Infrared Physics & Technology 123 (2022) 104203

Contents lists available at ScienceDirect

Infrared Physics and Technology

journal homepage: www.elsevier.com/locate/infrared

An impact analysis of pre-processing techniques in spectroscopy data to classify insect-damaged in soybean plants with machine and deep learning methods

Lucas Prado Osco [e,*], Danielle Elis Garcia Furuya [a], Michelle Taís Garcia Furuya [a], Daniel Veras Corrêa [g], Wesley Nunes Gonçalvez [b], José Marcato Junior [b], Miguel Borges [c], Maria Carolina Blassioli-Moraes [c], Mirian Fernandes Furtado Michereff [c], Michely Ferreira Santos Aquino [c], Raúl Alberto Laumann [c], Veraldo Lisenberg [h], Ana Paula Marques Ramos [a,f], Lúcio André de Castro Jorge [d]

[a] *Program of Environment and Regional Development, University of Western São Paulo, Presidente Prudente, Brazil*
[b] *Faculty of Engineering, Architecture, and Urbanism and Geography, Federal University of Mato Grosso do Sul, Campo Grande, Brazil*
[c] *National Research Center of Development of Genetic Research and Biotechnology, Brazilian Agricultural Research Agency, Brasília, Brazil*
[d] *National Research Center of Development of Agricultural Instrumentation, Brazilian Agricultural Research Agency, São Carlos, Brazil*
[e] *Faculty of Engineering and Architecture and Urbanism, University of Western São Paulo, Presidente Prudente, Brazil*
[f] *Post-graduate Program of Agronomy, University of Western São Paulo, Presidente Prudente, Brazil*
[g] *Agronomy, University of Western São Paulo, Presidente Prudente, Brazil*
[h] *Department of Forest Engineering, College of Agriculture and Veterinary, Santa Catarina State University (UDESC), Santa Catarina, Brazil*

ARTICLE INFO

Keywords:
Field spectroscopy
Remote sensing
Precision agriculture
Artificial intelligence

ABSTRACT

Spectroscopy is essential to understand a series of phenomena in multiple fields of study. In remote sensing, vegetation analysis is one of the most prominent fields to explore, aiming to improve a specific task. As a task, modeling insect damage in the plants is essential to establish the correct management of agricultural farmlands. Hyperspectral data, which can be acquired with field spectroscopy at plant or leaf level, is a non-direct, rapid, and trustworthy approach to indicate its health. However, the spectral redundancy inherent is a challenge for the information extraction process, making the pre-processing phase an essential part of the analysis. Currently, artificial intelligence techniques, mostly based on machine and deep learning methods, are a standard application in data processing, being pre-processing techniques an essential part of it. But few studies aimed to measure the impact of such processes in vegetation monitoring, specifically with insect damage and spectral data. Here, we provide an analysis of the impact of pre-processing techniques on machine learning algorithms' performance over said classification task. For this, we used a field spectroradiometer that operates within the 350–1,000 nm and 1,000–2,500 nm ranges. The dataset was composed of multiple spectral measurements that took place on different days in a controlled environment with soybean plants. As pre-processing techniques, methods like baseline removal, smoothing, first and second-order derivatives, standard normal variate (SNV), multiplicative scatter correction (MSC), and principal components analysis (PCA) were investigated. Several machine learning algorithms and one deep learning method were applied to model the datasets. The impact of the pre-processing techniques was measured within validation metrics relate to its accuracy. Our results indicated that the Extra-Tree (ExT) algorithm was better, mainly when first-order derivative data were extracted from the dataset (accuracy equal to 93.68%). A ranking approach indicated that the most contributive spectral region situates at the near-infrared, between 784 and 911 nm. Our investigation also demonstrates that a deep neural network (DNN) did not return a satisfactory result over raw reflectance data. However, when considering a combination of PCA over the 2nd derivative data, it achieved similar results to the ExT algorithm (accuracy of 91.95%). The implications of such, alongside the ranking approach, are discussed in this paper. We hope that the information presented here serves as a framework for future research when applying pre-processing techniques alongside the machine and deep learning methods over spectral data.

* Corresponding author.
E-mail address: lucasosco@unoeste.br (L.P. Osco).

https://doi.org/10.1016/j.infrared.2022.104203
Received 21 December 2021; Received in revised form 2 May 2022; Accepted 3 May 2022
Available online 13 May 2022
1350-4495/

1. Introduction

Soybean plants are one of the most important crops on the planet, mostly because they are great sources of protein and oil. Three countries are considered the largest producers on a global scale: Argentina, Brazil, and the United States of (USA), which account for approximately 16, 32, and 33%, respectively, of the entire market share [6]. In Brazil, for instance, plantations are concentrated in three main states, and to achieve effective production, ensuring high market share, producers need to carry out the proper management of their areas. However, a common problem faced by many farmers in this tropical environment is the presence of insects and larvae. As such, approaches that consider modeling insect-herbivore damage in the crops are essential to establish an appropriate control and the correct management of agricultural farmlands, maintaining its production rates. A strategy to minimize both qualitative and quantitative losses in crop yield refers to early and accurate detection of insect damage caused in plants [15].

To detect insect damage in plants, field visual inspection is still a direct method used in many agricultural farmlands. However, this approach for monitoring plants in the field is labor-intensive, being prone to be subjective, and generally shows low-efficiency [25,15]. The difficulties imposed by traditional methods promulgated the development of faster and less labor-intensive methods to infer certain diagnoses. As of recently, remote sensing approaches are being tested to detect the presence of these plagues by associating them with damaged plants. One approach that holds potential relates to field spectroscopy investigations. In this regard, the spectroscopy area refers to the method of obtaining the hyperspectral characteristics of a target regarding radiation flux intensity emitted or reflected by its constituents at different wavelengths [12]. For the last decades, many studies have proved the potential of remote sensing in the precision agriculture area, mainly for plant disease detection [14,25,7,24].

As state-of-the-art, spectral data analysis has been mostly conducted with more robust and intelligent algorithms. Since this type of dataset is overly complex for traditional statistical analysis to deal with, machine and deep learning methods have been incorporated to improve its evaluation. As examples of machine learning processing in hyperspectral data, a study [1] applied two algorithms, radial basis function (RBF) and K-nearest neighbor (KNN), in hyperspectral (400 to 1,000 nm) images for the detection of a citrus canker at different development stages on leaves and immature (green) fruit, obtaining an overall accuracy higher than 94%. [16] developed decision-tree algorithms to predict the level of P. truncatus infestation and damage on maize grain, but the model performance was weak ($r = 0.43$) because of the complicated sampling and measurements being conducted over a long period between events. As for a damage prediction on grains by insects, the model had a stronger correlation coefficient ($r = 0.93$) being considered a good estimator. [22] investigated several learning algorithms to predict cotton leafworm (Spodoptera littoralis) plant infestation in the greenhouses and found that the XGBoost algorithm was the most effective, achieving a prediction accuracy of 84%.

Regardless, most of these approaches implemented direct analysis over reflectance data, not exploring additional processes to ensure that their model is obtaining the best accuracy possible. In this regard, pre-processing techniques are may help to improve overall classification or regression tasks. In remote sensing, specifically in visible and near-infrared spectral regions, data processing techniques have been used more often, even before artificial intelligent methods. With the nature of the machine and deep learning methods, where the algorithms learn from the dataset itself, pre-processing techniques have been integrated as a standard procedure in any type of data. However, it is still unknown how much these techniques can impact the overall analysis of spectral data classification. Commonly implemented pre-process techniques in spectral data consist of baseline removal, smoothing, first and second-order derivatives, standard normal variate (SNV), multiplicative scatter correction (MSC), principal components analysis (PCA), and others [21,23,20]. These techniques are part of today's hyperspectral data processing, and to indicate the most suitable technique, in most cases, is a "trial-and-error" type of approach.

Monitoring vegetation insect damage over time, with hyperspectral data, is not a simple task. Mostly because of the overly redundant dataset produced by it. Up to the time of writing, few studies aimed to measure the importance of the aforementioned pre-processing techniques in vegetation spectra, specifically within this scenario. Here, we provide an analysis of the impact of commonly used pre-processing techniques on the machine learning algorithms' performance over said classification task. For this, we conducted an experiment in a controlled environment with soybean plants. To better understand the importance of such processes on hyperspectral data, we evaluated the impact of said techniques on validation metrics regarding the accuracy, recall, precision, and overall scores. The implications of such, alongside a ranking approach to indicate the individual contribution of each wavelength over the classification within the best technique and algorithm found, are discussed in this paper. We hope that the information presented here serves as a framework for future research when applying pre-processing techniques alongside the machine and deep learning methods over hyperspectral data.

2. Method

The method was divided into five main phases (Fig. 1): (1) spectral data acquisition in-field conditions of healthy and insect-damaged soybean plants throughout eight days of consecutive analysis; (2) pre-processing of the dataset with different techniques; (3) processing of the dataset using machine learning algorithms and a DNN method; (4) identification of the overall best-adjusted model considering each pre-processing dataset; (5) analysis of the most contributive wavelengths based on the ranking approach.

2.1. Experimental Design

To compose an appropriate dataset, soybean plants (Glycine max) were cultivated in a controlled environment. These plants were stored in a greenhouse facility maintained at Embrapa Genetic Resources and Biotechnology in Brasília, DF, Brazil. The plants were grown for approximately 2 weeks after seed emergence and had fully expanded leaves. Two types of insects were used in the experimental delineation. The first one was Spodoptera frugiperda (larvae), created in a separated environment at 7 ± 1 °C, with 65 ± 10% relative humidity and a 14-h photoperiod. The second one was Dichelops melacanthus (stink-bug) from a laboratory colony, kept in a room with 26 ± 0.3 °C, 70 ± 10% relative humidity, and L14:D10 photoperiod. Both insects were reared in containers and spread across soybean pots, labeled appropriately. The experiment was conducted for approximately 8 days, where the spectral behavior of the plants was measured.

