Food Chemistry 297 (2019) 124960

Contents lists available at ScienceDirect

Food Chemistry

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

Brown rice authenticity evaluation by spark discharge-laser-induced breakdown spectroscopy

Michael Pérez-Rodríguez[a,*], Pamela Maia Dirchwolf[b], Tiago Varão Silva[c], Roxana Noelia Villafañe[a], José Anchieta Gomes Neto[c], Roberto Gerardo Pellerano[a], Edilene Cristina Ferreira[c]

[a] Institute of Basic and Applied Chemistry of the Northeast of Argentina (IQUIBA-NEA), National Scientific and Technical Research Council (CONICET), Faculty of Exact and Natural Science and Surveying, National University of the Northeast – UNNE, Av. Libertad 5470, 3400 Corrientes, Argentina
[b] Faculty of Agricultural Sciences, UNNE, Sgto. Cabral 2131, 3400 Corrientes, Argentina
[c] São Paulo State University – UNESP, Chemistry Institute of Araraquara, R. Prof. Francisco Degni 55, 14800-900 Araraquara, SP, Brazil

A R T I C L E I N F O

Keywords:
Food authenticity
PDO
Brown rice
SD-LIBS
Pattern recognition

A B S T R A C T

Rice is the most consumed food worldwide, therefore its designation of origin (PDO) is very useful. Laser-induced breakdown spectroscopy (LIBS) is an interesting analytical technique for PDO certification, since it provides fast multielemental analysis requiring minimal sample treatment. In this work LIBS spectral data from rice analysis were evaluated for PDO certification of Argentine brown rice. Samples from two PDOs were analyzed by LIBS coupled to spark discharge. The selection of spectral data was accomplished by extreme gradient boosting (XGBoost), an algorithm currently used in machine learning, but rarely applied in chemical issues. Emission lines of C, Ca, Fe, Mg and Na were selected, and the best performance of classification were obtained using k-nearest neighbor (k-NN) algorithm. The developed method provided 84% of accuracy, 100% of sensitivity and 78% of specificity in classification of test samples. Furthermore, it is simple, clean and can be easily applied for rice certification.

1. Introduction

Rice is an important staple food around the world. The cereal is rich in vitamins, fibers and essential minerals, making it a high nutritional value food. The nutritional properties, and therefore the quality of rice is related to its geographical origin. Thus, it is highlighted the need for controlled labeling to protect the authenticity of rice (Ariyama, Shinozaki, & Kawasaki, 2012; Promchan, Günther, Siripinyanond, & Shiowatana, 2016). Many countries have adopted for different food a controlled labeling based on protected designation of origin (PDO). The PDO information is a guarantee of the quality, tradition, and authenticity, and therefore food safety. In addition, PDO is also used as a brand strategy influencing the customer final decision (Dias & Mendes, 2018).

Worldwide several studies have been carried out aiming to determine the geographic origin of agricultural products (Becerra-herrera et al., 2018; Canizo, Escudero, Pérez, Pellerano, & Wuilloud, 2018; Magdas, Feher, Dehelean, & Cristea, 2018; Monahan, Schmidt, & Moloney, 2018; Moncayo, Rosales, Izquierdo-hornillos, Anzano, & Caceres, 2016; She et al., 2019). Particularly for rice, the combination of pattern recognition algorithms with multielemental techniques has been predominantly applied. To this purpose the most commonly used analytical techniques are inductively coupled plasma optical emission spectrometry (ICP OES) (Chung, Kim, Lee, & Kim, 2015) and inductively coupled plasma mass spectrometry (ICP-MS) (Ariyama et al., 2012; Cheajesadagul, Arnaudguilhem, Shiowatana, Siripinyanond, & Szpunar, 2013; Maione, Batista, Campiglia, Barbosa, & Barbosa, 2016). Although these techniques require a complex sample pre-treatment, they provide access to the sample elemental composition supplying a kind of fingerprint of the samples, which is useful for the food authentication.

