
ECSA-7
7th International Electronic Conference on Sensors and Applications
15 – 30 November 2020
sensors
chemosensors
MDPI

Evaluation of Feature Selection Techniques in a Multifrequency Large Amplitude Pulse Voltammetric Electronic Tongue

Luis F. Villamil-Cubillos[1,†], Jersson X. Leon-Medina*[1], Maribel Anaya[2], Diego A. Tibaduiza [4]

1 Departamento de Ingeniería Mecánica y Mecatrónica, Universidad Nacional de Colombia, Cra 45 No. 26-85, Bogotá 111321, Colombia; jxleonm@unal.edu.co

2 MEM (Modelling-Electronics and Monitoring Research Group), Faculty of Electronics Engineering, Universidad Santo Tomás, Bogotá 110231, Colombia; maribelanaya@usantotomas.edu.co

3 Departamento de Ingeniería Eléctrica y Electrónica, Universidad Nacional de Colombia, Cra 45 No. 26-85, Bogotá 111321, Colombia; dtibaduizab@unal.edu.co

* Correspondence: jxleonm@unal.edu.co; Tel.: +57-311-476-3378

† Presented at the 7th Electronic Conference on Sensors and Applications, 15–30 November 2020; Available online: https://ecsa-7.sciforum.net/.

OUTLINE

1 Introduction

2 Problem statement

3 Theoretical background

4 Materials and Methods

5 Results

6 Discussion

7 Conclusions

Minciencias

UNIVERSIDAD NACIONAL DE COLOMBIA

1 Introduction

Electronic tongue sensor array

Minciencias
UNIVERSIDAD NACIONAL DE COLOMBIA

2 Problem statement

High amount of data

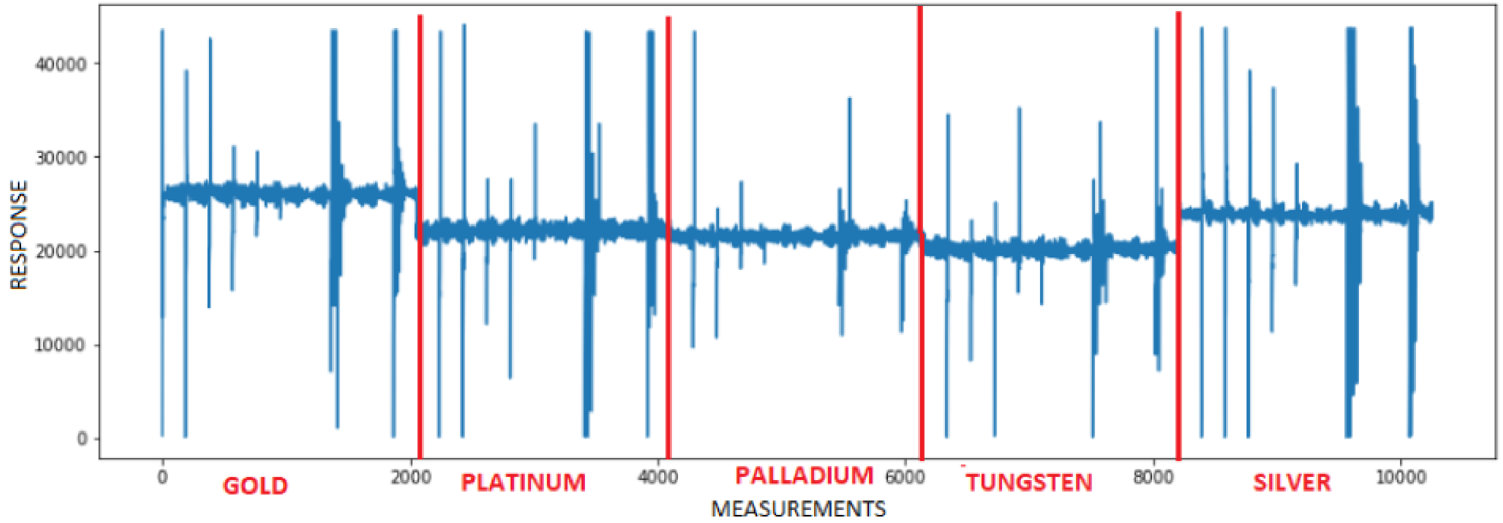
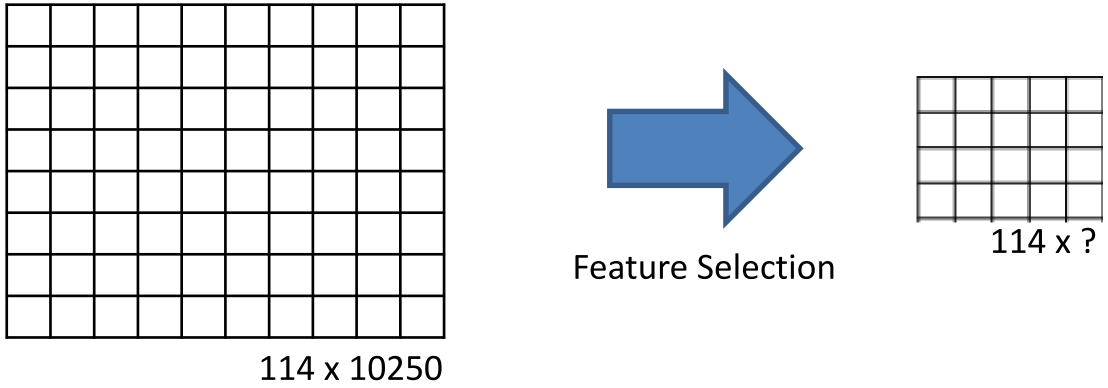


