Microchemical Journal 199 (2024) 109801

Contents lists available at ScienceDirect

Microchemical Journal

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

DrogoFinder: An open-source graphical user interface for the identification of illicit drugs using infrared spectroscopy

Jaume Béjar-Grimalt, Francesc A. Esteve-Turrillas, Sergio Armenta, Salvador Garrigues, David Pérez Guaita *

Departamento de Química Analítica, Universidad de Valencia, Valencia, Spain

ARTICLE INFO

Keywords:
Infrared spectroscopy
Chemometrics
New psychoactive substances
Seized drugs
Graphical User Interface
Open-Source Software

ABSTRACT

The development of Free and Open-Source Software (FOSS) is being increasingly recognized as an indispensable resource for the development, validation, and replication of analytical methodologies and tools. In this context, the use of Attenuated Total Reflectance Infrared (ATR-IR) spectroscopy for the in-situ identification of drugs requires software for the treatment of spectroscopic data, including spectral quality controls, the subtraction of possible excipients, and the comparison with well-defined libraries that can be complex and limits its application on the field analysis by non-expert users. Here we describe a user-friendly FOSS for drug identification that can be operated in an intuitive way, including automatic features such as i) quality controls that warn the user from possible technical problems during spectrum acquisition (e.g. water vapour, contaminated background), ii) the subtraction of the contribution of excipients and other substances present on the sample, and iii) the creation of custom and updated cloud-based libraries to deal with the great number of new psychoactive substances (NPS) that are constantly being developed and introduced for their consume. The proposed application was tested by expert and non-expert users with spectra from 42 drug samples measured in situ in a music festival and 39 seized samples measured in a laboratory. Non-expert users were able to flag technical problems in the spectra and to perform excipient subtractions automatically, obtaining drug identifications that matched those ones obtained by reference methods and dedicated commercial software. We share both, the code, and a standalone Windows app to reach worldwide users and scientists interested in the identification of drugs.

1. Introduction

The evolution of science towards an open-source mode is becoming evident by leaps and bounds. It is widely accepted that the scientific advancement depends on the discovery, access and reformulation or improvement of studies by other researchers. In this sense, publishers of scientific journals, as well as national governments and scientists, have clearly opted for the open dissemination of scientific results in open-access journals. The motivation of this "open-source" movement in analytical chemistry has been previously stated in a review article [1], which is, basically, make available to any scientist the tools to replicate a project together with the permission for users to use, improve, and share it with others. It generally implies the adoption of open hardware and software principles in analytical chemistry research.

Free and Open-Source Software (FOSS) has been around for decades, being its ubiquity the most obvious success. As it was stated in the aforementioned review paper [1], interest in open-source software in analytical chemistry has increased significantly since the turn of the century. As a result, the number of publications of open-source applications in spectroscopy [2–5], chromatography [6–8], and electrochemical analysis [9,10] has increased at a proportional rate. Here we propose the creation of "DrogoFinder" an open-source application for the identification of illicit drugs using their ATR-FTIR spectra. The app focuses on the identification of common drugs and new psychoactive substances (NPS), featuring a user-friendly interface; so non-expert users, such as health workers or security forces can easily process the data in order to obtain reliable results in a fast and simple way. This include the quality control of the spectra, the subtraction of potential interferents present in the sample, [11] and the comparison with custom and cloud-based libraries.

The identification of unknown illicit drugs in seized samples is one of the most active fields of research in forensics. In this context, IR

* Corresponding author.
E-mail address: david.perez-guaita@uv.es (D. Pérez Guaita).

https://doi.org/10.1016/j.microc.2023.109801
Received 28 June 2023; Received in revised form 6 December 2023; Accepted 11 December 2023
Available online 27 December 2023

spectroscopy is one of the most widely used techniques. Methodologies based in ATR-FTIR spectroscopy are rapid, sensitive, easy to handle, cost-effective and non-destructive. Because of that, ATR-FTIR have been widely applied in illicit drug analysis [12,13] for the identification of classical drugs like cocaine [14–16], MDMA [17], and also NPS [18–20]. The identification is done by the comparison of the spectrum of the unknown sample with spectra stored in spectral libraries, allowing its identification whenever the compound of interest is included in the library. However, this apparently simple operation can become very complicated in the case of samples that are complex mixtures. Bands from other components present in the sample such as other active components, fillers/adulterants or excipients, interfere with the identification, making it difficult to obtain a sufficiently reliable result.

In this sense, ATR-FTIR methodologies require an updated database containing spectra of all known psychoactive substances (as well as spectra of potential adulterants and excipients) and specialized software that allow the manipulation of spectra to increase the accuracy of the identifications (e.g., subtraction of the spectra of other psychoactive compounds or cellulose). Furthermore, in general, the resolution of these problems requires well-trained users who can detect potential experimental pitfalls that can hamper the identification and provide misleading results. Common examples of unsuitable spectra include the presence of bands associated with the rotational states of the water vapor, negative bands product of a contamination in the background scan, and low absorbances due to an ineffective contact of sample with the ATR crystal during measurement step.

Commercial companies provide dedicated software packages to control the IR instruments and to perform such identifications. These packages usually integrate extensive libraries with spectra from different substances, and generally increase the cost of the basic version of the software. However, depending on the supplier, each instrument uses different programs, with specific and protected procedures for data analysis (i.e. black boxes). This fact can make it even more difficult to compare results among different devices and software. Furthermore, these applications are generally not specific for drug identification and contain a large set of instructions and options, making them not very intuitive for non-expert user. Thus, there is a need for graphical user interfaces (GUI) that enable the implementation of complex routines and spectral processing pipelines on easy-to-use apps [5] dedicated to specific applications such as drug identifications.

The developed FOSS presented in this paper includes the quality control of the spectra, the subtraction of potential interferents present in the sample, and the comparison with custom libraries. The app also includes an advanced mode, which adds the possibility of performing advanced identification routines as well as the incorporation of new spectra into the library to be used in future identifications. Furthermore, the proposed software has also the potential to be implemented in a cloud-based environment, providing updated databases to keep pace with the continuous emergence of NPS. We tested the features of the app with expert and non-expert users to validate its use in both, laboratory and field scenarios. To that end 1 set of samples obtained at a music festival as well as 2 sets of seized samples were evaluated. Moreover, the code and a user manual are free shared as an open-source resource for researchers and institutions.