Laser-induced breakdown spectroscopy (LIBS) is also a multi-elemental technique able to supply a fingerprint of sample composition. The advantage of LIBS over the aforementioned techniques is that LIBS is able to perform direct analysis, demanding minimum or no sample preparation and no generation of waste after analysis (Cremers & Radziemski, 2006; Miziolek, Palleschi, & Schechter, 2006). LIBS data assisted by chemometric methods have been successfully applied in different troubleshoot of classification, including classification based in

* Corresponding author.
E-mail address: michaelpr1984@gmail.com (M. Pérez-Rodríguez).

https://doi.org/10.1016/j.foodchem.2019.124960
Received 29 March 2019; Received in revised form 17 May 2019; Accepted 7 June 2019
Available online 08 June 2019
0308-8146/

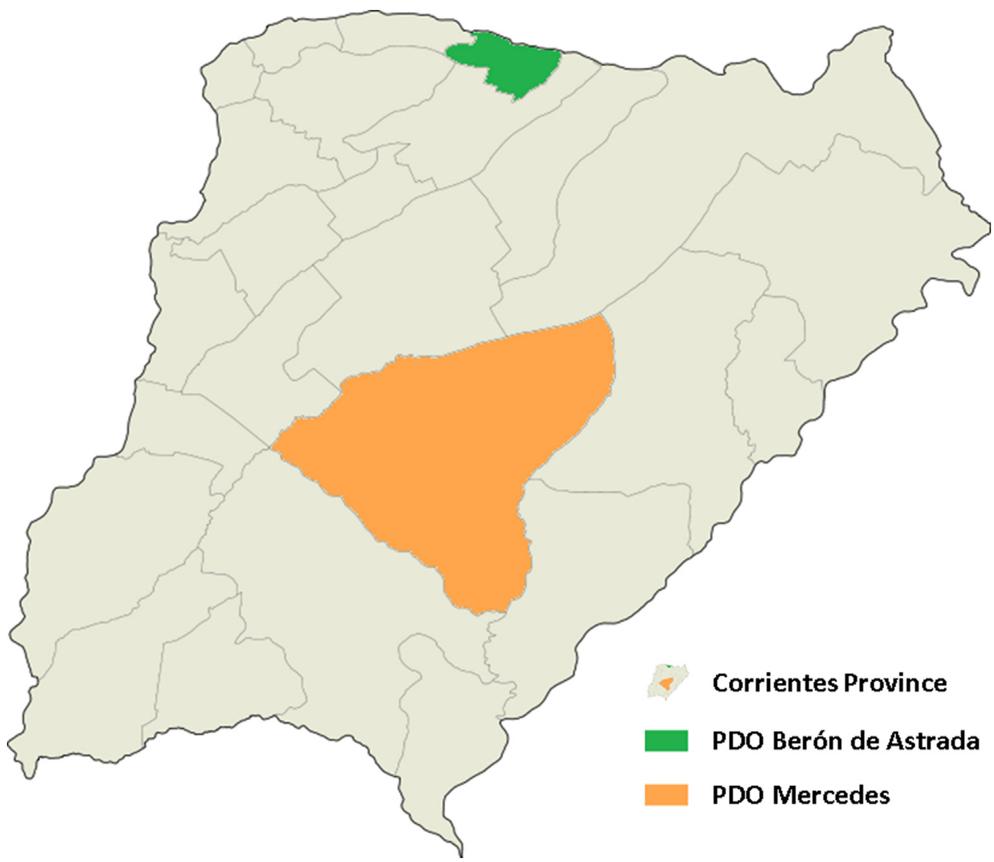

Fig. 1. Geographical regions selected for PDO rice sampling in Corrientes Province, Argentina.

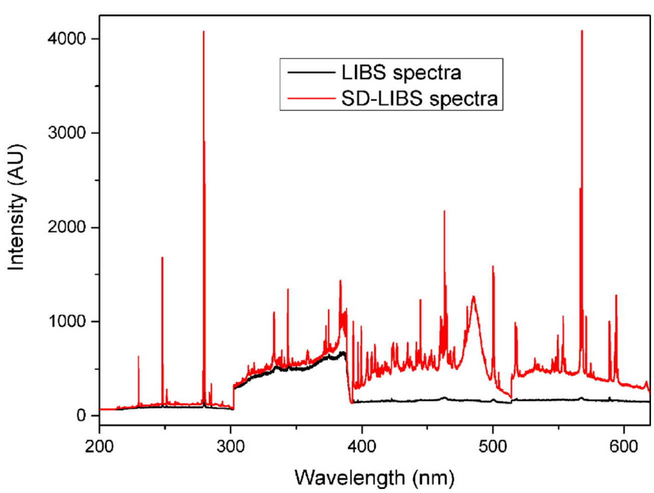

Fig. 2. LIBS (A) and SD-LIBS (B) spectra obtained from a brown rice sample.

