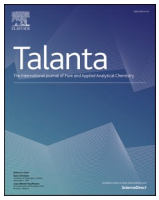
Talanta 243 (2022) 123327

Contents lists available at ScienceDirect

Talanta

journal homepage: www.elsevier.com/locate/talanta



Using machine learning and an electronic tongue for discriminating saliva samples from oral cavity cancer patients and healthy individuals

Daniel C. Braz [a,b], Mário Popolin Neto [c,d], Flavio M. Shimizu [b,e,f], Acelino C. Sá [b], Renato S. Lima [f,g,h,i], Angelo L. Gobbi [f], Matias E. Melendez [j,k], Lídia M.R. B. Arantes [j], André L. Carvalho [j], Fernando V. Paulovich [l], Osvaldo N. Oliveira Jr [b,*]

[a] Mato Grosso do Sul State University (UEMS), 79804-970, Dourados, MS, Brazil
[b] São Carlos Institute of Physics (IFSC), University of São Paulo (USP), 13566-590, São Carlos, SP, Brazil
[c] Federal Institute of São Paulo (IFSP), 14804-296, Araraquara, SP, Brazil
[d] Institute of Mathematics and Computer Sciences (ICMC), University of São Paulo (USP), 13566-590, São Carlos, SP, Brazil
[e] Department of Applied Physics, "Gleb Wataghin" Institute of Physics (IFGW), University of Campinas (UNICAMP), 13083-859, Campinas, SP, Brazil
[f] Brazilian Nanotechnology National Laboratory, Brazilian Center for Research in Energy and Materials, 13083-970, Campinas, SP, Brazil
[g] Institute of Chemistry, University of Campinas, 13083-970, Campinas, São Paulo, Brazil
[h] Federal University of ABC, 09210-580, Santo André, SP, Brazil
[i] São Carlos Institute of Chemistry, University of São Paulo, 09210-580, São Carlos, SP, Brazil
[j] Molecular Oncology Research Center, Barretos Cancer Hospital, 14784-400, Barretos, SP, Brazil
[k] Brazilian National Cancer Institute, 20231-091, Rio de Janeiro, RJ, Brazil
[l] Faculty of Computer Science, Dalhousie University, Halifax, Canada

ARTICLE INFO

Keywords:
Cancer diagnosis
Electronic tongue
Impedance spectroscopy
Machine learning
Multidimensional calibration space

ABSTRACT

The diagnosis of cancer and other diseases using data from non-specific sensors – such as the electronic tongues (e-tongues) - is challenging owing to the lack of selectivity, in addition to the variability of biological samples. In this study, we demonstrate that impedance data obtained with an e-tongue in saliva samples can be used to diagnose cancer in the mouth. Data taken with a single-response microfluidic e-tongue applied to the saliva of 27 individuals were treated with multidimensional projection techniques and non-supervised and supervised machine learning algorithms. The distinction between healthy individuals and patients with cancer on the floor of mouth or oral cavity could only be made with supervised learning. Accuracy above 80% was obtained for the binary classification (YES or NO for cancer) using a Support Vector Machine (SVM) with radial basis function kernel and Random Forest. In the classification considering the type of cancer, the accuracy dropped to ca. 70%. The accuracy tended to increase when clinical information such as alcohol consumption was used in conjunction with the e-tongue data. With the random forest algorithm, the rules to explain the diagnosis could be identified using the concept of Multidimensional Calibration Space. Since the training of the machine learning algorithms is believed to be more efficient when the data of a larger number of patients are employed, the approach presented here is promising for computer-assisted diagnosis.

1. Introduction

There have been considerable efforts to develop biosensors for early diagnosis of cancer [1–5] and other diseases [6–10], especially for screening at low cost with portable instruments, including for point-of-care diagnosis [11–15]. These biosensors may operate with various principles of detection, e.g., with electrical, optical, electrochemical methods (for a review see Ref. [16]) and are targeted at detecting specific biomarkers. For cancer, in particular, immunosensors and genosensors [17–24] have been reported where the biomarkers for diagnosis may be antigens (or antibodies) and genetic material (DNA, RNA), respectively. High sensitivity and selectivity can be achieved owing to the specificity in antibody-antigen interactions and hybridization involving DNA or RNA probes [25–27]. However, limitations related to the biorecognition element have impaired the commercial translation of biosensors into real-world point-of-care diagnostics [28,

* Corresponding author.
E-mail address: chu@ifsc.usp.br (O.N. Oliveira Jr).

https://doi.org/10.1016/j.talanta.2022.123327
Received 1 December 2021; Received in revised form 14 February 2022; Accepted 16 February 2022
Available online 22 February 2022

29]. Proteins denature due to their poor stability, and the production using heterologous cell expressions is complex and costly. Alternatively, receptor-mimicking peptides identified from structural analyses and computational modeling have garnered interest in developing the next-generation biosensors by overperforming whole proteins in terms of stability and resistance to harsh environments. These peptides can be easily and inexpensively produced by chemical synthesis. Nevertheless, their deployment to mimic binding pockets of natural receptors to bind to targets in complex fluids remains challenging [28–30]. These difficulties have sparked research into bioreceptor-free sensing platforms for point-of-care settings [31]. Also relevant for experiments with biological samples with intrinsic high variability, as in blood, urine and saliva, is to treat the data with statistical and computational methods [32–35]. Of special relevance in recent years has been the use of information visualization [36,37] and machine learning [38–41] techniques. With such methods, one may enhance the accuracy of diagnosis by combining the high specificity in the response of biosensors with pattern recognition strategies [42]. Examples of these applications can be found in the diagnosis of breast cancer based on impedance spectra analysis of a microfluidic chip [31], and for prostate cancer where image analysis of biosensing units was carried out with supervised machine learning [43].

The utilization of pattern recognition and machine learning for diagnosis is well established in some areas, which include radiology image analysis [44–48] and genomics [49–52]. This may serve as inspiration for similar applications with sensors and biosensors, for example with electronic tongues (e-tongues) [53–56] and electronic noses (e-noses) [57,58]. E-tongues and e-noses are generally made of sensor arrays that do not detect specific analytes but are rather based on the global selectivity concept [59]. Within this concept, the electrical responses from a few sensing units are combined via statistical methods to establish “fingerprints” of liquids and gases or vapors. The main challenge to using these devices and pattern recognition concepts for diagnosis with biological samples lies in the limited amount of data available to establish unequivocal patterns. Though the sensing data obtained with e-tongues and e-noses certainly have considerable volume, requiring statistical methods beyond a manual analysis, they are normally insufficient for a fully-fledged training procedure in supervised machine learning. One has therefore to be cautious in applying machine learning to avoid overfitting [60,61] and data leakage [62] and combine sensing data with other types of information that may be useful for diagnosis, if at all possible.

In this paper, we report on the use of an e-tongue based on impedance spectroscopy to detect oral cavity cancer, which belongs to the head and neck cancers group, with saliva samples from diagnosed patients. The choice of this type of cancer was motivated by the difficulties in their diagnosis, especially at early stages. The first approach to identify oral cancer remains the conventional oral examination, which consists of a white light visual examination and palpation of the oral cavity surfaces as well as the external facial and neck regions [63]. But this only happens when the disease is already at an advanced stage. The material extracted from the patient is further submitted to biopsy and histopathological examination, the gold standard in the diagnosis of oral cancer. For complex scenarios, the clinical cases are evaluated by multidisciplinary workstations (surgeons, clinical oncologists, radiologists, etc.). However, such methods are invasive and traumatic for patients. Alternative exams have been exploited (exfoliative cytology and Polymerase Chain Reaction (PCR)), but they lack sensitivity and are expensive [64]. Complementary studies using spectroscopy [65] and electrochemical [66] techniques to develop noninvasive and painless methods for oral cancer diagnosis would encourage routine screening tests and increase the chances of early detection. In the literature, there are a few examples of electronic tongues applied for prostate and bladder cancer screening [67–71] with a non-invasive methodology. Urine samples were analyzed by potentiometric and voltammetric techniques, in which chemometric and machine learning tools made the distinguishing task possible. In our study we used saliva from 27 patients (individuals diagnosed with cancer) and healthy volunteers (without any disease). Though this number is small, thus generating a limited amount of data, we were able to obtain a reasonable accuracy in diagnosis with supervised machine learning, especially upon combining impedance data and patient clinical information. Because different types of information were used, we employed the concept of multidimensional calibration space (MCS) [72] to generate the rules that explain the diagnosis results. It is also significant that multidimensional projection techniques and clustering methods with non-supervised machine learning were unable to provide an accurate diagnosis.