2. Methods

2.1. Software development

The proposed application was developed using MATLAB R2022a from MathWorks (Natick, MA, USA). According to preliminary tests, it is be compatible with MATLAB versions R2021b, R2022a and R2022b. The GUI was created using the "app designer" extension from MATLAB, and uses in-house written scripts for importing, pre-processing and evaluation of the spectra. Scripts contain functions from the *Bioinformatics*, *Image processing, Signal processing,* and *Statistics and Machine Learning* Toolboxes, as well as other functions available at the MathWorks file exchange such as the PerkinElmer IR data file import tool [21] and passwordUI [22]. The developed app can be used and modified from MATLAB as long as these toolboxes are available. Furthermore, a stand-alone version of the app was created to be used in the Windows environment without the need for a MATLAB license. This was performed by using the MATLAB compiler toolbox. The stand-alone version contains almost all the functions available in the MATLAB version, apart from saving data in a custom library. A user manual and installation guide are included in the software files, which can be downloaded at https://zenodo.org/records/10245755.

Within the included options, the app provides the ability to automatically subtract common adulterants typically found in the samples, such as cellulose or caffeine. This subtraction is performed by generating a vector 'C' ranging from 0.1 to 1, which simulates relative concentration of the adulterant in the spectrum.

$$C = [0.1\,0.2 \cdots 1] \tag{1}$$

Such vector is multiplied by the spectrum of the excipient (e), obtaining a matrix (Ce) of the excipient spectra at different levels of concentrations.

$$Ce = C^{*}e \tag{2}$$

Then, the generated spectra in Ce are subtracted to the sample spectrum, to obtain a new matrix of subtracted spectra (Ci).

$$Ci = i - Ce \tag{3}$$

Finally, a search is performed to select the spectrum with the minimum value in the matrix of spectra above 0. This search is performed over a predetermined range of wavenumbers.

$$S(x) = min_i((i - Ci) > 0) \tag{4}$$

Being S(x) the minimum positive absorbance value. The spectrum that contains this value is considered as the best subtracted spectrum.

The automatic subtraction was evaluated by using samples adulterated with different compounds, like cellulose or caffeine. The predetermined range of wavenumbers is optimized for different substances used as excipients or adulterants like cellulose, caffeine and lactose. Alternatively, a selected substance from the compound list (i.e. in the case of samples containing mixtures) can also be subtracted, enabling the determination of whether the sample contains more than one substance. For common excipients, this specific region is already registered within the app, while for other compounds of the library wavenumber interval for correction, taking into consideration the maximum absorbance band, is automatically selected. The range was optimized by performing subtractions with different seized samples and selecting those that provide better subtractions results, which included the regions with the higher absorbance bands for each compound (see Table 1).

The app also performs three different quality controls every time a new spectrum is loaded. The first one checks if there is a contribution from the presence of water vapour bands on the spectrum, which is common if the measure takes place in an environment where humidity is not well regulated (i.e., outside a laboratory), and reduces the quality of the spectrum. The second one checks for possible negative bands in the spectrum, which can be the product of cross-contamination of the crystal

Table 1
Wavenumber intervals used for the subtraction of the desired substance.

Compound	Wavenumbers region (cm^{-1})
Cellulose	970–1067
Lactose	1006–1070
Caffeine	1605–1675
Selected compound from the library	± 40 from the maximum band of the compound

during the background. The last quality check assesses whether the absorbance is enough for a satisfactory analysis since a spectrum with a low signal is more susceptible to noise, also hampering the identification of the drug. To that end, the software follows the criteria in the Table 2 to perform the calculations. For the absorbance and negative bands, the range between 800 and 3400 cm^{-1} was selected, and for the water vapour the region selected included one rotational band from the water vapour spectrum.

The quality control procedures were validated by measuring spectra under conditions that simulated typical experimental problems which could potentially skew the identification, including i) interferences caused by water vapour bands, ii) negative bands caused by an ineffective cleaning of the crystal when the background was measured, and iii) low absorbance caused by a weak contact of sample with the ATR crystal. The cut-off values were optimized by testing the quality controls with the spectra of different samples.

2.2. Spectral acquisition

The IR spectra were obtained using a Spectrum Two FT-IR spectrometer from Perkin-Elmer (Waltham, MA, USA) equipped with the UATR accessory, coadding 10 scans at 4 cm^{-1} spectral resolution over the range 4000–500 cm^{-1}. The atmospheric contribution was corrected using the build-in CO_2/H_2O correction provided by the software from Perkin-Elmer, Spectrum (version 10.03.06). A new background scan was acquired between each sample, which was obtained after cleaning the ATR crystal surface with ethanol. Components in the acquired spectra were identified by both, the proposed app and Spectrum. In both cases, the same strategy for the identification of illicit drug was employed, including the subtraction of probable excipients or other active compounds and the computation of "score" values based on the correlation between the sample spectrum and each spectrum of a library. A total of 3 different data sets of samples were used. Set 1 was used to evaluate the methodology and contained spectra of 17 samples of seized drugs that were identified by reference techniques such as gas chromatography mass-spectrometry (GC–MS) and ion mobility spectrometry (IMS) in the analytical laboratory. Set 2 was used to validate the software with samples measured in-situ and contained spectra of 42 sample drugs measured during a music festival that took place at Spain in July 2022. Set 3 was used to test the method's selectivity and the impact of fillers and excipients on drug identification. It consisted of 22 adulterated cocaine samples characterized by liquid chromatography diode array detection (LC-DAD) in the analytical laboratory. In our study we used the library from SWGDRUGS (Version 2.1, 624 NPS spectra, August 27th, 2019) in combination with spectra of standards and characterized samples measured by our research team at the University of Valencia (93 drug spectra plus 7 adulterants).

An additional dataset containing 15 spectra were used for the blind testing of the app by non-expert users. The set included spectra of different drugs from Set 1 and 2, as well as spectra of complex samples or with problems, like the presence of water vapour bands, negative bands, or the presence of strong contributions of excipients that should be subtracted. A total of 4 non-expert users which included a final year chemistry student, and 3 non-qualified users with technological degrees not related to chemistry, were employed for validating the app. Users were asked to identify the given spectra blindly after reading the app instruction manual (SM1) in order to prove if the designed GUI was easy to use, even if the users had no experience with drug analysis and/or ATR spectra treatment.

Table 2
Software criteria for the quality controls.

Quality Control	Criteria	Range (cm^{-1})	Cut-off value (u.a.)
Water vapour	Difference of intensity between the maximum and minimum of a rotational water band	3909 to 3917	< 0.001
Absorbance	Maximum value for the loaded spectrum	800 to 3400	> 0.01
Negative bands	Minimum value for the loaded spectrum	800 to 3400	< -0.01

3. Results

3.1. Software interface and options

When launching the app, the GUI shows five main panels (see Fig. 1): a) Spectrum panel, where the spectra can be loaded, and saved to the library, b) Subtraction panel, where the subtraction takes place by choosing the adulterating substance, c) Parameters panel, for modifying the parameters of the determinations, d) Quality control panel, for displaying the quality controls outcome, and e) Results panel, in which the processing is described and the results are displayed. The user can import the spectrum through the "Load" button on the Spectrum panel using .sp, .DX, .dx and .spc file extensions. The wavenumbers of the imported spectrum are then automatically aligned using the function *Interp1* from MATLAB to create a spectrum within the 3980–750 cm^{-1} range with a uniform spacing of 1 cm^{-1}. The same process is also applied to all the spectra of the databases to provide compatible comparisons.