PDO of different food samples (Bassbasi, Luca, Ioele, Oussama, & Ragno, 2014; Kumar et al., 2011; Moncayo, Manzoor, & Caceres, 2015; Moncayo et al., 2016; Zhang, Xia, Tang, Yang, & Li, 2016; Zhang, Shen, Liu, & He, 2018). Recently, Yang and collaborators have evaluated LIBS data processed by several chemometric algorithms to study the Chinese rice sample classification (Yang, Zhou et al., 2018; Yang, Zhu et al., 2018). In these studies, the number of inputted parameters for mathematical modeling was larger than the available samples. According to Hastie and collaborators (Hastie, Tibshirani, & Friedman, 2008), statistical models that contain more variables than can be justified by the data are prone to overfit, low capacity of generalization and weighting of spurious variables. Therefore, a suitable method for features selection is crucial, specially for spectra containing a large number of variables, like that provided by LIBS.

The extreme gradient boosting (XGBoost) is an algorithm also known as gradient boosted trees (GBT). Besides being an excellent classifier, it can be used to optimize and select features. Feature selection accomplished by XGboost is based on the action of boosting, where a weak learner can be modified to become better. In gradient boosting, weak learners are decision trees with a single split, called decision stumps. The great advantage of this method is that the importance of the variables for the classification models can be retrieved from the decision trees construction. In addition, the feature importance is estimated from trained predictive models and these tasks can be automatized. The variable importance means how useful it was to the model construction. The higher times the variable is used, the more important it is (Brownlee, 2017; Hastie et al., 2008). XGBoost is currently used by machine learning enthusiasts (Mustapha & Saeed, 2016; Torlay, Perrone-Bertolotti, Thomas, & Baciu, 2017; Zheng, Yuan, & Chen, 2017; Wang, Liu, & Deng, 2018), but very rarely used in chemical issues. In the best of our knowledge it has never been used to select variables from LIBS spectral data.

Considering the above, the aim of the present work was to develop a new method to determine PDO of Argentinean brown rice using LIBS as analytical technique and XGBoost for the spectral features selection to be used in the classification model fitting.

2. Material and methods

2.1. Rice samples

Brown rice of the species *Oryza sativa* L., coming from the northeastern of Argentine, was evaluated. A total of 66 samples were collected from rice fields located in two main producing regions of

Table 1
Performance evaluation measurements (Feature importance and F-Score metric) for variables selection by using XGBoost algorithm and identification of the selected spectral lines.

Variable	Importance	F-score	Maximum	Minimum	Spectral line	Element
1	0.215	61	X		335.3	Fe I
2	0.125	36		X	229.7	Fe I
3	0.112	34	X		247.8	C I/Fe II
4	0.075	22		X	588.9	Na I
5	0.070	21		X	409.8	Ca I
6	0.068	20		X	464.3	Fe I
7	0.060	18		X	374.8	Fe I
8	0.037	11		X	280.2	Mg II
9	0.035	10	X		463.1	Mg II
10	0.017	9	X		229.7	Fe I

Note: (I) indicates atomic spectral lines and (II) indicates ionic spectral lines.

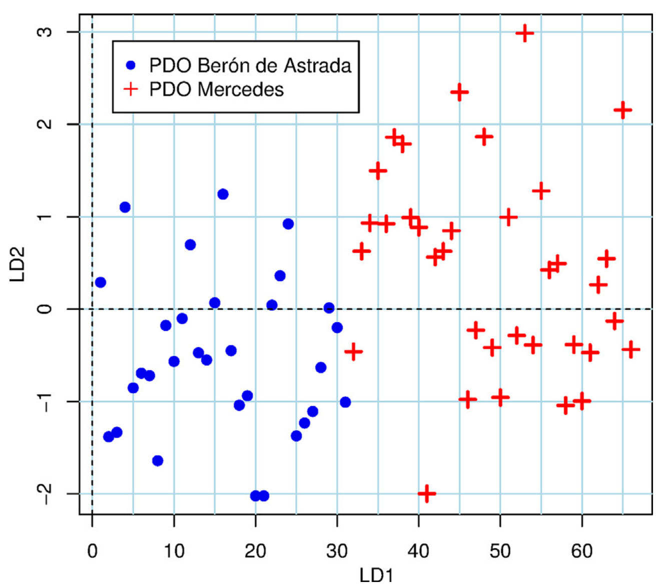

Fig. 3. Scatter plot of the first two discriminant functions obtained from LDA model to discriminate brown rice based on their PDO.

Table 2
Results of the classification metrics calculated for different evaluated models.