2.2. Spectral Measurement

To conduct the spectral measurement, plants were taken outside to direct sunlight exposition. The equipment used to register the spectral behavior of both damaged and undamaged plants was a handheld spectroradiometer ASD FieldSpect 3 (Analytical Spectral Devices Inc., Boulder, USA). This equipment posses two sensors, which operates within the visible and near-infrared regions, registering wavelengths from 350 to 1,000 nm and from 1,000 to 2,500 nm, with spectral resolutions of 1.4 nm and 2 nm, respectively. Prior to the vegetation analysis, a Spectralon white reference panel was used to calibrate the reflectance spectrum from the measured targets. The reflectance data (ρ) is obtained by the division of the radiance reflected by the measured target (LT) by the quantitie of radiance registered at the same wavelenght interval from the reference table (Lr). This is later multiplied by a known correction factor (K). This process is summarized in the Eq. (1).

The equipment is composed of a 1º lens, which was used to register the spectral behaviour and later on converted to reflectance values.

$$\rho_T = \frac{L_T}{L_r}\ x\ K \tag{1}$$

Since the experiment was conducted for 8 days, spectral measurements were also conducted consecutively. Here we used data registered between 09:00 and 15:00, and the reference Spectralon panel was often registered to ensure that the equipment stayed calibrated throughout the experiment. By the final day of register, a total of 991 spectra were collected, in which 465 samples were from healthy plants, and the remaining 526 samples were from insect-damaged plants. The data, in reflectance value (treated as ”raw data” in the following steps), was organized and the wavelengths, from 350 to 2,500 nm, were used as an input for the subsequent processes. After the pre-processing techniques were applied, we removed the initial (before 390 nm) and final spectral wavelengths(after 2,400 nm) mostly because of low-to-signal noise from the spectroradiometer. Also, to ensure that regions related to the atmospheric absorption of water vapor Jensen [12], we chose to remove the following regions from the analysis: 1,350 to 1,440 nm and 1,800 to 1,980 nm. This resulted in 1,693 wavelengths to be evaluated by the machine learning models.

2.3. Data Pre-Processing and Organization

For the pre-processing step, we organized the raw data into spreadsheets and implemented the mdatools and RNIR libraries in R language. These libraries offer lots of chemometrics methods and near-infrared data analysis with R 4.04. As techniques, we used the baseline removal, the smoothing method, both 1st and 2nd order-derivatives, the standard normal variate (SNV) and multiplicative scatter correction (MSC) techniques, and lastly the principal components analysis (PCA). Since data reduction is an interesting approach for deep learning methods, we also applied the PCA in conjunction with the remaining techniques, incorporating it into the neural network architecture constructed for this experiment.

The data created with said techniques was used as input for multiple learners (i.e. algorithms), state-of-the-art methods used in spectra data processes as of today. Model from the following algorithms were than produced: logistic regression (LoR); linear discrimination analysis (LDA); k-nearest neighbor (KNN); classification and regression trees (CART) (or simply decision tree); naive Bayes (NB); support vector machine (SVM); gradient boosting (GB); multi-layer perceptron (MLP) and; extra-tree (ExT). As mentioned, a deep neural network (DNN) was also built to investigate the impact of said techniques on a deep learning method. The impact of said techniques within all these algorithms was evaluated with validation metrics, as described in the following section. This organization was summarized in an illustrated form to assist in understanding its order (Fig. 2).

The pre-processing techniques chosen here were selected mainly because they have good relation with spectral data processing, being used numerous times previously, even without the integration with the machine or deep learning methods. The spectral smoothing technique reduces high-frequency noises, removing what is called ”spikes” in the data. However, a moving average window must be chosen properly to not interfere with high-frequency components related to important information [20]. Here we used a 7x7 window, which is a standard in smoothing pre-process. As for the baseline removal, this simple technique returns the spectral data to a common baseline. This is useful since mostly spectra have an offset caused by changes in illumination angle or optical path length [23]. Here we used a polynomial fitting method to perform a least-square fitting of a curve to its base. The SNV method consists of a normalization of the data by subtracting each spectrum by its meanwhile dividing it by its standard deviation value. This is important to compare differences between the spectral samples in terms of intensities and correct changes provoked by its optical path length and light scattering. As for the MSC method, it is often used to compensate for the same issues related to light and path length changes, by minimizing the deviations by fitting a linear model between a reference spectrum and others. Usually, it is similar to the SNV, producing proximal results in most cases [21,20].

The derivative analysis is another type of data processing commonly applied to spectra analysis. It helps to remove constant background signals, deal with overlapping peaks, and highlight absorption ranges. The problem with derivatives is the constant increase of noise, so only the first and second-order-derivatives are more often adopted in such

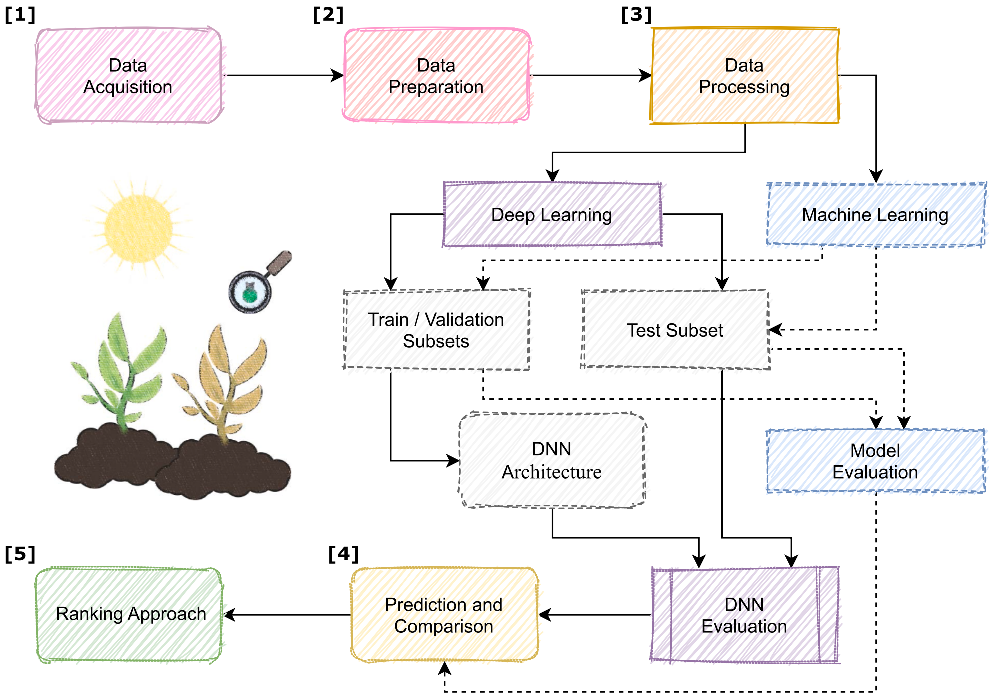

Fig. 1. Simplified scheme with the processing-steps implemented in this study.

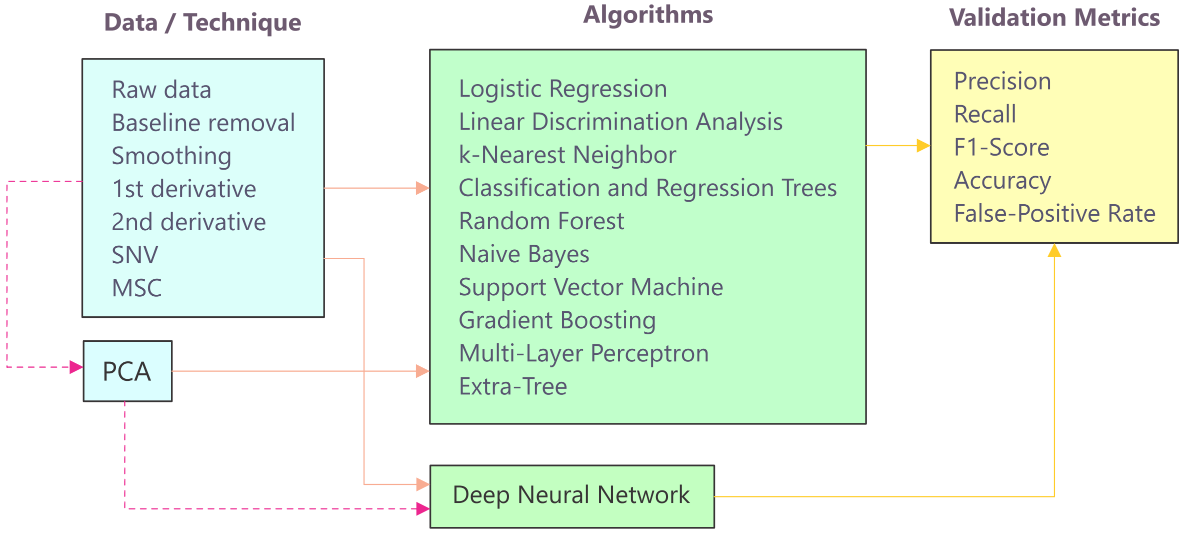

Fig. 2. Diagram simplifying the organization of the spectral data acquired during the study.

cases [21]. Here, to reduce noise increase, the Savitzky-Golay algorithm was used. Lastly, the PCA is one of the most adopted methods to reduce data dimensionality by decomposing data and returning unit vectors, known as ”loadings”. The components are oriented in a variable space and data points are projected into it, forming a score. The score distance is calculated between the projection of the variable and its origin, and the distance can be computed with standardized scores. Both scores and distances are important to indicate how well samples are described by the model [23,20]. PCA is a widely known technique in data analysis and can be used in conjunction with the other processes investigated here.

