BCSJ

Document type: Article

# Machine Learning Used to Create a Multidimensional Calibration Space for Sensing and Biosensing Data

**Mário Popolin Neto,[1,2] Andrey Coatrini Soares,[3] Osvaldo N. Oliveira, Jr.,*[4] and Fernando V. Paulovich[2,5]**

[1]Federal Institute of São Paulo (IFSP), 14804-296 Araraquara, Brazil

[2]Institute of Mathematics and Computer Sciences (ICMC), University of São Paulo (USP), 13566-590 São Carlos, Brazil

[3]Nanotechnology National Laboratory for Agriculture (LNNA), Embrapa Instrumentação, 13560-970 São Carlos, SP, Brazil

[4]São Carlos Institute of Physics (IFSC), University of São Paulo (USP), 13566-590 São Carlos, Brazil

[5]Faculty of Computer Science (FCS), Dalhousie University (DAL), B3H 4R2 Nova Scotia, Canada

E-mail: chu@ifsc.usp.br

Received: November 21, 2020; Accepted: March 15, 2021; Web Released: March 27, 2021

**Osvaldo N. Oliveira, Jr.**
Osvaldo N. Oliveira, Jr. is a professor at the São Carlos Institute of Physics, University of São Paulo, Brazil. He received his BSc. and MSc. in physics from the University of São Paulo in 1982 and 1984, respectively. In 1990 he got his PhD from the University of Wales, Bangor, UK. His main research areas are in nanostructured organic films and natural language processing. He is an associate editor of ACS Applied Materials & Interfaces.

## Abstract

Calibration curves are essential constructs in analytical chemistry to determine parameters of sensing performance. In the classification of sensing data of complex samples without a clear dependence on a given analyte, however, establishing a calibration curve is not possible. In this paper we introduce the concept of a multidimensional calibration space, which could serve as reference to classify any unknown sample as in determining an analyte concentration from a calibration curve. This calibration space is defined from a set of rules generated using a machine learning method based on trees applied to the dataset. The number of attributes employed in the rules defines the dimension of the calibration space and is established to warrant full coverage of the dataset. We demonstrate the calibration space concept with impedance spectroscopy data from sensors, biosensors and an e-tongue, but the concept can be extended to any type of sensing data and classification task. Using the calibration space should allow for the correct classification of unknown samples, provided that the data used to generate rules via machine learning can cover the whole range of sensing measurements. Furthermore, an inspection in the rules can assist in the design of sensing systems for optimized performance.

**Keywords:** Machine learning | Interpretable artificial intelligence | Sensors classification model

## 1. Introduction

Analytical curves are ubiquitous in analytical chemistry, with well-established procedures to determine analytical parameters recommended by IUPAC.[1,2] The same can be said of calibration curves and processes for instruments in general, some of which are utilized not only to determine a given physical quantity but also to adjust the instruments for correct functioning.[3,4] In sensors and biosensors, in particular, using analytical curves one can "transform" the task of classifying the set of samples under analysis into a predictive exercise where the concentration of a given analyte in an unknown sample can be determined precisely. In other types of sensors, as in the case of electronic tongues[5–11] and electronic noses,[9,12–14] this determination may not be possible and no analytical curves can be established. This happens because these sensors may be used to classify different types of liquids such as wine[15,16] or coffee[17] without determining the concentration of any specific analyte. Calibration procedures can nevertheless still be employed for e-tongues and e-noses,[18,19] but these are related to the multivariate calibration that allows for studying quantitative and qualitative aspects of simple and complex solutions.[20] For the



so-called complex samples that contain multiple analytes, use has been made of multivariate analysis[21] and other statistical and computational methods,[22] including information visualization and machine learning techniques. These methods are advantageous for the evaluation of large volumes of data, providing predictions about food and water contaminants, diagnosis and prognosis. For example, Daikuzono et al. applied an information visualization technique referred to as Interactive Document Mapping (IDMAP)[23] in electrical impedance spectra to detect gliadin in food samples contaminated by gluten. IDMAP was also used to treat data from a microfluidic electronic tongue to detect petrochemical compounds, heavy metals and basic flavors[5] and from a biosensor to detect pathogenic bacteria in food samples.[24] Examples of machine learning applied to diagnosis include data processing of very distinct natures. In image analysis, for instance, the superiority of computer-assisted diagnosis (compared to human experts) is well established, which includes the use of deep learning for analyzing magnetic resonance images[25] and biosensors surfaces.[26,27] For cardiac diseases, datasets with varied medical parameters and results from clinical exams, including from immunosensors, have been used for diagnosis with various machine learning strategies.[28] Optical biosensing data from paper-based point-of-care (POCs) have been used in conjunction with machine learning to diagnose cardio-vascular diseases.[29] Cancer biomarkers were detected with higher specificity in serum samples of patients by applying classification algorithms to SERS (surface-enhanced Raman scattering) data in a biosensor made with a microfluidic chip.[30] SERS spectra from other biosensors were treated with machine learning algorithms for the diagnosis of liver cancer and liver cirrhosis.[31] Even for glucose detection using amperometry has a genetic algorithm been useful to improve diagnosis.[32] Calibration curves are not very useful in these cases, though they are sometimes employed for comparison purposes and to obtain sensing performance.

The usefulness of the various statistical and computational methods abovementioned is irrefutable, but their limitation remains in not being interpretable in the classification tasks. In this paper, we introduce the concept of a multidimensional calibration space, which we believe addresses this limitation. The multidimensional space is defined from rules generated upon classifying the samples in a dataset with Decision Trees[33,34] machine learning algorithms. The definition and exemplification of the multidimensional calibration space is given in Section 2 with sensing data from a sensor to detect phytic acid[35] whereas Section 3 brings examples of the application of the concept for biosensors in the diagnosis of pancreatic cancer[36] and for e-tongue data.[35,37] The generality of the approach is discussed in the Discussion in Section 4 and in the Final Remarks.

## 2. Methodology

The concept of a multidimensional calibration space is introduced by applying Decision Tree (DT) models[33,34] to the impedance spectroscopy data of a sensor made with layer-by-layer films of polyelectrolytes to detect different concentrations of phytic acid.[35] Figure 1 shows the capacitance spectra obtained in ref 35.[35] This problem was chosen because we knew from the latter reference that the sensor was not selective for phytic acid and a calibration curve could not be established. On the other hand, there was some distinction between the spectra for different concentrations and therefore classification should be a simple task with a small number of rules. As it will be shown, this allows the calibration space to be represented with only three dimensions.

