http://journal.unoeste.br/index.php/ca/index
DOI: 10.5747/ca.2024.v20.h526
ISSN on-line 1809-8215

Colloquium Agrariae

Submetido: 16/10/2023 Revisado: 23/02/2024 Aceito: 20/03/2024

MACHINE LEARNING INTEGRATION, REMOTE SENSING DATA PREPROCESSING TECHNIQUES TO MAP PESTS COTTON CROPS

Daniel Veras Correa[1], Fabio Rafael Echer[1], Lucas Prado Osco[1], Ana Paula Marques Ramos[2]

[1]Universidade do Oeste Paulista – UNOESTE, Presidente Prudente, SP. [2]Universidade Estadual Paulista – UNESP, Presidente Prudente, SP. E-mail: danielveras248@gmail.com

Abstract

Cotton has a considerable economic impact on agribusiness. Strategies to reduce production loss due, for example, to pest attacks are increasingly required. *Spodoptera frugiperda*, known as fall armyworm, causes irreversible damage to cotton. In this context, a current approach is the use of hyperspectral measurements obtained by remote sensors and processed by machine learning algorithms. However, such measures generate data redundancy, making it difficult to extract information. An alternative is to apply pre-processing techniques, but little is known about the impact these generate on the learning ability of algorithms. This study evaluates the performance of machine learning algorithms in identifying cotton plants attacked by pests using pre-processed and raw hyperspectral measurements. Data are collected by EMBRAPA, and consist of hyperspectral measurements, in the range of 350-2500 nm, referring to eight days of collections in healthy cotton plants and attacked by *S. frugiperda*. Pre-processing techniques to try are baseline removal, smoothing, first and second order derivatives. A group of machine learning algorithms, such as Random Forest, Support Vector Machine, Extra Tree, was used to model pre-processed and non-pre-processed hyperspectral measurements. According to the proposed metric, the F-Score and the Extra Trees (ExT) algorithm performed better (0.77). So it overlapped the other results with the preprocessed dataset. In addition to obtaining the most important lengths for the algorithm to have its best performance. Concluding that machine learning with spectroscopy can help the field in a promising way. Studies in other crops and with other factors applied to the plant are recommended.

Keywords: field spectroscopy; supervised algorithms; precision agriculture.

INTEGRAÇÃO DE APRENDIZAGEM DE MÁQUINA, TÉCNICAS DE PRÉ-PROCESSAMENTO DE DADOS DE SENSORIAMENTO REMOTO PARA MAPEAR PRAGAS NA CULTURA DO ALGODÃO

Resumo

O algodão tem um considerável impacto econômico no agronegócio. Estratégias para redução de perda de produção devido, por exemplo, a ataques de pragas são cada vez mais requeridas. A *Spodoptera frugiperda*, conhecida como lagarta do cartucho, causa danos irreversíveis ao algodão. Neste contexto, uma abordagem atual é o uso de medidas hiperespectrais obtidas por sensores remotos e processadas por algoritmos de aprendizagem de máquina. Todavia, tais medidas geram redundância de dados, dificultando a extração de informações. Uma alternativa é aplicar técnicas de pré-processamento, mas pouco se sabe sobre o impacto que estas geram na capacidade de aprendizagem dos algoritmos. Este trabalho avalia o desempenho de algoritmos de aprendizagem de máquina ao identificarem plantas de algodão atacadas por pragas utilizando medidas hiperespectrais pré-processadas e brutas. Os dados são coletados pela EMBRAPA, e consistem em medidas hiperespectrais, no intervalo de 350-2500 nm, referentes a oito dias de coletas em plantas de algodão saudáveis e atacadas por *S. frugiperda*. As técnicas de pré-processamento a serem testadas são remoção de linha de base, suavização, derivadas de primeira e segunda ordem. Um grupo de algoritmos de aprendizagem de máquina, como *Random Forest, Support Vector Machine, Extra Tree,* foi utilizado para modelar as medidas hiperespectrais pré-processadas ou não. De acordo com a métrica proposta o *F-Score,* o algoritmo *Extra Trees (ExT)* obteve melhor desempenho (0.77). De maneira que se sobrepôs aos outros resultados com o conjunto de dados pré-processados. Além de obtermos os comprimentos de maior importância para o algoritmo ter seu melhor desempenho. Concluindo que o aprendizado de máquina com a espectroscopia pode auxiliar de modo promissor o campo. Recomenda-se estudos em outras culturas, e com outros fatores aplicados à planta.

Palavras-chave: espectroscopia de campo; algoritmos supervisionados; agricultura de precisão.