2.4. Machine Learning Analysis

Once obtained different sets of data were through the aforementioned pre-process techniques, the spectra were loaded into a computational environment built with the Python 3.9.9 language. In this environment, we applied an oversampling approach namely SMOTE to compensate for the differences between samples in each class. We also used a StandartScaler to fit the data into a min.max range, between 0 and 1 scale. After that, the scikit-learn library was used to import the algorithms. The dataset was randomly split into 70% for training and 30% for testing. From the training group, a cross-validation approach with $k = 10$ folders was adopted. This approach separates the training set into 10 parts, storing its data in a stratified manner to match the balanced conditions of the classes samples. From this, 9 parts of the data are used to train the algorithm and generate a model, while the remaining part is used as a test. To ensure that the training metrics were consistent, we also run the cross-validation method additional 10 times, thus creating 100 validation scores for each algorithm. The data was later plotted in box-plot graphics and analyzed. As for the algorithm’s parameters, most of them were left with their respective library default values, except the ones described in the following.

The CART, ExT, and RF are algorithms based on decision trees, where the ExT [9] is an ensemble method that builds randomized trees with independent structures, while CART consists of the classical decision tree model [4], providing support for bagged decision trees. RF [4], on the other hand, combines multiple tree predictors in a manner that each tree depends on values of a random independent vector. In our environment, the number of attributes decided at each node of the ExT was defined as random, and the RF used 100% of the training set as bagging size. The LoR and the NB are both based on a probabilistic concept, where LoR [5] is based on a sigmoid function, and NB [13] uses a naïve approach based on the Bayes Theorem, disregarding the correlation between input variables. Here we adopted a Ridge value equal to 0.00000001 in the log-likelihood for the LoR, and did not use any Kernel estimator nor supervised discretization for the NB method.

The MLP [11] uses hidden layers to perform a classification task, and executes it in a feed-forward manner, being dependent on its activation functions and solver adapted for optimizing its weights. The MLP used here adopted a learning rate of 0.05, a momentum of 0.1, Adam solver, and sigmoid functions. The SVM [3] separates an attribute space using a hyperplane and calculates a linear function while maximizing the margins between instances. We implemented a C-SVC type, with an eps and gamma equal to 0.001 and exp(-gamma*—u-v—^{2}), respectively, using a radial basis function as Kernel. The kNN [2] verifies the proximities of the data by adopting a set of weights and distance metrics. Here we set the number of neighbors to be equal to 5 and used a euclidean distance approach to measure it. Lastly, the GB, which is one of the most recognized algorithms by the machine learning community [10], implements a forward stage-wise ensemble method and computes second-order gradients of a loss function, generally over a decision tree type of learning.

For the deep neural network (DNN) method (Fig. 3) we perform an attribute normalization by standardizing our training data. We used the Adam optimizer, adopted an adaptive learning rate with a sparse categorical cross-entropy loss function, and used the batch size of 64. The accuracy was adopted as stop criteria, and 70% of the dataset was used for training the DNN. For the hidden layers, we used two dense layers with 256 neurons, one with 128 neurons, and another two layers with 32 neurons, adopting the Relu activation function with a dropout of 20% between each layer. A dense final layer was added with the softmax function with 2 units. A total of 1000 epochs were evaluated and the deviance criteria were used to determine the necessary amount of epochs. After 200 epochs the model did not improve in performance; so this number was considered ideal for the given task. To measure the network performance in the testing phase, the same metrics used for the evaluation of the shallow learners were used.

All algorithms were submitted to the same conditions of training and testing split, and from the cross-validation method, the following validation metrics were calculated from its confusion matrix: precision (P); recall (R); F-score; global accuracy (Acc.), and; false-positive rate (FPR). These metrics were obtained by Eqs. ()()()()()(2)–(6), in which TP means true-positive, FP is false-positive, TN is true-negative and FN is false-negative. These values are obtained with the confusion matrices generated at the end of every test conducted.

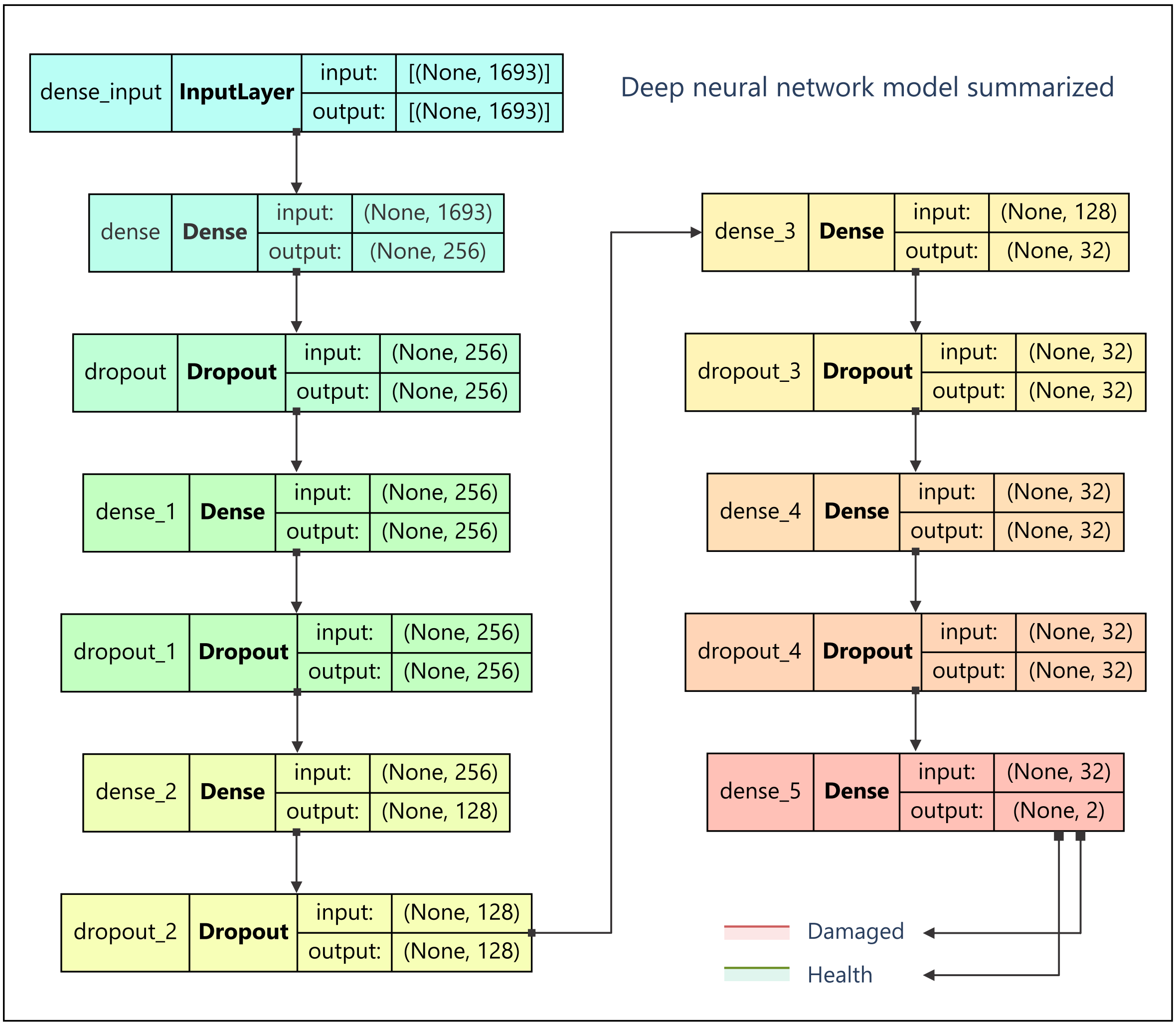
Deep neural network model summarized
dense_input InputLayer input: [(None, 1693)] output: [(None, 1693)]
dense Dense input: (None, 1693) output: (None, 256)
dropout Dropout input: (None, 256) output: (None, 256)
dense_1 Dense input: (None, 256) output: (None, 256)
dropout_1 Dropout input: (None, 256) output: (None, 256)
dense_2 Dense input: (None, 256) output: (None, 128)
dropout_2 Dropout input: (None, 128) output: (None, 128)
dense_3 Dense input: (None, 128) output: (None, 32)
dropout_3 Dropout input: (None, 32) output: (None, 32)
dense_4 Dense input: (None, 32) output: (None, 32)
dropout_4 Dropout input: (None, 32) output: (None, 32)
dense_5 Dense input: (None, 32) output: (None, 2)
Damaged
Health

Fig. 3. Deep neural network model visualization of the hidden layers configurations and parameters.

$$P = \frac{TP}{TP + FP} \tag{2}$$

$$R = \frac{TP}{TP + FN} \tag{3}$$

$$F.Score = 2 \times \frac{P \times R}{P + R} \tag{4}$$

$$Acc. = \frac{TP + TN}{TP + TN + FP + FN} \tag{5}$$

$$(FPR) = \frac{FP}{TP + FP} \tag{6}$$

The precision metric is related to the number of samples returned by the model as one class, divided by the total number of samples from said class. The recall consists of the relevant samples obtained by the model's classification divided by the total of existing samples. The F-score metric is the harmonic mean of precision and recall values, being useful to indicate the overall performance of the model. The global accuracy metric measures how many true samples were classified to the total number of samples. And the FT-rate is measured with how many samples were mistakenly classified as one of the classes. The impact of each pre-processing technique was evaluated regarding the difference between the values of these metrics to the testing results from the raw data (i.e. reflectance), which served as a baseline for comparison.

3. Results and Discussion

The dataset was composed of all the measure variables within the days of analysis and separated into two classes: Undamaged plants (i.e. ”Health”) and herbivory-damaged plants (i.e. ”Damaged”) (with both the S. frugiperda larvae and D. melacanthus bugs). The averaged and the standard deviation values of every wavelength indicated that both groups differentiate each other, in amplitude terms, in most of the near-infrared spectrum space (Fig. 4). The plot also describes the average and standard deviation values from the spectra with the pre-process techniques applied. Notice that some of the graphical differences between bands are mostly because of the removal of its portions during the data organization phase. Another observation is that, in the visible spectrum, the damaged group had a little deviation from the averaged values than the control group. This indicates that this region may not be interesting to separate both groups, which could indicate a possible hindrance for the early visual inspection of the individuals. Henceforward, investigative analysis of the near-infrared regions appears to be promissory for this task.

