.825
Average	0.887	0.157	0.886	0.887	0.886

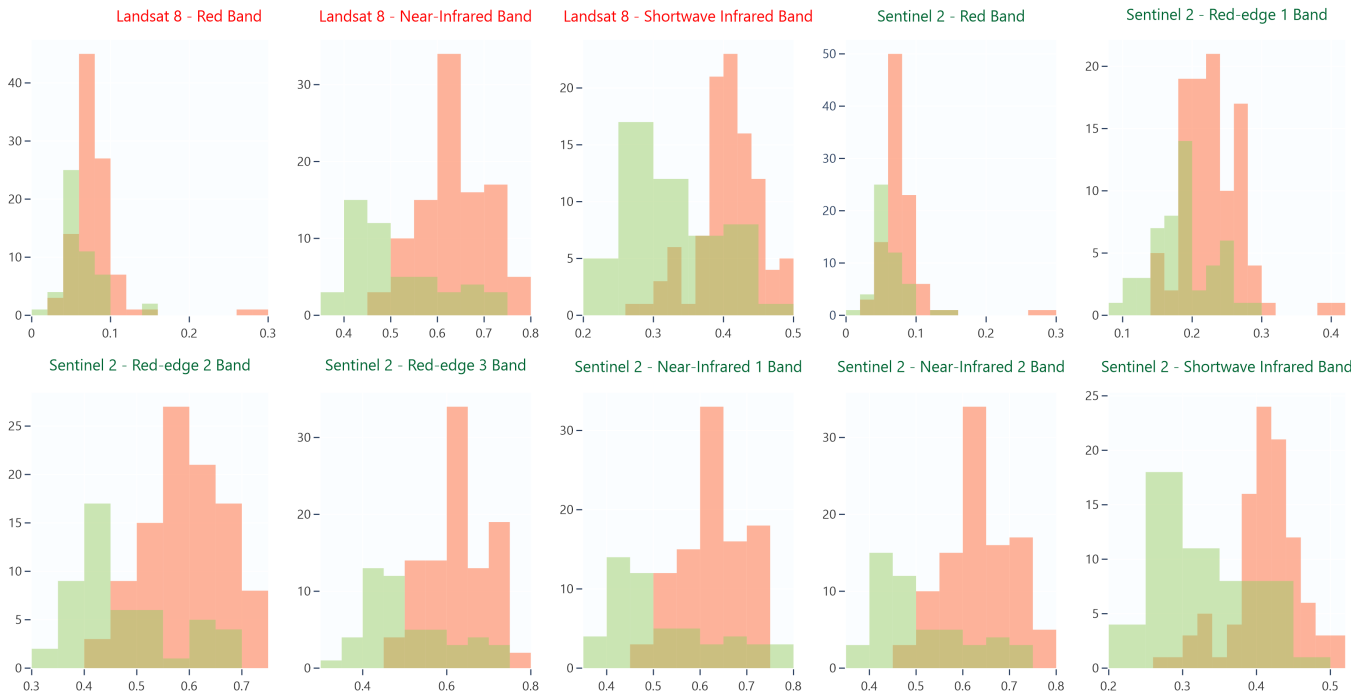
The main concept behind this proposal was to present a framework with methods capable of modeling the spectral response of the cotton plant under insect-herbivore attack and to develop a theoretical model to be implemented into novel research or practices that intend to investigate the performance of these methods in other multispectral sensors. The theoretical model prediction values are an indicator of how feasible the RF algorithm is to correctly separate cotton plants attacked by the fall armyworm from healthy cotton plants even at a multispectral level. In future research, a validation of the theoretical model with real in-field values, in known impacted plantations, could be conducted to determine the impact on the model's accuracy. Regardless, the approach shown here may assist in rapid and non-invasive

Self-Organizing Map for the individual contribution (Ranking) of each wavelength



**Fig. 6.** The ranking approach combined with the Self-Organizing Map (SOM) with the most contributive spectral regions for the classification of healthy and insect-damaged cotton plants.

