VITAE, REVISTA DE LA FACULTAD DE CIENCIAS FARMACÉUTICAS Y ALIMENTARIAS
ISSN 0121-4004 / ISSNe 2145-2660. Volumen 26 número 2, año 2019
Universidad de Antioquia, Medellín, Colombia. Págs. 94-103
DOI: http://dx.doi.org/10.17533/udea.vitae.v26n2a05

AUTOMATIC CLASSIFICATION OF WATER SAMPLES USING AN OPTIMIZED SVM MODEL APPLIED TO CYCLIC VOLTAMMETRY SIGNALS

CLASIFICACIÓN AUTOMÁTICA DE MUESTRAS DE AGUA USANDO UN MODELO SVM OPTIMIZADO APLICADO A SEÑALES DE VOLTAMETRÍA CÍCLICA

Hugo ROMERO BONILLA[1*], Iván RAMÍREZ-MORALES[2], Cinthia ROMERO FLORES[3]

Received: March 29 of 2019. Approved: August 28 of 2019.

ABSTRACT

Background: concern about the quality of the water for human consumption has become widespread among the population. The taste and some problems associated with drinking water have been the cause of increased demand for bottled water. Due to this, day to day, a large number of companies has manifested their interest in the production of bottled water. **Objective:** to evaluate a novel automatic classification model that differentiates bottled water from tap water. **Methods:** the voltammetric technique consisted of three electrode setup. The output current has been considered for data analysis. From the results of grid search, six pairs of values were pre-selected for the parameters of σ and C whose results were similar. High values of accuracy, specificity and sensitivity were achieved in test dataset. The final decision was made after performing an ANOVA test of 100 repetitions of 5-fold cross-validation, 3000 models were evaluated with the parameter combinations described above for the SVM. **Results:** the oxidation and reduction peaks of the water samples have been observed to be prominent. Absolute values of current (I) increased in the case of public water samples, possibly due to the largest concentration of chloride ions which have higher contributions to the conductivity. 5-fold cross-validation test mean specificity resulted in C parameters values greater than 0 and between 0 and 30; a σ value greater than 10 and between 0 and 15 were found for tap water and bottled water, respectively. The combination ($\sigma = 10$, $C = 30$) presented best results in accuracy 0.988 ± 0.037, specificity 0.973 ± 0.085 and sensitivity 1 ± 0.09. **Conclusions**: results of this research work have shown that voltammograms for values of current increased for tap water samples, 9.94e-6μA, compared to 7.99e-6μA due to higher chloride ions concentration in the former. The parameters combination ($\sigma = 10$, $C = 20$) was selected as optimal parameters since there were no significant difference between this and the former.

Keywords: Electronic tongue, water quality, authenticity, machine learning, voltammetry.

RESUMEN

Antecedentes: en la población hay una preocupación generalizada por la calidad del agua de consumo humano. El sabor y algunos problemas asociados con el agua potable han sido la causa del incremento en la demanda del agua embotellada. Debido a esto, un gran número de empresas han manifestado su interés en la producción de agua en botella. **Objetivo:** evaluar un nuevo modelo de clasificación automática que

[1] Research Group Electroanalytical Applications, Electroanalytical and Bioenergy Laboratory, Faculty of Chemistry, Universidad Técnica de Machala. El Oro, Ecuador.

[2] DINTA applied technologies research group, Universidad Técnica de Machala. El Oro, Ecuador.

[3] Faculty of Chemistry, Universidad Técnica de Machala. El Oro, Ecuador.

[*] Author of correspondence: hromero@utmachala.edu.ec

diferencie el agua embotellada del agua del grifo. **Metodología:** la técnica de voltametría consistió en la configuración de tres electrodos. Para el análisis de datos se consideró la corriente de salida y de los resultados de la búsqueda de cuadrícula y se seleccionaron seis pares de valores para los parámetros de σ y C, cuyos resultados fueron similares. Se lograron altos valores de precisión, especificidad y sensibilidad en el conjunto de datos de prueba. La decisión final se tomó después de realizar una prueba ANOVA de 100 repeticiones de validación cruzada de 5 veces y se evaluaron 3000 modelos con las combinaciones de parámetros descritas anteriormente para el SVM. **Resultados:** se observó que los picos de oxidación y reducción de las muestras de agua son prominentes. Los valores absolutos de corriente (I) aumentaron en el caso de muestras de agua pública, posiblemente debido a la mayor concentración de iones de cloruro que tienen una mayor contribución a la conductividad. La especificidad media de la prueba de validación cruzada 5 veces dio como resultado valores de parámetros C mayores que 0 y entre 0 y 30; se encontró un valor σ mayor que 10 y entre 0 y 15 para el agua de red pública y el agua embotellada, respectivamente. La combinación (σ = 10, C = 30) presentó los mejores resultados en precisión 0,988 ± 0,037, especificidad 0,973 ± 0,085 y sensibilidad 1 ± 0,09. **Conclusiones:** los resultados de este trabajo demostraron que los voltamogramas para valores de corriente, aumentaron para muestras de agua corriente, 9,94e-6 μA, en comparación con 7,99e-6 μA debido a una mayor concentración de iones de cloruro en el primer caso. La combinación de parámetros (σ = 10, C = 20) se seleccionó como parámetros óptimos, ya que no mostró diferencias significativas entre éste y el primer caso.

Palabras clave: Lengua electrónica, calidad del agua, autenticidad, aprendizaje automático, voltametría.

INTRODUCTION

Concern about the quality of the water for human consumption has become widespread among the population. The taste and some problems associated with drinking water have been the cause of increased demand for bottled water. Due to this, day to day, a large number of companies has manifested their interest in the production of bottled water. The consumption of bottled water has been increasing consistently over the last decade, even in countries where tap water quality is considered of excellent quality (1).

