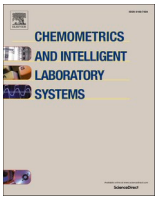
Chemometrics and Intelligent Laboratory Systems 241 (2023) 104938

Contents lists available at ScienceDirect

Chemometrics and Intelligent Laboratory Systems

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



GCIMS: An R package for untargeted gas chromatography – Ion mobility spectrometry data processing

S. Oller-Moreno [a,c,1], C. Mallafré-Muro [a,c,*,1], L. Fernandez [a,c], E. Caballero [a], A. Blanco [a], J. Gumà [b], S. Marco [a,c], A. Pardo [c,**]

[a] Signal and Information Processing for Sensing Systems, Institute for Bioengineering of Catalonia -The Barcelona Institute of Science and Technology (IBEC-BIST), Baldiri Reixac 10-12, 08028, Barcelona, Spain

[b] Oncology Department, Hospital Universitari Sant Joan de Reus, Institut d'Investigació Sanitària Pere Virgili (IISPV), Universitat Rovira i Virgili (URV), 43204, Reus, Spain

[c] Department of Electronics and Biomedical Engineering, Universitat de Barcelona, Marti i Franqués 1, 08028, Barcelona, Spain

A B S T R A C T

Gas-Chromatography coupled to Ion Mobility Spectrometry (GC-IMS) based metabolomics is an emerging technique for obtaining fast, reliable untargeted metabolic fingerprints of biofluids. The generated raw data is highly dimensional and complex, suffers from baseline problems, misalignments, long peak tails and strong non-linearities that must be corrected to extract chemically relevant features from samples. In this work, we present our GCIMS R package, which includes spectra loading, metadata handling, denoising, baseline correction, spectral and chromatographic alignment, peak detection, integration, and peak clustering to produce a peak table ready for multivariate data analysis. We discuss package design decisions, and, for illustration purposes, we show a case study of sex discrimination on the basis of the volatile compounds in urine samples. The GCIMS package provides a user-friendly workflow for non-code developers to process their raw data samples.

1. Introduction

Volatilomics studies the fraction of volatile metabolites (the volatilome) present in biological systems [1]. The human volatilome includes more than 1000 volatile organic compounds (VOCs) present in our body excretions [2,3]. Typical proportions of VOCs in the volatilome lead to what we consider normal smells for breath, saliva, sweat, milk, blood, semen, urine, faeces, etc. It is a well-known fact that health condition modifies human volatilome [4]. For this reason, the smell of biofluids has been used to diagnose diseases since the antiquity. For instance, fruity scented breath can indicate diabetes [5], typhus skin infection smells like fresh-baking bread [6], bladder infection (cystitis) by *E. coli* causes cloudy, foul-smelling urine [7], and so on. Several applications of volatilomics for early diagnose of disease been reported in recent times [8–10]. Also, volatilomics can be used to infer drug concentration in human body. More specifically, breath analysis enables a non-invasive but indirect monitoring of intravenous drugs based on the correlation of drug concentration in blood and breath [4].

Volatilome analysis is usually performed acquiring data from samples using hyphenated analytical chemistry techniques. Note that, for 'hyphenated', we understand a technique that is a combination of two independent analytical techniques. Commonly, chromatographic techniques to separate gas mixtures are coupled to spectrographic techniques to characterize the different compounds of the mixture. The most popular characterization techniques in volatilomics are gas chromatography-mass spectrometry (GC-MS), gas chromatography with flame ionization detection (GC-FID), two-dimensional gas chromatography combined by high resolution time of flight spectrometry (GC x GC-TOF-MS), and more recently, gas chromatography ion mobility spectrometry [11–13].

GC-IMS is a fast, sensitive and moderate-cost analytical technique for VOCs separation and detection [14]. In such type of instruments, chromatographic separation is generally achieved using multi-capillary columns (MCC) operating at isothermal conditions [15]. Then, the sample is ionized and accelerated by a constant electric field against a constant drift gas flow (typically nitrogen). As a result, the ions in the sample travel through the drift tube of the instrument at a constant speed that is proportional to the applied electric field. The constant of

* Corresponding author. Signal and Information Processing for Sensing Systems, Institute for Bioengineering of Catalonia -The Barcelona Institute of Science and Technology (IBEC-BIST), Baldiri Reixac 10-12, 08028, Barcelona, Spain.

** Corresponding author.

E-mail addresses: cmallafre@ibecbarcelona.eu (C. Mallafré-Muro), a.pardo@ub.edu (A. Pardo).

[1] Equally contributed to this work.

https://doi.org/10.1016/j.chemolab.2023.104938

Received 31 July 2023; Received in revised form 16 August 2023; Accepted 20 August 2023

Available online 25 August 2023

proportionality between the speed of the ion and the electric field is known as mobility, and it characterizes the ion. A sequence of ion mobility recordings constitutes an ion mobility spectrum. Alternatively, one can characterize the volatiles in a mixture by determining the time required for the different ions to traverse the drift tube (that is determining their drift time) [16,17]. GC-IMS has recently been explored in biomedical applications such as the recognition of bacterial growth and pathogen differentiation in blood cultures [18], the detection of bacterial respiratory tract infection in breath [19], as well as for COVID-19 diagnosis [20], among others [21,22]. The technique has been also successfully employed for assessing the quality of alimentary products such as honey, wine, olive oil, and for preventing labelling frauds [23–26].

Despite the advantages of GC-IMS, the technique still presents several drawbacks that limit its usability, especially if volatilome analysis is not performed by highly trained personnel. First, GC-IMS data are highly dimensional, and their chemical information content sparse [27]. The previous statement means that a single sample measurement can contain thousands of features, with only a small fraction of which providing chemical information (the two-dimensional peaks associated to ions). Second, peaks in GC-IMS spectra can be masked by high levels of noise for low ion concentrations [28]. Third, data readings can be affected by uncontrolled changes of experimental conditions such as humidity and temperature. That is, humidity modifies the shapes of peaks in IMS spectra [29], while temperature changes the position of peaks both in chromatographic and drift time axes [30]. So, daily and seasonal environmental variations make the instrument drift over time. Forth, GC-IMS data suffers from baseline problems due to several factors, namely, background contamination, chromatographic column bleeding, reactant ion peak tailing [28,30–32]. And fifth, the instrument response to metabolite concentration is highly non-linear [33], hampering the quantification of the volatilome. To overcome these problems, signal pre-processing techniques for feature extraction followed by machine learning are required [34,35]. The available tools for data treatment are usually provided by the instrument vendors (e.g. GAS Dortmund - VOCal Software [36]). However, commercial tools are non-versatile closed solutions linked to the instrument and offer simplified data processing workflows. Few attempts have been made to improve the quality of GC-IMS data processing by providing full workflows that take raw data and provide complete peak tables for further statistical analysis and the development of machine learning based predictive models. The authors previously described a full workflow for GC-IMS data processing implemented in MATLAB and demonstrated the application in foodomics [31]. Recently, a solution has been disclosed as open source for the research community as a Python package [37]. This package implements a simple pre-processing workflow to use the full

