Analytical and Bioanalytical Chemistry
https://doi.org/10.1007/s00216-023-04514-z

TRENDS

# Machine learning toward high-performance electrochemical sensors

**Gabriela F. Giordano[1] · Larissa F. Ferreira[1,2] · Ítalo R. S. Bezerra[1,3] · Júlia A. Barbosa[1,4] · Juliana N. Y. Costa[1,3] · Gabriel J. C. Pimentel[1,5] · Renato S. Lima[1,2,3,4]**

Received: 30 November 2022 / Revised: 30 December 2022 / Accepted: 2 January 2023


**Abstract**
The so-coined fourth paradigm in science has reached the sensing area, with the use of machine learning (ML) toward data-driven improvements in sensitivity, reproducibility, and accuracy, along with the determination of multiple targets from a single measurement using multi-output regression models. Particularly, the use of supervised ML models trained on large data sets produced by electrical and electrochemical bio/sensors has emerged as an impacting trend in the literature by allowing accurate analyses even in the presence of usual issues such as electrode fouling, poor signal-to-noise ratio, chemical interferences, and matrix effects. In this trend article, apart from an outlook for the coming years, we present examples from the literature that demonstrate how helpful ML algorithms can be for dispensing the adoption of experimental methods to address the aforesaid interfering issues, ultimately contributing to translate testing technologies into on-site, practical, and daily applications.

**Keywords** Artificial intelligence · Data treatment · Classification · Regression · Accuracy

## Introduction

Since the early days of science, classical heuristics have looked for patterns from limited data sets to achieve laws, models, and rules. Many of the chemical heuristics that have still been taught in chemistry courses date back at least a century ago, such as the concept of electronegativity, the Pauling rules, and the periodic table of elements. As a limitation, such heuristics suffer from extrapolation issues, being valid only under specific conditions. For instance, the periodic table and electronegativities can change drastically at high pressures. Alternatively, the traditional heuristic approach has recently been replaced by artificial intelligence (AI) and machine learning (ML) models trained on large data sets (these terms are briefly explained in the next topic of this article). Because of the significant amount of data available and the existence of open-source software able to perform high-throughput data processing, AI and ML methods have provided the so-called fourth paradigm in science, namely, data-driven scientific discoveries and improvements by applying robust algorithms that are valid at more extreme conditions when compared with the conventional models [1].

ML has contributed to advance diverse areas such as speech processing, finances, navigation control, locomotion, personality profiling, game playing, computer vision, organic synthesis, bioinformatics, drug discovery, material design, and sensors/biosensors (bio/sensors) [2, 3]. Bio/sensors have been widely developed in a period of sustained growth due to a series of intrinsic advantages like speed, low-cost, simplicity, nondestructive property, and ability to make on-site applications across environmental, food, and biomedical fields [4,

Published in the topical collection *Young Investigators in (Bio-) Analytical Chemistry 2023* with guest editors Zhi-Yuan Gu, Beatriz Jurado-Sánchez, Thomas H. Linz, Leandro Wang Hantao, Nongnoot Wongkaew, and Peng Wu.

✉ Renato S. Lima
renato.lima@lnnano.cnpem.br

1 Brazilian Nanotechnology National Laboratory, Brazilian Center for Research in Energy and Materials, Campinas, São Paulo 13083-100, Brazil

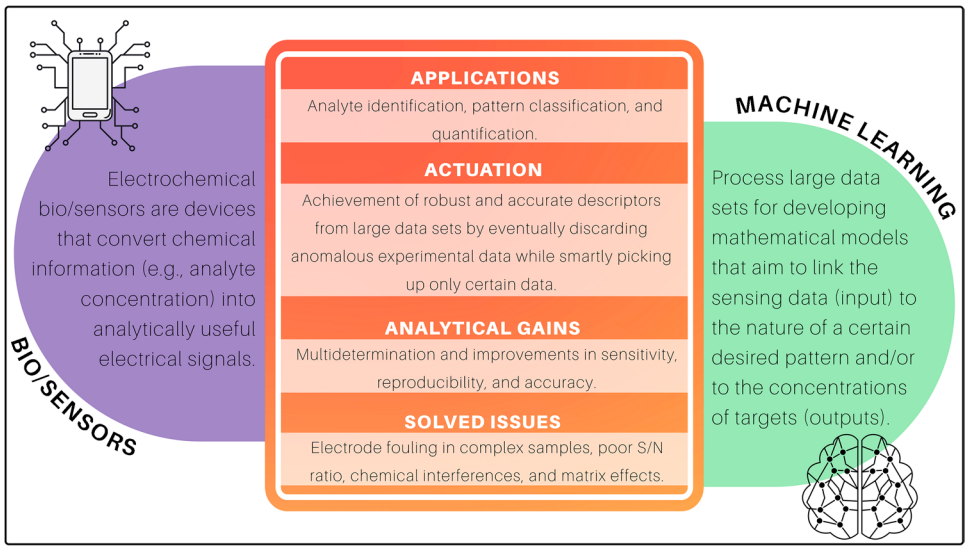
2 Institute of Chemistry, University of Campinas, Campinas, São Paulo 13083-970, Brazil

3 Center for Natural and Human Sciences, Federal University of ABC, Santo André, São Paulo 09210-580, Brazil

4 São Carlos Institute of Chemistry, University of São Paulo, São Carlos, São Paulo 13566-590, Brazil

5 School of Sciences, São Paulo State University, Bauru, São Paulo 17033-360, Brazil

Published online: 13 January 2023




**Scheme 1** Some highlights ML-wise and brief considerations on electrical and electrochemical bio/sensors and ML used in these platforms

5]. In particular, the electrical and electrochemical bio/sensors are powerful tools by combining low-cost, handheld, and user-friendly platforms with rapid and high-performance assays. However, these approaches can suffer from issues such as electrode fouling, poor signal-to-noise (S/N) ratio, chemical interferences, and matrix effect that undermine their precision and accuracy. As an alternative, ML algorithms can assist the performance of electrical and electrochemical bio/sensors even when facing those challenges [4].

In a pioneering work, Holmberg et al. [6] reported in 1996 the use of ML to reduce signal drifts in electronic noses for correctly classifying 85% of alcohol gas samples (1-propanol, 2-propanol, 1-butanol, and 2-butanol). Since then, the amalgamation of electrical and electrochemical bio/sensors with ML to process their data has proved to be an effective shortcut to reach accurate analyses, being a frontier trend in the sensing area [4, 5]. Indeed, ML methods bring relevant analytical gains by eventually removing anomalous experimental features while smartly picking up only certain data to obtain robust descriptors, which can deliver analytically useful information (outputs) from qualitative (analyte identification and pattern classification) and/or quantitative (regression) analyses with enhanced sensitivity, reproducibility, and accuracy even in the presence of issues (electrode fouling, poor S/N ratio, and matrix effects) as mentioned before [4, 5, 7–9]. Another ML-aided benefit is the capability of monitoring multiple parameters or targets from a single measurement (multidetermination), thus avoiding the need of separation methods or various selective sensors/biosensors (bio/sensors) [10, 11]. Such advances are summarized in Scheme 1, together with conceptual considerations on bio/sensors and ML. Some reviews have already been reported summarizing the use of ML in different bio/sensing applications [4, 5, 9, 11]. Specifically, this trend article is aimed at researchers in the field of electrochemical/electrical bio/sensors with focus on the general benefits of ML (particularly, supervised models) and how it can help us to overcome typical and decisive analytical issues found in the area.

