Integrating LIBS and Machine Learning to Identify Aphelenchoides Besseyi Infection in Asymptomatic Soybean Leaves

PDF

Luís Carlos Leva Borduchi ; Helga Maria Darezzo ; Daniele Souza ; Debora Marcondes Bastos Pereira Milori All Authors

Abstract:
This study evaluated the use of Double Pulse Laser-Induced Breakdown Spectroscopy (DP-LIBS) combined with machine learning to detect asymptomatic soybean leaves infected by Aphelenchoides besseyi, the causal agent of Green Stem and Foliar Retention Syndrome (GSFR). Spectral lines corresponding to macro- and micronutrients were selected as input features for classification models. The Support Vector Machine achieved 95.7% accuracy, while the Multilayer Perceptron reached 92.9%. The enhanced sensitivity of DP-LIBS allowed detection of additional micronutrient signals, enhancing classification accuracy. Results demonstrate the potential of LIBS as a rapid, non-destructive diagnostic tool for early disease detection in precision agriculture.

Published in: 2025 SBFoton International Optics and Photonics Conference (SBFoton IOPC)

Date of Conference: 21-24 September 2025

Date Added to IEEE *Xplore*: 31 October 2025

DOI: 10.1109/SBFotonIOPC66433.2025.11218396

Conference Location: São Pedro/SP, Brazil

ISBN Information:

ISSN Information:

Electronic ISSN: 2837-4967

Print on Demand(PoD) ISSN: 2837-4959

Funding Agency:

Integrating LIBS and Machine Learning to Identify *Aphelenchoides besseyi* Infection in Asymptomatic Soybean Leaves

Luís Carlos Leva Borduchi[a,b], Helga Maria Darezzo[a], Daniele Souza[a], Debora Marcondes Bastos Pereira Milori[a]

[a]Embrapa Instrumentação, Rua XV de Novembro, 1452, CEP 13560-970 São Carlos, SP, Brazil
[b]Instituto de Física de São Carlos, Universidade de São Paulo, Avenida Trabalhador São Carlense, 400, CEP 13566-590 São Carlos, SP, Brazil

This study evaluated the use of Double Pulse Laser-Induced Breakdown Spectroscopy (DP-LIBS) combined with machine learning to detect asymptomatic soybean leaves infected by Aphelenchoides besseyi, the causal agent of Green Stem and Foliar Retention Syndrome (GSFR). Spectral lines corresponding to macro- and micronutrients were selected as input features for classification models. The Support Vector Machine achieved 95.7% accuracy, while the Multilayer Perceptron reached 92.9%. The enhanced sensitivity of DP-LIBS allowed detection of additional micronutrient signals, enhancing classification accuracy. Results demonstrate the potential of LIBS as a rapid, non-destructive diagnostic tool for early disease detection in precision agriculture.

Keywords—*LIBS, Aphelenchoides besseyi, early diagnosis, machine learning, leaf analysis*

I. Introduction

Soybean (*Glycine max*) is currently one of the most important agricultural crops worldwide, holding a central role in the production of food, animal feed, biofuels, and various industrial derivatives. According to data from the Brazilian National Supply Company (Conab, 2024), Brazil remains the world's largest soybean exporter, accounting for approximately 52% of global export volume. However, efforts to increase productivity have faced significant challenges, particularly losses caused by emerging pests and diseases, such as green stem and foliar retention syndrome (GSFR) [1].

GSFR has been associated with the presence of the nematode *Aphelenchoides besseyi* and can lead to yield reductions of up to 60%, especially in Brazilian states such as Maranhão, Tocantins, Pará and Mato Grosso. Early detection of the disease remains a major challenge, as infected plants can remain visually healthy (asymptomatic) for extended periods, hindering effective management and increases the risk of disease spread and economic losses.