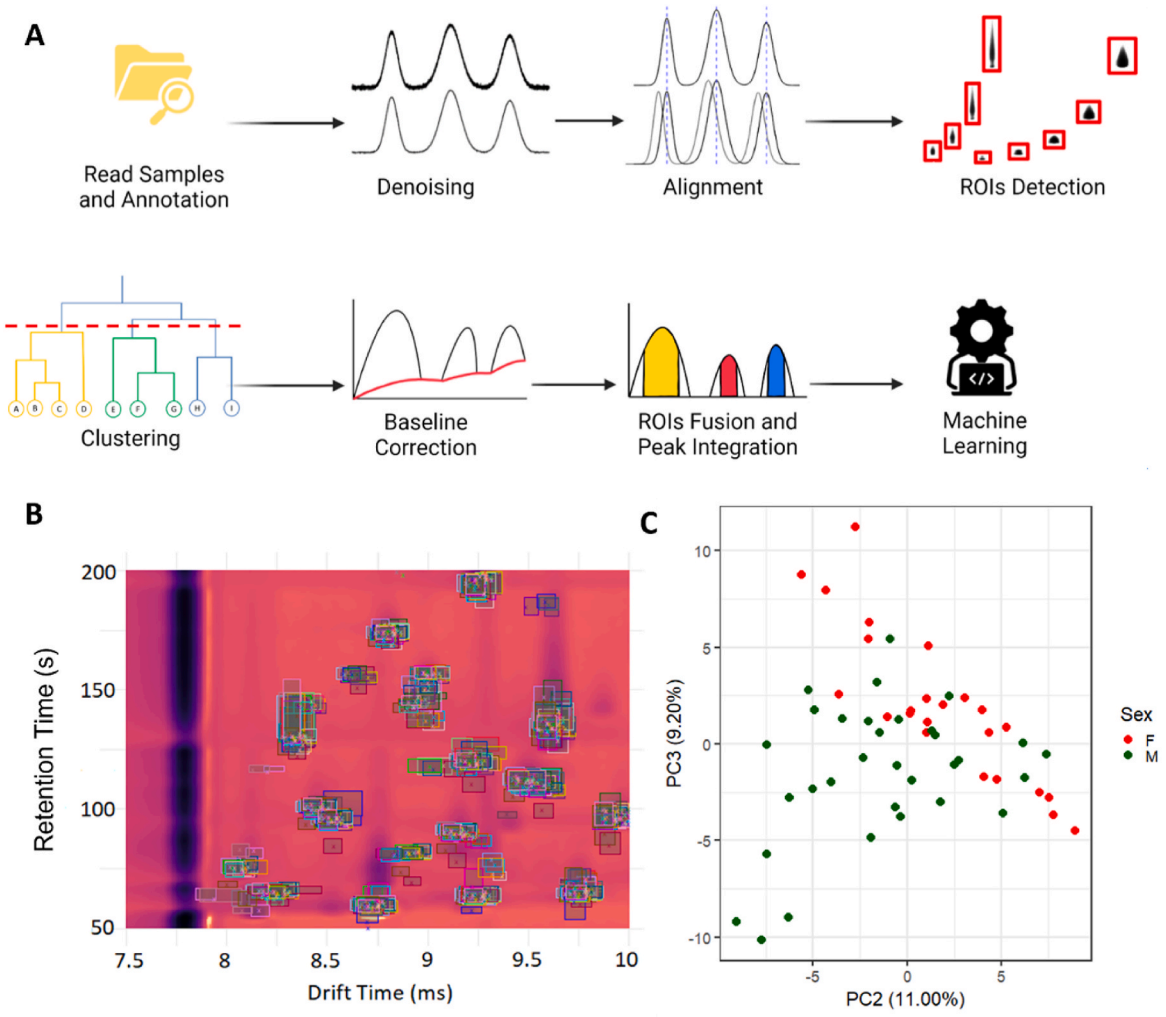


Fig. 1. A) Main steps of the GCIMS R package workflow; B) Image of the ROIs detected for all the samples, where each sample is represented by a different color; C) Score plot of the second and third Principal Components of the processed urine data. Red and green markers correspond to female and male individuals, respectively. The First Principal Component is mostly aligned with a batch effect in the dataset and it is discarded (see Fig. 3). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

GC-IMS raw data points as features for further analysis. This possibility was already explored by the authors, but it provides extremely high dimensionality that includes a large fraction of irrelevant characteristics, and it is very sensitive to misalignment errors. According to our previous experience, there is a need for open-source GC-IMS data processing tools that provide a more powerful approach with selective and robust extraction of chemical information.

In this work, we present the GCIMS R package, which is publicly available on GitHub (https://github.com/sipss/GCIMS), a general-purpose, customizable, open-source workflow for GC-IMS data treatment. The workflow includes a series of signal processing steps to correct nonidealities in data (denoising, chromatographic and spectral alignment and baseline correction), peak detection and matching, segmentation of regions of interest (peak boundaries), and finally peak integration.

