Food profiling using innovative voltammetric metallic-glassy carbon electrodes evaluated by clustering metrics

Szymon WÓJCICIK
AGH University of Science and Technology

Valentin MIRCESKI
University of Lodz

Bogusław BAŚ
AGH University of Science and Technology

Małgorzata JAKUBOWSKA (jakubows@agh.edu.pl)
AGH University of Science and Technology

Research Article

Keywords:

Posted Date: December 2nd, 2022

DOI: https://doi.org/10.21203/rs.3.rs-2318649/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License
Abstract

This work proved that the quadruple-disk electrodes - iridium-platinum, platinum-glassy carbon, and iridium-glassy carbon can successfully work as a single sensor and act as voltammetric electronic tongue in food profiling. Current samples obtained with square-wave voltammetry (SWV) were additionally interpreted by a novel technique termed double-sampled differential square-wave voltammetry (DSSWV), which allowed for increasing the resolution and extracting the full, hidden information available in the signals. Four clustering validity indices: Calinski – Harabasz index, Davies – Bouldin index, Silhouette index and gap statistics were applied to select the optimal sensor and the current samples interpretation strategy. These four coefficients indicated the best iridium-platinum electrode combined with overall cathodic potential modulation using the technique of DSSWV. Correct clustering of objects, i.e., samples of five different isotonic beverages from one producer, was confirmed using principal component analysis and polar dendrogram. At the same time, it has been shown that in some variants of measurements and signal interpretation, it is not possible to correctly group the samples.

1. Introduction

People of all ages consider a healthy lifestyle, combined with physical activity and a healthy diet, to be important. Gymnastics in various forms, running, cycling or other activities cause the loss of water from the body and a number of valuable minerals. Water and mineral salts are essential for the proper functioning of the human body, including most of all the cardiovascular and nervous systems. Adequate fluid intake is essential and recommended during physical activity to maintain water and electrolytes for body homeostasis (Lewis et al. 2013). Sport drinks (Sierra-Rosales et al. 2019) are flavoured and functional non-alcoholic drinks that are designed to promote rehydration and the rapid metabolic conversions that are required during exercise. Sport drinks can be classified depending on the concentrations of minerals with respect to the human body, named as hypotonic, isotonic, or hypertonic when the level is lower, similar and higher, respectively, than the found in the human body.

Consumption of isotonic drinks expands significantly due to their increased use by sportsmen and athletes what causes the development of the sports drinks market. In general, there is a lot of interest, the popularity and consumption of functional drinks. The global isotonic drinks market was estimated at 1870.9 million $ in 2020, and is projected to reach 7785.0 million $ by 2028 (https://www.alliedmarketresearch.com/isotonic-drinks-market). It is segmented on the basis of region of the world, the form of the drink, the type of packing and the distribution channel. In terms of form, it can be sold as a powder in sachets, packed in a ready-to-drink bottles, metal cans or pouches. The isotonic beverage contains water and minerals, mainly sodium, potassium, and magnesium. Its composition also includes carbohydrates that provide the body with energy necessary for exercise and increase the rate of regeneration after physical activity. An isotonic drink, unlike an energy drink, has a low sugar content and is not saturated (or slightly saturated) with carbon dioxide. It is well absorbed from the gastrointestinal tract and does not burden the stomach. Its role relies primarily in quenching thirst, replenishing water,
vitamins, and minerals such as sodium, potassium, calcium and magnesium ions (electrolytes) during physical activity because of a balanced level of electrolytes and appropriate osmotic pressure.

Beverages are considered isotonic when they have an osmotic concentration similar to that found in body fluids (Stasiuk and Przybyłowski 2017). This feature allows quick absorption of the drink after ingestion, improving the performance of athletes and preventing muscle fatigue. An isotonic drink should have a sodium concentration ranging from 460 to 1150 mg L$^{-1}$ and the appropriate content of carbohydrates, i.e., glucose, fructose or sucrose, should be 4–8%. Other ingredients typically present in isotonic drinks are food colourings, preservatives, artificial sweeteners, ascorbic acid, B-group vitamins, A and E vitamins, folic acid, as well as both essential and toxic minerals in trace concentrations, also caffeine and taurine (Leśniewicz et al. 2016; Styburski et al. 2020).

