



# Dairy products discrimination according to the milk type using an electrochemical multisensor device coupled with chemometric tools

Imam Tazi<sup>1,2</sup> · Kuwat Triyana<sup>1,3</sup> · Dwi Siswanta<sup>4</sup> · Ana C. A. Veloso<sup>5,6</sup> · António M. Peres<sup>7,8</sup> · Luís G. Dias<sup>7</sup>

Received: 9 March 2018 / Accepted: 12 June 2018  
© Springer Science+Business Media, LLC, part of Springer Nature 2018

## Abstract

This study shows the potential application of a potentiometric electronic tongue coupled with a lab-made DataLogger device for the classification of dairy products according to the type of milk used in their production, i.e., natural, fermented and UHT milk. The electronic tongue device merged a commercial pH electrode and 15 lipid/polymeric membranes, which were obtained by a drop-by-drop technique. The potentiometric signal profiles gathered from the 16 sensors, during the analysis of the 11 dairy products (with ten replicate samples), together with principal component analysis showed that dairy samples could be naturally grouped according to the three types of milk evaluated. To further investigate and verify this capability, a linear discriminant analysis together with a simulated annealing variable selection algorithm was also applied to the electrochemical data, which were randomly split into two datasets, one used for model training and internal-validation using a repeated K-fold cross-validation procedure (with 64% of the data); and the other for external validation purposes (containing the remaining 36% of the data). The multivariate supervised strategy used allowed establishing a classification model, based on the potentiometric information of four sensor lipid membranes, which enabled achieving a successful discrimination rate of 100% for both internal- and external-validation processes. The demonstrated versatility of the built electronic tongue for discriminating dairy products according to the type of milk used in their production combined with its simplicity, low-cost and fast time analysis may envisage a possible future application in dairy industry.

**Keywords** Dairy products · Electronic tongue · Principal component analysis · Linear discriminant analysis · Simulated annealing algorithm

## Introduction

Foods characterization, namely of dairy products is a hard task taking into account the complexity and inherent variability of the composition of these products. Among dairy

products, milk quality control during processing and storage is still of major interest for the dairy industry [1]. Milk and milk based products contribute to a healthy human diet being a key source of bio-available elements/compounds like calcium and proteins [1], being quite important to ensure their

✉ Kuwat Triyana  
triyana@ugm.ac.id

<sup>1</sup> Physics Department, Universitas Gadjah Mada, Sekip Utara BLS 21, Yogyakarta, Indonesia

<sup>2</sup> Physics Department, Universitas Islam Negeri Maulana Malik Ibrahim Malang, Malang, Indonesia

<sup>3</sup> Interdisciplinary Halal Research Group, Universitas Gadjah Mada, Jl. Kaliurang Km. 4, Sekip Utara, Yogyakarta, Indonesia

<sup>4</sup> Chemistry Department, Universitas Gadjah Mada, Sekip Utara BLS 21, Yogyakarta, Indonesia

<sup>5</sup> Instituto Politécnico de Coimbra, ISEC, DEQB, Rua Pedro Nunes, Quinta da Nora, 3030-199 Coimbra, Portugal

<sup>6</sup> CEB - Centre of Biological Engineering, University of Minho, Campus de Gualtar, 4710-057 Braga, Portugal

<sup>7</sup> Centro de Investigação de Montanha (CIMO), ESA, Instituto Politécnico de Bragança, Campus Sta Apolónia, 5300-253 Bragança, Portugal

<sup>8</sup> Laboratory of Separation and Reaction Engineering - Laboratory of Catalysis and Materials (LSRE-LCM), ESA, Instituto Politécnico de Bragança, Campus Santa Apolónia, 5300-253 Bragança, Portugal

safety and quality. To increase milk shelf-stability, its microbial population must be reduced, which is usually achieved by thermal processing of milk. However, these technological processes highly affect the sensory attributes of milk as well as physicochemical changes that may condition protein functionality and the bioavailability of the nutrients [2]. Usually, food analysis is carried out using standard analytical techniques such as chromatography and spectroscopy [1, 3]. However, these methodologies are non-portable, hardly miniaturized and difficult to implement for on-line analysis. Moreover, they are expensive, time-consuming and require specialized technicians, exciding the economic possibilities of the majority of small milk producers. The use of fluorescence sensor arrays for classifying commercial milk samples according to their thermal treatments (i.e. pasteurized milk, sterilized milk, UHT fresh milk and recombined milk) has been reported [2]. Also, near infrared (NIR) portable sensors [4] have been applied for monitoring, in situ, milk composition.

Sensor-arrays, namely electronic tongues (E-tongues), have been reported as potential, versatile and advantageous electrochemical tools within food analysis field, mainly since they are able to quantitatively estimate characteristic properties of food samples, to distinguish among several types of foods and to recognize taste attributes [3]. In fact, sensor-arrays allow collecting broad information on attributes like quality, condition or process instead of measuring single parameters using specific sensors [5]. Concerning dairy products, the successful performance of E-tongues in recognizing milk types may be attributed to the quite different composition of raw milk according to the animal origin (e.g., cow, goat, sheep, etc.), the different processing technologies applied in the production of milk, or the incorporation of different flavors in milk-based beverages. Indeed, in the last 10–15 years several papers have reported the use of E-tongues, based on different electrochemical approaches (e.g., potentiometric using ion selective electrodes or lipid polymeric membranes; voltammetric using one or more working electrodes; etc.), combined with different chemometric techniques (e.g., principal component analysis, PCA; linear discriminant analysis, LDA; artificial neural networks, ANN; support vector machines, SVM; partial least square, PLS; etc.) for qualitative, semi-quantitative or quantitative milk analysis. Among these works the main application concerned milk classification according to the brand, fat content or storage time, being some studies focused on the capability of monitoring milk quality.

