



## Learning algorithms for identification of whisky using portable Raman spectroscopy

Kwang Jun Lee<sup>a,b,d,1</sup>, Alexander C. Trowbridge<sup>a,b,d,1</sup>, Graham D. Bruce<sup>c</sup>, George O. Dwapanyin<sup>c</sup>, Kylie R. Dunning<sup>d,e</sup>, Kishan Dholakia<sup>a,c,d,\*,2</sup>, Erik P. Schartner<sup>a,b,e,2</sup>

<sup>a</sup> Centre of Light for Life (CLL) and Institute for Photonics and Advanced Sensing (IPAS), The University of Adelaide, Adelaide, 5005, SA, Australia

<sup>b</sup> School of Physics, Chemistry and Earth Sciences, The University of Adelaide, Adelaide, 5005, SA, Australia

<sup>c</sup> SUPA School of Physics and Astronomy, University of St Andrews, St Andrews, KY16 9SS, Fife, United Kingdom

<sup>d</sup> School of Biological Sciences, The University of Adelaide, Adelaide, 5005, SA, Australia

<sup>e</sup> Robinson Research Institute, School of Biomedicine, The University of Adelaide, Adelaide, 5005, SA, Australia

### ARTICLE INFO

#### Keywords:

Raman spectroscopy  
Machine learning  
Whisky  
Brand identification

### ABSTRACT

Reliable identification of high-value products such as whisky is vital due to rising issues of brand substitution and quality control in the industry. We have developed a novel framework that can perform whisky analysis directly from raw spectral data with no human intervention by integrating machine learning models with a portable Raman device. We demonstrate that machine learning models can achieve over 99% accuracy in brand or product identification across twenty-eight commercial samples. To demonstrate the flexibility of this approach, we utilized the same algorithms to quantify ethanol concentrations, as well as measuring methanol levels in spiked whisky samples. To demonstrate the potential use of these algorithms in a real-world environment we tested our algorithms on spectral measurements performed through the original whisky bottle. Through the bottle measurements are facilitated by a beam geometry hitherto not applied to whisky brand identification in conjunction with machine learning. Removing the need for decanting greatly enhances the practicality and commercial potential of this technique, enabling its use in detecting counterfeit or adulterated spirits and other high-value liquids. The techniques established in this paper aim to function as a rapid and non-destructive initial screening mechanism for detecting falsified and tampered spirits, complementing more comprehensive and stringent analytical methods.

### 1. Introduction

Recognizing a brand is important for the global growth of the whisky market as consumers are showing an increasing demand for exclusive, high quality products. As demand increases, concerns arise about counterfeit and adulterated products being sold, which violate laws related to alcohol labeling and fraud (Chaudhry et al., 2009; Green and Smith, 2002). Misrepresentation of lower quality commercial whiskies as premium products can harm a producer's reputation and financial performance. In 2018, a third of commercial Scotch whiskies tested were found to be fraudulent (The British Broadcasting Corporation, 2018; The Times, 2018; The Guardian, 2018). More broadly, the European Union

loses €3 billion annually in sales due to fake wine, beer, and spirits (Fraud Advisory Panel, 2022). Counterfeit spirits made with industrial alcohols or poor distillation may have high levels of methanol, causing serious illness. In 2019, toxic 'moonshine' killed 154 people in India (CNN, 2019), while in March 2020, Iranian media reported that nearly 300 people died, and over 1000 became ill from drinking methanol-laced bootlegged spirits (Tech Times, 2020).

A range of analytical techniques, such as mass spectrometry, nuclear magnetic resonance spectroscopy, gas chromatography, or liquid chromatography are used in laboratories to guarantee the quality, safety, and authenticity of spirits (Power et al., 2020). These techniques typically necessitate whisky sample preparation, which can be costly and

\* Corresponding author. Centre of Light for Life (CLL) and Institute for Photonics and Advanced Sensing (IPAS), The University of Adelaide, Adelaide, 5005, SA, Australia.

E-mail address: [kishan.dholakia@adelaide.edu.au](mailto:kishan.dholakia@adelaide.edu.au) (K. Dholakia).

<sup>1</sup> These authors contributed equally to this work and share first authorship.