The flowchart in Fig. 2 summarizes the different steps and options performed by the app. Once the file is loaded, the quality controls are automatically applied to the spectrum, checking if water vapour bands, negative signals or low absorbance values, are detected. Results are displayed as green or red lights in the quality controls panel If a quality control check fails, a message in the results panel suggests steps to address the issue. For example, if the Negative band QC fails, the software recommends acquiring a new background. Additionally, if a filler or excipient is detected in the sample within the top 5 identification results, the application performs the subtraction automatically, and shows a message in the results panel specifying the type of adulterant and providing the subtraction result. Moreover, the user can subtract the desired substance from the spectra selecting one of the common adulterating substances included in the dropdown menu of the Subtraction panel, which includes cellulose, caffeine, and lactose. Furthermore, the user can subtract any compound from the Results panel table by checking the "subtract" checkbox. After the subtraction, the identification of the resulting spectrum is automatically performed, and the new matches appears in the Result panel. After that, the user can generate a report by pressing the button "Do report" in the Results panel, which can be downloaded as a pdf file. An example of pdf report can be found in the electronic supplementary material (Report_example.pdf). A further explanation on how the app works is described in the user manual.

3.2. Data processing and parameters

After loading the spectrum and conducting quality controls and excipients subtractions, the spectrum is compared to the spectra in the library. The application also includes an advanced mode which can be used by checking the "Advance" box located in the bottom left corner of the GUI. It has been blocked by a password that can be set by the user. Within the advanced mode, the "Add to library" button and the "Parameters" panel will be unlocked, giving the user the ability to include a loaded spectrum in a new library called "Own library" (In the current version, this feature is only available when using the MATLAB version of the app). Additionally, the spectrum is plotted in the Results panel for visual inspection, and can be normalized by pressing the "Normalize" button on the Spectrum panel. Also, the spectrum with the higher match is displayed alongside the sample spectrum, and the user can manage which spectra are shown by selecting or deselecting the "display" checkboxes at the results panel. In this mode, the software also incorporates tuning parameters for the identification, such as the

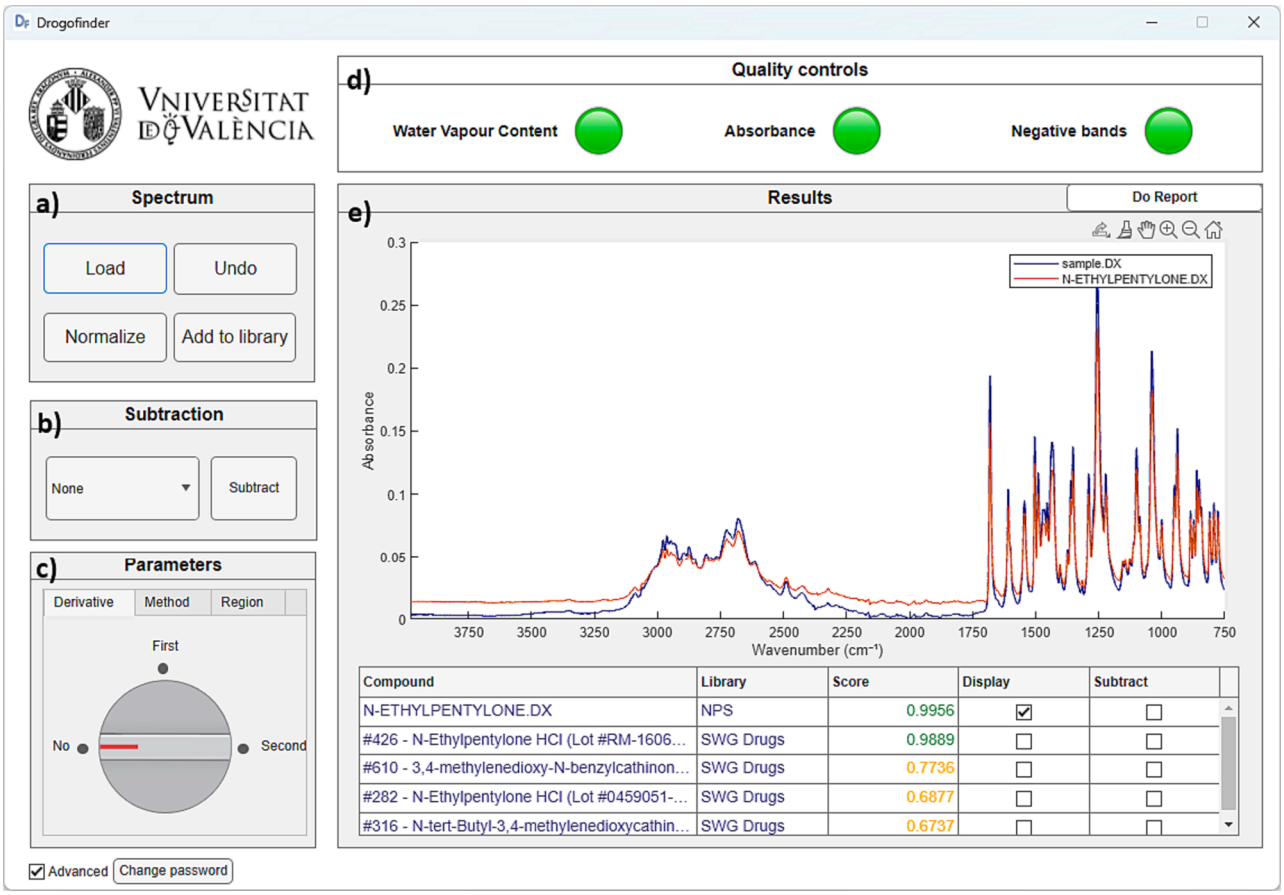

Fig. 1. DrogoFinder GUI after loading a spectrum, showing 5 main panels: a) Spectrum panel, where the spectra can be loaded, normalized, and saved to the library, b) Subtraction panel, where the subtraction takes place by choosing the adulterating substance, c) Parameters panel, for modifying the parameters of the determinations, d) Quality control panel, for displaying the quality controls outcome, and e) Results panel, the results are displayed.