During the processing of isotonic beverages, flavourings and synthetic dyes with sensory characteristics similar to fruit are used, which do not add any nutritional value to food products (Silva et al. 2022). Artificial food colorants are important additives because they have high stability to light, oxygen and pH, and their addition does not change the properties of the product. However, the problem of the toxicity of synthetic dyes and the influence of them on health is significant (Carocho, M., Barreiro, M. F., Morales, P., & Ferreira 2014). Almost all the approved colorants used by the food industry have been related to health issues, like skin irritations and bronchial constriction, hyperactivity, anxiety, allergic reactions in sensitive individuals and asthma (Ferreira et al. 2016). They are especially dangerous to the health of children who also eagerly eat food containing artificial dyes, in relatively greater amounts than adults (Stevens et al. 2014). Considering the health risks associated with these dyes, development is observed of reliable, sensitive, selective and robust analytical methods to analyse these colorants, to help mitigate their dangerous effects. Dyes used to food colouring can be successfully determined by voltammetric techniques using a range of different working electrodes (Shetti et al. 2018; Bessegato et al. 2019; Penagos-Llanos et al. 2019; Silva et al. 2020). Among the electrode materials used in these measurements, carbon (glassy carbon, carbon paste) modified electrodes are the most widely reported, while graphene oxide and carbon nanotubes are the most relevant modifiers (Bessegato et al. 2019). In particular, the work (Sierra-Rosales et al. 2019) concerns the voltammetric determination of synthetic dyes in isotonic drinks, which were performed with the use of a MWNCT-modified glassy carbon electrode.

The development of electrochemical methods and devices based on innovative sensor platforms for the to distinguish groups of objects with a complex organic matrix and different characteristics is a major challenge in the area of electroanalytical and electrochemical analysis. Square wave voltammetry being more powerful to discriminate Faradaic from non-Faradaic current (Mirceski et al. 2013), is typically applied for qualitative and quantitative analysis. It can also be used during data collection when research strategies referred to as voltammetric electronic tongue (VET) are implemented. In the SWV technique, an increase in the current response is observed with the increase in the parameters of the square-wave potential modulation, its frequency $f$ of the pulses defined as $f = 1/(2t_p)$, where $t_p$ is duration of a single pulse, the height of the step of the underlaying staircase potential $E_s$, and the height of the potential
pulses termed as square-wave amplitude $E_{sw}$. However, a compromise is being sought between the achieved Faradaic current amplification and the accompanying peak shape distortion and the non-Faradaic current amplification.

This work presents three quadruple-disk electrodes, bimetallic and glassy carbon-metallic, each working independently as a voltammetric electronic tongue. The Theory chapter presents a short introduction to machine learning in a modern perspective and more broadly refers to unsupervised learning methods, including a detailed description of clustering validity indices carried out with the use of these algorithms. The next section, Experimental, describes the apparatus, the design of the innovative quadruple-disk working electrodes, the measurement procedure and the test objects, which were samples of commercially available isotonic drinks. The results are presented in such a way that the directly recorded current samples are shown and then the voltammograms created using two strategies, the typical square-wave approach and the double-sampled differential square-wave voltammetry technique (Mirceski et al. 2020). The purpose of the chemometric calculations used, i.e., the k-means algorithm, cluster analysis and principal component analysis, was to automatically indicate which working electrode, overall potential direction (anodic or cathodic), and the method of interpreting direct current samples is the recommended strategy for profiling isotonic drinks.

2. Theory

Currently available and widely used machine learning and deep learning strategies are becoming more and more complex and advanced (Goodfellow et al. 2017; Andrew Ng 2018; Janet and Kulik 2020; Wei et al. 2020; Géron 2022). They provide many possibilities, and various trends and innovations are constantly emerging. Currently, these innovations, among others, include: hyperautomation - thanks to machine learning technology, companies use the automation of numerous repetitive processes which are based on huge volumes of data, to enhance the speeds, accuracy, and reliability of the work; Internet of Things (IoT) – this technology connects numerous small devices across a network and allows for seamless communication between each other, making them a lot more intelligent, cyber security – uses machine learning to identify cyber threats, to fight cybercrime, and enhance the current antivirus software; no-code ML – such platforms allow companies to work without requiring an engineer or developer, allow the user with less technical skills to create their own tools with a drag-and-drop interface, which reduces costs and time; deep learning - further intensive development of this technology using multilayer neural networks with different architectures, in particular for modelling multidimensional data (>$2D$), used for applications that require image recognition, autonomous movement, voice interaction; semi- and self-supervised learning - automate the traditional manual process, like data labeling; reinforcement learning - allows software to find solution by interacting with the environment, uses a reward and punishment system, and lets the machine learn by experimenting with a potential path and then deciding which one would have the best reward and is effective.

One of the constantly updated strategies is unsupervised machine learning, which does not require humans to intervene since the algorithms are designed to identify data groupings and patterns that are
unseen (Tripathy et al. 2021). Labels are not used in the calculation. This type of learning is able to look at the data and identify similarities. Unsupervised machine learning uses clustering approaches, which mine data to find groupings. The advantages of such methods include solving the problem by learning the data and classifying it without any labels. It is very helpful at the initial stage of data analysis in finding patterns in data, i.e. subsets of similar objects, and dimensionality reduction can be easily accomplished. Unsupervised learning is is the perfect tool for data scientists, as it can help to understand raw data, and also find up to what degree the data are similar. This task is defined in this work. The problem was finding which of the sensors or data preparation methods provide the correct answer about the grouping of objects. Experimental data, which were the signals obtained in the voltammetric measurements, were input for the models. It was checked whether the samples without labels are correctly assigned to subsets that are known but not used in calculations. This approach allowed for the development of an overall correct research strategy which in many aspects used the knowledge about the experiments and a lot of domain information in the field of analytical chemistry. As a result, recommendations were formulated on the correct acquisition of data for supervised modelling.

