Volume 13
Number 36
28 September 2021
Pages 4011–4180

Analytical Methods

rsc.li/methods

ISSN 1759-9679

ROYAL SOCIETY OF CHEMISTRY

PAPER
Leo Lebanov and Brett Paull
Smartphone-based handheld Raman spectrometer and machine learning for essential oil quality evaluation

Analytical Methods

ROYAL SOCIETY OF CHEMISTRY

PAPER

View Article Online
View Journal | View Issue

Cite this: *Anal. Methods*, 2021, **13**, 4055

Smartphone-based handheld Raman spectrometer and machine learning for essential oil quality evaluation†

Leo Lebanov[ab] and Brett Paull *[ab]

We present a method, utilising a smartphone-based miniaturized Raman spectrometer and machine learning for the fast identification and discrimination of adulterated essential oils (EOs). Firstly, the approach was evaluated for discrimination of pure EOs from those adulterated with solvent, namely benzyl alcohol. In the case of ylang–ylang EO, three different types of adulteration were examined, adulteration with solvent, cheaper vegetable oil and a lower price EO. Random Forest and partial least square discrimination analysis (PLS-DA) showed excellent performance in discriminating pure from adulterated EOs, whilst the same time identifying the type of adulteration. Also, utilising partial least squares regression analysis (PLS) all adulterants, namely benzyl alcohol, vegetable oil and lower price EO, were quantified based on spectra recorded using the smartphone Raman spectrometer, with relative error of prediction (REP) being between 2.41–7.59%.

Received 26th May 2021
Accepted 12th August 2021

DOI: 10.1039/d1ay00886b

rsc.li/methods

1 Introduction

Worldwide demand for pure natural products has led to an increase in the production of essential oils (EOs) over many years. The estimated global EO market in 2020 was 247.08 kilo tonnes, with expected annual growth of 7.5% until 2027.[1] Application/addition of natural extracts has added a strong marketing advantage to many manufactured goods, such as perfumes, cosmetics, food and beverages, and pharmaceuticals. Prices for natural EOs are often much higher than those of synthetic materials, which leads to many cases of adulteration.[2] Adulteration can be performed by dilution with cheaper oils, such are vegetable, coconut or mineral oils, as well as synthetic materials, benzyl alcohol, ethanol, *etc.*[2] Cheap oils are common EO adulterants because they are low cost, easily available and their density is similar to that of many EOs. Also, EO adulteration can be performed by partial or total substitution by other inferior EOs, which have lower price and quality, and can obviously impact the overall physicochemical properties. This can affect the EO compliance to specifications set by corresponding national and international standards. In order to avoid easy detection of EO adulteration, typically adulterants are added at relatively low levels (5–8%).[3]

Most commonly applied techniques for the quality assurance of EOs involve gas chromatography hyphenated with mass spectrometry or flame ionisation detector (GC-MS and GC-FID).[4–6] With GC-MS and GC-FID analysis the chemical profile of the examined EO is obtained, which is further compared to a corresponding ISO standard. In this way, addition of solvents, or EOs out of specification, can be detected. However, GC-based techniques are expensive and lab-based, and demand the attention of highly trained analysts. Also, due to the complexity of most EOs, achieving good resolution among most EO components, typically requires analysis times of between 30 to 120 min. Further, detecting dilution with cheaper oils, such as vegetable, coconut, or castor oils, requires the sample to be derivatised prior to either GC-MS or GC-FID analysis, which further increases time of the analysis, and can introduce potential experimental and procedural errors. To help with this workflow, automated high-performance thin-layer chromatography (HPTLC) techniques have been increasingly applied within pharmacopoeias and by regulatory agencies.[2] HPTLC allows detection of adulteration with cheaper oils, without any derivatisation step. However, it does involve additional solvent consumption and of course adds further cost and complexity.

Over recent years, there has been an increase in the application of vibrational spectroscopy techniques in the quality control of commercial natural products.[7,8] UV-Vis absorption spectroscopy, followed by non-linear modelling, has been applied for the detection and quantification of virgin olive oil adulterants,[9] characterisation of virgin olive oil blends[10] and characterisation of edible oil blends.[11] Increased application of

[a]*Australian Centre for Research on Separation Science (ACROSS), School of Natural Sciences, University of Tasmania, Hobart, TAS, Australia. E-mail: brett.paull@utas.edu.au*

[b]*ARC Industrial Transformation Research Hub for Processing Advanced Lignocellulosics Products (PALs), School of Natural Sciences, University of Tasmania, Hobart, TAS, Australia*

† Electronic supplementary information (ESI) available. See DOI: 10.1039/d1ay00886b

infrared (IR) and Raman spectroscopy is credited to their inherent advantages of little or minimal sample preparation requirements, non-destructive nature, combined with minimal solvent consumption.[12] In comparison to near infrared (NIR) based devices, where spectra arise from combination and overtone bands, which are broader bands with lower intensity compared to fundamental bands in mid infrared region, Raman spectroscopy offers a much more distinctive fingerprint of the analysed samples, which allows sample identification through spectra comparison with readily available compound spectral libraries. Over the past ten years, there have been significant developments in portable and handheld Raman devices, which allow analysis to be performed directly in-field.[12,13] Miniaturised Raman spectrometers have been applied in quality control of pharmaceuticals products,[12] meat products,[14] food additives[15] and spirits.[16] Within EO quality control, handheld Raman devices have already been used to distinguish different classes of EOs,[17] as well as to detect adulteration of clove EOs with benzyl alcohol.[18] Many commercial Raman spectrometers offer databases which are used for sample identification and purity evaluation. However, current capabilities are ineffective when comparing similar spectra, and when differences are very subtle, specifically in the case of low-level adulterations. Therefore, the application of multivariate statistical techniques and machine learning algorithms offers significant potential in this space, for revealing these small differences in Raman spectra and identifying adulterated and classifying lower quality commercial products.[16,19–21]