Traditional methodologies are commonly used to detect compounds and its characteristics. These methodologies show good precision, accuracy and reliability. However, this technologies are often destructive, time consuming, and require expensive equipments. To overcome above drawbacks, electronic tongues have emerged as rapid and ease to use tools very promising for evaluation of food quality (2).

With regard to water intended for human consumption, the problems associated with global warming leading to regional changes in climate and water availability are seriously affecting sustainability of supplies as well as seriously impacting on quality. Advances in chemical and microbial analysis have revealed that water contains many new contaminants that were previously undetectable or unknown, constantly presenting water utilities and regulators with new challenges (3).

As an alternative, the instrumental methods employing high end analytical instruments like high performance liquid chromatography, gas chromatography (4) and capillary electrophoresis (5) are prohibitively expensive and require skilled manpower (6). Nowadays, consumers are paying great attention to the characteristics of food such as smell, taste, and appearance. This motivates scientists to imitate human senses using devices known as electronic senses. These include electronic noses, electronic tongues, and computer vision. Thanks to the utilization of various sensors and methods of signal analysis, artificial senses are widely applied in food analysis for process monitoring and determining the quality and authenticity of foods (7).

Electronic tongues are sensor arrays combined with voltammetric methods and multi-variable analysis which are tested for large household applications (8, 9). The existent problem is the complex data analysis. Cyclic voltammetry (CV) is well-known analytical method and deliver more specific information (10). This technology is becoming extremely important, proof of this is the large number of bibliographic review works that have been published on this subject. Some of them are general and cover the different types of electronic tongues and their applications. Others, although they give an overview of this technology,

focus on some specific application such as the food industry, the pharmaceutical industry or environmental applications (11).

The most common application of electronic voltammetric tongues are beverages, as is the case of the extensive number of published works related to wine (11). Other emerging approach from voltammetric electronic tongue (VE-tongue) is the analytical characterization and geographical classification of honey (12). Water analysis is another of the applications in which the electronic voltammetric tongues are having a great development (13). For example, Tønning et al., (14) used 8 sensors created by serigraphy and modified with enzymes to classify waters with different degrees of contamination. For this purpose, the time evolution of the current is measured by applying voltage steps to the electrodes. The resulting data are treated with PCA, achieving clear discrimination between the samples of each category.

Another work related to the analysis of water quality is described by Gutés et al., (15). In this case, wastewater from paper mills is measured in a continuous flow by means of the voltammetric tongue with Au, Pt and Rh electrodes.

There exists a need for better on-line monitoring of water systems given that existing laboratory-based methods are too slow to develop operational response and do not provide a level of public health protection in real time. There is a clear need to be able to rapidly detect (and respond) to instances of accidental (or deliberate) contamination, due to the potentially severe consequences to human health (16).

The use of support vector machine (SVM) has been reported to work efficiently with VE-tonge in several chemometric applications (17, 18). In this sense, the aim of present research was to develop a model based on support vector machines algorithms, using data obtained by cyclic voltammetry in order to discriminate water from the public network and drinkable water quality.

MATERIALS AND METHODS

Sample collection

In this work, a total of 60 different samples of drinkable water have been considered, 30 of them from the public tap in different places around the south of Ecuador. The remaining 30 samples, were obtained from different production batches of bottled water of a recognized brand. All the samples have been subjected for evaluation.

The voltammetric technique involves a three electrode setup. The potential was applied between the reference electrode and the working electrode and the current was measured between the working electrode and the counter electrode. It has been investigated with working electrode made of glassy carbon, an Ag/AgCl reference electrode (saturated KCl, Gamry Instruments Inc.) and a high purity platinum foil was used as a counter electrode. A PAR 273-A Potentiostat, was used for electrochemical measurements.

The output current has been considered for data analysis. The number of data points generated from glassy carbon electrode are 2000. The scan rate was 0.2 V/s. The resulting current has been recorded and considered for data analysis. The oxidation and reduction peaks of the water samples have been observed to be prominent (Figure 1). These peaks are due to the combined redox potentials of the constituents of water. The waveform is applied with the range from 1.0V to - 1.0V (19).

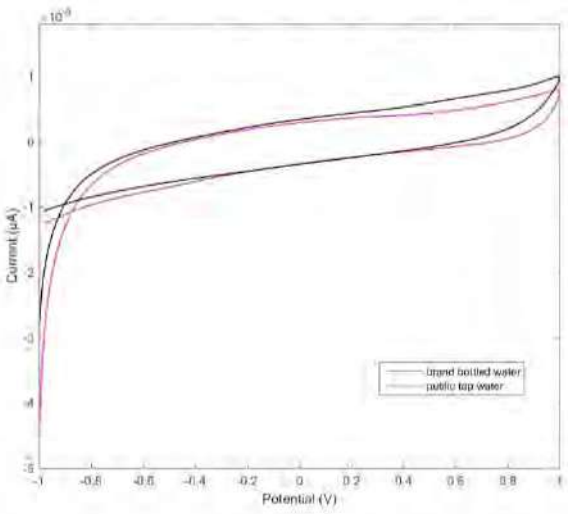

Figure 1. Mean of cyclic voltammograms of the public tap water and brand bottled water samples.

Automatic feature (potential range) selection

In chemometrics, due to the large amount of information provided by the instruments, it is necessary to substantially reduce the number of variables necessary for the construction of classification and calibration models (20). In the last decade, feature selection (FS) in the construction

of models has gone from illustrative examples in terms of their operation, to become a requirement, particularly by the nature of the problems with high dimensionality such as microarray analysis and spectral analysis (21).