<sup>2</sup> These authors contributed equally to this work and share last authorship.

<https://doi.org/10.1016/j.crfs.2024.100729>

Received 18 January 2024; Received in revised form 13 March 2024; Accepted 30 March 2024

Available online 1 April 2024

2665-9271/© 2024 Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

time-consuming. Importantly, these methods also require the sample to be removed for analysis, restricting their use in online monitoring and requiring end-products to be opened for analysis. This limits their widespread application. Portable sensors and methods for analyzing suspicious products at the point of sale or distribution are essential for widespread fraud prevention (Oliveira et al., 2018; Limm et al., 2018; Soon and Manning, 2019). In this paper, a new compact Raman technique which employs machine learning is presented to fulfill this burgeoning need.

Raman spectroscopy is a non-destructive analytical technique that uses laser light to interrogate sample and measure the inelastically scattered light from the constituent molecules. As a result, this technique yields a form of optical fingerprint with Raman features that correlate with the molecular species present in the sample. This has broad applicability and can offer both quantitative and qualitative analysis of whisky samples (Flack, 2022; Nordon et al., 2005). Raman has been demonstrated to be a very versatile technique, with applications including pharmaceutical analysis, explosives detection and biological sensing among many others (Vankeirsbilck et al., 2002; Moore and Scharff, 2009; Izake, 2010). Raman spectroscopy has previously been demonstrated on decanted samples in the determination total sugar content, pH levels, grape varieties, geographic origin, and aging duration (Dos Santos et al., 2018; Mandrile et al., 2016). This technique is not only capable of classifying whisky brands (Fleming et al., 2020; Ashok et al., 2011), but it also has the potential to estimate the concentrations of methanol (indicative of toxicity) and ethanol (a measure of quality) in alcoholic beverages (Ashok et al., 2013). Portable Raman spectroscopy allows for in-field analysis of whisky samples, which can be useful for monitoring the authenticity and quality of the product during production and distribution (Ellis et al., 2017, 2019; Kiefer and Cromwell, 2017). Importantly, techniques that can measure the Raman spectra of the contents of a bottle without opening it are also being developed (Fleming et al., 2020; Shillito et al., 2022).

The large amount of information from Raman spectroscopy may be efficiently handled by the use of machine learning models in the data analysis to extract valuable and subtle insights which can then be leveraged to make predictions. Standard statistical methods may find it challenging to fit subtle trends in complex, information rich data (Bzdok et al., 2018). Statistical methods applying Raman spectroscopy have been already published (Lednev and Sikirzhyski, 2012; Wang et al., 2020), suggesting the possibility to discriminate and analyze whiskies. However, a systematic comparison of different machine learning methods has not been performed for the application of Raman analysis to the brand identification of a variety of whiskies.

Furthermore, recent advancements in machine learning models have provided exciting new avenues for spectral data analysis in materials science (Lussier et al., 2020). Machine learning algorithms can analyze features and correlations within spectra, leading to various applications in Raman spectral analysis. Previous studies have employed various machine learning algorithms such as support vector machines (SVM),  $k$ -nearest neighbors (KNN), and random forest (RF) (Jimenez-Carvelo et al., 2017; Khan et al., 2017). These algorithms have demonstrated high prediction accuracy, but they have limitations such as poor model flexibility and generalization to new datasets, which can limit their applications in more challenging scenarios (Kotsiantis, 2007; Kotsiantis et al., 2006; Singh et al., 2016). In particular, previous studies using these methods have required manual preprocessing to enhance the performance of machine learning in spectral data analysis. This is a time consuming step which must be performed for each and every application (Singh et al., 2021). Furthermore, the improper usage of preprocessing methods may lead to errors and loss of information, thereby negatively affecting the accuracy of the results and making the analysis process more complex (Liland et al., 2010). Overcoming this barrier is a key aspect of the present work.

Deep learning is a subset of machine learning that is designed to handle large amounts of complex data, and can be employed to

automatically extract complex features and relationships between features and tasks. Previously it has been demonstrated that deep learning can outperform conventional machine learning methods in a range of challenging problems (Sarker, 2021), however this is typically at the cost of higher computational resource and data requirements compared to conventional machine learning (Janiesch et al., 2021).