Algorithms	Parameters[a]	Accuracy (%)	Sensitivity (%)	Specificity (%)
LDA	–	53	70	33
k-NN	k = 7	84	100	78
SVM	C = 0.8; ε = 0.075	74	90	44
RF	nt = 500; mtry = 1	57	70	44

[a] k: number of k neighbors; nt: number of trees; mtry: number of variables tried at each split; C: penalty factor; ε: ε -insensitive loss function.

Corrientes province: thirty-five samples came from PDO Mercedes (29° 12′ S, 58° 05′ W) and thirty-one samples came from PDO Berón de Astrada (27° 33′ S, 57° 32′ W). All samples coming from the harvest of 2017. In Fig. 1 are shown the geographical locations of the producing regions with protected designation of origin.

The collected rice samples were peeled and were singly ground using a cryogenic mill from Spex 6750 (Metuchen, NJ, USA). The set grinding program consisted of 2.0 min for pre-freezing, 2.0 min for grinding and 3.0 min for freezing between the two milling steps. Around 250 mg of each homogenized sample were made into pellets by applying 10 tons using a mechanical press (Solab SL – 10/15, Piracicaba, Brazil). Two pellets were prepared for each of the 66 samples.

2.2. LIBS experimental setup and measurements

The LIBS system used for pellets analysis comprises a Q-switched laser Nd:YAG (Quantel, Big Sky Ultra 50, Bozeman, USA) emitting pulses at 1064 nm, lens for laser focalization (focal distance of 10 cm), an ablation chamber, a mobile sample holder with adjustment on the x and y axes, an optical system made up of lens and optical fiber bundle to collect plasma emission and drive it to the spectrometers. The plasma emission was collected by means of an optical fiber bundle and conducted to four spectrometers, Ocean Optics (HR2000 + , Dunedin, USA). The spectrometers optical resolution was 0.1 nm FWHM (full width at half maximum) and the spectral range measured for this instrument is 200 to 630 nm. All measurements were performed using laser pulse energy around 50 mJ, with 20 ns pulse duration. The interaction laser-sample provides a spot diameter around 300 µm, resulting in an irradiance in the focal point around 0.35 TW cm^{-2}. The delay time (relative to a Q-switch delay) was set to 1 µs and the integration time is instrumentally fixed at 1 ms.

Aiming to improve the sensitivity of measured spectrum, a system able to promote a spark discharge (SD) on the plasma, previously described by Vieira et al. (2018), was used coupled to the LIBS system above described. The discharge output was set to 4.5 kV and SD-LIBS spectra were acquired by applying 40 laser pulses spread on the surface of each pellet.

2.3. Data processing

Aiming to minimize undesirable fluctuation effects in LIBS spectra, due to laser pulses fluctuation and different laser-sample interaction, which occurs at each pulse (Miziolek et al., 2006), the spectra were individually transformed by the first derivative. Each prepared pellet was represented by an average of derived spectra (n = 40). Further an average of two pellets spectra was assessed for statistical inference of sampling behavior. For features selection step, the derived spectral data were autoscaled, to equalize the ranges of variability in prediction results.

Thereafter, XGBoost, available in the XGBoost Python package, was applied to select descriptors from the entire spectral range. The descriptors corresponded to the maximum and/or minimum variation ratio of derived spectra. For XGBoost algorithm application, the data matrix was randomly divided into two subsets, which consisted of a training set (70% of the samples) and a test set (the remaining 30%). The hyperparameters needing optimization included the number of trees (nt) and the learning rate (lr). These parameters were calculated during the training step by ten-fold cross-validation. Then, the method performance was measured from the feature importance and F-Score metric values. A detailed description regarding the calculation of the F-Score statistic can be found from Maione et al. (2016).

After features selection, the samples spectra were again divided in sets of calibration (training) and validation (testing). For calibration set

Table 3
Analytical methods evaluating the geographical authenticity of rice samples.

