

ENGINEERING

A comparison of ATR-FTIR and Raman spectroscopy for the non-destructive examination of terpenoids in medicinal plants essential oils

Rahul Joshi¹, Sushma Kholiya², Himanshu Pandey², Ritu Joshi^{3,4}, Omia Emmanuel¹, Ameeta Tewari², Taehyun Kim⁵, Byoung-Kwan Cho^{1,6,*}

¹Department of Biosystems Machinery Engineering, College of Agricultural and Life Science, Chungnam National University, Daejeon 34134, Korea

²Department of Chemistry, Moti Ram Babu Ram Government Post Graduate College Haldwani, Kumaun University, Uttarakhand 263139, India

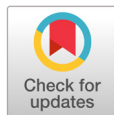
³Department of Biochemistry and Biophysics, Texas A&M University, College Station, Texas 77843, United States

⁴Department of Biomedical Engineering, Texas A&M University, College Station, Texas 77843, United States

⁵Department of Agriculture Engineering, National Institute of Agricultural Science, Rural Development Administration, Wanju 55365, Korea

⁶Department of Smart Agriculture Systems, College of Agricultural and Life Science, Chungnam National University, Daejeon 34134, Korea

*Corresponding author: chobk@cnu.ac.kr



OPEN ACCESS

Citation: Joshi R, Kholiya S, Pandey H, Joshi R, Emmanuel O, Tewari A, Kim T, Cho BK. A comparison of ATR-FTIR and Raman spectroscopy for the non-destructive examination of terpenoids in medicinal plants essential oils. Korean Journal of Agricultural Science 50:675-696. <https://doi.org/10.7744/kjoas.500408>

Received: August 14, 2023

Revised: September 10, 2023

Accepted: September 26, 2023

Copyright: © 2023 Korean Journal of Agricultural Science



This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Terpenoids, also referred to as terpenes, are a large family of naturally occurring chemical compounds present in the essential oils extracted from medicinal plants. In this study, a non-destructive methodology was created by combining ATR-FT-IR (attenuated total reflectance-Fourier transform infrared), and Raman spectroscopy for the terpenoids assessment in medicinal plants essential oils from ten different geographical locations. Partial least squares regression (PLSR) and support vector regression (SVR) were used as machine learning methodologies. However, a deep learning based model called as one-dimensional convolutional neural network (1D CNN) were also developed for models comparison. With a correlation coefficient (R^2) of 0.999 and a lowest RMSEP (root mean squared error of prediction) of 0.006% for the prediction datasets, the SVR model created for FT-IR spectral data outperformed both the PLSR and 1 D CNN models. On the other hand, for the classification of essential oils derived from plants collected from various geographical regions, the created SVM (support vector machine) classification model for Raman spectroscopic data obtained an overall classification accuracy of 0.997% which was superior than the FT-IR (0.986%) data. Based on the results we propose that FT-IR spectroscopy, when coupled with the SVR model, has a significant potential for the non-destructive identification of terpenoids in essential oils compared with destructive chemical analysis methods.

Keywords: ATR-FT-IR (attenuated total reflectance-Fourier transform infrared) spectroscopy, essential oils, Raman spectroscopy, spectral data analysis, terpenoids

Introduction

Medicinal plants have been used for thousands of years in developing countries. According to the World Health Organization (WHO), traditional healthcare systems serve 70 - 80% of the population in Africa, India, and other developing countries (Kumar and Jnanasha, 2016). The Indian state of Uttarakhand, also known as the herbal state, is a reservoir of MAPs. Traditional uses for MAPs included the treatment of leukorrhea, diabetes, kidney stones, kidney disease, cuts, wounds, fever, jaundice, stomach issues, rheumatism, and kidney stones. The industry is once again on the rise as a result of the recent significant increase in the use of natural essential oils. India is one of the few countries that can produce a majority of the essential oils required by the pharmaceutical, flavoring, and cosmetic industries (Chandra et al., 2022). Essential oils are complex mixtures of secondary metabolites that include phenylpropenes and terpenes with low boiling points (Greathead, 2003). Essential oils have distinct flavor and fragrance properties, as well as biological activities, and are widely used in aromatherapy and healthcare, as well as in cosmetics, flavorings and fragrances, spices, pesticides and repellents, and herbal beverages. Aromatic plants' antioxidant and antimicrobial activities have been extensively researched and were found to have health applications in the prevention and reduction in the risk of diseases such as inflammation, atherosclerosis, cardiovascular disease, and cancer (Gutteridge and Halliwell, 2010; Ndhlala et al., 2010). Various plant families, such as Asteraceae, Lamiaceae, Myrtaceae, Rutaceae and Verbenaceae are well known among essential oil containing plant families. These families possess various therapeutic properties and both traditional and modern uses (Samarth et al., 2017; Michel et al., 2020; Kholiya et al., 2022). Various distillation processes/methods such as gas chromatography–mass spectrometry (GC-MS) (Meng et al., 2014), high performance liquid chromatography (HPLC) (Porel et al., 2014), and hydro distillation (Irshad et al., 2020) are commonly used methods for the separation of essential oils from aromatic plant materials. These techniques enable the highest detection limits, but real-time examination of essential components in plants is very difficult due to their time-consuming, destructive, and complex sample preparation characteristics. Therefore, it is crucial to develop rapid, non-destructive technology to identify terpenoids in plant essential oils.

Molecular spectroscopic techniques such as Fourier transform infrared (FT-IR), and Raman spectroscopies have shown superior application in the handling of qualitative or quantitative researches in food and agricultural products due to their specific advantages when compare with NIR (near-infrared) spectroscopy. For example, both FT-IR and Raman spectroscopy has no influence of overtones and combination bands which are generally observed in NIR spectroscopy. Moreover, when compared with the traditional destructive chemical analysis procedures, both offer quick, simple sample preparation and non-destructive measurements. FT-IR and Raman spectroscopies have been used successfully in a number of studies to perform non-destructive evaluations of various products, such as red wines (Joshi et al., 2021), Sudan dye (Lohumi et al., 2017), Grignard reagent (Joshi et al., 2020), melamine and cyanuric acid-contaminated pet food (Joshi et al., 2023a), fabricated eggs determination (Joshi et al., 2022), and others. The identification of chemical components remains a difficult task due to the presence of various chemical compounds resulting in spectral variability. The direct observation of spectra results in the generation of incorrect information about the presence of different chemical constituents. To overcome this, it is essential to combine multivariate analysis techniques with spectrum data, thereby revealing the hidden chemical information present in the samples (Seo et al., 2021; Kim et al., 2022). For example, Huang et al. (2022) utilized Raman spectroscopy with chemometrics for the authentication and detection of adulterated agarwood essential oils. Sufriadi et al. (2021) also used principal component analysis with FT-IR spectroscopy for the discrimination of patchouli essential oils based on different geographical areas in Aceh. Also, Divyanth et al. (2022) used an application of chemometrics methods

the non-destructive prediction of Nicotine Content in Tobacco. Conventional machine learning techniques such as support vector machines, principal component regression, linear discriminant analysis, have certain limitations which are responsible for slowing down their performance during model analysis. However, recently developed deep learning techniques have resolved this problem, achieving excellent performance. Due to its superior abilities in terms of feature extraction, preprocessing, and the identification of information in a single architecture without the need for manual adjustments, the convolutional neural network (CNN), a widely used deep-learning approach, has established itself in several areas of scientific research (Chatzidakis and Botton, 2019; Jung et al., 2021; Sihalath et al., 2021; Putra et al., 2022). Fuentes et al. (2023) used Raman spectroscopy and convolutional neural networks for monitoring biochemical radiation response in breast tumor xenografts. . In another study, (Wu et al., 2022) utilized a convolutional neural network with Raman spectroscopy to tackle the complex problem of identifying and quantifying honey adulteration, highlighting the model's capability in food authenticity assessment. Furthermore, (Kawamura et al., 2021) demonstrated the potential of convolutional neural networks by combining visible and near-IR spectroscopy for soil phosphorus prediction, emphasizing the versatility of deep learning across various domains. However, it is critical to assess the advantages and limitations of using deep learning and chemometrics. Deep learning methods, such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs), on the other hand, excel at delivering highly accurate predictions by capturing complex patterns in spectral data, thus improving classification accuracy (Luo et al, 2022). They also automate the feature extraction procedure, saving time on human feature engineering. Furthermore, deep learning algorithms can effectively handle high-dimensional datasets prevalent in spectroscopy (Lotfollahi et al, 2019). Nonetheless, there are drawbacks to using deep learning and chemometrics. One major problem is the severe data requirements, as deep learning models normally thrive on huge labelled datasets, which can be sparse in chemometrics. Their fundamental complexity could hinder interpretability, making them less suited when understanding the reasoning behind forecasts is essential (Jindasa et al, 2021). Another issue is overfitting, which is especially problematic with little or noisy data (Lotfollahi et al, 2019). Furthermore, while deep learning excels in pattern identification, it may not provide precise insights into chemical interactions, implying that classical chemometric methods will remain relevant for in-depth research (Luo et al, 2022). To date, no work has been published that uses combination of machine and deep learning approach in conjunction with FT-IR, and Raman spectroscopy for the qualitative, and quantitative assessment of terpenoids in essential oils. The following two statements can, therefore, be used to describe the study's goals in more detail: 1) Comparative prediction analysis of terpenoids in medicinal plants essential oils utilizing partial least squares regression, support vector regression and one-dimensional convolutional neural networks. 2) Rapid classification of essential oils based on geographic locations using support vector machine classifier.

Materials and Methods

Plant Materials

In the present study, ten plant species from five families were selected. The five families were Lamiaceae and Myrtaceae, which each included a maximum number (three) of plant species; Asteraceae included two plant species, while both Rutaceae and Verbenaceae included one species each. The ten plant species were: *Melaleuca linariifolia* Sm (S1), *Melaleuca bracteata* F. Muell (S2), *Callistemon citratus* (S4), *Murraya koeingii* (S7), *Lanata Camara* (S10), *Ageratum conyzoides* (S12), *Wedelia chinesis* (T2), *Ocimum gratissimum* (O1), *Ocimum kilimandascharicum* (O2), and *Thymus linearis* (T1). Fresh plant material from the ten selected species was collected in the year 2022 from a wild region between Kathgodam (latitude 29.24°N and longitude 79.53°E, 554 m) and Pantnagar (latitude and longitude) from Nainital District of Uttarakhand, India. Fig. 1 below presents the images of some of the plants species which were collected during sample preparation and further utilized for essential oil extraction using chemical analysis methods.



Fig. 1. Images of some plants taken at the time of sample collection.