## Brief considerations on ML methods

ML is a class of statistical methods that automatically identifies patterns in data sets, even if they are present in high-dimensionality spaces, to obtain input–output algorithms. In the sensing field, such ML algorithms can predict unknown information (output) in qualitative, semi-quantitative, or quantitative assays [2, 12]. To date, ML is a subfield of AI that covers any computational tool capable of mimicking human intelligence, including "less intelligent" approaches such as decision trees, if–then rules, and computer logic. ML methods can be categorized into unsupervised or supervised learning. In the first case, patterns from unlabeled inputs (the output is unknown) are found for clustering, density estimation, and/or dimensionality reduction tasks. For instance, PCA is a commonly used approach to decrease the dimensionality of large data sets while preserving relevant information contained in the original data cloud. Conversely, supervised learning is trained on labeled inputs to achieve classification or regression algorithms, which are able to accurately provide discrete and continuous (numerical) outputs, respectively [2]. Details on the concept, advantages, and limitations of a set of ML algorithms can be found out in recent and didactic reviews [3, 4].

## ML-aided analytical gains in bio/sensors

### Capacity of classification

When bio/sensors lead to the generation of chemically diversified outputs, i.e., fingerprints, with pattern responses being attained to the samples, the use of ML becomes an effective strategy to assure accurate detection and/or classification tasks [13–21]. Impedimetric devices are attractive tools to provide diversified features as the variation of impedance (*Z*) with frequency depends on a set of distinguishable parameters, including resistive, capacitive, interface, and mass-transport phenomena [22]. Ali et al. [15] reported a disposable all-printed impedance sensor for fast detection and classification of three bacteria, *Salmonella typhimurium*, and *Escherichia coli* strains. Such a sensor consisted of interdigitated silver (Ag) electrodes coated with Ag nanowires. While the impedance data for forty samples of each strain were similar, unique sample-related fingerprints consisting of distinct input features, i.e., power, current–potential (*i*-*V*) curve, and first and second derivative of these curves, were extracted from the data and utilized as features in pattern recognition methods like linear discrimination analysis (LDA), linear maximum likelihood estimation (MLE), and nonlinear back propagation neural network (BPNN). These unsupervised approaches delivered the classification of the samples through randomized cross-validation tests with 100% accuracy.

In the work developed by Okur et al. [19], an electronic nose based on quartz crystal microbalance (QCM) sensing array was applied to distinguish five pairs of chiral odor molecules, with ten volatile organic compounds (VOCs) in total. These arrays were coated with six different metal–organic framework (MOF) thin films, three of them with chiral properties and the other three, achiral. Since the isomers have their intrinsic response patterns, these features were treated by a ML method toward a more detailed understanding of the sensor data and an enhanced performance of the nose. Supervised k-nearest neighbor (KNN) algorithm was used, and the mean classification accuracy for distinguishing all 10 isomers was 96.1%, indicating that it was possible to discriminate the compounds with high accuracy.

More recently, the coronavirus disease 2019 (COVID-19) pandemic showed us the necessity of developing quickly available tools to address emerging healthcare issues at the point of care (POC). In the testing area, different devices were reported to diagnose this infection [23], with ML methods proving to be essential in some works for clinical screening applications. For instance, Shan et al. [17] addressed a noninvasive approach to detecting and following up on individuals who are at risk or have an existing COVID-19 infection, with a potential ability to serve as a pandemic control tool. Specifically, a breathing device composed of a hybrid sensing array based on nanomaterial with multiplexed detection capability was described. Different gold (Au) nanoparticles bonded to organic ligands created electrical resistance-based fingerprints as these nanoconjugates undergo diversified levels of swelling or shrinking after exposure to volatile disease-specific biomarkers. ML methods were used to investigate the pattern of these signals to achieve the COVID-19 signature toward screening purposes. The study cohort included 49 confirmed COVID-19 patients, 58 healthy controls, and 33 non-COVID lung infection controls. Discriminant factor analyses (DFA) of the sensing data provided 94% and 76% accuracies in differentiating patients from controls for the training and test sets, respectively. The method further led to 90% and 95% accuracies in differentiating between patients with COVID-19 and those with other lung infections. ML models can benefit not only from diversified sensing data, but also from a significant training set size for creating descriptors with an enhanced prediction ability. In this way, analyses of a larger number of training samples are expected to boost the classification ability of the prior sensor in blinded sample-with applications (test set).

The combination of sensor with ML method toward COVID-19 diagnosis was also proposed by Beduk et al. [21]. In this case, laser-scribed graphene (LSG) devices coupled to Au nanoparticles (AuNPs) were developed as affinity-based biosensing platforms to probe novel variants of COVID-19, i.e., alpha, beta, and delta, as presented in Fig. 1A. The electrode was modified with angiotensin-converting enzyme 2 (ACE2) bioreceptor for detecting SARS-CoV-2 S1 and S2 antigen proteins. A homemade electrochemical analyzer, KAUSTat, was utilized for differential pulse voltammetry (DPV) experiments. This device is portable and allowed smartphone connection via micro-USB port. The KAUSTat platform was also able to provide ML processing through a neural algorithm, thus meaning a promising tool to deliver POC diagnostics. The dense neural network (DNN)–supervised architecture was used to validate such a self-diagnosis setup. A clinical study was conducted with nasopharyngeal swabs from 63 patients having the SARS-CoV-2 variants, patients without the mutation, and negative patients. Accuracies of 98.7%, 99.5%, 100.0%, and 99.4% were obtained for the inference of the beta, alpha, and delta variants and control patients, respectively. Particularly, apart from electrical devices as discussed before, these data reveal the ability of faradaic electrochemical methods to afford chemically diversified signals for ML-aided high-performance classifications as well.