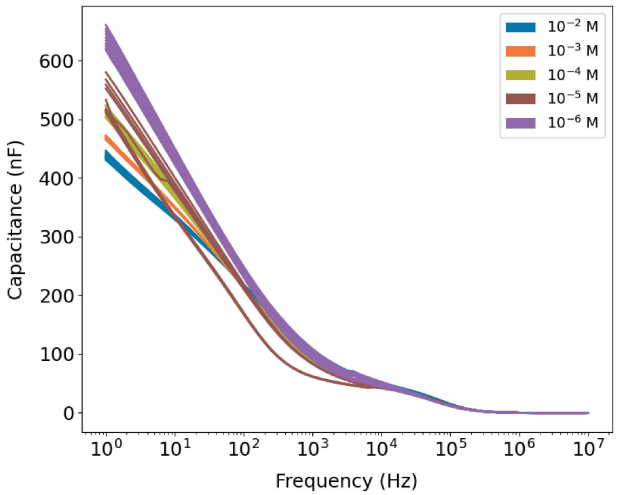



**Figure 1.** Capacitance spectra obtained with the sensor made with poly(allylamine hydrochloride) (PAH) and poly(vinyl sulfonic acid) (PVS) layer-by-layer films for different phytic acid concentrations. The values in the insert correspond to concentrations in M. Data extracted from ref 35.[35]

The choice of DT models is justified by the possibility of establishing predictive rules for calibration, as DTs are today the most prevalent interpretable classification approach.[38,39] In classification with machine learning techniques, a computational model is built to predict the class of a given data instance. A set of data instances $X = \{x_1, x_2, \ldots x_N\}$ and their associated classes $Y = \{y_1, y_2, \ldots y_N\}$ where $y_i \in C = \{c_1, c_2, \ldots c_M\}$, usually called the training set, are employed to infer a function f(.) that maps each training instance $x_i$ into its class $y_i$, that is $f(x_i) \rightarrow y_i$. If $X$ is comprehensive and represents the phenomena under analysis in their entirety, $f(.)$ can be used to predict the class of any new instance $x_j$ that was not originally in $X$. The techniques to infer $f(.)$ may be split into two distinct groups, *viz.* the blackboxes[38,39] such as Artificial Neural Networks[15] and Support Vector Machines,[40,41] and the inherently interpretable models[38,39] such as DTs[33,34] and Rule Sets.[38,39] Although typically less accurate, the techniques of the latter group allow for the reasoning of a given classification. In other words, interpretable models support the understanding of how the attributes of a given instance, that is, the values describing it, contributed to its classification. With this scheme one may generate interpretable models not only as predictive tools but also as descriptive strategies where intrinsic relationships among data attributes and classes can be revealed.[34]

Using DT models as the classification approach for impedance spectroscopy data, we may select a subset of available attributes (frequencies) without requiring dimensionality reduction as in data pre-processing or manual frequency selection. By using the visualization method ExMatrix,[42] the Multidimensional Calibration Space created by DT models can be

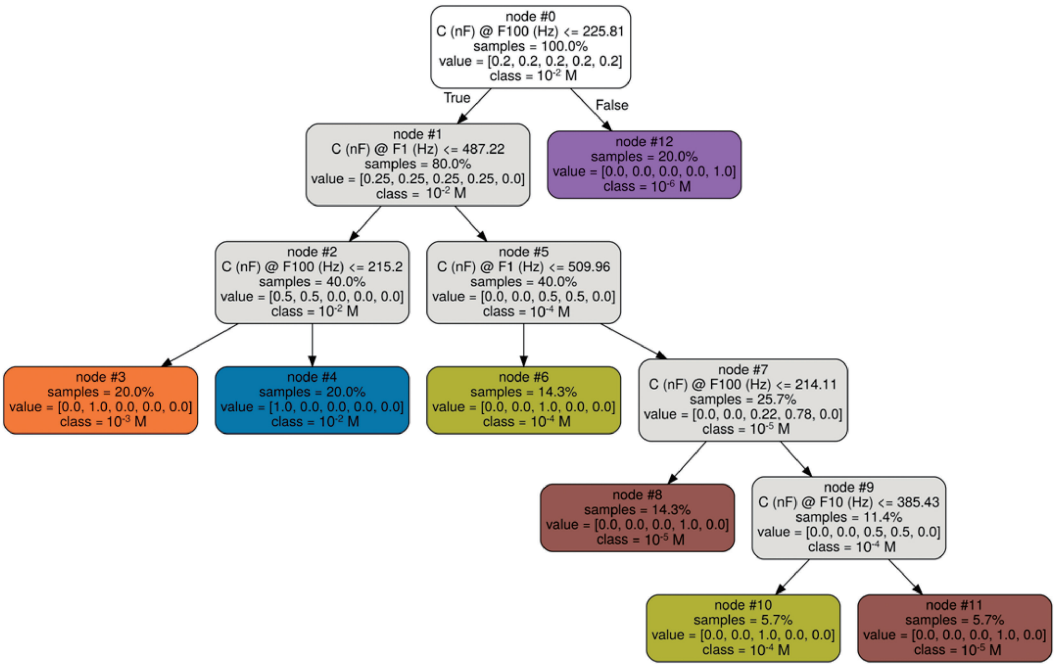
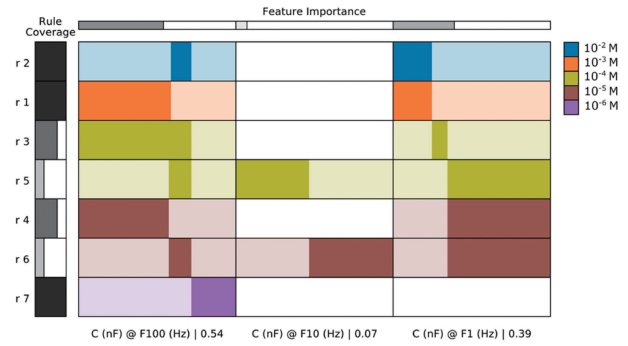


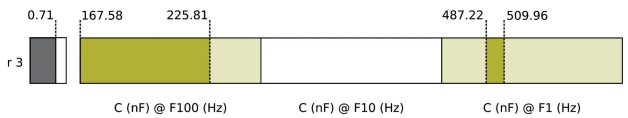

**Figure 2.** Example of DT for a dataset containing the capacitance at 3 frequencies (F100, F10, and F1), measured with a PAH/PVS sensor[35] on samples of phytic acid concentrations ($10^{-2}$, $10^{-3}$, $10^{-4}$, $10^{-5}$, and $10^{-6}$ M).

explored, displaying space ranges and class associations. The classification of individual instances/samples can be analyzed for reasoning about the class assignment. As the name indicates, DT techniques create a tree-like structure where internal nodes contain test functions (or predicates) based on the data attribute values to recursively split the training data into non-overlapping sub-groups so that each final sub-group contains only instances of the same class. Figure 2 displays an example of a DT inferred from a training dataset containing 35 instances (N = 35) of 5 phytic acid concentrations (M = 5, 7 instances per concentration at $10^{-2}$, $10^{-3}$, $10^{-4}$, $10^{-5}$, and $10^{-6}$ M). The experimental details for obtaining this impedance spectroscopy dataset are given in the Supporting Information. Each instance is described by its capacitance at 3 frequencies (100, 10 and 1 Hz, referred to as F100, F10, and F1, respectively) obtained with a sensor made of a layer-by-layer (LbL) film with poly-(allylamine hydrochloride) (PAH) and poly(vinyl sulfonic acid) (PVS) deposited onto an interdigitated gold electrode. The DT in Figure 2 is validated through a testing set containing 15 instances (3 instances per concentration) not present in the training set, reaching 100% accuracy. In the nomenclature adopted here, we use FX to refer to the capacitance value in nF at the frequency X in Hz. Using a DT, IF/THEN logic rules can be extracted to represent the combination of attribute ranges that best describe a specific class of instances. Each tree path from the root to a leaf defines one distinct rule, and the entire set of rules can be used as a descriptive model of the (training) data. For instance, in Figure 2 the path from node #0, node #1, node #5, and node #6 defines the rule **IF C(nF)@F100 ≤ 225.81 AND C(nF)@F1 > 487.22 AND C(nF)@F1 ≤ 509.96 THEN $10^{-4}$**. If a node predicate is true, we navigate to the left branch; right otherwise. Note that different rules can result in the same class, everything depending on the complexity of the input data.