Histogram distribution of the reflectance values at each simulated band grouped by class (x = reflectance; y = frequency) / (health = green; damaged = red)



**Fig. 7.** Histogram of the spectral response of the cotton plants under the simulated bands according to the class in which they were associated.

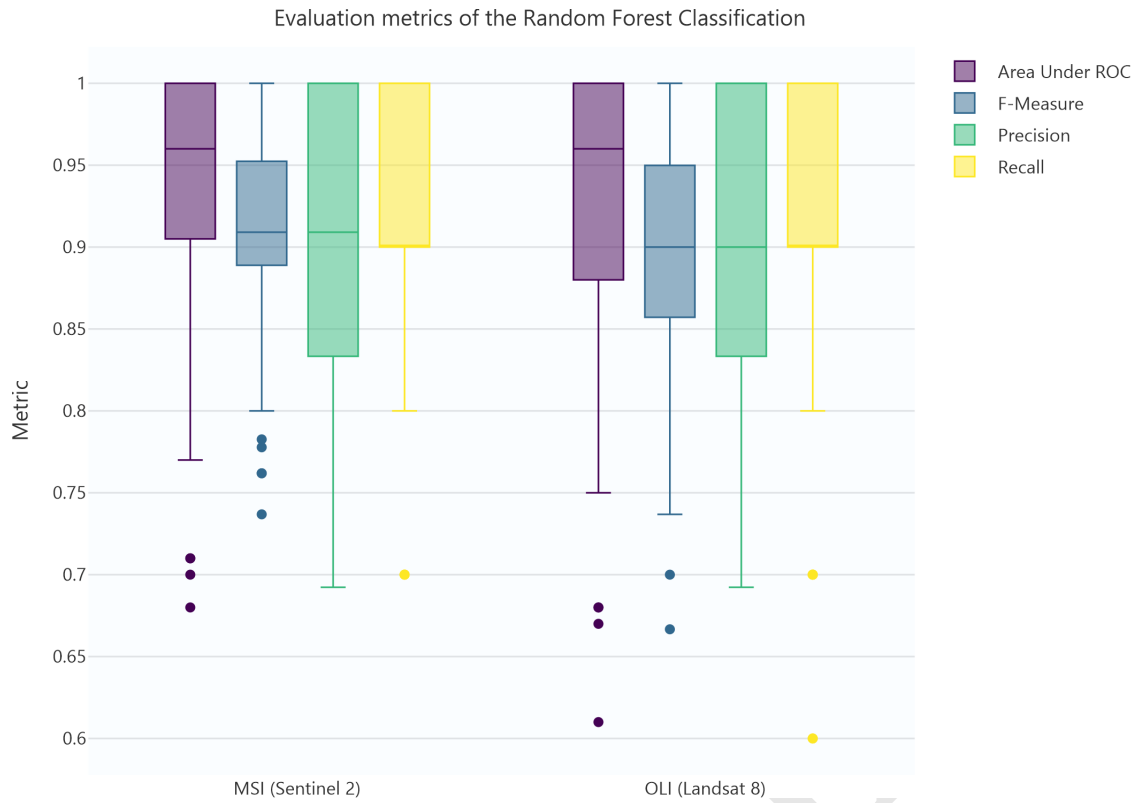


Fig. 8. The Random Forest (RF) performance considering the multiple validations for OLI and MSI sensors.

