Classification of As, Pb and Cd Heavy Metal Ions Using Square Wave Voltammetry, Dimensionality Reduction and Machine Learning

JERSSON X. LEON-MEDINA, DIEGO A. TIBADUIZA, JUAN C. BURGOS, MARTHA CUENCA, AND DREIDY VASQUEZ

1Departamento de Ingeniería Mecánica y Mecatrónica, Universidad Nacional de Colombia, Bogotá 111321, Colombia
2Departamento de Ingeniería Eléctrica y Electrónica, Universidad Nacional de Colombia, Bogotá 111321, Colombia
3Programa de Ingeniería Química, Universidad de Cartagena, Cartagena de Indias 130001, Colombia
4Escuela de Ingeniería Química, Pontificia Universidad Católica de Valparaíso, Valparaíso 2362854, Chile

Corresponding author: Dreidy Vasquez (dreidy.vasquez@pucv.cl)

This work was supported by the Escuela de Ingeniería Química of the Pontificia Universidad Católica de Valparaíso.

ABSTRACT The detection and classification of heavy metals is a growing need to guarantee the quality of process water in different industries. However, the official methodologies to evaluate the presence of these contaminants require samples pre-processing, making them time-consuming and expensive; these elements do not allow online monitoring. For this reason, new technologies are required for online monitoring and evaluation. In this work, a new methodology is presented for the detection and classification of different heavy metal ions such as: As, Pb and Cd. Commercial graphite sensors modified with 2D molybdenite were used applying an electroanalytical technique of square wave voltammetry. Subsequently, signal processing based on pattern recognition and machine learning methods was carried out. This classification methodology includes the following steps: data display and arrangement, dimensionality reduction through the t-distributed stochastic neighbor embedding (t-SNE) method, which serves as feature extraction, and the support vector machines (SVM) method as a classifier. The validation is carried out with a data set of 118 aqueous samples. Leave one out cross-validation (LOOCV) was used to obtain classification accuracy. The results showed a classification accuracy of 98.31% with only two errors of the experimental validation with this data set. It is concluded that this methodology is a useful tool for detecting the presence of these ions in aqueous samples with MoS2-2D.

INDEX TERMS Classification, heavy metals detection, square wave voltammetry, machine learning, feature extraction, t-SNE, SVM, multivariate data analysis.

I. INTRODUCTION

The presence of heavy metals in water resources has been widely reported by researchers and control entities [1]. This environmental issue is caused by the emission of contaminants and inadequate management of chemical products and industrial wastes. Fast and reliable detection and quantification of heavy metal ions are becoming a priority for communities near water bodies, because the stabilization of ions in water through biogeochemical processes [2] leads to toxicological consequences in the water ecosystem [3]. There are different methods to detect the presence of heavy metals based on instrumental analytical techniques [4], [5], but most of them demand long processing times of samples and equipment [6]. Thus, electroanalytical techniques, such as square wave voltammetry (SWV), arise as a fast and reproducible alternative to detect chemical substances, including heavy metals [7], with measurements made out of small samples and skipping sample preprocessing steps [8].

Square wave voltammetry (SWV) has been successfully used as electrochemical method for the anodic stripping analysis of heavy metal ions [9]. In order to perform classification tasks of heavy metal ions the experiments are done with different concentrations. However, after performing electrochemical experiments with SWV method, a set of different voltammograms is obtained. In this work,
three different heavy metal ions were classified Cd(II), Pb (II), and As (III); thus, data from each heavy metal ion at different concentrations were obtained. These data were obtained in a two-dimensional way, owing to the number of experiments performed with different concentrations in parts per million (ppm). The entire dataset must be arranged in such a way that multivariate analysis using machine learning can be performed [10]. In recent years, machine learning has demonstrated its powerful capability to accomplish different tasks, such as clustering, classification, and regression [11]. Data-driven paths have evolved greatly, obtaining useful insights about the acquired information. However, there are some drawbacks related to the large amount of data obtained in each experiment [12]. Moreover, these data can present some outliers and abnormalities [13]. To deal with these situations, dimensionality reduction algorithms have emerged as a successful alternative [14].

Dimensionality reduction and manifold learning algorithms transform raw data into a new low-dimensional data representation. This reduced data tends to form clusters in such a way that the intraclass distance is minimized and the interclass distance is maximized [15]. The dimensionality-reduced data serves as an input to a machine-learning classifier algorithm. Different machine learning classification algorithms have been developed, such as k-nearest neighbors, naive Bayes, classification trees, linear discriminant analysis, random forest, Xgboost, and support vector machines (SVM) [16]. Electrochemical sensors can suffer degradation over time, presenting drift phenomena, which can also be alleviated via machine learning approaches [17].

In this study, the classification of square wave voltammograms from aqueous samples containing heavy metals ions was performed. It was applied dimensional reduction, and machine learning. Validation of this methodology was done using a dataset of 118 aqueous samples containing Cd(II), Pb (II), and As (III). In general, this study includes a related work section that give information of electrochemical sensors reported, a materials and methods section describing the modified sensor with 2D-molybdenum disulfide procedure and square wave voltammetry experiments. Following, the section 4 explains the proposed heavy metal classification methodology with every one of its stages: data arrangement, dimensionality reduction using the t-distributed stochastic neighbor embedding method (t-SNE), machine learning classification via SVM and cross validation. The obtained results are detailed in section 5. Finally, the main conclusions are presented in the last Section.