Herein we evaluate the application of a new smartphone-based portable Raman spectrometer equipped with a 785 nm laser for the detection of adulteration in EOs at different levels. The obtained data is subjected to multivariate statistical analysis, namely partial least squares regression, and discriminatory analysis (PLS and PLS-DA), and application of a Random Forest (RF) machine learning algorithm. Several types of adulteration were examined, including the addition of benzyl alcohol (BA) and vegetable oil (VO), from 4.76–54.55%. For ylang–ylang (Yly) EO, discrimination of pure EOs from the oils adulterated with BA, VO and low grade EO was evaluated, seeking to simultaneously identify and quantify the adulterant. To the best of our knowledge, this is the first report on the application of machine learning for the purpose of EO quality control, adulteration detection, identification and quantification of the potential adulterant, based on spectra recorded with smartphone-based Raman spectrometer.

2 Experimental work

2.1 Chemicals and sample preparation

The following EOs, cinnamon (Cin, 4 samples), geranium bourbon (Geb, 5 samples), cypress (Cyp, 1 sample), rosemary (Ros, 5 samples), clove bud (Clb, 5 samples), frankincense serrata (Frs, 5 samples) and ylang–ylang (Yly, 10 samples), were all obtained from "Plant Therapy Inc." (Twin Falls, ID, USA). Vegetable oil was obtained from a local food store (Hobart, Tasmania). Benzyl alcohol (99.6%) was purchased from "Chem-Supply Pty" (Gillman, SA, Australia).

Table 1 Number of samples and essential oils used in creating two discriminatory models[a]

		Model #1				Model #2
Number of samples		Cin	Clb	Geb	Ros	YY
Training set	Pure EO	3	3	3	3	7
Training set	AD EO	6	6	6	6	18
Independent set	Pure EO	1	2	2	2	3
Independent set	AD EO	3	3	3	3	12

[a] EO – essential oil; AD – adulterated; Cin – cinnamon; Clb – clove bud; Geb – geranium bourbon; Ros – rosemary; Yly – ylang–ylang.

For the quantification of adulterants, adulterated EOs were created according to Table 2. For the discrimination of adulterated EOs and quantification of adulterants, adulterated EOs were created according to Tables 1 and 2. Two different discrimination models were created. Model #1 was developed to discriminate BA adulterated EOs from pure EOs. The model was created using four different EO classes, namely Cin, Clb, Geb and Ros, categorised as pure EOs, where one EO from each EO class was adulterated with 4.76, 9.09, 28.57, 37.5, 44.44 and 54.55% of BA, creating a training set with 36 EO samples, with 12 pure and 24 adulterated oils. Another EO from each class was adulterated with 5.0, 10.0 and 50.0% of BA to create independent data matrix with 7 pure and 12 adulterated EOs, for the evaluation of the models' predictive capacity. For model #2, three different Yly EOs, each with one adulterant, were adulterated with BA, VO and Cyp EO, to obtain EOs with 4.76, 9.09, 28.57, 37.5, 44.44 and 54.55% of adulterant. The independent set was prepared of three pure Yly EOs, where one sample was adulterated with BA, VO and Cyp EO to obtain EOs with 5.0, 10.0, 37.5 and 50.0% of adulterant. Here the training/validation set had in total 25 samples, while the independent set had 15

Table 2 Preparation of adulterated EOs for the training/validation set (VS) and for the independent set (IS) for the quantification of adulterants

		Pure oil (μL)	AD (μL)	% of AD
Training/validation set	VS1	500	25	4.76
	VS2	500	50	9.09
	VS3	500	75	13.04
	VS4	500	200	28.57
	VS5	500	300	37.50
	VS6	500	350	41.18
	VS7	500	450	47.37
	VS8	500	500	50.00
	VS9	500	600	54.55
Independent set	IS1	500	26	5.00
	IS2	500	55	10.00
	IS3	500	100	16.67
	IS4	500	250	33.33
	IS5	500	334	40.00
	IS6	500	400	44.44
	IS7	500	500	50.00
	IS8	500	550	52.38

samples, three pure and 12 adulterated Yly EOs, see Table 1. Training/validation and independent sets for the quantification of adulterants were prepared according to Table 2. For the quantification of BA in EOs from 4 different EO classes, one sample from each EO class was adulterated with BA according to Table 2. In the case of VO quantification, two Frs EOs were adulterated with VO according to Table 2. For the quantification of adulterants in Yly EOs, three different Yly EOs, each with different adulterant, namely BA, VO and Cyp EO, were adulterated according to Table 2.

2.2 Handheld Raman spectrometer procedure

Raman measurements of EOs were performed using a smartphone Raman spectrometer, specifically the Smart Raman XI™ (CloudMinds Inc., Wangjing SOHO Tower, Chaoyang District, Beijing, China), see Fig. S1.† The Raman spectra were recorded using the 785 nm laser and 500 mW laser power with acquisition time of 5–10 s, utilising the liquid vial holder. The spectral resolution was 2 cm^{-1} and the spectral range was from 1800 cm^{-1} to 200 cm^{-1}. Pure EOs were analysed in 2 mL glass vials. Three spectra were recorded for each EO, with auto baseline correction, and these spectra were used to obtain the average spectrum for each EO.