Voltammetric based approaches have been reported for milk evaluation namely to monitor milk deterioration process or its quality during storage, using ANN, PCA or cluster analysis approaches as well as to predict bacterial growth in the milk samples using ANN, PLS or least square-SVM models [6, 7]; to monitor, on-line, off-flavors

of raw milks as well as to monitor, using PCA, the cleaning process [8]; to classify fermented milk samples using PCA and ANN models validated through a leave-one-out (LOO) cross-validation (CV) procedure [5]; to classify pasteurized milk according to its brand and for assessing the milk's storage time based on PCA or SVM, which allowed 100% of correct classification for LOO-CV [1]; to classify pure milk samples by brand using a PLS-discriminant analysis method that enabled a classification rate of 100% for samples included in a test dataset [9]; or to successfully detect milk samples adulterated with urea by means of multi-way PCA and multi-way PLS discriminant analysis or PCA-least square SVM, allowing 89–94% of correctly identified adulterated or pure milk samples included in a test dataset [10, 11].

On the other hand, potentiometric based devices have also been used for milk assessment. In more detail, Sim et al. [12] proposed a disposable screen-printed multi-channel potentiometric E-tongue, with several types of lipid compounds as transducers, to discriminate between fresh and spoiled UHT or pasteurized milk as well as to follow the deterioration of the milk quality when it is stored at room temperature based on PCA. Other researchers [3, 13] showed that a miniaturized solid-state potentiometric E-tongue operating either in stationary or flow-through mode, together with SVM or ANN, could be successfully applied to classify UHT milk samples according to brand and fat contents (97% of correctly classified samples belonging to an external test dataset). Similar satisfactory results (successful recognition rate from 80 to 100% for a test dataset) could also be achieved when using an E-tongue with selective and partially selective electrodes together with PLS-discriminant analysis [14] or when using a miniaturized E-tongue with an integrated reference microelectrode [15]. Dias et al. [16] showed that a lab-made potentiometric E-tongue with cross-sensitivity sensors could be used to satisfactorily identify goat milk adulteration with bovine milk, using LDA models that were validated by LOO-CV (sensitivity and specificity from 70 to 87%). A commercial E-tongue device (i.e.,  $\alpha$ -Astree analyzer), which included sensors coated with lipid/polymer material was also able to satisfactorily classify, using a PCA, milk samples according to different producers [17], to monitor changes in probiotic fermented milk during storage, using PCA, as well as to discriminate probiotic fermented milk samples with different flavors according to brand (ANN models with sensitivities from 63 to 100%) and to predict (using PLS and/or ANN) the intensity flavor perception of a human sensory panel or aroma compounds of fermented milk [18, 19]. Recently, Tazi et al. [20] developed an E-tongue, comprising lipid/polymer sensor membranes coupled with a microcontroller-based data acquisition system and a moving average filter procedure, to recognize dairy products according to their brands or to detect taste

change of bovine and goat milks in room environment, using PCA and/or LDA.

The overall satisfactory performances of E-tongue based devices for milk analysis has motivated the present work. A potentiometric taste sensor array, comprising 16 different sensors (15 lipid/polymeric membranes and an external pH sensor), was built and applied to evaluate its capability for differentiating dairy products (natural milk, flavored natural milk, drinking-style yogurt, probiotic drink and flavored ultra-high temperature (UHT) milk) according to the type of milk used in their production (i.e., natural, fermented and UHT milks). The E-tongue performance was evaluated by unsupervised (e.g., PCA) and supervised (e.g., LDA) classification multivariate methods. For the LDA approach, data was split into two sets (training dataset: for learning purposes; and test dataset: for external validation). Moreover, a meta-heuristic variable selection algorithm (i.e., the simulated annealing algorithm, SA) was applied to establish the best sub-set of sensors required for improving the correct classification rate of the samples of the test set. To minimize possible overfitting issues, during the training step, a repeated K-fold cross-validation (CV) procedure, with 10-folds and 10 repeats, was implemented and the best E-tongue-LDA-SA classification model was further validated using a test dataset.

## Materials and methods

### Reagents

The main components of sensors or lipid/polymeric membranes were made of polymer matrix: polyvinyl chloride from Fluka, Switzerland; plasticizer compound:

2-nitrophenyl octyl ether, Bis (2-ethylhexyl) sebacate, Bis (2-ethylhexyl) phosphate, bis (1-butyl pentyl) adipate from Fluka; lipids as material additive: Octadecylamine, Oleyl alcohol, Methyltrioctylammonium chloride, and Oleic acid from Fluka.

### Lipid/polymeric membranes

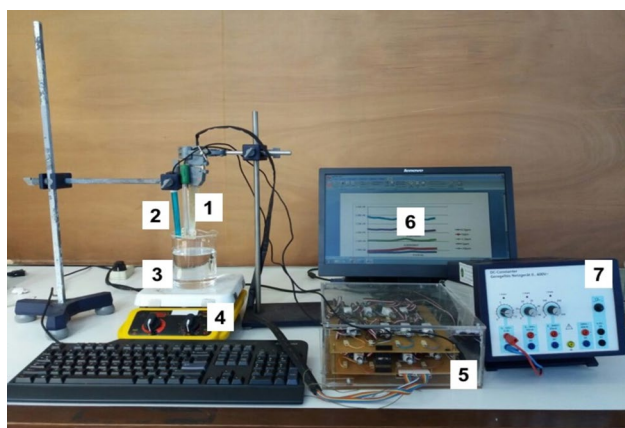
All components of the lipid/polymeric membrane were dissolved in tetrahydrofuran (THF) from Fluka. Table 1 shows the composition of 15 polymeric sensors included in the multi-sensor device. The type of additive compound and plasticizer as well as their relative proportions were selected since, previously, it has been demonstrated in the literature that, similar electrochemical devices, also comprising lipid polymeric sensor membranes, give qualitative and/or quantitative responses towards standard solutions mimicking basic tastes (e.g., acid, bitter, salty, sweet, and umami), positive sensory attributes (e.g., bitter, fruity, green and pungency) or sensory defects (e.g., butyric, musty, putrid, winey-vinegary and zapateria) [21–26], allowing successful application of such electrochemical devices in food analysis.