In recent years, laser-induced breakdown spectroscopy (LIBS) [2] has emerged as a promising technique for rapid, multielement analysis of biological and agricultural materials [3], [4]. The ability of LIBS to detect macro- and micronutrients [5] in plant tissues quickly, without chemical preparation, and with potential for field applications makes it a strategic tool for diagnosing diseases that impact the nutritional metabolism of crops.

Previous studies [6] suggest that imbalances in nutrients such as calcium (Ca), potassium (K), and magnesium (Mg) may be related to the onset of GSFR. However, distinguishing healthy from asymptomatic infected leaves still lacks objective, sensitive, and scalable methods. In this context, the integration of machine learning techniques with LIBS can provide innovative solutions for classifying samples based on their spectral profiles.

In this study, we propose the targeted selection of representative spectral lines of macro- and micronutrients in LIBS spectra of soybean leaves to enable the differentiation between healthy and asymptomatic GSFR-infected samples. To achieve this, supervised classification models were developed using artificial neural networks (ANN) and support vector machines (SVM), with a focus on early and reliable diagnosis.

II. Materials and Methods

A. Leaf Samples and Preparation

Soybean leaf samples were obtained from an experiment at Embrapa Soja (Londrina, PR, Brazil) using the *Brasmax Apolo RR* cultivar. Ninety pots were prepared: 45 inoculated with Aphelenchoides besseyi and 45 controls. Leaves were collected on the 4th, 7th, and 11th days after inoculation, selecting the uppermost fully expanded trifoliate leaf from each plant. In total, 135 asymptomatic samples from inoculated plants and 135 from healthy plants were collected.

In the laboratory, samples were washed, dried at 36 °C for 72 h, ground in liquid nitrogen, sieved (<250 µm), and pressed into 12.5 mm pellets (300 mg) under 2.4 kbar. Pellets were stored under controlled humidity.

B. LIBS Measurements

LIBS analysis used two Nd:YAG lasers: one at 1064 nm (50 mJ, 8 ns, 119 J/cm²) and another at 532 nm (70 mJ, 4 ns, 600 J/cm²), operated in a double-pulse (DP-LIBS) configuration. Emission was captured by an Echelle spectrometer (275–770 nm, 21–37 pm resolution) with an ICCD detector (500 ns gate width). One hundred shots were taken per pellet using an automated XY stage. Emission lines were identified via the NIST database. Data processing included outlier removal, spectral averaging, baseline correction, and peak fitting.

C. Data Processing and Analysis

The evaluation of the DP-LIBS spectra was conducted by quantifying the spectral areas corresponding to the following elements: Al II, B I, C I, Ca II, Fe II, K I, Mg II, Mn II, P I, Si I, Zn II, as well as the CN molecular band. In total, the analysis considered 12 emission lines.

This work was supported by CNPq (grants 440226/2021-0, 384173/2024-1 and 162231/2024-5) and FAPESP (grant 2013/07276-1 and 2022/05451-0).

2025 SBFoton International Optics and Photonics Conference (SBFoton IOPC) | 979-8-3315-9497-8/25/$31.00 ©2025 IEEE | DOI: 10.1109/SBFOTONIOPC66433.2025.11218396

Data processing was performed using machine learning algorithms implemented in Python within the Google Colab environment, where the support vector machine (SVM) and the multilayer perceptron (MLP) neural network achieved the best performance.

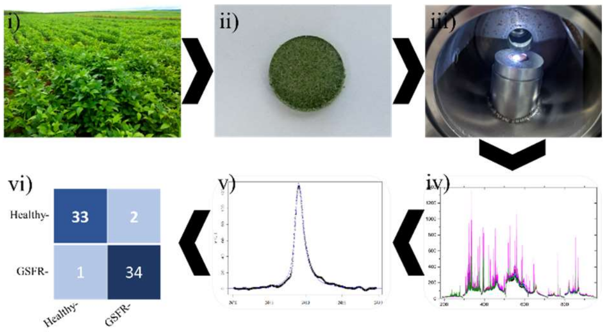

Fig. 1: Workflow illustrating the process of sample preparation, LIBS analysis, and classification. (i) Collection of soybean leaf samples, (ii) Pellet obtained from dried and ground leaf samples, (iii) DP- LIBS plasma generation on the pellet surface, (iv) Acquired emission spectra representing multiple elements and molecular bands, (v) Example of a processed spectral line selected for model input, (vi) Confusion matrix summarizing the classification performance of the machine learning models, differentiating healthy and asymptomatic GSFR-infected samples.