Isolation of essential oil

Each sample was made by first cleaning the fresh aerial part of the plant, followed by a 4-hour hydro-distillation process in a device of the Clevenger type (Clevenger, 1928). This process was done to extract the essential oil (Clevenger, 1928). Flowers were hydro-distilled separately from the plant's aerial portion (its stems and leaves). Essential oil was measured directly in the extraction burette, and contents (%) were determined as the volume (mL) of essential oil per 200 g of freshly weighed plant material. After being dried over anhydrous Na_2SO_4 , the crude oil was stored in a refrigerator until samples underwent GC-FID (gas chromatography-flame ionization detection) analysis.

GC-FID analysis

The 0.2 μL neat essential oils were analyzed by using GC Thermo Fischer Trace-1300 (Thermo Fisher Scientific Inc.,

USA). The capillary column type was TG-5MS (30 m × 0.25 mm, 0.25 μm film thickness). The carrier gas was Nitrogen at constant flow rate of 1.0 mL·min⁻¹ and average velocity of 30 mL·min⁻¹. The injector temperature was 240°C, the split ratio was 1 : 40, and the detector (flame ionization detector, FID) temperature was 250°C. The initial column oven temperature was set at 70°C to 220°C at the rate of 4°C·min⁻¹. The relative content of each constituent was calculated based on the % peak area (FID response) without using a correction factor. The composition of the terpenoids found in ten different essential oils, as determined by the GC-FID reference analysis method, is shown in Table 1 below.

Table 1. The reference values of terpenoids (%) in essential oils acquired through GC-FID analysis for ten categories (S1, S2, S4, S7, S10, S12, O1, O2, T1, T2) of samples.

Compounds name	S1	S2	S4	S7	S10	S12	O1	O2	T1	T2
α-Pinene	0.33	0.13	12.88	69.60	12.31	ND	t	2.06	0.4	33.6
β-Pinene	0.35	0.31	0.75	11.84	9.74	0.21	0.18	1.89	1.2	0.8
Myrcene	1.72	ND	0.21	ND	2.01	0.17	ND	ND	ND	ND
Limonene	3.59	0.56	2.24	6.08	1.93	0.07	ND	ND	ND	2.0
1,8-Cineole	62.36	0.21	72.81	0.30	3.04	ND	ND	12.68	ND	ND
Terpinolene	0.06	1.42	0.06	0.29	0.36	ND	ND	1.99	ND	ND
Linalool	0.30	1.19	0.31	0.25	0.73	ND	ND	1.84	ND	ND
Terpinene-4-ol	1.05	0.15	0.54	0.08	0.32	ND	0.07	3.10	ND	1.5
α-Terpineol	13.12	1.22	5.39	ND	0.43	ND	ND	ND	ND	ND
Methyl Eugenol	ND	75.82	ND	ND	ND	ND	ND	ND	ND	ND
(E)-Methyl cinnamate	ND	6.42	ND	ND	ND	ND	ND	ND	ND	ND
β-Caryophyllene	3.31	0.20	0.16	4.77	5.95	20.20	1.06	0.91	2.9	3.7
Germacrene-D	0.17	0.83	0.06	1.07	12.23	1.48	3.34	0.68	ND	1.9
Caryophyllene oxide	0.73	0.05	0.07	0.03	ND	ND	ND	ND	ND	ND
α-Copaene	0.18	ND	ND	ND	1.93	ND	ND	ND	ND	ND
β-Elementene	ND	ND	ND	0.23	1.91	ND	ND	ND	ND	ND
Davanone	ND	ND	ND	ND	2.51	ND	ND	ND	ND	ND
Bomyl acetate	ND	ND	ND	ND	ND	0.27	ND	ND	ND	ND
β-Cubebene	ND	ND	ND	ND	ND	1.56	ND	ND	ND	ND
Precocene II	ND	ND	ND	ND	ND	63.14	ND	ND	ND	ND
Precocene I	ND	ND	ND	ND	ND	3.38	ND	ND	ND	ND
Camphene	ND	ND	ND	ND	ND	ND	ND	8.44	0.2	0.1
β-Ocimene	ND	ND	ND	ND	ND	ND	7.53	5.11	ND	ND
Camphor	ND	ND	ND	ND	ND	ND	ND	52.89	ND	ND
Eugenol	ND	ND	ND	ND	ND	ND	85.66	1.06	ND	ND
p-Cymene	ND	ND	ND	ND	ND	ND	ND	ND	4.4	2.3
γ-Terpinene	ND	ND	ND	ND	ND	ND	ND	ND	11.2	ND
Thymol	ND	ND	ND	ND	ND	ND	ND	ND	ND	1.5
Carvacrol	ND	ND	ND	ND	ND	ND	ND	ND	55.7	ND
(E)-Nerolidol	ND	ND	ND	ND	ND	ND	ND	ND	ND	3.2
Total (%)	87.27	88.51	95.48	94.54	55.4	90.48	97.84	92.65	79	54.1

GC-FID, gas chromatography-flame ionization detection; ND, not detected; t, trace amount.

Analysis of essential oil

Identification of essential oil constituents was accomplished on the basis of retention index (RI), determined with reference to homologous series of n-alkanes, C8 - C24; Supelco Analytical, Bellefonte PA, USA, under same temperature-programmed conditions. The relative content of individual components of the oil is expressed as percent peak area relative to total peak area of the GC-FID chromatogram automated electronic integration without response factor correction.

FT-IR Spectroscopy

A Nicolet 6700 (Thermo Fisher Scientific Inc., USA) ATR-FT-IR spectrometer was used in a lab to examine the FT-IR

spectrum of the essential oil samples. The attenuated total reflectance (ATR) sample mode was installed in the spectrometer. The system also utilized a deuterated triglycine sulfate (DTGS) detector and a beam splitter made of potassium bromide (KBr), both of which were managed by the OMNIC software (Thermo Fisher Scientific Inc., USA). Spectral acquisition was performed at wavelengths from 400 to 4,000 cm^{-1} . For the spectral acquisition, each sample was placed on the surface of the diamond crystal sampling plate. Each sample was subjected to a total of 32 scans at 4 cm^{-1} spectral intervals, and the average spectral data were stored in Excel for-mat for future analysis.

Raman Spectroscopy

A portable i-Raman spectrometer (B&W TEK Inc., USA) outfitted with a charge-coupled device (CCD) detector, and with a pixel size of 14×900 m and a 785-nm laser was used to record the Raman spectra of the essential oils from different geographical locations. To prevent light interference during spectral acquisition, the chemical sample Raman spectra were all acquired in a dark environment. With a laser light source operating at a wavelength of 785 nm and a power of 200 mW, spectra were obtained for each sample with an exposure time of 1 second and a spectral resolution of 2 cm^{-1} . The experiment used the BAC100 model (B&W TEK Inc., USA) as a standard probe. Prior to measurement during spectral collection, the cuvette was dried. Each sample was then injected with a pipette into the cuvette from the top. The sample was then placed in front of the probe, which had been calibrated beforehand to make sure the laser would reach the sample inside the cuvette, at a distance of 2 mm. High-quality Raman spectra were produced using four scans with an integration duration of 10,000 ms, and the averaged spectra of each sample were used for model construction.

Spectral preprocessing, and chemometrics analysis

Since fluorescence background signals frequently affect Raman spectra, they contribute to masking valuable information about the chemical constitution of essential oils. The polynomial curve-fitting method was used to address this problem due to its quick processing speed and simplicity compared with other fluorescence correction methods such as wavelet transformation and Fourier transformation. The main idea behind the polynomial curve approach is to employ iterative calculations to determine the proper order of polynomials. Fig. 2a and b presents the fluorescence effected, and polynomial corrected Raman spectra. In this investigation, an 8th order polynomial with 100 iterations was selected to remove the fluorescence background.

The spectral data collected using both types of spectroscopic equipment required preprocessing procedures, essential for preventing undesirable scattering and noise effects. In this study, preprocessing processes included range normalization, multiplicative signal correction (MSC), standard normal variate (SNV), and Savitzky-Golay (SG) derivatives (first and second) to the acquired raw spectral data. Mean normalization is the preprocessing technique that is most frequently employed. Finding the mean values for each dataset is the main idea. In contrast, the maximum or range values are subtracted from each data point in the max and range normalization. On the other hand, MSC and SNV preprocessing played an important role for the elimination of background offset, slope, and scattering effect from the data. However, the overlapping peaks in the spectra were resolved and the additive effects were reduced by using the Savitzky-Golay filters (SG-first and second derivatives) (Rinnan et al., 2009). Later, machine learning and deep learning models were created using the preprocessed FT-IR and Raman spectral data. The deep learning model was created and executed using Python, whereas the machine learning model was created using MATLAB (version 7, The MathWorks Inc., USA). Fig. 3 shows a flowchart

for the FT-IR and Raman spectral analysis procedure for the samples of essential oils from different geographical locations.

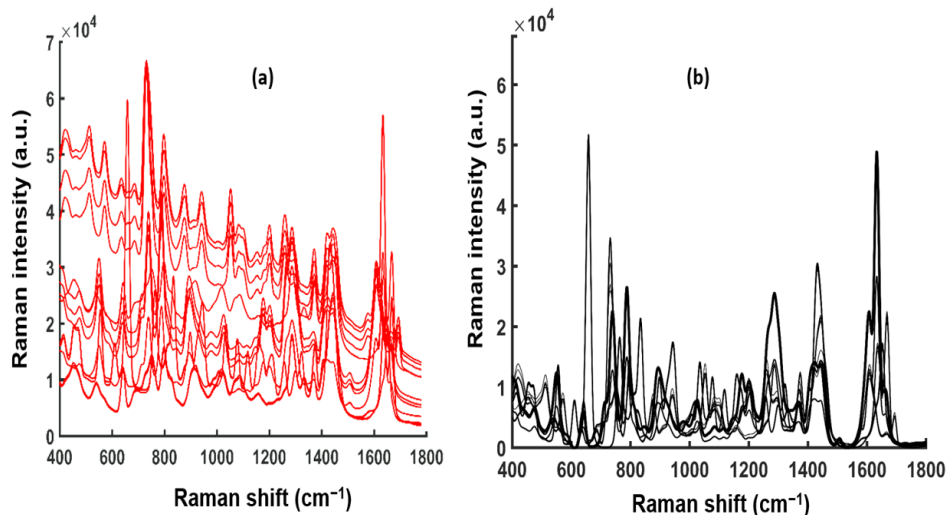


Fig. 2. Original Raman spectra (a), and fluorescence-corrected Raman spectra using the polynomial curve fitting method (b) for essential oil samples.