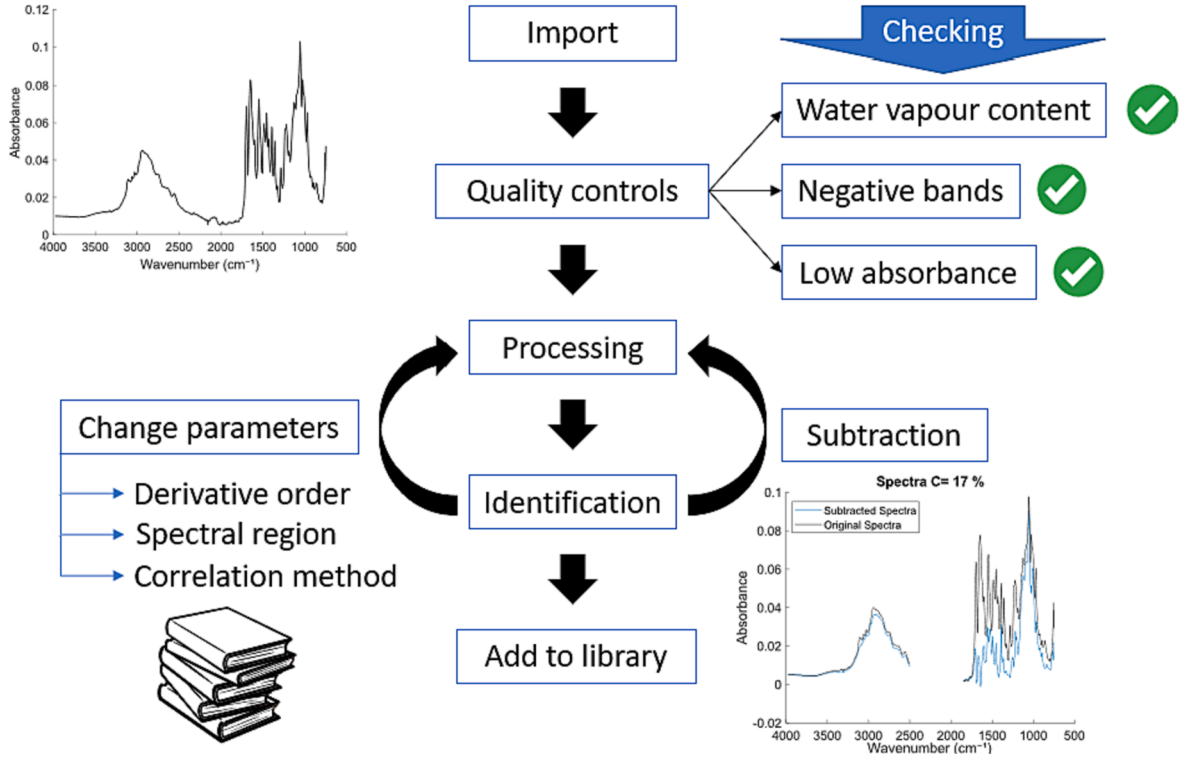

Fig. 2. Flowchart for the identification of the drug using the proposed GUI.

utilization of derivatives, the selection of specific spectral regions, and the choice of identification methods. The methods available for selection include the correlation coefficient (R^2) and the k-nearest neighbours (Knn). After modifying any of these parameters, an updated identification will take place automatically, and the resulting spectrum is displayed. Derivatives are performed using a Savitzky–Golay filter (15 points window, 2nd degree polynomial order). This is designed to provide experienced users with more options to explore the spectra and expand the library. Inexperienced users, on the other hand, can focus on the identification of the sample without the need to be concerned about additional parameters.

3.3. Evaluation of the quality controls

To check if the quality controls of the app were working as intended, a 'stress test' was performed by analysing spectra obtained under different circumstances imitating unfavourable experimental acquisition conditions. This included a spectrum with low absorbance, a spectrum obtained with a contaminated background, and a spectrum with contribution of water vapour. For the first spectrum, a minimal amount of sample was deposited onto the ATR crystal to simulate conditions where the spectrum exhibits very low absorbance (Fig. 3**a**). The second spectrum was obtained using a background registered with a contaminated ATR crystal, providing negative bands (Fig. 3**b**). In the third spectrum, the water vapour correction of the software was turned off and some water vapour was induced through the beam path, emulating high water vapour changes in the environment (Fig. 3**c**). The 3 spectra obtained under these circumstances were compared to 58 spectra from a database obtained by expert operators with appropriate control of experimental issues (i.e., without major water vapour