During the classification workflow, each sample was first transformed into a numerical feature vector representing the spectral intensities of the selected elements. The vectors were processed by the classification algorithms, which applied mathematical transformations to identify patterns distinguishing the classes. For example, in the case of the MLP, the input data are propagated through multiple layers of artificial neurons. Each neuron computes a weighted sum of the inputs and applies a nonlinear activation function, enabling the network to learn complex relationships among variables. The network is trained to minimize classification error by adjusting the internal connection weights iteratively through backpropagation.

In the case of the SVM, the algorithm seeks the optimal hyperplane that maximally separates vectors from different classes. To achieve this, it uses a kernel function to project the data into a higher-dimensional space, where the separation can be performed more effectively. The SVM then makes predictions based on the relative position of input vectors to this decision boundary.

Regardless of the model used, the final step consists of applying the learned transformations to new samples to determine, based on their position in the decision space, the class to which they most likely belong. This approach enables the classifiers to accurately identify whether a sample corresponds to a healthy or an asymptomatic leaf.

The original dataset contained 270 samples and was divided into two groups: 74% (200 samples) used for training and 26% (70 samples) reserved for testing. As the data were labeled, the task was framed as supervised classification. Model training was validated using 10-fold cross-validation to ensure robustness and reliability of the results.

III. RESULTS AND DISCUSSION

Initially, specific emission lines of chemical elements with recognized importance in soybean leaf analysis were selected: Al II, B I, C I, Ca II, Fe II, K I, Mg II, Mn II, P I, Si I, Zn II, and the CN molecular band. The areas of these spectral lines, extracted from LIBS measurements, were calculated and used as input variables for machine learning models. For model development and cross-validation, 74% of the samples were used for training, while the remaining 26% composed the test set for final validation.

These elements and the CN molecular band were chosen because they play essential roles in soybean physiology and metabolism. They are involved in processes such as photosynthesis, protein synthesis, nutrient transport, osmotic regulation, and defense against biotic and abiotic stresses. The central hypothesis was that healthy and infected leaves would exhibit subtle alterations in the concentrations or emission profiles of these nutrients, detectable through spectroscopic analysis. For example, carbon and nitrogen (reflected by the CN molecular band) are indicators of general metabolic status, while calcium, magnesium, and potassium are fundamental to osmotic balance and enzymatic activity. Iron, manganese, and zinc participate in redox processes and photosynthesis; phosphorus contributes to energy metabolism; boron and silicon are associated with structural integrity and stress tolerance; and aluminum, although potentially toxic, can signal changes in nutrient absorption or environmental stress. By selecting these chemically meaningful variables, the analysis focused on emission lines most likely to reflect physiological changes associated with disease progression.

The classification models achieved high accuracy in differentiating healthy soybean leaves from asymptomatic leaves infected by *Aphelenchoides besseyi*, the causal agent of green stem and foliar retention (GSFR). Among the algorithms evaluated, the support vector machine (SVM) model achieved the highest performance, reaching an accuracy of 95.7%, followed by the multilayer perceptron (MLP), which achieved 92.9% accuracy on the test set (Table II).

TABLE I. CONFUSION MATRICES FOR THE MLP AND SVM CLASSIFIERS. ROWS CORRESPOND TO THE ACTUAL CLASS LABELS, AND COLUMNS INDICATE PREDICTED LABELS FOR EACH MODEL.