In another work, our group [24] used a five–amino acid peptide (Asn-Asn-Ala-Thr-Asn-COOH, called PEP2003)

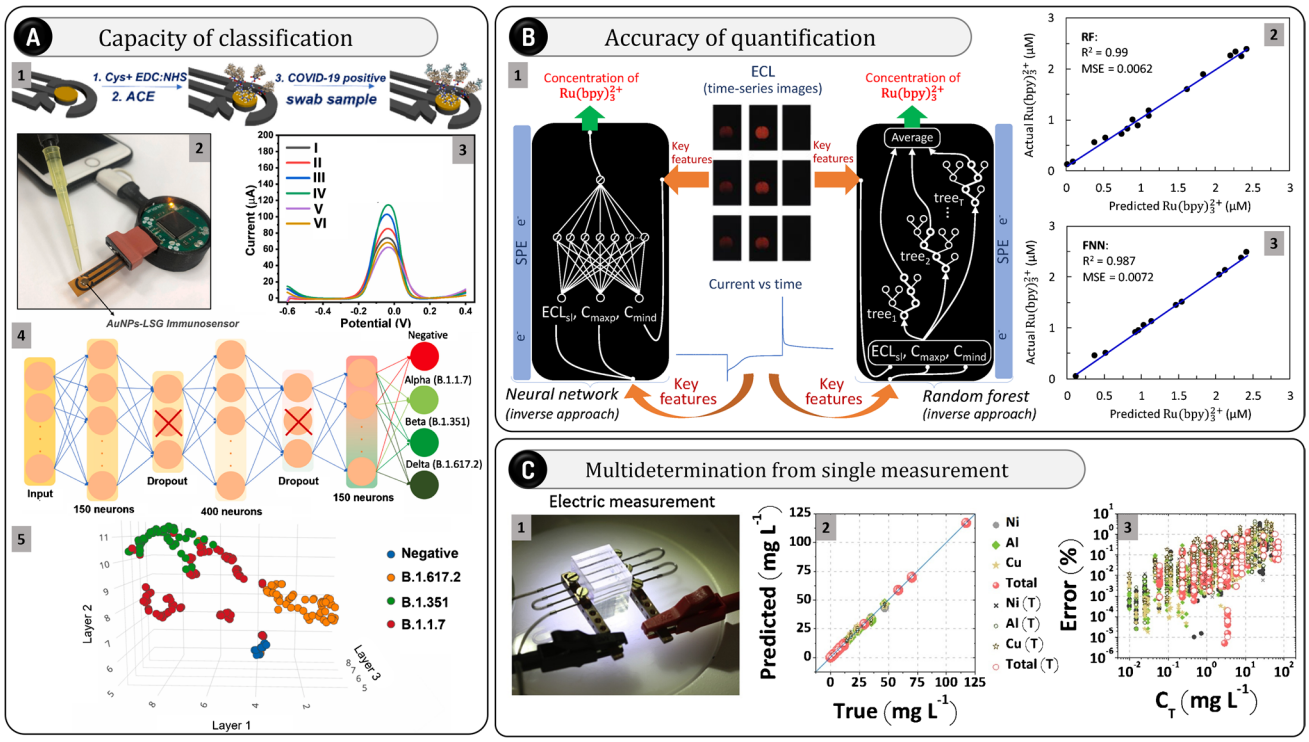



**Fig. 1** ML-aided analytical gains in electrical and electrochemical bio/sensors. **A** Immunoaffinity biosensor based on LSG/AuNPs electrodes to detect novel variants of COVID-19. Construction of the biosensing interface and SARS-CoV-2 detection (1), sensor attached on a portable electrochemical analyzer connected to smartphone by USB (2), DPV scans showing the oxidation current changes after each modification and detection of 200 ng mL$^{-1}$ SARS-CoV-2 S1 and S2 antigens (I: Bare LSG, II: AuNPs-LSG, III: AuNPs-LSG-Binders, IV: AuNPs-LSG Immunosensor, V: 200 ng mL$^{-1}$ S1 antigen, and VI: 200 ng mL$^{-1}$ S2 antigen) (3), DNN architecture (4), and resulting spatial representation of the dataset collected by measuring nasopharyngeal swabs of COVID-19-positive and negative patients (5). EDC, NHS, and Cys mean 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide, N-hydroxysulfosuccinimide, and cysteine, respectively. Reproduced from [21], Copyright 2020 with permission from Elsevier B.V. **B** Smartphone-based ECL sensor. Schematic diagrams of data-driven modeling using FNN and RF algorithms (1) and the parity plots of predicted vs actual Ru(bpy)$_3^{2+}$ concentrations using RF (2) and FNN (3). Reprinted from [26], Copyright 2020 with permission from MDPI Open Access Journals. **C** Multidimensional electrochemical sensor toward metal ion recognition. Microfluidic device comprising an association of double-layer capacitors in parallel (1), a parity plot of predicted vs true ion concentrations according to RF (2), and the errors related to multi-output regression (3). Reproduced from [10], Copyright 2020 with permission from Elsevier B.V

to recognize SARS-CoV-2 antibodies in a label-free (LF) biosensor designed for COVID-19 screening. The biosensor relied on glassy carbon electrodes (GCE) coated with AuNPs, which were used for electrochemical impedance spectroscopy (EIS) analyses. In contrast to big-size recognition elements (e.g., proteins and antibodies), this peptide can be easily prepared via chemical synthesis and it is not amenable to denaturation, hence meeting the trade-off between scalability, cost, and shelf-life. The biosensor preserved 95.1% of the initial signal for 20 days when stored dry at 4 °C. Concerning the discrimination of two types of diluted human sera, pre-pandemic individuals (15) and convalescent patients (24), false negatives (~ 10%) were noted when using a cutoff line based on univariate charge-transfer resistances ($R_{ct}$, extracted from Nyquist plots) for COVID-19 screening. To solve this issue, a supervised model named sure independence screening and sparsifying operator (SISSO) was used, and two simple equations fitted the $R_{ct}$ data. Remarkably, such equations led to the COVID-19 screening of sera into healthy and infected groups with no false positives or negatives, i.e., 100.0% accuracy. SISSO converts the input data into low-dimensional and easy-to-use mathematical equations aiming at accurate qualitative or quantitative analyses even from a small number of training sets, therefore meeting the trade-off between accuracy and simplicity/speed of computation. Such advantages favor the development of ML-aided sample-to-answer experiments on mobile phones, which would greatly facilitate detection at the point of care as no data treatment by the user is needed.

## Accurate quantification

Beyond their employment to improve the capacity of classification, ML methods have been applied to increase the quantification accuracy [25–27]. For instance, Rivera et al. [26] used supervised random forest (RF) and feedforward

neural network (FNN) algorithms to quantitatively investigate the relationship between the concentration of $Ru(bpy)_3^{2+}$ luminophore and the resulting electrochemiluminescence (ECL) and electrochemical signals, as exhibited in Fig. 1B. The multivariate character of this kind of experiment naturally imposes challenges against the fitting of accurate regression models, which were successfully attained by both the prior ML methods. Multimodal data consisting of ECL images and amperograms (recorded in + 1.2 V) and the $Ru(bpy)_3^{2+}$ concentrations were processed as the input and output features for ML models, respectively. High correlations (0.99 for RF and 0.96 for FNN) between real and predicted values were achieved in the detection range from 0.02 to 2.50 μmol $L^{-1}$. Thus, the RF and FNN regression models proved to be capable of directly inferring the $Ru(bpy)_3^{2+}$ concentration from diversified ECL and electrochemical responses.