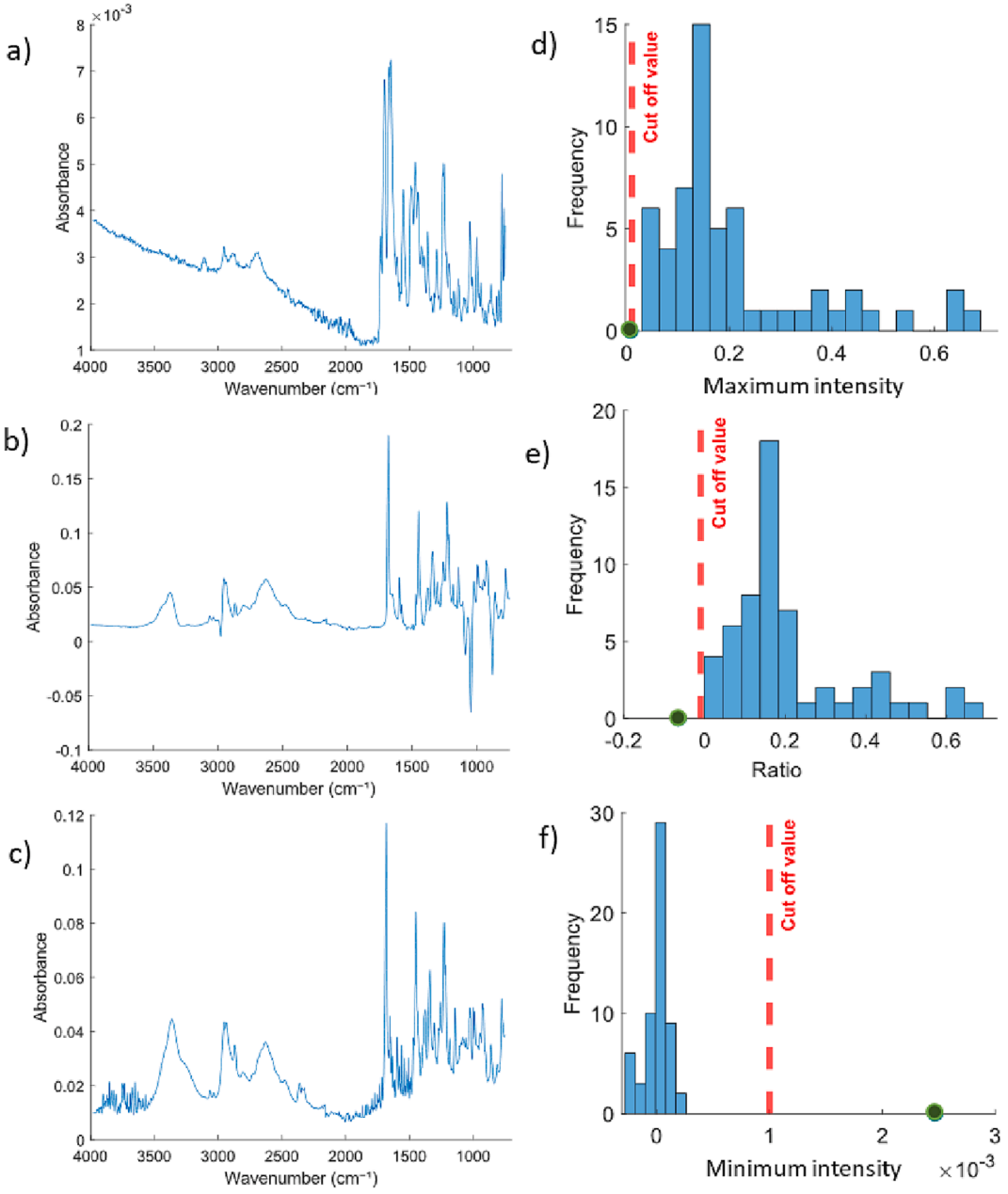

Fig. 3. Evaluation of the quality controls. Spectra measured in the "stress test" of the low absorbance (a), negative bands (b) and water vapour quality controls (c). Comparison of the ratios calculated for the spectra collected by expert operators under favourable conditions (blue histogram) and "stress test" samples (green dot), for the low absorbance, negative bands and water vapour quality controls, for the low absorbance (d), negative bands (e) and water vapour quality controls (f). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

contributions, with enough absorbance and with an optimal obtained background). The calculated values of the parameters of the quality controls (see Table 1) obtained for each spectrum are depicted in Fig. 3**d-f,** which compare a histogram of the results of 58 well-collected spectra with the spectrum obtained under unfavourable conditions and the cut off value selected in the quality control. As seen in the different histograms, all the 58 spectra measured under favourable conditions passed the quality controls, while the ones obtained under the stress test, represented as a green dot in the histograms, failed. Thus, the QC represented an effective and automatic way to flag possible experimental problems to users without spectroscopic expertise.

3.4. Subtraction

To test if the subtraction was working as intended, a sample adulterated with caffeine was assessed. The spectrum of the sample can be seen in the Fig. 4**a** along with the spectrum of the caffeine. The app provided a R^2 value of 0.74 for caffeine as the first option. The subtraction of the caffeine automatically removed the most intense caffeine bands, as seen in the subtracted spectrum in Fig. 4**b**. The algorithm computed an optimal C parameter of 78, indicating that the caffeine standard spectrum was multiplied by 0.78 before subtraction. This value is related to the actual proportion of cocaine and caffeine in the sample and can serve as a semiquantitative indicator of the concentration level of the excipient and illicit substance. Subsequent identification using the subtracted spectrum yielded an R^2 value of 0.96 for cocaine, which had the highest R^2 among all the compounds. This improvement could also be appreciated in the comparison of the subtracted spectrum with the cocaine spectrum (See Fig. 4**c)**. This fact proved that the sample spectra were being processed as intended, since the subtraction increased the R^2 to almost 1 after the subtraction.

3.5. Software comparison and validation by non-expert users

Results obtained after identifying the 42 drug spectra using the software Spectrum and the designed GUI DrogoFinder are shown in Table 3. As it can be seen, the identification provided by the DrogoFinder app matched the one obtained by the Spectrum software for almost all the samples. A discrepancy can be found which is the identification of 2C-B instead of amphetamine for Sample #24 and in the identification of 3MMC instead of MDMA for sample #20 (see Table 3). These mismatches can be caused by small differences in the procedures employed by both software, but in general, the results obtained from the app are comparable to the ones obtained from a commercial software.

For the blind test performed by the non-expert users, all the participants were asked to identify the substances from a set of spectra without any additional information about the samples (see Table 4). The users were previously trained with the DrogoFinder User Manual (available in **SM1**) and had to provide a table with the results, which was compared to the results obtained by expert users with the Spectrum software. The four participants were able to flag the two spectra that did not pass the quality controls (due to the presence of negative bands or water vapour). Furthermore, all the spectra were correctly identified by the users, which were able to automatically subtract the caffeine contribution in one of them. These results show the capability of the application to help with the identification of the substances, even when the users have no previous knowledge about drugs or spectroscopy analysis.

3.6. Validation of the methodology with well-known seized samples

To validate the methodology, the spectra obtained from the seized samples (Set 2) were identified using ATR-FTIR with DrogoFinder and Spectrum software, and the results were compared with the identifications obtained with reference methods (GC–MS and IMS). All seized samples were presented in the form of powders with different colours (white, pink, blue, or grey). The sample set contained a total of 17 of