As for the PCA, the information obtained within the raw data spectrum indicated that it is feasible to reduce the number of components to at least 3 since these account for almost 99% of the variance observed in the dataset (Fig. 5). The PCA is an important strategy for data processing, and since the 7 first components account for all of the explained variances, it is an important technique to reduce the dimensionality of the dataset. Hyperspectral data, specifically for said task, can return

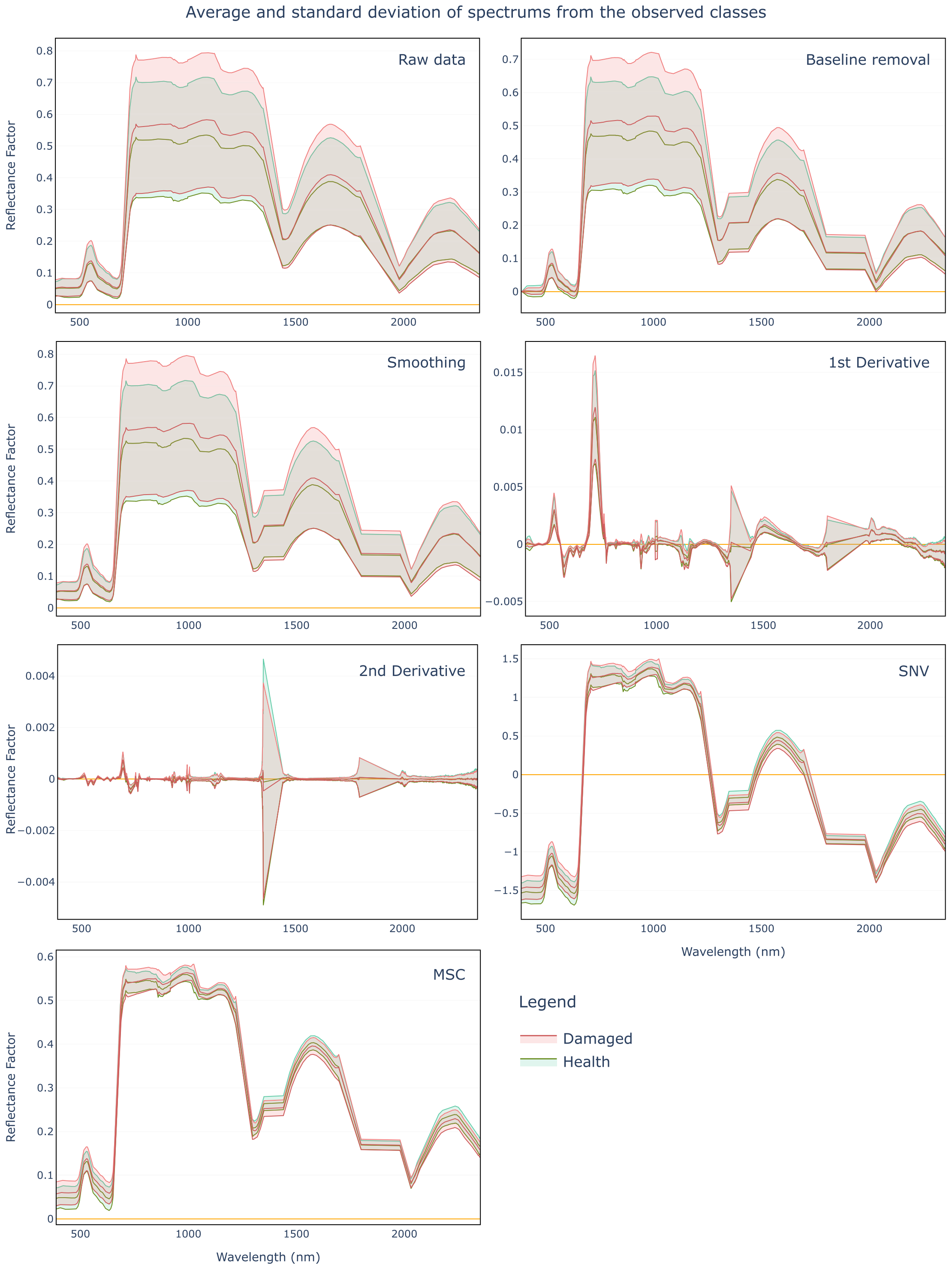

Fig. 4. Spectral averaged data of the implemented pre-processing techniques obtained from the experiment in soy plants.

highly-correlated information. This reduction, on the other hand, can be beneficial for algorithms like SVM or even neural networks, since they are sensitive to the high-dimensionality of the data.

To ensure that the models were properly trained, a box-plot indicating the variation in the F-score was used (Fig. 6). As explained, the F-score measurement is a robust metric to indicate the overall performance of a method. By evaluating this result in multiple consecutive runs from each algorithm, we were able to measure the impact of different training/validation sets. As an onverall observation of these results, it is noticeable some standard behavior between the algorithm's performance and the type of dataset used. We considered the raw data (i.e. the spectra in reflectance values) as a baseline. For most algorithms, any of the pre-processing techniques improved its accuracy. This first observation is an important indicator of the overall importance of pre-processing spectral data in a complex analysis such as this, over instead dealing only with the reflectance values.

Indeed, when measuring different targets with distinctly spectral behavior (i.e. vegetation, water bodies, bare soil, urban areas, etc.), one could easily use the reflectance values. But here we are considering the same target (soybean plants), being the only difference between each

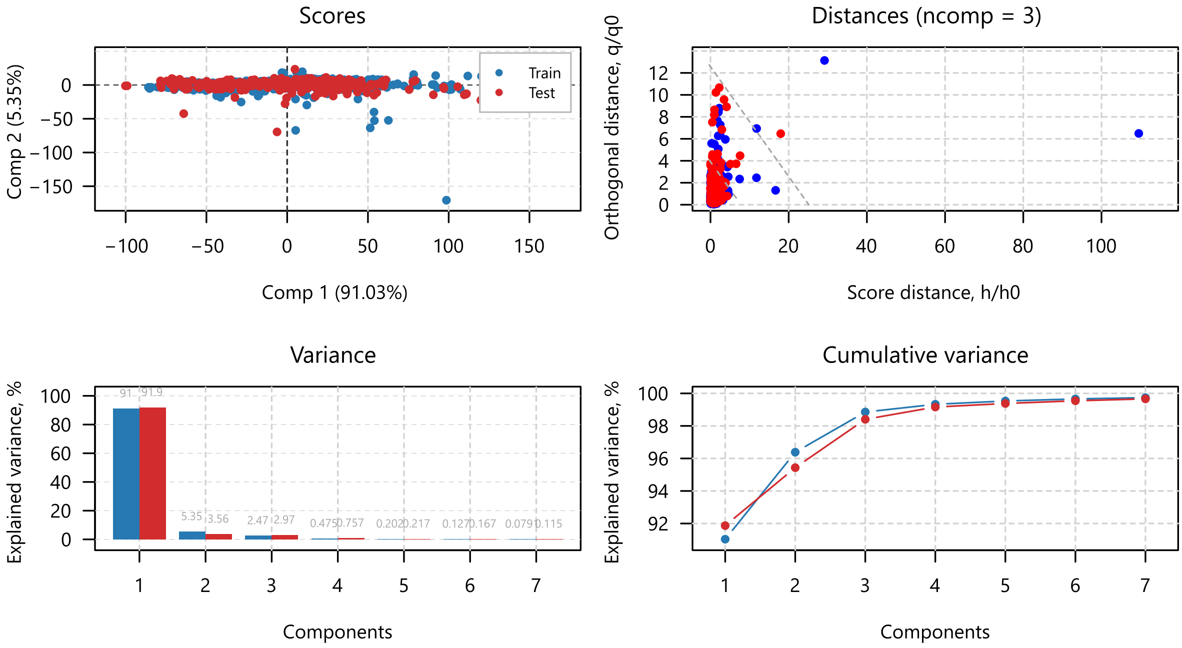

Fig. 5. PCA information obtained from the raw data spectrum.

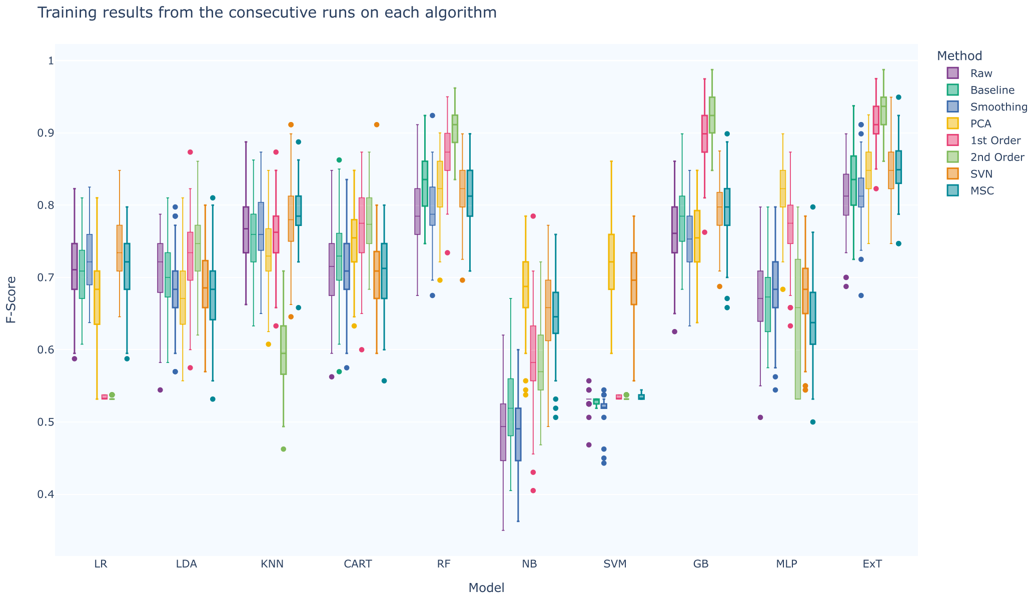

Fig. 6. Training results from the consecutive runs on each algorithm.

class the impact of the insect-damaged sustained by the plant. This creates a highly complex dataset, and the pre-processing techniques appear to better deal with this type of condition than the pure reflectance values. The importance of said techniques for the algorithms performances was also observed on the testing results (Tables 1–3).