More recently, Lu et al. [27] applied the artificial neural network (ANN) in the analysis of niclosamide (NA) using an electrochemical sensor. ANN was chosen due to its intrinsic abilities such as the high capacity of self-learning, solution of non-linear problems in arbitrary data, high-speed search for optimal modeling, and robustness against noise issues. The sensor consisted of a glassy carbon electrode modified with carbonized MOF. DPV scans were first recorded to quantitatively detect NA in the range from 1.0 nmol $L^{-1}$ to 9.0 μmol $L^{-1}$ by the traditional analytical curve method. In this case, the peak currents presented a linear relationship with the NA concentration square, and the root mean square error of calibration (RMSEC) was 2.7602. When ANN was applied to the DPV data, nonetheless, the RMSEC was reduced to 0.2788. For the analysis of NA in spiked real drug samples, the average relative standard deviation (RSD) of recoveries reduced from 1.9 to 1.6% with the ML method. More significative improvements in prediction capacity can be provided by ML when challenging the method in complex samples, as it will be discussed in the section "Supervised models solving specific challenges in chemical analyses and bio/sensors" of this trend article.

### Multidetermination from single measurements

Another analytical gain from the adoption of supervised ML models to treat the bio/sensing data is the capability of addressing the determination of multiple analytes, i.e., multidetermination, from a single measurement, thus bypassing the use of preparation methods (e.g., clean-up routine, extraction, chromatography, and electrophoresis) or different selective sensors as usually required. Accordingly, this strategy leads to important advances in cost reduction, throughput, and operational simplicity [10, 28, 29]. Some recent examples of this application of ML are presented below.

Bonet-San-Emeterio et al. [29] proposed a voltammetric device electrochemically modified with reduced graphene oxide (rGO) for the analysis of mixtures of dopamine (DA), serotonin (5-hydroxytryptamine, 5-HT), and their most common interferents, i.e., ascorbic acid (AA) and uric acid (UA). Although these compounds are electrochemically active, peak current overlapping can occur as their voltammetric responses are similar, damaging the accuracy of direct electrochemical analyses. Remarkably, ANN provided the accurate quantification of each component from a unique voltammetry scan. Methods such as principal component analysis (PCA), discrete wavelet transform (DWT), and fast Fourier transform (FFT) were further employed to decrease the dimensionality of the voltammetric data. The graphs acquired from DWT-ANN showed the lowest dispersion and greatest linearity (correlation coefficient > 0.974). Further, this model yielded a normalized root-mean-square error (NRMSE) of 0.088.

In another publication, our group [10] developed a multidimensional electrochemical sensing array that assures the multidetermination of metal ions in lake samples in a direct way from a single cross-reactive electrode and measurement, as presented in Fig. 1C. Microfluidic devices were prototyped by a scalable, cleanroom-free, green, and simple method, whereas commercial and ordinary stainless-steel capillaries were utilized as sensing probes. The latter formed an association of double-layer capacitors in parallel to generate multidimensional, i.e., chemically diversified, responses. This property was reached in a single response since the equivalent capacitances (fingerprints) encompass contributions of all the individual capacitors that were capable of delivering pattern responses because of the multidimensional nature of the frequency-function capacitance scans [30, 31], as aforesaid for impedimetric devices. This sensor led to the simultaneous quantification of the cations $Ni^{2+}$, $Al^{3+}$, and $Cu^{2+}$ from universal ac voltage tests by treating the capacitance values assuming the electrodes as ideally polarizable interfaces. Specifically, multi-output regressions based on RF algorithms showed high correlation coefficients ($R^2 > 0.99$) in all the cases. The overall mean absolute error (MAE) was revealed to be only 0.2 mg $L^{-1}$ (the concentrations of the ions ranged from 5.0 up to 50.0 mg $L^{-1}$).

## Supervised models solving specific challenges in chemical analyses and bio/sensors

### Electrode fouling

As recently demonstrated by Ferreira et al. [7], ML is capable of circumventing accuracy issues caused by electrode

fouling. Specifically, the authors developed a platform using a millifluidic impedimetric sensor to monitor the synthesis of silica nanoparticles ($SiO_2$NPs) for 24 h. These nanoparticles were selected due to their wide range of applications. Using linear regression–based analytical curves with univariate signals of $Z$ at specific frequencies as sensing features, the inter-synthesis accuracy (employing independent sensors as well) was poor. Specifically, the determined hydrodynamic diameter ($D_H$) of the nanoparticles presented discrepancies of up to 132.8% with regard to the true data (determined by dynamic light scattering, DLS). This poor reproducibility was shown to be triggered by the adsorption of $SiO_2$NPs on the Au electrode during the synthesis. Utilizing SISSO descriptors composed of only six $Z$ inputs (at different frequencies; the whole spectra contained 18 $Z$ data), this interference could be overcome with $D_H$ being determined in a real-time, accurate, and simple way without using antifouling layers on the electrodes. The global average accuracy was $103.7 \pm 1.9\%$, thereby demonstrating the robustness of the SISSO descriptor. The $SiO_2$NP concentration could also be attained with accuracy using the SISSO multi-output regression model. The root-mean-square errors (RMSE) were calculated as 2.0 nm and $2.6 \times 10^{10}$ nanoparticles $mL^{-1}$ for the size and $SiO_2$NPs concentration, respectively.

In another work, Aiassa et al. [32] developed a new method for electrochemical sensing of propofol with compensation of the fouling effect through the ML model, as shown in Fig. 2A. In this case, the passivation of the HB pencil lead, a carbon electrode, stemmed from the formation of a polymeric film coating, decreasing the signal and disturbing the accuracy as such a phenomenon is characterized by a strong non-linear response. As a consequence, the application of a univariate linear regression model to cyclic voltammetry (CV) features provided poor accuracy in classifying propofol samples diluted in phosphate buffered saline solution (PBS) and human serum, being only 69.8% and 33.3%, respectively. The compensation of this non-linear fouling effect was achieved by processing the CV data through the radial basis function support vector classifier (RBF-SVC), which is a non-linear ML algorithm. In this case, the accuracies were improved to 98.9% and 100% for samples in PBS and human serum, respectively.

As noted in the two prior works [7, 32] and following in this section, supervised ML models are a powerful strategy to afford direct analyses as the use of experimental methods to inhibit the interfering issue is avoided and ML-fitted algorithms can be automatically performed, e.g., in mobile phones toward the development of sample-to-answer fashions. The latter facilitates detection at the point of need because no data treatment or interpretation by the user is required. Concerning the approaches that are commonly described in the literature to prevent electrode fouling, they include self-assembled monolayers of polyethylene glycol, zwitterionic polymers, hybrid coating, and bovine serum albumin (BSA). Although valuable, these blocking layers generally hamper the redox reaction kinetics, hence impairing the analytical performance of the device. Further, these coatings may present low durability, repeatability, and scalability, mining POC tests and commercial manufacturing feasibility [33].