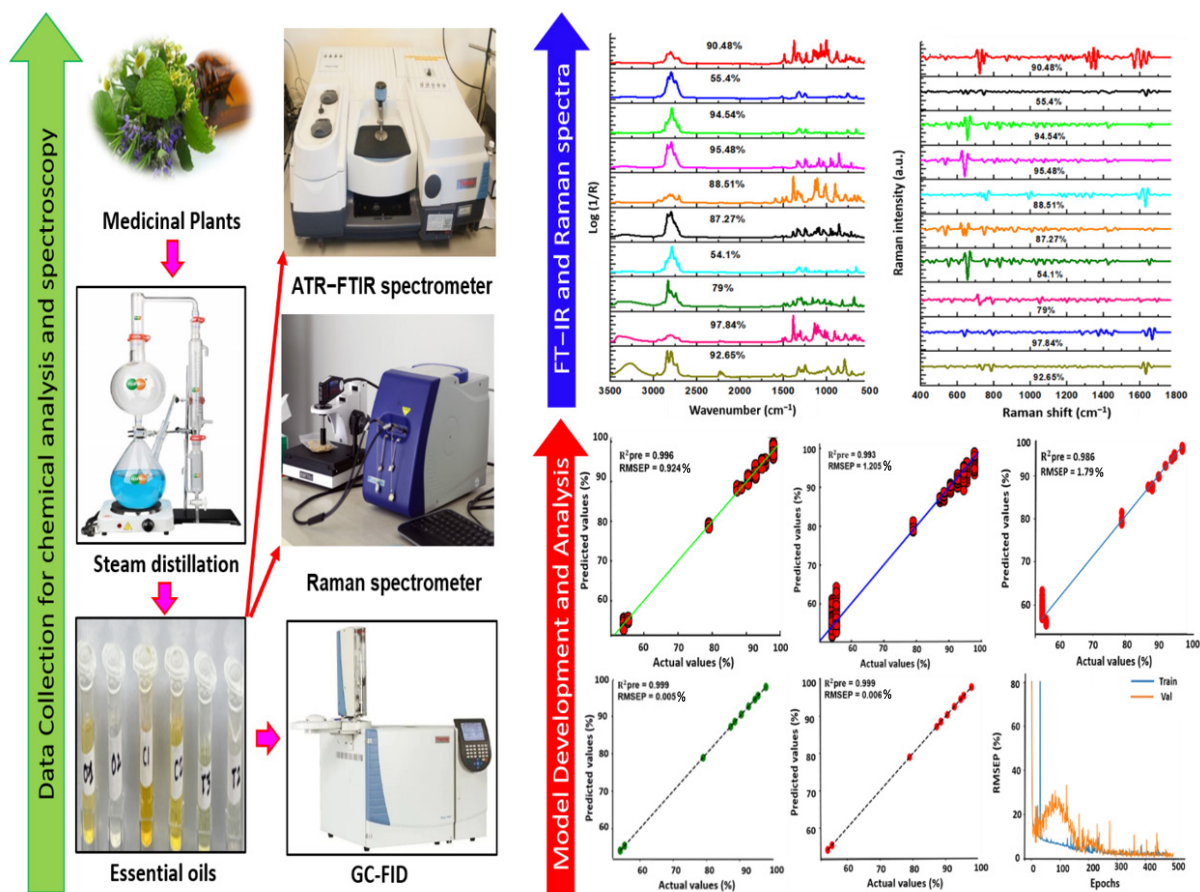


Fig. 3. Diagram illustrating the full procedure explaining the workflow for FT-IR (Fourier transform infrared) and Raman spectroscopy for the terpenoids prediction in essential oils analysis. ATR-FTIR, attenuated total reflectance-Fourier transform infrared; GC-FID, gas chromatography-flame ionization detection; R^2 pre, correlation coefficient for prediction; RMSEP, root mean squared error of prediction.

Regression and classification analysis

To perform the non-destructive quantitative evaluation of terpenoids in essential oils from different geographical locations, two widely used regression analysis methods i.e., partial least squares, and support vector machine regression were utilized in this study. Both model prediction analysis performance was compared with the GC-FID reference analysis method utilized in this research. PLSR (partial least squares regression) is very popular analytical vibrational spectroscopy methods for component quantification. It is extensively used in the spectroscopy community to relate spectral data to the physical or chemical feature being measured. In PLSR, the predictor X (predictors) and the dependent variable y (degree of adulteration) are both split into orthogonal structures referred to as latent variables (LVs), which indicate the largest correlation between X and Y (Wold et al., 2001). On the other hand, the support vector machine (SVM), a subclass of machine learning method, carries out both classification and regression tasks for the spectrum analysis. Support vectors were employed to find a hyperplane that closely corresponds to the relationship between continuous target variables and input variables. The SVM aims to increase the margin while reducing the error to a preset level (Cortes and Vapnik, 1995). SVR can manage nonlinear interactions using the kernel approach. In this work, 10 different geographic regions' worth of essential oils were collected, and we used a support vector machine (SVM) as a classification technique for the qualitative discrimination of those oils. Researchers have demonstrated the strong potential of PLSR, and SVR algorithm in the areas of plants and food products with different spectroscopic techniques like FT-IR (Elzey et al., 2016), and Raman spectroscopy (Li et al., 2021).

1D CNN model architecture

The inquiry inputs of the 1D CNN model were one-dimensional data. The features were obtained utilizing a one-dimensional convolution kernel. In this study, a deep learning-based quantitative analysis model named 1D CNN was created similar to those in our previously published report (Joshi et al., 2023b) with little modifications using FT-IR, and Raman spectral data for the quantitative evaluation of terpenoids in essential oils and its performance was compared to that of machine learning regression techniques such as PLSR, and SVR. The model comprised of three convolutional layers (Conv1D_1, 2, and 3), one max-pooling layer, one flattened layer, three dense layers (Dense 1, 2, and 3), rectified linear unit (ReLU) as an activation function, and a regression output layer. Fig. 4 depicts the 1D CNN model. Fig. 4a, b presents the FT-IR, and Raman spectral data for the essential oils from different geographical locations, which consists of terpenoids of varying concentrations. Fig. 4c represents the detailed information linked to the architecture of the designed 1D CNN model for this study. The Jupyter framework and Python 3.11 with Tensor Flow were used to implement the 1D CNN models. Table 2 lists the 1D CNN architecture specifications and all the parameters used during the training process for both chemicals added to pet food.

The input layer for the 1D CNN model is made up of FT-IR and Raman spectra with an input size of $3,800 - 500 \text{ cm}^{-1}$, and $400 - 1,800 \text{ cm}^{-1}$ for each component of the 1D CNN. Additionally, three distinct 1D CNN layers were used for feature extraction in the 1D CNN model, each with a different layer of filters of varied sizes. To improve the performance of the developed model, alternate kernel sizes and dropout layers were selected to minimize overfitting. Additionally, the ReLU activation function was used which often follows a convolutional layer. Further, max pooling a widely employed operation in CNN architectures was implemented to down sample feature maps, expand the receptive field, and give resistance to tiny spatial translations. Max pooling with a value of 2 was used, and root mean square error and loss function were used to evaluate performance.

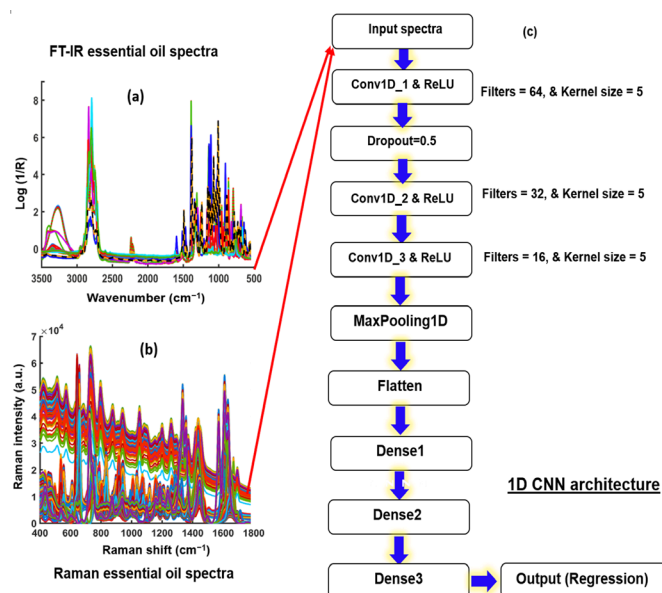


Fig. 4. Procedure used to process FT-IR (Fourier transform infrared), and Raman spectral data using a 1D CNN (one-dimensional convolutional neural network) model. (a, b) FT-IR, and Raman spectra for essential oils; (c) 1D CNN model structure details. ReLU, rectified linear unit.

Table 2. 1D CNN architecture details developed for both FT-IR and Raman spectroscopic data for plants essential oils.

Layer (type)	Output shape	Parameters	Activation function
Conv1D_1 (Conv1D)	(None, 2036, 64)	384	ReLU
dropout (Dropout)	(None, 2036, 64)	0	ReLU
Conv1D_2 (Conv1D)	(None, 2032, 32)	10,272	ReLU
Conv1D_3 (Conv1D)	(None, 2028, 16)	2,576	-
MaxPooling1D	(None, 1014, 16)	0	-
flatten	(None, 16224)	0	-
Dense_1	(None, 64)	1,038,400	ReLU
Dense_2	(None, 64)	2,080	ReLU
Dense_3	(None, 1)	33	-
Total parameters: 1,053,745			
Trainable parameters: 1,053,745			
Non-trainable parameters: 0			

1D CNN, one-dimensional convolutional neural network; FT-IR, Fourier transform infrared; ReLU, rectified linear unit.

Results and Discussion

Raw spectra for FT-IR and Raman spectroscopies

The raw spectra recorded using FT-IR and Raman spectroscopy are displayed in Fig. 5a, b. The figure shows multiple overlapping peaks caused by a variety of undesirable background characteristics, such as instrument drift, scattering effects, noises, etc. The aforementioned effects are capable of lowering the spectral quality and thereafter reducing prediction and classification performances. To avoid these problems and obtain the meaningful data required for terpenoids prediction in essential oils, it is essential to choose the best preprocessing steps. Many pre-processing techniques, including normalization

(mean and range), multiple scattering correction, standard normal variate (SNV), and Savitzky Golay derivatives (1st, and 2nd), were used in this work. Based on the root mean square error levels, a single optimal preprocessing was then selected from among these processes.