2. Experimental

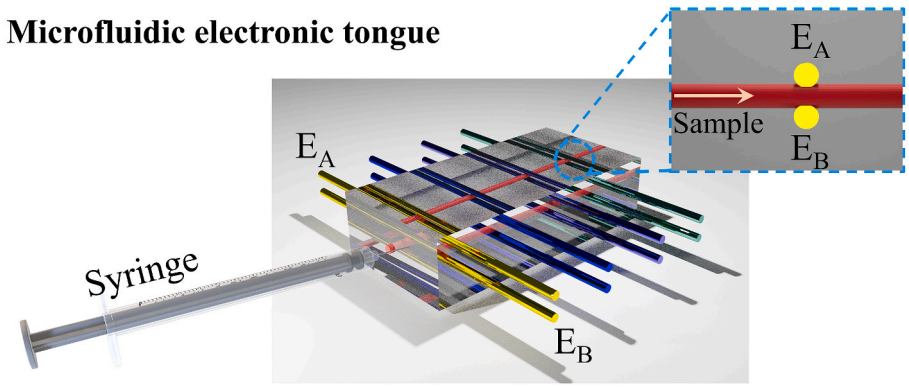
Collection of saliva from patients and clinical data. Saliva samples were collected from patients at the Barretos Cancer Hospital (SP - Brazil; ethics committee approval #468/2011). The collection was made after vigorous mouth washing with 10 mL NaCl (0.9%) aqueous solution during 1–2 min. The patient then spitted the saliva into a 50 mL Falcon tube, which was then centrifuged at 1500 rpm for 10 min at 4 °C. The supernatant was discarded, and the pellet was resuspended with residual leftover supernatant. This new suspension was poured into a 1.5 mL Eppendorf, which was centrifuged at 1500 rpm for 10 min at 4 °C. The supernatant was discarded, and the resulting pellet was stored at −80 °C for further analysis. At the moment of the measurements in the impedance analyzer, each pellet was resuspended in a 200 μL phosphate-buffered solution (PBS, Sigma-Aldrich). The demographic characteristics of patients are summarized in Table S1.

Measurements with the electronic tongue. The e-tongue used in this work is similar to that reported in Ref. [73] with a single response microfluidic device [74–76]. In short, a single piece of PDMS containing four pairs of 304 stainless steel microwires (Treficap, Sao Paulo, Brazil), with a diameter of 700 μm, modified with 800 nm of SiO_2, NiO_2, Al_2O_3, and Fe_2O_3 oxide films, using a UNIVEX 300 electron beam (Oerlikon Leybold Vacuum, Cologne, Germany). The sensing units were short-circuited establishing an array of capacitors connected in parallel, whose distance between the electrodes (E_A and E_B) is the diameter of the microwire used as a template for the microfluidic channel. Scheme 1 depicts the microfluidic electronic tongue device used in this work, here the sample is injected by a syringe that passes between the electrodes above (E_A) and below (E_B) the microchannel, as represented in the enlarged image. Electrical impedance spectroscopy experiments were performed using an impedance analyzer model 1260 A coupled to a dielectric interface model 1296 A (Solartron Analytical, Leicester, England), applying 25 mV ac voltage in the frequency range of 1 Hz to 1 MHz (19 frequencies). Moreover, the experiments were performed under a flow rate of 1000 μL h^{-1} using a syringe pump (New Era Pump Systems Inc., NE-1000, Farmingdale, NY) and a 1 mL plastic syringe. The method used around 500 μL of sample solution to execute the entire measurements. To avoid cross-contamination among samples, after each measurement a washing step was performed three times by injecting 1 mL of ultrapure water.

2.1. Data analysis

Dataset. The impedance spectroscopy data obtained with the e-tongue were analyzed with various methods from the areas of data visualization and machine learning. The raw data consists of 162 capacitance spectra (here just spectra) with samples from 27 individuals (6 measurements per sample), which were labeled as YES (either floor of mouth or other oral cavity tumor subsites) and NO (no tumor) based on prior clinical and pathological diagnoses at Barretos Cancer Hospital. In order to prevent some level of leakage because of the repetition in the spectra acquisition, the spectra of each patient were aggregated through the average, and Table 1 shows the size of the dataset used in all analyses.

Each sample data is composed of 23 features of which 19 come from



Scheme 1. Schematic representation of the microfluidic electronic tongue that comprises four sensing units, forming an association of capacitors in parallel, having one electrode above (E_A) and the other below (E_B) the microchannel through which the sample is injected.

Table 1
Dataset size for each label.

Labels		Number of samples
cancer	Sample	
NO	Control	14
YES	Floor of mouth	4
	Other oral cavity subsites	9
Total		27

the spectra and 4 are related to clinical information. The spectra features are capacitances measured over frequencies ranging from 1 Hz to 1 MHz. The clinical features are smoking (yes, no, former), alcoholism (yes, no, former), gender (male, female), and age (37–78), as given in Table 2. The ages histograms can be seen in Figures S1A through S1C in the Support Information. The dataset is balanced for the label cancer (YES or NO) but unbalanced for the label sample case. For the class 'floor of mouth,' there are no samples of patients with no smoking, yes/no alcoholism, and female gender.

Data Visualization. Dimensionality reduction and projection methods common in the literature were tried to visualize the intrinsic capacity of the spectra features (only-sensor) to exhibit the known group patterns structures (binary and multiclass). The methods used were Principal Component Analysis (PCA) [77], Neighborhood Components Analysis (NCA) [78], t-distributed Stochastic Neighbor Embedding (TSNE) [79], and Interactive Document Mapping (IDMAP) [80].

Machine learning. Clustering and classification machine learning algorithms were employed for the data group discrimination. In order to verify the influence of the clinical features, the analysis was accomplished with spectra features (only-sensor) and with the aggregation of clinical data (all-features). The clustering algorithms used were K-Means (KM) [80], Hierarchical Agglomerative Clustering (HAC) [80], and Spectral Clustering (SC) [81] with default hyperparameters in Scikit-learn module [82]. The metric used to evaluate the performance was the Average Silhouette Width (ASW) [83], which varies between −1 and +1. The closer to +1 the higher quality the clustering has. Each clustering analysis was repeated 100 times, and then the average value and the standard deviation of the AWS were obtained. The classification was performed with the following algorithms: Logistic Regression (LR), Linear Discriminant Analysis (LDA), K-Nearest Neighbor (KNN), Naive Bayes (NB), and Support Vector Machine with kernels: linear (SVML), polynomial (SVMP) and radial basis function (SVMR) [80]. Also applied was the Random Forest (RF) [84] algorithm as an ensemble method. To improve the performance avoiding overoptimistic bias and overfitting, the hyperparameters were tuned and the best model was evaluated with Nested K-Fold Cross-Validation [60,85,86] protocol, which provides average performance (e.g., accuracy) as an estimation of how the classification model will perform on new data instances (not available to the algorithm yet). This approach has been useful when few data instances (samples) are available [85]. It is preferable to a single K-Fold Cross-Validation [87], being a robust [88] and overzealous [89] performance estimation method. On the Nested K-Fold Cross-Validation, two K-Fold Cross-Validation procedures are enclosed. The inner K-Fold Cross-Validation loop is performed for model selection (tuning the model hyperparameters) [90], whereas model performance is carried out by the outer K-Fold Cross-Validation loop [85,86]. Here, there are the kouter and kinner configuration parameters for the outer (evaluation) and inner (tunning) loops respectively. Optimistic (overestimation) and biased performance can be an issue especially on small datasets [85,86], which might be avoided following the Nested K-Fold Cross-Validation procedure [85,86]. Nevertheless, more samples will produce a better and more reliable calibration.