The selection of features contrasts with other dimensionality reduction techniques, such as principal component analysis (PCA) since the former does not alter the original representation of the variables, but simply consists of selecting a subset of the best characteristics, preserving their nature original, which allows them to be easily interpreted by a field expert (21). Considering that many of the pattern recognition techniques were not designed in their origins to deal with large amounts of irrelevant information, the application of FS techniques has become a necessity in many applications today (22).

Pre-apply a FS technique, prevent overfitting of the model, improve performance and reduce computation time and obtain a deeper understanding of the data, this adds complexity to the modeling task, since instead of just optimizing the parameters of the model, now we have to find the optimal characteristics that define the model, and in the case of regression it is used to look for the variables that maximize the fit of the model (21,23,24).

A group of techniques commonly used for FS, are filters that evaluate the relevance of a characteristic, analyzing the intrinsic properties of the data, in general a score of relevance is calculated and those characteristics with lower score are eliminated (21,25), there are two types: univariate and multivariate (26).

The univariate filter is a simple but efficient paradigm, since the output ranking is easy to understand, it is usually defined a threshold method to select those that meet a condition above or below it. Filters work independently of the model and use the intrinsic properties of the data (25). Univariate filters have the advantage that they are fast, scalable and independent of the classification / regression technique used, however they have the disadvantage that they ignore the dependency and correlation between characteristics (27). The univariate filter method can be applied using a t-test (28), F-test (29) or the Wilcoxon rank sum test (25). Calculates a p-value that represents the statistical significance of each variable in the model, so the variables are ordered depending on their p-value (21).

Support vector machines algorithms

The supervised machine learning algorithms are used when there is sufficient knowledge about the desired results, in order to build a strong model capable of correctly assigning the class to which a new data entry belongs (30).

A trained algorithm is able to transfer the learned dependency between the input patterns and the expected results in the new data sets. The performance of the classifier can be evaluated by the proportion of correctly classified patterns in the test, ie the data that were not used during the training (31).

Support vector machines (SVM) are new supervised machine learning algorithms that are used to classify data sets into two different classes, separated by a hyperplane (30).

The SVM can be used either in regression or classification models, since its operation is based on learning from a set of training samples whose classes are known (32, 33). Developed the foundations of the SVM, in a study on the theories of statistical learning, later, Boser et al., (1992) proposed a method to create non-linear classifiers (34). Current SVM standard models were proposed by (35).

The SVM models are intended to obtain models with little structural risk of error with respect to future data, originally designed to solve binary classification problems but now its application has been extended to regression, multiclassification, clustering and other tasks (32). This technique is intended to find an optimal hyperplane capable of distributing the data in the classes to which they belong. Intuitively, it seems obvious to conclude that when faced with a linear classification problem there is a high probability of obtaining several solutions that correctly classify the data.

The optimal hyperplane is used to separate the two classes can be defined from a small amount of data from the training set called support vectors, which determine the optimal margin of separation (35). Figure 2 illustrates the aforementioned concepts.

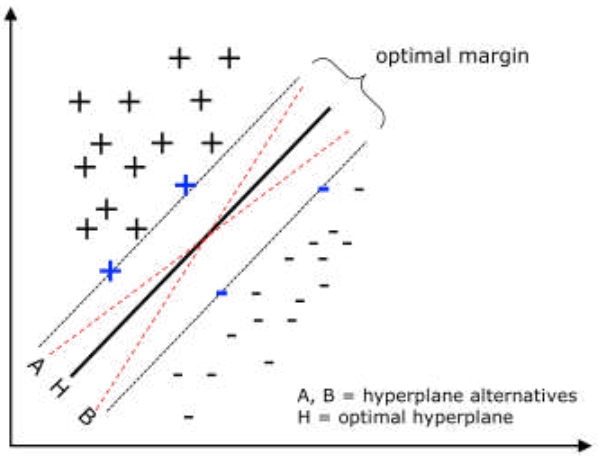

Figure 2. A problem separable in a two-dimensional space. Support vectors (highlighted in blue) define the margin of greatest separation between classes.

The choice of the best hyperplane was resolved in 1965 (36) with the criterion that the optimal hyperplane is defined as the linear decision function with the maximum margin between the vectors of the two classes. However, in most problems, the data is not linearly separable and the use of strategies such as the identification of other separation dimensions is required. Core functions are used to transform the original multidimensional space into another one, where the classes are linearly separable. In practice, support vector machines are trained using different cores to select the one that has the best performance for the problem posed (37).

In the present work, preliminary tests were carried out by the trial-error method (38), to determine the most suitable calculation kernel. A radial basis function - RBF (Gauss), was chosen as the kernel of the model, to perform the exhaustive evaluation. Normally polynomial kernels and RBF are among the most used; the latter has a sigma parameter (σ) that can be tuned to adjust the size of the kernel (39). Preliminary tests were carried out to select the range of best sigma values, which was between one and six; This range was used for the exhaustive evaluation.

SVM has a compensation parameter of C, which can be modified and affects the quality of the classification, since it determines the severity with any classification error should be penalized; in general, very high values of C can lead to problems of overfitting, which reduces the ability to generalize SVM (37). In order to evaluate this parameter without overfitting the classifier, values below 0.25 were selected.

Model optimization

A 5-fold cross-validation was applied in order to evaluate how the model would perform in a set of independent data. This also reduces the risk of overfitting the model (40). By partitioning the data in this way, the training and test subsets of each fold contain representative samples of the two classes in a random and stratified manner. The process of partitioning into 5-fold cross-validation, divides the data into k subsets; one of them is used as a subset of the test and the others (k-1) are used as training subsets (37). This process is repeated for k folds, with each of the possible subsets. In this work 100 repetitions of 5-fold cross-validation, were performed, so that for this study, 1.25 million models based on different configurations of the SVM parameters were evaluated in order to determine the optimal configuration.