## Matrix effects

ML can also be employed to solve matrix effects in chemical analysis, as recently reported by our group [8]. Microhole-structured and flexible Ni meshes acting simultaneously as gas diffusion membranes and electrodes were used for voltammetric determination of volatile compounds. The diffusion of gas from donor (samples) to receptor phases (electrolyte) was conducted in headspace (contactless) mode, thus minimizing issues related to the mesh fouling, as represented in Fig. 2B. The platform was challenged in sugar cane fermentation broths for ethanol determination. This application is relevant to detect unconformities and provide high efficiencies in the production of ethanol biofuel. $Ni(OH)_2$-modified Ni electrodes were interrogated with CV to probe ethanol vapor dissolved into the receptor solution. Briefly, $Ni(OH)_2$ was reversibly oxidized in alkaline media to the high-valence NiOOH, which promoted the irreversible oxidation of ethanol by acting as an electron mediator [34]. Using the univariate method of analytical curve–based interpolation with oxidation currents (+0.65 V) as responses, the attained ethanol concentrations variated in relation to the expected data with accuracies from 80% up to 105%. This accuracy range suffered from the susceptibility of gas diffusion to matrix effects as the medium composition alters the analyte evaporation rate. Nonetheless, a simple SISSO descriptor with CV scan–based inputs was once again able to boost accuracy, as discussed next.

Parity plots between the predicted and expected concentrations of ethanol exhibited ideal behavior, with the slope close to 1.0 and linear fitting (0.99 $R^2$), whereas the accuracies ranged from 97 to 102% for the test samples [8]. The use of a mathematical descriptor containing only seven features of current at different potentials provided direct assays with accuracy. In this sense, while the traditional method of standard addition may overcome matrix effect interferences, the necessity for determining spiked solutions before the analysis of every sample can compromise the practical accomplishment of daily, practical, and on-site applications.

## Chemical interferences

Torrecilla et al. [35] developed an amperometric biosensor to simultaneously determine glucose and its interferents,

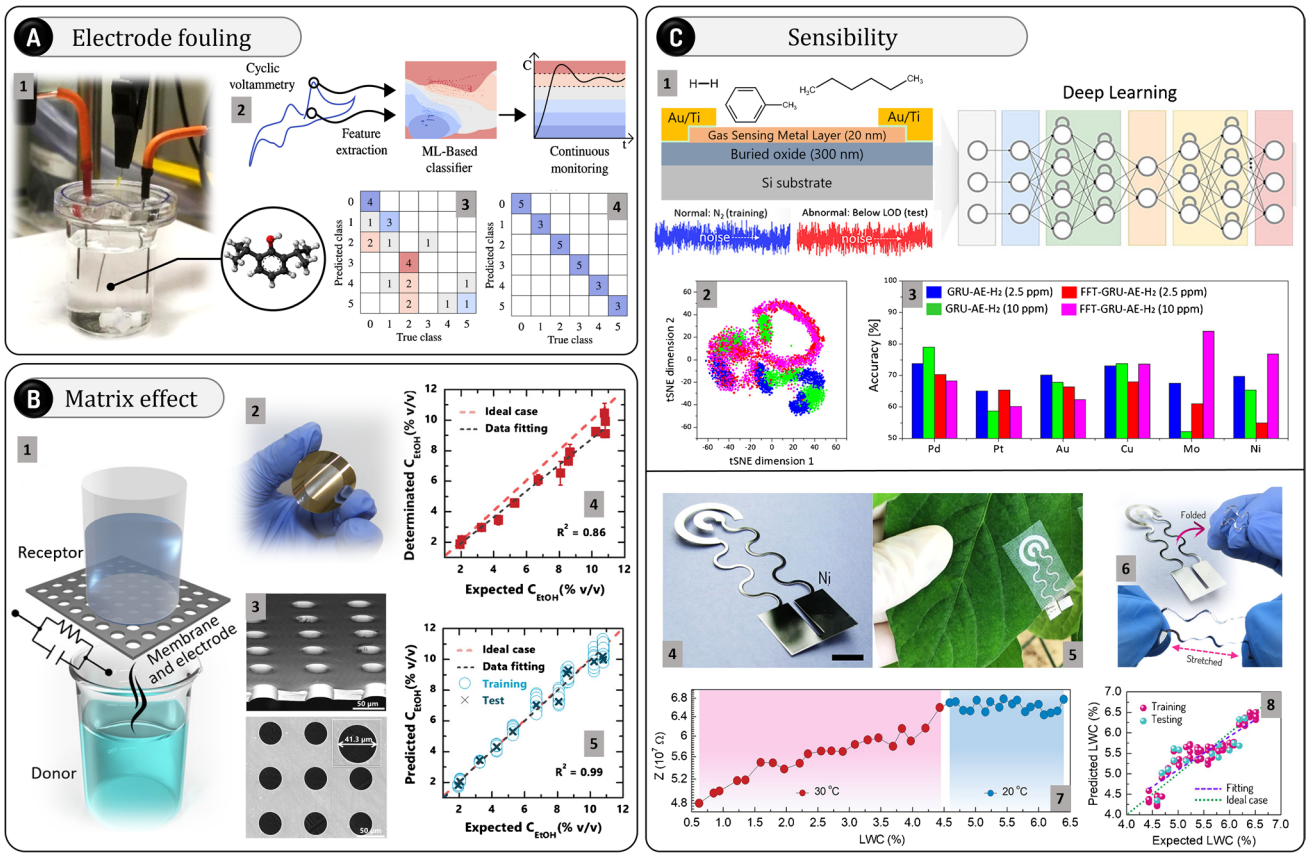



**Fig. 2** ML addressing specific challenges in the bio/sensing area. **A** Continuous voltametric monitoring of propofol. Experimental setup and the representation of a propofol molecule (1), scheme of the proposed ML approach that assured accurate determinations in spite of the electrode fouling (2), and the confusion matrices for the results in undiluted human serum with the standard linear analytical curve (3) and with the ML-based model (4). Adapted from [32], Copyright 2022 with permission from ACS. **B** Bifunctional metal mesh working as gas diffusion membrane and electrode for the accurate electrochemical quantification of ethanol in liquid samples (donor) by overcoming matrix effect–related interferences. Illustration of the developed protocol with the addition of an electrolytic receptor solution to dissolve ethanol vapor and provide low-ohmic drop faradaic electrochemical assays (1), picture (2) and SEM images (3) of the flexible Ni mesh with 20.0-µm microholes, and parity plots of predicted vs true concentrations of ethanol considering the sensing data treatment by the traditional univariate analytical curve method (4) and SISSO (5). Adapted from [8], Copyright 2021 with permission from ACS. **C** ML-aided sensitive monitoring of gas (1–3) and LWC from soy leaves (4–8). Layers of the device and the use of DNN to extract "hidden signals" from the raw resistance signals (1), DNN learning approach (2), and classification accuracy for two $H_2$ concentrations measured in distinct metallic electrodes as highlighted (3). Adapted from [43], Copyright 2022 with permission from ACS. Impedimetric wearable sensor composed of flexible free-standing Ni electrodes to assess LWC from soy leaves. Electrodes before (4) and after attachment on leaf using adhesive tape (5), electrodes under deformation (6), *Z* assays over 24 h at 30 and 20 °C (7), and the ensuing parity plot of predicted vs expected LWC at 20 °C using SISSO (8). The scale bar in (4) means 5.0 mm. Adapted from [44], Copyright 2022 with permission from ACS