Figure 2. Set of 5 MLAPV response signals that characterize a beer sample.



Zhang et al., 2018

3 Theoretical background

Principal concepts

- Electronic tongue
- Feature selection
- Filter Method
- Embedded Method
- 5 fold cross validation
- Hyperparameter tuning
- Accuracy as performance measure

4 Materials and methods

MLAPV ELECTRONIC TONGUE DATA SET

Label	1	2	3	4	5	6	7	8	9	10	11	12	13
Liquid type	Beer	Black tea	Coffee	Cola	Maofeng tea	Medicine	Milk	Oolong tea	Pu er tea	Redwine	Salt	Vinegar	Whitespirit
Samples	19	9	9	6	9	6	9	9	9	8	6	9	6

Figure 1. Dataset distribution

Zhang et al., 2018

Minciencias

UNIVERSIDAD NACIONAL DE COLOMBIA

4 Materials and methods

Feature selection

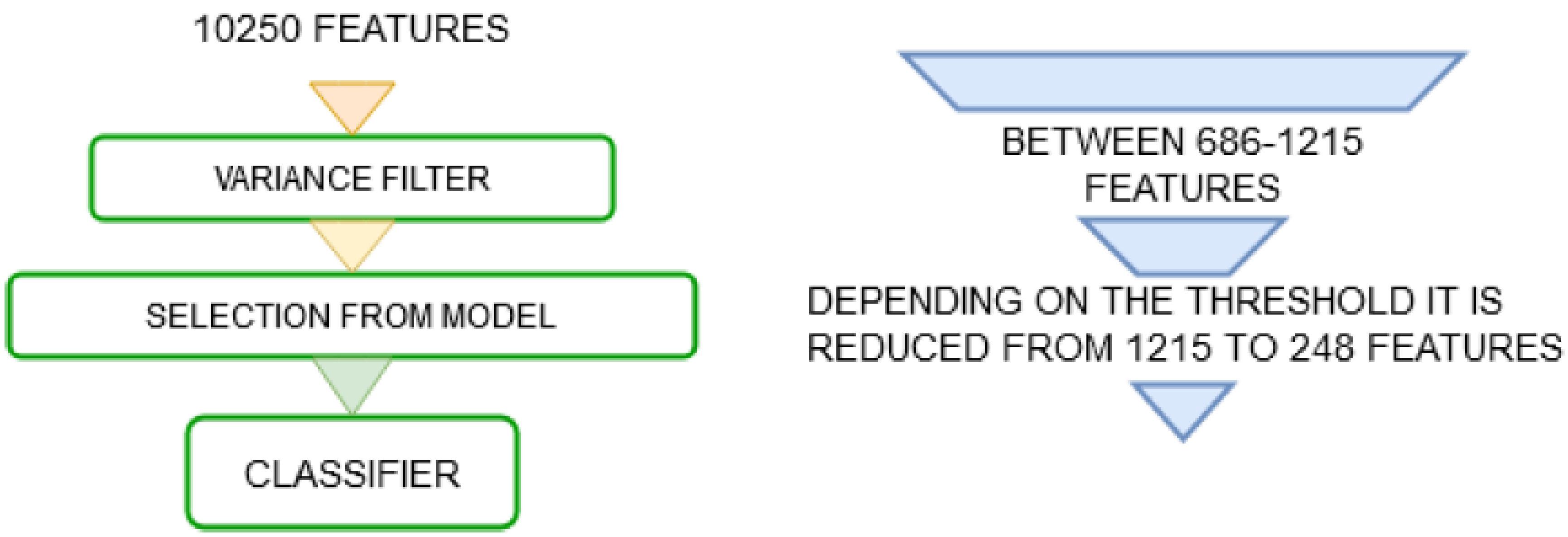


Figure 3. Combination between variance filter and selection from model (diagram)

4 Materials and methods

Feature selection

- **Combination between variance filter, ANOVA filter and selection from model:** A similar technique to the previous one is proposed using another intermediate feature selection method, the ANOVA technique as shown in Figure 4.

Figure 4. Combination between variance filter, ANOVA filter and selection from model (diagram)