2.3 Data analysis

Two different RF and PLS-DA models were developed. Model #1 was developed for the discrimination of pure EOs, from those adulterated with BA. The experimental data matrix, used for model training and validation was sized 36 × 800, sample × variables, with rows containing samples' spectra and columns of absorbance at each wavelength. The independent sample set was sized 19 × 800. Model #2 was created to discriminate pure Yly EOs from adulterated versions, whilst at the same time identifying the adulterant. The training/validation sample set was sized at 25 × 800, while the independent sample set was sized at 15 × 800 matrix. Four data matrices for BA quantification using PLS analysis were prepared, where the experimental data matrix, used for model training and validation was sized at 9 × 800, while the independent data matrix was sized at 8 × 800. For the quantification of VO in Frs EOs training/validation data matrix was sized at 18 × 800, while independent data matrix was sized 16 × 800. In the case of Yly EO adulteration with BA, VO and Cyp EO, three training/validation data matrices of 9 × 800 were prepared, while the three independent data matrices were sized at 8 × 800.

In order to evaluate discrimination power based on different pre-processing techniques, RF and PLS-DA models were created using raw Raman spectra, as well as using pre-processed data, specifically multiplicative scatter correction (MSC) and MSC followed by Savitzky–Golay first derivative (SG1), with the width window set to 7 points. Data pre-processing was performed using the "*mdatools*" Package in R software (R.3.4.4).[22]

Discrimination between adulterated and pure EOs was performed using RF and PLS-DA algorithms. RF models were developed using the "*randomForest*" Package in R software (R.3.4.4).[23] All RF models were created under the same parameters. The optimal number of variables (*mtry*) and number of trees (*ntree*) were set for all models (*ntree* = 1000 with *mtry* = 50), while other parameters were kept according to default. The performance of the model was evaluated based on the accuracy, presented as the out-of-the-bag error (OOBE).[24] Variable importance was evaluated through the normalised mean decrease in GINI (MDG), which measures how much a variable reduces the GINI impurity metric within a particular class.[24] PLS-DA and PLS models were created using the "*mdatools*" Package. In the case of BA adulteration, 4 separate PLS models were created for the quantification of BA in each individual EO class. PLS model was also created to quantify VO in Frs EO. In the case of Yly EO, three PLS models were created to quantify BA, VO and Cyp EO.

A training/validation sample set was used to create and optimise the models, through leave-one-out cross validation (LOO CV). In this step, it was determined optimal number of latent variables, as the lowest value for root mean square error of cross-validation (RMSECV) before it starts to rise, meaning that addition of the next LV does not increase the performance of the models. An independent test set was used to evaluate the performance of all models, RF, PLS-DA and PLS, providing a double cross-validation (DCV) procedure. PLS-DA models for the discrimination of adulterated EOs were assessed using a specificity and sensitivity confusion matrix.[25,26]

For the PLS models the root mean square error of cross-validation (RMSECV) was calculated. RMSECV estimates the prediction performance of CV, applying the equations:

$$\mathrm{RMSECV} = \sqrt{\frac{\sum_{i=1}^{M}(C_i - C'_i)^2}{M}} \quad (1)$$

where C_i is the known value, C'_i is the value calculated (predicted) for the validation sample set, and M is the number of samples in the validation set. Another parameter for the evaluation of the model is the relative error of cross-validation

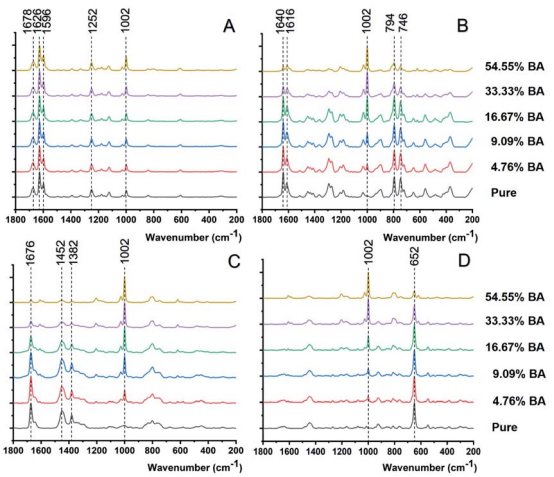


Fig. 1 Raman spectra of (A) cinnamon, (B) clove bud, (C) geranium bourbon and (D) rosemary pure essential oils and those adulterated with different quantity of benzyl alcohol.

(RECV). The RECV shows the predictive ability of the model and can be calculated as:

$$\mathrm{RECV} = \frac{\mathrm{RMSECV}}{\bar{y}} \times 100 \quad (2)$$

where $\bar{y}$ is the mean of the predicted values. The utility of PLS for quantification was assessed by calculating the residual predictive deviation (RPD). RPD can be defined as standard deviation of all values in the validation set divided by the RMSECV. If the RPD is greater than 1.5 and lower than 2.0, the model created is good for preliminary screening and initial predictions. An RPD of 2.0–2.5 is considered satisfactory for prediction, while an RPD greater than 3.0 indicates good predictive accuracy using the model.[27–29] Two parameters were used to evaluate performance in the independent test set, relative error of prediction (REP), and root mean squares error of prediction (RMSEP) using eqn (1) and (2), where C'_i is the value predicted for the independent sample set. The importance of variables was assessed using variable importance in projection (VIP).