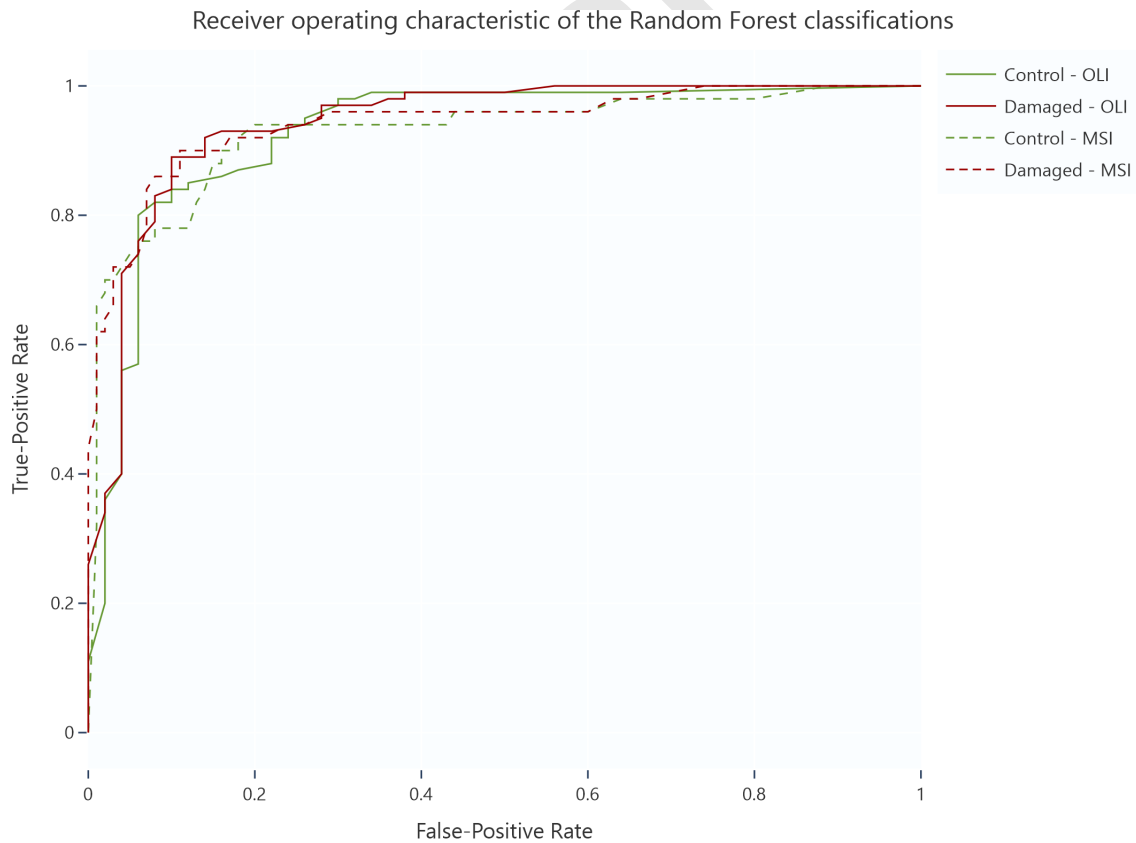


Fig. 9. The comparison between each class classification for the different sensors using their respective Receiver Operating Characteristics (ROC).

monitoring of insect-damaged cotton plants, and the related method may be replicated into similar cultures and cultivars. As spectral measurements may be simulated to other sensors such as multispectral-based ones, different theoretical models can be proposed and initially evaluated with this approach. Another important contribution is that, as the spectral measurements occurred at the plant level, the simulation process with this theoretical model may be scaled up to field and farm levels.

## Conclusions

This paper presented a framework to model the spectral response of the fall armyworm-damaged cotton-plants. It indicated the overall best learner obtained from a state-of-the-art comparison, the appropriate day of analysis, and the most contributive wavelengths and spectral regions to support this type of analysis. It also proposed a theoretical model to infer this damage with multispectral up-to-date orbital sensors, commonly used to map fields, crops, and farmlands in different scales. Among the machine learning methods evaluated, the results indicated that the Random Forests presented better performance to separate healthy from unhealthy cotton-plants, achieving Precision, Recall, and F1-measure values above 0.912 on the last day of analysis (8th day). According to the ranking approach and the SOM clustering method results, both the red to near-infrared (650 to 1,350 nm) and shortwave infrared (1,570 to 1,640 nm) regions are the more important spectral regions for mapping insect-damaged in cotton-plants. Additionally, the theoretical model for both multispectral sensors, OLI and MSI, demonstrated high performance in the evaluated task, returning an F-Measure of 0.865 and 0.886, respectively. In conclusion, the proposed framework can detect cotton plants under insect-herbivory attack with high accuracies and may be implemented by other studies to evaluate its feasibility for other types of insect-damage in similar cultivars. In future research, the evaluation of deeper machine learning methods like deep neural networks with much higher in-field collected information will be conducted.