### Potentiometric sensor array

For this work, a multi-sensor device was built with the possibility of comprising 16 sensors. The device comprised gold electrodes (thin layers), which were connected to silver wires (purity of 99.9%). The body of the sensor device was made with an acrylic board and, on both sides, eight gold electrodes were placed as described in a previous work [20]. At the other end, each silver wire was connected to a copper cable, allowing the connection to a self-made DataLogger. To prevent from deterioration, all the silver connections were

**Table 1** Lipid/polymeric membrane composition of each sensor

No. sensor	Additive compound (2.8–3.2%)	Plasticizer (64.7–65.2%)	Polymer (31.9–32.3%)
S1	Octadecylamine	2-Nitrophenyl octyl ether	Polyvinyl chloride
S2	Oleyl alcohol	2-Nitrophenyl octyl ether	Polyvinyl chloride
S3	Methyltrioctylammonium chloride	2-Nitrophenyl octyl ether	Polyvinyl chloride
S4	Oleic acid	2-Nitrophenyl octyl ether	Polyvinyl chloride
S5	Octadecylamine	Bis(2-ethylhexyl) sebacate	Polyvinyl chloride
S6	Oleyl alcohol	Bis(2-ethylhexyl) sebacate	Polyvinyl chloride
S7	Methyltrioctylammonium chloride	Bis(2-ethylhexyl) sebacate	Polyvinyl chloride
S8	Oleic acid	Bis(2-ethylhexyl) sebacate	Polyvinyl chloride
S9	Octadecylamine	Bis(2-ethylhexyl) phosphate	Polyvinyl chloride
S10	Oleyl alcohol	Bis(2-ethylhexyl) phosphate	Polyvinyl chloride
S11	Methyltrioctylammonium chloride	Bis(2-ethylhexyl) phosphate	Polyvinyl chloride
S12	Oleic acid	Bis(2-ethylhexyl) phosphate	Polyvinyl chloride
S13	Octadecylamine	Bis(1-butylpentyl) adipate	Polyvinyl chloride
S14	Oleyl alcohol	Bis(1-butylpentyl) adipate	Polyvinyl chloride
S15	Methyltrioctylammonium chloride	Bis(1-butylpentyl) adipate	Polyvinyl chloride



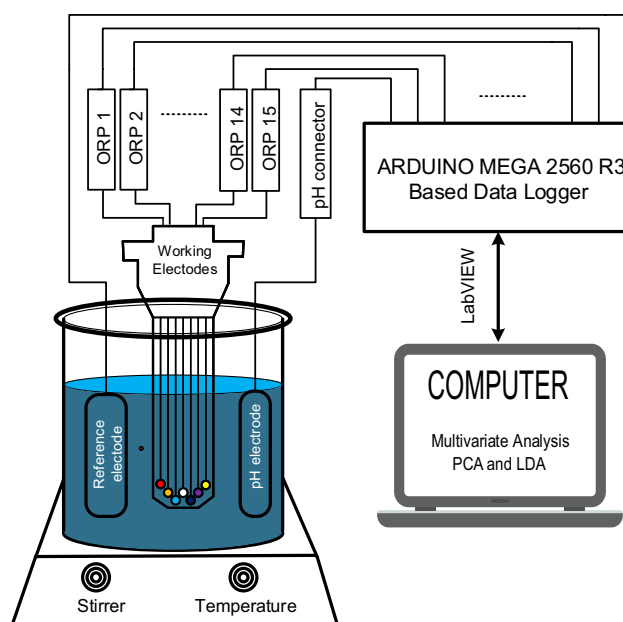
**Fig. 1** Photograph of the E-tongue system. (1) Sensor array comprising the working electrodes, (2) reference electrode, (3) sample container, (4) magnetic stirrer, (5) DataLogger, (6) personal computer, and (7) DC power supply

isolated with an acrylic resin. Figure 1 shows the photograph of the E-tongue system containing working and reference electrodes connected to the DataLogger. Since the DataLogger built only had 16 channels and it was intended to use an external pH electrode (SENPH001 from SFE electronics), in the multi-sensor device, the 15 lipid/polymeric membranes (described in Table 1) were placed into the gold electrodes 1–15, being electrode 16 not used. A double junction reference electrode Ag/AgCl (Hanna Instruments) was used as a reference electrode for all sensor inputs.

A DataLogger was built for recording the potentiometric signal profiles. As shown in Fig. 2, a microcontroller board of Arduino Mega 2560 R3 (based on the ATmega2560) having 16 channel input analog read and 54 digital input/output pins (of which 15 can be used as PWM outputs) was used [20]. Each working electrode was connected to an adapter interface (system based on the 1130-pH/ORP Adapter Interfaces, from Phidgets Inc.). The roles of each adapter interface are as a buffer and an amplifier. The output of each amplifier was then filtered using moving average method with intervals of 20 s directly during the measurement process. This filtering process was carried out using of LabView-based virtual instrumentation software.