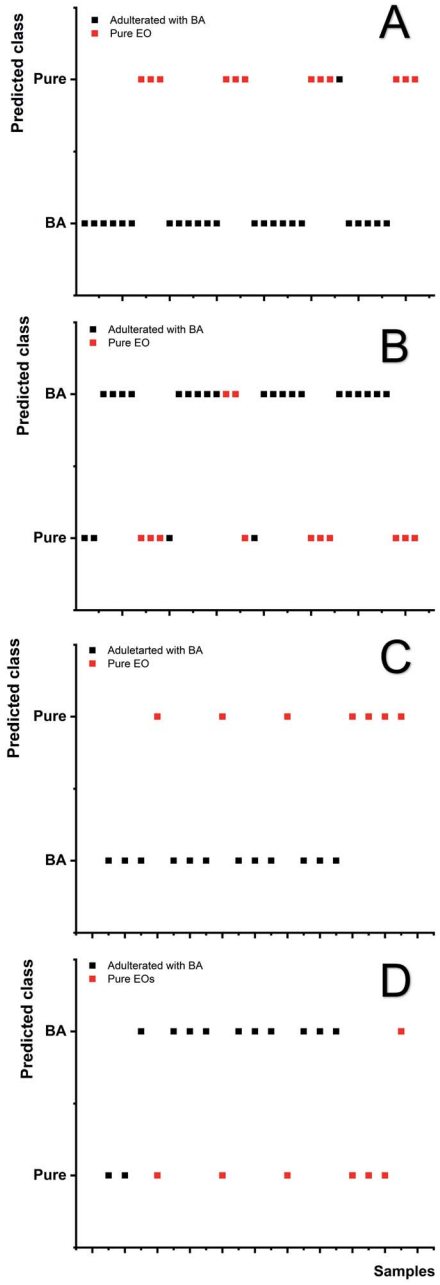


Fig. 2 Prediction classes using the Random Forest algorithm on the (A) training/validation set and (C) independent set and PLS-DA on the (B) training/validation set and (D) independent set.

3 Results and discussion

3.1 Adulteration of EOs by addition of benzyl alcohol

Addition of cheaper synthetic oils and solvents, such as BA, is a common practise in EO adulteration. Fig. 1 depicts the Raman spectra recorded on the handheld device for pure Cin, Clb, Geb and Ros EOs and adulterated with BA in range from 4.76–54.55%. As it is clear from Fig. 1A, addition of BA results in increase of bands at wavenumber 1032 and 1002 cm^{-1}, which are characteristic for benzyl alcohol and arise from the in-plane C–H stretching of the ring and 2, 4, 6 radial in-phase stretching,[30] respectively. Most prominent Raman bands for Cin EO are 1678 cm^{-1}, associated with C=C stretching, 1626 and 1596 cm^{-1}, coming from the ring stretching. Additionally, band at 1252 cm^{-1} is associated with CH deformation from *cis* –CH=CH–. Recorded bands in the spectral region 1678–1596 cm^{-1} match the Raman spectra of the *trans*-cinnamaldehyde, main compound in the Cin EO.[31,32]

Characteristic Raman bands for Clb EO are 1640 and 1616 cm^{-1}, associated with C=C stretching, 1296 cm^{-1} which arise from $=CH_2$ rocking, as well as 794 and 746 cm^{-1}, coming from C–O stretch in alcohols and ring quadrant in-plane bend, respectively. The recorded Raman spectra match with that already reported in the literature for Clb EO.[18] In the case of Geb EO, bands 1676 and 1654 cm^{-1} are associated with C=C stretching, as well as 1452 and 1382 cm^{-1}, associated with CH_3/CH_2 bend and CH_3 bend (attached to a C=C), respectively. Also,

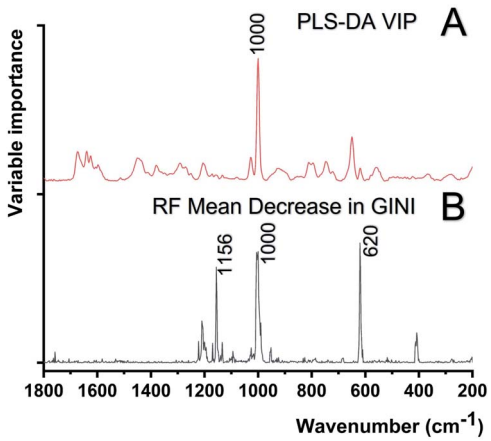


Fig. 3 Importance of Raman spectra bands for (A) PLS-DA model, presented through variable importance in projection (VIP), and (B) Random Forest model presented through mean decrease in GINI (MDG), in the discrimination of pure and EO adulterated with benzyl alcohol.

Table 3 Figures of merits for quantification of benzyl alcohol (BA) in cinnamon (Cin), rosemary (Ros), clove bud (Clb) and geranium bourbon (Geb) essential oils based on spectra recorded with smartphone-based Raman spectrometer[a]

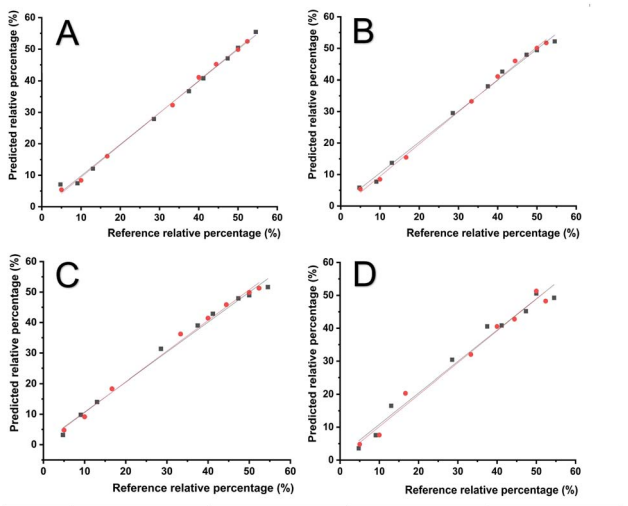
	Validation set						Independent set	
	LVs	RECV (%)	RMSECV	RPD	Slope	R^2	REP (%)	RMSEP
Cin	4	3.10	1.13	16.66	1.000	0.996	2.41	0.87
Ros	3	3.29	1.20	15.68	0.981	0.996	2.76	0.99
Clb	5	4.76	1.73	10.85	0.981	0.990	4.09	1.47
Geb	3	7.22	2.63	7.15	0.952	0.979	6.40	2.30

