Detection of Escherichia Coli in Tomatoes Using Laser-Induced Breakdown Spectroscopy and Machine Learning

PDF

Helga Maria Darezzo ; Luís Carlos Leva Borduchi ; Raphael Antonio Caface ; 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 Escherichia coli in grape tomatoes. A total of 216 samples were analyzed, with spectral data from elements like Mg, Zn, and P used as input features for classification. The Multilayer Perceptron (MLP) neural network achieved 92.4% accuracy in the test set, outperforming the Random Forest model. The results demonstrate the potential of DP-LIBS as a rapid, non-destructive, and accurate method for food safety monitoring, paving the way for broader applications in the detection of other pathogens in agricultural products.

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

Date of Conference: 21-24 September 2025

DOI: 10.1109/SBFotonIOPC66433.2025.11218567

Date Added to IEEE *Xplore*: 31 October 2025

Conference Location: São Pedro/SP, Brazil

ISBN Information:

Electronic ISBN:979-8-3315-9497-8

Print on Demand(PoD) ISBN:979-8-3315-9498-5

ISSN Information:

Electronic ISSN: 2837-4967

Print on Demand(PoD) ISSN: 2837-4959

Detection of *Escherichia coli* in Tomatoes Using Laser-Induced Breakdown Spectroscopy and Machine Learning

Helga Maria Darezzo[a], Luís Carlos Leva Borduchi[a,b], Raphael Antonio Caface[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 Escherichia coli in grape tomatoes. A total of 216 samples were analyzed, with spectral data from elements like Mg, Zn, and P used as input features for classification. The Multilayer Perceptron (MLP) neural network achieved 92.4% accuracy in the test set, outperforming the Random Forest model. The results demonstrate the potential of DP-LIBS as a rapid, non-destructive, and accurate method for food safety monitoring, paving the way for broader applications in the detection of other pathogens in agricultural products.

Keywords—LIBS, grape tomatoes, Escherichia coli, machine learning

I. Introduction

In 2024, agribusiness in Brazil was a key driver of the national economy, contributing 23.2% to the Gross Domestic Product (GDP) [1]. This underscores the sector's significance in both economic and social development, largely fueled by advancements in technology and research. However, increasing concerns about environmental quality, public health, and food safety, particularly regarding the quality of water used in irrigation, are gaining attention. Contaminated water can harbor harmful microorganisms such as *Escherichia coli* (*E. coli*), which is a well-known pathogen responsible for severe gastrointestinal infections. The contamination of vegetables by *E. coli* is a widespread concern, occurring both pre-harvest through irrigation and post-harvest due to improper cleaning and storage practices that facilitate bacterial proliferation [2].

Tomatoes, one of the most important agricultural crops globally [2], are the second most consumed vegetable worldwide, both fresh and processed. Given the difficulty in completely eradicating contaminants from food, regulatory agencies establish tolerance limits for each type. In Brazil, the National Health Surveillance Agency (ANVISA), following World Health Organization (WHO) guidelines, stipulates that the presence of *E. coli* in food should not exceed 10^2 CFU/g.

Ensuring compliance with these limits is crucial, yet traditional methods for controlling contamination are often slow, costly, and waste-intensive. This scenario highlights the urgent need for innovative analytical techniques that are faster, more accurate, and economically viable. Optical techniques like double-pulse laser-induced breakdown spectroscopy (DP-LIBS) have gained prominence in scientific research due to their high sensitivity, precision, and capacity for real-time analysis [3]. DP-LIBS, which analyzes emissions generated by plasma formation on the sample surface, has proven especially effective for identifying and characterizing bacterial cells. In recent years, LIBS has established itself as a powerful tool for rapid detection and analysis, providing unmatched precision in bacteriological studies [4], [5].

This study explores the potential and accuracy of the DP-LIBS technique in detecting biological contamination by *E. coli* in grape tomatoes using machine learning algorithms. The findings demonstrate the technique's capability to accurately identify the presence of the bacterium, paving the way for more efficient methods in food quality control.

