*Abstract*

# Development of a Pattern Recognition Tool for the Classification of Electronic Tongue Signals Using Machine Learning

**Edgar G. Mendez-Lopez [1], Jersson X. Leon-Medina [2,*] and Diego A. Tibaduiza [1]**

[1] Departamento de Ingeniería Eléctrica y Electrónica, Universidad Nacional de Colombia, Cra 45 No. 26-85, Bogotá 111321, Colombia; egmendezl@unal.edu.co; dtibaduizab@unal.edu.co
[2] Departamento de Ingeniería Mecánica y Mecatrónica, Universidad Nacional de Colombia, Cra 45 No. 26-85, Bogotá 111321, Colombia; jxleonm@unal.edu.co
* Correspondence: jxleonm@unal.edu.co
† This paper is an extended version of our paper published in Presented at the 1st International Electronic Conference on Chemical Sensors and Analytical Chemistry, 01—15 July 2021 ; Available online: https://csac2021.sciforum.net/.

**Citation:** Mendez-Lopez E.G.; Leon-Medina, J.X.; Tibaduiza, D.A. Development of a Pattern Recognition Tool for the Classification of Electronic Tongue Signals Using Machine Learning. *Journal Not Specified* **2021**, *1*, 0. https://doi.org/

Received:
Accepted:
Published: 1 July 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Abstract:** Electronic tongue type sensor arrays are made of different materials with the property of capturing signals independently by each sensor. The signals captured when conducting electrochemical tests often have high dimensionality, which increases when performing the data unfolding process. This unfolding process consists of arranging the data coming from different experiments, sensors, and sample times, thus the obtained information is arranged in a two-dimensional matrix. In this work, a description of a tool for the analysis of electronic tongue signals is developed. This tool is developed in Matlab® App Designer, to process and classify the data from different substances analyzed by an electronic tongue type sensor array. The data processing is carried out through the execution of the following stages: (1) data unfolding, (2) normalization, (3) dimensionality reduction, (4) classification through a supervised machine learning model, and finally (5) a cross-validation procedure to calculate a set of classification performance measures. Some important characteristics of this tool are the possibility to tune the parameters of the dimensionality reduction and classifier algorithms, and also plot the two and three-dimensional scatter plot of the features after reduced the dimensionality. This to see the data separability between classes and compatibility in each class. This interface is successfully tested with two electronic tongue sensor array datasets with multi-frequency large amplitude pulse voltammetry (MLAPV) signals. The developed graphical user interface allows comparing different methods in each of the mentioned stages to find the best combination of methods and thus obtain the highest values of classification performance measures.

**Keywords:** Electronic Tongue; Graphical User Interface; feature extraction; dimensionality reduction; classification; machine learning.

## 1. Introduction

The data set obtained from an MLAPV (multifrequency large amplitude pulse voltammetry) electronic tongue device comes from various types of sensors and their magnitudes can have different scales [1]. These signals are characterized by having high dimensionality [2]. This can cause problems in Machine Learning models, both in pattern recognition and in the accuracy of data classification [3]. Due to this, it is necessary to perform the correct processing of these data sets to obtain high precision values for the classification of liquid substances.

In 2020, Leon-Medina et al. [2] developed a methodology that seeks to improve the classification accuracy with an approach based on non-linear feature extraction of signals obtained with electronic tongue type sensor array devices. This methodology is composed of several stages: (1) Data unfolding, (2) Normalization, (3) Non-linear dimensionality reduction, (4) Classification by means of a supervised machine learning model and finally a (5) Cross validation [2]. The application of the methodology in each stage includes the execution of algorithms in the software Matlab®. These algorithms contain a series of parameters that must be configured. As a result of the application of the methodology, the value of the classification accuracy and the confusion matrix of the classification model used are obtained, together with their performance metrics.

Due to the number of stages and the different configuration options of the parameters in the algorithms, the need was generated to develop a tool that would facilitate the application of this methodology, guiding the user through the different stages and making the configuration of the algorithms more user-friendly. One of the main advantages of a graphical user interface (GUI) is that it makes an implemented system easy to use, understand and evaluate. [4].