### Dairy products samples and E-tongue analysis

Eleven milk types (natural and flavoured milk, drinking-style yogurts and probiotic drink), from 7 different brands, were purchased at a local Indonesian supermarket, being seven of them flavored (white tea, chocolate, and strawberry). All milk samples were produced in 2015, and for each of the 11 milk types, 10 different packages (i.e., independent samples) were analyzed. Three types of milk were used in the production of these dairy products: natural, fermented and UHT



**Fig. 2** Schematic diagram of E-tongue system with an array of 15 sensors and a pH electrode

milk. Details regarding all samples evaluated are given in Table 2. The probiotic drink used in this study was fermented milk obtained using *Lactobacillus casei* strain Shirota [27]. This probiotic drink is different from yogurt, which generally uses bacteria such as *Streptococcus salivarius* subsp. *thermophilus* and/or *L. delbrueckii* subsp. *Bulgaricus* [28]. As shown in Table 2, the milk samples were labeled with a letter considering the first letter of milk type (n, natural; f, fermented; u, ultra-high temperature milk, UHT) and, within each milk type, the sample number. According to the label information milks belonging to the natural milk group had fat contents from 3 to 11% and a sugar content of 3%, for those belonging to fermented milk group the fat contents varied between 0 and 3% and the sugar contents from 4 to 5%, and finally UHT milks had fat contents ranging from 4 to 9% and sugar contents varying from 5 to 10%.

For the E-tongue assays, 50 mL of each milk sample were placed in a 100-mL beaker to ensure that the sensor device was completely immersed in the sample solution and stirred during the measurement (~5 min) to obtain a homogenous solution. To ensure the stability and repeatability of each sensor, the potentiometric measurement was carried out during 10 min with 10 repeated measurements for each dairy sample.

### Data processing

The matrix data used for multivariate analysis included the E-tongue signal profiles (16 sensors comprising 15 lipid/polymeric membranes plus a pH electrode) recorded during

**Table 2** Milk samples evaluated: dairy product, brands, flavor and type of processed milk

Sample	Dairy products	Brand	Flavour	Types of milk
n1	Natural milk	1	None	Natural milk
n2	Flavored natural milk	1	White tea	Natural milk
f1	Drinking-style yogurt	2	None	Fermented milk
f2	Drinking-style yogurt	3	None	Fermented milk
f3	Probiotic drink	4	None	Fermented milk
u1	Flavored milk	5	Chocolate	Ultra-high temperature milk
u2	Flavored milk	5	Strawberry	Ultra-high temperature milk
u3	Flavored milk	6	Strawberry	Ultra-high temperature milk
u4	Flavored milk	6	Chocolate	Ultra-high temperature milk
u5	Flavored milk	7	Chocolate	Ultra-high temperature milk
u6	Flavored milk	7	Strawberry	Ultra-high temperature milk

each dairy sample (natural, fermented and UHT milk) analysis. A radar plot was used for illustrating the variability between average potentiometric sensor profiles for each dairy sample, which were centered and scaled to allow direct comparison between sensor profiles, enabling to evidence the existence of possible pattern differences within samples.

The PCA is an unsupervised pattern recognition method that can reduce the dimensionality of the data by orthogonal transformation into principal components (PCs), a new set of variables (uncorrelated) that are linear functions of the original variables (possibly correlated variables). The number of PCs depends on data variability but are usually less than or equal to the number of original variables. It is used mainly as an exploratory data analysis tool since it allows to reveal the internal structure of the data considering that the new variables represent, by decreasing order, the variability in the data. This means that the first PC gives the most relevant information because it represents the larger amount of information present in the original data. This technique was applied to the potentiometric signals data matrix in order to evaluate data variability. For this, data were previously centered and scaled.

Linear discriminant analysis (LDA), a supervised pattern recognition method, was also used to verify the capability of the E-tongue to correctly classify dairy samples according to the type of milk (natural, fermented and UHT milk) used in their production. The best subset of  $k$  sensors (variable selection) of the 16 sensors included in the potentiometric E-tongue was chosen using a meta-heuristic simulated annealing (SA) variable selection algorithm [29–31]. The SA algorithm searches for a global minimum that optimizes a system with the minimum variables, using an iterative procedure, based on the assumptions of the annealing physic process. In each iteration, using the new subset of variables, the obtained  $\tau^2$  quality criterion (a measure of the goodness of fitting) is compared, and the new solution is selected if it gave a better result than the initial one. For this, a 10,000 attempts were used to select the best subset of  $k$  variables

(best model), starting the process of selecting the best subsets of variables on each attempt in order to assure the true optimal solution. The selection process also included the evaluation of the optimal number of sensors in the LDA model, being the search carried out for subsets containing from 2 to 10 sensors. For each model, based on different subsets of sensors, the best model was chosen considering a repeated K-fold cross-validation procedure (repeated K-fold-CV: 10 repeats with tenfolds, ensuring that at each iteration 10% of the data was used for internal-validation purposes) in order to minimize possible overfitting issues and by that, reducing the risk of over-optimistic results usually reported when only LOO-CV based procedures are used. For implementing this procedure, data (centered and scaled) was first split into two data sets, the train set, for model development and the test set for external validation. The training dataset was obtained randomly in order to contain approximately 64% of the overall data (i.e., 70 of the 110 samples), ensuring that all electrochemical assays of the same milk sample were included in the set, and using the remaining data for external validation purposes (i.e., 36% of the data, resulting into a dataset with 40 samples). The overall best model was the one that used the lowest number of sensors and showed the maximum sensitivity and specificity when applied for evaluating the samples of the test group. Sensitivity corresponds to the proportion of positive, correct classifications and specificity, the proportion of negatives that are correctly identified as such. It should be remarked that the data generated from a potentiometric E-tongue usually show linear behavior, and so, in this work only multivariate linear models were applied. Also, it has been showed that potentiometric data could be satisfactorily using modeled linear models and, in some cases with a superior performance compared to other statistical regression strategies such as partial least squares (PLS) and principal component regression (PCR) [32].