II. Sample Preparation

Grape tomatoes were purchased from a local market, packaged in 150g polyethylene trays. The tomatoes were initially selected based on size and firmness, resulting in a final selection of 216 tomatoes. These selected tomatoes were then carefully washed with a sponge and neutral detergent, followed by a 20-minute disinfection in a 150 ppm sodium hypochlorite solution. After drying, the tomatoes were randomly divided into two groups with 108 fruits each.

In this study, we utilized the *Escherichia coli* strain ATCC 25922, obtained from Embrapa. Two saline solutions were prepared: one containing the bacterium at a concentration of 10^2 CFU/g and another without the bacterium. Then, considering the two groups with 108 tomatoes, one group was immersed in the saline solution containing *E. coli* (inoculated treatment) and the other in the solution without the bacterium (control), for 3min in each solution. After immersion, the tomatoes were air-dried for 24 hours on trays maintained at 23°C and 71% relative humidity. Subsequently, the tomatoes were sliced and stored in beakers for lyophilization. The samples were first ultra-frozen at -80°C for 12 hours and then lyophilized for 3 days. This lyophilization process removed moisture from the samples, facilitating the preparation of pellets for analysis using the DP-LIBS technique.

The lyophilized samples were first ground using a ball mill at 30 Hz for 60 seconds. After grinding, the samples were weighed and pressed into pellets using a hydraulic press at 2.4 kBar for 60 seconds, forming pellets

This work was supported by CNPq (grants 440226/2021-0, 308339/2021-5, 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/SBFOTONIOPC66433.2025.11218567

with a diameter of 12.5 mm each. A total of 108 pellets were produced for each treatment, resulting in 216 samples in total. The samples were then stored in a freezer at -10°C until analysis by the DP-LIBS technique.

III. LIBS Technique

LIBS is a multi-elemental analytical technique based on atomic, ionic, and molecular emission spectroscopy [6]. It involves focusing a high-energy laser pulse on the sample's surface to create a plasma and resolving the atomic, ionic, and molecular emissions using a spectrometer with appropriate resolution. In addition to its versatility in analyzing solid, liquid, and gaseous materials, LIBS offers both qualitative and quantitative capabilities for elemental analysis [7], [8]. The technique is extremely fast, precise, cost-effective, and does not generate polluting waste. In recent years, the international scientific community has reported the use of LIBS as an effective analytical tool not only for detection but also for rapid analysis and characterization of the elemental composition of specimens [5]. It has been described as an extremely precise technique for a wide range of bacteriological investigations. LIBS spectra have been used to identify different bacterial species through the concentrations observed in the spectral emission lines of elements such as Mg, P, K, Ca, and the CN molecular band, serving as spectral markers for bacterial identification and discrimination [4].

IV. LIBS Measurements

The samples were analyzed using a Q-switched Nd-YAG laser (Quantel, Brilliant) equipped with a second-harmonic generator operating at 532 nm, delivering a pulse duration of 4 ns, pulse energy of 60 mJ, and a fluence of 950 J/cm². Additionally, a Q-switched Nd-YAG laser (Quantel, Ultra) operating at 1064 nm (infrared) with an 8 ns pulse duration, 20 Hz repetition rate, 50 mJ pulse energy, and 510 J/cm² fluence was employed. For spectral analysis, an LTB Echelle spectrometer (Arielle model) with a resolution of 21–37 pm and a spectral range of 275 to 770 nm, coupled with an ICCD, was utilized. The spectra were collected with a 500 ns delay time, 750 ns interpulse delay, and a 10 μs gate width. A total of 30 laser shots were fired at different points on one side of each pellet using an automated XY linear stage, each spectrum was generated by accumulating ten laser shots at each position.

V. Data Processing and Analysis

The evaluation of the DP-LIBS spectra was carried out by quantifying the spectral areas corresponding to the elements Si, P, C, Mg, Zn, Al, K, Ca, and the CN molecular band, selected based on previous related studies [4]. For magnesium (Mg), two emission lines were considered (Fig 1), and for zinc (Zn), three emission lines were used. In total, the DP-LIBS analysis was based on 12 emission lines, representing 9 chemical elements (including the CN molecule), resulting in 12 evaluation parameters.