For $X$ corresponding to data from sensors or biosensors (or other types of data as explained later on), a multidimensional calibration space visualization can be created using the ExMatrix[42] technique to display an overview of the logic rules extracted from a DT model inferred from $X$. DT models can be complex to interpret as the number of nodes increases, producing deep trees. With the ExMatrix visual representations, a DT model is arranged into a matrix-like visualization where rules are rows, attributes are columns, and the predicates are the matrix cells. Figure 3-a shows a visual representation of the DT in Figure 2. The resulting matrix has 7 rows, one per rule, and 3 columns for the different attributes. The matrix cell is colored to reflect the inferred class and filled so that darker colors represent the range of each attribute used by a rule. In a cell, the left-most side represents the minimum value for an attribute considering the entire data set, whereas the right-most side is the maximum. For instance, in the rule depicted on the third matrix row (Figure 3-a) highlighted on Figure 3-b represents the rule **IF C(nF)@F100 ≤ 225.81 AND C(nF)@F1 > 487.22 AND C(nF)@F1 ≤ 509.96 THEN $10^{-4}$**. The F100 cell is filled representing the range [167.58, 225.81], where 167.58 is the minimum value admitted by the attribute F100. F1 is filled to represent the interval [487.22, 509.96]. In addition to obtaining the relationships among attribute ranges and classes conveyed by each rule, the coverage of the rules is also calculated. This coverage is the percentage of instances in the training set belonging to the same inferred class for which the rule is true. The rule coverage is mapped to one additional column on the left side of the matrix. The rule 3 on the third row in Figure 3 extracted from the path finishing at node #6 in Figure 2 has coverage of 0.71, whereas the rule on the fourth row from the path finishing at node #10 has coverage of 0.29. This indicates that the first is more generic, being valid for a higher number of instances. The last rule is more specific, valid for a small




(a) ExMatrix visualization from the DT of Figure 2.




(b) Rule 3 highlighted, showing the ranges values.

**Figure 3.** Multidimensional calibration space visualization using ExMatrix. (a) The seven rules defined in the DT of Figure 2 are represented as rows, the frequencies are in the columns and the cells indicate the ranges in each frequency "used" to predict the different concentrations. The leftmost column represents the rule coverage, with rules r2, r1, and r7 exhibiting maximum values, while r3 and r4 give intermediate values and r5 and r6 have small values. This indicates that for the concentrations $10^{-2}$, $10^{-3}$, and $10^{-6}$ M, the data is easier to separate (classify) since only one rule can represent those concentration classes on the three frequencies used. However, for the concentrations $10^{-4}$ and $10^{-5}$ M, multiple rules are necessary, i.e. the data is more complex for these concentrations. By comparing the ranges in different classes, one may infer the parts of the space that best define a concentration. (b) Rule 3 highlighted, presenting the range values, visually representing the text format: IF C(nF)@F100 $\leq$ 225.81 AND C(nF)@F1 $>$ 487.22 AND C(nF)@F1 $\leq$ 509.96 THEN $10^{-4}$. This rule provides 0.71 as coverage (between 0 and 1), where the higher the coverage the more generic is the rule.

number of instances. The attribute importance is added as a row on the top of the table and reflects attribute capability to differentiate classes.[43] The attribute name is placed at the bottom, along with the attribute importance value.

Although simple, ExMatrix visual representation allows for an informative analysis. For example, an instance/sample with a higher value of real capacitance at F100 has a high probability of being $10^{-6}$ M (lilac color), given the high coverage of the rule 7 at the last row (Figure 3-a). By analyzing rules at the first two rows, one notes that the $10^{-2}$ and $10^{-3}$ M concentrations (blue and orange colors) are similar at F1 and distinguished at F100, whereas $10^{-2}$ M (blue color) holds a small range of values, but higher than the values for $10^{-3}$ M (orange color). By inspecting rules predicting the $10^{-4}$ and $10^{-5}$ M concentrations (olive and brown colors), we observe overlaps of attribute ranges at F100 and F1, which are different in terms of F10. These overlaps and rules coverage reveal a certain space split complexity to separate the concentrations (instances/

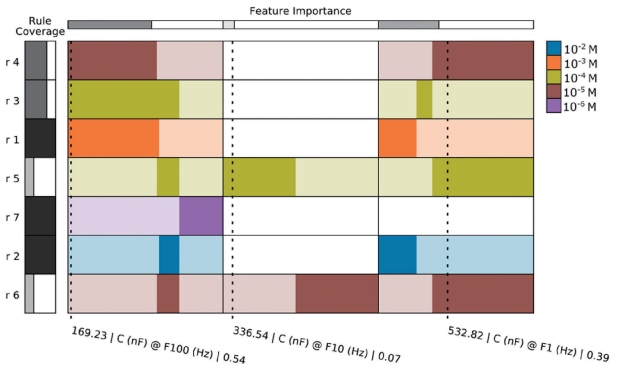



**Figure 4.** Multidimensional calibration space visualization using ExMatrix for the classification of a specific instance/sample. Dashed lines indicate the instance values in each attribute (frequency). In this matrix, the rules are ordered by proximity to the instance under analysis, where the rule in the first row (brown/rule 4) is used to classify the instance as class $10^{-5}$ M. To change the instance's classification from $10^{-5}$ to $10^{-4}$ M requires the smallest modification (olive/rule 3 on the second row), while the modification required to change the classification to $10^{-2}$ M is the largest (blue/rule 2 on the sixth row).

samples). It should be remarked that in simple cases feature selection can be made manually, for example noting the frequencies at which distinction among samples is higher. With DTs, on the other hand, this selection is performed in a systematic, non-arbitrary manner selecting features that present the best separability between classes. This is reflected on the feature (or attribute) importance values displayed at the top of the ExMatrix representation.

The calibration space in Figure 3 can also be represented by checking the ranges at which the different rules apply. As an example, Figure 4 shows dashed lines for an instance/sample with values 169.23, 336.54, and 532.82 for the frequencies F100, F10, and F1. This instance is classified as $10^{-5}$ M since it falls into the darker colored area of the rule in the first row (brown color/rule 4). In Figure 4 the rules are ordered according to the proximity to the instance under analysis, where the used rule to classify the instance is in the first row (brown/rule 4). Proximity here means the smallest modifications (gaps between dotted lines and ranges) needed to apply to the instance in order to change its class. For instance, in the second row (rule 3) one notes that a small decrease in capacitance at F1 would make the instance be classified as belonging to the $10^{-4}$ M class (olive). On the other hand, larger modifications are required to make it switch, for example, to the class $10^{-2}$ M (blue/rule 2 on the sixth row), where capacitance values at frequencies F100 and F1 need positive and negative increments, respectively.

In the example chosen, the number of rules to cover 100% of the dataset is seven, close to the minimum possible of 5 rules for the five classes. Since only three attributes (F100, F10, and F1) need to be used in these rules, one may establish a 3D calibration space as shown in Figure 5 for the rules in Figure 3 (or Figure 4). The colored boxes represent the space where each concentration is different from the others. From this plot one infers the difficulty in distinguishing the $10^{-4}$ and $10^{-5}$ concentrations (olive and brown), while the others occupy more

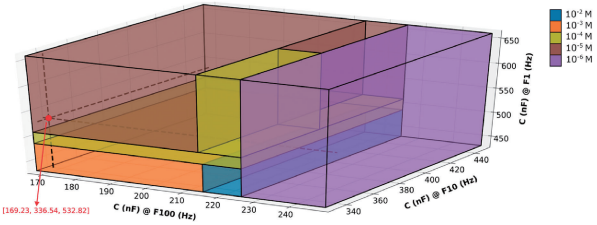



**Figure 5.** Multidimensional calibration space shown as a 3D plot. For most concentrations, the parts of the space "used" by the different concentrations are simple. Only for $10^{-4}$ and $10^{-5}$ concentrations (olive and brown) the space splitting is more complex.

defined parts of the 3D calibration space. Also interesting is to consider the instance classified as $10^{-5}$ M and discussed in connection with Figure 4, represented as a red circle in Figure 5. This circle needs to "travel" very little in the 3D space to move to class $10^{-4}$ M, whereas to move to class $10^{-2}$ M the distance is much larger (and in two axes).