The section 2 ***Materials and Methods*** describes two tests performed by the devleoped GUI, as well as the datasets used in each one and the operation of the GUI. Then, the section 3 ***Results*** illustrates the main findings obtained during the two tests applying the methodology of data processing through the GUI. Finally, the section 4 ***Conclusions*** shows the main conclusions in data processing through the GUI.

## 2. Materials and Methods

The measurements of the responses of an electronic tongue system are discretized currents in time. In this way, a measurement is obtained at each instant of time for each of the electrodes that make up the electronic tongue device, obtaining a matrix of size ***I x K*** where ***I*** are the experimental tests and ***K*** are the time instants of the signal collected by each electrode. Due to the electronic tongue system has an array of sensors and taking ***J*** as the number of electrodes. A data unfolding procedure is executed to convert the three dimensional matrix ***I x J x K***, in a two-dimensional matrix ***I x (J · K)*** [2]. Figure No 1 shows an illustrative graph of the Data Unfolding process.

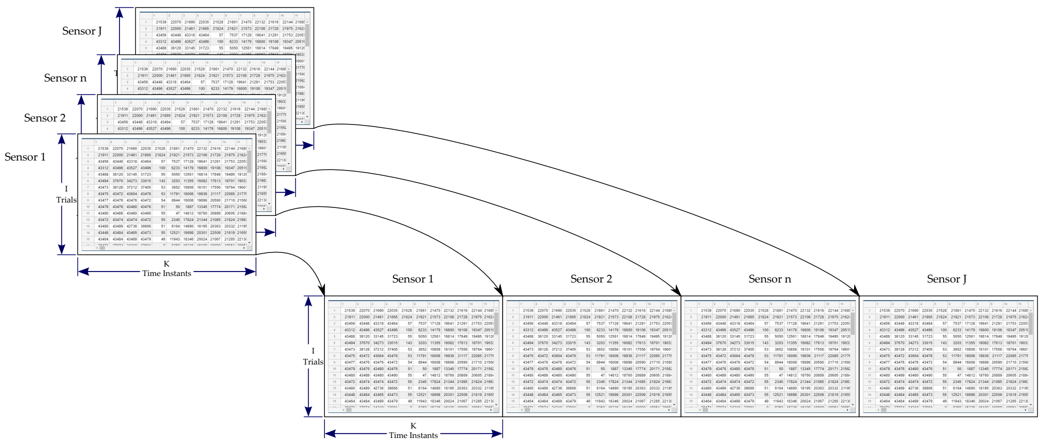


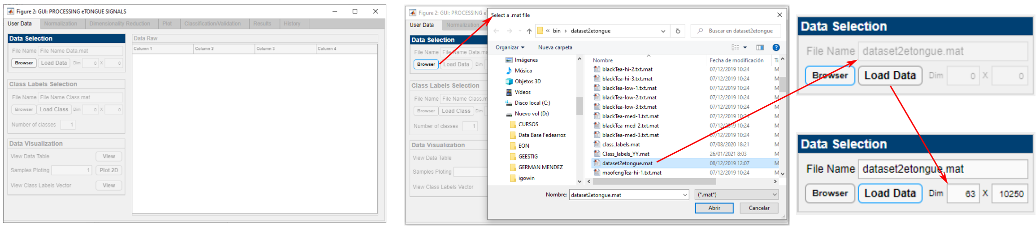

**Figure 1.** Data unfolding procedure.

In this work two tests with the developed tool are performed using two different datasets. These tests are described below:

For the first test, a dataset obtained by means of a MLAPV electronic tongue developed by Liu et al [5] is used. The electronic tongue consisted of a platinum pillar auxiliary sensor, an Ag / AgCl reference sensor, and six working electrodes made of different materials, gold, platinum, palladium, titanium, tungsten, and silver. In the experiment, the fourth titanium electrode was damaged, so it was not considered in the data analysis [5]. Seven liquids or aqueous matrices were used to collect the data from the ***first dataset***: 1) red wine, 2) Chinese liquor, 3) beer, 4) black tea, 5) oolog tea, 6 ) you maofeng and 7) you pu'er. Each one with three different concentrations (14%, 25% and 100%) of the original solution mixed with distilled water, to which three replications were made, that is, 9 samples for each liquid [2], for a total of 63 samples. With 2050 measurement points per sensor and 5 sensors in the electronic tongue, when performing the Unfolding procedure of the data (described above, see Figure 1), the dataset is composed of a matrix of size ***63 X 10250***.