Data processing was performed using machine learning algorithms, specifically an artificial neural network (ANN) and a random forest model, both implemented in python within the Google Colab environment. The workflow included two main steps: data standardization and classification. The standardization step prepared the dataset for classification algorithms by adjusting the scale and applying appropriate normalization techniques. The classification step aimed to identify sample classes (control and inoculated) using supervised learning models, determining how well the samples could be differentiated [9].

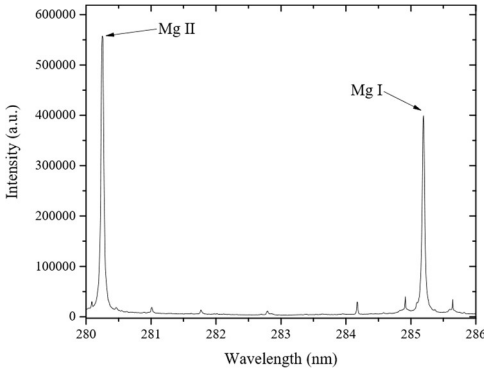

Fig. 1: LIBS spectrum with the Mg emissions used

The original dataset contained 216 samples and was split into two groups: 70% (150 samples) used to train the models, while the remaining 30% served as the test set (66 samples). Since the data were labeled, the task was a supervised classification. Model training was validated using 10-fold cross-validation to ensure robustness and reliability.

VI. Results and Discussion

First, the elements present in the LIBS spectra of tomato samples were identified, including Si, P, C, Mg, Zn, Al, K, Ca, and the molecular CN band. Spectral areas corresponding to these elements were calculated, and the dataset was constructed with all calculated areas for the 216 samples. Related studies have employed these elements as features for both the detection and differentiation of E. coli using the LIBS technique [4]. These elements are integral components of bacterial cells, and contamination by E. coli can modify their concentrations within tomato tissues. Such alterations can be accurately detected through the high-resolution analytical capabilities of LIBS.

In the search for the best accuracy in distinguishing between control and inoculated classes, several classification algorithms were tested. Initially, the models were trained using 150 samples (training set), followed by evaluation using 66 samples (test set). The performance metrics used included accuracy, which measures the percentage of correctly classified samples, and the confusion matrix, which provides detailed insight into the correct and incorrect classifications for each class (control and inoculated), in both training and test datasets.

The most effective algorithms were the multilayer perceptron (MLP) and random forest. The MLP, a feedforward neural network with hidden layers of 100 neurons, captures complex nonlinear patterns through weighted inputs and activation functions. Random Forest is an ensemble learning method that builds multiple decision trees and combines their outputs to improve classification accuracy.

The MLP model achieved the highest accuracy in classifying the control and *E. coli*-inoculated samples, with 99.3% accuracy in the training set and 92.4% in the test set (Table 1). The confusion matrix for the MLP model showed

high sensitivity in identifying *E. coli*-inoculated samples, with 31 correct predictions and only 2 misclassifications. In contrast, the Random Forest model showed signs of overfitting, with 100% accuracy in training but only 84.8% in testing, indicating the model may have memorized the training data and struggled to accurately classify new samples (test data).

TABLE I. STATISTICAL METRICS OF PREDICTED RESULTS

Metrics	Multilayer Perceptron	Random Forest
Correctly classified	61	56
Accuracy (%)	92.4	84.8

The strong performance of the MLP model highlights the potential of the DP-LIBS technique in detecting elemental alterations due to biological contamination by *E. coli* in tomatoes. In particular, the elements Mg, Zn, and P (Figure 2) were key discriminators between the control and inoculated groups.

These results hold significant value for the scientific community and for real-world applications. The demonstrated effectiveness of the DP-LIBS technique, especially when combined with robust machine learning models like the MLP neural network, reinforces its potential as a rapid, non-destructive, and reagent-free method for detecting biological contamination in food products. In practical terms, this approach could be integrated into quality control workflows in the food industry, improving early detection of pathogens such as *E. coli*, and reducing the risk of contaminated products reaching consumers. From a scientific perspective, the ability to differentiate sample classes with high accuracy using elemental composition alone highlights the sensitivity of LIBS and opens new avenues for its application in microbiological diagnostics, precision agriculture, and environmental monitoring.