2. Materials and methods

The proposed workflow for GC-IMS data analysis is a seven-step sequential process (Fig. 1A). In step 1, data are uploaded and annotated. GCIMS R package accepts data in three different input formats: .mea, .csv, and.mat. After that, signal to noise ratio is improved by using optimized Savitzky-Golay filters [38] in step 2. This task is performed in both drift and retention time axes. An optional but often recommended decimation process to reduce data dimensionality can be done at this point [31]. Note that the decimation factors in the two different time axes are usually different since drift time axis tends to have more resolution than retention time axis. To amend the effect of instrumental shifts on peak location among samples, spectral and chromatographic aligment is implemented in step 3. The alignment in retention time is a piecewise linear correction according to a set of reference peaks (e.g. internal standards) [39]. Drift time axis aligment is achieved through a linear correction. This technique transforms the spectra so that the positions of all Reactant Ion Positive (RIP) peaks coincide with the position of RIP of the reference spectrum [40]. It is worth to mentioning that the previous signal pre-processing step is crucial for the subsequent clustering task that will take place in step 5. Regions in data susceptible to containing chemical information (Regions of Interest: ROIs) are detected, that is peaks and their boundaries, in step 4. Each of these ROIs is a rectangle enclosing a peak. Peak detection is based on the continuous wavelet transform [41]. Next, we match the ROIs across samples to ensure that they belong to the same chemical species in step 5 (Fig. 1B). This can be done by applying hierarchical clustering [42] to the drift time – retention time coordinates of ROI representatives. Alternatively, it can be done with k-medoids [43]. Beyond peak clustering, a specific ROI size consensus is determined whose dimensions are the median of all the sample ROI's. In this manner, the area support for all the peak integrations is the same for all samples. Be aware that ROI size is different for each peak to accommodate specific peak shapes. Step 6 removes baselines in drift and retention time axes. Baselines are estimated as the *lowess* curves [44] obtained from the local minima of spectra and chromatograms. The volume of ROIs is computed in step 7 for all samples. Volume estimation consists of the double Riemman sum of peak intensities within the consensus ROI centered at each original peak position. The outcome of this workflow is a peak volume/ROI table with as many rows as samples and as many columns as distinct ROIs have been determined. Finally, a missing data imputation step is carried out by doing data integration on the consensus ROI in case the peak is not detected for a particular sample. Peak table rows can be normalized to reduce the effects of sample dilution and instrumental drift on peak intensities. This is done by using Probalilistic Quotient Normalization (PQN) [45], and normalization with respect to the Reactant Ion Positive (RIP) peak height [46] methods. The package provides several support functions to visualize the effect the different signal pre-processing steps on data. More specifically, the user can select and visualize the image of sample, plot all the chromatograms/spectra corresponding to a given drift/retention time, and check the results of the peak picking and clustering processes. A final machine learning predictive model can be developed for an ulterior classification of new samples according to their chemical signature. For a better understanding of GCIMS package functionality, please refer to the package vignette (https://sipss.github.io/GCIMS/articles/introduction-to-gcims.html). Also, you can find a detailed help for package functions and methods at https://sipss.github.io/GCIMS/reference/index.html.

3. Results and discussion

Urine is a very promising biofluid for volatilome analysis, due to its abundant availability and easy, non-invasive sample collection [47,48]. In the urine, there are many metabolites that can be sex, age, and condition dependent [49–51]. The effect of sex on the volatile phase of the urine metabolome has been previously studied with Gas Chromatography – Mass Spectrometry (GC-MS) [52,53], but to the best of our knowledge, there are no sex influence studies with GC-IMS.

To illustrate the operation of the GCIMS software, we conducted a subject discrimination study based on sex, using urine samples analysed with GC-IMS. A total of 56 urine samples were collected from 29 subjects (13 females and 16 males) in two different measurement campaigns, where the age, the size, and the weight of the subjects, were balanced among the 2 groups. The study protocol was approved by the Ethics Committee of Hospital de Reus (study approval no. 074/2018). The urine samples were obtained from the subjects at Hospital Universitari Sant Joan de Reus, and after collection, they were stored at −80 °C and transported to Barcelona with dried ice. For the sample preparation, 300 μL of a stock solution containing hydrochloric acid (HCl) 5 M, sodium chloride (NaCl), and sodium azide (NaN3) were added to each urine sample. The HCl was used to reach pH = 2 [54], as this pH captures more volatile organic compounds (VOCs) [55]. The NaCl was added to favour the volatile extraction, and the NaN3 served as a bacteriostatic agent, preventing bacterial growth in the urine samples [56]. Subsequently, the samples were incubated for 15 minutes at 60 °C, just before the GC-IMS analysis. The GC-IMS measurements were performed using a FlavourSpec® instrument from G.A.S. Dortmund (Dortmund, Germany). The flow rate of the drift gas was 200 ml/min, and the flow rate of the carrier gas was 11 ml/min, both using Nitrogen 5.0. The GC and IMS temperature were set at 60 °C, and the total analysis time lasted 33 minutes. Each sample acquired from the GC-IMS equipment resulted in a numeric matrix containing all the drift time spectra on one axis and all the retention time chromatograms on the other axis. Data collection was randomized. The dataset used in this study is available at Zenodo (https://zenodo.org/record/7941230).

Fig. 2 shows a raw GC-IMS urine sample from the dataset. The image exhibits the intrinsic complexity of GC-IMS data, where chemical information of volatiles is encoded in the form of two-dimensional peaks. Please observe the presence of the Reactant Ion Positive (RIP) peak for a drift time around 7.76 ms. This peak is responsible for transferring change to the rest of ions in a spectrum. Therefore, a reduction/increment in RIP peak height entails and increment/reduction of the height of peaks associated to volatile compounds. Interestingly, ion stability depends on volatile concentration in GC-IMS instruments [57]. At low concentrations, the most stable ion is the monomer. The height of the monomer peak increases with volatile concentration until a second ion generated from the same compound is the most stable (the dimer). Then, the dimer peak appears in the spectrum at a higher drift time value and the monomer peak suddenly vanishes. On some occasions, trimer ions can be also generated [58]. If volatile concentration is reduced, dimer peak disappears because the dominant ion is the monomer. Consequently, each time a volatile compound elutes from the chromatographic column the presence of their corresponding monomer and dimer peaks can be seen through adjacent spectra. From the figure, it is also evident that raw GC-IMS data are affected by baseline problems in both chromatographic and drift time axes. We can identify long peak tails in