II. RELATED WORK

The development and modification of materials for sensors in liquid matrices have been studied extensively, allowing the use of different techniques to obtain them, such as nanostructured hydroxyapatite-naphion-ionophore compounds, which are lipid-soluble molecules that can transport ions. Studies of bismuth, tin, and other metals with sensitivity for the detection of heavy metals have also been developed. One of the most straightforward procedures that have been found for the design and manufacture of these sensors is the deposition of thin films in carbon-screened electrodes, either by electrochemical techniques or by drop-casting [18], [19] [20].

Electrochemical detection of heavy metals is an alternative method to measure them, avoiding sample preparation, compared to conventional methods [21]. Voltammetry, potentiometry, impedemetry, amperometry, and conductometry are electrochemical techniques that have been used to detect heavy metals. Mercury electrodes have been successfully used to detect various heavy metals, but their toxicity represents a limitation. Therefore, other materials have been used to replace them [22]. One of them is bismuth, but it has a limitation because of its narrow cathodic potential range and low air stability, as it suffers from natural oxidation. Carbon nanotubes and graphene can improve the sensitivity and performance of heavy metal detection. However, multiple peaks can be observed because of the deposited films and different surface site interactions [9]. As an alternative, metallic nanoparticles are used to modify the surface and improve the sensitivity and selectivity. An electrode modified with a metallic nanoparticle has a higher surface area, improved electron and mass transfer rates, and enhanced analytical characteristics (limit of detection, sensitivity, response stability, and multi-detection capability). The catalytic properties of nanoparticles depend on the particle size. Gold is a nanoparticle that can improve heavy metal detection, but it has a high cost, which is a limitation for its use as an electrode modifier [23].

Another type of nanoparticle that can be used for sensing is 2D materials such as graphene and molybdenum disulfide. Until now, few studies have reported the electrochemical detection of heavy metals such as cadmium (II) [24] and mercury [25] with MoS2-2D functionalized with N atoms and with the T-MoS2 phase, respectively. Molybdenum disulfide as bulk material is an indirect semiconductor with a bandgap of 1.2 eV while as a monolayer, it turns into a direct semi-conductor with a bandgap of 1.8 eV with more active sites formed due to higher surface area, formation of channels for ion adsorption, and intercalation that could be useful for electrochemical sensing [26].

There are two main types of techniques for the fabrication of two-dimensional materials: bottom-up and top-down. For the first one, wet chemical, chemical vapor deposition (CVD), physical vapor deposition while for the second one chemical exfoliation and mechanical exfoliation, respectively. A more economical and scalable process is a chemical exfoliation that can be performed by ultrasound-assisted, ball milling, ultra-high mechanical mixing using a suitable solvent with a surface tension capable of overcoming the energy between the layers, which is approximately 40 mJ/m2, than at the moment the best result is with N-methyl pyrrolidinone (NMP) [27]. Exfoliation products, in general, are a mix of monolayers and few-layer nanoparticles with a broad size distribution depending on the assisted power used [28].
A. CHALLENGES

Electrochemical sensors for heavy metals detection have to overcome some challenges like limit of detection, sensitivity, response stability, and multi-detection capability. In this way, the use of new sensors or modified sensors as the used in this paper allow to continue the improvement of these features. However, the analysis of data to obtain the main features and information from the multisensor system still as a need and this paper provide a new methodology to this aim. As will be show in the next sections, this methodology combines the use of square wave voltammetry, data unfolding, t-sne for dimensionality reduction, SVM classification and LOOCV cross validation with excellent results.

III. MATERIALS AND METHODS

A. MODIFIED SENSOR WITH MOLYBDENUM

1) MOLYBDENUM DISULFIDE EXFOLIATION

Molybdenum disulfide exfoliated in an aqueous solution with dextrin as an additive that reduce hydrophobicity and promotes dispersion. Ultrasound tip gave the power necessary for the separations of the layered material and its conversion in few-layers materials, as it is possible to see in Figures 1 and 2. SEM images show the mineral in bulk state and the exfoliated particles. As was mentioned, this 2D particle change the semiconducting properties and the adsorption of these ions on the surface of this material during the detection process changes its electronic behavior. After the exfoliation, screen-printed electrodes modification consist in drop-casting the MoS2-2D solution on the working electrode for the detection of As+3, Cd+2 and Pb+2 in aqueous solution.

2) EXFOLIATION OF THE two-DIMENSIONAL MATERIAL AND SCREEN-PRINTED ELECTRODE MODIFICATION

Screen-printed electrodes are practical tools for the analysis of different sensible materials by modifying the working electrode. MoS2-2D nanolayers obtained by ultrasound-assisted exfoliation of molybdenite minerals in an aqueous dextrin solution for 60 min and isolated by centrifugation at 3000 rpm for 1 h to remove the unexfoliated particles. The obtained nanopowder redispersed in N, N-dimethylformamide (DMF), after washing by vacuum filtration, with an ultrasound tip for 15 min. With a drop of 3µL the screen-printed electrode modification is obtained. The raw material characterization by SEM (Hitachi US 3500), and the composition by XRD with a D8 diffractometer, are necessary before the exfoliation process is performed. The two-dimensional material required a characterization by scanning electron microscopy (SEM), and UV-vis spectrophotometry using a UV-vis spectrophotometer (Halo RB10) for the confirmation of the exfoliation. A PalmSens potentiostat allows the electrochemical characterization and heavy metal detection, by a first-cycle voltammetry from −1 to 1 V was executed in 1M NaCl aqueous solution. For heavy metal detection, standard solutions of Cd(II), Pb (II), and As (III) were synthesized using CdCl2.H2O (Merck), Pb(NO3)2 in 0.5M HNO3 (Merck), and H3AsO3 in 0.5M HNO3 (Merck), respectively. For every ion, an aqueous solution was prepared by dilution of the 1000 ppm ion solution to the concentration needed and adjusting to 1M NaCl solution. The optimized parameters for heavy metal detection by square-wave voltammetry use a 10 ppm solution. Finally, the detection of these ion by SWV measurement using different concentrations of heavy metals, from 1, 0.5, 0.1, 0.05, and 0.01 ppm concentration for each metal with 1M NaCl as buffer.