[a] LVs – latent variables, RECV/REP – relative error of cross validation/prediction, RMSECV/RMSEP – root mean square error of cross validation/prediction, RPD – residual predictive deviation.

a prominent band at 804 cm^{-1} comes from a C–O stretch in alcohols. This also matches with the Raman spectra for Geb EO reported elsewhere.[17] Ros EO has a single distinctive Raman band at 652 cm^{-1} associated with the ring deformation, which is distinctive for 1,8-cineole, the main compound in Ros EO.[33]

RF and PLS-DA models were created to discriminate between EOs adulterated with BA and pure EOs from 4 classes, Cin, Clb, Geb and Ros. Models were trained to discriminate between two classes, oils adulterated with BA, as one class, and pure EOs, as another. Firstly, the effect of data pre-processing was examined on the RF and PLS-DA models. An RF model created using the raw spectra had an OOBE of 2.78% (see Table S1†), whilst the PLS-DA model showed greatest accuracy with spectra processed using MSC and SG1, delivering an average accuracy of 0.833 (see Table S2†).

An RF model created using 1000 trees and 50 variables in each tree node, produced sensitivity = 0.972, where all pure EOs were positively identified as pure ones, while Ros EOs adulterated with 4.76% of BA was identified as pure EOs. In the independent test sample set all samples were correctly identified to their corresponding classes, see Fig. 2. The PLS-DA



	Cinnamon		Rosemary		Clove bud		Geranium bourbon	
	CV	DCV	CV	DCV	CV	DCV	CV	DCV
Slope	1.000	1.020	0.981	1.024	0.981	1.004	0.952	0.967
Intercept	-0.246	-0.784	0.692	-0.801	0.895	0.511	1.334	0.508
R^2	0.996	0.998	0.996	0.997	0.990	0.994	0.979	0.984

Fig. 4 Predicted relative percentage in the cross-validation (CV) (black square) and in independent test set (double cross validation (DCV)) (red circle) *vs.* reference relative percentage for the quantification of benzyl alcohol in (A) cinnamon, (B) rosemary, (C) clove bud and (D) geranium bourbon EOs.

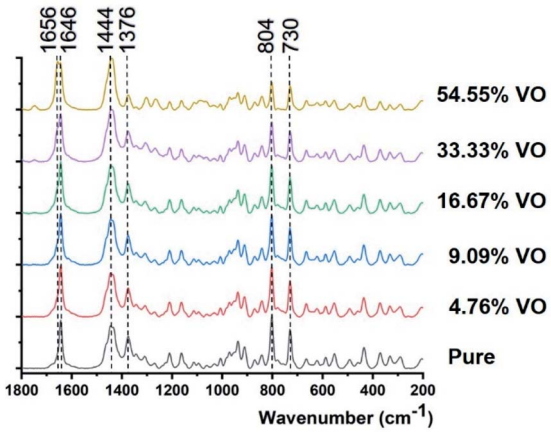


Fig. 5 Raman spectra of pure frankincense serrata (Frs) essential oil and adulterated with different percentages of vegetable oil (VO).

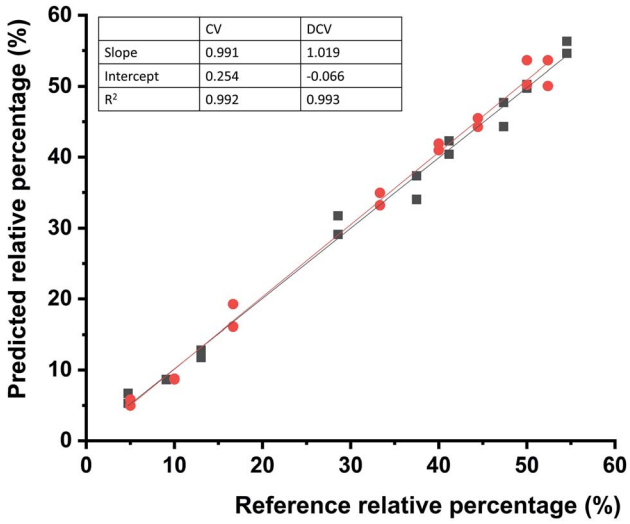


Fig. 6 Predicted relative percentage (PRP) in the cross-validation (black square) and in independent test set (red circle) *vs.* reference relative percentage for the quantification of vegetable oil in frankincense serrata EO.

model, created using 6 LVs, could not positively classify 6 samples, see Fig. 2. Accuracy of the prediction on the independent test set was 0.842, where two Cin EO adulterated with 5.0 and 10.0% of BA were identified as pure EOs, as well as one pure Ros EO was identified as adulterated.

Observing the variable importance, the spectral band at 1000 cm^{-1} associated with BA was most important for discrimination of adulterated EOs from pure ones, in both MDG and VIP, see Fig. 3. Band at 1156, indicated by RF MDG, had higher intensity in Clb EOs, while spectral band at 620 cm^{-1} was characteristic for EO adulterated with BA, see Fig. 3B.