## ACKNOWLEDGEMENTS

The authors acknowledge the support of the UFMS (Federal University of Mato Grosso do Sul) and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) (Finance Code 001).

## FUNDING

This research was partially funded by National Council for Scientific and Technological Development (CNPq), project number: 433783/2018-4, 303559/2019-5, and 304052/2019-1, and received financial support from the Brazilian Corporation of Agricultural Research (EMBRAPA), project number: 11.14.09.001.04.00.

## CONFLICTS OF INTEREST

The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

## Bibliography

- Jingcheng Zhang, Yanbo Huang, Ruiliang Pu, Pablo Gonzalez-Moreno, Lin Yuan, Kaihua Wu, and Wenjiang Huang. Monitoring plant diseases and pests through remote sensing technology: A review. *Computers and Electronics in Agriculture*, 165:104943, oct 2019. doi: 10.1016/j.compag.2019.104943.
- Vijai Singh, Namita Sharma, and Shikha Singh. A review of imaging techniques for plant disease detection. *Artificial Intelligence in Agriculture*, 4:229–242, 2020. doi: 10.1016/j.iaia.2020.10.002.
- Instituto Brasileiro de Geografia e Estatística. Indicadores IBGE. Contas nacionais trimestrais - indicadores de volume e valores correntes, 2019.
- CONAB. Monitoring of the brazilian harvest 2019/2020, 2020.
- J.G.P.W. Clevers. A simplified approach for yield prediction of sugar beet based on optical remote sensing data. *Remote Sensing of Environment*, 61(2):221–228, aug 1997. doi: 10.1016/S0034-4257(97)00004-7.
- Angel Maresma, Lindsay Chamberlain, Aristotelis Tagarakis, Tulsi Kharel, Greg Godwin, Karl J. Czymmek, Elson Shields, and Quirine M. Ketterings. Accuracy of NDVI-derived corn yield predictions is impacted by time of sensing. *Computers and Electronics in Agriculture*, 169:105236, feb 2020. doi: 10.1016/j.compag.2020.105236.
- H. Sembiring, W.R. Raun, G.V. Johnson, M. L. Stone, J. B. Solie, and S. B. Phillips. Detection of nitrogen and phosphorus nutrient status in winter wheat using spectral radiance. *Journal of Plant Nutrition*, 21(6):1207–1233, jun 1998. doi: 10.1080/01904169809365478.
- Lucas Prado Osco, Ana Paula Marques Ramos, Érika Akemi Saito Moriya, Maurício de Souza, José Marcato Junior, Edson Takashi Matsubara, Nilton Nobuhiro Imai, and José Eduardo Creste. Improvement of leaf nitrogen content inference in valencia-orange trees applying spectral analysis algorithms in UAV mounted-sensor images. *International Journal of Applied Earth Observation and Geoinformation*, 83:101907, nov 2019. doi: 10.1016/j.jag.2019.101907.
- P.L. Hatfield and P.J. Pinter. Remote sensing for crop protection. *Crop Protection*, 12(6):403–413, sep 1993. doi: 10.1016/0261-2194(93)90001-y.
- Jaafar Abdulridha, Ozgur Batuman, and Yiannis Ampatzidis. UAV-based remote sensing technique to detect citrus canker disease utilizing hyperspectral imaging and machine learning. *Remote Sensing*, 11(11):1373, jun 2019. doi: 10.3390/rs11111373.
- H Nilsson. Remote sensing and image analysis in plant pathology. *Annual Review of Phytopathology*, 33(1):489–528, sep 1995. doi: 10.1146/annurev.py.33.090195.002421.
- Nesreen M. Abd El-Ghany, Shadia E. Abd El-Aziz, and Shahira S. Marei. A review: application of remote sensing as a promising strategy for insect pests and diseases management. *Environmental Science and Pollution Research*, 27(27):33503–33515, jun 2020. doi: 10.1007/s11356-020-09517-2.
- Tom M. Mitchell. *Machine Learning*. McGraw-Hill, Inc.: New York, NY, USA, 1997.
- Jiawei Han and Micheline Kamber. *Data Mining Concept and Techniques*. San Francisco: Morgan Kaufman, 2006.
- Sandra M. Guzmán, Joel O. Paz, Mary Love M. Tagert, Andrew E. Mercer, and Jonathan W. Pote. An integrated SVR and crop model to estimate the impacts of irrigation on daily groundwater levels. *Agricultural Systems*, 159:248–259, jan 2018. doi: 10.1016/j.agry.2017.01.017.