The second test uses a dataset obtained from the study by Zhang et al. [6]. This second dataset contains the data collected from an MLAPV electronic tongue with five working electrodes made of gold, silver, palladium, tungsten and silver. The auxiliary electrode used is platinum pillar and the reference electrode is Ag / AgCl [7]. For this study, 13 liquids or aqueous matrices (number of samples) were used: 1) beer (19), 2) red wine (8), 3) white alcohol (6), 4) black tea (9), 5) tea Maofeng (9), 6) pu'er tea (9), 7) Oolong tea (9), 8) coffee (9), 9) milk (9), 10) cola (6), 11) vinegar (9) , 12) medicine (6) and 13) salt (6), for a total of 114 samples [6]. Like the first dataset, in the ***second dataset*** there are 2050 measurement points per sensor and 5 sensors in the electronic tongue, when performing the Unfolding procedure of the data, the second dataset has a size of ***114 X 10250***.

The developed GUI is an application made in Matlab® App Designer, it is made up of 7 tabs. Only the first tab is enabled at the beginning of the GUI, as shown in Figure 2 a). By means of the ***Browser*** button in the ***Data Selection*** section, the file containing the dataset previously ordered with the unfolding process is selected. Subsequently, the data is loaded in the GUI through the button ***Load***, after this, the size of the dataset is shown in the GUI, Figure 2 b) illustrates this process.




a) b)

**Figure 2.** a) GUI Initial state. b) dataset selection

With the dataset loaded, the ***Browser*** button is enabled in the ***Class Labels Selection*** to select, in the same way as was done with the dataset, the file ***Class Labels***. Once this vector is loaded, the number of classes used can be viewed, see Figure 3 1).

After selecting the data files, the ***Normalization*** tab is enabled, in which the method for data normalization can be selected, see Figure 3 2). With normalized data, the ***Dimensionality Reduction*** tab is enabled where the Feature Extraction technique [8] to reduce the dimensionality of the data can be selected, additionally there is a ***Parameters*** **section** where it is possible to configure certain parameters depending on the selected dimensionality reduction technique, see Figure 3 3). With the data in low dimensionality, the ***Plot*** tab is enabled for the selection and visualization of the variables in 2D and scatter plots, see Figure 3 4). Simultaneously, the ***Classification / Validation*** tab is enabled, where there are four classifiers, along with some parameters that can be configured depending on the selected classifier, see figure 3 5). Executing the classification stage, the ***Cross Validation*** section is enabled, which contains three validation techniques, see Figure 3 6). At the end of the procedure, the ***Results*** tab is enabled, where the classification performance metrics [9] and the confusion matrix are shown, see Figure 3 7). At the same time, the ***History*** tab is enabled, in which a summary of the different techniques and methods used in data processing is presented, see Figure 3 8). Figure 3 shows the sequence of enabling the GUI tabs throughout the data processing in each of the stages.

1. ***User Data Tab***: Selecting the dataset and vector from Class Labels;
2. ***Normalization Tab***: Data Normalization;
3. ***Dimensionality Reduction Tab***: Data dimensionality reduction;

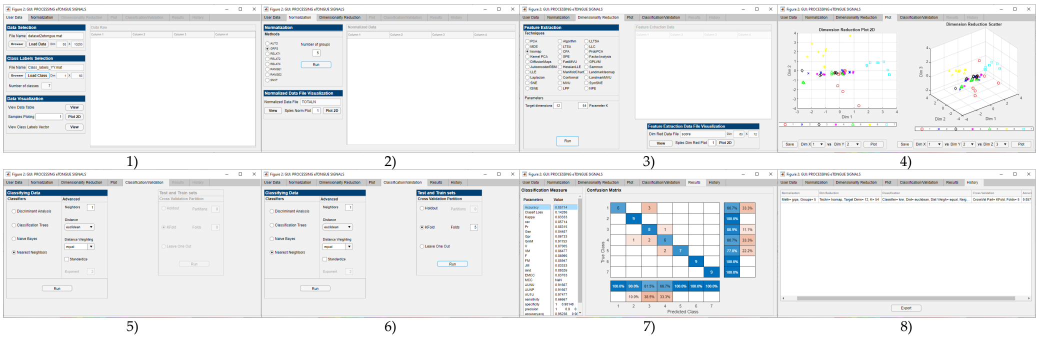


**Figure 3.** Sequence of enabling the stages in the developed GUI of the tool for classification of electronic tongue signals.