K-means is an iterative clustering algorithm, which states that similar data points should be in a close neighbourhood in a data space (Blokdyk 2021). The value of k is the number of data points located near centroids in the data set. Object clustering uses measurement of the distance between points and the centroid. Various mathematical distance functions are typically used in calculations. Also, miscellaneous stop criteria are applied, e.g. no change in the membership of objects to clusters in the next iteration ends the calculations. The k-means clustering method is the simplest and most commonly used algorithm, but it is limited by the required number of clusters, which has to be pre-determined.

Hierarchical clustering allows for the formation of multiple clusters, which are distinct from each other, but the contents inside the cluster are highly similar to each other (Everitt et al. 2011). Important parameters of this algorithm are the distance function and the agglomeration procedure. In the first step, the distance matrix is calculated, which is the basis for further iterative clustering. The algorithm would treat each observation as a separate cluster. Then it would find two most similar clusters and merge them. This step continues iteratively until all clusters merge together. For visual representation of the clusters, a dendrogram would be formed. It would show the hierarchical similarity between the clusters.

In the unsupervised learning group, one of the most commonly used algorithms is Principal Components Analysis (PCA) (Jolliffe 2002). This strategy makes it possible to reduce the dimensionality of the data space by transforming the correlated input features into new, mutually orthogonal principal variables, with little loss of information. Typically, the first few major components describe a significant percentage of the information contained in the original data. A significant limitation of the number of variables enables objects to be visualized in a space with a smaller number of dimensions. When a limitation to two dimensions is sufficient, hidden relationships between samples and original variables can be observed in the plane. PCA is often used in voltammetry to qualitatively evaluate complex samples from the obtained signals.
Clustering evaluation

In the case of supervised methods, different coefficients of model quality assessment are typically used, i.e., for classification models accuracy, sensitivity, precision, specificity, f1-score, receiver operating characteristic curve (ROC), area under the ROC curve (AUC), confusion matrix, while for regression models it is a root mean squared error (RMSE) and the assessment of the predicted / measured relationship (Powers 2020). Numerical coefficients allow for unambiguous decisions regarding the usefulness of defined models. There is a need to define such measures to assess grouping efficiency in clustering methods.

Clustering validity indices (CVIs) are utilized to validate the clustering results and find the correct number of clusters in a dataset. The CVIs are intended to indicate the intensity of separability, compactness among clusters and the geometric structural knowledge of the dataset. The indicators that can be used to estimate the number of clusters include Calinski–Harabasz (CH) index, Davies–Bouldin (DB) index, Silhouette (S) index or gap statistics.

Let’s assume that dataset \( \{x_{ij}\}, i = 1,...,n, j = 1,...,p \) consists of \( p \) features measured on \( n \) independent observations. Below variable descriptions in CVIs definitions (formulae (1)-(4)) are listed:

\( n \) is total number of observations in dataset,
\( k \) is ideal number of clusters,
\( C_i \) is the \( i \)-th cluster,
\( n_i \) is number of objects in the cluster \( C_i \),
\( d(x, y) \) is distance between \( x \) and \( y \) (the most common choice is the Euclidean distance),
\( c \) is centroid of dataset,
\( c_i \) is centroid of the cluster \( C_i \).

Calinski and Harabasz Index

The CH index technique was proposed by Calinski and Harabasz (Calinski and Harabasz 1974) (also known as the Variance Ratio Criterion) to determine the ideal number of clusters \( k \) and is defined as (1):

\[
CH \ (k) = \frac{\sum_i n_i d^2(c_i, c) / (k - 1)}{\sum_i \sum_{x \in C_i} d^2 (x, c_i) / (n - k)} \tag{1}
\]

The score determines ratio of the sum of between-cluster dispersion and of within-cluster dispersion for all clusters (where dispersion is defined as the sum of distances squared). The CH index is a measure of how similar an object is to its own cluster (compactness) compared to other clusters (separation).
Compactness is estimated based on the distances from the data points in a cluster to its cluster centroid and separation is based on the distance of the cluster centroids from the global centroid. Higher value of CH index means the clusters are compact and well separated, although there is no acceptable cut-off value. Typically, \( k \) is selected which gives a peak or at least an abrupt elbow on the line plot of CH indices. But if the line is horizontal then there is no such reason to prefer one solution over others.