B. SQUARE WAVE VOLTAMMETRY EXPERIMENTS

Square-wave voltammetry (SWV) is a fast electroanalytical technique. It depends on the frequency (Hz) and step height (mV). It generates a peak-shaped symmetrical voltamogram. The current is sampled twice during each cycle of the square wave, once at the end of the forward pulse and once at the end of the reverse pulse. The concentration of the studied substance is proportional to the peak of the current. It has a high sensitivity because the current peak is larger than the oxidation or reduction separated signals.
Very low detection limits can be achieved by applying effective discrimination against the charging background current [29].

The developed methodology to process the data captured with the SWV method was tested in samples at six different concentrations of As, Pb, and Cd heavy metal ions. These concentrations were 0.01, 0.05, 0.1, 0.5, 1 and 10 ppm. The following subsection presents the results and parameters used during the test and validation steps.

The evaluation of the samples with the sensors was performed by following the pre-treatment and square-wave voltammetry settings shown in Table 1 and Table 2, respectively.

**TABLE 1. Pre-treatment settings used in the SWV experiments.**

| Parameter       | Value   |
|-----------------|---------|
| E condition     | 0.65V   |
| t condition     | 45s     |
| E deposition    | -1.0V   |
| t deposition    | 240s    |

**TABLE 2. Square Wave Voltammetry settings used in the experiments.**

| Parameter               | Value   |
|-------------------------|---------|
| t equilibrium           | 15s     |
| E begin                 | -1.0V   |
| E end                   | 0.0V    |
| E step                  | 0.005V  |
| Amplitude               | 0.02V   |
| Frequency               | 50Hz    |

Figure 3 shows the results after performing square wave voltammetry with the molybdenite sensor for each of the three heavy metal ions analyzed in this work as As, Pb, and Cd.

**IV. PROPOSED HEAVY METAL IONS CLASSIFICATION METHODOLOGY**

The proposed classification methodology for the data obtained after perform the SWV experiments consisted of different stages, including data arrangement, dimensionality reduction, classification, cross-validation, and tuning parameters. Each of these stages are described below.

**A. DATA ARRANGEMENT**

The data acquired after performing the SWV voltammetry must be arranged. Six different concentrations in ppm were measured for each heavy metal ion. A 2-dimensional matrix is obtained by each heavy metal ion. Multiple matrices are available at this point because of the number of the three metal ions to be classified. From this point of view and to allow the analysis, all the matrices are organized in a process known as unfolding, where all data are organized in a 2-dimensional matrix. Additional information about this step can be found in [30], [31]. The total acquired data is arranged as follows [16]:

$$X = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1J} \\ \vdots & \vdots & & \vdots \\ x_{i1} & x_{i2} & \cdots & x_{ij} \\ \vdots & \vdots & & \vdots \\ x_{I1} & x_{I2} & \cdots & x_{IJ} \end{pmatrix} \in \mathcal{M}_{I \times J}(\mathbb{R})$$

This matrix $X \in \mathcal{M}_{I \times J}(\mathbb{R})$, where $\mathcal{M}_{I \times J}(\mathbb{R})$ is the vector space of $I \times J$ matrices over $\mathbb{R}$, which contains information from $I \in \mathbb{N}$ number of experiments and $J \in \mathbb{N}$ current measurement points in the SWV experiment.

**B. DIMENSIONALITY REDUCTION**

1) T-disTriBuTed STOCHASTIC NEIGHBOR EMBEDDING (T-SNE)

The previous step provides a 2-dimensional matrix where all the information from the sensors is organized. However, because of the high-dimensional data in this matrix and because it is possible to work with a reduced version of this matrix to decrease the computational cost of the classifier algorithm, a dimensionality reduction step is required. For this step, the t-SNE [32] algorithm was used.

In order to describe the dimensionality reduction process carried out by the t-SNE method the following recap based on [33] is performed. For more detailed information the reader is referred to the original work of t-SNE [32].
of t-SNE is to reduce the dimensionality of a collection of high dimensional data points $X$:

$$X = \{x^1, \ldots, x^v\} \subset \mathbb{R}^D, \quad v, D \in \mathbb{N},$$

(2)

The result is a low-dimensional representation map $Y$:

$$Y = \{y^1, \ldots, y^v\} \subset \mathbb{R}^d, \quad d \in \mathbb{N},$$

(3)

In this case, the data with original dimension $D$ is transformed into a new low dimensional subspace of dimension $d$ where $d \ll D$. t-SNE works in general terms as follows: First, a probability distribution is built using a pair of data from the original dataset. Similar data have a higher probability of being selected. This is done calculating the pairwise similarities as follows:

$$p_{ji} = \frac{\exp\left(\frac{-||x^i - x^j||^2}{2\sigma_i^2}\right)}{\sum_{l \neq i}^{N} \exp\left(\frac{-||x^i - x^l||^2}{2\sigma_i^2}\right)}, \quad i, j = 1, \ldots, v, \quad i \neq j,$$

(4)

By symmetrizing the conditional probability in Equation (4), the joint probability is [33]:

$$p_{ij} = \frac{p_{ji} + p_{ij}}{2v}, \quad i, j = 1, \ldots, v, \quad i \neq j,$$

$$p_{ii} = 0.$$

Then, $q_{ij}$ representing the local structure of the data points in the low-dimensional space is stated as follows:

$$q_{ij} = \frac{1}{\sum_{k=1}^{v} \sum_{l \neq k}^{v} \frac{1}{||y^i - y^l||^2}}, \quad i, j = 1, \ldots, v, \quad i \neq j,$$