AA and UA, in a mixture. The concentration of these analytes ranged from 0.1 to 1.0 mmol $L^{-1}$, and their analytical information was extracted from CV. The combination of biosensing responses with chemometric tools can solve the issue of complex analytical signals from the set of species with similar responses. In this sense, the authors used ANN to process the device signals aiming at the accurate quantification of glucose in the presence of interferents. The contents of glucose, AA, and UA could be estimated with mean prediction errors (MPE) of 0.007, 0.013, and 0.032, respectively. Moreover, in all these cases, $R^2$ was higher than 0.99. Using ML, accurate quantifications could be reached without the adoption of experimental methods to prevent chemical interferences in bio/sensors, such as the use of separation techniques and permselective membranes [36].

## Sensitivity

Improving the sensitivity is a relevant task to be pursued in the sensing area. In addition to contributing toward early monitoring, sensitive devices allow a high dilution of samples, hence providing the analysis of small-volume samples (a crucial benefit in biological assays) and preventing electrode fouling as the interferences present in the samples are

diluted to insignificant contents [37]. Despite the existence of a plethora of efficient strategies to enhance sensitivity such as the use of nanomaterials toward current amplification [38–41], the adoption of bare electrodes is desirable by supporting scalable, simple, and low-cost sensing methods [42]. In this case, we can resort to ML to guarantee sensitive analyses by smartly selecting specific input features. For instance, Cho et al. [43] developed a resistive array comprising six metals, i.e., Au, Cu, Mo, Ni, Pt, and Pd, for sensitive gas monitoring. DNN was used to extract "hidden signals" from the raw resistance signals in the error region, as shown in Fig. 2C. They found that the use of ML enabled a reduction in the limit of detection (LOD) for $H_2$ from 10.0 to 2.5 mg $L^{-1}$ with a recovery of 73.8% considering the Pd electrode.

Significant ML-assisted improvement in the S/N ratio was recently described by our group [44]. We proposed an impedimetric wearable sensor for determining the loss of water contents (LWC) from soy leaves at different temperatures along 24 h, namely, 12 h at 30 °C and then 12 h at 20 °C. Water content is a key marker of leaf health, and it can lend insights into daily practice in precision agriculture, toxicity studies, and the development of agricultural inputs. Ni films obtained by well-established microfabrication approaches (photolithography and electroplating) were used as flexible on-leaf electrodes, as displayed in Fig. 2C. While these electrodes were sufficiently sensitive to quantify LWC at 30 °C using a simple linear fitting from univariate single-frequency *Z* input data, these signals remained nearly unchanged over the next 12 h at 20 °C when the water loss rate was decreased 1.7 times ($6.2\times10^{-3}$% $min^{-1}$ at 30 °C and $3.7\times10^{-3}$% $min^{-1}$ at 20 °C). Remarkably, the SISSO descriptor picking up only six input features from the whole Bode plot (16 *Z* values vs frequency) led to the accurate monitoring of LWC at 20 °C, with RMSEs being 0.2% and 0.1% for the training and test sets, respectively. The ability of directly determine LWC from plant leaves at distinct temperatures through a simple ML descriptor is important for practical use of the method outside of laboratory facilities in outdoor or even indoor gardens.

## Outlook

As presented throughout this trend article, the convergence of electrical and electrochemical bio/sensors with machine learning methods provides a promising strategy aiming at the translation of testing technologies capable of affording point-of-need and extensive trials (many trials per thousand people) into practical use. In addition to allowing multidetermination and improving the analytical performance of devices through the discrimination of overlapping signals, supervised ML models may lead to accurate tests without the requirement for experimental methods to prevent common analytical issues such as electrode fouling, matrix effects, chemical interferences, and poor S/N ratio. These obstacles can delay time and increase costs facing the commercial adaptation of sensing technologies [23].

According to Clark [45]: "the young investigators in the field of sensors are coming from a myriad of backgrounds, including materials, analytical chemistry, and chemical biology." Probably, by considering the analytical gains provided by ML-fitted mathematical models as described above, advanced data treatment techniques trained on large data sets will be also a crucial topic to be dominated by the emerging generation of these scientists. In fact, the amalgamation of bio/sensors with ML has emerged as a relevant trend in the literature, adding even more interdisciplinarity to the exciting sensing area [23]. In practice, several commercial and open-source software packages, codes, and tools exist to implement the most common ML models and learning workflow tasks. A set of theoretical and experimental databases are also available for supporting various areas [2, 46].

As mentioned above, the prediction ability of supervised ML methods is expected to progressively increase with the number of training samples and the chemical diversification of the sensing data (inputs) [1]. Hence, instead of training the model with standard samples or a limited number of real samples as usually noted in the literature, the descriptors in future works should be extracted from large sets of real samples. Such a type of investigation will not only contribute to effectively advancing the platform across the technology readiness levels (TRLs) toward real-world applications, but also it will likely reinforce the engineering, biological, and chemical challenges into the research, including parameters such as reproducibility, scaling, stability (i.e., shelf-life), sensitivity, cross-reactivity, and fouling. In this way, one should also stress the relevance of dialogues with business-related entities for product development. The convergence among academic and industrial knowledge will probably work as a shortcut to speed up the commercial translation of sensing systems, which are likely to be better equipped to deal with on-demand economic, social, health, and environment challenges in the future.

However, one should stress that the two prior requirements, i.e., number of training samples and chemical diversification of the inputs, are not enough to guarantee the fitting of algorithms with a high generalization, i.e., models that can accurately predict the output for unknown samples (test set) outside the training set. The fitting of these algorithms from large data sets critically depends also on the quality of the experimental data that compose the inputs [2, 47]. Specifically, the measurements must be precise and reproducible to minimize systematic errors. Further, the experimental data must be representative of the target property even in the presence of the aforesaid analytical issues. In fact, ML will only be able to meet these issues if a minimum correlation between the sensing data and the target property is held.

In practice, as absolute signal values are amenable to non-specific variations, this correlation can be partially provided by parameters such as peak position and signal profile, i.e., relative variations in signals.

**Acknowledgements** This work was sponsored by the Brazilian Coordination for Improvement of Higher Education Personnel (CAPES; grant 88887.513140/2020-00), the Brazilian National Council for Scientific and Technological Development (CNPq; grant 407951/2021-0), and the São Paulo Research Foundation (FAPESP; grants 2022/04397-4, 2020/09102-4, and 2018/24214-3).