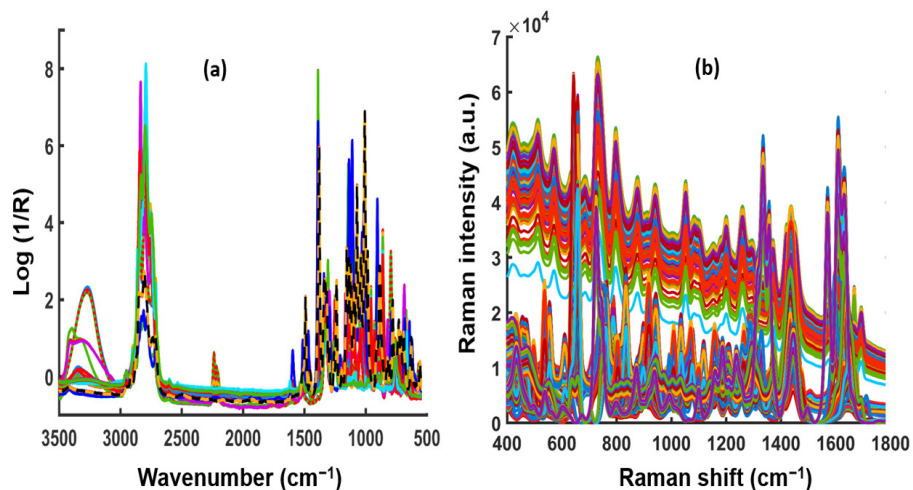


Fig. 5. Fourier transform infrared (FT-IR) (a), and Raman raw spectra (b) of essential oils.

Spectral interpretation for preprocessed FT-IR and Raman spectra for essential oils

The range-normalized FT-IR spectra for essential oils and the Savitzky-Golay 2nd derivative Raman spectroscopy preprocessed spectra are presented in Fig. 6a, b.

The FT-IR spectra were plotted in the 500 - 3,500 cm^{-1} wavenumber range, while the remaining regions i.e., below 500 cm^{-1} , and above 3,500 cm^{-1} , were omitted due to the lack of pertinent information required for terpenoids composition analysis. On the other hand, Raman spectra were plotted from 400 - 1,800 cm^{-1} , respectively. The results of both spectra represent several characteristics peaks for ten different concentrations of terpenoids present in the essential oils. Since several different

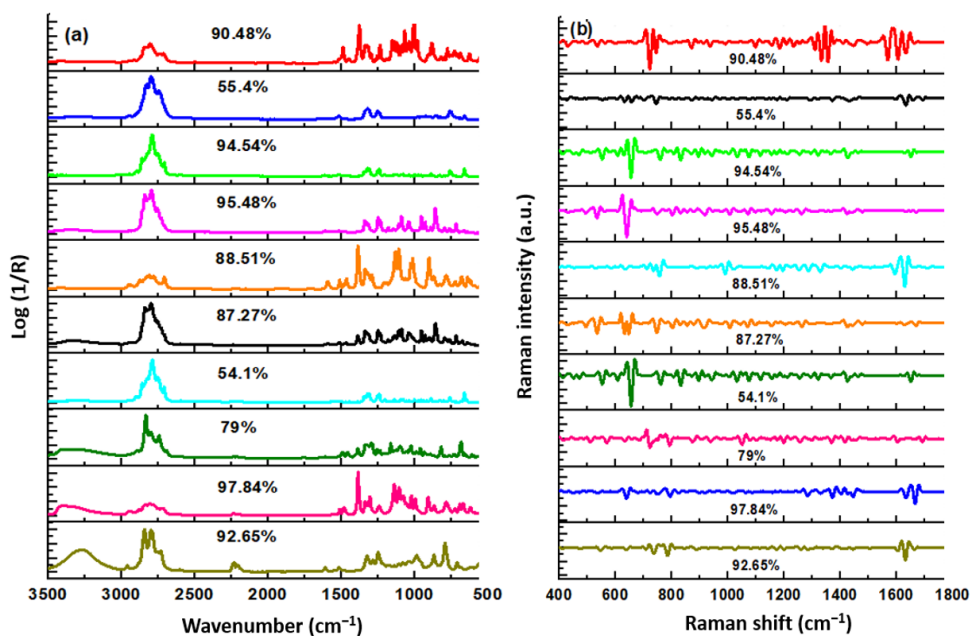


Fig. 6. FT-IR (Fourier transform infrared) range normalized spectra (a), and Savitzky-Golay 2nd derivative Raman preprocessed spectra (b) for plants essential oils.

terpenoids compounds were identified using the GC-FID analysis method in all the ten different categories of essential oils, highlighting their spectral peaks would be a time-consuming and difficult process. To overcome this problem, the most particular terpenoids that were present in higher concentrations were selected for spectrum interpretation and their chemical structures are presented in Fig. 7.

All the six different terpenoids identified through GC-FID reference analysis method deliver specific spectral signatures which are clearly presented in Tables 3, and Table 4 for the FT-IR, and Raman spectral data.

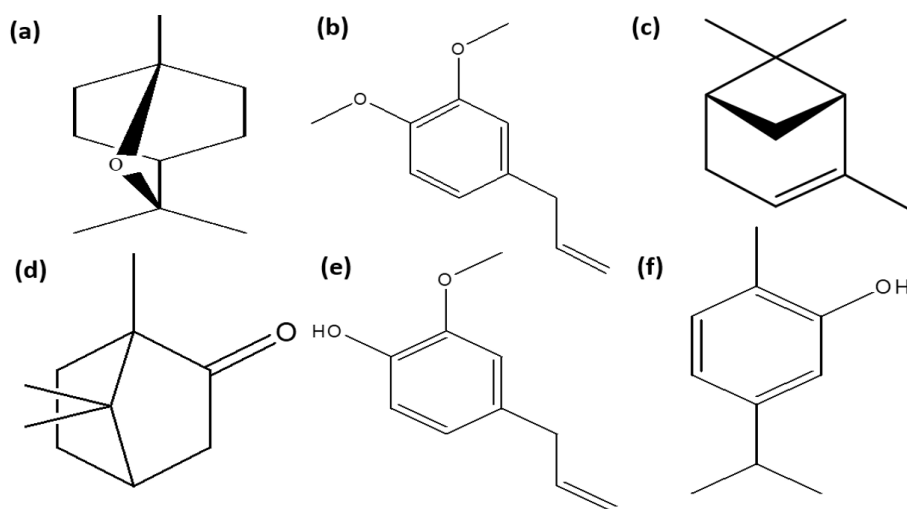


Fig. 7. Chemicals structures of (a) 1,8 cineole, (b) methyl eugenol, (c) α -pinene, (d) camphor, (e) eugenol, and (f) carvacrol.

Table 3. FT-IR and Raman characteristics vibrations of terpenoids in ten different categories of essential oils.

Terpenoids compounds	FT-IR spectral vibrations (cm^{-1})
1,8 cineole	985 allocated to symmetrical bending out of the CH_2 plane, 1,079 and 1,214 are attributed to symmetric and asymmetric stretches of group C-O-C, CH_3 deformation is discerned at 1,374 (García et al., 2017)
Methyl eugenol	1,604 - 1,640 C=C aromatic band, 810 indicates the presence of the aromatic group substitution, 1,458 shows presence of methyl groups (CH_3) (Riyanto et al., 2016)
α -pinene	1,650=C=C stretching vibration, 788 = out of plane bending, and 2,980 - 2,820 for symmetric and asymmetric stretching vibration (Masruri et al., 2016)
Eugenol	3,500 (broad band of -OH vibration in phenol groups), 3,100 (aromatic C-H stretching), 2,850 - 3,000 (symmetric and asymmetric stretching of methyl and methylene groups); the weak peak for the C=C bond is located at 3,000 - 3,040, peaks for aromatic C=C bonds at 1,510, 1,610, and 1,640 cm^{-1} (Medvecká et al., 2020)
Camphor	Significant spectral regions from 1,740 - 1,745 for C=O group (de Barros Fernandes et al., 2014)
Carvacrol	800 - 810 (C-H out of plane), 3,000 - 2,850 (C-H symmetric and asymmetric stretching), 3,360 - 3,330 (-OH stretching), 1,380, 1,360 (isopropyl group symmetric and asymmetric bending) (Valderrama and Rojas De, 2017)

FT-IR, Fourier transform infrared.

Table 4. Raman characteristics vibrations for terpenoids in ten different categories of essential oils.

Terpenoids compounds	Raman spectral vibrations (cm^{-1})
1,8 cineole	653 (attributed to the ring deformation mode). Strong band at 1,448, and weak bands at 1,275, 1,169, 807 (Baranska et al., 2005; Jentzsch and Ciobotă, 2014)
Methyl Eugenol	1,640 (stretching vibration of the exocyclic C=C bond) (Chowdhry et al., 2015)
α -pinene	1,654 (C=C stretching), 1,440 (CH_2/CH_3 bending), 666 (ring deformation) (Hanif et al., 2017)
Eugenol	1,640 (band corresponds to exocyclic C=C), 1,611 (attributed to the benzene ring) (Chowdhry et al., 2015)
Camphor	651 (ring deformation), characteristic bands for the C=O stretching and CH_3/CH_2 bending modes at 1,741 and 1,448 (Daferera et al., 2002)
Carvacrol	760 strong band due to ring vibration (Siatis et al., 2005)

Dirichlet distribution

For the purpose of this investigation, 30 samples were collected for each of the categories of essential oils using FT-IR and Raman spectroscopy. The number of samples chosen during the model creation process had a significant impact on how effectively the model was built. In this research, a phenomenon introduced by Dirichlet was exploited to avoid the impact of underfitting during machine learning. The mathematical description of this phenomenon is thoroughly detailed (Wang et al., 2019). This approach was used to create 3,000 artificial samples prior to the development of the models, which were then used to build machine learning models for both spectroscopic systems. According to previously published researches, this algorithm exhibits strong potential for the qualitative analysis of carbaryl pesticide in food products (Joshi et al., 2023b), the presence of melamine, and cyanuric acid in pet food [15], and nitrogen in soils (Patel et al., 2020). Fig. 8a, b describes the overall concept of the Dirichlet distribution used in our work. In this context, “Original sample” refers to the spectra produced for two replicates during FT-IR and Raman data collection (i.e., 79_1 and 79_5% samples, respectively), while “Sample without noise” refers to spectra that were preprocessed as a result of the production of false data in Fig. 8b.