1. Introduction

1 .1. Cotton cultivation and agricultural technologies

Brazil is one of the world's leading producers and exporters of cotton fiber (CONAB, 2024). The estimated planted area for this crop in 2022/23 was approximately 1.77 million hectares (CONAB, 2024), and the last season in 2019/20 achieved a record production with over 7,372 tons of cottonseed, resulting in a productivity of 1,774 kg/ha. Exports in 2020 amounted to 1.9 million tons, marking an 18% increase compared to 2019, setting a historical record. Mato Grosso (MT) and Bahia (BA) are the two main cotton-producing states in Brazil, cultivating over 88% of the currently planted cotton area (CONAB, 2020).

To achieve increasingly effective production and ensure a strong presence in the market, producers need to properly manage their fields. In this case, the use of geotechnologies, such as data obtained through remote sensing, combined with artificial intelligence methods, can contribute to improving crop monitoring in various aspects, such as plant nutritional conditions, yield estimates, identification of pest and disease attacks, and other related agronomic variables (Suradhaniwar, 2018). Pest attacks on plants are among the issues that have the most significant impact on crop productivity rates (Zhang *et al.*, 2019; Singh *et al.*, 2020). Plant diseases involve some form of physiological modification that disrupts their normal processes for healthy development (Singh *et al.* 2020).

Spectral data analysis collected by remote sensors emerges as a proposal to increase the frequency of crop monitoring (Sawicka; Egbuna. 2020), because such data makes it possible to identify potential physiological changes in the plant (Jensen, 2014). This is a quick and low-cost way to infer information about different crops (Li *et al.*, 2018; He *et al.*, 2016; Perry *et al.*, 2012; Thomason *et al.*, 2011). Hyperspectral sensors stand out for their ability to characterize the spectral response of targets (He *et al.*, 2016). Plants infected by diseases and pests exhibit differences in their spectral response compared to healthy plants because there are changes in the photosynthetic process and internal leaf structures (Jensen, 2014). These differences in spectral response can be particularly mapped using the visible, near infrared, and mid-infrared regions of the electromagnetic spectrum (Zhang *et al.*, 2019). In this sense, an important contribution is to identify the bands that best indicate such differences.

Herbivorous insects can affect cotton at different phenological stages, such as in the vegetative phase by consuming the leaves, or in the flowering and boll stages (Gomes; Santos; Ávila, 2017). Furthermore, they can affect the composition of amino acids, water content, and the oxidative state of cotton (Eisenring *et al.*, 2019). In cotton plants, an insect that causes irreversible damage is *Spodoptera frugiperda*, known as the fall armyworm.

1 .2. Machine learning context

In recent years, the integration of machine learning and remote sensing has been gaining attention in the agronomic sector, including for monitoring diseases and pests in crops. Machine learning is a subfield of artificial intelligence that can be applied to model various types of data (Raju, 2020). Machine learning algorithms are generally classified as supervised and unsupervised, and their main characteristic is the extraction of patterns in a dataset using learning attributes in a sample set (Han; Kamber, 2006). Machine learning methods can analyze hierarchical and non-linear relationships between independent variables and the dependent variable, often resulting in better performance compared to conventional data classification models (Guzmán *et al.*, 2018; Feng *et al.*, 2019).

Machine learning algorithms have been used in the analysis of hyperspectral measurements for crop mapping. As an example, there is a study (Abdulridha; Batuman; Ampatzidis, 2019) that used two algorithms, radial basis function (RBF) and k-nearest neighbor (KNN), on hyperspectral images (400 to 1000 nm) for disease detection in citrus crops. Another study (Nyabako *et al.*, 2020) used decision trees to predict the infestation level of *Prostephanus truncatus* in corn crops. In the context of cotton cultivation, Tageldin *et al.* (2020) investigated various machine learning algorithms to detect the fall armyworm (*S. littoralis*), but the prediction accuracy was relatively low, at 84%. A common feature of most of these approaches is the direct analysis of reflectance data, without exploring additional processes to improve the predictive capability of the models. In this regard, preprocessing techniques for spectral measurement data can be a strategy worth investigating. Among such techniques, we have baseline removal, smoothing, first and second-order derivatives, standard normal variate, multiplicative scatter correction, and principal component analysis (Rinnan; van den Berg; Engelsen, 2009; Yao; Lewis, 2010; Rinnan, 2014). To determine the most suitable technique, in most cases, it involves a "trial and error" approach.