**Silhouette Index**

The Silhouette index \( S \) was introduced by Kaufman and Rousseeuw (Rousseeuw 1987) and was built to show graphically how well each element is categorized in a given clustering output. It is defined as (2):

\[
S(k) = \frac{1}{k} \sum_i \left( \frac{1}{n_i} \sum_{x \in C_i} \frac{b(i) - a(i)}{\max[a(i), b(i)]} \right) \tag{2}
\]

where

\[
a(x) = \frac{1}{n_i - 1} \sum_{y \in C_i, x \neq y} d(x, y)
\]

\[
b(x) = \min_{j, j \neq i} \frac{1}{n_j} \sum_{y \in C_j} d(x, y)
\]

The optimal number of clusters will be the value for which the \( S(k) \) is maximum. The Silhouette value is a measure of how similar an object is to its own cluster (compactness) compared to other clusters (separation). It can be used to study the separation distance between the resulting clusters. \( S \) index validates the clustering performance based on the pairwise difference of between and within-cluster distances. The Silhouette plot displays a measure of how close each point in one cluster is to points in the neighbouring clusters and thus provides a way to assess parameters like number of clusters visually.

**Gap Statistics**

Tibshirani et al. (Tibshirani et al. 2001) proposed a way to determine the ideal number of clusters \( k \) in a dataset by the gap statistic. The idea of the gap statistic (3) is to compare the total within-clusters sum of squares around the cluster centroid for various numbers of \( k \) with their expected values generated from a reasonable reference null distribution.

\[
Gap(k) = E^*\log(w_k) - \log(w_k)
\]

where

\[
w_k = \sum_i \frac{1}{2n_i} \sum_{x \in C_i} d^2(x, c_i)
\]
where $E_n^*$ denotes the expectation under a sample size $n$ from the reference distribution. The gap statistic measures the deviation of the observed $w_k$ value from its expected value under the null hypothesis. A value of $k$ that maximizes the gap statistic will be the estimate of the ideal cluster number. The basic idea of the gap statistics is to choose the number of $k$, where the biggest jump in within-cluster distance occurred, based on the overall behaviour of uniformly drawn samples. It could be the case that only a very slight reduction in within-cluster distance occurred.

**Davies–Bouldin Index**

The Davies-Bouldin (DB) index is one of the clustering algorithms evaluation measures (Davies, D.L. Bouldin 1979). It is most commonly used to evaluate the goodness of split by a k-means clustering algorithm for a given number of clusters. The Davies–Bouldin index is calculated as the average similarity of each cluster with a cluster most similar to it. The lower the average similarity is, the better the clusters are separated and the better is the result of the clustering performed. DB can be defined using Eq. (4):

$$
DB(k) = \frac{1}{k} \sum_{i} \max_{j,j\neq i} \left( \frac{1}{n_i} \sum_{x \in C_i} d(x, c_i) + \frac{1}{n_j} \sum_{x \in C_j} d(x, c_j) \right) / d(c_i, c_j)
$$

The index is defined as the average similarity between each cluster $C_i$ for $i=1,...,k$ and its most similar one $C_j$. A lower Davies-Bouldin index relates to a model with better separation between the clusters. This index signifies the average similarity between clusters, where the similarity is a measure that compares the distance between clusters with the size of the clusters themselves. Zero is the lowest possible score. Values closer to zero indicate a better partition. The computation of Davies-Bouldin is simpler than that of Silhouette scores. The index is solely based on quantities and features inherent to the dataset as its computation only uses point-wise distances.

**3. Experimental**

**3.1. Apparatus and software**

All electrochemical profiles of isotonic drinks were performed with the use of the M161 electrochemical analyser and the M164 three-electrode stand (MTM-ANKO, Poland), which allows current recording at 1 kHz. A typical three-electrode measuring system with a novel quadruple-disk working electrode, a double junction Ag|AgCl|KCl (3M) with silver wire as reference electrode, and platinum wire as an auxiliary electrode were used in the experiments.

The measurements were also performed with the use of EAQt software dedicated to the analyser, which enables measurements, data collection, and initial signal processing. In addition, in order to ensure the highest quality of results, EACfg software was used, compatible with the measurement system with
which data were recorded in an automated manner in a strictly defined time regime. The Mathworks Matlab R2021a was used to preprocess the data for analysis and to realize chemometric modelling.

3.2. Construction of the quadruple-disk electrodes

For years, the team has been working intensively to propose completely new designs of solid electrodes, the signal of which is used as a data source for chemometric modelling. The combination of a modified and improved experimental approach with the use of machine learning strategies fits in the modern development trends of science and technology. The effect of evaluating these activities through experiments with the use of objects with a complex organic matrix is immediately visible. We design new electrodes for profiling different samples, with the main assumption that a single composite sensor acts as a voltammetric electronic tongue. In this way, we want to avoid the serious inconvenience of using expensive multi-channel analysers or performing multiple consecutive recordings for one sample, on different working electrodes. The verification of this approach with the use of various research objects confirms the usefulness of the adopted strategy.

The new proposed designs are electrodes with a disk-shaped working surface, each of which has glassy carbon disks and metallic disks made of platinum or iridium. These electrodes were specially constructed for this work and are presented for the first time. *In situ* activated sensors are easy to make and a very simple design. The idea of its construction is basically similar to that of our other sensor, i.e., quadruple-disk iridium electrode (q-DIrE) (Wójcik and Jakubowska 2021).