- **Combination between variance filter, ANOVA filter and RFE technique:** In this case, the recursive RFE elimination method will be used after applying the variance and ANOVA filters as show in Figure 5 . It is expected to reduce the number of features at the RFE input and in this way reduce the processing time and use a small step size, which can help to improve the final performance of the algorithm.

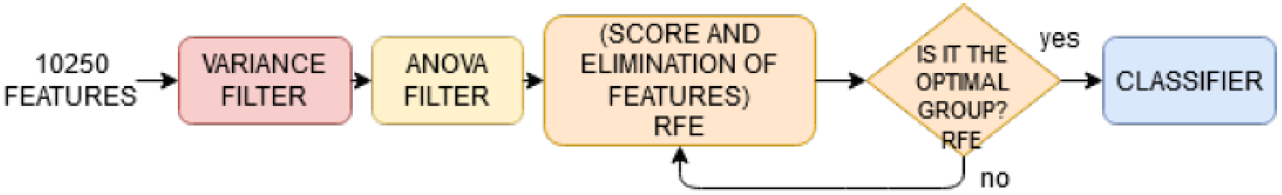


Figure 5. Combination between variance filter, ANOVA filter and RFE technique (diagram)

Minciencias

UNIVERSIDAD NACIONAL DE COLOMBIA

5 Results

Accuracy results

Table 1. Best results using variance filter.

Threshold	Classifier	Accuracy
0,0005	Multilayer perceptron(MLPC)	0,8762
0,002	Multilayer perceptron(MLPC)	0,8679
0	Linear SVC	0,8592

Table 2. Best results using ANOVA F-score

Features	Classifier	Accuracy
5000	Multilayer perceptron(MLPC)	0,9019
6000	Multilayer perceptron(MLPC)	0,8940
3000	Multilayer perceptron(MLPC)	0,8857

Table 3. Best results using 3 estimators and its optimal number of features

Estimator	Step	Classifier	Optimal features	Accuracy
Linear SVC	20	Multilayer perceptron(MLPC)	820	0.9385
	50	Multilayer perceptron(MLPC)	950	0.9035
	50	Linear SVC	950	0.8947