So far, little is known about the impact that preprocessing techniques for hyperspectral measurements have on the learning capability of machine learning algorithms in pest mapping in crops. In this regard, the objective of this study is to evaluate the performance of machine learning algorithms in identifying cotton plants attacked by fall armyworms using preprocessed and raw hyperspectral measurements. The specific objectives are as follows: characterize the spectral behavior of healthy cotton plants and those under attack by fall armyworms; determine the preprocessing technique that improves the machine learning algorithm's ability to identify cotton plants under attack by *S. frugiperda*; and identify the spectral ranges most strongly related to plants under attack by these insects.

1 .3. Machine learning algorithms

There are several machine learning algorithms available in open-source software that are suitable for both classification (separating features into classes) and regression (predicting values), such as: Decision Tree, Artificial Neural Network (ANN), Random Forests, K-Nearest Neighbors, and Support Vector Machine. Each of these algorithms has specific parameters to adjust during modeling.

The Decision Tree algorithm is a supervised and non-parametric machine learning method. Its principle is to determine values for a function, which is represented by a decision tree (Mitchell, 1997). In this type of tree, features are classified from the root node (the beginning of the tree) to a leaf node. Each node in the tree specifies a test of a feature attribute, and each descending branch from that node corresponds to one of the possible values for that attribute (Mitchell, 1997). Some of the advantages of working with decision trees are that they can be applied to both continuous and discrete datasets, do not require assumptions about the frequency distribution of data in each class, and can handle nonlinear relationships between features and classes (Mitchell, 1997).

An Artificial Neural Network or Artificial Neural Network (ANN) is a robust method for approximating real values, discrete values, or vectors of values (discrete or real) for a function (Mitchell, 1997). It is an abstraction of the biological neural network. An ANN is defined as a complex structure interconnected by processing elements called neurons. If the ANN consists of a single neuron, it is called a Perceptron, capable of expressing only linear decisions. In this type of network, there can be inputs but only a single output with a value of 0 and 1 or -1 and 1. If the ANN has two or more neurons, it is called a Multilayer ANN, capable of handling nonlinearly separable problems (Mitchell, 1997).

The operating principle of an ANN is to activate one or more neurons (x0, x1, ...xn) in the network; these input values are multiplied by weights (w0, w1, ...wn) that represent the importance of each input in relation to the desired output value (y). The result of the sum of weighted inputs is added to the activation threshold (θ),

$$u = \sum_{i=0}^{n} (w_i * x_i) - \theta$$

and this value (u) is then passed as an argument to the activation function g(u), which can be linear or nonlinear, resulting in the desired output (Mitchell, 1997). In the case of a Multilayer ANN, algorithms like backpropagation can use gradient descent to adjust the parameters (weights) of the ANN to improve the model's results (Mitchell, 1997).

The Random Forests algorithm is a supervised algorithm based on the principle of decision trees, but it constructs a varied number of trees during the training phase, combining them to make predictions with higher accuracy and stability (Han; Kamber, 2006). Each tree relies on values from a randomly sampled vector, independently and with the same distribution for all trees in the forest. During classification, each tree votes, and the most popular class is returned (Han; Kamber, 2006).

The K-Nearest Neighbors algorithm is a non-parametric method that assumes that all features (training data) are points in an n-dimensional space. The nearest neighbors to an instance are defined by calculating a distance, such as Euclidean distance (Mitchell, 1997). You need to parameterize a search radius (k) to execute the algorithm, which is often defined iteratively. The Support Vector Machine (SVM) algorithm separates the feature space using a hyperplane, which maximizes the margin between instances of different classes or values (Han; Kamber, 2006). This hyperplane is found using the so-called 'support vector.'

1 .4. Performance evaluation metrics

Different metrics can be used to evaluate the performance of a machine learning algorithm, such as (Han; Kamber, 2006): accuracy (or recognition rate), error rate (or misclassification rate), model sensitivity (or recall or true-positive rate), model specificity (or true-negative rate), precision, and the F1-score (harmonic mean between recall and precision).

These evaluation metrics determine how "good" or "accurate" a classifier is in predicting data and can be applied using a cross-validation system. In a cross-validation with k mutually exclusive subgroups (k-fold cross-validation), the dataset is randomly divided into k subsets of similar sizes, and the model is trained and tested k times (Han; Kamber, 2006). Cross-validation is an iterative process. For example, consider a 10-fold cross-validation, which means a cross-validation process with 10 repetitions. Using a total of 10 subgroups is recommended in cross-validation to estimate accuracy (Han; Kamber, 2006).