The coverage of the logic rules obtained from DT models is related to the dataset complexity and DT inference approach. In this paper, we use the Classification And Regression Trees (CART)[33] technique to derive DTs. It should be noted that different trees are obtained by varying the inference parameters. In a model selection experiment,[44–46] we intentionally varied these parameters to generate multiple (and many) trees and select the one that provides rules with the highest accuracy on KFold Cross Validation using the training set. Then, the parameters selected are used to create a DT model considering the whole training set, and the resulting DT is tested using the test set (unknown samples). In doing so, we selected the best tree in an unbiased manner, its performance can be evaluated, and through the ExMatrix method the Multidimensional Calibration Space can be analyzed.

One limitation that may be inferred from inspecting Figure 5 is in the discretization inherent in the classification rules defined by the DT algorithm, transforming regression into a classification problem.[47] The samples corresponding to $10^{-2}$, $10^{-3}$, $10^{-4}$, $10^{-5}$, and $10^{-6}$ M would be located on the "boxes" of round concentrations, and any new sample with an intermediate concentration would be assigned to one of these boxes. This limitation would obviously be circumvented if the dataset contained a much larger number of concentrations. It should be noted that regression techniques such as Partial Least Squares (PLS) have been used to assess calibration.[48] However, regression functions can only be transformed into visual representations to support interpretation if no more than two input attributes (frequencies in our case) are considered, and this reduces the precision of the regression model in complex datasets. For the sake of illustration, a PLS model for the phytic acid concentrations using F1 as the input attribute (feature) (Figure SM4 on Supplementary Material) attains a determination coefficient $R^2$ of 0.89. On the other hand, a multivariate PLS model using F100, F10, and F1 obtains 0.99, but it cannot be transformed into a visual representation. Hence, there is a trade-off between interpretability and performance in the use of regression models, especially for complex datasets. This limitation does not exist in our approach.

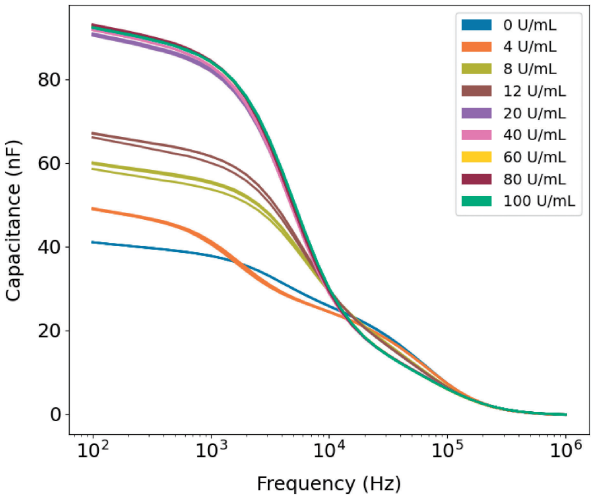



**Figure 6.** Capacitance spectra obtained with the immunosensor for 9 concentrations (0, 4, 8, 12, 20, 40, 60, 80, 100 U/mL) of pancreatic cancer biomarker CA19-9. The data were extracted from ref. 36

## 3. Results

The concept of multidimensional calibration space is now applied to two other sets of data, which are more difficult to classify. The first is from an immunosensor to detect a pancreatic cancer biomarker CA19-9,[36] while the other is from an electronic tongue used to distinguish between samples with antibodies related to Leishmaniasis and Chagas' disease.[37] The details of these experiments and how the data were obtained are given in the Supplementary Material. For each dataset, the calibration was created by building a DT over a training subdataset with 70% of all available data. The remaining subdataset with 30% was used for testing purposes. The DT parameters are selected using a model selection protocol[44–46] through a KFold Cross-Validation. We adopted as descriptors the capacitance value in nF at several frequencies in Hz (e.g., C(nF)@F100(Hz)), and the concentrations were taken as target variables. The capacitance values were used in nF scale, without any pre-processing.

**3.1 Pancreatic Cancer Biomarker CA19-9.** The detection of CA19-9 corresponds to a multiclass classification task where each class is represented by a different concentration of the biomarker in a phosphate buffer solution (PBS).[36] The capacitance spectra for 9 concentrations (M = 9, being 4, 8, 12, 20, 40, 60, 80, 100 U/mL) in Figure 6 were used, with 3 instances/samples for each concentration (N = 27).

The multidimensional calibration space for this case was created by a model selection experiment,[44–46] where the best DT parameters were chosen using a KFold Cross-Validation and the training sub-dataset (70% of the dataset). The DT created using the training sub-dataset with the best parameters provided 100% accuracy in the test sub-dataset (30% of the dataset). The technique SMOTE[49] was used for over-sampling to produce the required number of instances to split the dataset into two sub-datasets (70% training and 30% test) and perform a KFold Cross-Validation on the training sub-datset. Figure 7 presents the visualization of the multidimensional calibration space where 5 frequencies/attributes (F398107, F5011, F501,

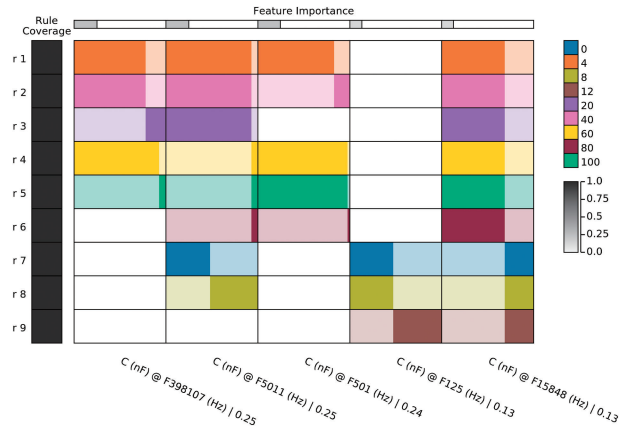



**Figure 7.** Multidimensional calibration space visualization using ExMatrix for the Pancreatic Cancer Biosensor dataset. The nine rules are defined by a DT model using 5 frequencies/attributes (F398107, F5011, F501, F125, and F15848) from 41 available (from F100 to F1000000). The nine rules provide 100% accuracy for both training (70%) and testing (30%) sub-datasets. Also, there is only one rule per class (maximum coverage) indicating a great separability capacity for the selected 5 frequencies, from where it is possible to create generic rules.