(5)

$$q_{ii} = 0.$$  

(6)

Subsequently, the probability distribution is defined, and a comparison of the probabilities is applied to reduce the divergence using the Cost Function $C$:

$$C = D_{KL}(P \parallel Q) = \sum_{i=1}^{v} \sum_{j=1}^{v} p_{ij} \log\left(\frac{p_{ij}}{q_{ij}}\right),$$

(7)

The t-SNE algorithm has proven its successful behavior in dimensionality reduction, data representation, and subspace learning tasks. This method allows the visualization of high-dimensional data. Among the main advantages of t-SNE are [33] the capability of reducing the tendency to crowd points in the center of the distribution, the use of Student’s t-distribution to find the similarity measure between two samples, and user simpler gradients to calculate its cost function with respect to the original stochastic neighbor embedding (SNE) algorithm [34]. A tuning parameter process in the t-SNE algorithm is performed. Two different parameters must be tuned i.e. perplexity parameter and target dimensions parameter.

### C. SUPERVISED MACHINE LEARNING CLASSIFICATION

1) SUPPORT VECTOR MACHINES (SVM)

Support vector machines is a supervised machine learning algorithm that aims to classify two different types of data; this type of classification is called binary.

For an initial dataset $\{(x_i, y_i)\}_{i=1}^{N}$ we define the $d$-dimensional data $x_i \in \mathbb{R}^d$ and the label $y_i \in \{-1, +1\}$ [35]. Ideally, between the two classes, a boundary or hyperplane can be found that can distinguish the region to which each class belongs. The hyperplane can be defined as:

$$h(x) = w^T x + b,$$

(8)

where $b$ is the bias term and $w$ is the so-called weight vector.

The optimal hyperplane can be found solving the following minimization problem:

$$\min_{w, b} \frac{1}{2} \|w\|^2 \quad \text{subject to} \quad h(x_i)y_i \geq 1, \ldots, N.$$  

(9)

If it is not possible to separate all the samples, and it is sought that at least most of them are separated. If the separation between classes is not clear a transformation to another space can be performed using the kernel trick [36]. Because SVM is a binary classification technique different strategies have been developed to solve multi-classification problems. In this work, the one vs. one [37] strategy is adopted to solve the three-class heavy-metal ion classification problem. A comparison of the behavior changing between the Gaussian and polynomial kernels was performed. In addition, tuning parameters of SVM classifier i.e. Kernel Scale and box constraint is carry out.

### D. CROSS VALIDATION

The validation process in the developed heavy metal ion classification methodology was executed via the leave-one-out cross-validation process (LOOCV). This procedure requires the use of different data for training and testing in an iterative manner. Specifically, the total number of iterations corresponded to the total number of experiments in the dataset. In each iteration, one experiment was defined as a test set, while the rest of the experiments were used to form the training set. LOOCV was executed to avoid overfitting. The stages of the developed heavy metal ion classification methodology are illustrated in figure 4.

### V. RESULTS AND DISCUSSION

A. DATA ARRANGEMENT AND T-SNE DIMENSIONALITY REDUCTION

The data arrangement resulting from the SWV tests was performed for the molybdenite sensor. For arsenic (As), there are 41 experiments, for lead (Pb) there are 42 experiments, and finally, for Cd, there are 35 experiments. The process used to process the data was as follows: 3 data matrices were created for each heavy metal of the following sizes described as follows:

- As = 41(rows) experiments $\times$ 301 columns (data points)
• Pb = 42 rows (experiments) × 201 columns (data points)
• Cd = 35 rows (experiments) × 201 columns (data points)

An illustrated 3D line series of each of the experiments performed in the three heavy metal ions are depicted in Figure 5. These plots explain the differences between each signal due to the changes in the concentration of the heavy metal in ppm.

Subsequently, to perform the classification process using an SVM supervised machine learning algorithm, a dimensionality reduction process was carried out using the t-distributed stochastic neighbor embedding (t-SNE) algorithm [32]. The perplexity parameter of the t-SNE algorithm is 5. For this dimensionality reduction, a target dimension \( d = 8 \) was set such that the new size of the feature matrices was as follows:

• As = 41 rows (experiments) × 8 columns (features)
• Pb = 42 rows (experiments) × 8 columns (features)
• Cd = 35 rows (experiments) × 8 columns (features)

These three matrices are subsequently concatenated to form a feature matrix \( X \) of size 118 rows × 8 columns. This feature matrix is used as an input to an SVM-supervised machine learning classifier algorithm.

B. TUNING PARAMETERS IN THE SVM AND T-SNE METHODS
Different kernels can be used in the SVM classifier in order to map the data in other space and identify different classification regions. In this work four different kernels are tested in order to find the best classification regions delimitated by the support vectors. The tested kernels were: Gaussian kernel with kernel scale (KS) = 2.80, another gaussian kernel with KS = 1.19, polynomial cubic kernel and finally polynomial quadratic kernel. The classification regions are illustrated in Figure 6, detailing the particular region and data in a two dimensional contour plot of the t-SNE dimensions 5 and 6. The colors of each class are defined below: As(red), Pb(white) and Cd(blue). The SVM algorithm with Gaussian kernel was selected with the kernel scale set to \( \sqrt{P} \), where \( P \) is the number of predictors. In this case, the number of predictors \( P \) is equal to eight; thus, the kernel scale has a value of \( \approx 2.80 \). In addition, a box constraint level \( c = 1 \) and a one-vs-one strategy for multiclass classification are used.