3. Results and discussion

Data Visualization and clustering with non-supervised machine learning. Fig. 1A and B shows the 162 capacitance spectra for all samples, where different colors are used to distinguish the prevalence of cancer and type of cancer. Two important observations can be made from a visual inspection of these figures: it is hard to distinguish the samples by solely inspecting the spectra, and the dataset appears to contain two separate sets of measurements. We found that this latter observation was due to a drift in the impedance spectroscopy measurements which was not related to the samples. It simply occurred because of a drift in the second batch of measurements, performed a few days after the first batch. No reason could be established for the drift, which is an artifact. Under normal circumstances, we would have to perform a novel set of measurements to verify reproducibility and

Table 2
Dataset distribution for each clinical feature and label.

Cancer	case	patients	smoking			alcoholism			gender	
			no	former	yes	no	former	yes	male	female
NO	Control	14	3	3	8	9	4	1	11	3
YES	Floor of mouth tumor	4	0	2	2	0	4	0	4	0
	Other oral cavity tumor subsite	9	1	3	5	1	2	6	7	2

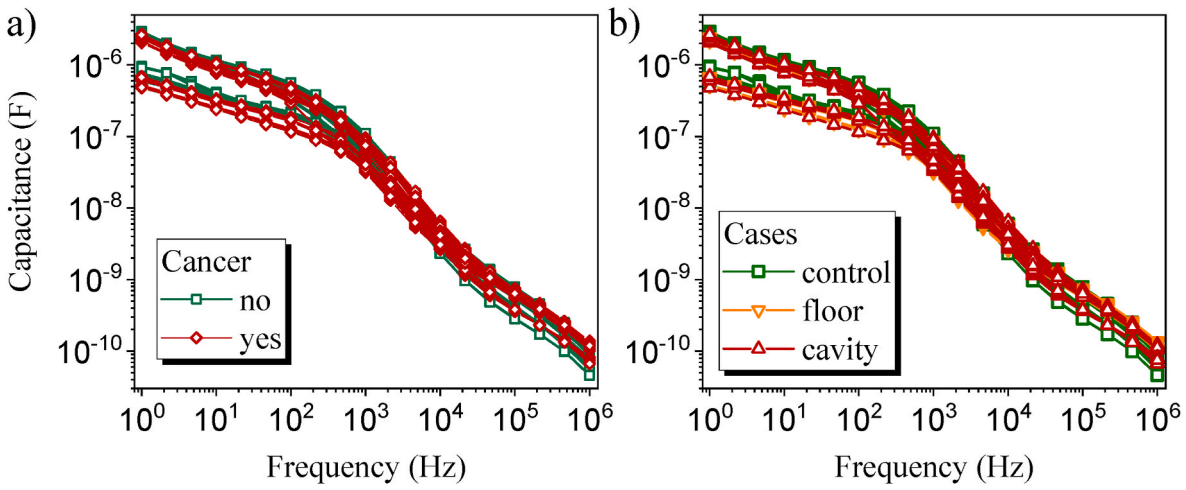


Fig. 1. Capacitance spectra for a) cancer labels, i.e., YES or NO, and b) case, i.e., control, and cancer on the floor or cavity of mouth.

remove the drift. However, our objective in this study is precise to exploit data analysis methods that should be sufficiently robust to classify complex samples, such as those of saliva here, and eliminate possible interferences from experimental artifacts and changes in the environment. Hence, the undesired drift is considered here as a happy coincidence to test the robustness of our analysis approaches.

Attempts to distinguish between different samples using dimensional reduction and multidimensional projection methods, as normally done with e-tongue data, failed. None of the techniques used, viz. PCA, NCA, TSNE, and IDMAP, provided reasonable distinction, as indicated in Figs. S2 and S3 in the Supporting Information. The results from clustering with the non-supervised machine learning algorithms KM, HAC, and SC are shown in Table 3 where the quality of classification was evaluated using ASW. The spectra features were scaled with standardization. The performances of the algorithms were low, even for binary classification, and there is a small improvement when using all features (spectra features and clinical data).

Classification using supervised machine learning. Table 4 shows the average accuracy (standard deviation, SD) values obtained with the algorithms LR, LDA, GNB, KNN, SVML, SVMP, SVMR and RF applied to the situations only-sensor and all-features in binary and multiclass analysis. For all models, the average accuracy was obtained by a 10×5 Nested K-Fold Cross-Validation (kouter = 10 and kinner = 5). As in the clustering, the spectra features were scaled with the standardization method, except for the RF models which were built upon data not pre-processed. For the binary classification, high accuracy values were obtained with SVMR and RF (similar accuracy considering the dispersion).

Table 3
ASW for clustering with KM, HAC, and SC algorithms. The best results for 2- and 3- cluster organizations are highlighted.

Features	Algorithms	Number of clusters	
		2	3
only-sensor	KM	0.431	**0.459**
	HAC	0.443	0.452
	SC	0.443	0.394
all-features	KM	**0.462**	0.426
	HAC	**0.462**	0.428
	SC	0.453	0.401

As expected, for the multiclass analysis the accuracy was considerably lower, which is seen in the last column in Table 4. The most efficient algorithm was RF and the inclusion of clinical features provided a small enhancement in performance.

Since RF was among the most efficient algorithms, it was possible to employ the concept of Multidimensional Calibration Space (MCS) [72] with which one may allow for some degree of predictability in the analysis of new data because rules are generated that provide the reasons for classification. This is especially important for the diagnosis based on the limited body of e-tongue results reported here. The proof-of-principle results do indicate that one may use an e-tongue to distinguish saliva samples from cancer patients from healthy individuals. With an MCS one can go one step further and establish the conditions for classification when a new set of data are analyzed. The concept behind MCS and its use in simple examples are described in the Supporting Information (Section 3).

Fig. 2 presents the MCS for the binary problem (classes NO or YES, for negative and positive for cancer, respectively) using both sensor and clinical data, with ExMatrix [91] where the RF model is represented as logic rules into a matrix visual metaphor. In such representation, rows are rules, columns are features, and cells are the rule predicates. The rule predicates specify range values of capacitance for frequencies obtained from the sensor, as well as ranges for the two possible values (0 and 1) of the clinical features (e.g., 0 or 1 for the feature "alcoholism_no" means negative or positive for a non-alcoholic patient). We employed the complementary features "alcoholism_yes" and "alcoholism_no" as separate features for the convenience of the algorithm implementation. The ranges defined by the rules are related to one of the two classes (NO and YES) mapped as category colors (blue and orange). With rules (rows) ordered by class and coverage and features (columns) by importance, this MCS is composed of 26 dimensions corresponding to 26 selected features (frequencies in the sensing measurements with the e-tongue and clinical data), which provide the best distinguishing ability among samples. The two most important features (first two columns) for RF are frequency 215 Hz and "alcoholism_no", with importance values of 0.156 and 0.123, respectively. According to their high coverage rules, low capacitance values at frequency 215 Hz are related to the class YES (orange color). In the first column, small orange ranges are found at the leftmost, and there are no equally positioned blue ranges. Leftmost orange ranges can also be seen in the second column for high coverage rules, matching the value 0 (negative) for clinical feature "alcoholism_no", while the leftmost blue ranges are not found for such rules. This means that from the point of view of the most generic knowledge with the RF model, patients with alcoholism issues and with low capacitance values at frequency 215 Hz are prone to be classified positive for cancer.

Table 4
Average accuracy (standard deviation) of the classification with LR, LDA, GNB, KNN, SVM-linear, SVM-poly, SVM-rbf, and RF algorithms. The best results for each label are highlighted.