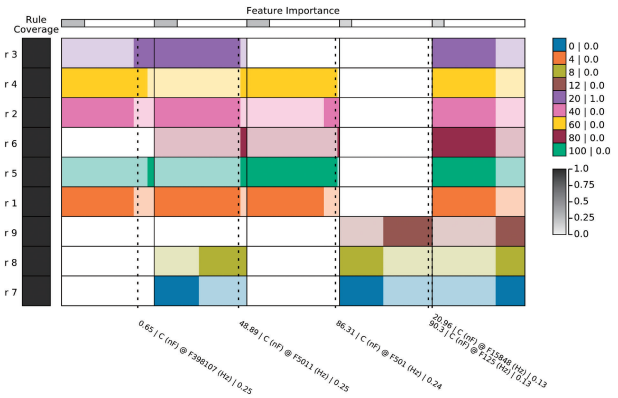



**Figure 8.** Multidimensional calibration space visualization using ExMatrix showing the classification of test instance 5. Dashed lines indicate the instance values in each frequency/attribute. In this matrix, the rules are ordered by proximity to the instance, which was classified as concentration class 20 (purple/rule 3 on the first row). To change the instance's classification from class 20 to 60 (yellow/rule 4 on the second row) requires the smallest modification, while to change to concentration class 0 (blue/rule 7 on the ninth row) the modification is the largest.

F125, F15848) are used from 41 available (F1000000 to F100). This means that the calibration space for this dataset has 5 dimensions. F398107 and F5011 are the most important (feature importance of 0.25) for the calibration (DT model), whereas F15848 is the least important (feature importance of 0.13). By inspecting the multidimensional calibration space visualization in Figure 7, one notes that the group formed by concentration classes 4, 40, 20, 60, 100, and 80 U/mL are distinguished from the group formed by concentration classes 0, 8, and 12 U/mL at F15848 (fifth column). The group formed by the concentration classes 4, 40, 20, 60, 100, and 80 U/mL (rule 1/first row to rule 6/sixth row) are distinguished into two subgroups at F5011 (second column). In the first subgroup (concentration classes 4, 40, and 20 U/mL), the concentration class 20 (purple/rule 3 on the third row) is distinguished from concentration classes 4 and 40 U/mL at F398107 (first column). Concentration class 4 (orange/rule 1 on the first row) is distinguished from concentration class 40 (magenta/rule 2 on the second row) at F501 (third column). At the second subgroup (concentration classes 60, 100, and 80 U/mL), the concentration class 80 (maroon/rule 6 on the sixth row) is distinguished from concentration classes 60 and 100 U/mL at F501 (third column). Concentration class 60 (yellow/rule 4 on the fourth row) is distinguished from concentration class 100 (emerald/rule 5 on the fifth row) at F398107 (first column). Finally, in the group formed by the concentration classes 0, 8, and 12 U/mL (rule 7/seventh row to rule 9/ninth row), the concentration class 12 (brown/rule 9 on the ninth row) is distinguished from concentration classes 0 and 8 U/mL at F125 (fourth column). The concentration class 0 (blue/rule 7 on the seventh row) is distinguished from concentration class 8 (olive/rule r8 on the eight row) at F5011 (second column).

Figure 8 illustrates the classification of the test instance 5 as concentration class 20 (purple) using rule 3 (first row), since the instance values 0.65, 48.89, 86.31, 90.3, and 20.96 for frequencies/attributes F398107, F5011, F501, F125, and F15848 are within ranges defined by rule 3. The rules are ordered by proximity to instance 5. Observing the gaps between instance values and attributes ranges for other classes, one notes that the smallest change in instance 5 that would turn its original classification is found at F5011 (second column), leading to concentration class 60 U/mL (yellow/rule 4 on the second row). The classification as concentration class 0 (blue) requires the highest change by being the last rule (rule 7 on the ninth row), having gaps between instance values and ranges on F5011 (second column), F125 (fourth column), and F15848 (fifth column).

**3.2 Detection of Chagas and Leishmaniasis with an E-Tongue.** A multidimensional calibration space was created for the data obtained with an electronic tongue (e-tongue), whose aim was to distinguish samples containing antibodies of two distinct diseases, namely Leishmaniasis (leish) and Chagas' disease (cruzi).[37] The dataset consists of 4 classes (M = 4), with 36 instances/samples for class cruzi, leish, and neg, while mist has 27 (N = 135). This e-tongue had four sensing units, two of which are biosensors for they contained antigens specific for the two diseases.[37] Figure 9 shows the capacitance spectra obtained with the e-tongue.

A DT model was chosen from a model selection experiment,[44–46] where the best DT parameters were selected using a KFold Cross-Validation over the training sub-dataset (70% of the dataset). The chosen DT model is employed to classify among the classes cruzi, leish, mist (presence of antibodies of both cruzi and leish), and neg (absence of either cruzi or leish), providing 100% accuracy over the test sub-dataset (30% of the dataset). Figure 10 shows the visualization of the multidimensional calibration space, with only 3 frequencies/attributes (F397 with Sensor 2, F158 with Sensor 3, and F1 with Sensor

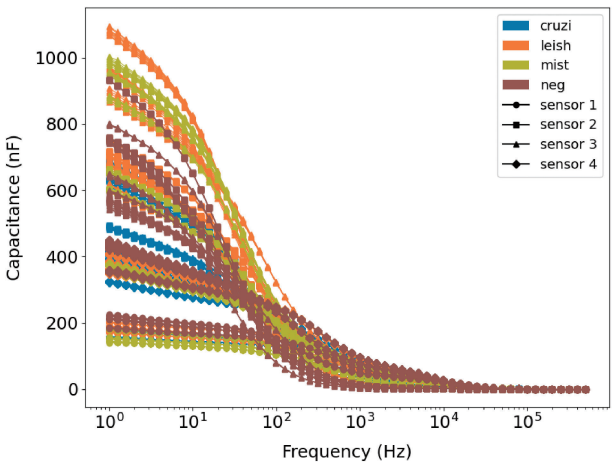



**Figure 9.** Capacitance spectra for samples with antibodies for cruzi, leish, mist (presence of antibodies of both cruzi and leish), and neg (absence of either cruzi or leish). They were obtained with 4 sensing units of an e-tongue, which explains why the plot is so crowded. These data were extracted from the work in ref. 37

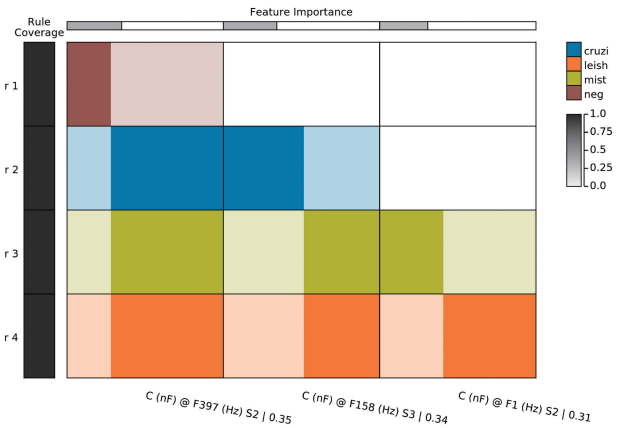



**Figure 10.** Multidimensional calibration space visualization using ExMatrix for the e-tongue data to distinguish samples with antibodies related to Chagas' disease and Leishmaniasis. The four rules are defined by a DT model, using 3 frequencies/attributes (F397 with Sensor 2, F158 with Sensor 3, and F1 with Sensor 3) from 232 available (58 frequencies from F500000 to F1 for each one of the four sensing units). The four rules provide 100% accuracy for both training (70%) and testing (30%) sub-datasets. Also, all four rules (one rule per class) have maximum coverage, indicating a high distinguishing capacity for the selected 3 frequencies, from where it is possible to create generic rules.