- Puyu Feng, Bin Wang, De Li Liu, and Qiang Yu. Machine learning-based integration of remotely-sensed drought factors can improve the estimation of agricultural drought in south-eastern australia. *Agricultural Systems*, 173:303–316, jul 2019. doi: 10.1016/j.agry.2019.03.015.
- Zhan-Yu Liu, Jia-Guo Qi, Nan-Nan Wang, Zeng-Rong Zhu, Ju Luo, Li-Juan Liu, Jian Tang, and Jia-An Cheng. Hyperspectral discrimination of foliar biotic damages in rice using principal component analysis and probabilistic neural network. *Precision Agriculture*, 19(6):973–991, apr 2018. doi: 10.1007/s11119-018-9567-4.
- Lucas Prado Osco, Ana Paula Marques Ramos, Danilo Roberto Pereira, Érika Akemi Saito Moriya, Nilton Nobuhiro Imai, Edson Takashi Matsubara, Nayara Estrabris, Maurício de Souza, José Marcato Junior, Wesley Nunes Gonçalves, Jonathan Li, Veraldo Liesenberg, and José Eduardo Creste. Predicting canopy nitrogen content in citrus-trees using random forest algorithm associated to spectral vegetation indices from UAV-imagery. *Remote Sensing*, 11(24):2925, dec 2019. doi: 10.3390/rs11242925.
- Lucas Prado Osco, Ana Paula Marques Ramos, Mayara Maezozo Fanta Pinheiro, Érika Akemi Saito Moriya, Nilton Nobuhiro Imai, Nayara Estrabris, Felipe Ianczyk, Fábio Fernando de Araújo, Veraldo Liesenberg, Lúcio André de Castro Jorge, Jonathan Li, Lingfei Ma, Wesley Nunes Gonçalves, José Marcato Junior, and José Eduardo Creste. A machine learning framework to predict nutrient content in valencia-orange leaf hyperspectral measurements. *Remote Sensing*, 12(6):906, mar 2020. doi: 10.3390/rs12060906.
- Ana Paula Marques Ramos, Lucas Prado Osco, Danielle Elis Garcia Furuya, Wesley Nunes Gonçalves, Dthenifer Cordeiro Santana, Larissa Pereira Ribeiro Teodoro, Carlos Antonio da Silva Junior, Guilherme Fernando Capristo-Silva, Jonathan Li, Fábio Henrique Rojo Baio, José Marcato Junior, Paulo Eduardo Teodoro, and Hemerson Pistori. A random forest ranking approach to predict yield in maize with uav-based vegetation spectral indices. *Computers and Electronics in Agriculture*, 178:105791, nov 2020. doi: 10.1016/j.compag.2020.105791.
- Tinashe Nyabako, Brighton M. Mvumi, Tanya Stathers, Shaw Mlambo, and Macdonald Mubayiwa. Predicting prostrephanus truncatus (horn) (coleoptera: Bostrichidae) populations and associated grain damage in smallholder farmers' maize stores: A machine learning approach. *Journal of Stored Products Research*, 87:101592, may 2020. doi: 10.1016/j.jspr.2020.101592.
- Ahmed Tageldin, Dalia Adly, Hassan Mostafa, and Haitham S Mohammed. Applying machine learning technology in the prediction of crop infestation with cotton leafworm in greenhouse. *bioRxiv*, 2020. doi: 10.1101/2020.09.17.301168.
- Tingting Chen, Ruier Zeng, Wenxuan Guo, Xueying Hou, Yubin Lan, and Lei Zhang. Detection of stress in cotton (gossypium hirsutum L.) caused by aphids using leaf level hyperspectral measurements. *Sensors*, 18(9):2798, aug 2018. doi: 10.3390/s18092798.
- M. Prabhakar, Y.G. Prasad, M. Thirupathi, G. Sreedevi, B. Dharajothi, and B. Venkateswarlu. Use of ground based hyperspectral remote sensing for detection of stress in cotton caused by leafhopper (hemiptera: Cicadellidae). *Computers and Electronics in Agriculture*, 79(2):189–198, nov 2011. doi: 10.1016/j.compag.2011.09.012.
- Mathyam Prabhakar, Yenumula G. Prasad, Sengottaiyan Vennila, Merugu Thirupathi, Gudapati Sreedevi, G. Ramachandra Rao, and Bandi Venkateswarlu. Hyperspectral indices for assessing damage by the solenopsis mealybug (hemiptera: Pseudococcidae) in cotton. *Computers and Electronics in Agriculture*, 97:61–70, sep 2013. doi: 10.1016/j.compag.2013.