Finding the best combination of parameters of an SVM is key in the construction of a prediction model to be highly accurate and stable (41, 42). The parameters of the RFB kernel are adjustable in the SVMs in order to control the complexity of the resulting hypothesis, and in this way avoid the overfitting of the model (43, 44). The combination of parameters C and σ, determines the performance of the model, to find the optimal combination can be used several approaches (44, 45).

In the present work, a grid search approach was applied for parameter optimization. This method is widely used in machine learning techniques and uses a brute force approach in which all combinations of parameters in a range are tested (46), in order to find the best combination, based on the performance measures, explained above.

Classification performance analysis

The accuracy is usually the only metric performance, which is used to evaluate the performance of machine learning techniques; this precision value is a statistical measure that is used to determine if a binary classification test (true or false) is able to identify or exclude a condition correctly (47). The accuracy is usually the only metric performance, which is used to evaluate the performance of machine learning techniques; this precision value is a statistical measure that is used to determine if a binary classification test (true or false) is able to identify or exclude a condition correctly (48). However, in order to compare the rate of false positives and false negatives, other

performance measures are used (49). The specificity in this context is the ability to detect water samples from the public network as false and the sensitivity is the ability to detect bottled water samples as true.

To select the optimal configuration of the models, Analysis of variance (ANOVA) and multiple range tests (MRT) were performed with the Tukey Honest Significant Difference (HSD) method for a value of $p < 0.01$. In each step of the methodology, a selection of the parameters that provided the best performance was made. In order to improve the ability of the model to generalize, the metrics were calculated from the confusion matrix of the test set, that is, data different from those used for the training, reducing the possibility of overtraining.

Results and Discussion

In this work a methodology is presented to optimize a classification model based on SVM, for the authenticity evaluation of bottled water. The voltammograms were labeled according to the type of sample in order to perform the supervised training of the classification algorithm. Figure 3 shows the voltammograms of the bottled water samples and the tap water samples. In this figure you can see the voltammograms of public water was combined to compare with brand bottled water samples. The cyclic voltammograms of public water represent similar curves; however, the redox curves obtained in negative potential range (-1.0 V ~ 0.0V) gradually shifted to the direction of lower current density (Figure 1). This is because, the amount of minerals in public water affects its ability to dissolve oxygen and therefore decrease the cathodic peaks.

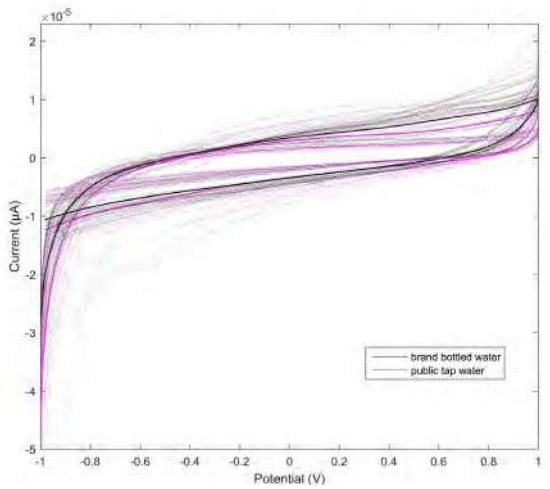

Figure 3. Cyclic voltammetry graph, black lines belongs to brand bottled water samples, magenta lines belongs to public potable water samples.

A univariate filter algorithm (28) was used for the selection of the potential ranges in which the differences between the voltammograms of the two groups evaluated are better evidenced. In Figure 4, the result of applying the univariate filter for the automatic selection of the potential range in which the differences between voltages of different groups are accentuated, which is from -1V to -0.94V and from 0.45V to 1V.

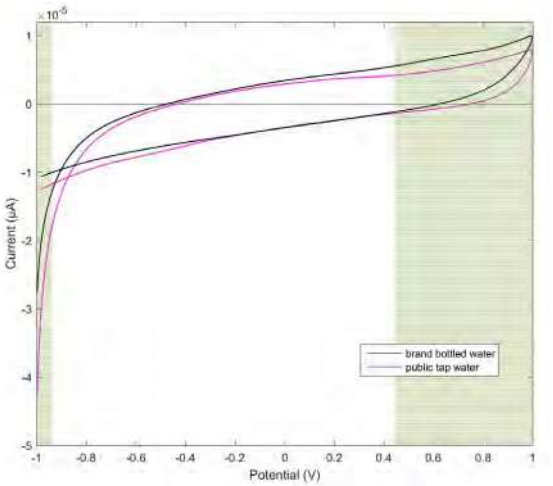

Figure 4. Univariate filter selected potential range (1).

Voltammetric electronic tongue signals and support vector machines for pattern recognition are used for classifying public tap water and brand bottled water samples in order to determine its authenticity. The resulting voltammograms of tap water and bottled water, are in good agreement with those reported in the literature for tap water (50) and for acidulated water (51). The water molecules presents in drinkable water represents a redox property different from tap water. If we analyze the cyclic voltammograms obtained in the potential range between 0.7V to 1.0V, we can observe a gradual displacement of the current to lower values comparing bottled water with tap water.

Most chemical species present in tap water, results in a difference in current densities compared to those observed in samples drinkable water. This behavior is accentuated due to the differences in the currents of reduction of the dissolved oxygen that is found in greater quantity in bottled water. Meanwhile, Figure 1 shows absolute values of current (I) that increases in the case of public water samples $9.94e^{-6}$ μA compared with $7.99e^{-6}$ μA, both at 0.9916V. The reason could be the largest concentration of chloride ions in the public water samples that have a higher contributions to the conductivity (10).