Features	Algorithms	Classification	
		YES/NO for cancer	Control/Floor/Cavity
only-sensor	LR	0.700 (±0.221)	0.481 (±0.105)
	LDA	0.717 (±0.248)	0.519 (±0.105)
	GNB	0.433 (±0.249)	0.370 (±0.052)
	KNN	0.783 (±0.224)	0.630 (±0.105)
	SVML	0.667 (±0.197)	0.481 (±0.052)
	SVMP	0.717 (±0.248)	0.519 (±0.052)
	SVMR	**0.867 (±0.208)**	0.519 (±0.052)
	RF	0.800 (±0.256)	**0.630 (±0.052)**
all-features	LR	0.650 (±0.252)	0.556 (±0.091)
	LDA	0.633 (±0.306)	0.556 (±0.240)
	GNB	0.567 (±0.291)	0.556 (±0.157)
	KNN	0.617 (±0.373)	0.519 (±0.139)
	SVML	0.733 (±0.238)	0.556 (±0.091)
	SVMP	0.733 (±0.186)	0.519 (±0.139)
	SVMR	0.767 (±0.200)	0.519 (±0.052)
	RF	**0.800 (±0.256)**	**0.667 (±0.091)**

*Logistic Regression (LR), Linear Discriminant Analysis (LDA), K-Nearest Neighbor (KNN), Gaussian Naive Bayes (GNB), Support Vector Machine [kernel linear (SVML), kernel polynomial (SVMP), kernel radial basis function (SVMR)], Random Forest (RF).

The MCS in Fig. 2 gives 80% average accuracy with the RF model. The average sensitivity and specificity can also be calculated, being 65% and 90% respectively. The sensitivity regards how well the calibration recognizes true positives, while specificity how well it identifies true negatives. Hence, the calibration will potentially recognize (on average) 65% of the patients that have cancer, and 90% of the healthy patients. An MCS can also be established for the multiclass problem (sample case: control, floor, and cavity), but the average accuracy drops to 66.7%. In this case, many and more complex rules are required, as one should expect.

Several issues should be discussed about the results obtained with the calibration space. First of all, the calibration space found for the binary classification had 26 dimensions, which is considerably higher than for the datasets of other e-tongues. For example, in Ref. [72] calibration spaces had up to 5 dimensions for the full coverage of the dataset, i.e., the prediction accuracy was 100% for multiclass classification. With the dataset analyzed here, the calibration space had only 80% accuracy, despite its 26 dimensions. These results do indicate that more data would be necessary for full coverage of the space, which should be expected because e-tongues do not contain biosensors that could detect cancer biomarkers specifically. Furthermore, the shift in a part of the impedance spectra made it more difficult to achieve an accurate classification. Based on sensitivity and specificity results (65% and 90%), the calibration is more suitable for identifying patients that do not have cancer (healthy). On the other hand, there is a clear indication that e-tongue data can be combined with another type of data (as clinical features used here) to provide a successful diagnosis of cancer and other diseases. With more and better representative samples, higher average accuracy may be achieved, probably with a simpler RF model (i.e. with an MCS with fewer dimensions).

4. Conclusions

We have demonstrated that e-tongue data can be used in cancer diagnosis, even without detecting a specific biomarker. This is made possible because pattern recognition can be applied within the global selectivity paradigm. The difficulty in diagnosing was highlighted by the poor performance of statistical methods and non-supervised learning in distinguishing between the saliva samples of cancer patients and healthy individuals. With supervised machine learning, on the other hand, a reasonable accuracy of ca. 80% for the binary classification (YES or NO for cancer) and ca. 70% when the three classes were considered (floor/cavity cancer, and control). These accuracy values are expected to increase when a larger number of samples are used, from which a more efficient training can be made with the machine learning algorithms. The accuracy tended to increase when clinical information from the patients was used in conjunction with the e-tongue impedance data. This is particularly encouraging for further studies as the combination of data from different natures is a hallmark of the new paradigm of computer-assisted diagnosis [92]. Also promising for future developments is the robustness of the classification approach based on machine learning applied to e-tongue data. The approach may be used in any type of application with reasonable performance even when there are problems and limitations in the data, as was the case here.

Credit author statement

D.C.B., M.P.N., Software, Formal analysis, Investigation, Writing; F. M.S., Conceptualization, Formal analysis, Investigation, Writing; A.C.S., Investigation, Writing; A.L.G., R.S.L., Resources, Writing, M.E.M., L.M. R.B.A., A.L.C., Resources, Methodology, Writing; F.V.P., Writing, Supervision; O.N.O.Jr, Conceptualization, Resources, Writing,

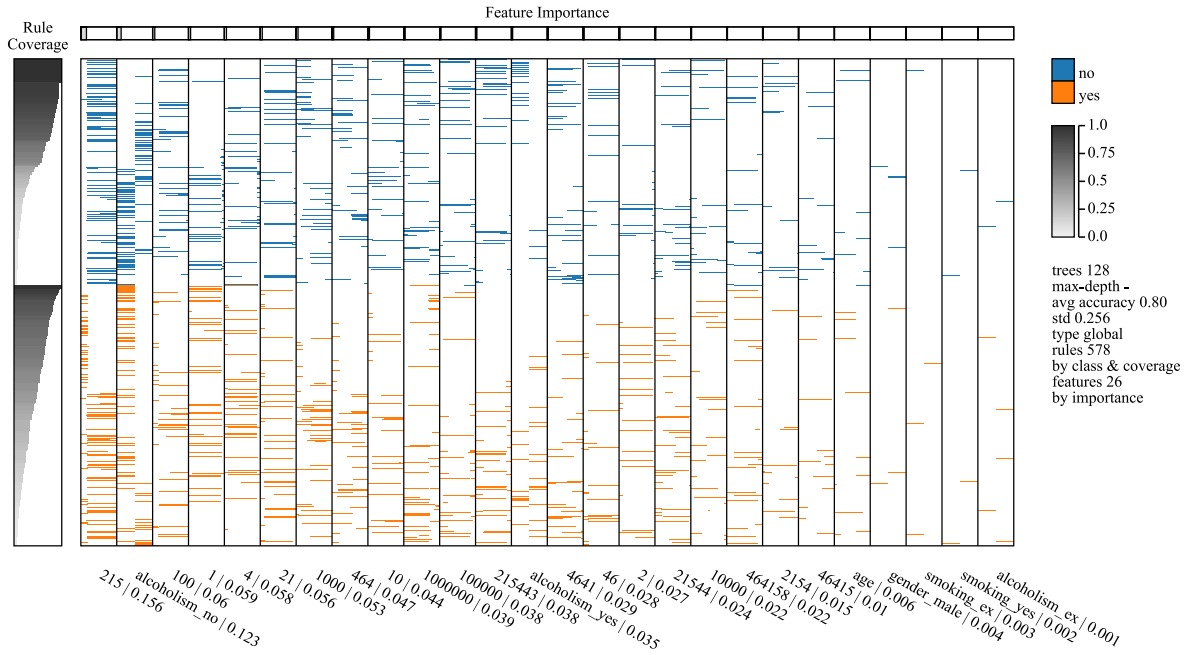


Fig. 2. Multidimensional Calibration Space (MCS) via RF model (128 Decision Trees – 578 logic rules) for the binary problem, with classes NO or YES for negative and positive for cancer, respectively. The space has 26 dimensions which correspond to 19 frequencies (1–1,000,000 Hz) and 7 selected clinic features ("age", "smoking_ex", "smoking_yes", "alcoholism_ex", "alcoholism_no", "alcoholism_yes", and "gender_male"). In this ExMatrix representation for the RF model, logic rules (rows) are ordered by class and coverage, while features (columns) are ordered by importance. Rules predicates are displayed into cells, where features range values are delimited as rectangular shapes and assigned to one of the two possible classes (NO and YES) coded as categorical colors (blue and orange). The first two columns represent the two most important features, namely frequency 215 Hz and the clinic feature "alcoholism_no" (0 or 1 regarding negative or positive), with importance values 0.156 and 0.123. A pattern can be seen for these two features on high coverage rules: the ones with darker and most filled coverage (left column legend). Orange ranges (class YES) are found with leftmost values, while blue ranges are not found in such a region. This means that low values of capacitance at frequency 215 Hz and value 0 (negative) for clinic feature "alcoholism_no" appear to be related to class YES (orange). In summary, from the point of view of the most generic knowledge of the RF model, a patient with alcoholism issues that presents a low capacitance value at frequency 215 Hz from the e-tongue is prone to be diagnosed as positive for cancer. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

Supervision.