- 07.004.
26. Diego Bedin Marin, Marcelo de Carvalho Alves, Edson Ampélio Pozza, Leônidas Leoni Belan, and Marcelo Loran de Oliveira Freitas. Multispectral radiometric monitoring of bacterial blight of coffee. *Precision Agriculture*, 20(5):959–982, dec 2018. doi: 10.1007/s11119-018-09623-9.
  27. J R Jensen. *Remote sensing of the environment: an earth resource perspective second edition*, volume 1. Prentice Hall, 2014. ISBN 9780131889507.
  28. Federico Martinelli, Riccardo Scalenghe, Salvatore Davino, Stefano Panno, Giuseppe Scuderi, Paolo Ruisi, Paolo Villa, Daniela Stroppiana, Mirco Boschetti, Luiz R. Goulart, Cristina E. Davis, and Abhaya M. Dandekar. Advanced methods of plant disease detection. a review. *Agronomy for Sustainable Development*, 35(1):1–25, sep 2014. doi: 10.1007/s13593-014-0246-1.
  29. Elias S. Gomes, Viviane Santos, and Crébio J. Ávila. Biology and fertility life table of helioverpa armigera (lepidoptera: Noctuidae) in different hosts. *Entomological Science*, 20(1): 419–426, jan 2017. doi: 10.1111/ens.12267.
  30. J. L. Bi, J. B. Murphy, and G. W. Felton. Antinutritive and oxidative components as mechanisms of induced resistance in cotton to helicoverpa zea. *Journal of Chemical Ecology*, 23(1):97–117, jan 1997. doi: 10.1023/b:joec.0000006348.62578.f.d.
  31. LILIAN SCHMIDT, ULRICH SCHURR, and URSULA S. R. RÖSE. Local and systemic effects of two herbivores with different feeding mechanisms on primary metabolism of cotton leaves. *Plant, Cell & Environment*, 32(7):893–903, jul 2009. doi: 10.1111/j.1365-3040.2009.01969.x.
  32. Michael Eisenring, Steven E. Naranjo, Sven Bacher, Angelique Abbott, Michael Meissle, and Jörg Romeis. Reduced caterpillar damage can benefit plant bugs in bt cotton. *Scientific Reports*, 9(1), feb 2019. doi: 10.1038/s41598-019-38917-9.
  33. Linyi Liu, Yingying Dong, Wenjiang Huang, Xiaoping Du, Binyuan Ren, Linsheng Huang, Qiong Zheng, and Huiqin Ma. A disease index for efficiently detecting wheat fusarium head blight using sentinel-2 multispectral imagery. *IEEE Access*, 8:52181–52191, 2020. doi: 10.1109/access.2020.2980310.
  34. Joel Segarra, Maria Luisa Buchailot, Jose Luis Araus, and Shawn C. Kefauver. Remote sensing for precision agriculture: Sentinel-2 improved features and applications. *Agronomy*, 10(5):641, may 2020. doi: 10.3390/agronomy10050641.
  35. George Deroco Martins, Maria de Lourdes Bueno Trindade Galo, and Bruno Sergio Vieira. Detecting and mapping root-knot nematode infection in coffee crop using remote sensing measurements. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 10(12):5395–5403, dec 2017. doi: 10.1109/jstars.2017.2737618.
  36. Leo Breiman. Random forests. *Machine Learning*, 45(1):5–32, 2001. doi: 10.1023/a:1010933404324.
  37. J. Ross Quinlan. *C4.5: Programs for Machine Learning*. Morgan Kaufmann Publishers, San Mateo, CA, 1993.