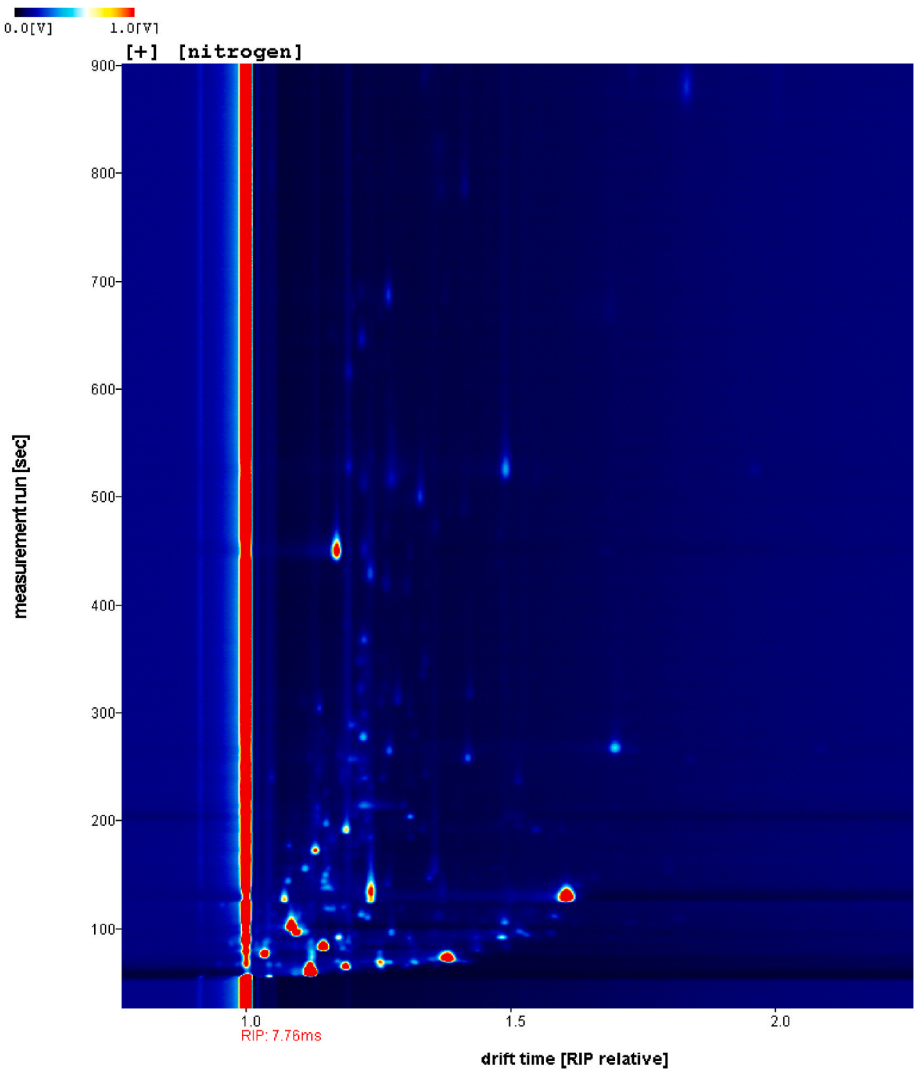


Fig. 2. Image of the raw GC-IMS data for a urine sample X-axis is the normalized drift time, and the Y-axis is the chromatographic retention time. The Reactant Ion Positive (RIP) appears as a strong intense (red) band parallel to the Y-axis. Visual inspection reveals the presence of numerous ion peaks. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

the retention time axis, but also strong variations in baselines of spectra, specially right after a new compound is eluting from the column. Finally, the position of peaks associated to volatiles varied considerably across samples (not shown).

After applying our pre-processing workflow on the dataset, raw data was denoised, aligned, and the baselines were corrected. Peak picking, peak clustering and peak integration stages were employed to extract data features. Therefore, a peak table summarizing all chemically relevant information from data was obtained. The total process for the 65 samples takes about 60 minutes to complete. We conducted an exploratory Principal Component Analysis [59] in order to identify trends in the corrected data. Fig. 1C shows the scores of the second and third principal components of a PCA analysis, coloured by sex. A certain tendency to separate the classes can be observed. Fig. 3 shows a remarkable batch effect in PC1 but negligible for higher Principal Components. Next, we conducted a classification task to assess the quality of the information extracted by the workflow. A leave one subject out (LOSO) double cross-validation process [60] was performed to train PLS-DA models [61] from urine data, that were able to reject the batch effect on this dataset. To optimize model complexity the area under the Receiver Operating Characteristic (ROC) curve was used in cross-validation [62]. The area under the curve (AUC) for the test samples was equal to 0.76 (CI 95% = 0.64–0.89). From these results, we concluded that with the GCIMS workflow proposed, differences between male and female urines can be detected. These conclusions were also validated with a permutation test which p-value was 0.005, fact that suggest that these results cannot be obtained by chance (Fig. 4). To compare the performance of our workflow with existing tools, we analysed the data with the gc-ims-tools for Python, and an AUC equal to 0.57 (CI 95% = 0.41–0.72) was obtained (Fig. 5). The AUC of both ROC curves were significantly different according to the DeLong's test (p-value = 0.03) [63]. The main differences between the two workflows were that gc-ims-tools Python package 1) did not correct chromatographic peak misalignments across samples, and 2) used the whole set of features in a matrix to characterize a sample [64] instead of a feature vector containing only the volumes of the detected peaks [65,66]. This showcases the importance of a proper signal processing/feature extraction workflow in the analysis of GC-IMS data.

In conclusion, the study showed promising results regarding the possibility of discriminating sex based on the volatile phase of the urine metabolome using GC-IMS, providing valuable insights for future research and potential diagnostic applications. For further information, the reader can consult the scripts used to perform this analysis at https://github.com/sipss/GCIMS_Case_Study.

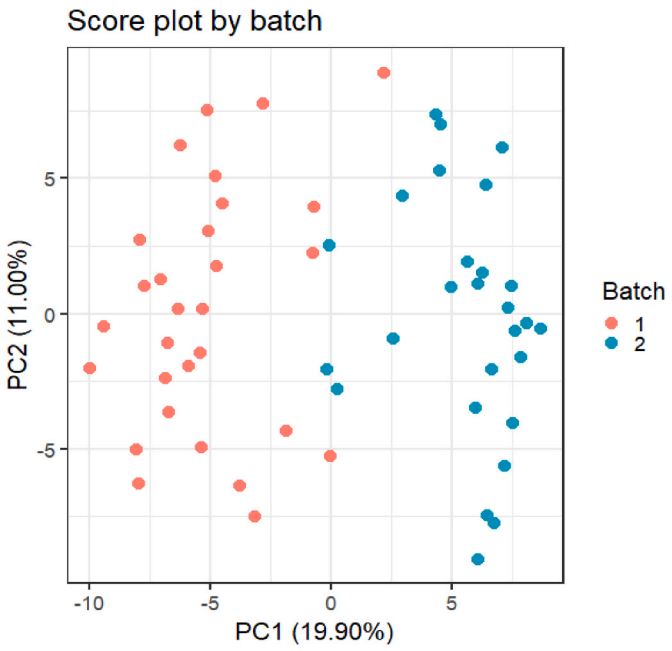


Fig. 3. Score plot of the first two Principal Components for the PCA model built from the complete set of samples after pre-processing and extracting data features. Batch effect is evident when colouring PC1 and PC2 scores according to the measurement campaign at which samples were acquired.