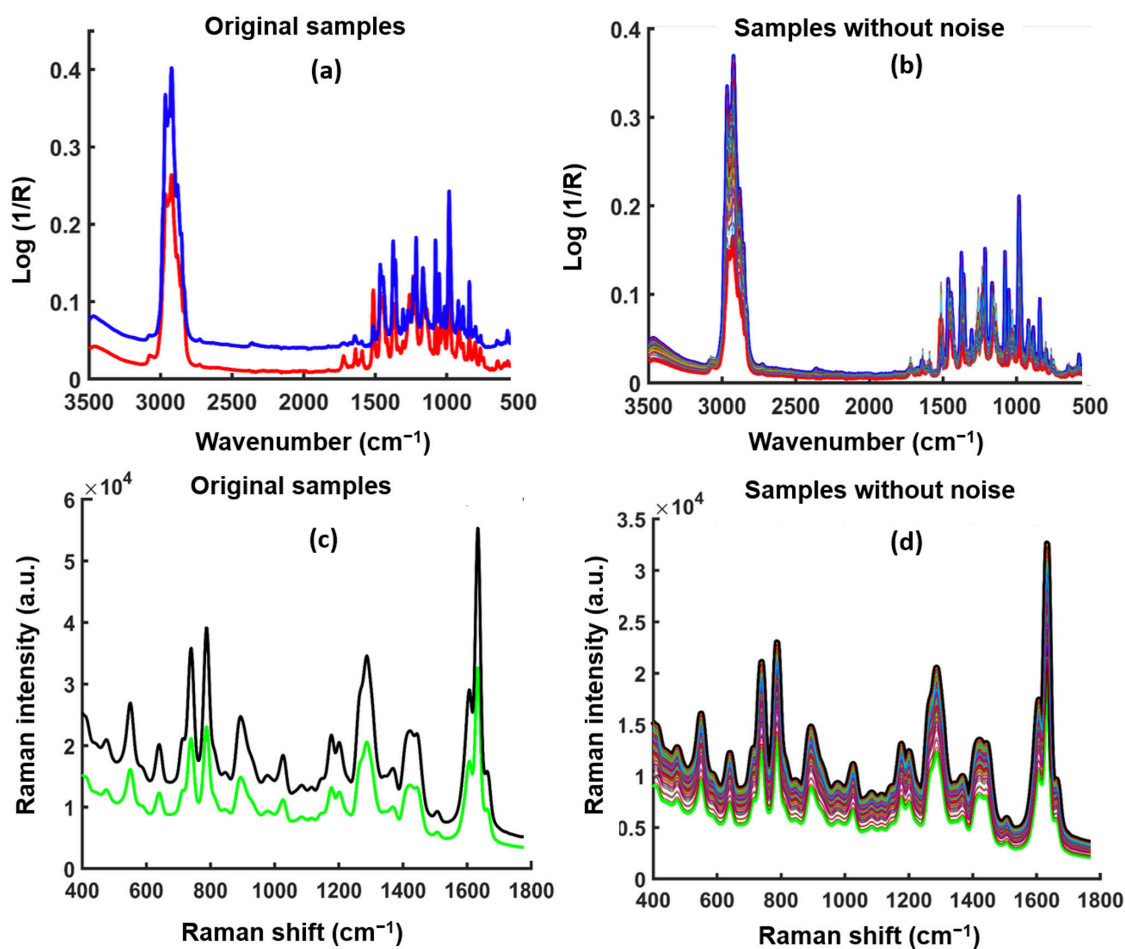


Fig. 8. FT-IR (Fourier transform infrared) spectra created between two replicates 1_1, and 1_2 for one category of essential oil (a). A total of 300 artificial samples were generated for FT-IR spectra for one category of essential oil using Dirichlet distribution (b). Raman spectra created between two replicates O1_1, and O_5 for one category of essential oil (c). A total of 300 artificial samples were generated for FT-IR spectra for one category of essential oil using Dirichlet distribution (d).

FT-IR and Raman spectroscopy model construction

The Dirichlet distribution technique produces 3,000 artificial samples that were later used to create qualitative and quantitative models. The data were splitted into calibration and prediction sets in a ratio of 70 : 30, where 2,000 samples were used as the training or calibration set and the remaining 1,000 samples were shifted to prediction or testing set. After this process, both the classification and regression models like SVC, PLSR, SVR, and 1 D CNN were developed. During the model development process, the selection of an ideal number of latent variables (LVs) is crucial to prevent overfitting issues within the model. The LVs selection was performed based on the root mean square error method during the cross-validation (leave-one out) process. Table 5 provides the complete details regarding the datasets used during model construction:

Table 5. Datasets used during classification and prediction model development.

Techniques	Total samples	Calibration	Prediction
FT-IR spectroscopy	3,000	2,000	1,000
Raman spectroscopy	3,000	2,000	1,000

FT-IR, Fourier transform infrared.

SVC classification results for FT-IR, and Raman spectroscopy

We performed a multi-class classification task on ten plant species to predict the class labels of the samples. In total, the dataset comprised of 3,000 Raman spectral observations and 3,000 FTIR spectral observation analyzed separately. Each set of data contained 2,040 spectral variables. We used a Support Vector Machine (SVM) classifier with the radial basis function (RBF) kernel and optimized the hyper parameters C and gamma through a grid search approach. The optimized hyper parameters were C = 0.0083 and gamma = scaler. The RBF kernel enabled the creation of a nonlinear decision boundary. By mapping the feature space to a higher-dimensional space, this kernel effectively captures complex relationships within the spectral data. The decision boundary created by the RBF kernel resulted in a smooth separation between the classes, allowing for accurate classification. The kernel is mathematically written as:

$$\kappa(\chi_i, \chi_j) = \exp\left(-\frac{d(\chi_i, \chi_j)^2}{2l^2}\right) \quad (1)$$

where l is the length scale of the kernel and $d(\chi_i, \chi_j)$ is the Euclidean distance (Duvenaud, 2014).

Fig. 9a, b shows the confusion matrices for the classification analysis results obtained by the Support Vector Machine (SVM) classifier for the FT-IR and Raman spectroscopic data. The results clearly show that the Raman spectroscopy based SVC model performed better than the FT-IR model, an acquired overall accuracy of 0.997 for the classification of essential oils from ten different geographical locations.

PLSR, and SVR analysis results for FT-IR and Raman spectroscopy

To accurately predict the composition of the terpenoids present in the essential oils, PLSR models were first developed using reference analysis findings obtained using GC-FID chemical analysis methods for the preprocessed FT-IR and Raman spectroscopy data. The models were subjected to several spectral treatments including normalization, MSC, SNV, and first and second SG derivatives. Of these techniques, range normalization were selected as an optimum, for FT-IR, while Savitzky-Golay 2 nd derivative was selected for Raman spectral data. With the use of seven, and eight LVS, respectively,

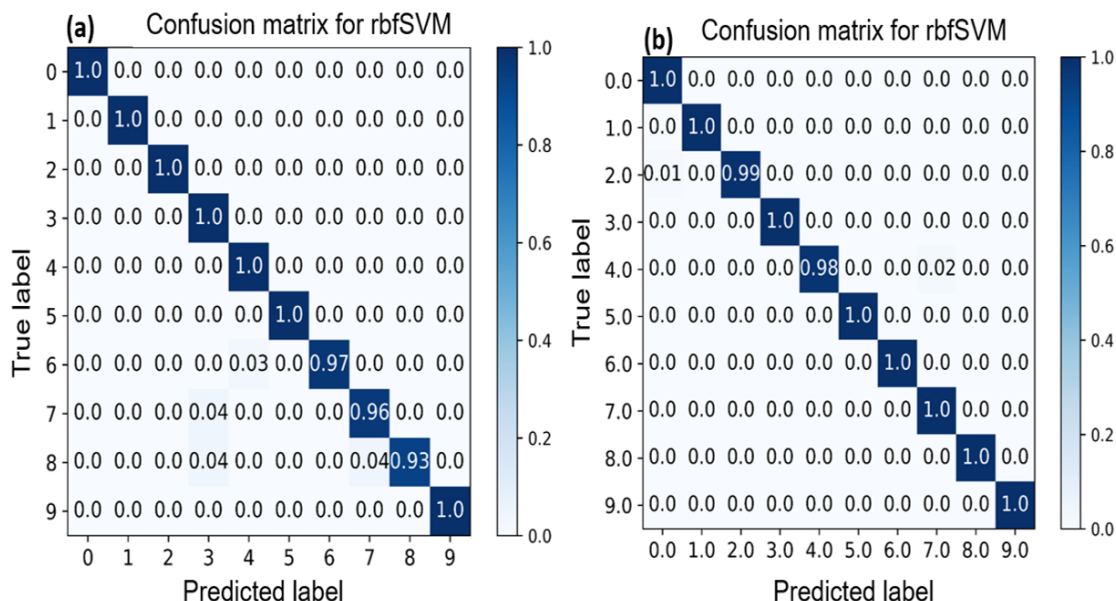


Fig. 9. Classification results using support vector classification for FT-IR (Fourier transform infrared) (a), and (b) Raman spectral data for the essential oils. rbfSVM, radial basis function support vector machine.

both PLSR models produced prediction correlation coefficients (R^2) of 0.996 and 0.993 together with RMSEP values of 0.924% and 1.205%. For FT-IR and Raman data, Fig. 10a, b shows a prediction analysis plot created by the PLSR model, demonstrating a significant correlation between actual and predicted values and showing that the model was appropriate for the quantitative analysis of terpenoids in essential oils.

Another widely used regression model SVR was created for both FT-IR and Raman spectral data to compare its prediction performances with the previously developed PLSR model. During SVR model construction, choosing an optimum kernel is very important. In this study, two different routinely used kernels namely linear, and radial basis function (rbf), were utilized. Of these two kernels, the linear kernel acquired the highest prediction performance for both spectroscopic techniques and resulted in the highest R^2 of 0.999 with the lowest RMSEP value of 0.005%, and 0.006%. The overall prediction analysis findings for both machine learning models are displayed in Table 6 below. Fig. 10c, d presented above also displays the prediction analysis plots created using the SVR models for the FT-IR and Raman datasets, which illustrate the linear correlations created between actual and predicted values. According to the results, shown in Table 6, it is abundantly obvious that the SVR models based on FT-IR spectroscopy outperformed the above developed PLSR models and achieved lower RMSE values during the quantitative analysis of terpene compositions in plant essential oils.

Table 6. Predicted terpenoids in essential oils using PLSR, and SVR models.

Models	FT-IR spectroscopy			Raman spectroscopy		
	R^2	RMSEP (%)	LVs	R^2	RMSEP (%)	LVs
PLSR	0.996	0.924	7	0.993	1.205	8
SVR	0.999	0.005	9	0.999	0.006	7

PLSR, partial least squares regression; SVR, support vector regression; R^2 , correlation coefficient for prediction; RMSEP, root mean squared error of prediction; LVs, latent variables.