Following a sensibility study, the change in the target dimensions of the t-SNE algorithm is described. Figure 7 depicts the behavior of classification accuracy when the t-SNE target dimensions parameter changes. The best number of target dimensions was 5 and 14.

Similarly, a study that changed the perplexity parameter of the t-SNE method was performed. The total number of Cd samples was found to be 35. Therefore, the perplexity parameter varies between 5 and 35. Figure 8 shows the resulting behavior, evidencing a fluctuation between a minimum of 0.68 and a maximum of 0.9322 for classification accuracy. The highest value of 0.9322 was reached when the perplexity parameter \( p \) was equal to 5.

C. HEAVY METAL IONS CLASSIFICATION
As a result of the classification process, the confusion matrices shown in Figure 9 were obtained when carrying out the LOOCV cross-validation process. This cross-validation method was used because of the low number of existing samples for each class in the dataset [30]. Both confusion matrices in Figure 9 represent a classification accuracy of 93.22%. This is an important aspect to highlight in the methodology developed because the 3D scatter plot and the confusion matrix of the target dimension = 5 are shown on the left, while the results for the target dimension = 14 are shown on the right.
FIGURE 5. Three dimensional line series plot of each acquired signals for a) As, b) Pb and c) Cd heavy metal ions.

FIGURE 6. Contour plot of the SVM classification regions for four different kernel variants a) Gaussian kernel $K_S = 2.80$, b) Gaussian kernel $K_S = 1.19$, c) polynomial cubic kernel and d) polynomial quadratic kernel. This behavior was obtained when target dimensions were equal to 17.

FIGURE 7. Classification accuracy behavior with t-SNE target dimensions variation.

FIGURE 8. Accuracy behavior due to the perplexity parameter variation in t-SNE.

the target dimension parameter, resulting in different representations of the data by plotting their first three dimensions. For the case when target dimensions are equal to 5 (left), it is observed how the data from As and Pb are grouped toward the center while the Cd data are spread throughout the domain in the form of a sphere. In contrast, when the target dimensions are equal to 14 (right), the Pb and Cd data are located in an elongated set, and the As data appear perpendicular to them.

Detailing the confusion matrix found when target dimensions are equal to 5 (left), the 41 experiments of As are correctly classified in their entirety, while for Pb there are six errors and for Cd, there are two errors. In contrast, for the confusion matrix obtained when using target dimensions equal to 14 (right), the Pb class is entirely classified, while As and Cd exhibit seven and one error, respectively. These errors in arsenic detection could be due to the aggregation of MoS$_2$-2D layers during the drop deposition on the screen-printed electrode, which could reduce the sensitivity or the electron transfer for this metal, which is a very common problem in electrochemical detection of this particular ion [38].
Both confusion matrices represent a classification accuracy of 93.22%.

Parameter tuning was executed in the SVM classifier. To obtain the best results, the parameters box constraint (C) and Kernel Scale (KS) were tuned comparing their behavior in the following ranges: for C six different values starting from 0.01 and ending to 1000. The KS parameter varied from 1.19 to 10000 in six different values. The results of these tuning parameters are shown in the heatmap in Figure 10. As observed in the heatmap figure, the best classification
accuracy values were obtained when the kernel scale took values of 1.19. As the KS values increased, the classification accuracy decreased. In contrast, the behavior exhibited when changing the box constraint parameter oscillated because when $C = 0.01$, the accuracy was the worst. Then, when $C$ increases, the accuracy also increases until $C = 1$; then, when $C$ increases, the accuracy value decreases slightly and remains constant. Therefore, the best value of $C$ is equal to 1.

After all the tuning processes, the best classification accuracy of 98.31% represents only two errors of the 118 samples in the dataset. The confusion matrix shown in Figure 11 represents this result. It was obtained after using the SVM classifier with a Gaussian kernel with $KS = 1.19, C = 1$. The target dimensions in t-SNE were equal to five, and the perplexity parameter was equal to five.

### VI. CONCLUSION

Molybdenum disulfide from raw ore was successfully exfoliated in aqueous solution to transform the bulk particle in 2D form that it was capable of electrochemical detection of As$^{3+}$, Pb$^{2+}$ and Cd$^{2+}$. After SWV test the development of a heavy-metal ion classification methodology was necessary for a better understanding of the measurements obtained with this sensor. The three heavy metal ions were satisfactorily classified by obtaining 98.31% of accuracy in the classification which represents only a poor classification of two samples. This methodology uses square wave voltammetry as an electroanalytical technique to obtain a set of signals that were subsequently successfully processed using pattern recognition and machine learning techniques.

Signals from the experiments with As, Pb, and Cd using SWV present some differences that result in signals of different sizes. The multivariate analysis took advantage of a data arrangement to perform a dimensionality reduction stage that brought the acquired data to a common low-dimensional space. The t-SNE dimensionality reduction method showed a favorable behavior when grouping the data of each class correctly, as well as separating the data between classes. This reduced feature matrix was used as the input for the SVM machine-learning classifier. A leave-one-out cross-validation procedure was used to avoid overfitting and to find the final confusion matrix in the classification process.

A tuning parameter procedure was performed to find the best t-SNE and SVM parameters in the classification process. In particular, the perplexity t-SNE parameter exhibited oscillatory behavior in terms of the classification accuracy. Perplexity expresses a smooth measure of the effective number of neighbors in the dimensionality reduction process. The results indicate that when the perplexity parameter was equal to five, the best accuracy was achieved. In contrast, the variation in the target dimensions t-SNE parameter also affects the average accuracy. As the target dimensions increase, the average accuracy also increases. The maximum accuracy value was reached when the target dimensions were equal to five. The methodology achieved a high classification accuracy of 98.31%. The above is evidence of the correct behavior of the methodology used to classify heavy metal ions.