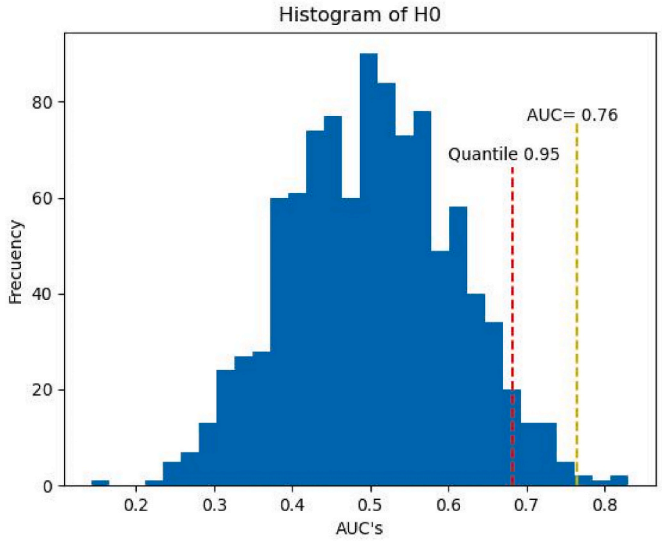


Fig. 4. Histogram of the Permutation test for the AUCs of PLS-DA models. The red line shows the value for quantile 0.95 (0.68) of the null hypothesis distribution, and the yellow one the AUC obtained after the LOSO double cross-validation process (0.76). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

4. Independently tested by Dr. Jia Yan

The GCIMS R Package was reviewed by Dr. Jia Yan, an outside reviewer from our institution. During the review process, Dr. Jia Yan used the GCIMS R package available in GitHub https://github.com/sipss/GCIMS with the dataset, also public available at https://zenodo.org/record/7941230.

After his review, Dr. Jia Yan declare.

Declaration and comments

With the provided code and guidance from authors' documents, I am able to implement the authors' program and obtain identical results. The code structure is clear, emphasizing code quality, and the program design is reasonable with a logical flow. This demonstrates that the authors are profound understanding of the problem and showcases their proficiency in applying solutions.

Dr. Jia Yan, Associate professor, Master supervisor, Director of Experimental Center.

College of Artificial Intelligence, Southwest University.

Address: No.2 Tiansheng Rd., Beibei District, Chongqing, 400715, P. R. China.

Telephone: +86 23 68250728 Cell phone: +86 15023308330.

Email: yanjia119@163.com yanjia119@swu.edu.cn.

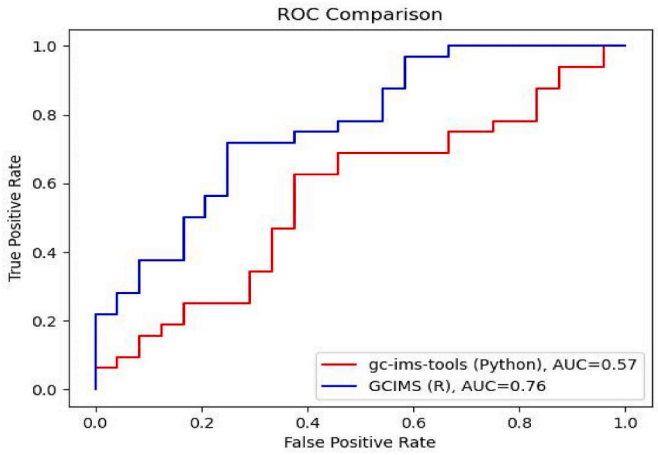


Fig. 5. Plot comparison of the ROC curves of the GCIMS R v0.1.0 package in blue (AUC 0.76 95%CI 0.64–0.89); and in red the ROC curve of the gc-ims-tools v0.1.2 for Python (AUC = 0.57 95%CI 0.41–0.72). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

5. Conclusion

GCIMS R package provides a fully automated, open-source workflow for GC-IMS data processing. This workflow is aimed at enhancing chemical information from the raw data. We have applied the package to a set of 29 subjects urine samples acquired with a GC-IMS instrument and corresponding two different measurement campaigns. The resulting ROI table was used to perform a LOSO double cross-validation process (AUC = 0.76, CI 95% = 0.64–0.89, p-value of the permutation test equal to 0.005). This result suggests that the proposed workflow was able to capture trends in data responsible for sample sex separation, that are hidden in simpler data processing approaches.

Availability

Source code is freely available at https://github.com/sipss/GCIMS under the GPL license. Dataset used in the presented case study is available at https://zenodo.org/record/7941230.

Funding

This research was supported by the Spanish Ministerio de asuntos económicos y transformacion digital (MINECO) project TensorChrom (TENSOMICS RTI2018-098577-B-C22) and TargetML (PID2021-126543OB-C21). Additional financial support was provided by the Institut de Bioenginyeria de Catalunya (IBEC). IBEC is member of the CERCAProgramme/Generalitat de Catalunya.

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.

Data availability

The data used in this study is publicly avaialble at Zenodo (https://zenodo.org/record/7941230).

Acknowledgements

We would like to acknowledge the Departament d'Universitats, Recerca i Societat de la Informació de la Generalitat de Catalunya (expedient 2021 SGR 01393).

We would like to acknowledge Dr. Jia Yan from College of Artificial Intelligence, Southwest University in China, for helping in the package review and testing.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.chemolab.2023.104938.

References

[1] A.E. Lytou, E.Z. Panagou, G.J.E. Nychas, Volatilomics for food quality and authentication, Curr. Opin. Food Sci. 28 (2019) 88–95, https://doi.org/10.1016/j.cofs.2019.10.003.