Reference	Analytical technique[a]	Chemometric technique[b]	Classification criteria[c]	Rice geographical origin
Ariyama et al. (2012)	HR-ICP-MS	LDA–SIMCA–k-NN	ACC: 97.0%	Japan, United States, China and Thailand
Promchan et al. (2016)	LA-ICP-MS	LDA	ACC: 91.1%	Thailand
Chung et al. (2015)	ICP OES	PLS–DA	–	Korea, China and Philippines
Maione et al. (2016)	ICP-MS	RF	ACC: 93.8%	Brazil
Cheajesadagul et al. (2013)	HR-ICP-MS	DA	OCI: 90.3–100%	Asia, Europe and Thailand
Chung et al. (2018)	IRMS and ICP-MS	OPLS–DA	Q^2Y: 0.834–0.963	Cambodia, China, Japan, Korea, Philippines and Thailand
Kukusamude and Kongsri (2018)	INAA and EA-IRMS	LDA	ACC: 95.5%	Thailand
Gonzálvez, Armenta, and Guardia (2011)	ICP OES	LDA	ACC: 91.3%	Brazil, India, Japan and Spain
Li et al. (2012)	ICP-MS and ICP OES	DFA and FIA	–	China
Qian et al. (2019)	HR-ICP-MS	DA	OCI: 89.6–96.6%	China
Teye, Amuah, McGrath, and Elliott (2019)	NIR	k-NN	ACC: 90.6%	Ghana, Thailand and Vietnam
Hwang, Kang, Lee, and Chung (2012)	Raman	LDA	DE: 1.61–9.97%	China and Korea
Zhu et al. (2018)	Raman	SIMCA	RR: 95.4%	China
Li et al. (2018)	XRF Raman	KMC–	ACC: 60–85.7%	China
Proposed method	SD-LIBS	XGBoost and k-NN	ACC: 84.0%	Argentina

[a] EA-IRMS: elemental analyzer isotope ratio mass spectrometry; HR-ICP-MS: high-resolution inductively coupled plasma mass spectrometry; ICP OES: inductively coupled plasma optical emission spectrometry; INAA: instrumental neutron activation analysis; LA-ICP-MS: laser ablation inductively coupled plasma mass spectrometry; NIR: near infrared spectrometry; Raman: Raman spectroscopy; SD-LIBS: spark discharge-laser-induced breakdown spectroscopy; XRF: X-ray fluorescence.
[b] DA: discriminant analysis; DFA: discriminant function analysis; FIA: Fibonacci index analysis; HC: hierarchical clustering; k-NN: k-nearest neighbors; KMC: k-means clustering; LDA: linear discriminant analysis; OPLS–DA: orthogonal projection to latent structure-discriminant analysis; PLS–DA: partial least-squares discriminant analysis; RF: random forest; SIMCA: soft independent modeling of class analogy; SVM: and support vector machine; XGBoost: extreme gradient boosting.
[c] ACC: Accuracy rate; DE: discrimination error; OCI: overall correct identification; Q^2Y: proportion of variance in the data predicted by a model; RR rejection rate (number of identified non-similar samples divided by the total number of unknown samples (non – class)).

25 spectra from PDO Mercedes plus 22 from PDO Berón de Astrada were randomly selected while for validation it was selected 10 from PDO Mercedes plus 9 from PDO Berón de Astrada. The division into subgroups was performed in a stratified manner, so that the proportion of each class in the original matrix was kept in the new subgroups. In addition, the samples included in each set were randomly changed for each reproduced model. This sampling scheme guaranteed the random sampling consistency.

Algorithms based on pattern recognition were evaluated to fit a suitable PDO classifier. The following supervised classification algorithms were evaluated: linear discriminant analysis (LDA), support vector machine (SVM), k-nearest neighbors (k-NN), and random forest (RF). The different classification models were validated by ten-fold cross-validation. An unknown group of samples belonging to known classes were used to assess the performance of the fitted models. All calculations for classification modeling were made using R-project software version 3.3.3 (R Core Team, 2017). The performance of the classification models were compared using the metrics of overall accuracy (all correct predictions divided by the total number of examined cases), sensitivity (correct positive predictions divided by the number of positive cases), and specificity (correct negative predictions divided by the number of negative cases) (Lantz, 2015).

3. Results and discussion

An initial assessment of the LIBS performance in rice analysis indicated low sensitivity of the available LIBS system. Aiming to improve the sensitivity of measurements a SD device previously proposed (Vieira et al., 2018), was coupled to the LIBS system. The coupling has reached its goal as can be seen in Fig. 2, therefore it was set as instrumental analytical condition.

After data acquisition spectral preprocessing was performed to minimize characteristic fluctuations of LIBS spectra that could affect the performance of the classification models. The first derivative transformation was applied to each individual spectrum because this preprocessing method has been demonstrated effectiveness and simplicity of application (Silva et al., 2017).