4. ***Plot Tab***: 2D and scatter graphics display;
5. ***Classification Tab***: Classifier selection;
6. ***Validation Tab***: Selection of the cross-validation method;
7. ***Results Tab***: Visualization of the Confusion Matrix and metrics of the classification model;
8. ***History Tab***: Summary of tests carried out;

In relation with the plots int the GUI, the data loaded in the GUI, as well as the normalized data and data after the dimensionality reduction process can be visualized. A table or graph visualization can be obtained by each experiment in the corresponding tabs. In Figure4, the original data are observed, in the Normalization and Dimensionality Reduction stage, the graphs are made for sample 7 in the same stages.

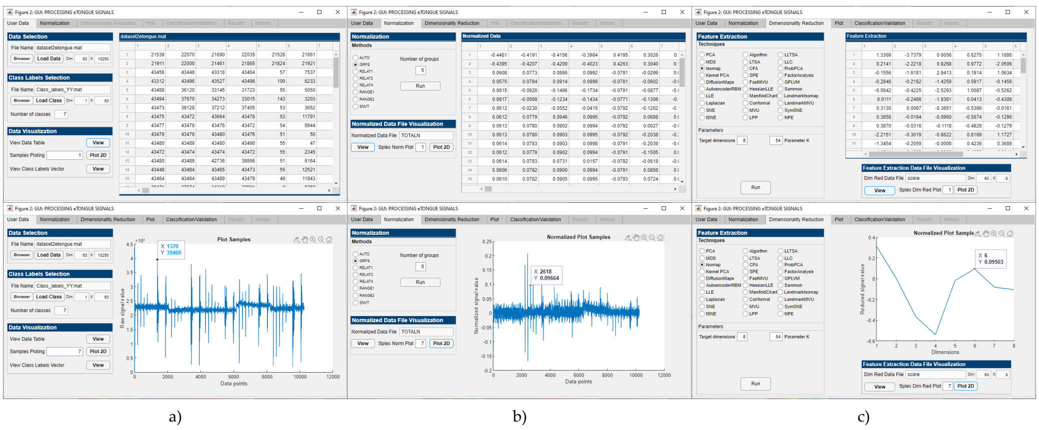


**Figure 4.** Viewing data as a sample chart or table. a) Original data; b) Standardized data; c) Data after dimensionality reduction.

## 3. Results

Through the GUI, the following tests are performed with the datasets described above, see section 2.

### 3.1. Comparison plots 2D and scatter

In the ***Plot*** tab the 2D and scatter graphs obtained after applying a dimensionality reduction technique are displayed. Figure 5 shows the graphs obtained with three different dimensionality reduction techniques applied to the first dataset. Additionally, new graphs generated by selecting different dimensions are observed, the label corresponds to each class of liquid in the dataset. Each graph can be saved in a file independently. The parameters used for each dimensionality reduction technique are described below:

a)

b)

c)