	MLP Predicted: Healthy	MLP Predicted: GSFR	SVM Predicted: Healthy	SVM Predicted: GSFR
Actual: Healthy	34	1	33	2
Actual: GSFR	2	33	3	32

These results are particularly relevant given the challenge of early GSFR diagnosis, especially during the asymptomatic infection stage ($\leqslant 11$ days). In this early phase, visual inspection is ineffective, as infected plants often appear normal, making timely control measures difficult and increasing the risk of disease spread and economic losses. The ability of the SVM and MLP models to detect subtle physiological alterations solely from spectral data underscores the potential of LIBS, combined with machine learning, as a promising tool for early, rapid, and non-destructive plant health monitoring.

Furthermore, the literature already highlights the effective use of LIBS for quantifying nutrients in soybean leaves [5], even under the influence of matrix effects—physical and chemical variations among samples that can interfere with spectral emission. This evidence supports the robustness of the technique for both qualitative and quantitative analysis of plant tissues, consolidating its applicability within precision agriculture.

TABLE II. STATISTICAL METRICS OF PREDICTED RESULTS

Metrics	MLP	SVM
Correctly classified	67	65
Accuracy (%)	95.7	92.9

A key methodological advance of this study lies in the targeted selection of spectral lines corresponding to macro- and micronutrients relevant to soybean metabolism. Unlike conventional approaches that rely on the full-spectrum LIBS —which may include noise and non-informative emissions— the strategy adopted here prioritizes only chemically meaningful spectral regions. This filtering reduces the inclusion of spurious variables, enhances model interpretability, and improves classification accuracy.

It is important to highlight that all misclassified samples (Table I) corresponded to the first collection carried out on the 4th day after inoculation. In contrast, the classification performance for samples collected on the 7th and 11th days was 100% accurate for both models. This finding suggests that the earliest stages of infection may result in less distinct or consistent spectral features, making differentiation more challenging. As the infection progresses, however, the physiological and biochemical changes become more detectable by LIBS, significantly improving the reliability of the classification process.

The high performance achieved (Table II)—95.7% accuracy for SVM and 92.9% for MLP—confirms the discriminative power of the selected variables and reinforces the hypothesis that nutritional imbalances induced by early-stage infection generate detectable spectral signatures, even in the absence of visible symptoms. Accordingly, this study validates the use of DP-LIBS, combined with supervised learning algorithms and intelligent feature selection, as an effective, precise, and scalable approach for plant disease diagnosis.

One of the main advantages of using the DP-LIBS system is its enhanced sensitivity compared to conventional single-pulse configurations [7], [8]. The application of a pre-ablation laser pulse followed by a second excitation pulse significantly increases the plasma temperature and lifetime, resulting in stronger and more stable emission signals. This improved sensitivity enabled the detection of a broader range of elements in the samples, particularly micronutrients that typically produce weaker spectral lines under single-pulse conditions. Consequently, the inclusion of these additional micronutrient emissions in the model contributed to higher discriminatory power and improved the overall classification performance.

Overall, this work demonstrates that integrating LIBS with careful spectral line selection and machine learning algorithms offers a diagnostic alternative that is more robust, accurate, and interpretable than approaches relying on full-spectrum analysis. The results contribute significantly to advancing precision agriculture and developing innovative tools for plant health management.

IV. CONCLUSIONS

This study demonstrated the potential of LIBS, combined with machine learning algorithms, as an effective tool for the early diagnosis of GSFR in soybean plants. The proposed approach, based on the careful selection of spectral emissions from both macro- and micronutrients, enabled the development of highly accurate classification models capable of differentiating healthy and asymptomatic infected leaves.

The SVM and MLP algorithms achieved accuracies of 95.7% and 92.9%, respectively, demonstrating the ability of selected spectral features to capture subtle nutritional imbalances induced by *Aphelenchoides besseyi* infection. By avoiding the indiscriminate use of the entire LIBS spectrum and focusing on agronomically relevant variables, this strategy also improved the interpretability and robustness of the models.