Considering many inputted variables can hinders the successful performance of the classification models (Hastie et al., 2008), the XGBoost algorithm was evaluated for spectral feature selection. The results obtained by applying the XGBoost algorithm in the set of training samples are shown in Table 1. The best performance for this feature selection was reached when the hyperparameters nt = 100 and lr = 0.1 were used. In this particular case, since the learning rate value was less than 1.0, the modeling have the effect of making the corrections for each tree added of the model (Brownlee, 2017). As can be observed in Table 1, ten variables obtained from maximum and minimum values from derived spectra were selected. According to the database of the National Institute of Standards and Technology (NIST, 2018), the selected variables corresponded to the emission lines of C, Fe, Ca, Na and Mg. These findings corroborate those observed by Maione et al. and Promchan et al. wherein Ca, Fe, Mg, and Na were also important elements for classification of Brazilian (Maione et al., 2016) and Thai (Promchan et al., 2016) rice, according to geographic origin. Moreover, the F-Score shows that Fe is the most important element to distinguish the evaluated classes with 59.5% of total importance, followed by C (plus Fe) with 14%, Na, 9.1% and Ca and Mg, both showing 8.7% of total importance.

Thereafter, the selected data were used to fit four different classifiers (LDA, RF, SVM and k-NN). In Fig. 3 are shown the patterns distribution of brown rice samples according to their PDO defined by the first two discriminant functions of LDA fitted model. A notably discrimination between the two studied groups can be observed in the whole set of samples with PDO Mercedes showing positive and negative scores on LD2 and PDO Berón de Astrada showing the majority of negative scores with a few positive scores on LD2. Nevertheless, the prediction accuracy in the testing set was 53%, indicating the PDO brown rice samples were not successfully classified.

Therefore, k-NN, SVM and RF non-linear algorithms were also evaluated. In Table 2 are shown the classification metrics calculated for each fitted model, including LDA linear algorithm.

The better classification metrics were obtained for k-NN, SVM and RF. These findings suggest non-linear models fit better the behavior of C, Fe, Ca, Na and Mg in the two PDO evaluated. The highest accuracy in prediction of test samples were given by SVM (74%) with radial basis kernel function using hyperparameters C = 0.8 and ε = 0.075 and k-NN (84%) with an optimized neighbońs number of seven. The best model for differentiating brown rice samples according to their PDO was k-NN. This model also showed the highest sensitivity (100%) and specificity (78%), compared to the other methods studied.

A comparison of the performance of the proposed method, with methods previously described for the rice geographical authenticity is given in Table 3.

Each method uses as variables specific properties provided by the different employed analytical techniques. All classification methods are multivariate, suggesting more than one feature is frequently required to perform geographical classification of rice. For the several evaluated methods, analytical performance in terms of accuracy ranged from 60 to 97%, indicating the suitability of the proposed method. Compared to the methods using multielement techniques (e.g. ICP OES and ICP MS) the SD-LIBS is faster and more environmentally friendly, since it dispenses complexes sample pretreatments and the use of any chemical reagent.

4. Conclusions

Spectral fingerprints of brown rice obtained by SD-LIBS were investigated to discriminate samples according to their PDO. XGBoost algorithm were used for the first time to select relevant features from LIBS spectra. The analysis of F-Scores of the variables suggested that Fe were the most important element to discriminate the studied PDOs. The other elements presented the following order of importance: C > Na > Ca = Mg. The evaluated algorithms for classification suggested the amounts of selected elements changes in a non-linear fashion for the evaluated PDOs. Moreover, a non-linear classifier based on k-NN showed the best classification performance, achieving 84% of accuracy in prediction of test samples. The developed method is simple, fast and reliable approach, demands minimal sample preparation and does not produce waste after analysis. Its application can help to identify rice mislabeling protecting market rice brands and preventing consumers and producers from financial damage.

Declaration of Competing Interest

The authors have no conflicts of interest to declare.

Acknowledgements

The authors would like to thank National Scientific and Technical Research Council (LH: 172645 CONICET) for granting a postdoctoral scholarship to M. Pérez-Rodríguez and the Conselho Nacional de Desenvolvimento Tecnológico (CNPq) for the research grant to E.C.F. (308200/2018-7).

References

Ariyama, K., Shinozaki, M., & Kawasaki, A. (2012). Determination of the geographic origin of rice by chemometrics with strontium and lead isotope ratios and multielement concentrations. *Journal of Agricultural and Food Chemistry, 60*, 1628–1634.

Bassbasi, M., De Luca, M., Ioele, G., Oussama, A., & Ragno, G. (2014). Prediction of the geographical origin of butters by partial least square discriminant analysis (PLS-DA) applied to infrared spectroscopy (FTIR) data. *Journal of Food Composition and Analysis, 33*, 210–215.

Becerra-herrera, M., Vélez-martín, A., Ramos-merchante, A., Richter, P., Beltrán, R., & Sayago, A. (2018). Characterization and evaluation of phenolic pro fi les and color as potential discriminating features among Spanish extra virgin olive oils with protected designation of origin. *Food Chemistry, 241*, 328–337.