Figure 1 shows construction details and photos of the new carbon-metallic electrodes designed specifically for this work and used in the experiments. They consist of one iridium and one glassy-carbon or one platinum and one glassy carbon wire (iridium wire: Alfa Aesar, Germany, 99.8%, No. 11430.BQ, CAS: 7439-88-5, $\phi = 0.5$ mm; platinum wire: Alfa Aesar, Germany, 99.95%, No. 43288.BU, CAS: 7440-06-4, $\phi = 0.5$ mm; GC rod: GoodFellow, England, 99.95%, ultrafine 1–5 µm, No. LS566678 AH, C-00-RD-000108, CAS: 7440-44-0, $\phi = 1.0$ mm). The length of the wires, which formed four working surfaces, was 3 mm. They were placed in holes prepared by drilling appropriate holes in a silver rod, which also constituted an electrical contact. The diameter of the silver rod (Mint of Poland) was 1.6 mm and its length was about 8 cm. The arrangement of wires placed in a special form has been secured with TRANSLUX D180 epoxy resin (AKSON, France). This material is characterized by high chemical stability. The next necessary step was to polish the surfaces of the working electrodes, which had previously been removed from the mold and that needed to be secured with resin. Emery paper and the next alumina slurry in which the particle sizes ranged from 0.05 to 1.0 µm (Buehler) were used for this purpose. Sensor cleaning was performed placing the sensors in an ultrasonic bath for 10 minutes. Fixing the surface of the working electrodes on the vertical wall of the body caused the gas bubbles, which are formed during the conditioning of the electrode and as a product of the electrode reaction, to quickly leave the solution. This resulted in high repeatability of signals and long usability of the sensor without the need for mechanical cleaning and preparation of its surface. This special design of the electrodes also allows for easy surface preparation by polishing when necessary. In many preliminary experiments, it was found that the use of mixed electrode materials and the proposed design of the sensors results in their multifaceted usefulness in
profiling samples with a complex organic matrix. The use of these electrodes does not require connection to multichannel analysers because of the common electrical contact, which is a silver rod.

3.3. Chemical reagents and samples

In this study, five isotonic drinks were used from the Polish company 4move, available in stores throughout Poland and more than 50 countries around the world. All drinks came from a "4move isotonic drink" line, which was purchased from the supermarket. The beverages have been selected to meet the difference in colour and taste. During the tests using undiluted samples, no supporting electrolyte was added because the amount of substances included in the isotonic drink ensured sufficient conductivity for these types of measurement. The beverage composition data are summarized in the Table 1.

<table>
<thead>
<tr>
<th>Isotonic drink</th>
<th>White</th>
<th>Yellow</th>
<th>Orange</th>
<th>Green</th>
<th>Blue</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flavour</td>
<td>Grapefruit</td>
<td>Lemon</td>
<td>Orange</td>
<td>Lime - mint</td>
<td>Multifruit</td>
</tr>
<tr>
<td>Composition</td>
<td>Water, sugar, maltodextrin, glucose, citric acid</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Acidity regulator</td>
<td>Sodium citrates, Potassium citrates</td>
<td>Sodium citrates</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aroma</td>
<td>Natural</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Preservatives</td>
<td>Potassium sorbate, sodium benzoate</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stabilizers</td>
<td>Arabic gum, esters of glycerol and plant resin</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sweeteners</td>
<td>Acesulfame K, aspartame</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vitamins</td>
<td>Niacin, vitamin B6, vitamin B12</td>
<td></td>
<td>Folic acid, biotin, pantothenic acid</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dyes</td>
<td>-</td>
<td>Tartrazine E102</td>
<td>Tartrazine E102, sunset yellow FCF E110</td>
<td>Tartrazine E102, brilliant blue FCF E133</td>
<td>Brilliant blue FCF E133</td>
</tr>
</tbody>
</table>

As a preliminary stage of the research on isotonic drinks, an attempt was made to investigate the effect of changing the concentration of the dye on the electrochemical profile of the isotonic drink. Therefore, considered dyes E102, E110 and E133 (Food Colours, Poland) were added to the white isotonic drink, which, according to the manufacturer's declaration, did not contain any dye. To ensure appropriate conditions for this type of experiments, dye solutions with a concentration of 1 g L⁻¹ were prepared using the tested white isotonic drink as a solvent. According to the regulation of the European Commission (European Parliament and Council 2011), there is no specific maximum concentration of food colorant additives in nonalcoholic beverages, and the amount of dye concentration allowed is defined as
satisfying. At the same time, according to the standards established by the Food and Agriculture Organization of the United Nations (Food and Agriculture Organization of The United Nations 2011), the daily intake of these dyes should not exceed 100 mg/kg of adult body weight.