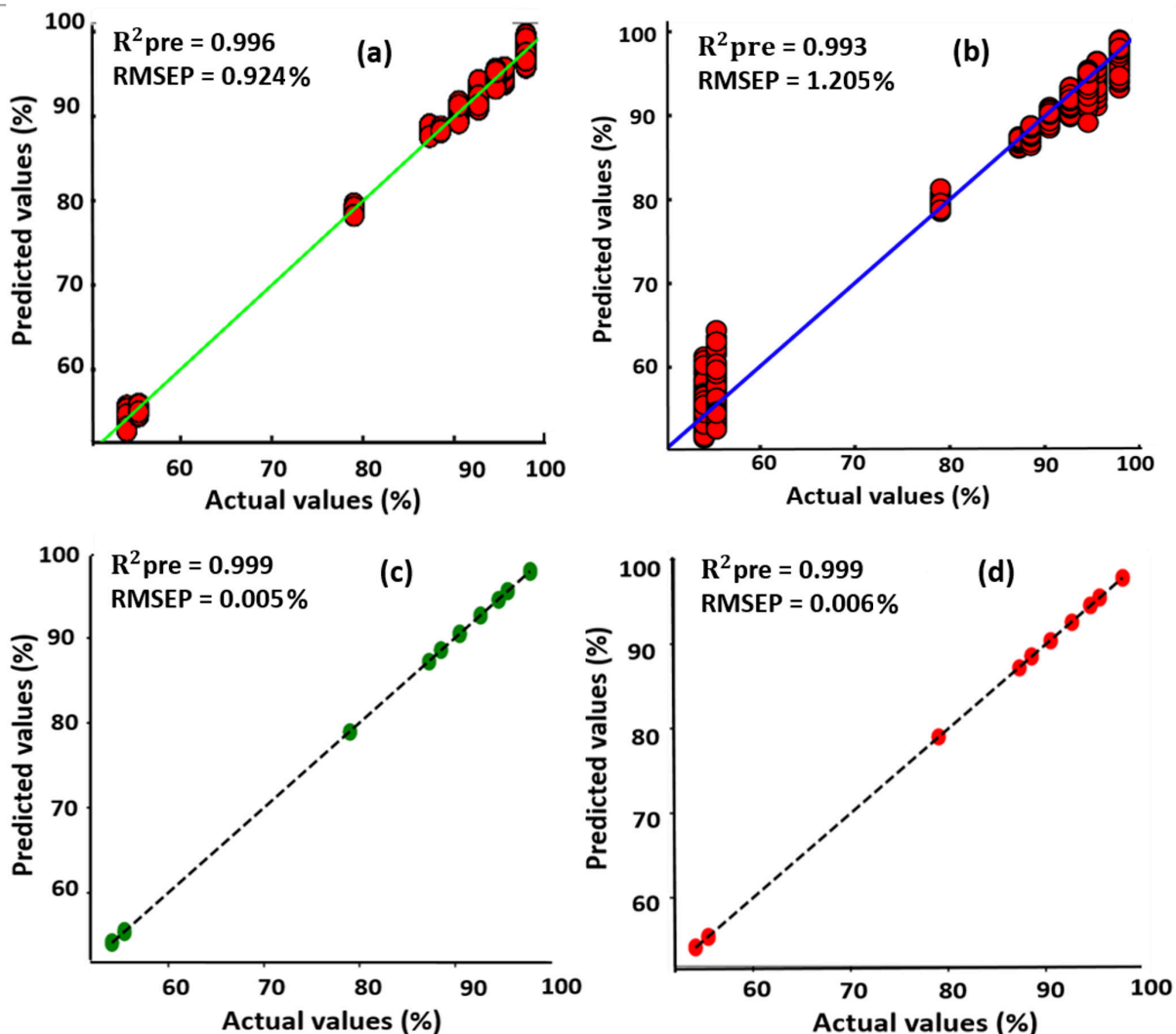


Fig. 10. PLSR (partial least squares regression) and SVR (support vector regression) prediction analysis plot between actual and predicted concentrations of terpenoids in essential oils for (a, c) FT-IR (Fourier transform infrared), and (b, d) Raman spectroscopy. R^2_{pre} , correlation coefficient for prediction; RMSEP, root mean squared error of prediction.

1 D CNN analysis results for FT-IR and Raman spectroscopy

In this study, a deep learning-based model named 1 D CNN was created in order to assess how well it performed in comparison to the machine learning models created in Section 3.6 for the quantitative prediction of terpenoids in essential oils. The model's full design is presented in depth in Section 2.9. Equal numbers of plant essential oil samples were employed in the creation of the 1D CNN model, and the correlation coefficient (R^2) and RMSEP were utilized as well as performance evaluation indicators. A total of 3,000 samples that resembled machine learning models were employed; 2,000 of these samples were moved to the calibration set, and the remaining samples were used for the prediction dataset for both FT-IR and Raman spectroscopic data. The model comprises three separate convolution layers with a dropout value of 0.5 make up the model. The ReLU activation function was chosen, and the learning rate for 1D CNN models was 0.001. The values of the epochs were changed and the best 500 epochs during the training process were selected; a total of 1,053,745 training

parameters were used. The parameters were carefully adjusted in order to maximize prediction performance. The prediction analysis results and loss-function curve obtained using the developed 1 D CNN model for the FT-IR spectroscopic data are shown in Fig. 11a, b, respectively. The developed deep learning model resulted in a R^2 value of 0.999 and with an RMSEP value of 1.10 % which is comparatively higher than those of previously constructed machine learning models. Further, the resulting loss-function curve be attaining convergence at 400 epochs for the training and validation datasets.

Similarly, another 1D CNN model was constructed for the acquired Raman spectral data using a similar number of samples for use the during training and validation process. The prediction analysis results and loss function curve obtained using the developed 1D CNN model for Raman spectroscopic data are shown in Fig. 11c, d above respectively. The developed 1D CNN model yielded a higher correlation coefficient (R^2) value of 0.986 with the lowest RMSEP value of 1.79 %. Further, for the training and validation datasets, convergence began after 400, which is shown in loss-function curve.

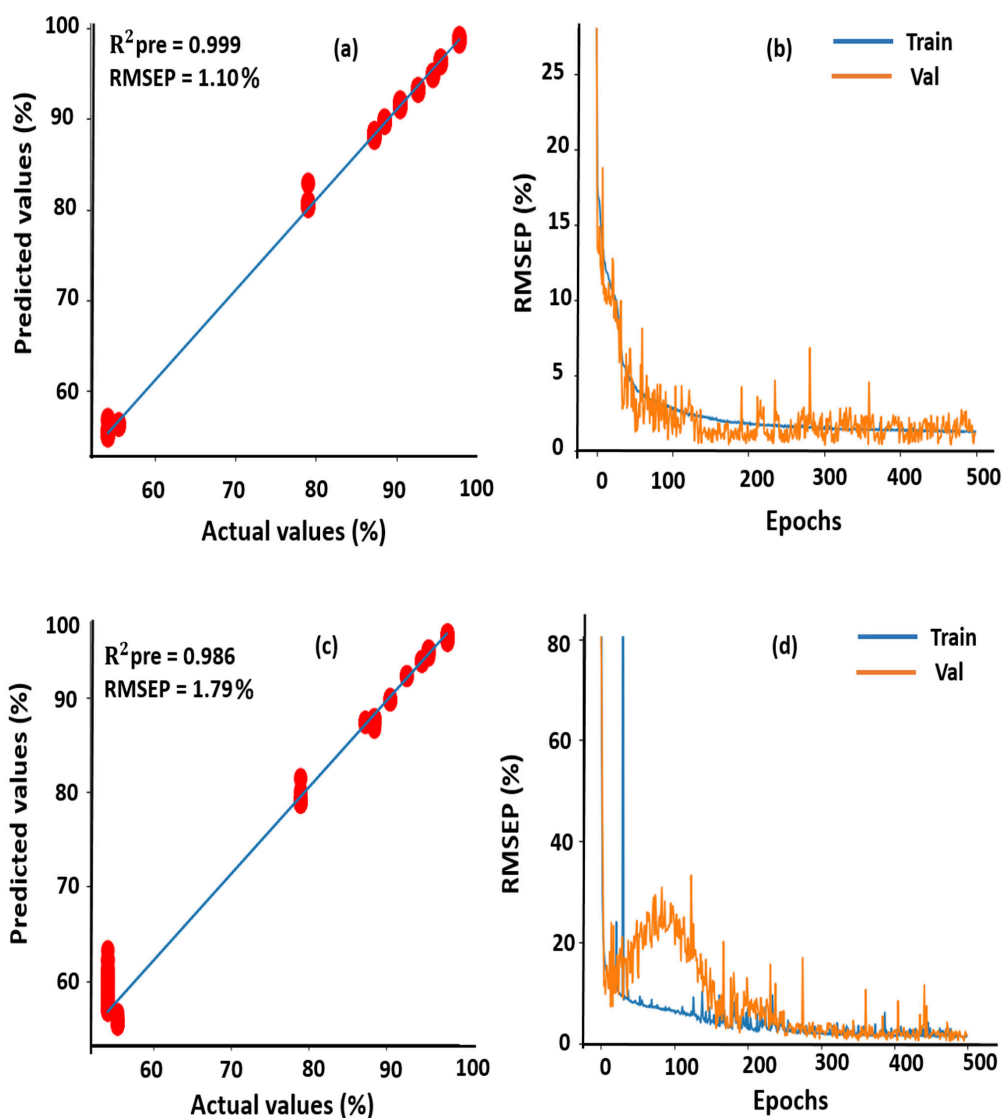


Fig. 11. 1D CNN (one-dimensional convolutional neural network) models prediction analysis results for terpenoids prediction in essential oils for (a, b) FT-IR (Fourier transform infrared), and (c, d) Raman spectroscopy. Here the figures (b, d) are loss function curves. R^2 pre, correlation coefficient for prediction; RMSEP, root mean squared error of prediction.

The two bar charts shown in Fig. 12a, b display the overall performance assessment results of the machine learning (PLSR, SVR) and 1 D CNN models for FT-IR and Raman spectroscopic data. The RMSEP values and correlation coefficient (R^2) variations for the prediction datasets are shown in both bar charts. In terms of R^2 and RMSEP values, the bar charts unambiguously demonstrate that support vector regression models outperformed both PLSR and 1D CNN models. Table 7 displays the outcomes of the prediction analysis for each of the three regression analysis models that were used on the prediction dataset produced by this study.