Declaration of 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.

Acknowledgements

This work was supported by CAPES, CNPq, INEO, and FAPESP (2018/22214-6). Flavio M. Shimizu thanks the support by CNPq and FAPESP (2012/15543-7). Acelino C. Sá thanks the support by CNPq (153855/2018-5). Daniel C. Braz thanks the Mato Grosso do Sul State University (UEMS) for the training program.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.talanta.2022.123327.

References

[1] G.C. Jensen, C.E. Krause, G.A. Sotzing, J.F. Rusling, Inkjet-printed gold nanoparticle electrochemical arrays on plastic. Application to immunodetection of a cancer biomarker protein, Phys. Chem. Chem. Phys. 13 (2011) 4888, https://doi.org/10.1039/c0cp01755h.

[2] P.M. Kosaka, V. Pini, J.J. Ruz, R.A. da Silva, M.U. González, D. Ramos, M. Calleja, J. Tamayo, Detection of cancer biomarkers in serum using a hybrid mechanical and optoplasmonic nanosensor, Nat. Nanotechnol. 9 (2014) 1047–1053, https://doi.org/10.1038/nnano.2014.250.

[3] M. Soler, M.-C. Estevez, R. Villar-Vazquez, J.I. Casal, L.M. Lechuga, Label-free nanoplasmonic sensing of tumor-associate autoantibodies for early diagnosis of colorectal cancer, Anal. Chim. Acta 930 (2016) 31–38, https://doi.org/10.1016/j.aca.2016.04.059.

[4] B. Wei, K. Mao, N. Liu, M. Zhang, Z. Yang, Graphene nanocomposites modified electrochemical aptamer sensor for rapid and highly sensitive detection of prostate specific antigen, Biosens. Bioelectron. 121 (2018) 41–46, https://doi.org/10.1016/j.bios.2018.08.067.

[5] D. Feng, J. Su, Y. Xu, G. He, C. Wang, X. Wang, T. Pan, X. Ding, X. Mi, DNA tetrahedron-mediated immune-sandwich assay for rapid and sensitive detection of PSA through a microfluidic electrochemical detection system, Microsyst. Nanoeng. 7 (2021) 33, https://doi.org/10.1038/s41378-021-00258-x.

[6] W. Gao, S. Emaminejad, H.Y.Y. Nyein, S. Challa, K. Chen, A. Peck, H.M. Fahad, H. Ota, H. Shiraki, D. Kiriya, D.-H. Lien, G.A. Brooks, R.W. Davis, A. Javey, Fully integrated wearable sensor arrays for multiplexed in situ perspiration analysis, Nature 529 (2016) 509–514, https://doi.org/10.1038/nature16521.

[7] O. Parlak, S.T. Keene, A. Marais, V.F. Curto, A. Salleo, Molecularly selective nanoporous membrane-based wearable organic electrochemical device for noninvasive cortisol sensing, Sci. Adv. 4 (2018), https://doi.org/10.1126/sciadv.aar2904.

[8] M.A. Zamzami, G. Rabbani, A. Ahmad, A.A. Basalah, W.H. Al-Sabban, S. Nate Ahn, H. Choudhry, Carbon nanotube field-effect transistor (CNT-FET)-based biosensor for rapid detection of SARS-CoV-2 (COVID-19) surface spike protein S1, Bioelectrochemistry 143 (2022) 107982, https://doi.org/10.1016/j.bioelechem.2021.107982.

[9] S.-S. Li, C.-W. Lin, K.-C. Wei, C.-Y. Huang, P.-H. Hsu, H.-L. Liu, Y.-J. Lu, S.-C. Lin, H.-W. Yang, C.-C.M. Ma, Non-invasive screening for early Alzheimer's disease diagnosis by a sensitively immunomagnetic biosensor, Sci. Rep. 6 (2016) 25155, https://doi.org/10.1038/srep25155.

[10] H.J. Kim, W. Choi, J. San Lee, J. Choi, N. Choi, K.S. Hwang, Clinical application of serological Alzheimer's disease diagnosis using a highly sensitive biosensor with hydrogel-enhanced dielectrophoretic force, Biosens. Bioelectron. 195 (2022) 113668, https://doi.org/10.1016/j.bios.2021.113668.

[11] A.E. Cetin, A.F. Coskun, B.C. Galarreta, M. Huang, D. Herman, A. Ozcan, H. Altug, Handheld high-throughput plasmonic biosensor using computational on-chip imaging, Light Sci. Appl. 3 (2014), https://doi.org/10.1038/lsa.2014.3 e122–e122.

[12] X. Fu, Z. Cheng, J. Yu, P. Choo, L. Chen, J. Choo, A SERS-based lateral flow assay biosensor for highly sensitive detection of HIV-1 DNA, Biosens. Bioelectron. 78 (2016) 530–537, https://doi.org/10.1016/j.bios.2015.11.099.

[13] G.V. Martins, A.P.M. Tavares, E. Fortunato, M.G.F. Sales, Paper-based sensing device for electrochemical detection of oxidative stress biomarker 8-Hydroxy-2′-