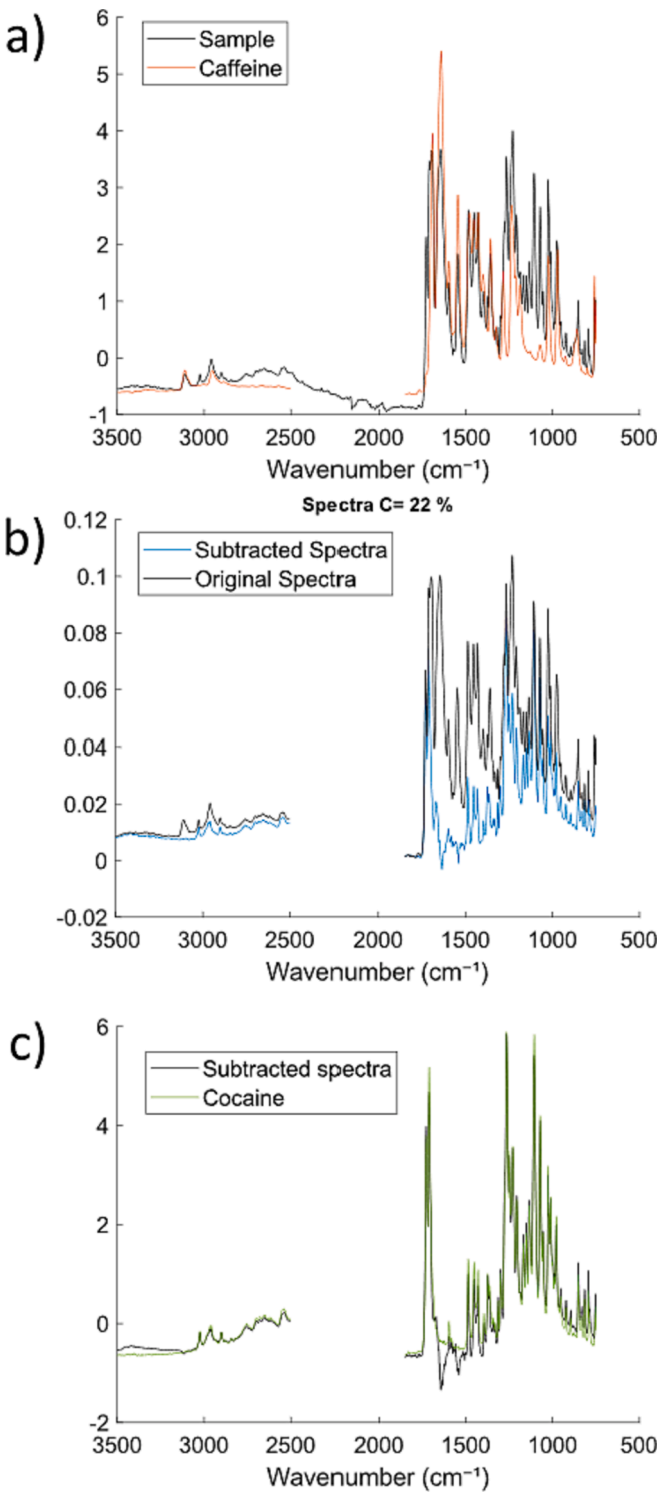

Fig. 4. Evaluation of the subtraction. First the original spectrum is plotted along the spectrum of the caffeine after being normalized(a), then the original spectrum is represented with the subtracted one (b) and in the last place the processed spectrum is normalized and represented alongside the normalized spectra of the cocaine which gave the highest score (R^2) (c).

common drugs (1 sample of cocaine and 4 samples of ketamine) as well as design drugs such as ADB-PINACA and 11 samples of 4-chloromethcathinone (4 CMC). As seen in Table 5, the DrogoFinder app was able to find matches of the ATR-FTIR spectra obtained within the SWG library (R^2 values between 0.51 and 0.98). Most importantly, the results obtained are in well agreement with the main drug found in the samples

Table 3
Comparison of substance identification provided by Spectrum software (by Perkin Elmer) and the DrogoFinder app for spectra of different seized drug samples.

Sample Number	Compound identified	
	Spectrum Software	DrogoFinder
1	MDMA	MDMA
2	MDMA	MDMA
3	Amphetamine	Amphetamine
4	MDMA	MDMA
5	MDMA	MDMA
6	MDMA	MDMA
7	MDMA	MDMA
8	MDMA	MDMA
9	MDMA	MDMA
10	MDMA	MDMA
11	Ketamine	Ketamine
12	MDMA	MDMA
13	MDMA	MDMA
14	Ketamine	Ketamine
15	Amphetamine	Amphetamine
16	Lidocaine	Lidocaine
17	Lidocaine	Lidocaine
18	MDMA	MDMA
19	MDMA	MDMA
20	MDMA	3MMC
21	MDMA	MDMA
22	Cocaine	Cocaine
23	Cocaine	Cocaine
24	Amphetamine	2C-B
25	MDMA	MDMA
26	Amphetamine	Amphetamine
27	Lidocaine	Lidocaine
28	MDMA	MDMA
29	MDMA	MDMA
30	MDMA	MDMA
31	MDMA	MDMA
32	2C-C	2C-C
33	Amphetamine	Amphetamine
34	MDMA	MDMA
35	MDMA	MDMA
36	MDMA	MDMA
37	MDMA	MDMA
38	Cocaine	Cocaine
39	Cocaine	Cocaine
40	2C-B	2C-B
41	MDMA	MDMA
42	MDMA	MDMA

Note: MDMA, 3,4-methylenedioxymethamphetamine; 2C-B, 4-Bromo-2,5-dimethoxyphenethylamine.

Table 4
Seized drug samples used for the blind test by non-expert users.

Sample	Main Component	Observations
M1	*N*-ethyl-pentylone	
M2	alpha-PHP	Water vapour bands
M3	MDMA	Negative bands
M4	Cocaine	
M5	Ketamine	
M6	Amphetamine	Caffeine before subtraction
M7	MDMA	
M8	Cocaine	
M9	Lidocaine	
M10	Amphetamine	
M11	Cocaine	
M12	MDMA	
M13	Amphetamine	
M14	2C-B	
M15	MDMA	

Note: alpha-PHP, alpha-pyrrolidinohexiophenone; 2C-B, 4-bromo-2,5-dimethoxyphenethylamine; MDMA, 3,4-methylenedioxymethamphetamine.

Table 5
Comparison between the identifications obtained using ATR-FTIR (for both DrogoFinder and the one obtained from commercial software and reference methods.

ATR-FTIR-DrogoFinder software	ATR-FTIR-Spectrum software	GC–MS	IMS
4-CMC (N = 11) Ketamine (N = 4) ADB-PINACA Cocaine	4-CMC Ketamine ADB-PINACA Cocaine	4-CMC Ketamine ADB-PINACA Cocaine	4-CMC Ketamine ADB-PINACA Cocaine

Note: ADB-PINACA, *N*-(1-amino-3,3-dimethyl-1-oxobutan-2-yl)-1-pentyl-1H-indazole-3-carboxamide; 4-CMC, 4-cloromethcathinone; N, number of samples.

by GC–MS and IMS. Similar results were obtained when the ATR-FTIR data was treated with the commercial Spectrum software.

Finally, set 3, which included 22 seized samples containing different concentrations of cocaine as well as other 6 excipients, was used to evaluate the prediction performance of the app in complex samples with low concentrations of drugs. Sample components were identified using their ATR-FTIR spectra and the proposed software, and the outcomes were compared with the results obtained by LC-DAD [23] (See Table 6). Samples in Set 3 exhibited a cocaine content ranging from 0.50 % to 17.17 % by weight, according to the reference method. These samples were adulterated with one to five substances, including levamisole, caffeine, phenacetin, tetracaine, procaine, and lidocaine. Following the initial identification (Table 6, hit 1), an automatic subtraction of the substance that gave the higher hit was applied obtaining a second result (Table 6, hit 2). A third subtraction was conducted aiming to identify another substance (Table 6, hit 3). A fourth subtraction was not feasible due to the significantly reduced spectral quality. Table 6 shows that out of the 22 samples, cocaine was detected in 19, but not in samples 1, 5, and 6. Sample 1, containing only 0.99 % cocaine, likely falls below the method's selectivity threshold. In this case, the only identified substance was procaine, comprising about 76.15 % of the sample. In sample 5, despite having a high cocaine content of 11.72 %, cocaine was not detected, highlighting cases where the app struggled to resolve mixtures. Sample 6 presented a unique challenge as it contained five adulterants, with four of them in higher concentrations than cocaine, rendering the app unable to identify cocaine.