Since multiple algorithms were used for each one of the data acquired with the pre-processing techniques, it is important to also consider that these results are influenced not only for the nature of the data itself, but also from the algorithm's characteristics and preferences. Classifiers that are characteristically more linear, as well as those related to probabilistic approaches, performed worse than classifiers that were based on decision trees (like ExT and RF). These trees were not affected by the highly redundant data nor its high dimensionality. This condition appears to affect more linear algorithms, and may also explain why most of the remaining algorithms performed better when introducing the PCA previously on its training phase. While the RF is considered one of the most important and overall powerful algorithms out there, ExT, which consists of a unique tree model, returns high metrics, thus being a quicker and an efficient solution for this problem, being a good method even when more datasets are introduced into the system.

By comparing the F-score, which is the harmonic mean between precision and recall, as well as the global accuracy and an FP-rate, there's a practical gain for almost all of the algorithms regarding the pre-processing technique used. Still, what appears to be the overall best approach, especially considering the decision tree-based learners (CART, RF, GB, and ExT), are both the usage of 1st and 2nd order-derivatives (Table 2), being the ExT the overall best method. This could be related to two things: first, the high-dimensionality of the dataset is not a problem for decision tree learners, and; second, since we are measuring the same target (i.e. soybean plants), these derivative

Table 1
Test dataset results for the raw data, baseline removal, and smoothing of spectrum using different models to classify health (control) plants from insect-damaged plants.

Raw data	LoR	LDA	KNN	CART	RF	NB	SVM	GB	MLP	**ExT**
Precision (%)	63.44	63.44	70.97	66.67	73.12	56.99	0.0	69.89	52.69	**74.19**
Recall (%)	69.41	68.60	74.16	66.67	76.40	43.09	0.0	71.43	63.64	**82.14**
F1-score (%)	66.29	65.92	72.53	66.67	74.73	49.07	0.0	70.65	57.65	**77.97**
Accuracy (%)	69.85	69.35	74.87	68.84	76.88	44.72	53.27	72.86	63.82	**80.40**
FP Rate (%)	29.82	30.09	24.55	29.25	22.73	52.63	46.73	25.93	36.07	**20.87**
Baseline Rem.	LoR	LDA	KNN	CART	RF	NB	SVM	GB	MLP	**ExT**
Precision (%)	75.82	70.33	81.32	70.33	85.71	58.24	0.0	79.12	79.12	**87.91**
Recall (%)	80.23	63.37	77.89	73.56	86.67	54.08	0.0	84.71	66.67	**91.95**
F1-score (%)	77.97	66.67	79.57	71.91	86.19	56.08	0.0	81.82	72.36	**89.89**
Accuracy (%)	80.40	67.84	80.90	74.87	87.44	58.29	54.27	83.92	72.36	**90.95**
FP Rate (%)	19.47	27.55	16.35	24.11	11.93	37.62	45.73	16.67	20.88	**09.82**
Smoothing	LoR	LDA	KNN	CART	RF	NB	SVM	GB	MLP	**ExT**
Precision (%)	66.28	77.91	72.09	67.44	84.88	62.79	0.0	70.93	76.74	**83.72**
Recall (%)	62.64	59.29	64.58	62.37	70.87	40.91	0.0	66.30	50.77	**74.23**
F1-score (%)	64.41	67.34	68.13	64.80	77.25	49.54	0.0	68.54	61.11	**78.69**
Accuracy (%)	68.34	67.34	70.85	68.34	78.39	44.72	56.78	71.86	57.79	**80.40**
FP Rate (%)	26.85	22.09	23.30	26.42	13.54	47.76	43.22	23.36	28.99	**13.73**

Table 2
Test dataset results for the first and the second derivative of spectrum using different models to classify health (control) plants from insect-damaged plants.

PCA	LoR	LDA	KNN	CART	RF	NB	SVM	GB	MLP	**ExT**
Precision (%)	68.00	68.00	77.00	73.00	84.00	68.00	77.00	76.00	87.00	**85.00**
Recall (%)	71.58	71.58	78.57	77.66	87.50	71.58	70.64	81.72	87.88	**86.73**
F1-score (%)	69.74	69.74	77.78	75.26	85.71	69.74	73.68	78.76	87.44	**85.86**
Accuracy (%)	70.35	70.35	77.89	75.88	85.93	70.35	72.36	79.40	87.44	**85.93**
FP Rate (%)	30.77	30.77	22.77	25.71	15.53	30.77	25.56	22.64	13.00	**14.85**
1st Derivativa	LoR	LDA	KNN	CART	RF	NB	SVM	GB	MLP	**ExT**
Precision (%)	0.0	71.58	87.37	80.00	90.53	63.16	0.0	94.74	80.00	**93.68**
Recall (%)	0.0	71.58	72.17	87.36	91.49	62.50	0.0	93.75	81.72	**93.68**
F1-score (%)	0.0	71.58	79.05	83.52	91.01	62.83	0.0	94.24	80.85	**93.68**
Accuracy (%)	52.26	72.86	77.89	84.92	91.46	64.32	52.26	94.47	81.91	**93.88**
FP Rate (%)	47.74	25.96	14.29	16.96	08.57	33.98	47.74	04.85	17.92	**05.77**
2nd Derivative	LoR	LDA	KNN	CART	RF	NB	SVM	**GB**	MLP	ExT
Precision (%)	0.0	67.37	63.16	72.63	89.47	69.47	0.0	92.63	68.42	**93.68**
Recall (%)	0.0	71.11	50.85	76.67	89.47	55.93	0.0	92.63	70.65	**90.82**
F1-score (%)	0.0	69.19	56.34	74.59	89.47	61.97	0.0	92.63	69.52	**92.23**
Accuracy (%)	52.26	71.36	53.27	76.38	89.95	59.30	52.26	92.96	71.36	**92.46**
FP Rate (%)	47.74	28.44	43.21	23.85	09.62	35.80	47.74	06.73	28.04	**05.94**

Table 3
Test dataset results for the Standard Normal Variate (SNV) and Multiplicative Scatter Correction (MSC) of spectrum using different models to classify health (control) plants from insect-damaged plants.

SNV	LoR	LDA	KNN	CART	RF	NB	SVM	GB	MLP	**ExT**
Precision (%)	69.57	75.00	71.74	72.83	80.43	68.48	54.35	73.91	66.30	**81.52**
Recall (%)	71.91	65.71	81.48	69.79	84.09	58.33	69.44	73.12	67.03	**83.33**
F1-score (%)	70.72	70.05	76.30	71.28	82.22	63.00	60.98	73.51	66.67	**82.42**
Accuracy (%)	73.37	70.35	79.40	72.86	83.92	62.81	67.84	75.38	69.35	**83.92**
FP Rate (%)	25.45	24.47	22.03	24.27	16.22	31.87	33.07	22.64	28.70	**15.60**
MSC	LoR	LDA	KNN	CART	RF	NB	SVM	GB	MLP	**ExT**
Precision (%)	69.07	74.23	67.01	63.92	81.44	70.10	0.0	80.41	17.53	**83.51**
Recall (%)	77.91	70.59	86.67	74.70	84.04	67.33	0.0	82.98	73.91	**90.00**
F1-score (%)	73.22	72.36	75.58	68.89	82.72	68.69	0.0	81.68	28.33	**86.63**
Accuracy (%)	75.38	72.36	78.89	71.86	83.42	68.84	51.26	82.41	56.78	**87.44**
FP Rate (%)	26.55	25.77	25.81	30.17	17.14	29.59	48.74	18.10	45.45	**14.68**

processes helps to highlight absorption ranges, thus improving separability between the classes.

The remaining techniques also returned interesting outcomes, but not as well as the derivatives within the decision tree models. Smoothing helped improve LDA, NB, and MLP algorithms results, while the removal of the baseline better improved LoR and KNN, exclusively, and also others like LDA, GB, MLP, and ExT (Table 1). As for the SNV and MSC methods, performances were similar in most cases, noticeably affecting more the GB and MLP models, respectively (Table 3). However, one interesting processing, that even though not by much, improved all the algorithm's performance equally, was the PCA (Table 2). Although it was not able to beat the combined framework of decision trees and 1st and 2nd order-derivatives, it helped improve learners like SVM considerably. As mentioned, because of the SVM characteristic of creating hyperplanes to separate data, one could assume that the dimensionality reduction obtained with the PCA technique was substantial to help this algorithm learn the dataset.

As aforementioned, since the best combination of data + algorithm was the 1st order-derivative with the ExT learner, a ranking approach was implemented to indicate the most contributive wavelengths (Fig. 7). This type of ranking approach differentiates from the traditional method implemented in machine learning evaluation since normally ranking methods are used to firstly remove unimportant variables from the dataset. However, in recent discoveries, specifically within decision tree-based models, machine learning libraries are also providing measurements from the practical importance metric of each input parameter (in this case, the wavelengths) after the model is generated [19]. This helps to analyze data and also reduce the number of variables used during the implementation of the algorithm, thus reducing processing time. In our study case, we discovered that the initial range of the 1st order-derivative of the near-infrared region, with wavelengths situated between 784 and 911 nm, is the most indicated to highlight the differences between health and insect-damaged soybean plants.

By the nature of the algorithm itself, decision trees are capable of ignoring non-practical or important variables and basing their decision ultimately on the best possible route. So when conducting a rank approach, it is interesting to verify the overall importance of all wavelengths to the algorithm, and then separate the most prominent ones. To demonstrate the feasibility of this method, we also evaluate the performance of the algorithm when considering only the 20 most contributive wavelengths (Fig. 8). The confusion matrix did not demonstrate high differences, either from the global accuracy standard-point or between classes. The small improvement between considering the 1,693 waves or only the 20 best waves was not affected. One hypothesis for that is that, as mentioned, decision-tree learners are not highly affected by non-important variables. As said, this results in lower computational cost, and it’s important to process higher quantities of data.