In preliminary tests it was determined that the classification algorithm to be used is SVM with a radial basis function (RBF) kernel. A grid search algorithm was used to select the best combination of the parameters σ and C of the SVM, a range between 0 and 30 with intervals of 1 was defined for both parameters. The results of the average of the 100 repetitions of 5-fold cross-validation in the test set can be reviewed in figures 5, 6 and 7. Figure 5 shows the mean accuracy in the test set of 100 repetitions of the 5-fold cross-validation, it can be seen that with a C parameter value greater than 5 and a value of σ between 6 and 18, good results are achieved accuracy results.

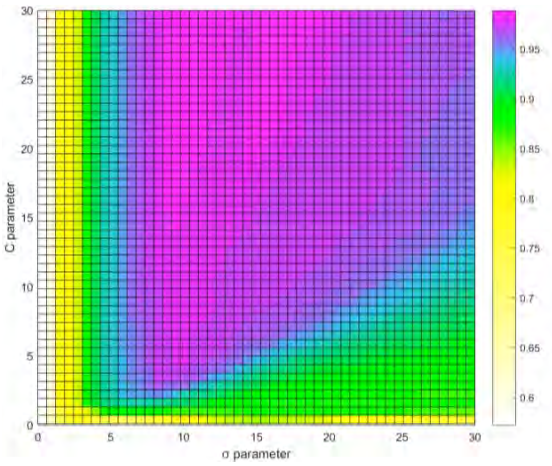

Figure 5. Grid search of accuracy according the values of σ and *C* parameter.

Figure 6 shows the mean specificity in the test set of 100 repetitions of the 5-fold cross-validation, it can be seen that with a value of the C parameter greater than 0 and a value of σ greater than 10, good results are achieved of specificity.

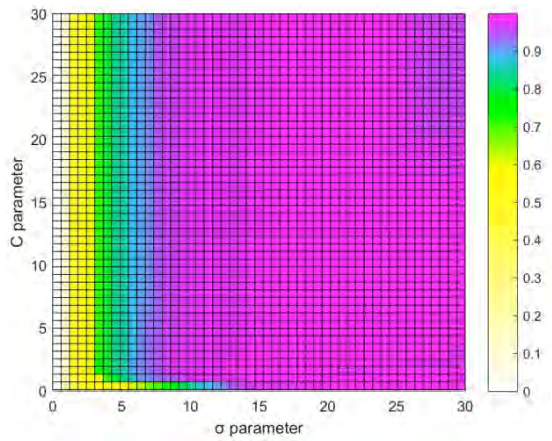

Figure 6. Grid search of specificity according the values of σ and *C* parameter.

Figure 7, shows the mean sensitivity in the test set of 100 repetitions of the 5-fold cross-validation, it can be seen that with a value of the C parameter between 0 and 30 and a value of σ between 0 and 15, it is reached good sensitivity results.

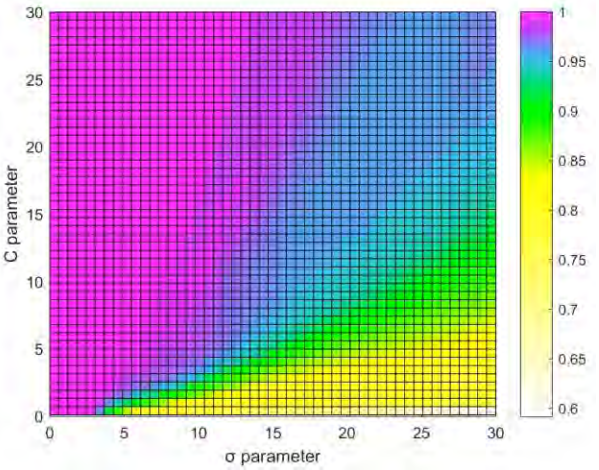

Figure 7. Grid search of sensitivity according the values of σ and *C* parameter.

From the results of grid search, six pairs of values have been pre-selected for the parameters of σ and C, (10, 15); (10, 20); (10, 30); (15, 15); (15, 20) and (15, 30) whose results are similar. The final decision was made after performing an ANOVA test of 100 repetitions of 5-fold cross-validation, 3000 models were evaluated with the parameter combinations described above for the SVM (Table 1).

Table 1. Multiple comparison of different values of σ and *C* parameter for the performance metrics assessed.

	Accuracy			Specificity			Sensitivity		
σ, *C*	mean	⋆	std dev	mean	⋆	std dev	mean	⋆	std dev
10, 15	0.978	b	0.051	0.970	a	0.088	0.985	b	0.064
10, 20	0.986	ab	0.039	0.972	a	0.085	0.997	a	0.026
10, 30	0.988	a	0.037	0.973	a	0.085	1.000	a	0.009
15, 15	0.978	b	0.054	0.984	a	0.069	0.973	c	0.081
15, 20	0.983	ab	0.050	0.984	a	0.069	0.982	b	0.071
15, 30	0.984	ab	0.047	0.984	a	0.069	0.984	b	0.065

Rows with different letters differ significantly according to Tukey's Honest Significant Difference method for a value of $p < 0.01$.

Although the combination (σ = 10, C = 30) presented the best results in accuracy 0.988 ± 0.037, specificity 0.973 ± 0.085 and sensitivity 1 ± 0.09. Finally, a combination of parameters (σ = 10, C = 20) was selected as optimal parameters, since

there was no significant difference between this and the combination (σ = 10, C = 30). According to (52) it is better to select lower values of the parameter C in order to avoid overfitting.

The univariate filter applied to the voltammograms for the automatic selection of the potential range in which the samples of both classes differ best, was from -1V to -0.94V and from 0.45V to 1V. There are no previous works that indicate the range of potential in which it is possible to differentiate samples of bottled water from any tap water samples.