Brownlee, J. (2017). XGBoost with Paython.

Canizo, B. V., Escudero, L. B., Pérez, M. B., Pellerano, R. G., & Wuilloud, R. G. (2018). Intra-regional classi fi cation of grape seeds produced in Mendoza province (Argentina) by multi-elemental analysis and chemometrics tools. *Food Chemistry, 242*, 272–278.

Cheajesadagul, P., Arnaudguilhem, C., Shiowatana, J., Siripinyanond, A., & Szpunar, J. (2013). Discrimination of geographical origin of rice based on multi-element fingerprinting by high resolution inductively coupled plasma mass spectrometry. *Food Chemistry, 141*, 3504–3509.

Chung, I., Kim, J., Lee, J., & Kim, S. (2015). Discrimination of geographical origin of rice (Oryza sativa L.) by multielement analysis using inductively coupled plasma atomic emission spectroscopy and multivariate analysis. *Journal of Cereal Science, 65*, 252–259.

Chung, I., Kim, J., Lee, K., Park, S., Lee, J., Son, N., ... Kim, S. (2018). Geographic authentication of Asian rice (Oryza sativa L.) using multielemental and stable isotopic data combined with multivariate analysis. *Food Chemistry, 240*, 840–849.

Cremers, D., & Radziemski, L. J. (2006). *Handbook of laser-induced breakdown spectroscopy* (2nd Ed.). Chichester: John Wiley & Sons.

Dias, C., & Mendes, L. (2018). Protected designation of origin (PDO), protected geographical indication (PGI) and traditional speciality guaranteed (TSG): a bibiliometric analysis. *Food Research International, 103*, 492–508.

Gonzálvez, A., Armenta, S., & Guardia, M. (2011). Geographical traceability of "Arròs de Valencia" rice grain based on mineral element composition. *Food Chemistry, 126*, 1254–1260.

Hastie, T., Tibshirani, R., & Friedman, J. (2008). *The elements of statistical learning: Data mining, inference, and prediction* (2nd Ed.). Springer: Stanford University.

Hwang, J., Kang, S., Lee, K., & Chung, H. (2012). Enhanced Raman spectroscopic discrimination of the geographical origins of rice samples via transmission spectral collection through packed grains. *Talanta, 101*, 488–494.

Kukusamude, C., & Kongsri, S. (2018). Elemental and isotopic profiling of Thai jasmine rice (Khao Dawk Mali 105) for discrimination of geographical origins in Thung Kula Rong Hai area, Thailand. *Food Control, 91*, 357–364.

Kumar, A., Sreedhar, S., Barman, I., Chari, N., Rao, S. V., Kiran, P. P., ... Kumar, G. M. (2011). Laser-induced breakdown spectroscopy-based investigation and classification of pharmaceutical tablets using multivariate chemometric analysis. *Talanta, 87*, 53–59.

Lantz, B. (2015). *Machine learning with R* (Second Ed.). Birmingham: Packt Publishing Ltd.

Li, G., Nunes, L., Wang, Y., Williams, P. N., Zheng, M., Zhang, Q., & Zhu, Y. (2012). Profiling the ionome of rice and its use indiscriminating geographical origins at the regional scale, China. *Journal of Environmental Sciences, 25*, 144–154.

Li, F., Wang, J., Xu, L., Wang, S., Zhou, M., Yin, J., & Lu, A. (2018). Rapid screening of cadmium in rice and identification of geographical origins by spectral method. *International Journal of Environmental Research and Public Health, 15*, 312–324.

Magdas, D. A., Feher, I., Dehelean, A., & Cristea, G. (2018). Isotopic and elemental markers for geographical origin and organically grown carrots discrimination. *Food Chemistry, 26*, 231–239.

Maione, C., Batista, B. L., Campiglia, A. D., Barbosa, F., & Barbosa, R. M. (2016). Classification of geographic origin of rice by data mining and inductively coupled plasma mass spectrometry. *Computers and Electronics in Agriculture, 121*, 101–107.

Miziolek, A. W., Palleschi, V., & Schechter, I. (2006). *Laser-induced breakdown spectroscopy (LIBS) – Fundamentals and Applications* (1st Ed.). New York – USA: Cambridge University Press.

Monahan, F. J., Schmidt, O., & Moloney, A. P. (2018). Meat provenance: Authentication of geographical origin and dietary background of meat. *Meat Science, 144*, 2–14.