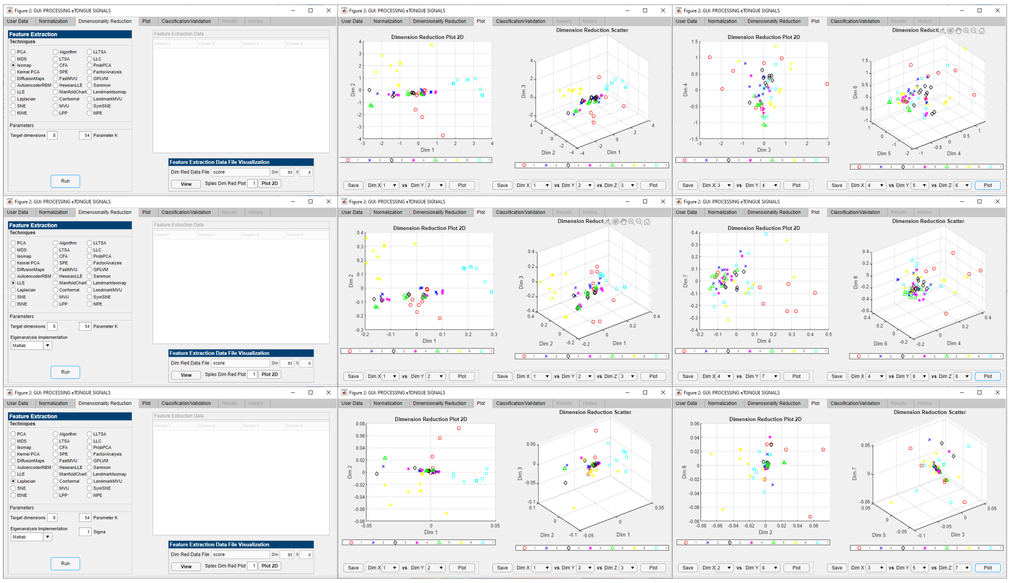


**Figure 5.** Data representation after dimensionality reduction. The labels are described in the numbered list.

a). Dimensionality reduction method=Isomap, Dim=8, K=54, Plot Dim1 2D=1,2 (Default), Plot Dim1 scatter=1,2,3 (Default), Plot Dim2 2D= 3,4, Plot Dim2 scatter=4,5,6;

b). Dimensionality reduction method=Locally Linear Embedding (LLE), Dim=8, K=54, Plot Dim1 2D=1,2 (Default), Plot Dim1 scatter=1,2,3 (Default), Plot Dim2 2D= 4,7, Plot Dim2 scatter=4,6,8;

c). Dimensionality reduction method=Laplacian Eigenmaps, Dim=8, K=54, Plot Dim1 2D=1,2 (Default), Plot Dim1 scatter=1,2,3 (Default), Plot Dim2 2D=2,8, Plot Dim2 scatter=3,5,7;

### 3.1.1. Classification accuracy behavior

Two tests are described below to observe the behavior of the classification accuracy. These tests are applied to the second dataset based on the developed methodology of [7]. First, the number of neighbors ***(k)*** of the *k*-NN Classifier is modified, varying its value from 1 to 16, but keeping the number of dimensions fixed at 8 in the PCA dimensionality reduction technique. In the second test, the number of PCA dimensions is varied from 3 to 16, but the number of neighbors (*k*) is fixed equal to 2. In both tests, the group scaling method (GRPS) is used to normalize the data and 5-Fold Cross validation is perfromed. In the Figure 6, the results of the tests carried out are described below:

a). Confusion matrix and performance metrics of the classification model for the Accuracy of 94.73 % obtained in the first test with a parameter *k* **= 2**;

b). Confusion matrix and performance metrics of the classification model for the Accuracy of 50 % obtained in the first test with a parameter *k* **=** ***16***;

c). Summary of the trials of the first trial displayed in the ***History tab*** of GUI;

d). Excel file exported from ***History tab*** for the first test;

e). Graph of Accuracy vs Number of *k* Neighbors, obtained from the results in the first test.

f). Graph of Accuracy vs Number of Dimensions, obtained from the results in the second test.

## 4. Conclusions

This work showed the development of a tool for the processing of data contained acquired by an electronic tongue type sensor array. First, the GUI design allows the user