The model that obtained the best performance in terms of accuracy (0.986), specificity (0.972) and sensitivity (0.997), was the one that uses the RBF kernel, with a parameter σ equal to 10, and a parameter C equal to 20. The results were obtained in the test set, which gives validity to the model given that the training was performed with different data to this set, in this way, the proposed model is able to classify new samples from the dependency learned between the patterns of entry in accordance with the raised by (31).

The optimization of the parameter σ and C in the SVM models applied to voltammograms of water samples generate an improvement in sensitivity, accuracy and specificity. however, considering that the higher the value of C, the greater the probability of overfitting (37), the authors selected values of σ equal to 10, and of C equal to 20, even though the combination that produced results with a higher mean was (σ = 10, C = 30).

CONCLUSIONS

In this paper, glassy carbon working electrode cyclic voltammetry is used for discriminating between drinkable bottled water and water from the public network, using an optimized SVM. The novelty of the proposed method is that of the model classifies automatically the water samples using the learned pattern contained in several points of voltammogram.

The cyclic voltammetry applied as a method of electrochemical analysis, shows that the water molecules present in the samples of a brand of bottled water, have redox properties different from those of tap water. The range of the potential in which the samples of bottled water and tap water differs best is from -1V to -0.94V and from 0.45V to 1V.

Our results show that it is possible to evaluate the authenticity of bottled water, by classifying cyclic voltammetric signals. It was observed that the absolute values of current (I) increases in the case of public water samples $9.94e^{-6}$ μA compared with $7.99e^{-6}$ μA, both at 0.9916 V. The reason could be the largest concentration of chloride ions in the public water samples that have higher contributions to the conductivity.

The optimal SVM model consisted of an RBF kernel, σ = 10, C = 20, whose results in Accuracy, Specificity and Sensitivity were 0.986, 0.972 and 0.997, respectively. The optimization of the parameters of the SVM is fundamental to improve the accuracy, specificity and sensitivity in the problems of classification of water samples using voltammograms.

ACKNOWLEDGEMENTS

The authors wish to thank to the Electroanalytical Applications research group and DINTA-UTMACH group for the provision of the necessary means for the development of the present investigation. Iván Ramírez-Morales also like to thank the support provided by the NVIDIA Research Scholarship Program.

This study was funded with the project number 420/2017 of the Research Center of the Technical University of Machala.

CONFLICTS OF INTEREST

The authors declare that there is no conflict of interest with the published results.

AUTHORS´CONTRIBUTIONS

Hugo Romero contributed to the experimental electroanalytical part through the analysis of the samples, both tap water and bottled water with cyclic voltammetry. In addition, I perform the electrochemical writing of the manuscript in relation to the results and discussion. Iván Ramírez Morales develop the programing, adjustment of the hyperparameters and fine tuning of the model, also collaborated in the writing of the manuscript. Cinthia Romero collaborated in the sampling and in the statistical validation of the results obtained. In addition, I write the introduction of the manuscript.