The statistical analysis was carried out using the open source statistical program R (version 2.15.1) and applying

the Subselect [29, 33] and MASS [34] packages at a significance level of 5%.

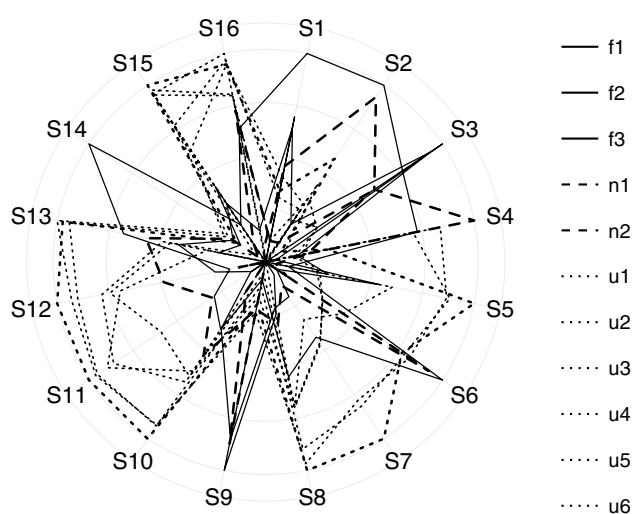
## Results and discussion

### Potentiometric signal profiles

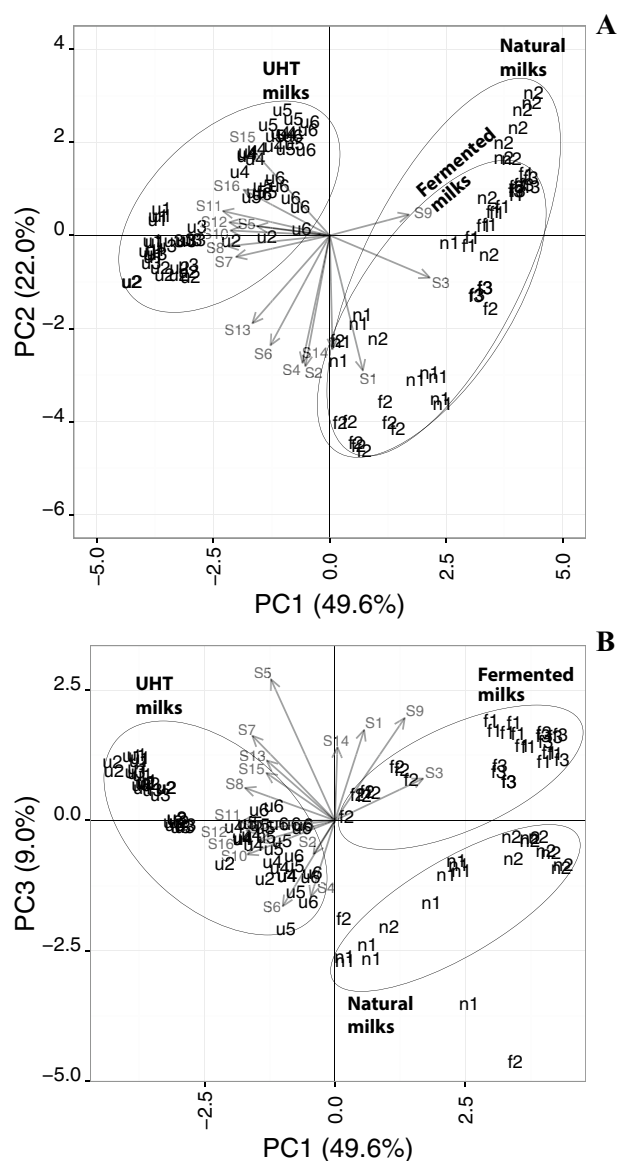
All samples were analyzed 10 times by the E-tongue in random order providing a potentiometric signals data set (110 assays  $\times$  16 signal sensors). The mean signal profiles (centered and scaled) obtained from E-tongue analysis for each dairy sample is shown in Fig. 3, using a radar plot. This figure shows the variability of the signal profiles recorded by each sensor of E-tongue for the dairy samples evaluated as well as within each type of milk used in their production. These variabilities could be attributed to the different chemical composition of the dairy samples. To confirm this hypothesis, PCA was performed using E-tongue signals data matrix in order to verify the spatial distribution of all assays.

### Principal component analysis

The PCA applied to the E-tongue signal profiles (signals from 15 potentiometric lipid/polymeric sensor membranes) obtained during 110 analyses of 11 dairy products (11 milk type products  $\times$  10 samples  $\times$  1 electrochemical assay) showed that eight PC functions explained 96% of the total data variance. The first three PCs allowed explaining a total of 80.6% of the data variability (being the values of each PC, from the first to the third, 49.6, 22.0 and 9.0%, respectively). In Fig. 4a, the 2D spatial sample distribution considering the two first PCs is shown. The natural distribution of all assays



**Fig. 3** Radar plot of the average responses obtained with the potentiometric sensor array for each milk sample



**Fig. 4** Principal component analysis of the milk samples resulted from E-tongue signals profile: **a** PC2 versus PC1; **b** PC3 versus PC1. Abbreviations: *n* natural milk, *f* fermented milk, *u* UHT milk

(analysis of the 10 samples of each milk type) showed that the replicates of each dairy product are located close to each other (showing acceptable precision) with the exception of dairy products n1, n2, and f2, for which there was a wider distribution.