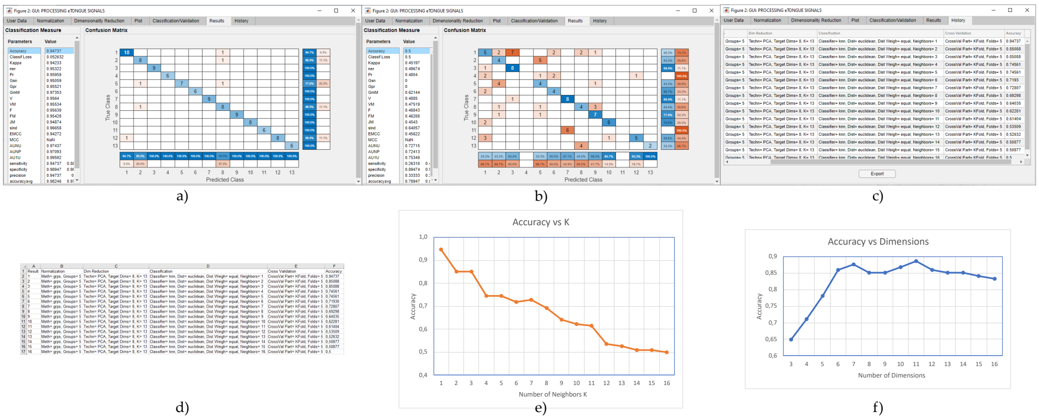



**Figure 6.** Confusion matrix results, and behacior of accuracy varying the number of target dimensions and number of *k* neighbors.

to be guided intuitively through the signal processing methodology by enabling the tabs, but at the same time it allows the user to choose the different techniques and methods, as well as the parameter configuration. Second, the GUI offers the visualization of the data, by means of tables or graphically, both the original data and those transformed in the Normalization and dimensionality reduction stages. Another advantage is the visualization of 2D and 3D scatter graphics, where the user can observe the distribution of the samples, according to the selected feature extraction technique, choosing between different combinations of dimensions. In the same way, this tool offers the visualization of the results in the confusion matrix and the performance classification metrics of the classification model, finally it provides a summary table of the tests carried out in such a way that the user can easily compare the results obtained.

**Funding:** This research was funded by FONDO DE CIENCIA TECNOLOGÍA E INNOVACION FCTeI DEL SISTEMA GENERAL DE REGALÍAS SGR. The authors express their gratitude to the Administrative Department of Science, Technology and Innovation—Colciencias with the grant 779—"Convocatoria para la Formación de Capital Humano de Alto Nivel para el Departamento de Boyacá 2017" for sponsoring the research presented herein.

**Acknowledgments:** Jersson X. Leon-Medina is grateful with Colciencias and Gobernación de Boyacá for the PhD fellowship..

**Conflicts of Interest:** "The authors declare no conflict of interest.".

## References

1. Leon-Medina, J.X.; Vejar, M.A.; Tibaduiza, D.A. Signal Processing and Pattern Recognition in Electronic Tongues: A Review. In *Pattern Recognition Applications in Engineering*; IGI Global, 2020; pp. 84–108. doi:10.4018/978-1-7998-1839-7.ch004.
2. Leon-Medina, J.X.; Anaya, M.; Pozo, F.; Tibaduiza, D. Nonlinear Feature Extraction Through Manifold Learning in an Electronic Tongue Classification Task. *Sensors* **2020**, *20*, 4834.
3. Ayesha, S.; Hanif, M.K.; Talib, R. Overview and comparative study of dimensionality reduction techniques for high dimensional data. *Information Fusion* **2020**, *59*, 44–58.
4. Djelouat, H.; Ali, A.A.S.; Amira, A.; Bensaali, F. An interactive software tool for gas identification. *Journal of Natural Gas Science and Engineering* **2018**, *55*, 612–624.
5. Liu, T.; Chen, Y.; Li, D.; Wu, M. An active feature selection strategy for DWT in artificial taste. *Journal of Sensors* **2018**, *2018*.
6. Zhang, L.; Wang, X.; Huang, G.B.; Liu, T.; Tan, X. Taste recognition in E-tongue using local discriminant preservation projection. *IEEE transactions on Cybernetics* **2018**, *49*, 947–960.
7. Leon-Medina, J.X.; Cardenas-Flechas, L.J.; Tibaduiza, D.A. A data-driven methodology for the classification of different liquids in artificial taste recognition applications with a pulse voltammetric electronic tongue. *International Journal of Distributed Sensor Networks* **2019**, *15*, 1550147719881601.
8. Van der Maaten, L. An introduction to dimensionality reduction using matlab. *Report MICC 07-07 Universiteit Maastricht* **2007**, *1201*, 1–44.
9. Ballabio, D.; Grisoni, F.; Todeschini, R. Multivariate comparison of classification performance measures. *Chemometrics and Intelligent Laboratory Systems* **2018**, *174*, 33–44.