In future work, it will be desirable to analyze other types of heavy metals and develop a portable device that can be used in situ for the detection and classification of heavy metal ions. In terms of data processing, the use of deep learning will be explored to solve heavy-metal ion classification tasks.

### ACKNOWLEDGMENT

Jersson X. Leon-Medina would like to thank Prof. Miryam Rincón Joya from the Department of Physics, National University of Colombia and Prof. Leydi Julieta Cardenas Flechas, who is currently a Ph.D. Student, for their introduction to the electronic tongue sensor array field of research.

### CONFLICTS OF INTEREST

The authors declare no conflict of interest.

### REFERENCES

[1] L. M. Nollet and F. Toldrá, *Handbook Analysis Edible Animal by-Products*. Boca Raton, FL, USA: CRC Press, 2019.

[2] G. Zhang, Z. Pan, X. Hou, X. Wang, and X. Li, “Distribution and bioaccumulation of heavy metals in food web of Nansi lake, China,” *Environ. Earth Sci.*, vol. 73, no. 5, pp. 2429–2439, Aug. 2014, doi: 10.1007/s12665-014-3592-z.

[3] J. Cheng, X. Zhang, Z. Tang, Y. Yang, Z. Nie, and Q. Huang, “Concentrations and human health implications of heavy metals in market foods from a Chinese coal-mining city,” *Environ. Toxicol. Pharmacol.*, vol. 50, pp. 37–44, Mar. 2017, doi: 10.1016/j.etap.2017.01.011.

[4] J. A. Buleli, S. Amin, S. I. Haider, M. I. Bhamer, and A. R. Solangi, “A review on detection of heavy metals from aqueous media using nanomaterial-based sensors,” *Environ. Sci. Pollut.*, vol. 8, pp. 1–9, Feb. 2020.

[5] S. Sawan, R. Maalouf, A. Errachid, and N. Jaffrezic-Renault, “Metal and metal oxide nanoparticles in the voltammetric detection of heavy metals: A review,” *Trac-Trend Anal. Chem.*, vol. 4, Aug. 2020, Art. no. 116014.

[6] M. Jin, H. Yuan, B. Liu, J. Peng, L. Xu, and D. Yang, “Review of the distribution and detection methods of heavy metals in the environment,” *Anal. Methods*, vol. 12, no. 48, pp. 5747–5766, 2020.

[7] J. Wang, *Analytical Electrochemistry*. Hoboken, NJ, USA: Wiley, 2006, doi: 10.1002/047190303.

[8] A. Escarpa, “Food electroanalysis: Sense and simplicity,” *Chem. Rec.*, vol. 12, no. 1, pp. 72–91, Nov. 2011, doi: 10.1002/crcr.201100033.

[9] J.-J. Ye, C.-H. Lin, and X.-J. Huang, “Analyzing the anodic stripping square wave voltammetry of heavy metal ions via machine learning: Information beyond a single voltammetric peak,” *J. Electroanal. Chem.*, vol. 872, Sep. 2020, Art. no. 113934, doi: 10.1016/j.jelechem.2020.113934.

[10] F. Pozo, Y. Vidal, and Ó. Salgado, “Wind turbine condition monitoring strategy through multway PCA and multivariate inference,” *Energies*, vol. 11, no. 4, p. 749, 2018, doi: 10.3390/en11040749.

[11] J.-J. Ye, C.-H. Lin, and X.-J. Huang, “Analyzing the anodic stripping square wave voltammetry of heavy metal ions via machine learning: Information beyond a single voltammetric peak,” *J. Electroanal. Chem.*, vol. 872, Sep. 2020, Art. no. 113934, doi: 10.1016/j.jelechem.2020.113934.

[12] J. Qiu, Q. Wu, G. Ding, Y. Xu, and S. Feng, “A survey of machine learning for big data processing,” *EURASIP J. Adv. Signal Process.*, vol. 2016, no. 1, pp. 1–16, May 2016, doi: 10.1186/s13634-016-0355-x.

[13] A. D. Tibahdija, L. E. Mujeica, and J. Rodellost, “Damage classification in structural health monitoring using principal component analysis and self-organizing maps,” *Struct. Control Health Monitor.*, vol. 20, no. 10, pp. 1303–1316, Dec. 2012, doi: 10.1002/stc.1540.

[14] S. L. France and U. Akkucuk, “A review framework, and R toolkit for exploring, evaluating, and comparing visualization methods,” *Viz. Comput.*, vol. 37, no. 3, pp. 457–475, Feb. 2020, doi: 10.1007/s00371-020-01817-5.
[15] L. Zhang and F.-C. Tian, “A new kernel discriminant analysis framework for electronic nose recognition,” Anal. Chim. Acta, vol. 816, pp. 8–17, Mar. 2014, doi: 10.1016/j.aca.2014.01.049.

[16] J. X. Leon-Medina, L. J. Cardenas-Flechas, and D. A. Tibbondua, “A data-driven methodology for the classification of different ligands in artificial taste recognition applications with a pulse voltammetric electronic tongue,” Int. J. Distriht Sensor Netw., vol. 15, no. 10, Oct. 2019, Art. no. 15504771988160, doi: 10.1177/15504771981601.

[17] J. X. Leon-Medina, W. A. Pineda-Mu noz, and D. A. T. Burgos, “Joint distribution adaptation for drift correction in electronic nose type sensor arrays,” IEEE Access, vol. 8, pp. 134443–134421, 2020.