REFERENCES

1. Doria MF. Bottled water versus tap water: understanding consumers' preferences. J Water Health [Internet]. 2006 Jun;4(2):271–6. Available from: https://www.ncbi.nlm.nih.gov/pubmed/16813019
2. Bueno L, de Araujo W, Salles M, Kussuda M, Paixão T. Voltammetric Electronic Tongue for Discrimination of Milk Adulterated with Urea, Formaldehyde and Melamine. Chemosensors [Internet]. 2014 Nov 14 [cited 2018 Sep 5];2(4):251-66. Available from: DOI: org/10.3390/chemosensors2040251
3. Gray NF. Drinking Water Quality: Problems and Solutions [Internet]. Cambridge University Press; 2008. 537 p. Available from: https://market.android.com/details?id=book-qirl4ysxvgcC
4. Togari N, Kobayashi A, Aishima T. Pattern recognition applied to gas chromatographic profiles of volatile components in three tea categories - ScienceDirect [Internet]. Elsevier; 1995 [cited 2017 Feb 24]. Available from: DOI: Org/10.1016/0963-9969(95)00029-1
5. Horie H, Mukai T, Kohata K. Simultaneous determination of qualitatively important components in green tea infusions using capillary electrophoresis. J Chromatogr A [Internet]. 1997 Jan 17;758(2):332–5. Available from: https://doi.org/10.1016/S0021-9673(96)00764-9
6 Zuo Y, Chen H, Deng Y. Simultaneous determination of catechins, caffeine and gallic acids in green, Oolong, black and pu-erh teas using HPLC with a photodiode array detector. Talanta [Internet]. 2002 May 16;57(2):307-16. Available from: DOI:10.1016/s0039-9140(02)00030-9
7. Sliwinska M, Wisniewska P, Dymerski T, Namiesnik J, Wardencki W. Food analysis using artificial senses. J Agric Food Chem [Internet]. 2014;62(7):1423-48. Available from: DOI. ORG/10.1021/JF403215Y
8. Ivarsson P, Krantz-Rülcker C, Winquist F, Lundström I. A voltammetric electronic tongue. Chem Senses [Internet]. 2005 Jan;30 Suppl 1:i258-9. Available from: DOI:10.1093/chemse/bjh213
9. Winquist F, Olsson J, Eriksson M. Multicomponent analysis of drinking water by a voltammetric electronic tongue. Anal Chim Acta [Internet]. 2011 Jan 10;683(2):192-7. Available from: DOI:10.1016/j.aca.2010.10.027
10. Gruden R, Buchholz A, Kanoun O. Electrochemical analysis of water and suds by impedance spectroscopy and cyclic voltammetry. Journal of Sensors and Sensor Systems [Internet]. 2014;3(2):133. Available from: DOI: 10.5194/jsss-3-133-2014
11. Alcañiz Fillol M. Diseño de un sistema de lengua electrónica basado en técnicas electroquímicas voltamétricas y su aplicación en el ámbito agroalimentario [Internet]. riunet.upv.es; 2011. Available from: DOI:10.4995/Thesis/10251/11303
12. El Alami El Hassani N, Tahri K, Llobet E, Bouchikhi B, Errachid A, Zine N, et al. Emerging approach for analytical characterization and geographical classification of Moroccan and French honeys by means of a voltammetric electronic tongue. Food Chem [Internet]. 2018 Mar 15;243:36-42. Available from: http://dx.doi.org/10.1016/j.foodchem.2017.09.067
13. Fuentes Pérez E. Aplicación de la lengua electrónica voltamétrica a alimentos líquidos [Internet]. riunet.upv.es; 2017. Available from: DOI:10.4995/Thesis/10251/90446
14. Tønning E, Sapelnikova S, Christensen J, Carlsson C, Winther-Nielsen M, Dock E, et al. Chemometric exploration of an amperometric biosensor array for fast determination of wastewater quality. Biosens Bioelectron [Internet]. 2005 Oct 15;21(4):608-17. Available from: DOI: 10.1016/j.bios.2004.12.023
15. Gutés A, Cespedes F, del Valle M, Louthander D, Krantz-Rülcker C, Winquist F. A flow injection voltammetric electronic tongue applied to paper mill industrial waters. Sens Actuators B Chem [Internet]. 2006 May 23;115(1):390-5. Available from: DOI: 10.1016/j.snb.2005.09.024
16. Storey MV, van der Gaag B, Burns BP. Advances in on-line drinking water quality monitoring and early warning systems. Water Res [Internet]. 2011 Jan;45(2):741–7. Available from: DOI: 10.1016/j.watres.2010.08.049.
17. Wei Z, Zhang W, Wang Y, Wang J. Monitoring the fermentation, post-ripeness and storage processes of set yogurt using voltammetric electronic tongue. J Food Eng [Internet]. 2017 Jun 1;203:41–52. Available from: DOI: 10.1016/j.jfoodeng.2017.01.022
18. Saidi T, Moufid M, Zaim O, El Bari N, Bouchikhi B. Voltammetric electronic tongue combined with chemometric techniques for direct identification of creatinine level in human urine. Measurement [Internet]. 2018 Feb 1;115:178-84. Available from: DOI:10.1016/j.measurement.2017.10.044
19. Bhattacharyya R, Tudu B, Das SC, Bhattacharyya N, Bandyopadhyay R, Pramanik P. Classification of black tea liquor using cyclic voltammetry. J Food Eng [Internet]. 2012 Mar;109(1):120-6. Available from: DOI:10.1016/j.jfoodeng.2011.09.026
20. Blanco M, Villarroya I. NIR spectroscopy: a rapid-response analytical tool. Trends Analyt Chem [Internet]. 2002 Apr 1;21(4):240-50. Available from: DOI: 10.1016/S0165-9936(02)00404-1
21. Saeys Y, Inza I, Larrañaga P. A review of feature selection techniques in bioinformatics. Bioinformatics [Internet]. 2007 Oct 1;23(19):2507–17. Available from: DOI:10.1093/bioinformatics/btm344
22. Guyon I, Elisseeff A. An Introduction to Variable and Feature Selection. J Mach Learn Res [Internet]. 2003 Mar;3:1157-82. Available from: http://dl.acm.org/citation.cfm?id=944919.944968
23. Keleş S, van der Laan M, Eisen MB. Identification of regulatory elements using a feature selection method. Bioinformatics [Internet]. 2002 Sep;18(9):1167-75. Available from: DOI:10.1093/bioinformatics/18.9.1167
24. Guyon I, Gunn S, Nikravesh M, Zadeh LA. Feature Extraction: Foundations and Applications [Internet]. Springer Berlin Heidelberg; 2008. (Studies in Fuzziness and Soft Computing). Available from: https://books.google.com.ec/books?id=FOTzBwAAQBAJ
25. Szymańska E, Gerretzen J, Engel J, Geurts B, Blanchet L, Buydens LMC. Chemometrics and qualitative analysis have a vibrant relationship. Trends Analyt Chem [Internet]. 2015 Jun;69:34-51. Available from: DOI: 10.1016/j.trac.2015.02.015
26. Guyon I, Elisseeff A. An Introduction to Feature Extraction. In: Guyon I, Nikravesh M, Gunn S, Zadeh LA, editors. Feature Extraction [Internet]. Springer Berlin Heidelberg; 2006 [cited 2016 Jun 2]. 1-25p. (Studies in Fuzziness and Soft Computing). Available from: http://link.springer.com/chapter/10.1007/978-3-540-35488-8_1
27. Liu H, Motoda H. Feature Selection for Knowledge Discovery and Data Mining [Internet]. Springer US; 2012. (The Springer International Series in Engineering and Computer Science). Available from: https://books.google.com.ec/books?id=aaDbBwAAQBAJ
28. Jafari P, Azuaje F. An assessment of recently published gene expression data analyses: reporting experimental design and statistical factors. BMC Med Inform Decis Mak [Internet]. 2006 Jun 21;6:27. Available from: DOI:10.1186/1472-6947-6-27
29. Bhanot G, Alexe G, Venkataraghavan B, Levine AJ. A robust meta-classification strategy for cancer detection from MS data. Proteomics [Internet]. 2006;6(2):592-604. Available from: DOI:10.1002/pmic.200500192/full
30. Kotsiantis SB, Zaharakis I, Pintelas P. Supervised machine learning: A review of classification techniques [Internet]. books.google.com; 2007. Available from: https://books.google.com.ec/books?hl=en&lr=&id=vLiTXDHr_sYC&oi=fnd&pg=PA3&dq=supervised+machine+learning&ots=CXpwvB0Con&sig=-JhVl5fviVAnLXtZptt-aBJr8oo