Regarding the limit of detection of the method, the app successfully identified cocaine with a minimum concentration of 4.24 % by weight, even in the presence of two or more adulterants. However, it is important to note that this result may vary depending of the complexity of the samples. Binary ternary samples could range better sensitivity if the interferents can be easily subtracted, while multi-component samples would have a higher limit of detection. It is important to note that excipients not included in the database cannot be accounted for, and as a result, the limit of detection for samples containing these interferences will be significantly high. This highlights again the need for a comprehensive database of spectra including a wide range of drugs and excipients.

4. Conclusions and outlook

Drug identification though IR methodologies can be a complicated task, especially for inexperienced users with barely any knowledge of spectroscopy. A GUI like DrogoFinder offers a user-friendly tool with an easy way to identify these substances from their IR-ATR spectra without the need of a commercial software. To that end, the quality controls included in the developed app can assist non-experienced users to identify errors that can hamper the identification of the sample during the obtention of the spectra. It also provides an intuitive tool for the subtraction of an adulterating substance or additional compounds in the same sample in order to improve the identification of compounds in seized samples. Furthermore, the app can be easily updated to keep up

Table 6
Identification of the adulterated cocaine samples using Drogofinder with successive subtractions with the obtained hits from before and after the 2 subtractions.

Sample	%Coc	%Lev	%Caff	%Phe	%Tet	%Pro	%Lid	Hit 1	Hit 2	Hit 3
1	0.99	0	18.14	0	1.93	76.15	0	Pro 0.99		
2	10.3	0	4.72	20.08	0.29	0	0	Phe 0.66	**Coc 0.51**	
3	17.16	10.12	19.59	26.59	0	10.51	0	Phe 0.76	Caff 0.50	**Coc 0.43**
4	8.3	4.91	0	10.32	0	0	0	**Coc 0.81**	Phe 0.83	
5	11.72	5.45	0	86.96	0	0	0	Phe 0.96		
6	3.8	6.58	5.39	68.92	3.8	7.3	0	Phe 0.96	Pro 0.46	
7	15.55	9.14	8.08	10.51	6.12	23.6	0	**Coc 0.88**	Pro 0.94	
8	13.25	29.18	45.79	0	0	0	0	Caff 0.96	**Coc 0.70**	
9	11.22	0	26.29	56.07	0	0	0	Phe 0.77	Caff 0.71	**Coc 0.57**
10	14.95	49.6	0	0	0	9.04	0	**Coc 0.79**	Lev 0.75	
11	14.98	0	25.66	0	0	6.84	0	Caff 0.73	**Coc 0.91**	
12	4.88	98.99	0	0	0	0	0	Lev 0.68	**Coc 0.77**	
13	9.85	24.66	0	12.09	0	5.57	0	**Coc 0.76**	Phe 0.58	Lev 0.57
14	9.82	95.86	0	0	0.31	0	0	Lev 0.70	**Coc 0.79**	
15	14.01	37.34	0	11.42	0	32.85	0	**Coc 0.76**	Phe 0.84	
16	9.7	35.34	0	10.79	0	41.56	0	Phe 0.75	**Coc 0.9**	
17	9.57	0	29.13	0	0	0	72.06	Caff 0.86	**Coc 0.70**	Lid 0.64
18	10.53	7.73	0	17.43	0	0	0	Phe 0.79	**Coc 0.82**	Lev 0.36
19	9.21	14.03	30.93	0	16.78	0	0	Caff 0.85	Tet 0.80	**Coc 0.76**
20	6.9	9.13	25.28	0	17.54	0	0	Caff 0.78	Tet 0.88	**Coc 0.69**
21	4.24	6.33	24.78	0	14.88	0	0	Caff0.85	Tet 0.90	**Coc 0.4**
22	8.09	9.89	19.43	0	16.01	0	0	Caff 0.74	Tet 0.88	**Coc 0.72**

Note: Coc, Cocaine; Lev, Levamisole; Caff, Caffeine; Phe, Phenacetin; Tet, Tetracaine; Pro, Procaine; Lid, Lidocaine.

with the constant development of NPS. The app was successfully validated with a set of real samples by comparing their identification with identifications performed by commercial software and reference methodologies, being able to identify the presence of cocaine is complex samples with a lower limit of concentration around 5 %. A blind test was also conducted with inexperienced users, who were able to identify compounds of interest, subtract the spectra of excipients and detect spectra with poor quality. The developed software, including the source code, is free and open-access, resulting in a tool that can be used or improved by anyone in the field. Examples of future developments that can be added to the app include the addition of more complete drug spectra databases and the inclusion of more sophisticated search/quantification algorithms, increasing the potential value of the FOSS.

CRediT authorship contribution statement

Jaume Béjar-Grimalt: Data curation, Investigation, Methodology, Software, Writing – original draft. **Francesc A. Esteve-Turrillas:** Conceptualization, Investigation, Project administration, Writing – review & editing. **Sergio Armenta:** Conceptualization, Investigation, Project administration, Writing – review & editing. **Salvador Garrigues:** Conceptualization, Investigation, Project administration, Writing – review & editing. **David Pérez Guaita:** Conceptualization, Funding acquisition, Investigation, Methodology, Project administration, Software, Supervision, Writing – review & editing.

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

Data is shared in a link that can be found in the "Experimental Section"

Acknowledgments

D.P.-G. acknowledges the financial support from the projects RYC2019- 026556-I, PID2019-110788GB-I00 and RPID2020-119326RA-I00 funded by MCIN/AEI/ 10.13039/501100011033. JB acknowledges financial support from the INVESTIGO program (INVEST/2022/95) funded by the Next Generation EU recovery plan.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.microc.2023.109801.