[2] A. Amann, B.D.L. Costello, W. Miekisch, J. Schubert, B. Buszewski, J. Pleil, N. Ratcliffe, T. Risby, The human volatilome: volatile organic compounds (VOCs) in exhaled breath, skin emanations, urine, feces and saliva, J. Breath Res. 8 (2014), https://doi.org/10.1088/1752-7155/8/3/034001.

[3] N. Drabińska, C. Flynn, N. Ratcliffe, I. Belluomo, A. Myridakis, O. Gould, M. Fois, A. Smart, T. Devine, B.D.L. Costello, A literature survey of all volatiles from healthy human breath and bodily fluids: the human volatilome, J. Breath Res. 15 (2021), https://doi.org/10.1088/1752-7163/abf1d0.

[4] J. Beauchamp, C. Davis, J. Pleil, Breathborne Biomarkers and the Human Volatilome, 2020.

[5] T.D.C. Minh, D.R. Blake, P.R. Galassetti, The clinical potential of exhaled breath analysis for diabetes mellitus, Diabetes Res. Clin. Pract. 97 (2012) 195–205, https://doi.org/10.1016/j.diabres.2012.02.006.

[6] M. Shirasu, K. Touhara, The scent of disease: volatile organic compounds of the human body related to disease and disorder, J. Biochem. 150 (2011) 257–266, https://doi.org/10.1093/jb/mvr090.

[7] Z. Chen, M.D. Phan, L.J. Bates, K.M. Peters, C. Mukerjee, K.H. Moore, M. A. Schembri, The urinary microbiome in patients with refractory urge incontinence and recurrent urinary tract infection, Int Urogynecol J 29 (2018) 1775–1782, https://doi.org/10.1007/s00192-018-3679-2.

[8] J. Pinto, F. Amaro, A.R. Lima, C. Carvalho-Maia, C. Jerónimo, R. Henrique, M.D. L. Bastos, M. Carvalho, P. Guedes De Pinho, Urinary volatilomics unveils a candidate biomarker panel for noninvasive detection of clear cell renal cell carcinoma, J Proteome Res 20 (2021) 3068–3077, https://doi.org/10.1021/acs.jproteome.0c00936.

[9] C.V. Berenguer, F. Pereira, J.A.M. Pereira, J.S. Câmara, Volatilomics: an emerging and promising avenue for the detection of potential prostate cancer biomarkers, Cancers 14 (2022), https://doi.org/10.3390/cancers14163982.

[10] F. Amaro, J. Pinto, S. Rocha, A.M. Araújo, V. Miranda-Gonçalves, C. Jerónimo, R. Henrique, M. de L. Bastos, M. Carvalho, P.G. de Pinho, Volatilomics reveals potential biomarkers for identification of renal cell carcinoma: an in vitro approach, Metabolites 10 (2020), https://doi.org/10.3390/metabo10050174.

[11] B. De Lacy Costello, A. Amann, H. Al-Kateb, C. Flynn, W. Filipiak, T. Khalid, D. Osborne, N.M. Ratcliffe, A review of the volatiles from the healthy human body, J. Breath Res. 8 (2014), https://doi.org/10.1088/1752-7155/8/1/014001.

[12] A. Tangerman, Highly sensitive gas chromatographic analysis of ethanol in whole blood, serum, urine, and fecal supernatants by the direct injection method, Clin. Chem. 43 (1997) 1003–1006.

[13] A.W. Jones, Excretion of low-molecular weight volatile substances in human breath: focus on endogenous ethanol, J. Anal. Toxicol. 9 (1985) 246–250, https://doi.org/10.1093/JAT/9.6.246.

[14] G.A. Eiceman, Z. Karpas, Ion Mobility Spectrometry, CRC-Press, 1994. https://books.google.es/books?id=XAXwAAAAMAAJ.

[15] J.I. Baumbach, Ion mobility spectrometry coupled with multi-capillary columns for metabolic profiling of human breath, J. Breath Res. 3 (2009), https://doi.org/10.1088/1752-7155/3/3/034001.

[16] V. Gabelica, E. Marklund, Fundamentals of ion mobility spectrometry, Curr. Opin. Chem. Biol. 42 (2018) 51–59, https://doi.org/10.1016/j.cbpa.2017.10.022.

[17] H. Borsdorf, G.A. Eiceman, Ion mobility spectrometry: principles and applications, Appl. Spectrosc. Rev. 41 (2006) 323–375, https://doi.org/10.1080/05704920600663469.

[18] C. Drees, W. Vautz, S. Liedtke, C. Rosin, K. Althoff, M. Lippmann, S. Zimmermann, T.J. Legler, D. Yildiz, T. Perl, N. Kunze-Szikszay, GC-IMS headspace analyses allow early recognition of bacterial growth and rapid pathogen differentiation in standard blood cultures, Appl. Microbiol. Biotechnol. 103 (2019) 9091–9101, https://doi.org/10.1007/s00253-019-10181-x.

[19] J.M. Lewis, R.S. Savage, N.J. Beeching, M.B.J. Beadsworth, N. Feasey, J. A. Covington, Identifying volatile metabolite signatures for the diagnosis of bacterial respiratory tract infection using electronic nose technology: a pilot study, PLoS One 12 (2017) 1–10, https://doi.org/10.1371/journal.pone.0188879.

[20] D.M. Ruszkiewicz, D. Sanders, R. O'Brien, F. Hempel, M.J. Reed, A.C. Riepe, K. Bailie, E. Brodrick, K. Darnley, R. Ellerkmann, O. Mueller, A. Skarysz, M. Truss, T. Wortelmann, S. Yordanov, C.L.P. Thomas, B. Schaaf, M. Eddleston, Diagnosis of COVID-19 by analysis of breath with gas chromatography-ion mobility spectrometry - a feasibility study, EClinicalMedicine (2020) 29–30, https://doi.org/10.1016/j.eclinm.2020.100609.

[21] M. Allers, J. Langejuergen, A. Gaida, O. Holz, S. Schuchardt, J.M. Hohlfeld, S. Zimmermann, Measurement of exhaled volatile organic compounds from patients with chronic obstructive pulmonary disease (COPD) using closed gas loop GC-IMS and GC-APCI-MS, J. Breath Res. 10 (2016), https://doi.org/10.1088/1752-7155/10/2/026004.

[22] R. Gasparri, R. Capuano, A. Guaglio, V. Caminiti, F. Canini, A. Catini, G. Sedda, R. Paolesse, C. Di Natale, L. Spaggiari, Volatolomic urinary profile analysis for diagnosis of the early stage of lung cancer, J. Breath Res. 16 (2022), https://doi.org/10.1088/1752-7163/ac88ec.