2. Material and methods

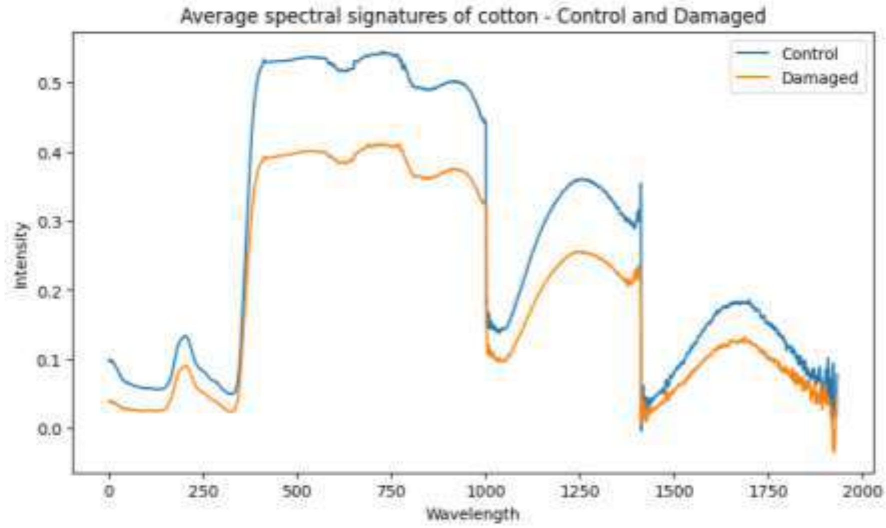
2 .1. Data collection and preprocessing

To create an appropriate dataset, cotton plants (*Gossypium* L.) were cultivated in a controlled environment. These plants were housed in a facility (greenhouse) located at Embrapa Recursos Genéticos e Biotecnologia in Brasília, DF, Brazil. The plants were grown for approximately 14 Days After Emergence (DAE), with fully expanded leaves. Two types of insects were used in the experimental setup. *S. frugiperda* (fall armyworms) were reared in a separate environment at 7±1°C, with 65±10% relative humidity and a 14-hour photoperiod. *Dichelops melacanthus* (stink bugs) were sourced from a laboratory colony maintained in a room at 26±0.3°C, 70±10% relative humidity, and a photoperiod of L14:D10. Both insects were placed in containers and distributed among cotton pots, properly labeled. The experiment spanned 8 days, during which the spectral behavior of the plants was measured.

The spectral measurements were conducted over the course of 8 days, from 9:00 AM to 3:00 PM, with the first day of plant exposure to pests designated as day 1 and the eighth day of exposure as day 8. The experiment resulted in 991 collected spectra, with 465 from samples of healthy cotton plants and 526 from cotton plants attacked by the insects *S. frugiperda* or *D. melacanthus*. The average of these signatures for both classes is illustrated in Figure 1.

The hyperspectral measurements were carried out in a laboratory with ambient light, using a portable spectroradiometer, the ASD FieldSpec 3 (Analytical Spectral Devices Inc., Boulder, USA). This equipment records wavelengths from 17350 to 1000 nm with a resolution of 1.4 nm and from 1000 to 2500 nm with a spectral resolution of 2 nm. A Spectralon white panel was used to calibrate the instrument before conducting the measurements.

Figure 1. Average spectral signatures of cotton.



Source: Author’s own work.

The figure 1 shows the graph with the spectral signatures of healthy cotton plants (blue) and plants under pest attack (orange).

The spectroradiometer was positioned over the leaf, carefully considering the plant's height and the equipment's Field of View (FOV). The data recorded by the spectroradiometer were used to calculate the estimate of the Hemispherical Conical Reflectance Factor (HCRF) of the plant. HCRF is the measurement of interest, meaning it involves the spectral signature of the plant at different wavelengths and is calculated according to the following equation (Anderson *et al.*, 2011),

$$HCRF\,(\omega_i \omega_r) = \frac{dL\,(\theta_r, \Phi_r)\,(alvo)}{dL\,(\theta_r, \Phi_r)\,(referência)}\,K\,(\theta_i, \Phi_i, \theta_r, \Phi_r) \qquad (1)$$

where: dL corresponds to radiance, ω is the solid angle, θ and Φ are, respectively, the zenith and azimuth angles; i corresponds to the incident flux, and r is the reflected energy flux. The value of K is the correction factor specific to the equipment itself.

2 .2. Treatment and organization of hyperpectral data

The data treatment consists of removing spectral regions with noise, as recommended by Jensen (2014), and excluding data collected outside the time interval from 9:00 AM to 3:00 PM, aiming to minimize the influence of sunlight variation on capturing the spectral measurements of targets (cotton plants), as recommended by Jensen (2014). Next, each of the preprocessing techniques, such as baseline removal, smoothing, first and second-order derivatives, standard normal variable, multiplicative scatter correction, and principal component analysis, was applied, resulting in the preprocessed dataset. The open-source software SpectraGraph was used for this purpose. This allowed for comparing the performance of algorithms when processing both non-preprocessed and preprocessed measurements. Subsequently, the data was organized into training, validation, and test sets.