3) used from 232 available (58 frequencies from F500000 to F1 for each one of the four sensing units). That is to say, in spite of the apparent complexity of the data, a 3-dimensional calibration space is sufficient. F397 with Sensor 2 is the most important (feature importance of 0.35) for the calibration (DT model), whereas F1 with Sensor 2 is the least important (feature importance of 0.31). By inspecting the multidimensional calibration space, the class neg (brown/rule 1 on the first row) is distinguished from classes cruzi, mist, and leish at F397 Sensor

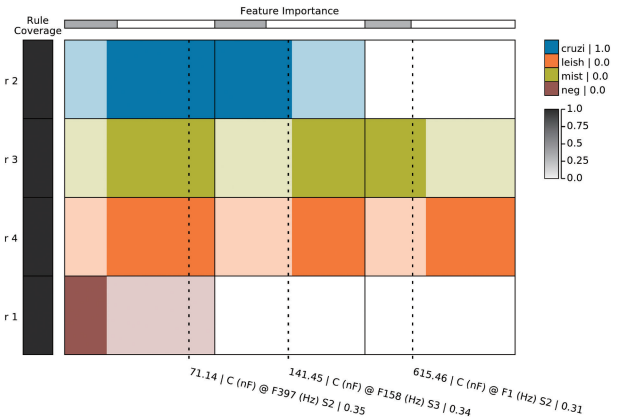



**Figure 11.** Multidimensional calibration space visualization using ExMatrix showing the classification of test instance 24. Dashed lines indicate the instance values in each frequency/attribute. In this matrix, the rules are ordered by proximity to the instance, which was classified as class cruzi (blue/rule 2 on the first row). It can be noted that to change the instance's classification from class cruzi to mist (olive/rule 3 on the second row) requires the smallest modification, while to change to class neg (brown/rule 1 on the fourth row) the largest modification is required.

2 (first column). The class cruzi (blue/rule 2 on the second row) is distinguished from classes mist and leish at F158 Sensor 3 (second column). Finally, the class mist (olive/rule 3 on the third row) is distinguished from class leish (orange/rule 4 on the fourth row) at F1 Sensor 2 (third column).

Figure 11 illustrates the classification of the test instance 24 assigned as class cruzi (blue), using rule 2 (first row) since the instance values 71.14 and 141.45 for frequencies/attributes F397 with Sensor 2 and F158 with Sensor 3 are within the ranges defined by rule 2. By having rules ordered by proximity to the instance 24, Figure 11 presents the gaps between the instance values to the ranges for other classes, where the smallest change in instance 24 that would turn its original classification is found for attribute F158 with Sensor 3 (second column), leading to the class mist (olive/rule 3 on the second row). The classification as class neg (brown) requires the largest change, represented in the figure as the last rule (rule 1 on the fourth row).

## 4. Discussion

We start the discussion by making a connection with standard calibration curves in analytical chemistry, for which we created a synthetic dataset with the signal varying linearly with a concentration. For impedance spectroscopy data used in the examples in this paper, this would correspond to the impedance (or capacitance) varying linearly with an analyte concentration at one fixed frequency. Figure S1 in the Supporting Information shows the linear dependence and the ExMatrix representation of the rules inferred upon applying a DT algorithm to the dataset is given in Figure S2. As one should expect, a one-dimensional calibration space is retrieved from the algorithm, depicted in Figure S3. By inference we state that a two-dimensional space would be obtained if the synthetic dataset were generated with calibration equations involving

two frequencies (attributes). Hence, the concept of multidimensional calibration space naturally leads to the expected calibration curves in simple cases.

Let us dwell upon the choices of the datasets employed to introduce the concept of multidimensional calibration space. The first one involving determination of phytic acid concentrations was chosen because the sensor made with a layer-by-layer film of polyelectrolytes lacked specificity. Indeed, from the visualization maps in ref 35,[35] one can infer that distinction of the different concentrations was not possible using only this sensor. However, the limited overlap observed for some concentrations meant that perhaps full distinction would be possible if machine learning methods were used. This expectation was fulfilled here and a calibration space with only 3 dimensions was sufficient to account for the whole data. In fact, this is demonstration of one of the advantages of employing DTs to generate a calibration space, for a problem that was not soluble using multidimensional projection methods (such as Principal Component Analysis (PCA)[50] and IDMAP) can now be resolved. It remains to be proven whether the calibration space would be robust enough to avoid effects from interferents, but the data available do not allow us to probe this possibility. As for the other two examples, the biosensor to detect the cancer biomarker CA19-9 and the electronic tongue (e-tongue) to distinguish the tropical diseases, they were chosen as more challenging cases of classification. Then, the need to have calibration spaces with 5 and 3 dimensions was not surprising. What may seem surprising is the need of fewer dimensions (3) for the e-tongue dataset. This is probably because the combination of different sensing units in the e-tongue concept made it easier to establish a "fingerprint" for each type of sample analyzed, and because one important unit was a biosensor.

Another advantage of the approach is the interpretability of the rules, including the coverage. Although we have not explored it here, a multidimensional calibration space may be used to analyze the parameters for sensor construction, such as the materials employed in the sensing units, types of film deposition, thickness, etc. This could be done by producing various sensors with varying parameters to yield a dataset of sensors described by their parameters (the resulting ExMatrix representation has parameters as columns). Hence, rules linking the different parameters and determining their importance for the sensing task could be established. The search for optimized conditions for a given task may be assisted by a high-throughput strategy[51] using robots to produce the sensors allowing for consistent and massive results. Also, one may analyze in detail some differences between the calibration spaces for E-Tongue Disease Detection and Pancreatic Cancer Biosensor datasets. In both calibration spaces, the selected frequencies/attributes hold distinct feature importance values. For the e-tongue in Figure 10, F397 with Sensor 2 (first column) is the most important for calibration. On the other hand, for the multidimensional calibration space for the CA19-9 dataset, F398107 is the most important, as shown in Figure 7. It should be noted that the feature importance values are related to a particular DT model, and not directly to a specific domain. Thus, they should be viewed as a model result and not a general rule for the domain, since feature importance may vary from model to model. The selected frequencies/attributes for e-tongue calibration are in the region of 100 Hz (F397, F158, and F1), while on the calibration for CA19-9 dataset the selected ones do not tend to any particular spectrum region, since frequencies from different regions are used. Both calibrations contain only rules with maximum coverage, indicating an efficient distinguishing capacity for the selected frequencies. In the rules from these calibrations, some ranges distinguish groups of classes, as for the CA19-9 dataset, where F15848 (fifth column) distinguishes the group formed by concentration classes 4, 40, 20, 60, 100, and 80 U/mL from the group formed by concentration classes 0, 8, and 12 U/mL.

All the examples discussed here derived from sensors and biosensors based on impedance spectroscopy. The same principles could be applied if detection (or more generally classification) were to be made with other experimental (or theoretical or simulation) data. For instance, in the various types of optical and vibrational spectroscopies the frequencies would be the attributes employed as input into the DT algorithms. In electrochemical methods, e.g. cyclic voltammetry and amperometry, and in detection based on electrical measurements the attributes could be current or voltages. In datasets representing time series then distinct times could be the attributes. Furthermore, the concept of the multidimensional space can be extended to datasets containing non-structured data (such as text and images). The only limitation in the approach is whether the data are sufficient to cover the whole space. That is to say, the ability to predict and classify unknown data may not reach 100% if the data available for the learning are not representative of all possibilities. In addition, because the number of rules and dimensions may be too large, the utility of the calibration space could be questionable. Though the choice of DT as the machine-learning paradigm along ExMatrix was made precisely to make the classification interpretable, the interpretation task by a human would be nearly impossible if there are too many rules and dimensions. In this context, one should comment on the possibility of using other methods. Blackboxes[38,39] such as Artificial Neural Networks[15] and Support Vector Machines[40,41] are not interpretable, and therefore not suitable for our purpose to establish a calibration space. Random Forest (RF)[43] models could generate too many rules and dimensions, but they may be used in scenarios where single DT models are not able to provide good performance. This was not the case of the results presented in this paper. Once the DT models provided good performance, they are preferable based on the Occam's Sharp Razor,[52] according to which the simpler model (more comprehensible) should be chosen when different models provide the same performance.