Minciencias UNIVERSIDAD NACIONAL DE COLOMBIA

5 Results

Optimal number of features

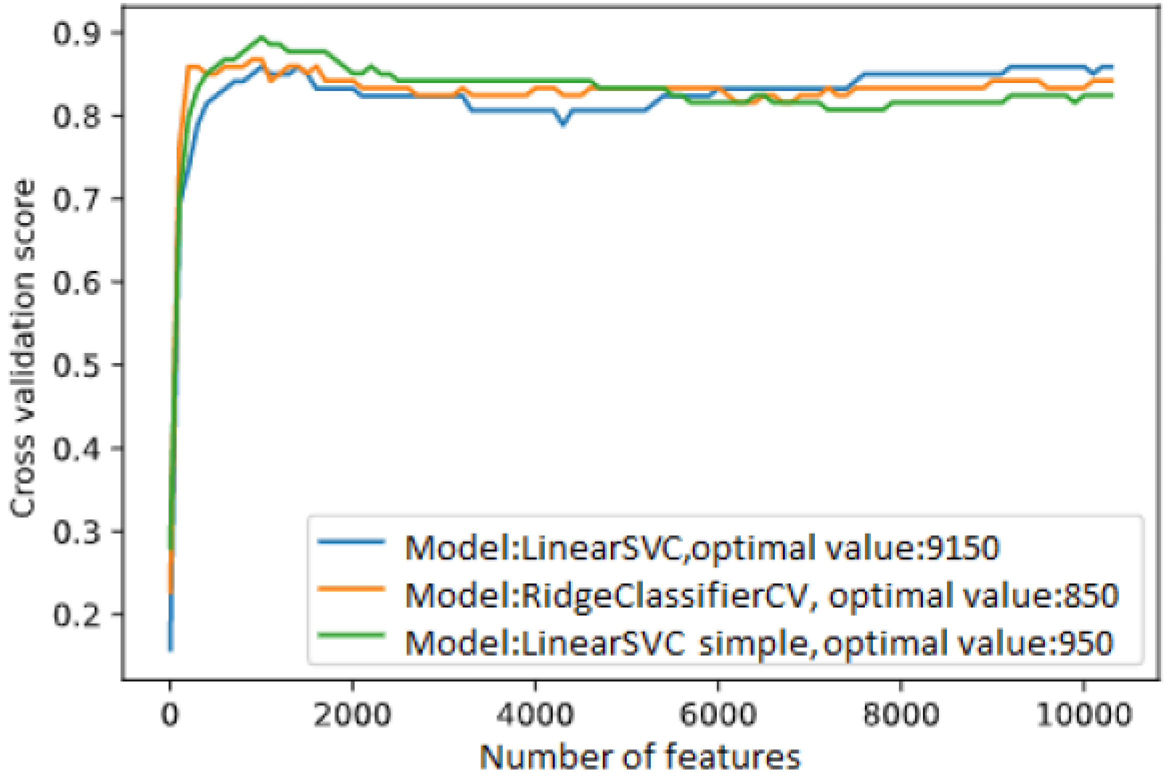


Figure 6. Accuracy vs number of features selected using RFE and LinearSVC as classifier