31. Friedman J, Hastie T, Tibshirani R. The elements of statistical learning [Internet]. Vol. 1. Springer series in statistics Springer, Berlin; 2001. Available from: http://statweb.stanford.edu/~tibs/book/preface.ps
32. Palma J, Marín R. Inteligencia artificial. Técnicas, métodos y aplicaciones. García Jurado JL, editor. Murcia: McGraw Hill; 2013.
33. Benítez R, Escudero G, Kanaan S. Inteligencia artificial avanzada [Internet]. España: Editorial UOC; 2013. Available from: http://site.ebrary.com/lib/alltitles/docDetail.action?docID=10832185
34. Boser BE, Guyon IM, Vapnik VN. A Training Algorithm for Optimal Margin Classifiers. Proceedings of the 5th Annual ACM Workshop on Computational Learning Theory [Internet]. 1992;144–52. Available from: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.21.3818
35. Cortes C, Vapnik V. Support-vector networks. Mach Learn [Internet]. 1995;20(3):273–97. Available from: DOI: 10.1007/BF00994018
36. Vapnik VN, Kotz S. Estimation of dependences based on empirical data. Vol. 41. Springer-Verlag New York; 1982.
37. Mucherino A, Papajorgji PJ, Pardalos PM. Data Mining in Agriculture [Internet]. New York, NY: Springer New York; 2009. (Springer Optimization and Its Applications; vol. 34). Available from: DOI: 10.1007/978-0-387-73669-3
38. Mollazade K, Omid M, Arefi A. Comparing data mining classifiers for grading raisins based on visual features. Comput Electron Agric [Internet]. 2012 Jun;84:124-31. Available from: DOI: 10.1016/j.compag.2012.03.004
39. Bennett KP, Campbell C. Support vector machines. ACM SIGKDD Explorations Newsletter [Internet]. 2000 Dec 1;2(2):1–13. Available from: http://portal.acm.org/citation.cfm?doid=380995.380999
40. Hsu C-W, Chang C-C, Chih-Jen L. A practical guide to support vector classification. 2003;(1):1–16. Available from: DOI: 10.1.1.224.4115
41. Wu C-H, Tzeng G-H, Lin R-H. A Novel hybrid genetic algorithm for kernel function and parameter optimization in support vector regression. Expert Syst Appl [Internet]. 2009/4;36(3, Part 1):4725-35. Available from: http://www.sciencedirect.com/science/article/pii/S095741740800300X
42. Prasoona RK, Jyoti A, Mukesh Y, Nishant S, Anuraj NS, Shobha J. Optimization of Gaussian Kernel Function in Support Vector Machine aided QSAR studies of C-aryl glucoside SGLT2 inhibitors. Interdiscip Sci [Internet]. 2013 Mar;5(1):45-52. Available from: DOI: 10.1007/s12539-013-0156-y
43. Cristianini N, Shawe-Taylor J. An Introduction to Support Vector Machines (Cambridge. Cambridge Univ. Press; 2000.
44. Devos O, Ruckebusch C, Durand A, Duponchel L, Huvenne J-P. Support vector machines (SVM) in near infrared (NIR) spectroscopy: Focus on parameters optimization and model interpretation. Chemometrics Intellig Lab Syst [Internet]. 2009 Mar 15;96(1):27–33. Available from: http://www.sciencedirect.com/science/article/pii/S0169743908002086
45. Jeng J-T. Hybrid approach of selecting hyperparameters of support vector machine for regression. IEEE Trans Syst Man Cybern B Cybern [Internet]. 2006 Jun;36(3):699–709. Available from: http://www.ncbi.nlm.nih.gov/pubmed/16761822
46. Ma X, Zhang Y, Wang Y. Performance evaluation of kernel functions based on grid search for support vector regression. In: 2015 IEEE 7th International Conference on Cybernetics and Intelligent Systems (CIS) and IEEE Conference on Robotics, Automation and Mechatronics (RAM) [Internet]. ieeexplore.ieee.org; 2015. p. 283-8. Available from: DOI: 10.1109/ICCIS.2015.7274635
47. Martens D, Baesens B. Building Acceptable Classification Models. In 2010. p. 53–74. Available from: DOI: 10.1007/978-1-4419-1280-0_3
48. Venkatesan M, Thangavelu A, Prabhavathy P. Proceedings of Seventh International Conference on Bio-Inspired Computing: Theories and Applications (BIC-TA 2012) [Internet]. Bansal JC, Singh P, Deep K, Pant M, Nagar A, editors. India: Springer India; 2013. (Advances in Intelligent Systems and Computing; vol. 202). Available from: http://www.scopus.com/inward/record.url?eid=2-s2.0-84875106244&partnerID=tZOtx3y1
49. Hastie T, Tibshirani R, Friedman J. The Elements of Statistical Learning [Internet]. New York, NY: Springer New York; 2009. (Springer Series in Statistics; vol. 18). Available from: http://www.springerlink.com/index/D7X7KX6772HQ2135.pdf
50. Choi WK. Electrochemical Characterizations of the Reducibility and Persistency of Electrolyzed Reduced Water Produced from Purified Tap Water. Int J Electrochem Sci [Internet]. 2014; Available from: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.656.3159&rep=rep1&type=pdf
51. Rajkumar M, Thiagarajan S, Chen S-M. Electrochemical detection of arsenic in various water samples. Int J Electrochem Sci [Internet]. 2011;6:3164–77. Available from: http://www.academia.edu/download/32571188/RajKumar_2.pdf
52. Ramírez I, Rivero Cebrián D, Fernández Blanco E, Pazos Sierra A. Early warning in egg production curves from commercial hens: A SVM approach. Comput Electron Agric [Internet]. 2016 Feb;121:169-79. Available from: http://www.sciencedirect.com/science/article/pii/S0168169915003919