As for the utility of the multidimensional calibration space, we may list a few possibilities. It can be used to classify samples in complex cases or in complex matrices which would not be possible by using only multivariate statistics, as already mentioned with the demonstration for detection of phytic acid here. It may be used to integrate data from different natures — as in clinical diagnostics where images, text and sensing data are combined.[53] With the coverage of the whole information in the datasets, one is likely to obtain a more efficient classification. The multidimensional calibration space can also be used in materials design, which has been an important application of machine learning in chemistry and materials science.[54] Indeed,

chemical elements described by their properties could be combined to generate new materials according to rules generated and visualized with ExMatrix. Perhaps one the main uses of the concept will be in e-tongues and e-noses, which have been demonstrated in a large number of papers, but only a few commercial products are available.[15,55] They can in principle be used to build libraries of responses of different liquids and vapours,[15,19,55] in many cases providing excellent ability to classify similar samples. However, this high sensitivity and selectivity come at a price since the electrical response is too sensitive to any change in the measuring conditions and indeed on the sensing units. If one sensor of the array has to be changed, which is inevitable after many measurements in some applications, then the whole library built with the previous set of sensors is useless. This is one of the main reasons why it has been so difficult for e-tongues and e-noses to reach the market, for an entire procedure of recalibration would be necessary to recover the information in the library. With the multidimensional calibration space introduced here it may be possible to retrieve such a library with relatively few experiments as most of the rules are unlikely to change when one sensor is replaced by a similar (generally nominally identical) one. Furthermore, the strategy based on the calibration space could help one understand why nominal identical biosensors produce different responses, assisting researchers in facing biological variability problems.

## 5. Final Remarks

The concept of a multidimensional calibration space introduced here exploits the immense potential from the use of machine learning methods to analyze data. It allows for a predictive power that is unprecedented for some types of data, as demonstrated in this paper with simple examples of sensors, biosensors and e-tongues. In some aspects this calibration space resembles the multivariate calibration space, and indeed for simple cases with small dimensions they may coincide. However, the concept of a multidimensional calibration space is broader, not only because rules are generated with machine learning but also because different types of data may be treated (see below) and visualized. Two features to be highlighted are the use of Decision Trees (DT) algorithms, which permit the generation of interpretable rules, and the visualization of such rules with ExMatrix software. Based on the cases of sensing data we analyzed, it seems that the number of rules — and dimensions of the space — is likely to be less than 10 or a few tens for complex, large datasets. Visualization and interpretation of the rules should therefore be manageable with ExMatrix (or a similar visualization tool), an essential requirement for the usability of the approach. Since the multidimensional calibration space may be applied to any type of data, including images, videos, text, in addition to scientific data, applications beyond analytical chemistry can be envisaged. Examples can be surveillance and monitoring systems of various kinds, computer-assisted clinical diagnostics and natural language processing. As for the limitations of the approach, we have already mentioned those associated with DTs yielding assignment into classes, rather than having real numbers as output. Another difficulty is to prove that a given dataset is sufficient for covering the whole space for the problem under analysis. This is frequent in clinical diagnosis, being a limitation that harms the predictive power of the calibration space.

This work was supported by São Paulo Research Foundation (FAPESP) (Grants #2018/22214-6 and #2018/18953-8). The authors also wish to express thanks for the support received from the Qualification Program of the Federal Institute of São Paulo (IFSP), as well as from the Natural Sciences and Engineering Research Council of Canada (NSERC).

## Supporting Information

Details about experimental procedures for the e-tongues/biosensor fabrication and an example of a one-dimensional calibration space! This material is available on https://doi.org/10.1246/bcsj.20200359.

## References

1 L. A. Currie, *Pure Appl. Chem.* **1995**, *67*, 1699.
2 L. A. Currie, *Anal. Chim. Acta* **1999**, *391*, 105.
3 S. M. Moosavi, S. Ghassabian, in *Calibration and Validation of Analytical Methods - A Sampling of Current Approaches*, ed. by M. T. Stauffer, InTech, **2018**.
4 M. Valcárcel Cases, Á. I. López-Lorente, M. Á. López-Jiménez, *Foundations of Analytical Chemistry: a Teaching-Learning Approach*, Springer, Cham, **2018**.
5 F. M. Shimizu, F. R. Todão, A. L. Gobbi, O. N. Oliveira, Jr., C. D. Garcia, R. S. Lima, *ACS Sens.* **2017**, *2*, 1027.
6 M. Braunger, F. Shimizu, M. Jimenez, L. Amaral, M. Piazzetta, Â. Gobbi, P. Magalhães, V. Rodrigues, O. N. Oliveira, Jr., A. Riul, *Chemosensors* **2017**, *5*, 14.
7 F. M. Shimizu, M. L. Braunger, A. Riul, *Chemosensors* **2019**, *7*, 36.
8 C. M. Daikuzono, C. A. R. Dantas, D. Volpati, C. J. L. Constantino, M. H. O. Piazzetta, A. L. Gobbi, D. M. Taylor, O. N. Oliveira, Jr., A. Riul, *Sens. Actuators, B* **2015**, *207*, 1129.
9 *Electronic noses and tongues in food science*, ed. by V. Preedy, M. L. R. Méndez, Elsevier: Academic Press, Amsterdam; Boston; Heidelberg; London; New York; Oxford; Paris; San Diego; San Francisco; Singapore; Sydney; Tokyo, **2016**.
10 A. TermehYousefi, *Nanocomposite-based electronic tongue: carbon nanotube growth by chemical vapor deposition and its application*, Springer, Cham, **2018**.
11 J. E. Oliveira, V. Grassi, V. P. Scagion, L. H. C. Mattoso, G. M. Glenn, E. S. Medeiros, *IEEE Sens. J.* **2013**, *13*, 759.
12 M. V. Farraia, J. Cavaleiro Rufo, I. Paciência, F. Mendes, L. Delgado, A. Moreira, *Porto Biomed. J.* **2019**, *4*, e42.
13 H. K. Patel, *The electronic nose: artificial olfaction technology*, Springer, New Delhi, **2014**.
14 C. Di Natale, R. Paolesse, A. Macagnano, A. Mantini, A. D'Amico, A. Legin, L. Lvova, A. Rudnitskaya, Y. Vlasov, *Sens. Actuators, B* **2000**, *64*, 15.
15 A. Riul, H. C. de Sousa, R. R. Malmegrim, D. S. dos Santos, A. C. P. L. F. Carvalho, F. J. Fonseca, O. N. Oliveira, Jr., L. H. C. Mattoso, *Sens. Actuators, B* **2004**, *98*, 77.
16 A. Rudnitskaya, L. M. Schmidtke, A. Reis, M. R. M. Domingues, I. Delgadillo, B. Debus, D. Kirsanov, A. Legin, *Food Chem.* **2017**, *229*, 20.
17 P. Alessio, C. J. Leopoldo Constantino, C. M. Daikuzono, A. Riul, O. N. Oliveira, Jr., in *Electronic Noses and Tongues in Food Science*, Elsevier, **2016**, pp. 171–177.
18 A. Legin, A. Rudnitskaya, Y. Vlasov, in *Comprehensive*