### 3.4. Measuring procedure

A recently introduced new voltammetric technique, that is, double-sampled differential square wave voltammetry (DSSWV), was used to interpret the signals recorded according to the square wave pattern (Mirceski et al. 2019, 2020). It consists in subtracting the values of two current samples in the final segment of each step and enables a new look at the recorded data (Fig. 2).

This approach fits the general tendency to look for hidden information conveyed by signals, that is, information that is not conveyed by individual variables but is contained in relationships between variables. This is an innovative approach that is being developed on a large scale. An example of a typical algorithm that relies on hidden variables is the Partial Least Squares Regression (PLSR) algorithm. In the case of this algorithm, the model is defined using an optimized number of hidden variables. The application of DSSWV is also a step in this direction, using the properties of the voltammetric signal and the knowledge of the characteristics of the Faradaic and capacitive components in voltammetric experiments. The SWV parameters that were used were the following: potential range of -1000 to 1000 mV, SW frequency of potential change 12.5 Hz, 25 mV amplitude, and 2 mV step potential.

In the DSSWV interpretation, the problem is the low signal level, due to the fact that the current difference is calculated in the area of its low variability. Additionally, this approach results in a high level of noise. Therefore, in this work, two strategies are used to counteract these problems and their impact on the outcome of modelling using machine learning. First, to calculate the difference, averaged values of current samples were used, and a digital filtering algorithm was applied. In the device used in this work, the current sampling is performed at a current sampling frequency of 1 kHz. With the measurement parameters, after several optimization steps, the averaging was finally performed in such a way that the difference from the average of two sets of 5 measuring points (5 ms each) at the end of the pulse was calculated. The process of reducing the noise from such data by averaging the data points with their neighbours may be done using many techniques like simple moving average, weighted moving average, kernel smoother, etc. After testing various variants, digital filtration was performed using the *Gaussian-weighted moving average filter* (other name *Gaussian kernel smoother*) (Smith 2002). The kernels define the shape of the function that is used to take the average of the neighbouring points. The Gaussian kernel has the shape of the Gaussian curve. As seen in the Gaussian curve, the points around zero will be weighted higher and the farther points will be weighted lower. Filtration is performed in a window that moves along the signal at one point on the voltammogram. For each interval, the kernel function values are calculated and then the weighted average of data points is calculated weighted by kernel function values. It is also important to note that the Gaussian filter that is used is normalized, so weights sum to 1. With regards to how the endpoints are handled, namely, if a window has fewer valid points than specified by the window size, the window will be re-scaled so that the contributing coefficients sum to 1.
Smoothing attenuated sharp dips and spikes. The expected effect was obtained in relation to the recorded signals.

To maintain compatibility, the typical interpretation of the SWV signal was carried out on the basis of the averaged values of the last five current samples (5 ms) on each pulse. In this case, no digital filtering algorithm was used.

In general, the operating strategy was to record cyclic SWV voltammograms (i.e., by applying the overall SW potential modulation in both anodic and cathodic direction) for five isotonic drinks, using three different working electrodes. Signals recorded at 1 kHz were interpreted according to the typical SWV scheme and the new DSSWV approach. Furthermore, different clustering assessment coefficients were used to indicate which electrode, overall potential modulation (anodic or cathodic), or the method of interpretation of current samples (SWV/DSSWV) have the greatest potential in differentiating the tested isotonic drinks. The optimization results were verified by hierarchical cluster analysis (polar dendrogram) and principal component analysis (PCA). Figure 3 shows schematically the applied strategy.

4. Results And Discussion

Examples of measured curves are shown in Figs. 4 (overall SW anodic potential direction) and Fig. S1 (cathodic direction). These signals are directly recorded current samples (raw data) using the SWV technique (with current sampling frequency of 1 kHz), without averaging the measured values and without using any signal processing procedure, in particular digital filtering or baseline correction. These current-data form the basis for all further considerations and calculations. Figures 4a and 4b show the anodic signals obtained for three different quadruple-disk bimetallic and carbon-metallic electrodes. They were registered for one of the tested isotonic agents, i.e., yellow isotonic. Significant differences between the signals can be observed (including, in particular, the signal of the bimetallic electrode that differs from the others), and therefore this picture initially illustrates the need to compare different sensors for profiling purposes. The initial tests were then carried out to establish the ability to differentiate individual samples (types of isotonic drinks). It was confirmed that the applied working electrodes can be useful for the intended purpose; however, the differences in the waveforms are not always clearly visible. An example of the anodic signals of a bimetallic electrode for five isotonic drinks is presented in Figs. 4c and 4d. The responses recorded in the overall cathodic potential direction were compared in an identical manner and similar preliminary observations were confirmed (Fig. S1).