2 .3. Application of machine learning algorithms

In a computational environment, the class (healthy vs. damaged) is selected as the target variable. As input parameters, the hyperspectral curves (bands) were used, and the performance of different algorithms in their prediction was evaluated using the default hyperparameters of each tested algorithm. To configure and run the algorithms, the open-source computer program Weka 3.9.5 was used, which is based on specific Python libraries. Some of the algorithms used for the proposed framework included k-nearest neighbor (kNN), support vector machine (SVM), artificial neural network (ANN), decision tree (DT), and random forest (RF). These algorithms were used to model both non-preprocessed and preprocessed measurements.

The method used to calculate the accuracy metrics of the algorithms is cross-validation. This method involves dividing the data into training and validation sets, so that, given the defined number of folds, K-1 folds are used for training, and the remaining fold is used for validating the algorithm. This process is repeated until each fold has been used to validate the algorithm once. Finally, the contribution of each band or spectral index to the algorithm's performance is calculated by displaying its Relief-F value. Relief-F uses a kNN score to handle noisy data while dealing with incomplete data. It is considered a reliable metric for assessing feature importance and is then applied to rank the features based on their higher scores.

2 .4. Identification of wavelengths related to insect infestation

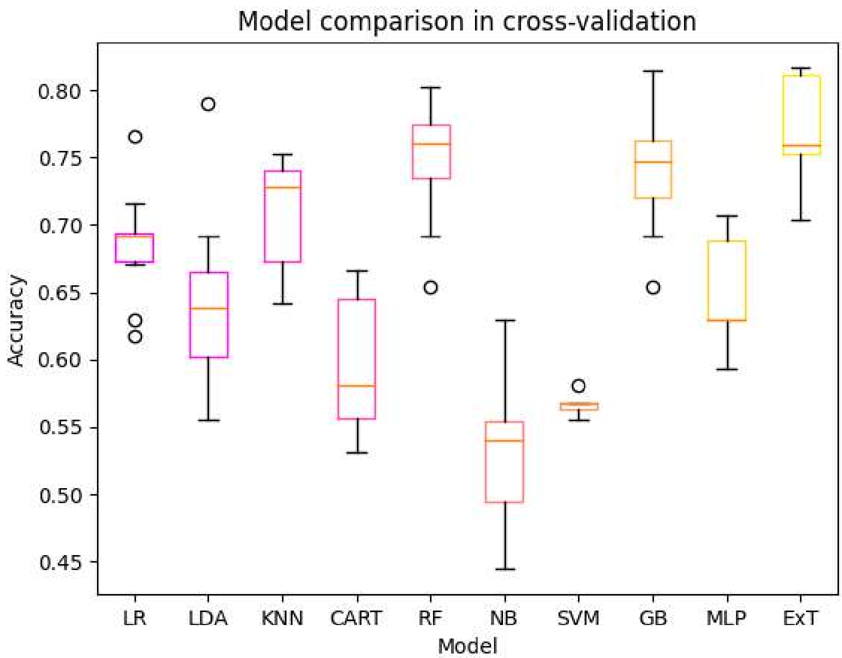
To identify the contribution of each wavelength in the separation between healthy cotton plants and plants damaged by insect infestation, a comparison approach was adopted between the best-performing algorithm identified and a baseline algorithm. The baseline algorithm used for this comparison is ZeroR, which calculates the average value of the measured variables and uses it as a prediction. This machine learning algorithm is considered the baseline for the Weka software. A metric score related to this difference in performance between the algorithms is obtained, which can be positive or negative and may even return a number above 1 since the improvement can exceed 100%. This metric score will indicate the most contributive spectral wavelengths for prediction, i.e., separating healthy plants from damaged plants.

To determine the most contributive spectral regions, instead of just the individual contribution of each band (wavelength), the Self-Organizing Map (SOM) clustering algorithm was used. SOM is based on an unsupervised artificial neural network (Osco *et al.*, 2021). This algorithm applies a competitive learning approach using a neighborhood function. It helps preserve the topological properties of input variables and is useful for visualization because it creates a low-dimensional (2D) representation using high-dimensional datasets (such as hyperspectral data). SOM can be executed within Weka 3.9.5. With this, it was possible to identify the spectral regions of greatest contribution used by the machine learning algorithm to separate healthy cotton plants from plants damaged by insects.