deoxyguanosine (8-OHdG) in point-of-care, Sci. Rep. 7 (2017) 14558, https://doi.org/10.1038/s41598-017-14878-9.
[14] S. Mavrikou, G. Moschopoulou, A. Zafeirakis, K. Kalogeropoulou, G. Giannakos, A. Skevis, S. Kintzios, An ultra-rapid biosensory point-of-care (POC) assay for prostate-specific antigen (PSA) detection in human serum, Sensors 18 (2018) 3834, https://doi.org/10.3390/s18113834.
[15] D. Song, J. Liu, W. Xu, X. Han, H. Wang, Y. Cheng, Y. Zhuo, F. Long, Rapid and quantitative detection of SARS-CoV-2 IgG antibody in serum using optofluidic point-of-care testing fluorescence biosensor, Talanta 235 (2021) 122800, https://doi.org/10.1016/j.talanta.2021.122800.
[16] V. Naresh, N. Lee, A review on biosensors and recent development of nanostructured materials-enabled biosensors, Sensors 21 (2021) 1109, https://doi.org/10.3390/s21041109.
[17] K. Kadimisetty, I.M. Mosa, S. Malla, J.E. Satterwhite-Warden, T.M. Kuhns, R. C. Faria, N.H. Lee, J.F. Rusling, 3D-printed supercapacitor-powered electrochemiluminescent protein immunoarray, Biosens. Bioelectron. 77 (2016) 188–193, https://doi.org/10.1016/j.bios.2015.09.017.
[18] R.C.B. Marques, E. Costa-Rama, S. Viswanathan, H.P.A. Nouws, A. Costa-García, C. Delerue-Matos, M.B. González-García, Voltammetric immunosensor for the simultaneous analysis of the breast cancer biomarkers CA 15-3 and HER2-ECD, Sensor. Actuator. B Chem. 255 (2018) 918–925, https://doi.org/10.1016/j.snb.2017.08.107.
[19] J.C. Soares, L.E.O. Iwaki, A.C. Soares, V.C. Rodrigues, M.E. Melendez, J.H.T. G. Fregnani, R.M. Reis, A.L. Carvalho, D.S. Corrêa, O.N. Oliveira Jr, Immunosensor for pancreatic cancer based on electrospun nanofibers coated with carbon nanotubes or gold nanoparticles, ACS Omega 2 (2017) 6975–6983, https://doi.org/10.1021/acsomega.7b01029.
[20] A.B. Ganganboina, R.-A. Doong, Graphene quantum dots decorated gold-polyaniline nanowire for impedimetric detection of carcinoembryonic antigen, Sci. Rep. 9 (2019) 7214, https://doi.org/10.1038/s41598-019-43740-3.
[21] E. Er, A. Sánchez-Iglesias, A. Silvestri, B. Arnaiz, L.M. Liz-Marzán, M. Prato, A. Criado, Metal nanoparticles/MoS $_2$ surface-enhanced Raman scattering-based sandwich immunoassay for α-fetoprotein detection, ACS Appl. Mater. Interfaces 13 (2021) 8823–8831, https://doi.org/10.1021/acsami.0c22203.
[22] S. Ramanathan, S.C.B. Gopinath, M.K.M. Arshad, P. Poopalan, P. Anbu, T. Lakshmipriya, F.H. Kasim, Aluminosilicate nanocomposite on genosensor: a prospective voltammetry platform for epidermal growth factor receptor mutant analysis in non-small cell lung cancer, Sci. Rep. 9 (2019) 17013, https://doi.org/10.1038/s41598-019-53573-9.
[23] R. Sánchez-Salcedo, R. Miranda-Castro, N. de-los-Santos-Álvarez, M.J. Lobo-Castañón, Dual electrochemical genosensor for early diagnosis of prostate cancer through lncRNAs detection, Biosens. Bioelectron. 192 (2021) 113520, https://doi.org/10.1016/j.bios.2021.113520.
[24] P.U. Alves, R. Vinhas, A.R. Fernandes, S.Z. Birol, L. Trabzon, I. Bernacka-Wojcik, R. Igreja, P. Lopes, P.V. Baptista, H. Águas, E. Fortunato, R. Martins, Multifunctional microfluidic chip for optical nanoprobe based RNA detection – application to Chronic Myeloid Leukemia, Sci. Rep. 8 (2018) 381, https://doi.org/10.1038/s41598-017-18725-9.
[25] A. Hidalgo, M. Baudis, I. Petersen, H. Arreola, P. Piña, G. Vázquez-Ortiz, D. Hernández, J. González, M. Lazos, R. López, C. Pérez, J. García, K. Vázquez, B. Alatorre, M. Salcedo, Microarray comparative genomic hybridization detection of chromosomal imbalances in uterine cervix carcinoma, BMC Cancer 5 (2005) 77, https://doi.org/10.1186/1471-2407-5-77.
[26] H. Kawanishi, T. Takahashi, M. Ito, J. Watanabe, S. Higashi, T. Kamoto, T. Habuchi, T. Kadowaki, G. Tsujimoto, H. Nishiyama, O. Ogawa, High throughput comparative genomic hybridization array analysis of multifocal urothelial cancers, Cancer Sci. 97 (2006) 746–752, https://doi.org/10.1111/j.1349-7006.2006.00259.x.
[27] M. Freitas, H.P.A. Nouws, C. Delerue-Matos, Electrochemical sensing platforms for HER2-ECD breast cancer biomarker detection, Electroanalysis 31 (2019) 121–128, https://doi.org/10.1002/elan.201800537.
[28] G. Liu, J.F. Rusling, COVID-19 antibody tests and their limitations, ACS Sens. 6 (2021) 593–612, https://doi.org/10.1021/acssensors.0c02621.
[29] Y.Y. Broza, X. Zhou, M. Yuan, D. Qu, Y. Zheng, R. Vishinkin, M. Khatib, W. Wu, H. Haick, Disease detection with molecular biomarkers: from chemistry of body fluids to nature-inspired chemical sensors, Chem. Rev. 119 (2019) 11761–11817, https://doi.org/10.1021/acs.chemrev.9b00437.
[30] O.S. Kwon, H.S. Song, T.H. Park, J. Jang, Conducting nanomaterial sensor using natural receptors, Chem. Rev. 119 (2019) 36–93, https://doi.org/10.1021/acs.chemrev.8b00159.
[31] C.Y.N. Nicoliche, R.A.G. de Oliveira, G.S. da Silva, L.F. Ferreira, I.L. Rodrigues, R. C. Faria, A. Fazzio, E. Carrilho, L.G. de Pontes, G.R. Schleder, R.S. Lima, Converging multidimensional sensor and machine learning toward high-throughput and biorecognition element-free multidetermination of extracellular vesicle biomarkers, ACS Sens. 5 (2020) 1864–1871, https://doi.org/10.1021/acssensors.0c00599.
[32] B.M.H. Kumar, P.C. Srikanth, A.M. Vaibhav, A novel computation method for detection of Malaria in RBC using Photonic biosensor, Int. J. Inf. Technol. 13 (2021) 2053–2058, https://doi.org/10.1007/s41870-021-00782-z.
[33] D. Lin, S. Feng, J. Pan, Y. Chen, J. Lin, G. Chen, S. Xie, H. Zeng, R. Chen, Colorectal cancer detection by gold nanoparticle based surface-enhanced Raman spectroscopy of blood serum and statistical analysis, Opt Express 19 (2011) 13565, https://doi.org/10.1364/OE.19.013565.
[34] A. Kamińska, K. Winkler, A. Kowalska, E. Witkowska, T. Szymborski, A. Janeczek, J. Waluk, SERS-based immunoassay in a microfluidic system for the multiplexed recognition of interleukins from blood plasma: towards picogram detection, Sci. Rep. 7 (2017) 10656, https://doi.org/10.1038/s41598-017-11152-w.