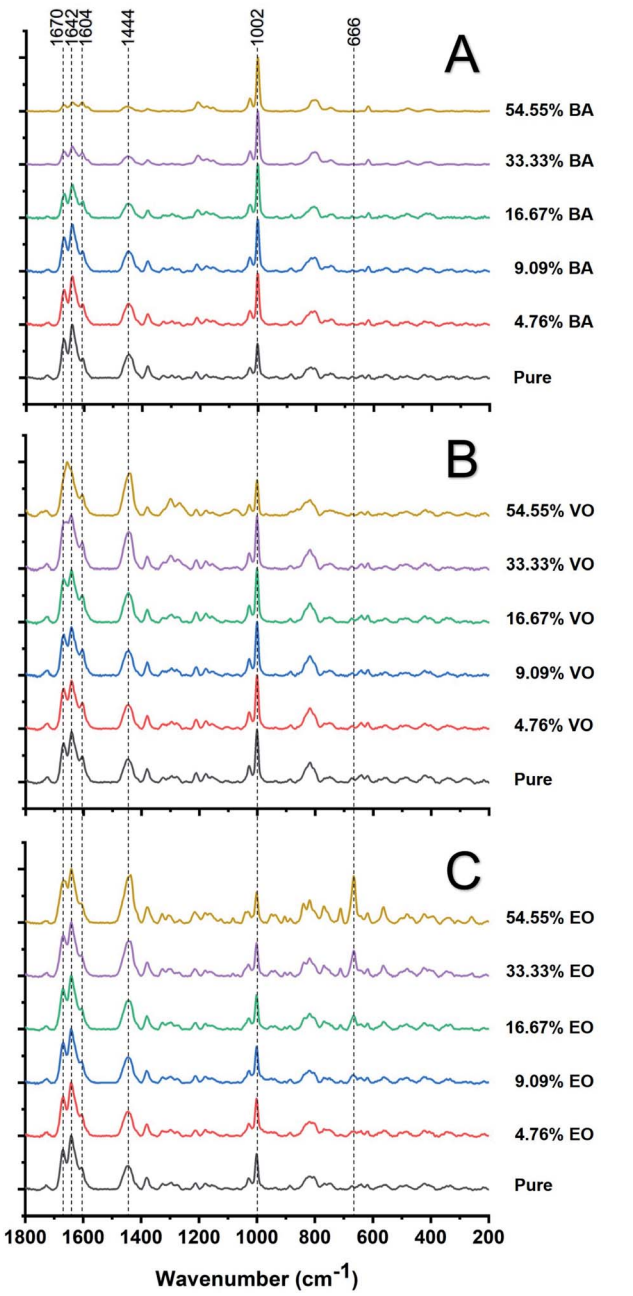


Fig. 7 Raman spectra of pure ylang–ylang EO and adulterated with different percentages of (A) benzyl alcohol (BA), (B) vegetable oil (VO) and (C) cypress EO.

Quantification of BA was performed using a PLS algorithm, where individual models were created for each of the EO classes adulterated with BA. As can be seen from Table S3,† spectra processed for MSC and SG1 showed the greatest accuracy. The created models showed excellent performance with both validation and independent test sets, with REP = 2.41–6.40% and RMSEP = 0.87–2.30, see Table 3 and Fig. 4. Also, with RPD being greater than 6.0%, the PLS model showed great prediction power for the quantification of BA.

3.2 Adulteration with vegetable oil

As it can be observed in Fig. 5, adulteration of Frs EOs with VO caused an increase in the intensity of bands at 1656 and 1444 cm^{-1}. One of the characteristic bands for Frs EO is 1646 cm^{-1}, which is associated with C=C stretching. Proximity of the 1656 and 1646 cm^{-1} shifts, as well as Raman band at 1444 cm^{-1}, being present in both Frs EO and pure VO, does not allow clear observations of adulterations with VO, especially at concentrations of VO lower than 10%, see Fig. 5. Additional clear bands observed in pure Frs EOs are at 1376 cm^{-1}, associated with CH_3 bend attached to a C=C, as well as at 804 and 730 cm^{-1}, which come from C–O stretch in alcohols and ring quadrant in-plane bending, respectively.

A PLS model created using raw spectra showed the highest accuracy, see Table S4.† Good accuracy was observed for the validation set, with RECV = 4.25%, RMSECV = 1.55, and RPD = 11.77, while for the independent sample set, figures of merit showed good accuracy, with REP = 4.39%, and RMSEP = 1.58 see Fig. 6. With the RPD = 11.77, the PLS model created using spectra recorded with smartphone-based Raman spectrometer revealed excellent performance for quantification of VO in these EOs.

3.3 YY adulteration with vegetable oil, benzyl alcohol and lower price EO

On the example of Yly EO adulteration, discrimination between pure and adulterated oils was attempted, whilst at the same

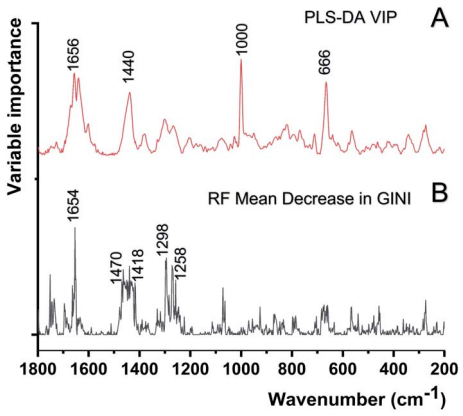


Fig. 8 Importance of Raman spectra bands for (A) PLS-DA model, presented through variable importance in projection (VIP), and (B) Random Forest model presented through mean decrease in GINI (MDG), in the discrimination of pure ylang–ylang EOs from adulterated EOs.