Minciencias

UNIVERSIDAD NACIONAL DE COLOMBIA

5 Results

Feature importance

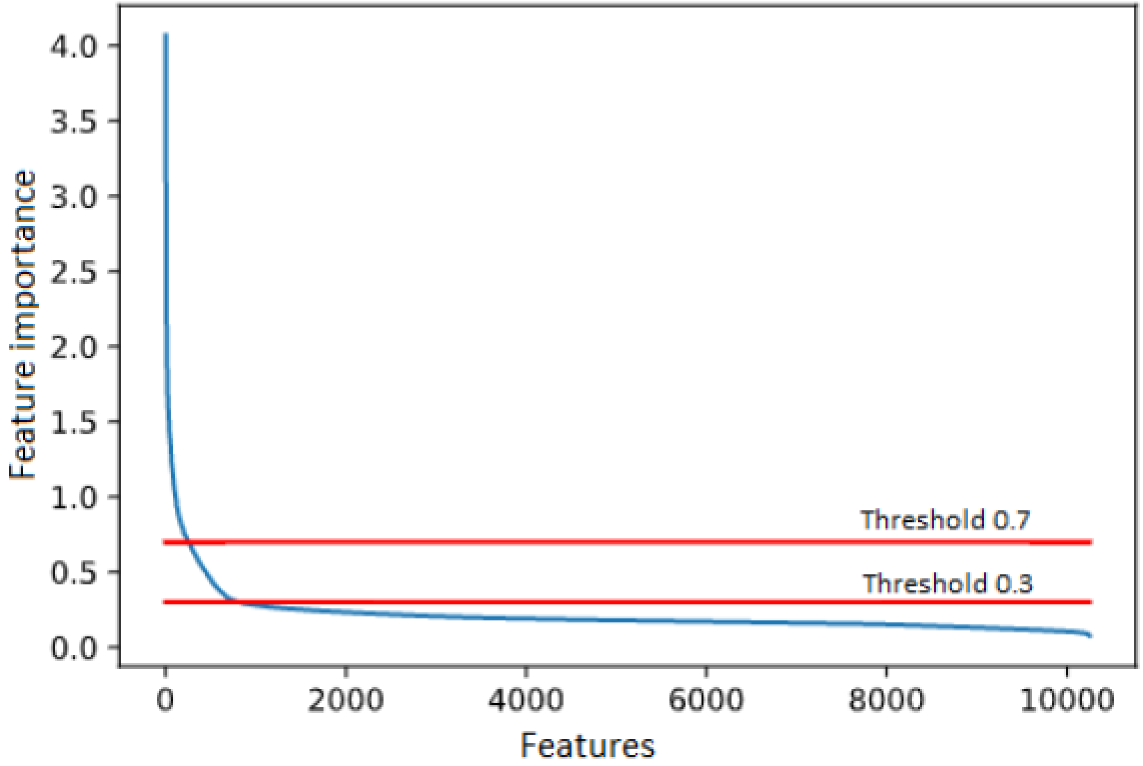


Figure 7. Feature importance from model using logistic regression

5 Results

Table 4. Best results using selection from model

Classifier	Estimator	Threshold	Accuracy
Multilayer perceptron(MLPC) (adjusted)	Logistic Regression	0,4	0,9114
	LinearSVC	0,7	0,8940

Table 5. Best results of the combination between the variance filter and selection from model

Classifier	Threshold Selection from model	Accuracy
Multilayer perceptron(MLPC) (adjusted)	0,4	0.9032
	0,6	0.9032
Multilayer perceptron(MLPC)	0,2	0.9028

Table 6. Best results of the combination between variance filter, ANOVA filter and selection from model

Classifier	Threshold Selection from model	Threshold variance	Features	Accuracy
Multilayer perceptron(adjusted)	0,5	0,0001	5200	0.9285
	0,4	0,0001	5200	0.9285
	0,3	0,0001	5200	0.9123

5 Results

Best confusion matrix

Actual Class \ Predicted Class	1	2	3	4	5	6	7	8	9	10	11	12	13
1	17	0	0	0	0	0	0	2	0	0	0	0	0
2	0	9	0	0	0	0	0	0	0	0	0	0	0
3	0	0	9	0	0	0	0	0	0	0	0	0	0
4	0	0	0	6	0	0	0	0	0	0	0	0	0
5	0	0	0	0	8	0	0	0	1	0	0	0	0
6	1	0	0	0	0	5	0	0	0	0	0	0	0
7	0	0	0	0	0	0	9	0	0	0	0	0	0
8	0	1	0	0	0	0	0	6	2	0	0	0	0
9	0	0	0	0	0	0	0	0	9	0	0	0	0
10	0	0	0	0	0	0	0	0	0	8	0	0	0
11	0	0	0	0	0	0	0	0	0	0	6	0	0
12	0	0	0	0	0	0	0	0	0	0	0	9	0
13	0	0	0	0	0	0	0	0	0	0	0	0	6