3. Results and discussion

Figure 2 displays the algorithm performance using the first derivative dataset, with Extra Trees notably excelling. Table 1 corroborates this, highlighting Extra Trees during testing. Despite assessing both derivatives, the second derivative didn't offer significant improvement over the first. Thus, further study focused solely on the first derivative. Extra Trees demonstrated the highest efficiency with an F-Score of 0.77, a 31.17% advantage over the lowest-performing algorithm, NB. These results are attributed to redundancy with high dimensionality and the random generation of decision trees with the dataset.

Figure 2. Box-plot comparison graph of cross-validation models.



Source: Author’s own work.

In Figure 2, a box plot compares the performance of the 10 algorithms used in this study.

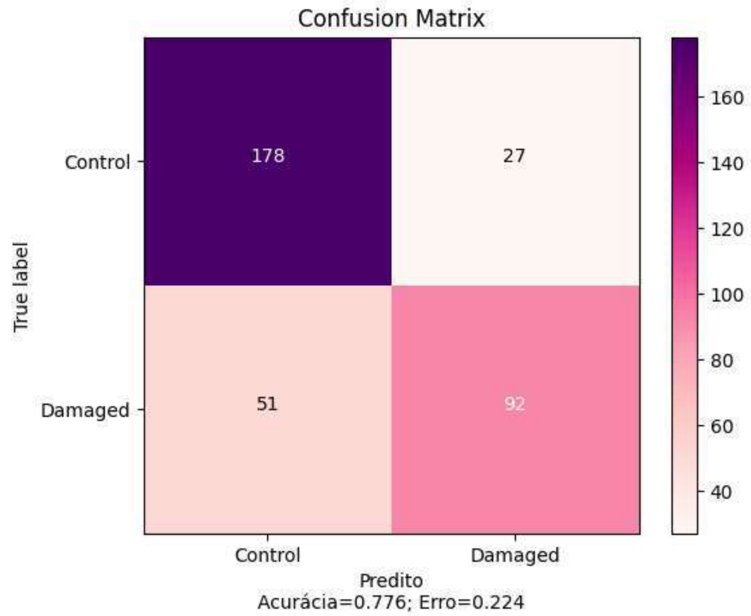
Table 1. Test set results for each algorithm.

Algorithms	F-Score
LR	0.687
LDA	0.713
KNN	0.716
CART	0.661
RF	0.750
NB	0.532
SVM	0.586
GB	0.739
MLP	0.552
ExT	0.773

Source: Author's own work.

The confusion matrix of the Extra Trees algorithm (Figure 3) illustrates TP - true positive, FP - false positive, TN - true negative, and FN - false negative. Examination of this matrix reveals that ExT encountered comparable challenges in misclassifying healthy plants as damaged (FN = 37%) and damaged plants as healthy (FP = 37%).

Figure 3. Confusion matrix for the ExT algorithm.



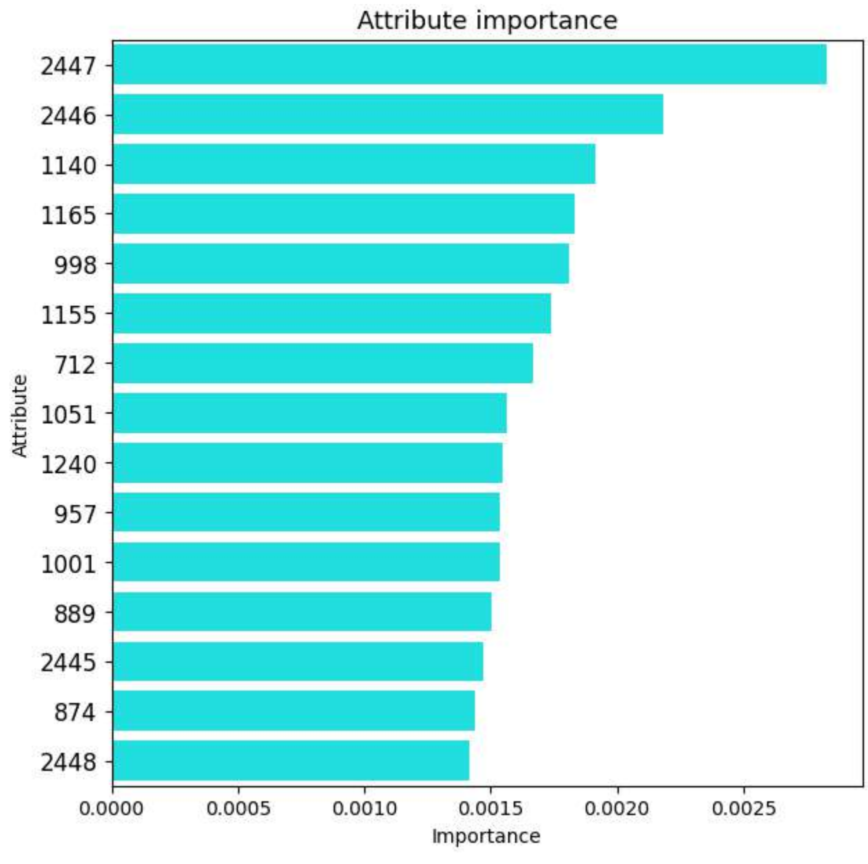
Source: Author's own work.