Table 4 Figures of merits for quantification of benzyl alcohol (BA), vegetable oil (VO) and cypress essential oil (Cyp EO) in ylang–ylang essential oils based on spectra recorded with smartphone-based Raman spectrometer[a]

	Validation set						Independent set	
	LVs	RECV (%)	RMSECV	RPD	Slope	R^2	REP (%)	RMSEP
BA	1	4.94	1.80	11.13	0.993	0.990	7.59	2.73
VO	2	4.31	1.57	12.77	0.982	0.992	3.37	1.21
Cyp EO	2	2.82	1.17	15.89	1.002	0.997	4.41	1.83

[a] LVs – latent variables, RECV/REP – relative error of cross validation/prediction, RMSECV/RMSEP – root mean square error of cross validation/prediction, RPD – residual predictive deviation.

time identifying the type of adulteration (VO, BA or other EO). Cyp EO was used as the EO adulterant, due to its lower price, compared to the price of Yly EOs.[34,35] Fig. 7 presents Raman spectra of the pure Yly samples and adulterated with BA, VO and Cyp EO at different concentrations. Characteristic Raman bands for the pure Yly EOs were 1670, 1642, 1444 and 1002 cm^{-1}. 1670 and 1642 cm^{-1} were assigned to C=C stretching, while 1444 cm^{-1} was associated with CH_3/CH_2 bending and 1002 cm^{-1} with ring 2, 4, 6 radial in-phase stretching. Bands 1670, 1642 and 1444 cm^{-1} are characteristic for Raman spectra of β-caryophyllene,[36] one of the main compounds in the Yly EOs.[6] The band at 1002 cm^{-1} is characteristic for the benzyl derivatives, such are benzyl benzoate[37] and benzyl acetate,[38] high abundant compounds in Yly EOs.[39,40] As already mentioned, addition of BA causes the increase in two bands, at 1032 and 1002 cm^{-1}, while addition of VO causes an increase in bands at 1656 and 1444 cm^{-1}. In the case of addition of Cyp EO, intensities of two bands were increased. First one was 1444 cm^{-1}, while the other one was 666 cm^{-1}, associated with the $=CH_2$ twist in the ring.

RF and PLS-DA models were created on the raw Raman spectra, as they showed greatest accuracy (Tables S5 and S6†). The created RF model accurately identified all EOs. Also, all 15 samples in the independent test set, were correctly identified. In the case of PLS-DA, created on 7 LVs, one pure sample was identified as adulterated with EO. All EOs in the independent test set were correctly identified within corresponding classes.

The importance of the variables was assessed through normalised MDG for RF and VIP for the PLS-DA model. The main contribution to the PLS-DA model came from the spectral region 1656–1640 cm^{-1}, and bands 1440, 1000 and 666 cm^{-1} (Fig. 8A), which can be directly connected with the bands in the Raman spectra of the pure Yly EOs and spectra of adulterated Yly EOs with VO, BA and Cyp EO, respectively. In the case of the RF model, MDG indicated bands 1654 cm^{-1}, as well as spectral regions 1470–1418, and 1298–1258 cm^{-1}, as the most important for the discrimination among pure and adulterated YY EOs (Fig. 8B). Increase in the intensity of the spectral band at 1654 and spectral region 1470–1418 cm^{-1} was characteristic for both addition of VO and Cyp EO, while a decrease in its intensity was associated with addition of BA. Spectral region 1298–1258 cm^{-1} showed increase in the EOs adulterated with VO.

PLS models for the quantification of adulterants were created on data pre-processed by MSC, which showed highest accuracy in prediction (Table S7†). As it can be observed from Table 4 and Fig. S3,† quantitative performance of all three models was again excellent with both RECV and REP being <7.6%, and RMSECV and RMSEP < 3.0. Also, with RPD being greater than 10.0 for all three models, it could be shown that the PLS models had suitable performance for the quantification of adulterants in Yly EOs based on the recorded Raman spectra.

4 Conclusions

In this work presented herein fast non-invasive methods for the detection of adulteration and quantification in EOs based upon spectra from a portable handheld Raman spectrometer are presented. Three different adulterations were examined, dilution with solvent (benzyl alcohol), vegetable oil and a cheaper EO. Two discriminatory algorithms were used for this task, RF and PLS-DA. In all the examples RF machine learning algorithm showed equal or better performance than PLS-DA, being able to discriminate pure from the adulterated EOs. On the example of ylang–ylang EOs, it was demonstrated that spectra recorded with the smartphone-based Raman spectrometer can be used to discriminate adulterated samples, at the same time as identifying the adulterant, even when the adulterant was present at less than 10%. Spectra recorded with smartphone-based device followed by analysis using the PLS algorithm showed great accuracy in quantification of the adulterant in the adulterated samples, with RECV being <8.0% and RMSECV < 3.0. With RPD > 6.0, it also achieved excellent performance in quantifying different adulterants.

Author contributions

Leo Lebanov: investigation, conceptualization, methodology, visualisation, data curation, writing – original draft; Brett Paull: conceptualization, methodology, writing – review & editing.

Conflicts of interest

The authors have declared that there is no conflict of interest.

Acknowledgements

Authors would like to thank the University of Tasmania and PALs Research Hub for their generous provision of co-funded PhD scholarship and project support for Leo Lebanov. Authors would also like to thank Plant Therapy company for

providing the essential oil samples, and Dr Stuart Rumble (Quark Photonics) for the loan of the portable spectrometer used within this study.