Even though the ranking approach is an interesting take to reduce input data, data dimensionally still is a problem for some algorithms to deal with. This initial exploration with shallow learners also returned indicatives of how one might deal with such data when considering a deep learning approach. In this regard, we tested whether such preprocessing techniques can also impact a deep neural network (DNN). By comparison against the raw data, the baseline removal, smoothing, SNV and PCA techniques improved substantially the DNN’s performance (Table 4). This may be related to two explanations: such techniques are capable of standardizing and/or normalizing the entire dataset, which is a must for DNN learning; and, with the PCA, by reducing the dimensionality of the spectra. As of recently, DNN’s are incapable of properly dealing with hyperspectral data, mostly because of the high volumes of highly-correlated variables.

Upon such observations, we theorized that a combination of preprocessing techniques with the PCA could help improve the overall classification task performed by the DNN. When doing so, we performed the same tests as the previous approaches, and an improvement was achieved, mostly on the remaining data that did not return satisfactory results during the initial analysis (being raw data, 1st, and 2nd order-derivatives, and MSC). The overall best result, however, was acquired with the 2nd order derivative + PCA combination (Table 5), returning similar accuracies as of the overall best combination from the shallow learners (1st order derivative with the ExT algorithm). The historical loss of the network was also used to indicate the overall importance of multiple epochs of training to achieve such results (Fig. 9).

DNNs are an important method for data processing, mostly because, in comparison against shallow learners, could return even better results providing that the necessary amount of observations are given. Our experiment consisted of 991 plant samples in total, observed between 8 days of analysis. But even with a reduced number of samples, the DNN method was capable of achieving similar performance in comparison against the overall best shallow learner, providing that the PCA and 2nd order derivatives processes are used in combination. Such techniques demonstrate not only the feasibility of implementing both shallow and deep models but how important the pre-process techniques are in their impact on the overall classification of the algorithms. While shallow learners have the advantage of rapid training, they are limited to a certain extension, not being able to, in most cases, update the models with newer data. DNNs on the other hand, are able of adjusting their weights with transfer-learning capability, and are being used in even other types of spectra data by domain adaptation methods [17].

The ranking approach, however, is an important metric obtained with some shallow learners, and it may even help to construct or isolate important wavelengths from all the spectra to smaller ranges. As of the time of writing, although some approaches are being theorized in the computer vision communities, forums and discussions, and even implemented into defining the most contributive variables of a deep network, it is still difficult to indicate it. Differently from a decision tree, every parameter will be mixed up along the network. Because of that, the initial layers may not be able to indicate how important each variable is, since its importance can vary between the subsequent layers, affecting the importance of another variable. Also, when considering deeper networks, it is important to notice that a lot of its learning occurs at the deeper levels. Our network has a more superficial structure than state-of-the-art deeper networks. However, if the analysis was able to impact the performance of a network of the proposed level, it may as

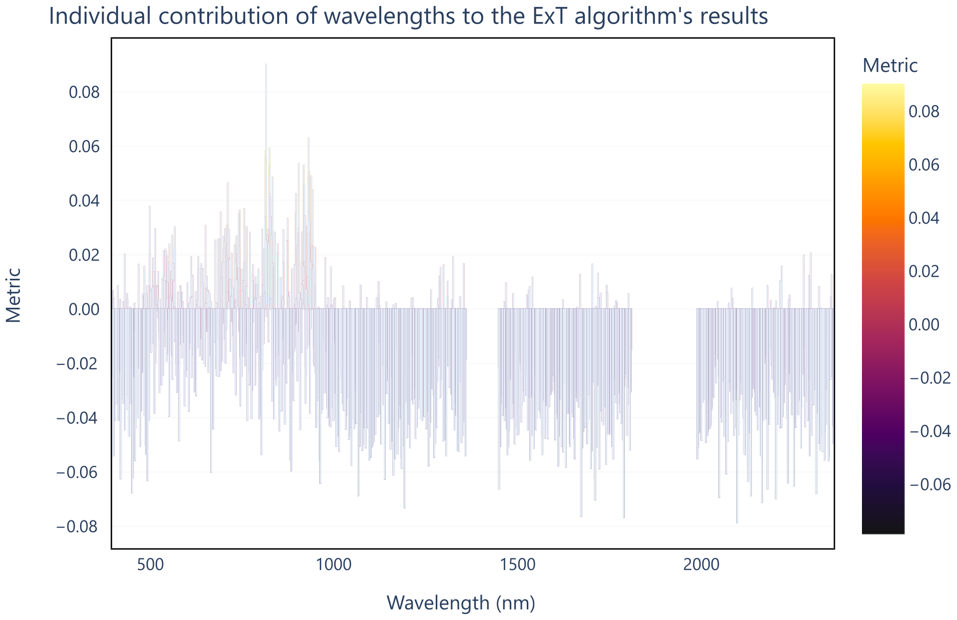

Fig. 7. Individual contribution of wavelengths to the ExT algorithm’s results.

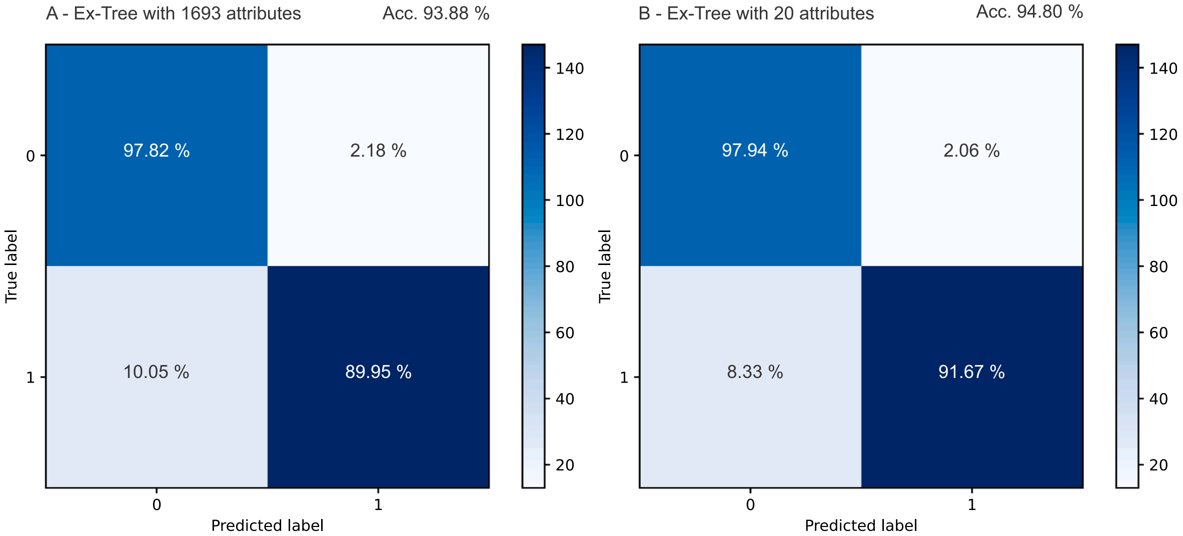

Fig. 8. Confusion matrices comparing the performance of the ExT algorithm with 1st order-derivative data using all the 1,693 wavelengths against the overall 20 most contributive wavelengths.

Table 4
Comparison between the DNN performance over raw reflectance spectrum data and other pre-processes to classify health (control) plants from insect-damaged plants.

	Precision (%)	Recall (%)	F1-score (%)	Accuracy (%)	FP Rate (%)
Raw data	1.44	66.67	2.82	53.69	46.44
Baseline removal	61.59	68.89	65.03	66.44	35.58
Smoothing	94.89	52.42	67.53	58.05	14.00
PCA	48.46	71.59	57.80	69.13	31.90
1st Derivative	0.00	0.00	0.00	51.68	48.32
2nd Derivative	0.00	0.00	0.00	53.02	46.98
SNV	88.72	50.86	64.66	56.71	22.73
MSC	0.00	0.00	0.00	55.37	44.63

Table 5
Testing results returned by the combination of a pre-processing technique and the PCA for the DNN approach.

	Precision (%)	Recall (%)	F1-score (%)	Accuracy (%)	FP Rate (%)
Raw data	48.46	71.59	57.80	69.13	31.90
Baseline removal	49.32	70.21	55.78	67.83	32.11
Smoothing	48.65	50.35	49.48	50.67	49.03
1st Derivative	73.44	56.97	64.16	64.77	25.56
2nd Derivative	**97.74**	**86.09**	**91.72**	**91.95**	**2.04**
SNV	47.06	69.57	56.14	66.44	34.95
MSC	100.00	50.34	66.97	51.34	0.00

well influence more complex architectures. Future studies should take this pre-processing technique into consideration, even when training more deep structures.

As such, it is difficult to understand some aspects of a DNN in modeling a given dataset. However, to help discuss this issue, specifically on the results achieved during this particular experiment, the dimensionality reduction obtained with the PCA may serve as an important point for future research that aims to improve its performance with spectra data analysis to consider. Mostly because with the PCA we can understand some features from our dataset. Initially, an overall analysis of the components was conducted, demonstrating that, differently from the raw data spectrum (i.e. reflectance), the 2nd order-derivative PCA did only accumulated % variance of over 95% after the 100 principal components (PC) (Fig. 10). By analyzing this data, at least 3 limitation points could be observed, being: data with 3 PCs, accumulating 62.1% of the variance; data with 7 PCs, accumulating 71.9%; and data with the 100 PCs, as said, accumulating 94.9%. The 7th component was also the beginning point for the inclination of the cumulative curve, while the first 3 components were chosen because they represent one of the most cumulative variance per ratio.