Moncayo, S., Manzoor, S., & Caceres, J. O. (2015). Chemometrics and Intelligent Laboratory Systems Evaluation of supervised chemometric methods for sample classi fi cation by Laser Induced Breakdown Spectroscopy. *Chemometrics and Intelligent Laboratory Systems, 146*, 354–364.

Moncayo, S., Rosales, J. D., Izquierdo-hornillos, R., Anzano, J., & Caceres, J. O. (2016). Talanta Classi fi cation of red wine based on its protected designation of origin (PDO) using Laser-induced Breakdown Spectroscopy (LIBS). *Talanta, 158*, 185–191.

Mustapha, I. B., & Saeed, F. (2016). Bioactive molecule prediction using extreme gradient boosting. *Molecules, 21*, 983–994.

NIST Atomic Spectra Database (2018). https://physics.nist.gov/PhysRefData/ASD/lines_form.html.

Promchan, J., Günther, D., Siripinyanond, A., & Shiowatana, J. (2016). Elemental imaging and classifying rice grains by using laser ablation inductively coupled plasma mass spectrometry and linear discriminant analysis. *Journal of Cereal Science, 71*, 198–203.

Qian, L., Zuo, F., Liu, H., Zhang, C., Chi, X., & Zhang, D. (2019). Determination of geographical origin of Wuchang rice with the geographical indicator by multielement analysis. *Journal of Food Quality*.

R Core Team (2017). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing Accessed from https://www.r-project.org/.

She, S., Chen, L., Song, H., Lin, G., Li, Y., & Zhou, J. (2019). Discrimination of geographical origins of Chinese acacia honey using complex $^{13}C/^{12}C$ oligosaccharides and polyphenols. *Food Chemistry, 272*, 580–585.

Silva, T. V., Hubinger, S. Z., Neto, J. A. G., Milori, D. M. B. P., Ferreira, E. J., & Ferreira, E. C. (2017). Potential of Laser Induced Breakdown Spectroscopy for analyzing the quality of unroasted and ground coffee. *Spectrochimica Acta Part B, 135*, 29–33.

Teye, E., Amuah, C. L. Y., McGrath, T., & Elliott, C. (2019). Innovative and rapid analysis for rice authenticity using hand-held NIR spectrometry and chemometrics. *Spectrochimica Acta Part A, 217*, 147–154.

Torlay, L., Perrone-Bertolotti, M., Thomas, E., & Baciu, M. (2017). Machine learning–XGBoost analysis of language networks to classify patients with epilepsy. *Brain Informatics, 4*, 159–169.

Vieira, A. L., Silva, T. V., Sousa, F. S. I., Senesi, G. S., Santos, D., Ferreira, E. C., & Neto, J. A. G. (2018). Determinations of phosphorus in fertilizers by spark discharge-assisted laser-induced breakdown spectroscopy. *Microchemical Journal, 139*, 322–326.

Wang, H., Liu, C., & Deng, L. (2018). Enhanced prediction of hot spots at protein-protein interfaces using extreme gradient boosting. *Scientific Reports, 8*, 14285–14298.

Yang, P., Zhou, R., Zhang, W., Tang, S., Hao, Z., Li, X., ... Zeng, X. (2018). Laser-induced breakdown spectroscopy assisted chemometric methods for rice geographic origin classification. *Applied Optics, 57*, 8297–8302.

Yang, P., Zhu, Y., Yang, X., Li, J., Tang, S., Hao, Z., ... Lu, Y. (2018). Evaluation of sample preparation methods for rice geographic origin classification using laser-induced breakdown spectroscopy. *Journal of Cereal Science, 80*, 111–1188.

Zhang, C., Shen, T., Liu, F., & He, Y. (2018). Identification of coffee varieties using laser-

induced breakdown spectroscopy and chemometrics. *Sensors, 18*, 95–107.
Zhang, T., Xia, D., Tang, H., Yang, X., & Li, H. (2016). Chemometrics and intelligent laboratory systems classification of steel samples by laser-induced breakdown spectroscopy and random forest. *Chemometrics and Intelligent Laboratory Systems, 157*, 196–201.
Zheng, H., Yuan, J., & Chen, L. (2017). Short-term load forecasting using EMD-LSTM neural networks with an XGBoost algorithm for feature importance evaluation. *Energies, 10*, 1168–1188.
Zhu, L., Sun, J., Wu, G., Wang, Y., Zhang, H., Wang, L., ... Qi, X. (2018). Identification of rice varieties and determination of their geographical origin in China using Raman spectroscopy. *Journal of Cereal Science, 82*, 175–182.