References

1 https://www.grandviewresearch.com/industry-analysis/essential-oils-market.
2 T. K. T. Do, F. Hadji-Minaglou, S. Antoniotti and X. Fernandez, *TrAC, Trends Anal. Chem.*, 2015, **66**, 146–157.
3 F. Pellati, G. Orlandini, K. A. van Leeuwen, G. Anesin, D. Bertelli, M. Paolini, S. Benvenuti and F. Camin, *Rapid Commun. Mass Spectrom.*, 2013, **27**, 591–602.
4 P. Marriott, R. Shellie and C. Cornwell, *J. Chromatogr. A*, 2001, **936**, 1–22.
5 L. Lebanov, L. Tedone, M. Kaykhaii, M. R. Linford and B. Paull, *Chromatographia*, 2019, **82**, 399–414.
6 L. Lebanov, S. Chatterjee, L. Tedone, S. C. Chapman, M. R. Linford and B. Paull, *J. Chromatogr. A*, 2020, **1618**, 460853.
7 D. Cozzolino, *Planta Med.*, 2009, **75**, 746–756.
8 C. W. Huck, *Phytochem. Lett.*, 2015, **11**, 384–393.
9 R. Aroca-Santos, J. C. Cancilla, G. Matute and J. S. Torrecilla, *J. Agric. Food Chem.*, 2015, **63**, 5646–5652.
10 R. Aroca-Santos, J. C. Cancilla, E. S. Pariente and J. S. Torrecilla, *Talanta*, 2016, **161**, 304–308.
11 X. Wu, X. Bian, E. Lin, H. Wang, Y. Guo and X. Tan, *Food Chem.*, 2021, **342**, 128245.
12 R. Deidda, P.-Y. Sacre, M. Clavaud, L. Coïc, H. Avohou, P. Hubert and E. Ziemons, *TrAC, Trends Anal. Chem.*, 2019, **114**, 251–259.
13 K. Carron and R. Cox, *Anal. Chem.*, 2010, **82**, 3419–3425.
14 A. Beganovic, L. M. Hawthorne, K. Bach and C. W. Huck, *Foods*, 2019, **8**, 49.
15 P. Vargas Jentzsch, S. Torrico-Vallejos, S. Mendieta-Brito, L. A. Ramos and V. Ciobotă, *Vib. Spectrosc.*, 2016, **83**, 126–131.
16 D. I. Ellis, H. Muhamadali, Y. Xu, R. Eccles, I. Goodall and R. Goodacre, *Analyst*, 2018, **144**, 324–330.
17 P. Jentzsch, L. Ramos and V. Ciobotă, *Cosmetics*, 2015, **2**, 162–176.
18 P. Vargas Jentzsch, F. Gualpa, L. A. Ramos and V. Ciobotă, *Flavour Fragrance J.*, 2018, **33**, 184–190.
19 C. Berghian-Grosan and D. A. Magdas, *Talanta*, 2020, **218**, 121176.
20 L. Lebanov, L. Tedone, A. Ghiasvand and B. Paull, *Talanta*, 2020, **208**, 120471.
21 L. Lebanov, L. Tedone, A. Ghiasvand and B. Paull, *Talanta*, 2020, **219**, 121208.
22 S. Kucheryavskiy, *Chemom. Intell. Lab. Syst.*, 2020, **198**, 103937.
23 A. Liaw and M. Wiener, *R News*, 2002, vol. 2, pp. 18–22.
24 M. Belgiu and L. Drăguț, *ISPRS J. Photogramm. Remote Sens.*, 2016, **114**, 24–31.
25 D. I. Broadhurst and D. B. Kell, *Metabolomics*, 2006, **2**, 171–196.
26 J. A. Westerhuis, H. C. J. Hoefsloot, S. Smit, D. J. Vis, A. K. Smilde, E. J. J. van Velzen, J. P. M. van Duijnhoven and F. A. van Dorsten, *Metabolomics*, 2008, **4**, 81–89.
27 A. Villar, S. Fernández, E. Gorritxategi, J. I. Ciria and L. A. Fernández, *Chemom. Intell. Lab. Syst.*, 2014, **130**, 68–75.
28 A. Rudnitskaya, S. M. Rocha, A. Legin, V. Pereira and J. C. Marques, *Anal. Chim. Acta*, 2010, **662**, 82–89.
29 W. Saeys, A. M. Mouazen and H. Ramon, *Bioprocess Eng.*, 2005, **91**, 393–402.
30 P. Vargas Jentzsch, *Am. J. Biomed. Sci.*, 2019, **4**, 10–11.
31 A. M. M. da Silva, A. M. Amado, P. J. A. Ribeiro-Claro, J. Empis and J. J. C. Teixeira-Dias, *J. Carbohydr. Chem.*, 1995, **14**, 677–684.
32 Y.-Q. Li, D.-X. Kong and H. Wu, *Ind. Crops Prod.*, 2013, **41**, 269–278.
33 H. Schulz, G. Özkan, M. Baranska, H. Krüger and M. Özcan, *Vib. Spectrosc.*, 2005, **39**, 249–256.
34 https://www.youngliving.com/en_AU/products/cypress-essential-oil-15ml.
35 https://www.youngliving.com/en_AU/products/ylang-ylang-essential-oil-15ml.
36 J. A. Boaroto, A. B. Fernandes, L. M. Moreira, L. Silveira and C. José de Lima, *Instrum. Sci. Technol.*, 2017, **46**, 265–276.
37 https://pubchem.ncbi.nlm.nih.gov/compound/Benzyl-benzoate#section=Raman-Spectra.
38 https://pubchem.ncbi.nlm.nih.gov/compound/8785#section=Raman-Spectra.
39 C. Benini, M. Ringuet, J.-P. Wathelet, G. Lognay, P. Jardin and M.-L. Fauconnier, *Flavour Fragrance J.*, 2012, **27**, 356–366.
40 M. Brokl, M. L. Fauconnier, C. Benini, G. Lognay, P. du Jardin and J. F. Focant, *Molecules*, 2013, **18**, 1783–1797.