The tridimensional scatter graphic can help analyze how well data distribution is when considering the first three components (Fig. 11). This representation indicates how well both classes (”Health” and ”Damaged” plants) can be visually separated from the other. Damaged plants appear to have higher component values in PC1 and PC2 than Health plants. This clustering may be one of the key aspects to optimize learning for the DNN model. To also ensure how well the impact of the dimensionality reduction occurs, we also conducted training and testing with these three conditions (3, 7, and 100 PCs) for the 2nd order-derivative (Fig. 12). As such, the overall best result was obtained with 7 PCs, achieving the accuracy acquired at the previous phase. When considering 100 or 3 components, little difference was obtained. For this, we hypothesize that while 100 components may be still much for the DNN model, the 3 components do not offer enough explanation of the dataset. The point of inclination curve at the 7th component 10) might be the most appropriate for this case. Because of that, we encourage that even if novel research aims to implement PCA alongside another pre-processing technique, also evaluate the impact of different components on the model’s accuracy.

It is not an easy task to indicate what is the most appropriate or correct pre-processing task for vegetation spectral data. Here, we conducted one experiment with a highly redundant and complex dataset aiming to solve an agricultural-related problem. As such, while the practical value of said task was already discussed in previous research [18,8], it is still important to note that the appropriate approach to deal with these datasets necessitates a critical investigation. This paper aimed to highlight some of these aspects. With the best-defined model, we were able to indicate the most contributive wavelengths or spectral regions to deal with it (being the beginning of the near-infrared region the most appropriate in this case). Insect damage is known for provoking stress over the plants, and its implications are indicated by their differences in their spectral behavior (Fig. 4). Nonetheless, it is not often that raw data processing results in the low practical use of a model to deal with this problem. By considering the framework and its outcomes

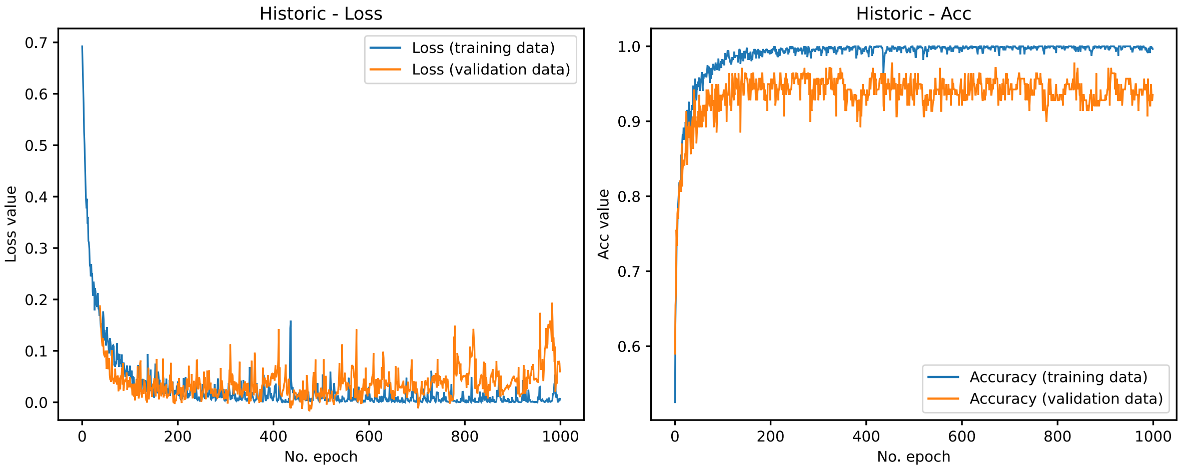

Fig. 9. Historical metrics for the loss and accuracy measurements during the DNN training for the PCA of the 2nd Derivative data.

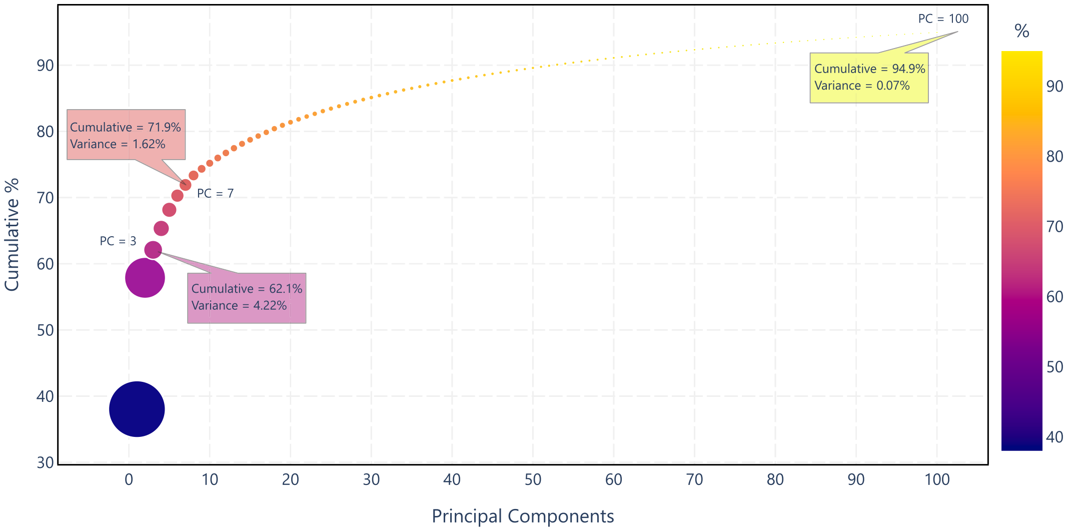

Fig. 10. Scatter plot of the first 100 principal components cumulative variance value.

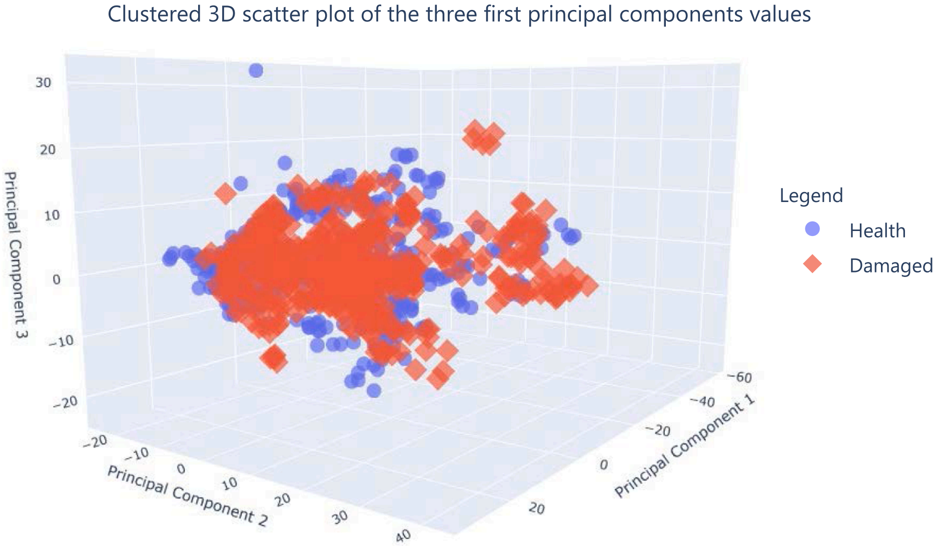

Fig. 11. Clustered 3D scatter plot of the three first principal components values.

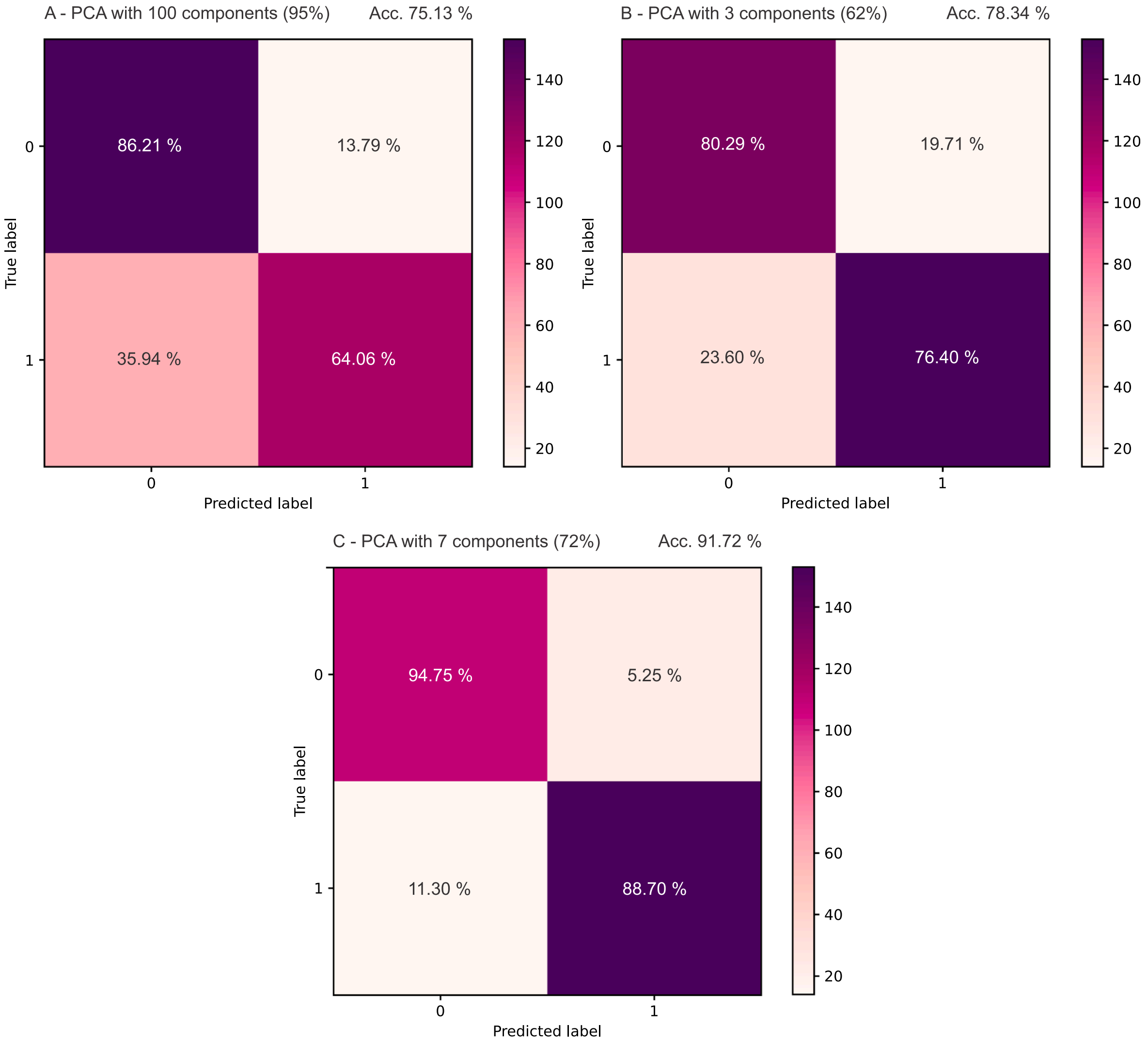

Fig. 12. Confusion matrices of different tests conducted with the DNN method over three specific conditions of the 2nd order-derivative: A- PCA with 100 components; B - PCA with 3 components and; C - PCA with 7 components.

presented here, one may lead to a higher accurate result without the need for additional experiments, which may prove a time-consuming and onerous task.