The description of the tested samples (Experimental section) shows that individual drinks differ in colour, which is important from the point of view of the profiling task. Therefore, before applying various machine learning unsupervised algorithms, the voltammetric response to the addition of dyes important in this work, i.e., E102, E110, E133, was verified. The experiment was designed in such a way that the base solution was colourless isotonic white and the prepared dye solutions were successively added to it. All three working electrodes were used, and similar observations were obtained. A detailed description of the experiments with qDptGCE was only presented.
The signals recorded on the carbon-platinum electrode, using the SWV and DSSWV interpretation, are shown in Fig. 5. It was found that the addition of each dye affects the SWV and DSSWV signals and the curve differs from the background signal, i.e., the white isotonic signal. For the E102 dye, for voltammograms recorded in anodic potential direction, net SW voltammetric peak was observed at -22 mV and 876 mV, and for DSSWV at -46 mV, 488 mV and 880 mV. For the E110 dye, a net SW voltammetric peak at 74 mV and 706 mV was observed, and for DSSWV at -20 mV, 448 mV and 734 mV. For the dye E133, a net SW voltammetric peak was observed at 732 mV and for DSSWV at 720 mV and 900 mV. These observations were another confirmation that different isotonic colours can be distinguished using the proposed approach. The three dyes considered also have an effect on the signals recorded in the overall cathodic potential direction at the carbon-platinum electrode (Fig. S2). It should be noted that when interpreting DSSWV, each dye gives a signal different from the background, while in the case of SWV the response of E102 basically coincides with the background (in the concentration range typical for food products). The dye tests performed were the basis for making a decision on the registration of cyclic SW voltammograms for the purposes of profiling and comparing the constructed electrodes.

The profiling of isotonic drinks was performed independently on each of the three working electrodes to indicate which of them had the greatest potential to distinguish between five different beverages. The tests were performed without the supporting electrolyte by immersing the sensor in the measuring vessel containing the drink directly taken from the bottle (manufacturer's packaging). The registrations were made using a cyclic procedure and with the use of control software, implemented especially for this purpose. Automatic recording of successive curves ensured perfectly comparable intervals between individual signals; therefore, the manual approach did not introduce an additional factor of variability. Numerous comparative studies have shown that this type of automatic controlled recording plays an important role in experiments with solid electrodes. The voltammograms recorded using individual electrodes are shown in Figs. 6–7 and S3 – S4. The net-voltammograms recorded in anodic and cathodic potential directions are presented separately, as they were processed and interpreted independently throughout the project. There was no need to concatenate the signals, as the voltammograms recorded in a one potential direction turned out to be sufficient to distinguish the samples. Figures 6 and 7 show the anodic signals using two current sample interpretation strategies, i.e., SWV and DSSWV. At this stage, it is possible to observe significant and decisive differences between the shape of the curves in these two approaches, obtained from exactly the same current samples measured every 1 ms. The signal level of the DSSWV technique is much lower, but it provides much more information about the sample. It is the fulfilment of a significant challenge posed to analytics to obtain as much information as possible about the objects studied in the least possible number of simple experiments. The aim was also to confirm whether the DSSWV technique, working in conjunction with unsupervised machine learning algorithms, would prove more useful. Next, Figs. S3 and S4 show the same effect for cathodic signals. A much more complex, i.e., informative signal was observed when using the DSSWV technique.

As a result of the experiments performed, 12 data matrices were obtained, which were processed and interpreted independently. The datasets were created from signals recorded with the use of three working electrodes, two techniques for interpreting samples of direct current sampling, recorded in both anodic
and cathodic potential directions. Each matrix had a dimension of 25 x 1000, where 25 corresponds to 5 tested isotonic drinks tested in five replications, and 1000 is the number of measurement points in the potential window from −1000 to 1000 mV with the step of 2 mV.

Further verification was carried out on which of the 12 variants gives the optimal and acceptable profiling effect of the five isotonic drinks. A strategy was developed to assess individual approaches using four coefficients (clustering metrics), for which the k-means procedure was the base clustering algorithm. During the comparative research, the experimental variant was considered optimal, in which each of the four coefficients, based on different theoretical foundations, would indicate five subsets of objects (Table 2). An exemplary comparison of values for IrPtCE and DSSWV current samples interpretation is shown in Fig. 8. The number of clusters indicated by the coefficient is marked in red - in this case five, i.e., as many as expected. Furthermore, it can be noticed that the variants of 2–7 clusters were tested for comparison.