In Figure 3, the confusion matrix graph of the ExT algorithm, which exhibited the best performance, is illustrated.

The algorithm aims to assimilate a spectral signature for each case. Figure 4 displays the 15 most important spectra utilized by the algorithm, arranged in descending order from top to bottom. The colored bars represent the spectra, with the corresponding wavelength indicated by the numbers in front.

This focused approach enables efficient allocation of computational resources, facilitating continuous improvement. By concentrating efforts on the most important wavelengths, resources can be redirected to enhance algorithm results and accuracy, leading to more efficient machine learning models. Identifying key wavelengths allows for resource allocation to other aspects, potentially improving outcomes and guiding future research and optimizations.

Figure 4. First 15 wavelengths used by the Ext.



Source: Author's own work.

Figure 4 displays the 15 most relevant spectra used for decision-making by the ExT algorithm.

While promising, further research is necessary to ensure that agriculture can fully benefit from these advancements, minimizing unnecessary impacts on both productivity and the environment.

4. Conclusion

Considering that this work aims to evaluate machine learning algorithms and how preprocessing can assist in analyzing hyperspectral data collected at the leaf level under attack by pests, specifically *S. frugiperda* and *D. melacanthus*, the conclusion based on the provided data is that the Extra Trees algorithm performed the best with an F-Score accuracy of 0.77. Additionally, the preprocessing technique proved to be less effective in this dataset.

This study highlights the possibility of mapping factors that affect plants through spectroscopy, coupled with machine learning, which enables more efficient assessments and decision-making. It is recommended to apply this approach to other crops and consider additional factors that may influence them.

Agradecimentos

Os autores agradecem ao CNPq pela bolsa PIBITIC/UNOESTE 2022-2023 Processo número 157949/2022-2.

References

ABDULRIDHA, J., BATUMAN, O., AMPATZIDIS, Y. UAV-based remote sensing technique to detect citrus canker disease utilizing hyperspectral imaging and machine learning. **Remote Sensing**, v.11, n.11, p.1373, 2019. doi:10.3390/rs11111373

ANDERSON, K., DUNGAN, J. L., MACARTHUR, A., On the reproducibility of field-measured reflectance factors in the context of vegetation studies. **Remote sensing of environment,** v. 115, n. 8, p.1893-1905, 2011. https://doi.org/10.1016/j.rse.2011.03.012

CONAB. **With a new reduction, the estimate for the grain harvest of 2023/24 is 306.4 million tons**. Brasília: CONAB, 2024.

CONAB. **Monitoring of the brazilian harvest, grains**. Brasília: CONAB, 2020.

EISENRING, M.; NARANJO, S.E.; BACHER, S.; ABBOTT, A.; MEISSLE, M.; ROMEIS, J. Reduced caterpillar damage can benefit plant bugs in bt cotton. **Scientific Reports**, v.9, n.1, feb. 2019. doi: 10.1038/s41598-019-38917-9.

FENG, P.; WANG, B.; LIU, D.L.; YU, Q. 2019. **Machine learning-based integration of remotely-sensed drought factors can improve the estimation of agricultural drought in South-Eastern Australia.** Agric. Syst, 173, 303–316. https://doi.org/10.1016/j.agsy.2019.03.015

GOMES, E.S.; SANTOS, V.; ÁVILA, C.J. Biology and fertility life table of Helicoverpa armigera (Lepidoptera: Noctuidae) in different hosts. **Entomological Science**, v.20, n.1, p.419–426, jan. 2017. https://doi.org/10.1111/ens.12267

GUZMÁN, S.M.; PAZ, J.O.; TAGERT, M.L.M.; MERCER, A.E.; POTE, J.W. **An integrated SVR and crop model to estimate the impacts of irrigation on daily groundwater levels.** Agricultural Systems, n.159, p.248–259, 2018. https://doi.org/10.1016/j.agsy.2017.01.017

HAN, J.D.; KAMBER, M. **Data mining concept and tehniques.** San Fransisco: Morgan Kauffman, 2006.