Overall, the samples of the dairy products evaluated can be grouped into two main groups; the first group (negative quadrant of the PC1) included samples from dairy products produced with UHT milk; and the second group (positive quadrant of the PC1) formed by the samples of dairy products produced with natural or fermented milk. The observed unsupervised samples split could be tentatively attributed to the different milk types characteristics such fat and sugar

contents (natural and fermented milks had, in general, lower fat and sugar contents compared to UHT milks). In fact, it is known that several compound families may contribute to the milk discrimination by an electronic tongue such as minerals, sugars, proteins and fats [1]. In each of these main groups, it also can be considered two smaller groups. In the group of UHT milk, it was possible to visualize that a sub-group containing samples from dairy products u1, u2 and u3, and a smaller second sub-group formed by samples of dairy products u4, u5, and u6. For the natural and fermented milk group, two smaller sub-groups can also be considered but with a mixture of milk types. In Fig. 4b, the inclusion of PC3 allowed the separation between natural and fermented milk. In conclusion, the first three PCs (using the full sensor array, i.e., the signal profiles gathered from the 15 sensors comprised in the potentiometric E-tongue) enabled to split the eleven dairy products (including the replicate samples) according to the type of milk used in their production: natural, fermented or UHT milks. Furthermore, Fig. 4 also showed that the variability found in the analytical results cannot be directly explained by the brand or type of flavor incorporated in the dairy product. Indeed, the PCA results allowed concluding that, within each group, there is an intrinsic variability between replicate samples of the same dairy product but more evident among groups. So, these results are indicative that the signal profiles recorded with the E-tongue contain representative information that could be further used to classify dairy samples according to the three types of milk (natural, fermented and UHT milk) using a supervised multivariate statistical technique.

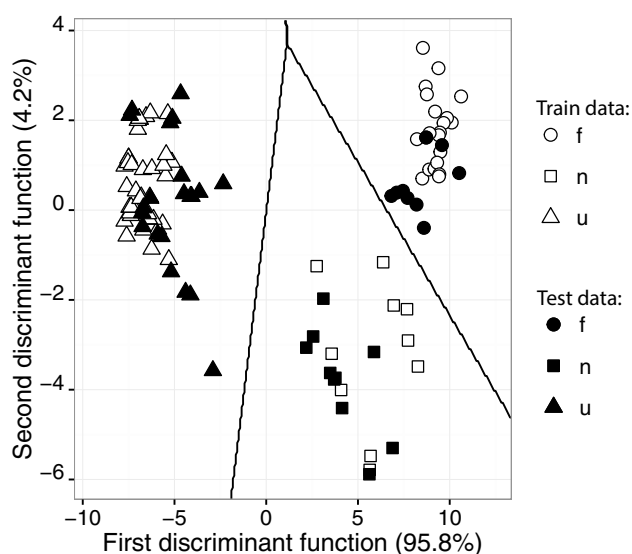
### Linear discriminant analysis

In order to verify the E-tongue capability to correctly classify samples from dairy products according to milk type, an LDA was applied coupled with the SA algorithm, which is a meta-heuristic variable selection algorithm. For this, replicate samples of the dairy products were randomly split into two groups. The training dataset was used for establishing the E-tongue-LDA-SA model, being the predictive performance evaluated by applying a repeated K-fold-CV technique (10 repeats  $\times$  10 folds). The tenfolds were randomly formed and allowed guaranteeing that 10% of the data of the training dataset was used for internal-validation purposes. With this procedure, the occurrence of possible overfitting issues is minimized and, consequently, overoptimistic correct classification performances, usually reported for LOO-CV, may be overcome. The test dataset, included data of all the replicate samples regarding 4 dairy products representative of the three milk groups considered. These dairy products were selected randomly and included replicate samples of natural milk (dairy product number 2; n2), fermented milk (dairy product number 4; f4) and UHT

milk (dairy product number 7 and 10; u7 and u10). Also, as already referred, to simplify and increase the accuracy of the predictive E-tongue-LDA model, the SA algorithm was applied to select the best subset of independent variables (with the minimum number of sensors) that gave the best prediction performance (i.e., the highest correct classification rate for the repeated K-fold-CV procedure), by removing sensors that could enhance noise issues due to multicollinearity issues. Prior to LDA modelling, the sensors signal matrix was scaled and centered.

The results obtained for the training dataset showed that the potentiometric E-tongue could discriminate dairy products according to the 3 pre-established milk groups (natural, fermented and UHT milks used in the dairy products production) using the best E-tongue-LDA-SA model established based on the predictive performance achieved for the repeated K-fold-CV procedure (i.e., internal-validation evaluation). As previously discussed, this discrimination capability could be tentatively attributed to the differences on fat and sugar contents of the milk types, which may be responsible of the different potentiometric signal profiles gathered by the E-tongue during milk analysis. The best E-tongue-LDA-SA model had two significant discriminant functions that explained 100% of the original data variability (95.8 and 4.2%, respectively) and was based on the signal profiles recorded by 4 E-tongue sensors selected by the SA algorithm (S1, S3, S13, and S15). This model allowed 100% of sensitivity and specificity for both original grouped data and for the repeated K-fold-CV, which reflects the average predictive performance achieved for the 100 internal-validation datasets randomly formed (10 folds  $\times$  10 repeats), containing each 10% of the training data (i.e., 40 replicate samples from 4 dairy products produced with the 3 different milk types studied). The predictive performance of the selected E-tongue-LDA-SA model was further verified using the test group data. The results also showed that 100% of correct classifications (corresponding to sensitivity and specificity of 100%), were obtained for the external validation procedure, demonstrating the predictive robustness of the selected electrochemical classification based model. The satisfactory classification performance achieved is clearly evidenced by Fig. 5, where the two linear discriminant functions of the best E-tongue-LDA-SA model are shown allowing to visualize the contribution of the 4 selected sensors to the complete separation of the dairy product according to the 3 different milk groups for the training dataset (different shapes of line markers) as well as for the test dataset (different shapes of solid markers). The lines of the decision boundaries for groups classification (probabilities of class membership with  $\alpha=0.05$ ) are also represented in Fig. 5.