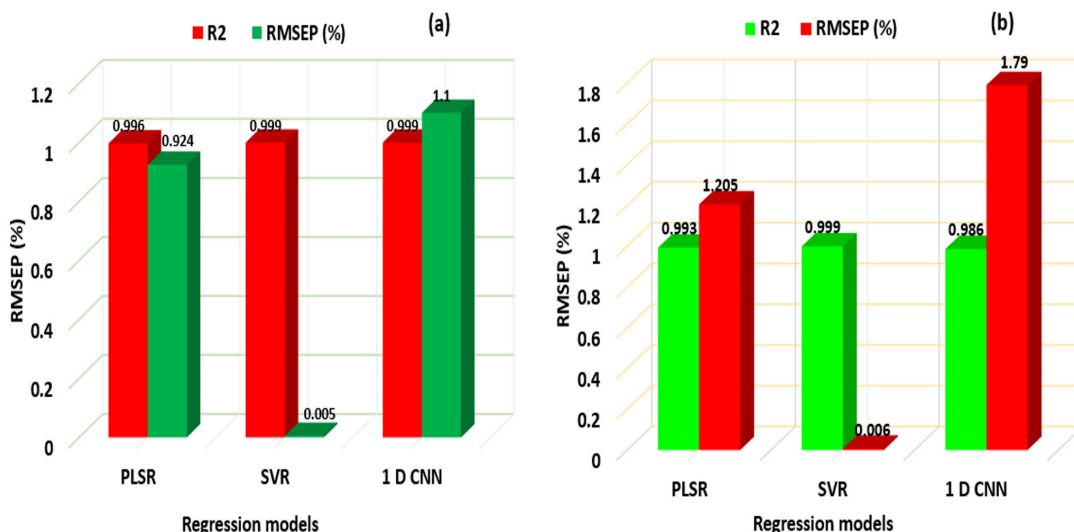


Fig. 12. Performance assessment of the PLSR (partial least squares regression), SVR (support vector regression), and 1D CNN (one-dimensional convolutional neural network) machine and deep learning models in terms of correlation coefficient and prediction error for (a) FT-IR (Fourier transform infrared), and (b) Raman spectroscopic data using bar-chart figures. R^2 , correlation coefficient for prediction; RMSEP, root mean squared error of prediction.

Table 7. Prediction analysis results for all three models using FT-IR and Raman spectroscopy.

Regression models	FT-IR spectroscopy		Raman spectroscopy	
	R^2	RMSEP (%)	R^2	RMSEP (%)
PLSR	0.996	0.924	0.993	1.205
SVR	0.999	0.005	0.999	0.006
1 D CNN	0.999	1.10	0.986	1.79

FT-IR, Fourier transform infrared; R^2 , correlation coefficient for prediction; RMSEP, root mean square of prediction; PLSR, partial least squares regression; SVR, support vector regression; 1D CNN, one-dimensional convolutional neural network.

The acquired findings presented in Table 7, confirm the performance of the SVR model. Hence, it could be clearly suggested that FT-IR spectroscopy, when combined with the support vector regression model is very effective for the non-destructive examination of terpenoids in essential oils, requiring less sample preparation and achieving faster evaluation. In earlier research, Rodríguez-Maecker et al. (2017) used a static headspace gas chromatography-ion mobility spectrometry method to identify terpenes with an R^2 correlation coefficient of 0.999. The measurement of terpenes in essential oils obtained from conifers was also carried out by Allenspach et al. (2020) utilizing gas chromatography in conjunction with a flame ionization detector with a higher LOD, yielding a strong R^2 value of 0.999. Due to their sophisticated sample preparation processes, the practical use of these approaches in the real world is laborious and time-consuming. However, because this study analyses more samples while concurrently creating machine learning and deep learning models, the sample preparation

process is quick and easy. Consequently, the results unambiguously demonstrate that support vector regression (SVR) in combination with ATR-based FT-IR spectroscopy is particularly capable of the quick and nondestructive evaluation of terpenoids in various kinds of medicinal plants essential oils. Because the spectrum data includes multiple variables that are nonetheless highly significant and connected to functional groups found in chemical entities. However, the inclusion of these variables causes the multivariate analysis computing time to increase. Future works will expand on this research by using band selection techniques like variable importance in projection (VIP), successive projection algorithm (SPA), and genetic algorithms (GA) to remove undesirable variables and enhance performance by speeding up data processing. Additionally, additional samples of plant essential oils in various combinations with other phytochemicals will be examined to see how well our established method works for non-destructive quantitative evaluation.

Conclusion

In this study, two different vibrational spectroscopy methods—ATR-FT-IR and Raman spectroscopy were used to evaluate terpenes composition in medicinal plant essential oils for comparison against gas chromatography coupled with a flame ionization detector (GC-FID), a standard analysis technique. Both spectroscopic methods produced interesting findings, however FT-IR spectroscopy outperformed Raman spectroscopy. The non-destructive quantitative prediction of terpenoids, was performed using PLSR, SVR, and 1D CNN models. The SVR model outperformed other models using FT-IR spectroscopy and acquired the highest correlation coefficient of 0.999 with an RMSEP error of 0.006%, which is somewhat better than the Raman data. On the other hand, the qualitative separation of essential oils from ten distinct geographic locations was conducted using the support vector classification method, which obtained the greatest accuracy of 0.986 and 0.997% utilizing FT-IR and Raman spectroscopic data. The acquired R^2 and RMSEP values demonstrate the potential of our developed model. Based on these results, it can be concluded that ATR-FT-IR spectroscopy in conjunction with the SVR machine learning technique offers a strong capability for the rapid, exact, and nondestructive identification of terpenoids presents in different varieties of essential oils. In contrast to more established destructive techniques such as HPLC, and GC-MS, which have low detection limits, the developed method shows promise for the quick, chemical-free, and effective detection of product adulteration because it does not require a sophisticated lab or trained personnel to conduct analyses.

Conflict of Interests

No potential conflict of interest relevant to this article was reported.

Acknowledgements

This work was supported by Korea Institute of Planning and Evaluation for Technology in Food, Agriculture and Forestry (IPET) through Smart Farm Innovation Technology Development Program, funded by Ministry of Agriculture, Food and Rural Affairs (MAFRA) (421005-04). Further, we also give deep thanks for Department of Chemistry, M.B.G.P.G College Haldwani, Kumaun University Nainital-263139, Uttarakhand, India for their collaboration during plants sample collection and their essential oils extraction and later terpenoids determination using GC-FID chemical analysis methods.

Authors Information

Rahul Joshi, <https://orcid.org/0000-0002-5834-2893>

Sushma Kholiya, <https://orcid.org/0009-0007-4066-0539>

Himanshu Pandey, <https://orcid.org/0000-0001-5419-8903>

Ritu Joshi, Texas A&M University

Omia Emmanuel, <https://orcid.org/0000-0001-8105-3304>

Ameeta Tewari, <https://orcid.org/0009-0007-5571-5635>

Taehyun Kim, National Institute of Agricultural Science, Rural Development Administration

Byoung-Kwan Cho, <https://orcid.org/0000-0002-8397-9853>

References

- Allenspach MD, Valder C, Steuer C. 2020. Absolute quantification of terpenes in conifer-derived essential oils and their antibacterial activity. *Journal of Analytical Science and Technology* 11:12. DOI:10.1186/s40543-020-00212-y.
- Baranska M, Schulz H, Krüger H, Quilitzsch R. 2005. Chemotaxonomy of aromatic plants of the genus *Origanum* via vibrational spectroscopy. *Analytical and Bioanalytical Chemistry* 381:1241-1247. DOI:10.1007/s00216-004-3018-y.
- Chandra N, Singh G, Lingwal S, Bisht MPS, Tewari LM, Joshi VC. 2022. Ecological status of alpine medicinal and aromatic plants of western Himalaya. *Journal of Herbs, Spices and Medicinal Plants* 28:73-88. DOI:10.1080/10496475.2021.1963903.
- Chatzidakis M, Botton GA. 2019. Towards calibration-invariant spectroscopy using deep learning. *Scientific Reports* 9:2126. DOI:10.1038/s41598-019-38482-1.
- Chowdhry BZ, Ryall JP, Dines TJ, Mendham AP. 2015. Infrared and Raman spectroscopy of eugenol, isoeugenol, and methyl eugenol: Conformational analysis and vibrational assignments from density functional theory calculations of the anharmonic fundamentals. *Journal of Physical Chemistry A* 119:11280-11292. DOI:10.1021/acs.jpca.5b07607.
- Clevenger JF. 1928. Apparatus for the determination of volatile oil. *The Journal of the American Pharmaceutical Association* (1912) 17:345-349.
- Cortes C, Vapnik V. 1995. Support-vector networks. *Machine Learning* 20:273-297.
- Daferera DJ, Tarantilis PA, Polissiou MG. 2002. Characterization of essential oils from Lamiaceae species by Fourier transform Raman spectroscopy. *Journal of Agricultural and Food Chemistry* 50:5503-5507. DOI:10.1021/jf0203489.
- de Barros Fernandes RV, Borges SV, Botrel DA, de Oliveira CR. 2014. Physical and chemical properties of encapsulated rosemary essential oil by spray drying using whey protein-inulin blends as carriers. *International Journal of Food Science and Technology* 49:1522-1529. DOI:10.1111/ijfs.12449.
- Divyanth LG, Chakraborty S, Li B, Weindorf DC, Deb P, Gem CJ. 2022. Non-destructive prediction of nicotine content in tobacco using hyperspectral image-derived spectra and machine learning. *Journal of Biosystems Engineering* 47:106-117. DOI:10.1007/s42853-022-00134-0.
- Duvenaud D. 2014. The kernel cookbook: Advice on covariance functions. Accessed in <https://www.cs.toronto.edu/~duvenaud/cookbook/> on 13 August 2023.
- Elzey B, Pollard D, Fakayode SO. 2016. Determination of adulterated neem and flaxseed oil compositions by FTIR spectroscopy and multivariate regression analysis. *Food Control* 68:303-309.
- Fuentes AM, Narayan A, Milligan K, Lum JJ, Brolo AG, Andrews JL, Jirasek A. 2023. Raman spectroscopy and convolutional neural networks for monitoring biochemical radiation response in breast tumour xenografts. *Scientific Reports* 13:1530. DOI:10.1038/s41598-023-28479-2.