Figure 8. Confusion matrix accuracy=93.86%

Minciencias UNIVERSIDAD NACIONAL DE COLOMBIA

5 Results

Combination of variance filter, ANOVA filter and RFE technique

Table 7. Best result of the combination between variance filter , ANOVA filter and RFE technique

Classifier	Threshold variance	Features	Accuracy
Multilayer perceptron(adjusted)	0	6200	0.9032
	0	5200	0.9028
	0	4800	0.8937

Minciencias UNIVERSIDAD NACIONAL DE COLOMBIA

6 Discussion

- The application of the feature selection techniques increases the accuracy of the classification in most cases (initially 81.54% simply using the MLP classifier), as well as reducing the time of algorithm prediction by reducing the number of features in each test instance.

- The two best results obtained were 93.86% using the RFE technique and MLP classifier and 92.85% using combination between variance filter, ANOVA filter and selection from model with an MLP classifier.

- Although in the first case the accuracy is greater, the time required for the selection of features and training is almost 117 times greater.

7 Conclusions

- The use of this type of method is useful to analyze data from sensor arrays, achieving an increase in the accuracy of the classification of up to about 12%, in addition, machine learning models diminish their training and prediction time by reducing the number of features.

- It is noteworthy that the use of combined feature selection techniques can achieve high precision, achieve a faster model construction and become very stable, compared to the recursive feature elimination RFE method, which, although it is more precise, the last is slow to select the optimal set of features.

- Although the best results are obtained with MLP classifier, several iterations are necessary to obtain the best performance, since the definition of the weights of each feature changes after the construction of each model, therefore, the results vary.

REFERENCES

1. 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** , *PP* , 1-–14.
2. Kumar , M. ; Kumar Rath , N.; Swain , A. ; Kumar Rath , S. Feature Selection and Classification of Microarray Data using MapReduce based ANOVA and K-Nearest Neighbor.*Procedia Computer Science* **2015** , *54* , 301–310.
3. Ding, H. ; Feng , P.-M. ; Chen , W.; Lin , H. Identification of bacteriophage virionproteins by the anova feature selection and analysis.*Molecular BioSystems* **2014** , *10*.
4. Duan, K.-B. ; Rajapakse , J. ; Wang , H. ; Azuaje , F. Multiple svm-rfe for gene selection in cancer classification with expression data. *IEEE transactions on nanobioscience* **2005** , *4*, 228—34.
5. Chen , Q. ; Meng, Z. ; Liu , X. ; Jin , Q. ; Su , R. Decision variants for the automatic determination of optimal feature subset in rf-rfe.*Genes* **2018** , *9* , 301.
6. 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 applica-tions with a pulse voltammetric electronic tongue. *International Journal of Distributed Sensor Networks* **2019** , *15 no. 10* , 1550147719881601.
7. Chandrashekar , G. ; Sahin , F. A survey on feature selection methods. *ComputersElectrical Engineering* **2014** , *40 no. 1* , 16 – 28.
8. 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.
9. Podrażka , M. ; Baczynska , E. ; Kundys , M. ; Jeleń , P. S. ; Nery , E. W. Electronic tongue—a tool for all tastes? . *Biosensors* **2017** , *8*.
10. Valle , M. Bioinspired sensor systems . *Sensors (Basel, Switzerland)* **2011** , *11* , 10180-–6.
11. Pedregosa , F.; Varoquaux , G.; Gramfort , A.; Michel , V.; Thirion , B.; Grisel , O.; Blon-del , M.; Prettenhofer , P.; Weiss , R.; Dubourg , V.; Vanderplas , J.; Passos , A.; Cournapeau , D.; Brucher ,M.; Perrot ,M.; Duchesnay E. Scikit-learn: Machine learning in Python . *Journal of Machine Learning Research* **2011** , *12* , 2825—2830.

Minciencias

UNIVERSIDAD NACIONAL DE COLOMBIA

Thank You

jxleonm@unal.edu.co