## Declarations

**Conflict of interest** The authors declare no conflict of interest.

## References

1. George J, Hautier G. Chemist versus machine: traditional knowledge versus machine learning techniques. Trends Chem. 2021. https://doi.org/10.1016/j.trechm.2020.10.007.
2. Schleder GR, Padilha ACM, Acosta CM, Costa M, Fazzio A. From DFT to machine learning: recent approaches to materials science–a review. J Phys: Mater. 2019. https://doi.org/10.1088/2515-7639/ab084b.
3. Ayres LB, Gomez FJV, Linton JR, Silva MF, Garcia CD. Taking the leap between analytical chemistry and artificial intelligence: a tutorial review. Anal Chim Acta. 2021. https://doi.org/10.1016/j.aca.2021.338403.
4. Cui F, Yue Y, Zhang Y, Zhang Z, Zhou HS Advancing biosensors with machine learning. ACS Sens. 2020. https://doi-org.ez106.periodicos.capes.gov.br/, https://doi.org/10.1021/acssensors.0c01424.
5. Puthongkham P, Wirojsaengthong S, Suea-Ngam A. Machine learning and chemometrics for electrochemical sensors: moving forward to the future of analytical chemistry. Analyst. 2021. https://doi.org/10.1039/d1an01148k.
6. Holmberg M, Winquist F, Lundström I, Davide F, DiNatale C, D'Amico A. Drift counteraction for an electronic nose. Sens Actuators B Chem. 1996. https://doi.org/10.1016/S0925-4005(97)80124-4.
7. Ferreira LF, Giordano GF, Gobbi AL, Piazzetta MHO, Schleder GR, Lima RS. Real-time and *in situ* monitoring of the synthesis of silica nanoparticles. ACS Sens. 2022. https://doi.org/10.1021/acssensors.1c02697.
8. Giordano GF, Freitas VMS, Schleder GR, Santhiago M, Gobbi AL, Lima RS. Bifunctional metal meshes acting as a semipermeable membrane and electrode for sensitive electrochemical determination of volatile compounds. ACS Appl Mater Interfaces. 2021. https://doi.org/10.1021/acsami.1c07874.
9. Haick H, Tang N. Artificial intelligence in medical sensors for clinical decisions. ACS Nano. 2021. https://doi.org/10.1021/acsnano.1c00085
10. da Silva GS, de Oliveira LP, Costa GF, Giordano GF, Nicoliche CYN, da Silva AA, Khan LU, da Silva GH, Gobbi AL, Silveira J v., Filho AGS, Schleder GR, Fazzio A, Martinez DST, Lima RS. Ordinary microfluidic electrodes combined with bulk nanoprobe produce multidimensional electric double-layer capacitances towards metal ion recognition. Sens Actuators B Chem. 2020. https://doi.org/10.1016/j.snb.2019.127482.
11. Ballard Z, Brown C, Madni AM, Ozcan A. Machine learning and computation-enabled intelligent sensor design. Nat Mach Intell. 2021. https://doi.org/10.1038/s42256-021-00360-9.
12. Debus B, Parastar H, Harrington P. Kirsanov D Deep learning in analytical chemistry. TrAC - Trends Anal Chem. 2021. https://doi.org/10.1016/j.trac.2021.116459.
13. Shehada N, Cancilla JC, Torrecilla JS, Pariente ES, Brönstrup G, Christiansen S, Johnson DW, Leja M, Davies MPA, Liran O, Peled N, Haick H. Silicon nanowire sensors enable diagnosis of patients via exhaled breath. ACS Nano. 2016. https://doi.org/10.1021/acsnano.6b03127.
14. Rong Y, Padron A v., Hagerty KJ, Nelson N, Chi S, Keyhani NO, Katz J, Datta SPA, Gomes C, McLamore ES. Post hoc support vector machine learning for impedimetric biosensors based on weak protein-ligand interactions. Analyst. 2018. https://doi.org/10.1039/c8an00065d.
15. Ali S, Hassan A, Hassan G, Eun CH, Bae J, Lee CH, Kim IJ. Disposable all-printed electronic biosensor for instantaneous detection and classification of pathogens. Sci Rep. 2018. https://doi.org/10.1038/s41598-018-24208-2.
16. Dean SN, Shriver-Lake LC, Stenger DA, Erickson JS, Golden JP, Trammell SA. Machine learning techniques for chemical identification using cyclic square wave voltammetry. Sensors. 2019. https://doi.org/10.3390/s19102392.
17. Shan B, Broza YY, Li W, Wang Y, Wu S, Liu Z, Wang J, Gui S, Wang L, Zhang Z, Liu W, Zhou S, Jin W, Zhang Q, Hu D, Lin L, Zhang Q, Li W, Wang J, Liu H, Pan Y, Haick H. Multiplexed nanomaterial-based sensor array for detection of COVID-19 in exhaled breath. ACS Nano. 2020. https://doi.org/10.1021/acsnano.0c05657.
18. Kim H, Park S, Jeong IG, Song SH, Jeong Y, Kim CS, Lee KH. Noninvasive precision screening of prostate cancer by urinary multimarker sensor and artificial intelligence analysis. ACS Nano. 2021. https://doi.org/10.1021/acsnano.0c06946.
19. Okur S, Qin P, Chandresh A, Li C, Zhang Z, Lemmer U, Heinke L. An enantioselective e-nose: an array of nanoporous homochiral MOF films for stereospecific sensing of chiral odors. Angew Chem Int Ed. 2021. https://doi.org/10.1002/anie.202013227.
20. Leon-Medina JX, Tibaduiza DA, Burgos JC, Cuenca M, Vasquez D. Classification of As, Pb and Cd heavy metal ions using square wave voltammetry, dimensionality reduction and machine learning. IEEE Access. 2022. https://doi.org/10.1109/ACCESS.2022.3143451.
21. Beduk D, Ilton de Oliveira Filho J, Beduk T, Harmanci D, Zihnioglu F, Cicek C, Sertoz R, Arda B, Goksel T, Turhan K, Salama KN, Timur S. "All In One" SARS-CoV-2 variant recognition platform: machine learning-enabled point of care diagnostics. Biosens Bioelectron X. 2022. https://doi.org/10.1016/j.biosx.2022.100105.
22. Orazem ME, Tribollet B. Electrochemical impedance spectroscopy. 2nd ed. Wiley; 2017.
23. Rosati G, Idili A, Parolo C, Fuentes-Chust C, Calucho E, Hu L, Castro E Silva CDC, Rivas L, Nguyen EP, Bergua JF, Alvárez-Diduk R, Muñoz J, Junot C, Penon O, Monferrer D, Delamarche E, Merkoçi A. Nanodiagnostics to face SARS-CoV-2 and future pandemics: from an idea to the market and beyond. ACS Nano. 2021;15:17137–49. https://doi.org/10.1021/acsnano.1c06839.
24. Castro ACH, Bezerra ÍRS, Pascon AM, da Silva GH, Philot EA, de Oliveira VL, Mancini RSN, Schleder GR, Castro CE, de Carvalho LRS, Fernandes BH v., Cilli EM, Sanches PRS, Santhiago M, Charlie-Silva I, Martinez DST, Scott AL, Alves WA, Lima RS. Modular label-free electrochemical biosensor loading nature-inspired peptide toward the widespread use of COVID-19 antibody tests. ACS Nano. 2022. https://doi.org/10.1021/acsnano.2c04364.
25. Xu Y, Li C, Jiang Y, Guo M, Yang Y, Yang Y, Yu H. Electrochemical impedance spectroscopic detection of E. coli with machine learning. J Electrochem Soc. 2020. https://doi.org/10.1149/1945-7111/ab732f.
26. Rivera EC, Swerdlow JJ, Summerscales RL, Uppala PPT, Filho RM, Neto MRC, Kwon HJ. Data-driven modeling of smartphone-based electrochemiluminescence sensor data using artificial intelligence. Sensors. 2020. https://doi.org/10.3390/s20030625.
27. Lu X, Liu P, Bisetty K, Cai Y, Duan X, Wen Y, Zhu Y, Rao L, Xu Q, Xu J. An emerging machine learning strategy for electrochemical sensor and supercapacitor using carbonized metal–organic