The overall satisfactory results clearly demonstrated the potential use of the E-tongue as a practical analytical tool for correctly classifying dairy product according to the type



**Fig. 5** E-tongue's classification of dairy samples according to the milk type (natural, fermented and UHT milk) using an LDA model. Abbreviations: *n* natural milk, *f* fermented milk, *u* UHT milk

of milk used in the production, regardless the intrinsic samples' variability related to the brand and the presence or absence of flavor additives. This qualitative potential can be tentatively attributed to the representative potentiometric fingerprints gathered by the E-tongue device, which allowed identifying the differences of organoleptic sensations and physicochemical contents of the milks studied induced by the different technological processes used in their production (Mungkarndee et al. [2]). Indeed, the capability of similar E-tongue systems, also containing lipid polymeric membranes, for differentiating pleasant and unpleasant sensory attributes of food matrices have been reported [16, 21, 22, 24–26].

## Conclusions

The reported study showed the potential application of an E-tongue analytical device, which comprised a DataLogger and a multi-sensor system both of them specifically built at lab-scale. The analytical performance of the electrochemical device was evaluated for dairy products classification according to the type of milk used in their production. The E-tongue multi-sensor tool included a commercial pH electrode and 15 lipid/polymeric membranes. The overall results showed that using the potentiometric fingerprints gathered from the 16 sensors it was possible to naturally differentiate all dairy products by type of milk, using a PCA. It was also demonstrated that using a supervised multivariate strategy with combined with a variable selection algorithm it was possible to establish a E-tongue-LDA-SA model based on

the signals of only 4 sensor lipid membranes, which had an overall sensitivity and specificity of 100% for both internal-validation (i.e., repeated K-fold-CV procedure) and external-validation procedures. Thus, it could be concluded that the E-tongue based strategy represented a feasible and accurate classification tool that could be used to identify the type of milk (natural, fermented or UHT milks) used in the production of different dairy products, and so its use could be foreseen at the quality and control level in dairy industries.

**Acknowledgements** This study was funded in part by the Ministry of Research, Technology and Higher Education, the Republic of Indonesia, Project 001/SP2H/LT/DRPM/IV/2017. The authors also thank the Directorate General of Islamic Education and Instituto Politécnico de Bragança, Portugal, for their support throughout the completion of this work. This work was financially supported by Project POCI-01-0145-FEDER-006984 – Associate Laboratory LSRE-LCM, Project UID/BIO/04469/2013 – CEB and strategic project PEst-OE/AGR/UI0690/2014 – CIMO all funded by FEDER - Fundo Europeu de Desenvolvimento Regional through COMPETE2020 - Programa Operacional Competitividade e Internacionalização (POCI) – and by national funds through FCT - Fundação para a Ciência e a Tecnologia, Portugal.

## References

1. M. Bougrini, K. Tahri, Z. Haddi, N. El Bari, E. Llobet, N. Jaffrezic-Renault, B. Bouchikhi, Aging time and brand determination of pasteurized milk using a multisensor e-nose combined with a voltammetric e-tongue. *Mater. Sci. Eng. C* **45**, 348–358 (2014)
2. R. Mungkarndee, I. Techakriengkrai, G. Tumcharern, M. Sukwattanasinitt, Fluorescence sensor array for identification of commercial milk samples according to their thermal treatments. *Food Chem.* **197**, 198–204 (2016)
3. P. Ciosek, K. Brudzewski, W. Wróblewski, Milk classification by means of an electronic tongue and support vector machine neural network. *Meas. Sci. Technol.* **17**, 1379–1384 (2006)
4. B. Roza-Delgado, A. Garrido-Varo, A. Soldado, A.G. Arrojo, M.C. Valdés, F. Maroto, D. Pérez-Marín, Matching portable NIRS instruments for in situ monitoring indicators of milk composition. *Food Cont.* **76**, 74–81 (2017)
5. F. Winquist, S. Holmin, C. Krantz-Rülcker, P. Wide, I. Lundström, A hybrid electronic tongue. *Anal. Chim. Acta* **406**, 147–157 (2000)
6. F. Winquist, C. Krantz-Rülcker, P. Wided, I. Lundström, Monitoring of freshness of milk by an electronic tongue on the basis of voltammetry. *Meas. Sci. Technol.* **9**, 1937–1946 (1998)
7. Z. Wei, J. Wang, X. Zhang, Monitoring of quality and storage time of unsealed pasteurized milk by voltammetric electronic tongue. *Electrochim. Acta* **88**, 231–239 (2013)
8. F. Winquist, R. Bjorklund, C. Krantz-Rülcker, I. Lundström, K. Östergren, T. Skoglund, An electronic tongue in the dairy industry. *Sens. Actuators B* **111–112**, 299–304 (2005)
9. Y. Yu, H. Zhao, R. Yang, G. Dong, L. Li, J. Yang, T. Jin, W. Zhang, Y. Liu, Pure milk brands classification by means of a voltammetric electronic tongue and multivariate analysis. *Int. J. Electrochem. Sci.* **10**, 4381–4392 (2015)
10. L. Li, Y. Yu, J. Yang, R. Yang, G. Dong, T. Jin, Voltammetric electronic tongue for the qualitative analysis of milk adulterated with urea combined with multi-way data analysis. *Int. J. Electrochem. Sci.* **10**, 5970–5980 (2015)