[18] M. Li, H. Hou, I. Al-Ogaidi, and N. Wu, “Nanostructured sensors for detection of heavy metals: A review,” ACS Sustain. Chem. Eng., vol. 1, no. 7, pp. 713–723, 2013, doi: 10.1021/sc30019a.

[19] N. Ruecha, N. Rodthongkum, D. Cate, J. Volckens, O. Chailapakul, and C. Henry, “Sensitive electrochemical sensor using a graphene–polyaniline nanocomposite for simultaneous detection of Zn (II), Cd (II), and Pb (II),” Anal. Chim. Acta, vol. 874, pp. 8–40, May 2015.

[20] J. Ping, Y. Wang, J. Wu, and Y. Ying, “Development of an electrochemically reduced graphene oxide modified disposable bismuth film electrode and its application for stripping analysis of heavy metals in milk,” Food Chem., vol. 151, pp. 65–71, May 2014, doi: 10.1016/j.foodchem.2013.11.026.

[21] A. Waheed, M. Mansha, and N. Ullah, “Nanomaterials-based electrochemical detection of heavy metals in water: Current status, challenges and future direction,” TrAC Trends Anal. Chem., vol. 105, pp. 37–51, Aug. 2018, doi: 10.1016/j.trac.2018.04.012.

[22] Y. Lu, X. Liang, C. Niyungeko, J. Zhou, J. Xu, and G. Tian, “A review of the identification and detection of heavy metal ions in the environment by voltammetry,” Talanta, vol. 178, pp. 324–338, Feb. 2018, doi: 10.1016/j.talanta.2017.08.033.

[23] N. Liu, G. Zhao, and G. Liu, “Coupling square wave anodic stripping voltammetry with support vector regression to detect the concentration of lead in soil under the interference of copper accurately,” Sensors, vol. 20, no. 23, pp. 1–19, 2020.

[24] X. Gan, H. Zhao, K.-Y. Wong, D. Y. Lei, Y. Zhang, and X. Quan, “Covalent functionalization of MoS2 nanosheets synthesized by liquid phase exfoliation to construct electrochemical sensors for Cd (II) detection,” Talanta, vol. 182, pp. 38–48, May 2018, doi: 10.1016/j.talanta.2017.01.059.

[25] J. Dai, L. Yao, X. Gao, S. Bai, X. Chen, L. Li, J. Song, and H. Yang, “To achieve ultrasensitive electrochemical detection of mercury ions employing metallic IT-MoS2 nanosheets,” Electrochimica Acta, vol. 355, Sep. 2020, Art. no. 136800, doi: 10.1016/j.electacta.2020.136800.

[26] J. Wu, J. Dai, Y. Shao, M. Cao, and X. Wu, “Carbon dot-assisted hydrothermal synthesis of flower-like MoS2 nanospheres constructed by few-layered multiphase MoS2 nanosheets for supercapacitors,” RSC Adv., vol. 6, no. 81, pp. 77999–78007, 2016, doi: 10.1039/c6ta05074h.

[27] Z. Li, X. Meng, and Z. Zhang, “Recent development on MoS2-based photocatalysis: A review,” J. Photochem. Photobiol., C: Photocem. Rev., vol. 35, pp. 39–55, Jun. 2018, doi: 10.1016/j.jphotochemrev.2017.12.002.

[28] E. Varrla, C. Backes, K. R. Paton, A. Harvey, Z. Gholamvand, J. McCauley, and J. N. Coleman, “Large-scale production of size-controlled MoS2 nanosheets by shear exfoliation,” Chem. Mater., vol. 27, no. 3, pp. 1129–1139, Jun. 2015, doi: 10.1021/cm5044864.

[29] J. Wang, Analytical Electrochemistry. Hoboken, NJ, USA: Wiley, 2000, doi: 10.1002/0471228230.

[30] J. X. Leon-Medina, M. Anaya, F. Pozo, and D. Tibbondua, “Nonlinear feature extraction through manifold learning in an electronic tongue classification task,” Sensors, vol. 20, no. 17, p. 4834, Aug. 2020.

[31] J. Vitola, F. Pozo, D. Tibbondua, and M. Anaya, “A sensor data fusion system based on k-Nearest neighbor pattern classification for structural health monitoring applications,” Sensors, vol. 17, no. 2, p. 417, Feb. 2017.

[32] L. Van der Maaten and G. Hinton, “Visualizing data using t-SNE,” J. Mach. Learn. Res., vol. 9, no. 11, pp. 2579–2605, 2008.

[33] D. Agis and F. Pozo, “A frequency-based approach for the detection and classification of structural changes using t-SNE,” Sensors, vol. 19, no. 23, p. 5097, Nov. 2019.

[34] G. Hinton and S. T. Roweis, “Stochastic neighbor embedding,” in Proc. Adv. Neural Inf. Process. Syst., vol. 15, Dec. 2002, pp. 833–840.

[35] Y. Vidal, G. Aquino, F. Pozo, and J. E. M. Gutiérrez-Arias, “Structural health monitoring for jacket-type offshore wind turbines: Experimental proof of concept,” Sensors, vol. 20, no. 7, p. 1855, Mar. 2020.

[36] B. Scholkopf and A. J. Smola, Learning With Kernels: Support Vector Machine, Regularization, Optimization, Beyond (Adaptive Computation and Machine Learning Series). Cambridge, MA, USA: MIT Press, 2018.

[37] E. L. Allwein, R. E. Schapire, and Y. Singer, “Reducing multiclass to binary: A unifying approach for margin classifiers,” J. Mach. Learn. Res., vol. 1, pp. 113–141, Sep. 2001.