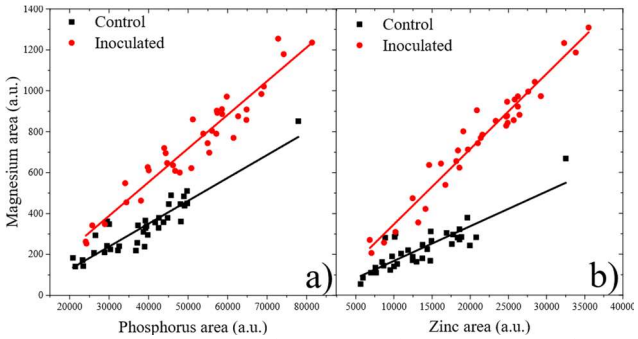

Fig. 2: Correlation of areas between two elements in test samples.

VII. CONCLUSION

The proposed methodology using DP-LIBS demonstrated high effectiveness in identifying and classifying *Escherichia coli* contamination in tomato samples. By analyzing elemental spectral data, the technique enabled clear differentiation between control and inoculated groups, achieving impressive classification performance. Among the tested algorithms, the Multilayer Perceptron (MLP) neural network provided the highest accuracy, with 99.3% in training and 92.4% in testing, while Random Forest reached 100% accuracy in training but exhibited signs of overfitting, with reduced performance on the test set (84.8% accuracy). These findings highlight the sensitivity of DP-LIBS to elemental changes associated with biological contamination, particularly involving key elements such as Mg, Zn, and P.

The results underscore the potential of DP-LIBS as a powerful, fast, non-destructive, and reagent-free tool for food safety applications, capable of detecting contamination through elemental signatures. Its integration with machine learning models enhances classification performance, making it suitable for implementation in quality control systems in the food industry.

This study represents a preliminary step toward the broader use of DP-LIBS in microbiological diagnostics. Future research will focus on extending the methodology to other pathogens, such as *Salmonella spp*., commonly found in contaminated produce. Additionally, the feasibility of identifying bacterial genera based on elemental patterns will be explored, further advancing the use of LIBS as a diagnostic tool in agri-food and environmental monitoring.

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, 308339/2021-5, 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] Confederation of Agriculture and Livestock of Brazil (CNA). (2024). Performance in the fourth quarter reverses annual downward trend and agribusiness GDP grows 1.81%in2024.https://www.cnabrasil.org.br/publicacoes/desempenho-do-quarto-trimestre-reverte-a-tendencia-de-queda-anual-e-pib-do-agronegocio-avanca-1-81-em-2024

[2] Empresa Brasileira de Pesquisa Agropecuária (EMBRAPA), Brazil in 50 Foods. Brasília, Brazil: EMBRAPA, 2023.

[3] 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.

[4] S. Rehse, "A review of the use of laser-induced breakdown spectroscopy for bacterial classification, quantification, and identification", Spectrochimica Acto Part B: Atomic Spectroscopy, vol. 154, pp. 50-69, April 2019.

[5] V. K. Singh, J. Sharma, A. K. Pathak, C. T. Ghany, and M. A. Gondal, "Laser-induced breakdown spectroscopy (LIBS): a novel technology for identifying microbes causing infectious diseases," Biophys. Rev., vol. 10, no. 5, pp. 1221–1239, 2018, doi: 10.1007/s12551-018-0465-9.

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

[7] 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.

[8] D. M. B. P. Milori, L. Martin-Neto, H. T. Gollany, D. V. de Babos, L. C. L. Borduchi, and P. R. Villas-Boas, Soil analysis using laser-induced breakdown spectroscopy, 2nd ed. Elsevier Inc., 2023. doi: 10.1016/B978-0-12-822974-3.00211-1.

[9] D. Queiroz, et al. "Digital Agriculture", 1 e.d Viçosa, MG, pp 258-285, 2020.