11. Y. Yu, H. Zhao, G. Dong, R. Yang, L. Li, Y. Liu, H. Wu, W. Zhang, Discrimination of milk adulterated with urea using voltammetric electronic tongue coupled with PCA-LSSVM. *Int. J. Electrochem. Sci.* **10**, 10119–10131 (2015)
12. M.Y.M. Sim, T.J. Shya, M.N. Ahmad, A.Y.M. Shakaff, A.R. Othman, M.S. Hitam, Monitoring of milk quality with disposable taste sensor. *Sensors* **3**, 340–349 (2003)
13. P. Ciosek, Z. Brzózka, W. Wróblewski, Electronic tongue for flow-through analysis of beverages. *Sens. Actuators B* **118**, 454–460 (2006)
14. P. Ciosek, W. Wróblewski, Performance of selective and partially selective sensors in the recognition of beverages. *Talanta* **71**, 738–746 (2007)
15. P. Ciosek, W. Wróblewski, Miniaturized electronic tongue with an integrated reference microelectrode for the recognition of milk samples. *Talanta* **76**, 548–556 (2008)
16. L.A. Dias, A.M. Peres, A.C.A. Veloso, F.S. Reis, M. Vilas-Boas, A.A.S.C. Machado, An electronic tongue taste evaluation: identification of goat milk adulteration with bovine milk. *Sens. Actuators B* **136**, 209–217 (2009)
17. M. Hruškar, N. Major, M. Krpan, I.P. Krbavčić, G. Šarić, K. Marković, N. Vahčić, Evaluation of milk and dairy products by electronic tongue. *Mljekarstvo* **59**, 193–200 (2009)
18. M. Hruškar, N. Major, M. Krpan, Application of a potentiometric sensor array as a technique in sensory analysis. *Talanta* **81**, 398–403 (2010)
19. M. Hruškar, N. Major, M. Krpan, N. Vahčić, Simultaneous determination of fermented milk aroma compounds by a potentiometric sensor array. *Talanta* **82**, 1292–1297 (2010)
20. I. Tazi, A. Choiriyah, D. Siswanta, K. Triyana, Detection of taste change of bovine and goat milk in room ambient using electronic tongue. *Indones. J. Chem.* **17**, 422–430 (2017)
21. Í Marx, N. Rodrigues, L.G. Dias, A.C.A. Veloso, J.A. Pereira, D.A. Drunkler, A.M. Peres, Sensory classification of table olives using an electronic tongue: analysis of aqueous pastes and brines. *Talanta* **162**, 98–106 (2017)
22. Í.M.G. Marx, N. Rodrigues, L.G. Dias, A.C.A. Veloso, J.A. Pereira, D.A. Drunkler, A.M. Peres, Quantification of table olives' acid, bitter and salty tastes using potentiometric electronic tongue fingerprints. *LWT - Food Sci. Technol.* **79**, 394–401 (2017)
23. S. Slim, N. Rodrigues, L.G. Dias, A.C.A. Veloso, J.A. Pereira, S. Oueslati, A.M. Peres, Application of an electronic tongue for Tunisian olive oils' classification according to olive cultivar or physicochemical parameters. *Eur. Food Res. Technol.* **243**, 1459–1470 (2017)
24. A.C.A. Veloso, L.G. Dias, N. Rodrigues, J.A. Pereira, A.M. Peres, Sensory intensity assessment of olive oils using an electronic tongue. *Talanta* **146**, 585–593 (2016)
25. A.C.A. Veloso, L.M. Silva, N. Rodrigues, L.P.G. Rebello, L.G. Dias, J.A. Pereira, A.M. Peres, Perception of olive oils sensory defects using a potentiometric taste device. *Talanta* **176**, 610–618 (2018)
26. U. Harzalli, N. Rodrigues, A.C.A. Veloso, L.G. Dias, J.A. Pereira, S. Oueslati, A.M. Peres, A taste sensor device for unmasking admixing of rancid or winey-vinegary olive oil to extra virgin olive oil. *Comp. Electron. Agric.* **144**, 222–231 (2018)
27. E.L. Chuayana Jr., C.V. Ponce, C.MaR.B. Rivera, E.C. Cabrera, Antimicrobial activity of probiotics from milk products. *Phil. J. Microbiol. Infect. Dis.* **32**, 71–74 (2003)
28. R. Talon, D. Walter, C. Viallon, J.L. Berdagué, Prediction of *Streptococcus salivarius* subsp. *thermophilus* and *Lactobacillus delbrueckii* subsp. *bulgaricus* populations in yoghurt by Curie point pyrolysis-mass spectrometry. *J. Microbiol. Method* **48**, 271–279 (2002)
29. D. Bertsimas, J. Tsitsiklis, Simulated annealing. *Stat. Sci.* **8**, 10–15 (1993)
30. J. Cadima, J.O. Cerdeira, M. Minhoto, Computational aspects of algorithms for variable selection in the context of principal components. *Comp. Stat. Data Anal.* **47**, 225–236 (2004)
31. S. Kirkpatrick, C.D. Gelatt, M.P. Vecchi, Optimization by simulated annealing. *Science* **220**, 671–680 (1983)
32. N. Rodrigues, L.G. Dias, A.C.A. Veloso, J.A. Pereira, A.M. Peres, Monitoring olive oils quality and oxidative resistance during storage using an electronic tongue. *LWT Food Sci. Technol.* **73**, 683–692 (2016)
33. J. Cadima, J.O. Cerdeira, P.D. Silva, M. Minhoto, The subselect R package, <http://cran.rproject.org/web/packages/subselect/vignettes/subselect.pdf>. Accessed 15 February 2016
34. W.N. Venables, B.D. Ripley, *Modern Applied Statistics with S (Statistics and Computing)*, 4th edn. (Springer, New York, 2002)