[38] A. Salimi, H. Mamkhezri, R. Hallaj, and S. Soltanian, “Electrochemical detection of trace amount of arsenic (III) at glassy carbon electrode modified with cobalt oxide nanoparticles,” Sens. Actuators B, Chem., vol. 129, no. 1, pp. 246–254, Jan. 2008, doi: 10.1016/j.snb.2007.08.017.

JERSON X. LEON-MEDINA received the B.S. degree in electromechanical engineering from the Universidad Pedagógica y Tecnológica de Colombia, in 2013, the M.Sc. degree in mechanical engineering and the Ph.D. degree in mechanical and mechatronics engineering from the Universidad Nacional de Colombia, Sede Bogotá, in 2016 and 2021, respectively. He has been linked to the research sector, belonging to the young researchers program with the “Departamento Administrativo de Ciencia, Tecnología e Innovación de Colombia–Colciencias,” from 2013 to 2015. From 2016 to 2018, he was a Professor with the Mechanical Engineering Faculty, Universidad Santo Tomas Seccional Tunja, Colombia. He is a member of the Research Group “Grupó de Investigación en Electrónica de Alta Frecuencia y Telecomunicaciones–CMUN,” Universidad Nacional de Colombia, Sede Bogotá. He is the author and coauthor of more than 30 publications in journals, conferences, and chapter books. His research interests include pattern recognition and signal processing in sensor array-type electronic noses and tongues. He is part of the International Society for Structural and Multidisciplinary Optimization–ISSMO. He is a member of the IEEE Instrumentation and Measurement Society (IMS).

DIEGO A. TIBADUIZA received the B.S. and M.S. degrees in electronics engineering from the Universidad Industrial de Santander, Bucaramanga, in 2003 and 2006, respectively, and the Ph.D. degree in earthquake engineering and structural dynamics from the Universitat Politècnica de Catalunya, Barcelona, Spain, in 2013. He is currently a Professor with the Department of Electrical and Electronics Engineering, Universidad Nacional de Colombia, Bogotá, Colombia, and currently works as the Program Coordinator of the Electronics Engineering Program. He joined the Department of Electrical and Electronics Engineering, in November 2016. Previously, he worked with Universidad Santo Tomas, Bogotá, the Universidad Autónoma de Bucaramanga, and the Universidad Industrial de Santander. He received the award of the best researcher from the Universidad Santo Tomas, in 2015, the Kobori Prize for Best Paper, in 2013, by the Structural Health Monitoring journal, International Journal (Wiley) and has been a Senior Researcher from the Ministry of Science of Colombia (Colciencias), since 2017. He is currently a Guest Editor for the Sensors (MDPI) journal and has collaborations as guest editors for other special issues and collections in different journals. He has published more than 70 papers in journals, conferences, books, and chapter books. His research interests include pattern recognition, structural health monitoring, artificial intelligence, control systems, robotics, data analysis, and signal processing.
JUAN C. BURGOS received the B.S. degree in chemical engineering from the Universidad Industrial de Santander, Bucaramanga, Colombia, in 2007, and the Ph.D. degree in materials science and engineering from Texas A&M University, College Station, TX, USA, in 2014. He is currently an Assistant Professor with the Chemical Engineering Program, Universidad de Cartagena, Cartagena, Colombia. He completed three postdoctoral fellowships with Texas A&M University, Pontificia Universidad Javeriana, Bogotá, Colombia, and the Universidad de Cartagena, before joining the School of Engineering, Universidad de Cartagena, in 2018. In 2019, he was granted a full scholarship within the Queen Elizabeth Scholars Program to participate in a multidisciplinary research project on climate change at Ryerson University, Canada, as a Visiting Early Career Professor. He has published a book chapter and more than 20 technical papers in scientific journals and conferences, and has served as a Reviewer for the Journal of Physics D: Applied Physics and SN Applied Sciences. His research interests include computational characterization of materials, surfaces and interfaces, adsorption phenomena, electron properties of materials, electrochemistry, sensors, and heterogeneous catalysis.

MARTHA CUENCA received the B.S., M.S., and Ph.D. degrees in chemical engineering from the Universidad Nacional de Colombia, Bogotá, in 2003, 2005, and 2015, respectively. She is currently an Assistant Professor with the Program of Chemical Engineering, Universidad de Cartagena, Cartagena, Colombia. She joined the Program of Chemical Engineering, in January 2018. Previously, she worked with the Pontificia Universidad Católica de Valparaíso, Chile, the Free University of Bozen, Italy, the Universidad de Bogotá Jorge Tadeo Lozano, Bogotá, Colombia, Universidad Nacional Abierta y a Distancia, Bogotá, and Quala S.A., Bogotá. She has been a Junior Researcher from the Ministry of Science and Technology of Colombia (Minciencias), since 2017. She is currently a Reviewer of the International Journal of Food Research. She has published more than ten technical papers in journals and conferences. Her research interests include functionalized materials, ionic liquids, semiconductors, nanomaterials, graphene, photocatalysis, and sensors.

DREIDY VASQUEZ received the bachelor’s degree in materials engineering from Simón Bolívar University, Venezuela, in 2005, and the Ph.D. degree in materials science and technology from the Politecnico di Torino, Italy, in 2012. She was a Postdoctoral Fellow with the Pontificia Universidad Católica de Valparaíso, Chile, from 2012 to 2015. Since 2017, she has been a Full Professor with the Chemical Engineering Department and the Laboratory of Advanced Materials and Nanomaterials, Pontificia Universidad Católica de Valparaíso. She has published more than ten technical papers in journals and conferences. Her research interests include functionalized materials, ionic liquids, semiconductors, nanomaterials, graphene, photocatalysis, and sensors.

* * *