[35] S.J. Ward, R. Layouni, S. Arshavsky-Graham, E. Segal, S.M. Weiss, Morlet wavelet filtering and phase Analysis to reduce the limit of detection for thin film optical biosensors, ACS Sens. 6 (2021) 2967–2978, https://doi.org/10.1021/acssensors.1c00787.
[36] F.V. Paulovich, M.L. Moraes, R.M. Maki, M. Ferreira, O.N. Oliveira Jr, M.C.F. De Oliveira, Information visualization techniques for sensing and biosensing, Analyst 136 (2011) 1344–1350, https://doi.org/10.1039/c0an00822b.
[37] S. Jafarinejad, M. Ghazi-Khansari, F. Ghasemi, P. Sasanpour, M.R. Hormozi-Nezhad, Colorimetric fingerprints of gold nanorods for discriminating catecholamine neurotransmitters in urine samples, Sci. Rep. 7 (2017) 8266, https://doi.org/10.1038/s41598-017-08704-5.
[38] H.M. Robison, C.A. Chapman, H. Zhou, C.L. Erskine, E. Theel, T. Peikert, C. S. Lindestam Arlehamn, A. Sette, C. Bushell, M. Welge, R. Zhu, R.C. Bailey, P. Escalante, Risk assessment of latent tuberculosis infection through a multiplexed cytokine biosensor assay and machine learning feature selection, Sci. Rep. 11 (2021) 20544, https://doi.org/10.1038/s41598-021-99754-3.
[39] H. Kim, S. Park, I.G. Jeong, S.H. Song, Y. Jeong, C.-S. Kim, K.H. Lee, Noninvasive precision screening of prostate cancer by urinary multimarker sensor and artificial intelligence analysis, ACS Nano 15 (2021) 4054–4065, https://doi.org/10.1021/acsnano.0c06946.
[40] K.J. Squire, Y. Zhao, A. Tan, K. Sivashanmugan, J.A. Kraai, G.L. Rorrer, A.X. Wang, Photonic crystal-enhanced fluorescence imaging immunoassay for cardiovascular disease biomarker screening with machine learning analysis, Sensor. Actuator. B Chem. 290 (2019) 118–124, https://doi.org/10.1016/j.snb.2019.03.102.
[41] R. Fernandez Rojas, X. Huang, K.-L. Ou, A machine learning approach for the identification of a biomarker of human pain using fNIRS, Sci. Rep. 9 (2019) 5645, https://doi.org/10.1038/s41598-019-42098-w.
[42] N. Banaei, J. Moshfegh, A. Mohseni-Kabir, J.M. Houghton, Y. Sun, B. Kim, Machine learning algorithms enhance the specificity of cancer biomarker detection using SERS-based immunoassays in microfluidic chips, RSC Adv. 9 (2019) 1859–1868, https://doi.org/10.1039/C8RA08930B.
[43] V.C. Rodrigues, J.C. Soares, A.C. Soares, D.C. Braz, M.E. Melendez, L.C. Ribas, L.F. S. Scabini, O.M. Bruno, A.L. Carvalho, R.M. Reis, R.C. Sanfelice, O.N. Oliveira Jr, Electrochemical and optical detection and machine learning applied to images of genosensors for diagnosis of prostate cancer with the biomarker PCA3, Talanta 222 (2021) 121444, https://doi.org/10.1016/j.talanta.2020.121444.
[44] X.-P. Zhang, Z.-L. Wang, L. Tang, Y.-S. Sun, K. Cao, Y. Gao, Support vector machine model for diagnosis of lymph node metastasis in gastric cancer with multidetector computed tomography: a preliminary study, BMC Cancer 11 (2011) 10, https://doi.org/10.1186/1471-2407-11-10.
[45] N. Lay, Y. Tsehay, M.D. Greer, B. Turkbey, J.T. Kwak, P.L. Choyke, P. Pinto, B. J. Wood, R.M. Summers, Detection of prostate cancer in multiparametric MRI using random forest with instance weighting, J. Med. Imaging 4 (2017), 024506, https://doi.org/10.1117/1.JMI.4.2.024506.
[46] A. Masood, B. Sheng, P. Li, X. Hou, X. Wei, J. Qin, D. Feng, Computer-assisted decision support system in pulmonary cancer detection and stage classification on CT images, J. Biomed. Inf. 79 (2018) 117–128, https://doi.org/10.1016/j.jbi.2018.01.005.
[47] Y.-J. Yu-Jen Chen, K.-L. Hua, C.-H. Hsu, W.-H. Cheng, S.C. Hidayati, Computer-aided classification of lung nodules on computed tomography images via deep learning technique, OncoTargets Ther. 2015 (2015), https://doi.org/10.2147/OTT.S80733.
[48] C. Zhang, J. Li, J. Huang, S. Wu, Computed tomography image under convolutional neural network deep learning algorithm in pulmonary nodule detection and lung function examination, J. Healthc. Eng. 2021 (2021) 1–9, https://doi.org/10.1155/2021/3417285.
[49] F. Hosseinzadeh, M. Ebrahimi, B. Goliaei, N. Shamabadi, Classification of lung cancer tumors based on structural and physicochemical properties of proteins by bioinformatics models, PLoS One 7 (2012), e40017, https://doi.org/10.1371/journal.pone.0040017.
[50] S.-W. Chang, S. Abdul-Kareem, A.F. Merican, R.B. Zain, Oral cancer prognosis based on clinicopathologic and genomic markers using a hybrid of feature selection and machine learning methods, BMC Bioinf. 14 (2013) 170, https://doi.org/10.1186/1471-2105-14-170.
[51] F.M. Alakwaa, K. Chaudhary, L.X. Garmire, Deep learning accurately predicts estrogen receptor status in breast cancer metabolomics data, J. Proteome Res. 17 (2018) 337–347, https://doi.org/10.1021/acs.jproteome.7b00595.
[52] J. Abraham, A.B. Heimberger, J. Marshall, E. Heath, J. Drabick, A. Helmstetter, J. Xiu, D. Magee, P. Stafford, C. Nabhan, S. Antani, C. Johnston, M. Oberley, W. M. Korn, D. Spetzler, Machine learning analysis using 77,044 genomic and transcriptomic profiles to accurately predict tumor type, Transl. Oncol. 14 (2021) 101016, https://doi.org/10.1016/j.tranon.2021.101016.
[53] I. Yaroshenko, D. Kirsanov, L. Kartsova, A. Sidorova, I. Borisova, A. Legin, Determination of urine ionic composition with potentiometric multisensor system, Talanta 131 (2015) 556–561, https://doi.org/10.1016/J.TALANTA.2014.08.030.
[54] G.O. Silva, Z.P. Michael, L. Bian, G.V. Shurin, M. Mulato, M.R. Shurin, A. Star, Nanoelectronic discrimination of nonmalignant and malignant cells using nanotube field-effect transistors, ACS Sens. 2 (2017) 1128–1132, https://doi.org/10.1021/acssensors.7b00383.
[55] D. Ortiz-Aguayo, X. Cetó, K. De Wael, M. del Valle, Resolution of opiate illicit drugs signals in the presence of some cutting agents with use of a voltammetric sensor array and machine learning strategies, Sensor. Actuator. B Chem. 357 (2022) 131345, https://doi.org/10.1016/j.snb.2021.131345.