*Analytical Chemistry*, Elsevier, **2003**, Vol. 39, pp. 437–486.
19 A. M. Graboski, C. Zakrzevski, F. M. Shimizu, R. T. Paschoalin, A. C. Soares, J. Steffens, N. Paroul, C. Steffens, *ACS Sens.* **2020**, acssensors.0c00636.
20 Y. Vlasov, A. Legin, A. Rudnitskaya, *Anal. Bioanal. Chem.* **2002**, *373*, 136.
21 M. Podrażka, E. Bączyńska, M. Kundys, P. Jeleń, E. Witkowska Nery, *Biosensors* **2018**, *8*, 3.
22 A. R. Di Rosa, F. Leone, F. Cheli, V. Chiofalo, *J. Food Eng.* **2017**, *210*, 62.
23 C. M. Daikuzono, F. M. Shimizu, A. Manzoli, A. Riul, M. H. O. Piazzetta, A. L. Gobbi, D. S. Correa, F. V. Paulovich, O. N. Oliveira, Jr., *ACS Appl. Mater. Interfaces* **2017**, *9*, 19646.
24 D. Wilson, E. M. Materón, G. Ibáñez-Redín, R. C. Faria, D. S. Correa, O. N. Oliveira, Jr., *Talanta* **2019**, *194*, 611.
25 A. S. Lundervold, A. Lundervold, *Z. Med. Phys.* **2019**, *29*, 102.
26 V. da C. Rodrigues, C. H. Comin, J. C. Soares, A. C. Soares, M. E. Melendez, J. H. T. G. Fregnani, A. L. Carvalho, L. da F. Costa, O. N. Oliveira, Jr., *ACS Appl. Mater. Interfaces* **2017**, *9*, 5885.
27 V. C. Rodrigues, J. C. Soares, A. C. Soares, D. C. Braz, M. E. Melendez, L. C. Ribas, L. F. S. Scabini, O. M. Bruno, A. L. Carvalho, R. M. Reis, R. C. Sanfelice, O. N. Oliveira, Jr., *Talanta* **2021**, *222*, 121444.
28 R. Vashistha, A. K. Dangi, A. Kumar, D. Chhabra, P. Shukla, *3 Biotech* **2018**, *8*, 358.
29 Z. S. Ballard, H.-A. Joung, A. Goncharov, J. Liang, K. Nugroho, D. Di Carlo, O. B. Garner, A. Ozcan, *npj Digital Med.* **2020**, *3*, 66.
30 N. Banaei, J. Moshfegh, A. Mohseni-Kabir, J. M. Houghton, Y. Sun, B. Kim, *RSC Adv.* **2019**, *9*, 1859.
31 X. Li, T. Yang, S. Li, L. Jin, D. Wang, D. Guan, J. Ding, *Opt. Express* **2015**, *23*, 18361.
32 F. Gonzalez-Navarro, M. Stilianova-Stoytcheva, L. Renteria-Gutierrez, L. Belanche-Muñoz, B. Flores-Rios, J. Ibarra-Esquer, *Sensors* **2016**, *16*, 1483.
33 *Classification and regression trees*, Repr ed., ed. by L. Breiman, Chapman & Hall [u.a.], Boca Raton, **1998**.
34 P.-N. Tan, M. Steinbach, A. Karpatne, V. Kumar, *Introduction to data mining*, Second edition ed., Pearson, NY NY, **2019**.
35 M. L. Moraes, R. M. Maki, F. V. Paulovich, U. P. Rodrigues Filho, M. C. F. de Oliveira, A. Riul, N. C. de Souza, M. Ferreira, H. L. Gomes, O. N. Oliveira, Jr., *Anal. Chem.* **2010**, *82*, 3239.
36 A. C. Soares, J. C. Soares, F. M. Shimizu, M. E. Melendez, A. L. Carvalho, O. N. Oliveira, Jr., *ACS Appl. Mater. Interfaces* **2015**, *7*, 25930.
37 Â. C. Perinoto, R. M. Maki, M. C. Colhone, F. R. Santos, V. Migliaccio, K. R. Daghastanli, R. G. Stabeli, P. Ciancaglini, F. V. Paulovich, M. C. F. de Oliveira, O. N. Oliveira, Jr., V. Zucolotto, *Anal. Chem.* **2010**, *82*, 9763.
38 P. Hall, *arXiv:1810.02909 [cs, stat]* **2020**.
39 R. Guidotti, A. Monreale, S. Ruggieri, F. Turini, D. Pedreschi, F. Giannotti, *arXiv:1802.01933 [cs]* **2018**.
40 H. Song, Y. Wang, J. M. Rosano, B. Prabhakarpandian, C. Garson, K. Pant, E. Lai, *Lab Chip* **2013**, *13*, 2300.
41 R. Kumar, A. P. Bhondekar, R. Kaur, S. Vig, A. Sharma, P. Kapur, *Sens. Actuators, B* **2012**, *171–172*, 1046.
42 M. Popolin Neto, F. V. Paulovich, *IEEE Trans. Vis. Comput. Graph.* **2021**, *27*, 1427.
43 L. Breiman, Manual On Setting Up, Using, And Understanding Random Forests V3.1, https://www.stat.berkeley.edu/~breiman/Using_random_forests_V3.1.pdf.
44 *An introduction to statistical learning: with applications in R*, ed. by G. James, D. Witten, T. Hastie, R. Tibshirani, Springer, New York, **2013**.
45 A. Jahangiri, H. A. Rakha, *IEEE Trans. Intell. Transp. Syst.* **2015**, *16*, 2406.
46 R. Kohavi, *IJCAI'95: Proceedings of the 14th international joint conference on Artificial intelligence* **1995**, *2*, 1137.
47 R. Salman, V. Kecman, 2012 Proceedings of IEEE Southeastcon, Orlando, FL, USA, 1–6.
48 D. Kirsanov, O. Mednova, V. Vietoris, P. A. Kilmartin, A. Legin, *Talanta* **2012**, *90*, 109.
49 N. V. Chawla, K. W. Bowyer, L. O. Hall, W. P. Kegelmeyer, *J. Artif. Intell. Res.* **2002**, *16*, 321.
50 I. T. Jolliffe, J. Cadima, *Philos. Trans. R. Soc., A.* **2016**, *374*, 20150202.
51 S. Michael, D. Auld, C. Klumpp, A. Jadhav, W. Zheng, N. Thorne, C. P. Austin, J. Inglese, A. Simeonov, *Assay Drug Dev. Technol.* **2008**, *6*, 637.
52 P. Domingos, *KDD'98: Proceedings of the Fourth International Conference on Knowledge Discovery and Data Mining* **1998**, 37.
53 J. F. Rodrigues, F. V. Paulovich, M. C. de Oliveira, O. N. Oliveira, Jr., *Nanomedicine* **2016**, *11*, 959.
54 G. M. Dalpian, O. N. Oliveira, Jr., *ACS Appl. Mater. Interfaces* **2019**, *11*, 24823.
55 E. J. Ferreira, R. C. T. Pereira, A. C. B. Delbem, O. N. Oliveira, Jr., L. H. C. Mattoso, *Electron. Lett.* **2007**, *43*, 1138.