- García C, Montero G, Coronado MA, Valdez B, Stoytcheva M, Rosas N, Torres R, Sagaste CA. 2017. Valorization of eucalyptus leaves by essential oil extraction as an added value product in Mexico. *Waste and Biomass Valorization* 8:1187-1197. DOI:10.1007/s12649-016-9695-x.
- Greathead H. 2003. Plants and plant extracts for improving animal productivity. *Proceedings of the Nutrition Society* 62:279-290.
- Gutteridge JMC, Halliwell B. 2010. Antioxidants: Molecules, medicines, and myths. *Biochemical and Biophysical Research Communications* 393:561-564.
- Hanif MA, Nawaz H, Naz S, Mukhtar R, Rashid N, Bhatti IA, Saleem M. 2017. Raman spectroscopy for the characterization of different fractions of hemp essential oil extracted at 130°C using steam distillation method. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy* 182:168-174. DOI:10.1016/j.saa.2017.03.072.
- Huang X, Li H, Ruan Y, Li Z, Yang H, Xie G, Yang Y, Du Q, Ji K, Yang M. 2022. An integrated approach utilizing Raman spectroscopy and chemometrics for authentication and detection of adulteration of agarwood essential oils. *Frontiers in Chemistry* 10:1036082. DOI:10.3389/fchem.2022.1036082.
- Irshad M, Ali Subhani MA, Ali S, Hussain A. 2020. Biological importance of essential oils. *Essential Oils-Oils of Nature*. IntechOpen Limited, London, UK. DOI:10.5772/intechopen.87198.
- Jentzsch PV, Ciobotă V. 2014. Raman spectroscopy as an analytical tool for analysis of vegetable and essential oils. *Flavour and Fragrance Journal* 29:287-295. DOI:10.1002/ffj.3203.
- Jindasa MHW, Kahawalage AC, Halstensen M, Skeie NO, Jens KJ. 2021. Deep learning approach for Raman spectroscopy. *Recent Developments in Atomic Force Microscopy and Raman Spectroscopy for Materials Characterization*. IntechOpen Limited, London, UK. DOI:10.5772/intechopen.99770.
- Joshi R, Adhikari S, Son JP, Jang Y, Lee D, Cho BK. 2023a. Au nanogap SERS substrate for the carbaryl pesticide determination in juice and milk using chemometrics. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy* 297:122734. DOI:10.1016/j.saa.2023.122734.
- Joshi R, Baek I, Joshi R, Kim MS, Cho BK. 2022. Detection of fabricated eggs using Fourier transform infrared (FT-IR) spectroscopy coupled with multivariate classification techniques. *Infrared Physics and Technology* 123:104163. DOI:10.1016/j.infrared.2022.104163.
- Joshi R, Joshi R, Amanah HZ, Faqeerzada MA, Jayapal PK, Kim G, Baek I, Park ES, Masithoh RE, Cho BK. 2021. Quantitative analysis of glycerol concentration in red wine using Fourier transform infrared spectroscopy and chemometrics analysis. *Korean Journal of Agricultural Science* 48:299-310. DOI:10.7744/kjoas.20210023.
- Joshi R, Joshi R, Mo C, Faqeerzada MA, Amanah HZ, Masithoh RE, Kim MS, Cho BK. 2020. Raman spectral analysis for quality determination of grignard reagent. *Applied Sciences* 10:3545. DOI:10.3390/app10103545.
- Joshi R, Lakshmi Priya GG, Faqeerzada MA, Bhattacharya T, Kim MS, Baek I, Cho BK. 2023b. Deep learning-based quantitative assessment of melamine and cyanuric acid in pet food using Fourier transform infrared spectroscopy. *Sensors* 23:5020. DOI:10.3390/s23115020.
- Jung DH, Kim NY, Moon SH, Kim HS, Lee TS, Yang JS, Lee JY, Han X, Park SH. 2021. Classification of vocalization recordings of laying hens and cattle using convolutional neural network models. *Journal of Biosystems Engineering* 46:217-224. DOI:10.1007/s42853-021-00101-1.
- Kawamura K, Nishigaki T, Andriamananjara A, Rakotonindrina H, Tsujimoto Y, Moritsuka N, Rabenarivo M, Razafimbelo T. 2021. Using a one-dimensional convolutional neural network on visible and near-infrared spectroscopy to improve soil phosphorus prediction in Madagascar. *Remote Sensing* 13:1519. DOI:10.3390/rs13081519.
- Kholiya S, Pandey H, Pargaian N, Tiwari A. 2022. A survey of some medicinal trees in Nandhaur valley region of district Nainital in Uttarakhand, India, and study on their phytochemistry and ethnobotanical importance. *Academia Journal of Medicinal Plants* 10:149-164.
- Kim G, Lee H, Baek I, Cho BK, Kim MS. 2022. Short-wave infrared hyperspectral imaging system for nondestructive evaluation of powdered food. *Journal of Biosystems Engineering* 47:223-232. DOI:10.1007/s42853-022-00141-1.
- Kumar A, Jnanasha AC. 2016. Medicinal and aromatic plants biodiversity in India and their future prospects: A review. *Indian Journal Unani Medicine* 9:10-17.

- Li P, Ma J, Zhong N. 2021. Raman spectroscopy combined with support vector regression and variable selection method for accurately predicting salmon fillets storage time. *Optik* 247:167879.
- Lohumi S, Joshi R, Kandpal LM, Lee H, Kim MS, Cho H, Mo C, Seo YW, Rahman A, Cho BK. 2017. Quantitative analysis of Sudan dye adulteration in paprika powder using FTIR spectroscopy. *Food Additives & Contaminants: Part A* 34:678-686. DOI:10.1080/19440049.2017.1290828.
- Lotfollahi M, Berisha S, Daeinejad D, Mayerich D. 2019. Digital staining of high-definition FTIR images using deep learning. *Applied Spectroscopy* 73:556-564. DOI:10.1177/0003702818819857.
- Luo R, Popp J, Bocklitz T. 2022. Deep learning for Raman spectroscopy: A review. *Analytica* 3:287-301. DOI:10.3390/analytica3030020.
- Masruri, Rahman MF, Ramadhan BN. 2016. Acidity-controlled selective oxidation of alpha-pinene, isolated from Indonesian pine's turpentine oils (*pinus merkusii*). In *Proceeding of IOP Conference Series: Materials Science and Engineering* 107:012060. DOI:10.1088/1757-899X/107/1/012060.
- Medvecká V, Mošovská S, Mikulajová A, Valík L, Zahoranová A. 2020. Cold atmospheric pressure plasma decontamination of allspice berries and effect on qualitative characteristics. *European Food Research and Technology* 246:2215-2223. DOI:10.1007/s00217-020-03566-0.
- Meng J, Chen X, Yang W, Song J, Zhang Y, Li Z, Yang X, Yang Z. 2014. Gas chromatography-mass spectrometry analysis of essential oils from five parts of Chaihu (*Radix Bupleuri Chinensis*). *Journal of Traditional Chinese Medicine* 34:741-748. DOI:10.1016/s0254-6272(15)30090-x.
- Michel J, Abd Rani NZ, Husain K. 2020. A review on the potential use of medicinal plants from asteraceae and lamiaceae plant family in cardiovascular diseases. *Frontiers in Pharmacology* 11:1-26. DOI:10.3389/fphar.2020.00852.
- Ndhlala AR, Moyo M, Van Staden J. 2010. Natural antioxidants: Fascinating or mythical biomolecules? *Molecules* 15:6905-6930.
- Patel AK, Ghosh JK, Pande S, Sayyad SU. 2020. Deep-learning-based approach for estimation of fractional abundance of nitrogen in soil from hyperspectral data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 13:6495-6511. DOI:10.1109/JSTARS.2020.3039844.
- Porel A, Sanyal Y, Kundu A. 2014. Simultaneous HPLC determination of 22 components of essential oils; Method robustness with experimental design. *Indian Journal of Pharmaceutical Sciences* 76:19-30.
- Putra BTW, Amirudin R, Marhaenanto B. 2022. The evaluation of deep learning using convolutional neural network (CNN) approach for identifying arabica and robusta coffee plants. *Journal of Biosystems Engineering* 47:118-129. DOI:10.1007/s42853-022-00136-y.
- Rinnan Å, van den Berg F, Engelsen SB. 2009. Review of the most common pre-processing techniques for near-infrared spectra. *TrAC Trends in Analytical Chemistry* 28:1201-1222. DOI:10.1016/j.trac.2009.07.007.
- Riyanto, Sastrohamidjojo H, Fariyatun E. 2016. Synthesis of methyl eugenol from crude cloves leaf oil using acid and based chemicals reactions. *IOSR Journal of Applied Chemistry* 9:105-112.
- Rodríguez-Maecker R, Vyhmeister E, Meisen S, Martinez AR, Kuklya A, Telgheder U. 2017. Identification of terpenes and essential oils by means of static headspace gas chromatography-ion mobility spectrometry. *Analytical and Bioanalytical Chemistry* 409:6595-6603. DOI:10.1007/s00216-017-0613-2.
- Samarth RM, Samarth M, Matsumoto Y. 2017. Medicinally important aromatic plants with radioprotective activity. *Future Science OA* 3: FSO247. DOI:10.4155/fsoa-2017-0061.
- Seo Y, Mo C, Lim J, Lee A, Kim B, Jang J, Kim G. 2021. Detection of spinach juice residues on stainless steel surfaces using VNIR hyperspectral images. *Journal of Biosystems Engineering* 46:173-181. DOI:10.1007/s42853-021-00097-8.
- Siaty NG, Kimbaris AC, Pappas CS, Tarantilis PA, Daferera DJ, Polissiou MG. 2005. Rapid method for simultaneous quantitative determination of four major essential oil components from oregano (*Oreganum* sp.) and thyme (*Thymus* sp.) using FT-Raman spectroscopy. *Journal of Agricultural and Food Chemistry* 53:202-206. DOI:10.1021/jf048930f.

- Sihalath T, Basak JK, Bhujel A, Arulmozhi E, Moon BE, Kim HT. 2021. Pig identification using deep convolutional neural network based on different age range. *Journal of Biosystems Engineering* 46:182-195. DOI:10.1007/s42853-021-00098-7.
- Sufriadi E, Meilina H, Munawar AA, Idroes R. 2021. Fourier transformed infrared (FTIR) spectroscopy analysis of patchouli essential oils based on different geographical area in Aceh. In *Proceeding of IOP Conference Series: Materials Science and Engineering* 1087:012067. DOI:10.1088/1757-899X/1087/1/012067.
- Valderrama ACS, Rojas De GC. 2017. Traceability of active compounds of essential oils in antimicrobial food packaging using a chemometric method by ATR-FTIR. *American Journal of Analytical Chemistry* 8:726-741. DOI:10.4236/ajac.2017.811053.
- Wang M, Zhao M, Chen J, Rahardja S. 2019. Nonlinear unmixing of hyperspectral data via deep autoencoder networks. *IEEE Geoscience and Remote Sensing Letters* 16:1467-1471. DOI:10.1109/lgrs.2019.2900733.
- Wold S, Sjöström M, Eriksson L. 2001. PLS-regression: A basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems* 58:109-130. DOI:10.1016/S0169-7439(01)00155-1.
- Wu X, Xu B, Ma R, Niu Y, Gao S, Liu H, Zhang Y. 2022. Identification and quantification of adulterated honey by Raman spectroscopy combined with convolutional neural network and chemometrics. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy* 274:121133. DOI:10.1016/j.saa.2022.121133.