[23] N. Gerhardt, M. Birkenmeier, D. Sanders, S. Rohn, P. Weller, Resolution-optimized headspace gas chromatography-ion mobility spectrometry (HS-GC-IMS) for non-targeted olive oil profiling, Anal. Bioanal. Chem. 409 (2017) 3933–3942, https://doi.org/10.1007/s00216-017-0338-2.

[24] N. Arroyo-Manzanares, A. Martín-Gómez, N. Jurado-Campos, R. Garrido-Delgado, C. Arce, L. Arce, Target vs spectral fingerprint data analysis of Iberian ham samples for avoiding labelling fraud using headspace – gas chromatography–ion mobility spectrometry, Food Chem. 246 (2018) 65–73, https://doi.org/10.1016/j.foodchem.2017.11.008.

[25] N. Arroyo-Manzanares, M. García-Nicolás, A. Castell, N. Campillo, P. Viñas, I. López-García, M. Hernández-Córdoba, Untargeted Headspace Gas Chromatography-Ion Mobility Spectrometry Analysis for Detection of Adulterated Honey, (n.d).

[26] R. Garrido-Delgado, L. Arce, A.V. Guamán, A. Pardo, S. Marco, M. Valcárcel, Direct coupling of a gas-liquid separator to an ion mobility spectrometer for the classification of different white wines using chemometrics tools, Talanta 84 (2011) 471–479, https://doi.org/10.1016/j.talanta.2011.01.044.

[27] B. Bödeker, W. Vautz, J.I. Baumbach, Peak finding and referencing in MCC/IMS-data, Int. J. Ion Mobil. Spectrom. 11 (2008) 83–87, https://doi.org/10.1007/s12127-008-0012-7.

[28] S. Bader, W. Urfer, J.I. Baumbach, Preprocessing of ion mobility spectra by lognormal detailing and wavelet transform, Int. J. Ion Mobil. Spectrom. 11 (2008) 43–49, https://doi.org/10.1007/s12127-008-0005-6.

[29] M. Mäkinen, M. Sillanpää, A.K. Viitanen, A. Knap, J.M. Mäkelä, J. Puton, The effect of humidity on sensitivity of amine detection in ion mobility spectrometry, Talanta 84 (2011) 116–121, https://doi.org/10.1016/j.talanta.2010.12.030.

[30] A.B. Kanu, H.H. Hill, Ion mobility spectrometry detection for gas chromatography, J. Chromatogr. A 1177 (2008) 12–27, https://doi.org/10.1016/j.chroma.2007.10.110.

[31] R. Freire, L. Fernandez, C. Mallafré-Muro, A. Martín-Gómez, F. Madrid-Gambin, L. Oliveira, A. Pardo, L. Arce, S. Marco, Full workflows for the analysis of gas chromatography—ion mobility spectrometry in foodomics: application to the analysis of iberian ham aroma, Sensors 21 (2021), https://doi.org/10.3390/s21186156.

[32] E. Szymańska, A.N. Davies, L.M.C. Buydens, Chemometrics for ion mobility spectrometry data: recent advances and future prospects, Analyst 141 (2016) 5689–5708, https://doi.org/10.1039/c6an01008c.

[33] Z. Karpas, Y.F. Wang, G.A. Eiceman, Qualitative and quantitative response characteristics of a capillary gas chromatograph/ion mobility spectrometer to halogenated compounds, Anal. Chim. Acta 282 (1993) 19–31, https://doi.org/10.1016/0003-2670(93)80348-O.

[34] E. Szymańska, G.H. Tinnevelt, E. Brodrick, M. Williams, A.N. Davies, H.J. van Manen, L.M.C. Buydens, Increasing conclusiveness of clinical breath analysis by improved baseline correction of multi capillary column – ion mobility spectrometry (MCC-IMS) data, J. Pharm. Biomed. Anal. 127 (2016) 170–175, https://doi.org/10.1016/j.jpba.2016.01.054.

[35] A.C. Hauschild, T. Schneider, J. Pauling, K. Rupp, M. Jang, J.I. Baumbach, J. Baumbach, Computational methods for metabolomic data analysis of ion mobility spectrometry data-Reviewing the state of the art, Metabolites 2 (2012) 733–755, https://doi.org/10.3390/metabo2040733.

[36] GAS Dortmund (n.d.), https://www.gas-dortmund.de/Products/Software/VOCal—Software-for-GC-IMS-Data/1_524.html. (Accessed 29 July 2023).

[37] J. Christmann, S. Rohn, P. Weller, gc-ims-tools – a new Python package for chemometric analysis of GC–IMS data, Food Chem. 394 (2022), 133476, https://doi.org/10.1016/j.foodchem.2022.133476.

[38] M.J.E. Savitzky, A.; golay, smoothing and differentiation, Anal. Chem. 36 (1964) 1627–1639, https://doi.org/10.1021/ac60214a047.

[39] T. Perl, B. Bödeker, M. Jünger, J. Nolte, W. Vautz, Alignment of retention time obtained from multicapillary column gas chromatography used for VOC analysis