framework. J Electroanal Chem. 2022. https://doi.org/10.1016/j.jelechem.2022.116634.
28. González-Calabuig A, Guerrero D, Serrano N, del Valle M. Simultaneous voltammetric determination of heavy metals by use of crown ether-modified electrodes and chemometrics. Electroanalysis. 2016. https://doi.org/10.1002/elan.201500512.
29. Bonet-San-Emeterio M, González-Calabuig A, del Valle M. Artificial neural networks for the resolution of dopamine and serotonin complex mixtures using a graphene-modified carbon electrode. Electroanalysis. 2019. https://doi.org/10.1002/elan.201800525.
30. Vadhva P, Hu J, Johnson MJ, Stocker R, Braglia M, Brett DJL, Rettie AJE. Electrochemical impedance spectroscopy for all-solid-state batteries: theory, methods and future outlook. ChemElectroChem. 2021. https://doi.org/10.1002/celc.202100108.
31. Yang L, Li Y, Griffis CL, Johnson MG. Interdigitated microelectrode (IME) impedance sensor for the detection of viable Salmonella typhimurium. Biosens Bioelectron. 2004. https://doi.org/10.1016/j.bios.2003.10.009
32. Aiassa S, Ny Hanitra I, Sandri G, Totu T, Grassi F, Criscuolo F, de Micheli G, Carrara S, Demarchi D. Continuous monitoring of propofol in human serum with fouling compensation by support vector classifier. Biosens Bioelectron. 2021. https://doi.org/10.1016/j.bios.2020.112666.
33. Nicoliche CYN, Pascon AM, Bezerra ÍRS, de Castro ACH, Martos GR, Bettini J, Alves WA, Santhiago M, Lima RS. *In situ* nanocoating on porous pyrolyzed paper enables antibiofouling and sensitive electrochemical analyses in biological fluids. ACS Appl Mater Interfaces. 2022. https://doi.org/10.1021/acsami.1c18778.
34. Toghill KE, Xiao L, Stradiotto NR, Compton RG. The determination of methanol using an electrolytically fabricated nickel microparticle modified boron doped diamond electrode. Electroanalysis. 2010. https://doi.org/10.1002/elan.200900523.
35. Torrecilla JS, Mena ML, Yáñez-Sedeño P, García J. A neural network approach based on gold-nanoparticle enzyme biosensor. J Chemom. 2008. https://doi.org/10.1002/cem.1100.
36. Sung WJ, Na K, Bae YH. Biocompatibility and interference eliminating property of pullulan acetate/polyethylene glycol/heparin membrane for the outer layer of an amperometric glucose sensor. Sens Actuators B Chem. 2004. https://doi.org/10.1016/j.snb.2003.12.005.
37. Liu G, Rusling JF. COVID-19 antibody tests and their limitations. ACS Sens. 2021. https://doi.org/10.1021/acssensors.0c02621.
38. Farka Z, Juřík T, Kovář D, Trnková L, Skládal P. Nanoparticle-based immunochemical biosensors and assays: recent advances and challenges. Chem Rev. 2017. https://doi.org/10.1021/acs.chemrev.7b00037.
39. Wongkaew N, Simsek M, Griesche C, Baeumner AJ. Functional nanomaterials and nanostructures enhancing electrochemical biosensors and lab-on-a-chip performances: recent progress, applications, and future perspective. Chem Rev. 2019. https://doi.org/10.1021/acs.chemrev.8b00172.
40. Welch EC, Powell JM, Clevinger TB, Fairman AE, Shukla A. Advances in biosensors and diagnostic technologies using nanostructures and nanomaterials. Adv Funct Mater. 2021. https://doi.org/10.1002/adfm.202104126.
41. Mariani F, Gualandi I, Schuhmann W, Scavetta E. Micro- and nano-devices for electrochemical sensing. Microchim Acta. 2022. https://doi.org/10.1007/s00604-022-05548-3.
42. Shimizu FM, Pasqualeti AM, Nicoliche CYN, Gobbi AL, Santhiago M, Lima RS. Alcohol-triggered capillarity through porous pyrolyzed paper-based electrodes enables ultrasensitive electrochemical detection of phosphate. ACS Sens. 2021. https://doi.org/10.1021/acssensors.1c01302.
43. Cho SY, Lee Y, Lee S, Kang H, Kim J, Choi J, Ryu J, Joo H, Jung HT, Kim J. Finding hidden signals in chemical sensors using deep learning. Anal Chem. 2020. https://doi.org/10.1021/acs.analchem.0c00137.
44. Barbosa JA, Freitas VMS, Vidotto LHB, Schleder GR, de Oliveira RAG, da Rocha JF, Kubota LT, Vieira LCS, Tolentino HCN, Neckel IT, Gobbi AL, Santhiago M, Lima RS. Biocompatible wearable electrodes on leaves toward the on-site monitoring of water loss from plants. ACS Appl Mater Interfaces. 2022. https://doi.org/10.1021/acsami.2c02943.
45. Clark HA. Has sensing become an engineering discipline? ACS Sens. 2020. https://doi.org/10.1021/acssensors.0c00227.
46. Butler KT, Davies DW, Cartwright H, Isayev O, Walsh A. Machine learning for molecular and materials science. Nature. 2018. https://doi.org/10.1038/s41586-018-0337-2.
47. Mistry A, Franco AA, Cooper SJ, Roberts SA, Viswanathan V. How machine learning will revolutionize electrochemical sciences. ACS Energy Lett. 2021. https://doi.org/10.1021/acsenergylett.1c00194.