The condition that all clustering metrics indicate five subsets satisfied the following experimental variants: anodic SWV using qDlrPtE, anodic DSSWV using qDPtGCE or qDlrPtE, and cathodic DSSWV using qDlrPtE. A summary of one selected CVI, i.e., Silhouette index, is shown in Fig. S5. There are significant fluctuations in its value for each of the 12 matrices, including the experimental results. Additionally, we only considered the approaches that lead to the detection of 5 subsets. They are marked with arrows. The most favourable variant was cathodic DSSWV using qDlrPtE, and to confirm the selection, PCA was performed to show the grouping of objects (Fig. 9a). The grouping effect can be clearly assessed as favourable because the samples of one isotonic drink are located close to each other, they form homogeneous clusters, and the distances between clusters of similar objects are significant. The most favourable variant was compared with the one indicated by the proposed method based on clustering metrics as the worst (Fig. 9b). In this variant, it was not possible to distinguish three drinks, i.e., white, yellow, and orange. Finally, for the optimal variant, hierarchical cluster analysis (Euclidean distance, Ward agglomeration algorithm) was applied, and the polar dendrogram confirms the expected detection of five homogeneous clusters (Fig. S6).
Table 2
Summary of the hits of each validity index for considered experimental datasets.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Technique</th>
<th>Electrode</th>
<th>CH</th>
<th>DB</th>
<th>Gap</th>
<th>Silhouette</th>
<th>Success</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>qDPtGC</td>
<td>5</td>
<td>5</td>
<td>6</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>SWV (overall anodic potential direction)</td>
<td></td>
<td>qDIrGC</td>
<td>7</td>
<td>3</td>
<td>5</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>qDIrPt</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>qDPtGC</td>
<td>6</td>
<td>5</td>
<td>6</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>SWV (overall cathodic potential direction)</td>
<td></td>
<td>qDIrGC</td>
<td>7</td>
<td>3</td>
<td>7</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>qDIrPt</td>
<td>7</td>
<td>5</td>
<td>7</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>DSSWV (overall anodic potential direction)</td>
<td></td>
<td>qDPtGC</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>qDIrGC</td>
<td>3</td>
<td>2</td>
<td>6</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>qDIrPt</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>DSSWV (overall cathodic potential direction)</td>
<td></td>
<td>qDPtGC</td>
<td>5</td>
<td>5</td>
<td>6</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>qDIrGC</td>
<td>4</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>qDIrPt</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>4</td>
</tr>
</tbody>
</table>

Conclusions

This work showed that the new working quadruple-disk bimetallic and glassy carbon-metallic electrodes can work as voltammetric electronic tongues. Thus, an optimal approach for the use of two electrode materials in one sensor has been proposed, which replaces the typical application of expensive multi-channel analysers or sequential testing with multiple sensors. The success of profiling has been shown to depend on the type of electrode material, the choice of overall anodic or cathodic potential direction, and the way in which directly measured current samples are processed. Particularly important in this work was the use of DSSWV, which extends the possibilities of typical interpretation of direct experimental data and can be useful in various aspects of research based on the voltammetric experiment.

Using, in addition to the classic SWV approach, the DSSWV technique, completely new in this type of application, a significant improvement in signal resolution was achieved and the lower amplitude of the current values was not a disadvantage that had an impact on the final profiling effect. Two strategies for interpreting the measured values did not require extending the measurement range, as they were based on the same measurement results, i.e., current samples obtained at the frequency of 1 kHz. These values have been shown to differ significantly when using three described sensors and samples from different
groups of objects. To verify the proposed approach, five isotonic drinks from one producer were selected, which differed in composition, colour, and taste.

To automate the selection of the optimal sensor, waveform, and interpretation of current samples, an unsupervised chemometric approach was used to calculate four clustering validity indices. Four variants of the measurement and interpretation configuration have been shown to indicate the expected number of five clusters in the set of objects. The best variant was the cathodic DSSWV recorded with qDlrPtE. Using the PCA algorithm and the projection of points on PC1/PC2, the correct grouping of samples in this variant was confirmed, i.e., small distances between samples of the same type and large between samples of different types. It was shown, however, that in three of the considered cases (Table 2), none of the clustering assessment coefficients indicate the correct number of clusters, and the PC1/PC2 projection confirms the inability to correctly distinguish the subsets of samples corresponding to individual isotonic drinks. These observations confirm the need to optimize the measurement and interpretation strategy in the voltammetric object profiling procedure.

Declarations

Competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Author contribution

Szymon Wójcik: Data curation; Formal analysis; Investigation; Resources; Visualization; Writing - original draft. Valentin Mirceski: Inspiration, Conceptualization; Methodology; Funding acquisition; Supervision; Writing - original draft. Bogusław Baś: Design and manufacture of voltammetric sensors. Małgorzata Jakubowska: Conceptualization; Formal analysis; Funding acquisition; Investigation; Methodology; Machine learning; Project administration; Resources; Software; Supervision; Validation; Writing - original draft.

References


Unsectioned Paragraphs

Szymon WÓJCiK: szwojcik@agh.edu.pl, ORCID 0000-0003-3124-3067
Valentin MIRCESKI: valentin@pmf.ukim.mk, ORCID 0000-0002-9191-3926

Bogusław BAŚ: bas@agh.edu.pl, ORCID 0000-0002-5129-2942

Małgorzata JAKUBOWSKA: jakubows@agh.edu.pl, ORCID 0000-0002-0192-4083

Declarations

Founding

Research project supported by program „Excellence initiative – research university” for the AGH University of Science and Technology.

VM acknowledges with gratitude the support from the National Science Centre of Poland through the Opus Lap grant no 2020/39/I/ST4/01854.

Appendix A. Supplementary Information

Supplementary Information to this article can be found online.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryInformation.docx
- SupplementaryInformation.docx