with ion mobility spectrometry, Anal. Bioanal. Chem. 397 (2010) 2385–2394, https://doi.org/10.1007/s00216-010-3798-1.
[40] M. Tabrizchi, Temperature Corrections for Ion Mobility Spectrometry, 2001.
[41] P. Du, W.A. Kibbe, S.M. Lin, Improved peak detection in mass spectrum by incorporating continuous wavelet transform-based pattern matching, Bioinformatics 22 (2006) 2059–2065, https://doi.org/10.1093/bioinformatics/btl355.
[42] S.C. Johnson, Hierarchical clustering schemes, Psychometrika 32 (1967) 241–254, https://doi.org/10.1007/BF02289588.
[43] H.S. Park, C.H. Jun, A simple and fast algorithm for K-medoids clustering, Expert Syst. Appl. 36 (2009) 3336–3341, https://doi.org/10.1016/j.eswa.2008.01.039.
[44] W.S. Cleveland, S.J. Devlin, Locally Weighted Regression: an Approach to Regression Analysis by Local Fitting, 1988.
[45] F. Dieterle, A. Ross, G. Schlotterbeck, H. Senn, Probabilistic quotient normalization as robust method to account for dilution of complex biological mixtures. Application in1H NMR metabonomics, Anal. Chem. 78 (2006) 4281–4290, https://doi.org/10.1021/ac051632c.
[46] M. del M. Contreras, N. Jurado-Campos, L. Arce, N. Arroyo-Manzanares, A robustness study of calibration models for olive oil classification: targeted and non-targeted fingerprint approaches based on GC-IMS, Food Chem. 288 (2019) 315–324, https://doi.org/10.1016/j.foodchem.2019.02.104.
[47] C. Mallafré-muro, M. Llambrich, R. Cumeras, A. Pardo, J. Brezmes, S. Marco, J. Gumà, Comprehensive volatilome and metabolome signatures of colorectal cancer in urine: a systematic review and meta-analysis, Cancers 13 (2021), https://doi.org/10.3390/cancers13112534.
[48] B. de Lacy Costello, O. Gould, N.M. Ratcliffe, Biomarkers in urine and stool, in: Breathborne Biomarkers and the Human Volatilome, Elsevier, 2020, pp. 379–408, https://doi.org/10.1016/B978-0-12-819967-1.00024-4.
[49] C.M. Slupsky, H. Steed, T.H. Wells, K. Dabbs, A. Schepansky, V. Capstick, W. Faught, M.B. Sawyer, Urine metabolite analysis offers potential early diagnosis of ovarian and breast cancers, Clin. Cancer Res. 16 (2010) 5835–5841, https://doi.org/10.1158/1078-0432.CCR-10-1434.
[50] E.J. Saude, B.D. Sykes, Urine stability for metabolomic studies: effects of preparation and storage, Metabolomics 3 (2007) 19–27, https://doi.org/10.1007/s11306-006-0042-2.
[51] S. Siegert, Sex dependency of human metabolic profiles revisited, J. Postgenom.: Drug Biomark. Develop. 2 (2012), https://doi.org/10.4172/2153-0769.1000115.
[52] S. Fan, A. Yeon, M. Shahid, J.T. Anger, K.S. Eilber, O. Fiehn, J. Kim, Sex-associated differences in baseline urinary metabolites of healthy adults, Sci. Rep. 8 (2018) 1–11, https://doi.org/10.1038/s41598-018-29592-3.
[53] M. Caterino, M. Ruoppolo, G.R.D. Villani, E. Marchese, M. Costanzo, G. Sotgiu, S. Dore, F. Franconi, I. Campesi, Influence of sex on urinary organic acids: a cross-sectional study in children, Int. J. Mol. Sci. 21 (2020) 1–17, https://doi.org/10.3390/ijms21020582.
[54] S. Smith, H. Burden, R. Persad, K. Whittington, B. De Lacy Costello, N.M. Ratcliffe, C.S. Probert, A comparative study of the analysis of human urine headspace using gas chromatography-mass spectrometry, J. Breath Res. 2 (2008), https://doi.org/10.1088/1752-7155/2/3/037022.
[55] P. Porto-Figueira, J. Pereira, W. Miekisch, J.S. Câmara, Exploring the potential of NTME/GC-MS, in the establishment of urinary volatomic profiles. Lung cancer patients as case study, Sci. Rep. 8 (2018) 1–11, https://doi.org/10.1038/s41598-018-31380-y.
[56] W. Wu, D. Yang, H.G. Tiselius, L. Ou, Z. Mai, K. Chen, H. Zhu, S. Xu, Z. Zhao, G. Zeng, Collection and storage of urine specimens for measurement of urolithiasis risk factors, Urology 85 (2015) 299–303, https://doi.org/10.1016/j.urology.2014.10.030.
[57] P.C. Moura, V. Vassilenko, Gas Chromatography – ion Mobility Spectrometry as a tool for quick detection of hazardous volatile organic compounds in indoor and ambient air: a university campus case study, Eur. J. Mass Spectrom. 28 (2022) 113–126, https://doi.org/10.1177/14690667221130170.
[58] A.K. Viitanen, T. Mattila, J.M. Mäkelä, M. Marjamäki, O. Anttalainen, J. Keskinen, Experimental study of the effect of temperature on ion cluster formation using ion mobility spectrometry, Atmos. Res. 90 (2008) 115–124, https://doi.org/10.1016/j.atmosres.2007.12.003.
[59] M. Ringnér, What is principal component analysis?. http://www.nature.com/naturebiotechnology, 2008.
[60] P. Filzmoser, B. Liebmann, K. Varmuza, in: J. Chemom (Ed.), Repeated Double Cross Validation, John Wiley and Sons Ltd, 2009, pp. 160–171, https://doi.org/10.1002/cem.1225.
[61] L.C. Lee, C.Y. Liong, A.A. Jemain, Partial least squares-discriminant analysis (PLS-DA) for classification of high-dimensional (HD) data: a review of contemporary practice strategies and knowledge gaps, Analyst 143 (2018) 3526–3539, https://doi.org/10.1039/c8an00599k.
[62] J.N. Mandrekar, Receiver operating characteristic curve in diagnostic test assessment, J. Thorac. Oncol. 5 (2010) 1315–1316, https://doi.org/10.1097/JTO.0b013e3181ec173d.
[63] E.R. DeLong, D.M. DeLong, D.L. Clarke-Pearson, Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach, Biometrics 44 (1988) 837–845.
[64] G. Sammarco, D. Bardin, F. Quaini, C. Dall'Asta, J. Christmann, P. Weller, M. Suman, A geographical origin assessment of Italian hazelnuts: gas chromatography-ion mobility spectrometry coupled with multivariate statistical analysis and data fusion approach, Food Res. Int. 171 (2023), https://doi.org/10.1016/j.foodres.2023.113085.
[65] M. García-Nicolás, N. Arroyo-Manzanares, L. Arce, M. Hernández-Córdoba, P. Viñas, Headspace gas chromatography coupled to mass spectrometry and ion mobility spectrometry: classification of virgin olive oils as a study case, Foods 9 (2020), https://doi.org/10.3390/foods9091288.
[66] M.P. Segura-Borrego, A. Martín-Gómez, R. Ríos-Reina, M.J. Cardador, M. L. Morales, L. Arce, R.M. Callejón, A non-destructive sampling method for food authentication using gas chromatography coupled to mass spectrometry or ion mobility spectrometry, Food Chem. 373 (2022), https://doi.org/10.1016/j.foodchem.2021.131540.