References

[1] M.D.M. Dryden, R. Fobel, C. Fobel, A.R. Wheeler, Upon the shoulders of giants: open-source hardware and software in analytical chemistry, Anal. Chem. 89 (8) (2017) 4330–4338, https://doi.org/10.1021/acs.analchem.7b00485.
[2] I.A. Lewis, S.C. Schommer, J.L. Markley, rNMR: open source software for identifying and quantifying metabolites in NMR spectra: rNMR: metabolomics software, Magn. Reson. Chem. 47 (S1) (2009) S123–S126, https://doi.org/10.1002/mrc.2526.
[3] J. Hu, Y. Zou, B. Sun, X. Yu, Z. Shang, J. Huang, S. Jin, P. Liang, Raman spectrum classification based on transfer learning by a convolutional neural network: application to pesticide detection, Spectrochim. Acta. A. Mol. Biomol. Spectrosc. 265 (2022) 120366, https://doi.org/10.1016/j.saa.2021.120366.
[4] Z.-M. Zhang, S. Chen, Y.-Z. Liang, Z.-X. Liu, Q.-M. Zhang, L.-X. Ding, F. Ye, H. Zhou, An intelligent background-correction algorithm for highly fluorescent samples in raman spectroscopy: background-correction algorithm for highly fluorescent samples in Raman spectroscopy, J. Raman Spectrosc. 41 (6) (2010) 659–669, https://doi.org/10.1002/jrs.2500.
[5] D. Perez-Guaita, Z. Richardson, A. Rajendra, H.J. Byrne, B. Wood, From bench to worktop: rapid evaluation of nutritional parameters in liquid foodstuffs by IR spectroscopy, Food Chem. 365 (2021) 130442, https://doi.org/10.1016/j.foodchem.2021.130442.
[6] J.P. Grinias, J.T. Whitfield, E.D. Guetschow, R.T. Kennedy, An inexpensive, open-source USB arduino data acquisition device for chemical instrumentation, J. Chem. Educ. 93 (7) (2016) 1316–1319, https://doi.org/10.1021/acs.jchemed.6b00262.
[7] J. Capellades, M. Navarro, S. Samino, M. Garcia-Ramirez, C. Hernandez, R. Simo, M. Vinaixa, O. Yanes, geoRge: A computational tool to detect the presence of stable isotope labeling in LC/MS-based untargeted metabolomics, Anal. Chem. 88 (1) (2016) 621–628, https://doi.org/10.1021/acs.analchem.5b03628.
[8] X. Domingo-Almenara, J. Brezmes, M. Vinaixa, S. Samino, N. Ramirez, M. Ramon-Krauel, C. Lerin, M. Díaz, L. Ibáñez, X. Correig, A. Perera-Lluna, O. Yanes, eRah: A computational tool integrating spectral deconvolution and alignment with quantification and identification of metabolites in GC/MS-based metabolomics, Anal. Chem. 88 (19) (2016) 9821–9829, https://doi.org/10.1021/acs.analchem.6b02927.
[9] M. Son, N. Yoon, S. Park, A. Abbas, K.H. Cho, An open-source deep learning model for predicting effluent concentration in capacitive deionization, Sci. Total Environ. 856 (2023) 159158, https://doi.org/10.1016/j.scitotenv.2022.159158.
[10] S.D. Curtis, K.L. Ploense, M. Kurnik, G. Ortega, C. Parolo, T.E. Kippin, K.W. Plaxco, N. Arroyo-Currás, Open source software for the real-time control, processing, and visualization of high-volume electrochemical data, Anal. Chem. 91 (19) (2019) 12321–12328, https://doi.org/10.1021/acs.analchem.9b02553.

[11] C. Cole, L. Jones, J. McVeigh, A. Kicman, Q. Syed, M. Bellis, Adulterants in illicit drugs: A review of empirical evidence, Drug Test. Anal. 3 (2) (2011) 89–96, https://doi.org/10.1002/dta.220.
[12] Long, D. A. Handbook of Vibrational Spectroscopy, Volumes 1-5.Edited by J. M. Chalmers and P. R. Griffiths. John Wiley & Sons, Chichester, 2002, Pp. 3862. *J. Raman Spectrosc.* **2005**, *36* (3), 271–271. https://doi.org/10.1002/jrs.1309.
[13] C. Liu, Y. Han, S. Min, W. Jia, X. Meng, P. Liu, Rapid Qualitative and Quantitative Analysis of Methamphetamine, Ketamine, Heroin, and Cocaine by near-Infrared Spectroscopy, Forensic Sci. Int. 290 (2018) 162–168, https://doi.org/10.1016/j.forsciint.2018.07.008.
[14] S. Armenta, M. de la Guardia, M. Alcalà, M. Blanco, Noninvasive Double Confirmation of Cocaine Abuse, Anal. Chem. 85 (23) (2013) 11382–11390, https://doi.org/10.1021/ac4023583.
[15] D.K. John, K. dos Santos Souza, M.F. Ferrão, Overview of cocaine identification by vibrational spectroscopy and chemometrics, Forensic Sci. Int. 342 (2023) 111540, https://doi.org/10.1016/j.forsciint.2022.111540.
[16] C.A.F.O. Penido, M.T.T. Pacheco, R.A. Zângaro, L. Silveira, Identification of different forms of cocaine and substances used in adulteration using near-infrared raman spectroscopy and infrared absorption Spectroscopy, J. Forensic Sci. 60 (1) (2015) 171–178, https://doi.org/10.1111/1556-4029.12666.
[17] E. Deconinck, R. Van Campenhout, C. Aouadi, M. Canfyn, J.L. Bothy, L. Gremeaux, P. Blanckaert, P. Courselle, Combining attenuated total reflectance- infrared spectroscopy and chemometrics for the identification and the dosage estimation of MDMA tablets, Talanta 195 (2019) 142–151, https://doi.org/10.1016/j.talanta.2018.11.027.
[18] D. Gallart-Mateu, J. Béjar-Grimalt, F.A. Esteve-Turrillas, S. Armenta, S. Garrigues, M. de la Guardia, A Synergetic Approach Based on Infrared Spectroscopy and Ion Mobility Spectrometry for the Analysis of Seized Blotters: Boosting Performance, Microchem. J. 181 (2022) 107810, https://doi.org/10.1016/j.microc.2022.107810.
[19] M. Custódio, L. Magalhães, L. Arantes, J. Braga, Identification of synthetic drugs on seized blotter papers using ATR-FTIR and PLS-DA: routine application in a forensic laboratory, J. Braz. Chem Soc. (2021) https://doi.org/10.21577/0103-5053.20200205.
[20] J. Coelho Neto, Rapid Detection of NBOME's and Other NPS on Blotter Papers by Direct ATR-FTIR Spectrometry, Forensic Sci. Int. 252 (2015) 87–92, https://doi.org/10.1016/j.forsciint.2015.04.025.
[21] Perston, B. *PerkinElmer IR data file import tools*. https://www.mathworks.com/matlabcentral/fileexchange/22736-perkinelmer-ir-data-file-import-tools (accessed 2023-03-30).
[22] Doke, J. *passwordUI*. https://www.mathworks.com/matlabcentral/fileexchange/29616-passwordui (accessed 2023-03-30).
[23] A. Brancaccio, P. Maresca, S. Albrizio, M. Fattore, M. Cozzolino, S. Seccia, Development and validation of a diode array high performance liquid chromatography method to determine seized street cocaine sample purity, Anal. Methods 5 (10) (2013) 2584, https://doi.org/10.1039/c3ay40141c.