4. Conclusion

This investigation conducted here brings an original contribution about how the pre-processing techniques can impact machine and deep learning models' performance to separate the insect-damaged from health plants based on hyperspectral reflectance measurements. Our results indicate that the ExT algorithm is the overall best shallow learner to deal with this issue, with an F1-score upper to 93% (Precision and Reall equal to 93%), improving approximately %16 in relation to the raw reflectance data. Among the pre-processing applied techniques, first-order derivative data is the most indicated to segregate the insect-damaged soybean from the control group with machine learning models. We also found out that the spectral region related to the initial range of the near-infrared region (between 784 and 911 nm) is the most contributive wavelength to map insect-damaged soybean plants, and these findings might support the future proximal sensor development to deal specifically with this type of crop monitoring.

Another discovery relates to how the DNN model presents better performance when the PCA method is applied in combination with second-order derivative data (reflectance measurements) are adopted as input data for the network, returning high accuracy values like the best shallow learners. Since DNNs are known to preserve knowledge stored in their weights, transfer learning and domain adaptation of such tasks could be used in continuous modeling. As such, we suggest that the proposal herein presented be tested with other types of crops in the future, highlighting the generalization capabilities of the models hither revised. We also suggest that the information presented, obtained with proximal measurements at wavelength scale, can be implemented in other projects that aim to evaluate the impact of the spectral regions on detecting insect-damaged using imagery acquired by sensors embedded in UAV (Unmanned Aerial Vehicle) platforms by the process of band simulation.

Funding

This research was partially funded by National Council for Scientific and Technological Development (CNPq), project number: 433783/2018–4, 303559/2019–5, 304052/2019–1, and 310517/2020–6, also received financial support from the Brazilian Corporation of Agricultural Research (EMBRAPA), project number: 11.14.09.001.04.00.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors acknowledge the support of the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) (Finance Code 001).

References

[1] J. Abdulridha, O. Batuman, Y. Ampatzidis, UAV-based remote sensing technique to detect citrus canker disease utilizing hyperspectral imaging and machine learning, Remote Sens. 11 (2019) 1373, https://doi.org/10.3390/rs11111373. URL: https://doi.org/10.3390%2Frs11111373.

[2] N.S. Altman, An introduction to kernel and nearest-neighbor nonparametric regression, Am. Stat. 46 (1992) 175–185, https://doi.org/10.1080/00031305.1992.10475879. URL: https://doi.org/10.1080.

[3] B.E. Boser, I.M. Guyon, V.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, pp. 144–152. doi:10.1145/130385.130401. URL: https://doi.org/10.1145.

[4] L. Breiman, Random forests. Mach. Learn. 45 (2001) 5–32. URL: https://doi.org/10.1023 doi:10.1023/a:1010933404324.

[5] S.L. Cessie, J.C.V. Houwelingen, Ridge estimators in logistic regression. Appl. Stat. 41 (1992) 191. URL: https://doi.org/10.2307 doi:10.2307/2347628.

[6] CONAB, 2020. Monitoring of the brazilian harvest 2019/2020.

[7] N.M.A. El-Ghany, S.E.A. El-Aziz, S.S. Marei, A review: application of remote sensing as a promising strategy for insect pests and diseases management, Environ. Sci. Pollut. Res. 27 (2020) 33503–33515, https://doi.org/10.1007/s11356-020-09517-2. URL: https://doi.org/10.1007.

[8] D.E.G. Furuya, L. Ma, M.M.F. Pinheiro, F.D.G. Gomes, W.N. Gonçalvez, J. M. Junior, D. de Castro Rodrigues, M.C. Blassioli-Moraes, M.F.F. Michereff, M. Borges, R.A. Alaumann, E.J. Ferreira, L.P. Osco, A.P.M. Ramos, J. Li, L.A. de Castro Jorge, Prediction of insect-herbivory-damage and insect-type attack in maize plants using hyperspectral data, Int. J. Appl. Earth Obs. Geoinf. 105 (2021) 102608, https://doi.org/10.1016/j.jag.2021.102608.

[9] P. Geurts, D. Ernst, L. Wehenkel, Extremely randomized trees, Mach. Learn. 63 (2006) 3–42, https://doi.org/10.1007/s10994-006-6226-1.

[10] S. González, S. García, J.D. Ser, L. Rokach, F. Herrera, A practical tutorial on bagging and boosting based ensembles for machine learning: Algorithms, software tools, performance study, practical perspectives and opportunities, Inform. Fusion 64 (2020) 205–237, https://doi.org/10.1016/j.inffus.2020.07.007. URL: https://doi.org/10.1016.

[11] S. Haykin, Neural Networks: A Comprehensive Foundation, Prentice-Hall, 1993.

[12] J.R. Jensen, Remote sensing of the environment: an earth resource perspective second edition. volume 1. Prentice Hall, 2014.

[13] G.H. John, Estimating Continuous Distributions in Bayesian Classifiers, Robotics (1995).

[14] Z.Y. Liu, J.G. Qi, N.N. Wang, Z.R. Zhu, J. Luo, L.J. Liu, J. Tang, J.A. Cheng, Hyperspectral discrimination of foliar biotic damages in rice using principal component analysis and probabilistic neural network, Precision Agric. 19 (2018) 973–991, https://doi.org/10.1007/s11119-018-9567-4. URL: https://doi.org/10.1007.

[15] A.K. Mahlein, Plant disease detection by imaging sensors – parallels and specific demands for precision agriculture and plant phenotyping, Plant Dis. 100 (2016) 241–251, https://doi.org/10.1094/pdis-03-15-0340-fe.

[16] T. Nyabako, B.M. Mvumi, T. Stathers, S. Mlambo, M. Mubayiwa, Predicting prostephanus truncatus (horn) (coleoptera: Bostrichidae) populations and associated grain damage in smallholder farmers' maize stores: A machine learning approach, J. Stored Prod. Res. 87 (2020) 101592, https://doi.org/10.1016/j.jspr.2020.101592. URL: https://doi.org/10.1016.

[17] L.P. Osco, J. Marcato Junior, A.P. Marques Ramos, L.A. de Castro Jorge, S. N. Fatholahi, J. de Andrade Silva, E.T. Matsubara, H. Pistori, W.N. Gonçalves, J. Li, A review on deep learning in uav remote sensing, Int. J. Appl. Earth Obs. Geoinf. 102 (2021) 102456, https://doi.org/10.1016/j.jag.2021.102456. URL: https://www.sciencedirect.com/science/article/pii/S030324342100163X.

[18] A.P.M. Ramos, F.D.G. Gomes, M.M.F. Pinheiro, D.E.G. Furuya, W.N. Gonçalvez, J. M. Junior, M.F.F. Michereff, M.C. Blassioli-Moraes, M. Borges, R.A. Alaumann, V. Liesenberg, L.A. de Castro Jorge, L.P. Osco, Detecting the attack of the fall armyworm (spodoptera frugiperda) in cotton plants with machine learning and spectral measurements, Precision Agric. (2021), https://doi.org/10.1007/s11119-021-09845-4. URL: 10.1007/s11119-021-09845-4.

[19] A.P.M. Ramos, L.P. Osco, D.E.G. Furuya, W.N. Gonçalves, D.C. Santana, L.P. R. Teodoro, C.A. da Silva Junior, G.F. Capristo-Silva, J. Li, F.H.R. Baio, J.M. Junior, P.E. Teodoro, H. Pistori, A random forest ranking approach to predict yield in maize with uav-based vegetation spectral indices, Comput. Electron. Agric. 178 (2020) 105791, https://doi.org/10.1016/j.compag.2020.105791. URL: https://doi.org/10.1016.

[20] Åsmund Rinnan, 2014. Pre-processing in vibrational spectroscopy-when, why and how. doi:10.1039/c3ay42270d.

[21] Åsmund Rinnan, F. van den Berg, S.B. Engelsen, Review of the most common pre-processing techniques for near-infrared spectra, 2009. doi:10.1016/j.trac.2009.07.007.

[22] A. Tageldin, D. Adly, H. Mostafa, H.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.

[23] H. Yao, D. Lewis, Spectral preprocessing and calibration techniques, 2010. doi:10.1016/B978-0-12-374753-2.10002-4.

[24] D. Zhang, Y. Ding, P. Chen, X. Zhang, Z. Pan, D. Liang, Automatic extraction of wheat lodging area based on transfer learning method and deeplabv3/mathplus network, Comput. Electron. Agric. 179 (2020) 105845, https://doi.org/10.1016/j.compag.2020.105845. URL: https://doi.org/10.1016.

[25] J. Zhang, Y. Huang, R. Pu, P. Gonzalez-Moreno, L. Yuan, K. Wu, W. Huang, Monitoring plant diseases and pests through remote sensing technology: A review, Comput. Electron. Agric. 165 (2019) 104943, https://doi.org/10.1016/j.compag.2019.104943. URL: https://doi.org/10.1016.