  38. Sergio González, Salvador García, Javier Del Ser, Lior Rokach, and Francisco Herrera. A practical tutorial on bagging and boosting based ensembles for machine learning: Algorithms, software tools, performance study, practical perspectives and opportunities. *Information Fusion*, 64:205–237, dec 2020. doi: 10.1016/j.inffus.2020.07.007.
  39. George H John. Estimating Continuous Distributions in Bayesian Classifiers. *Robotics*, 1995.
  40. S. Le Cessie and J. C. Van Houwelingen. Ridge estimators in logistic regression. *Applied Statistics*, 41(1):191, 1992. doi: 10.2307/2347628.
  41. Bernhard E. Boser, Isabelle M. Guyon, and Vladimir N. Vapnik. A training algorithm for optimal margin classifiers. In *Proceedings of the fifth annual workshop on Computational learning theory - COLT '92*. ACM Press, 1992. doi: 10.1145/130385.130401.
  42. Simon Haykin. *Neural Networks: A Comprehensive Foundation*. Prentice-Hall, 1993.
  43. N. S. Altman. An introduction to kernel and nearest-neighbor nonparametric regression. *The American Statistician*, 46(3):175–185, aug 1992. doi: 10.1080/00031305.1992.10475879.
  44. Teuvo Kohonen. Self-organized formation of topologically correct feature maps. *Biological Cybernetics*, 43(1):59–69, 1982. doi: 10.1007/bf00337288.
  45. Teuvo Kohonen. *Self-Organizing Maps*. Springer Berlin Heidelberg, 2001. doi: 10.1007/978-3-642-56927-2.
  46. Julia Barsi, Kenton Lee, Geir Kvaran, Brian Markham, and Jeffrey Pedelty. The spectral response of the landsat-8 operational land imager. *Remote Sensing*, 6(10):10232–10251, oct 2014. doi: 10.3390/rs61010232.
  47. Yingjie Li, Jing Chen, Qingmiao Ma, Hankui K. Zhang, and Jane Liu. Evaluation of sentinel-2a surface reflectance derived using sen2cor in north america. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 11(6):1997–2021, jun 2018. doi: 10.1109/jstars.2018.2835823.
  48. Alex Okiemute Onojeghuo and Ajoke Ruth Onojeghuo. Object-based habitat mapping using very high spatial resolution multispectral and hyperspectral imagery with LiDAR data. *International Journal of Applied Earth Observation and Geoinformation*, 59:79–91, jul 2017. doi: 10.1016/j.jag.2017.03.007.
  49. Qingxia Zhao, Shichuan Yu, Fei Zhao, Linghong Tian, and Zhong Zhao. Comparison of machine learning algorithms for forest parameter estimations and application for forest quality assessments. *Forest Ecology and Management*, 434:224–234, feb 2019. doi: 10.1016/j.foreco.2018.12.019.
  50. Zhulin Chen, Kun Jia, Chenchao Xiao, Dandan Wei, Xiang Zhao, Jinhui Lan, Xiangqin Wei, Yunjun Yao, Bing Wang, Yuan Sun, and Lei Wang. Leaf area index estimation algorithm for GF-5 hyperspectral data based on different feature selection and machine learning methods. *Remote Sensing*, 12(13):2110, jul 2020. doi: 10.3390/rs12132110.