[56] D. Ortiz-Aguayo, K. De Wael, M. del Valle, Voltammetric sensing using an array of modified SPCE coupled with machine learning strategies for the improved identification of opioids in presence of cutting agents, J. Electroanal. Chem. 902 (2021) 115770, https://doi.org/10.1016/j.jelechem.2021.115770.
[57] W. Liao, A. Zhang, S. Shih, Machine learning methods applied to predict ventilator-associated pneumonia with Pseudomonas aeruginosa infection via sensor array of electronic nose in intensive care unit, Sensors 19 (2019) 1866, https://doi.org/10.3390/s19081866.
[58] C.-Y. Chen, W.-C. Lin, H.-Y. Yang, Diagnosis of ventilator-associated pneumonia using electronic nose sensor array signals: solutions to improve the application of machine learning in respiratory research, Respir. Res. 21 (2020) 45, https://doi.org/10.1186/s12931-020-1285-6.
[59] A. Riul Jr, C.A.R. Dantas, C.M. Miyazaki, O.N. Oliveira Jr, Recent advances in electronic tongues, Analyst 135 (2010) 2481, https://doi.org/10.1039/c0an00292e.
[60] G.C. Cawley, N.L.C. Talbot, On over-fitting in model selection and subsequent selection bias in performance evaluation, J. Mach. Learn. Res. 11 (2010) 2079–2107.
[61] P. Domingos, A few useful things to know about machine learning, Commun. ACM 55 (2012) 78–87, https://doi.org/10.1145/2347736.2347755.
[62] S. Kaufman, S. Rosset, C. Perlich, O. Stitelman, Leakage in data mining, ACM Trans. Knowl. Discov. Data 6 (2012) 1–21, https://doi.org/10.1145/2382577.2382579.
[63] T. Walsh, J.L. Liu, P. Brocklehurst, A.-M. Glenny, M. Lingen, A.R. Kerr, G. Ogden, S. Warnakulasuriya, C. Scully, Clinical assessment to screen for the detection of oral cavity cancer and potentially malignant disorders in apparently healthy adults, Cochrane Database Syst. Rev. (2013), https://doi.org/10.1002/14651858.CD010173.pub2.
[64] I.F.P. Lima, L.M. Brand, J.A.P. de Figueiredo, L. Steier, M.L. Lamers, Use of autofluorescence and fluorescent probes as a potential diagnostic tool for oral cancer: a systematic review, Photodiagnosis Photodyn. Ther. 33 (2021) 102073, https://doi.org/10.1016/j.pdpdt.2020.102073.
[65] W. Tan, L. Sabet, Y. Li, T. Yu, P.R. Klokkevold, D.T. Wong, C.-M. Ho, Optical protein sensor for detecting cancer markers in saliva, Biosens. Bioelectron. 24 (2008) 266–271, https://doi.org/10.1016/j.bios.2008.03.037.
[66] Y.-T. Lin, S. Darvishi, A. Preet, T.-Y. Huang, S.-H. Lin, H.H. Girault, L. Wang, T.-E. Lin, A review: electrochemical biosensors for oral cancer, Chemosensors 8 (2020) 54, https://doi.org/10.3390/chemosensors8030054.
[67] V. Deev, S. Solovieva, E. Andreev, V. Protoshchak, E. Karpushchenko, A. Sleptsov, L. Kartsova, E. Bessonova, A. Legin, D. Kirsanov, Prostate cancer screening using chemometric processing of GC–MS profiles obtained in the headspace above urine samples, J. Chromatogr. B. 1155 (2020) 122298, https://doi.org/10.1016/j.jchromb.2020.122298.
[68] E. Martynko, E. Oleneva, E. Andreev, S. Savinov, S. Solovieva, V. Protoshchak, E. Karpushchenko, A. Sleptsov, V. Panchuk, A. Legin, D. Kirsanov, Non-invasive prostate cancer screening using chemometric processing of macro and trace element concentration profiles in urine, Microchem. J. 159 (2020) 105464, https://doi.org/10.1016/j.microc.2020.105464.
[69] R. Belugina, E. Karpushchenko, A. Sleptsov, V. Protoshchak, A. Legin, D. Kirsanov, Developing non-invasive bladder cancer screening methodology through potentiometric multisensor urine analysis, Talanta 234 (2021) 122696, https://doi.org/10.1016/j.talanta.2021.122696.
[70] S. Solovieva, M. Karnaukh, V. Panchuk, E. Andreev, L. Kartsova, E. Bessonova, A. Legin, P. Wang, H. Wan, I. Jahatspanian, D. Kirsanov, Potentiometric multisensor system as a possible simple tool for non-invasive prostate cancer diagnostics through urine analysis, Sensor. Actuator. B Chem. 289 (2019) 42–47, https://doi.org/10.1016/j.snb.2019.03.072.
[71] L. Pascual, I. Campos, J.-L. Vivancos, G. Quintás, A. Loras, M.C. Martínez-Bisbal, R. Martínez-Máñez, F. Boronat, J.L. Ruiz-Cerdà, Detection of prostate cancer using a voltammetric electronic tongue, Analyst 141 (2016) 4562–4567, https://doi.org/10.1039/C6AN01044J.
[72] M. Popolin Neto, A.C. Soares, O.N. Oliveira Jr, F.V. Paulovich, Machine learning used to create a multidimensional calibration space for sensing and biosensing data, Bull. Chem. Soc. Jpn. 94 (2021) 1553–1562, https://doi.org/10.1246/bcsj.20200359.
[73] F.M. Shimizu, A.M. Pasqualeti, F.R. Todão, J.F.A. de Oliveira, L.C.S. Vieira, S.P. C. Gonçalves, G.H. da Silva, M.B. Cardoso, A.L. Gobbi, D.S.T. Martinez, O. N. Oliveira Jr, R.S. Lima, Monitoring the surface chemistry of functionalized nanomaterials with a microfluidic electronic tongue, ACS Sens. 3 (2018) 716–726, https://doi.org/10.1021/acssensors.8b00056.
[74] R.A.G. de Oliveira, C.Y.N. Nicoliche, A.M. Pasqualeti, F.M. Shimizu, I.R. Ribeiro, M.E. Melendez, A.L. Carvalho, A.L. Gobbi, R.C. Faria, R.S. Lima, Low-cost and rapid-production microfluidic electrochemical double-layer capacitors for fast and sensitive breast cancer diagnosis, Anal. Chem. 90 (2018) 12377–12384, https://doi.org/10.1021/acs.analchem.8b02605.
[75] C.Y.N. Nicoliche, G.F. Costa, A.L. Gobbi, F.M. Shimizu, R.S. Lima, Pencil graphite core for pattern recognition applications, Chem. Commun. (2019), https://doi.org/10.1039/C9CC01595G.
[76] G.S. da Silva, L.P. de Oliveira, G.F. Costa, G.F. Giordano, C.Y.N. Nicoliche, A.A. da Silva, L.U. Khan, G.H. da Silva, A.L. Gobbi, J.V. Silveira, A.G.S. Filho, G. R. Schleder, A. Fazzio, D.S.T. Martinez, R.S. Lima, Ordinary microfluidic electrodes combined with bulk nanoprobe produce multidimensional electric double-layer capacitances towards metal ion recognition, Sensor. Actuator. B Chem. 305 (2020) 127482, https://doi.org/10.1016/j.snb.2019.127482.
[77] R.O. Duda, P.E. Hart, D.G. Stork, Pattern Classification, second ed., second ed., Wiley-Interscience, 2000.
[78] J. Goldberger, S.T. Roweis, G.E. Hinton, R.R. Salakhutdinov, Neighbourhood Components analysis, in: Proc. 17th Int. Conf. Neural Inf. Process. Syst., MIT Press, 2004, pp. 513–520.
[79] L. van der Maaten, G. Hinton, Visualizing Data using t-SNE, J. Mach. Learn. Res. 9 (2008) 2579–2605. http://jmlr.org/papers/v9/vandermaaten08a.html.
[80] A.R. Webb, Statistical Pattern Recognition, second ed., Wiley, 2002 https://doi.org/10.1002/0470854774.
[81] A.Y. Ng, M.I. Jordan, Y. Weiss, On spectral clustering: analysis and an algorithm, in: Proc. 14th Int. Conf. Neural Inf. Process. Syst. Nat. Synth., MIT Press, Cambridge, MA, USA, 2001, pp. 849–856.
[82] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, É. Duchesnay, Scikit-learn: machine learning in Python, J. Mach. Learn. Res. 12 (2011) 2825–2830.
[83] P.J. Rousseeuw, Silhouettes, A graphical aid to the interpretation and validation of cluster analysis, J. Comput. Appl. Math. 20 (1987) 53–65, https://doi.org/10.1016/0377-0427(87)90125-7.
[84] L. Breiman, Random forests, Mach. Learn. 45 (2001) 5–32, https://doi.org/10.1023/A:1010933404324.
[85] S. Varma, R. Simon, Bias in error estimation when using cross-validation for model selection, BMC Bioinf. 7 (2006) 91, https://doi.org/10.1186/1471-2105-7-91.
[86] I. Tsamardinos, A. Rakhshani, V. Lagani, Performance-Estimation Properties of Cross-Validation-Based Protocols with Simultaneous Hyper-Parameter Optimization, 2014, pp. 1–14, https://doi.org/10.1007/978-3-319-07064-3_1.
[87] R. Kohavi, A study of cross-validation and bootstrap for accuracy estimation and model selection, in: Proc. 14th Int. Jt. Conf. Artif. Intell., vol. 2, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1995, pp. 1137–1143.
[88] M. Ellinger, I. Merbach, U. Werban, M. Ließ, Error propagation in spectrometric functions of soil organic carbon, SOIL 5 (2019) 275–288, https://doi.org/10.5194/soil-5-275-2019.
[89] J. Wainer, G. Cawley, Nested cross-validation when selecting classifiers is overzealous for most practical applications, Expert Syst. Appl. 182 (2021) 115222, https://doi.org/10.1016/j.eswa.2021.115222.
[90] G. James, D. Witten, T. Hastie, R. Tibshirani, An Introduction to Statistical Learning, Springer New York, New York, NY, 2013, https://doi.org/10.1007/978-1-4614-7138-7.
[91] M. Popolin Neto, F.V. Paulovich, Explainable matrix - visualization for global and local interpretability of random forest classification ensembles, IEEE Trans. Visual. Comput. Graph. 27 (2021) 1427–1437, https://doi.org/10.1109/TVCG.2020.3030354.
[92] J.F. Rodrigues, F.V. Paulovich, M.C. de Oliveira, O.N. Oliveira Jr, On the convergence of nanotechnology and Big Data analysis for computer-aided diagnosis, Nanomedicine 11 (2016) 959–982, https://doi.org/10.2217/nnm.16.35.