HE, L.; SONG, X.; FEND, W.; GUO, B. B.; ZHANG, Y. S.; WANG, Y. H.; WANG, C. Y.; GUO, T. C. Improved remote sensing of leaf nitrogen concentration in winter wheat using multi-angular hyperspectral data. **Remote Sensing of Environment**, v.174, p.122-133, 2016. https://doi.org/10.1016/j.rse.2015.12.007

JENSEN, J.R. **Remote sensing of the environment:** an earth resource perspective. 2th. Pearson Education, 2014. 619p.

LI, Z.; JIN, X.; YANG, G.; DRUMMOND, J.; YANG, H.; CLARK, B.; LI, Z.; ZHAO, C. **Remote Sensing of Leaf and Canopy Nitrogen Status in Winter Wheat (*Triticum aestivum* L.) Based on N-PROSAIL Model. Remote Sensing,** vol. 10, pp. 1463, 2018. https://doi.org/10.3390/rs10091463

MITCHELL, T. **Machine learning.** Nova York: McGrawill, 1997.

NYABAKO, T.; MVUMI, B.M.; STATHERS, T.; MLAMBO, S.; MUBAYIWA, M. Predicting *Prostephanus truncatus* (Horn) (Coleoptera: Bostrichidae) populations and associated grain damage in smallholder farmers' maize stores: a machine learning approach. **Journal of Stored Products Research**, v.87, p.101592, 2020. https://doi.org/10.1016/j.jspr.2020.101592

OSCO, L. P., JUNIOR, J. M., RAMOS, A. P. M., de CASTRO JORGE, L. A., FATHOLAHI, S. N., de ANDRADE SILVAM, J., ... & LI, J. (2021). A review on deep learning in UAV remote sensing. **International Journal of Applied Earth Observation and Geoinformation.** 102, 02456. https://doi.org/10.1016/j.jag.2021.102456

PERRY, E.M.; FITZGERALD, G.J.; NUTTALL, J.G.; O'LEARY, G.J.; SCHULTHESS, U.; WHITLOCK, A. Rapid estimation of canopy nitrogen of cereal crops at paddock scale using a Canopy Chlorophyll Content Index. **Field Crops Research**, v.134, p.158-164, 2012. https://doi.org/10.1016/j.fcr.2012.06.003

RAJU, B., JUMAH, F., ASHRAF, O., NARAYAN, V., GUPTA, G., SUN, H., … & NANDA, A. (2020). Big data, machine learning, and artificial inteligence: a field guide for neurosurgeons. **Journal of neurosurgery,** 1(aop), 1-11.

RINNAN, A. Pre-processing in vibrational spectroscopy-when, why and how. **Analytical Methods,** n.18, 2014. https://doi.org/10.1039/C3AY42270D

RINNAN, A.; van den Berg, F.; Engelsen, S.B. **Review of the most common pre-processing techniques for near-infrared spectra.** 2009.

SAWICKA, B.; EGBUNA, C. Pests of agricultural crops and control measures. *In*: **Natural remedies for pest, disease and weed control**. Academic Press, 2020. p.1-16. https://doi.org/10.1016/B978-0-12-819304-4.00001-4

SINGH, V.; SHARMA, N.; SINGH, S. A review of imaging techniques for plant disease detection. **Artificial Intelligence in Agriculture**, 2020. https://doi.org/10.1016/j.aiia.2020.10.002

SURADHANIWAR, S. *et al.* Geo-ICDTs: Principles and applications in agriculture. *In*: **Geospatial technologies in land resources mapping, monitoring and management**. Springer, 2018. p.75-99. https://doi.org/10.1007/978-3-319-78711-4_5

TAGELDIN, A.; MOSTAFA, H.; MOHAMMED, H.S. Applying machine learning technology in the prediction of crop infestation with cotton leafworm in greenhouse. **bioRxiv**, 2020. https://doi.org/10.1101/2020.09.17.301168

THOMASON, W.E.; PHILLIPS, S.B.; DAVIS, P.H.; WARREN, J.G.; ALLEY, M.M.; REITER, M.S. Variable nitrogen rate determination from plant spectral reflectance in soft red winter wheat. **Precision Agriculture**, v.12, p.666-681, 2011. https://doi.org/10.1007/s11119-010-9210-5

YAO, H.; LEWIS, D. Spectral preprocessing and calibration techniques. *In*: **Hyperspectral imaging for food quality analysis and control**. Elsevier, 2010. https://doi.org/10.1016/B978-0-12-374753-2.10002-4.

ZHANG, J.; HUANG, Y.; PU, R.; GONZALEZ-MORENO, P.; YUAN, L.; WU, K.; HUANG, W. Monitoring plant diseases and pests through remote sensing technology: a review. **Computers and Electronics in Agriculture**, n.165, p.104943, oct. 2019. https://doi.org/10.1016/j.compag.2019.104943