Beyond confirming the feasibility of LIBS as a rapid, non-destructive, reagent-free technique for precision agriculture applications, the results pave the way for the development of portable field diagnostic systems. Overall, this approach represents a significant advancement in plant health monitoring and integrated disease management, with strong potential to directly impact the productivity and sustainability of soybean cultivation.

ACKNOWLEDGMENT

We thank Embrapa Instrumentation (grants 20.24.00.059.00.00 and 30.20.90.029.00.00), the National Laboratory of Agri-photonics (Lanaf), the Brazilian Agency for Research and Industrial Innovation (EMBRAPII) for grants PIAG-2406.0001 and PIAG-2406.0002, the National Photonics Laboratories System (Sisfoton) and the National Council for Scientific and Technological Development (CNPq) for grant 440226/2021-0, 384173/2024-1 and 162231/2024-5, and the Center for Research on Optics and Photonics (CEPOF) and the São Paulo Research Foundation (FAPESP) for grant 2013/07276-1 and 2022/05451-0, for their financial support of this project.

REFERENCES

[1] M. C. Meyer, L. Favoreto, D. Klepker, and F. C. Marcelino-Guimarães, “Soybean green stem and foliar retention syndrome caused by Aphelenchoides besseyi,” *Trop. Plant Pathol.*, vol. 42, no. 5, pp. 403–409, 2017, doi: 10.1007/s40858-017-0167-z.

[2] D. A. Cremers and L. J. Radziemski, *Handbook of Laser-Induced Breakdown Spectroscopy*. Oxford, UK: John Wiley & Sons Ltd, 2013. doi: 10.1002/9781118567371.

[3] G. S. Senesi, J. Cabral, C. R. Menegatti, B. Marangoni, and G. Nicolodelli, “Recent advances and future trends in LIBS applications to agricultural materials and their food derivatives: An overview of developments in the last decade (2010–2019). Part II. Crop plants and their food derivatives,” *TrAC - Trends Anal. Chem.*, vol. 118, pp. 453–469, 2019, doi: 10.1016/j.trac.2019.05.052.

[4] V. K. Singh *et al.*, “Review: Application of libs to elemental analysis and mapping of plant samples,” *At. Spectrosc.*, vol. 42, no. 2, pp. 99–113, 2021, doi: 10.46770/AS.2020.201.

[5] L. C. L. Borduchi, D. M. B. P. Milori, M. C. Meyer, and P. R. Villas-Boas, “Reducing matrix effects on the quantification of Ca, Mg, and Fe in soybean leaf samples using calibration-free LIBS and one-point calibration,” *Spectrochim. Acta - Part B At. Spectrosc.*, vol. 198, Dec. 2022, doi: 10.1016/j.sab.2022.106561.

[6] A. C. Ranulfi *et al.*, “Nutritional characterization of healthy and Aphelenchoides besseyi infected soybean leaves by laser-induced breakdown spectroscopy (LIBS),” *Microchem. J.*, vol. 141, no. February 2018, pp. 118–126, 2018, doi: 10.1016/j.microc.2018.05.008.

[7] L. C. Leva Borduchi, C. R. Menegatti, D. M. Bastos Pereira Milori, H. J. Izário Filho, and P. R. Villas-Boas, “Application of one-point calibration LIBS for quantification of analytes in samples with distinct matrix characteristics: a case study with Hg,” *J. Anal. At. Spectrom.*, vol. 38, no. 5, pp. 1155–1163, Mar. 2023, doi: 10.1039/d2ja00399f.

[8] P. R. Villas-Boas and L. C. L. Borduchi, “A statistical definition of limit of detection for calibration-free laser-induced breakdown spectroscopy,” *Spectrochim. Acta - Part B At. Spectrosc.*, vol. 205, Jul. 2023, doi: 10.1016/j.sab.2023.106690.