Convolutional neural networks (CNN) are a variant of deep learning models, and have been successfully applied in Raman spectroscopy for the component identification of complex mixture materials (Fan et al., 2019; Pan et al., 2021). Deep learning networks offer an advantage as they do not require manual tuning and can be trained as end-to-end networks that handle both feature extraction and classification or regression. This reduces the need for separate preprocessing or feature engineering steps, as the network can automatically convert the features into a more advanced representation (LeCun et al., 2015). As a result, deep learning networks can accommodate variations in samples that were previously unknown. Conventional machine learning methods, on the other hand, that require a more rigid protocol, may be unable to handle unseen data and result in inaccurate measurements (Taye, 2023). Therefore, in situations where the samples are not well-characterized or have a high degree of variability, more flexible analysis methods, such as deep learning, may be preferred. However, it should be noted that skilled preprocessing of the input data can still greatly assist the success of such models depending on the subject matter and dataset (Chollet, 2017).

In this paper, we focus on methods that obviate the need for spectral preprocessing, such as smoothing, baseline correction, normalization, and spectral windowing. It can be difficult to reproduce the same results if these subjective processing steps are not well documented. In addition, bias may be introduced into the data unknowingly. Here we used principal component analysis (PCA) rather than the extensive data processing for our analysis of the spectral data. The aim is to conserve analytical resources and eliminate subjective steps from the procedure. This choice was made because PCA is an unsupervised method that can be relatively robust and less resource-intensive compared to manual preprocessing.

In this study, we investigated whisky brand identification and determination of both the ethanol and methanol concentrations of whisky samples. This study offers a rapid (real-time), non-invasive technique for whisky brand identification. The approach leverages portable Raman spectroscopy, including the use of a novel through bottle geometry, and machine learning algorithms to achieve high accuracy without the need for manual spectral preprocessing. Deep learning results were compared to conventional techniques. The machine learning methods used in this study are summarized in Fig. S1. We achieved over 99% accuracy in identifying the brands of whisky samples, and the measurement of ethanol and methanol levels were within 2.47% vol/vol (v/v) and 0.05% v/v of the actual value, respectively. Finally, we utilized machine learning techniques to analyze the Raman spectra of the sample within an unopened glass bottle without removing the contents. Raman spectra are collected through the bottle using beam shaping to suppress the bottle contribution. It is noteworthy that this has previously not been applied in the context of whisky brand identification using machine learning. The accuracy of predicting the brand remained over 99% even when the method was applied to datasets obtained through the bottle.

## 2. Material and methods

### 2.1. Sample preparation

A total of 28 commercially obtained whisky samples, and a reference sample of 40% ethanol in distilled water were used for classification and ethanol quantification, as summarized in Table S1. The whiskies were chosen to represent a variety of distilleries, flavours, cask types, and ages. The ethanol content ranges from 40% to 63% vol. To assess how

well the trained ethanol regression models could generalize, additional test samples of twenty whiskies and three gins were used. Two pure whiskies (Talisker and Cragganmore) and a sample of 40% ethanol/water were spiked with HPLC grade methanol. Methanol concentrations ranging from 0 to 3% in 0.3% increments were used to generate the training set for methanol quantification. Additional test samples with 0, 0.3, 1, and 2% methanol concentrations were prepared using Caol Ila and Cynelish via the same protocol and used to evaluate the generalization performance of the trained methanol regression models.

## 2.2. Raman analysis

### 2.2.1. Through-vial

Initial studies were performed on 2 mL samples of whiskies which were pipetted from bottles and placed into 4 mL Wheaton sample vials for interrogation. The vial is made from clear type 1 A borosilicate glass with a diameter of 15 mm, a height of 46 mm, and a thickness of 1 mm. All through-vial Raman spectra were collected using a compact Wasatch Photonics WP 785 Raman spectrometer (WP-785-R-SR-LMMFC-IC) using an integrated 785 nm laser and 25  $\mu\text{m}$  slit giving a resolution of 7  $\text{cm}^{-1}$ . The laser was coupled into a Raman probe, and the emission collected by the probe was transferred to the spectrometer. The probe was focused directly into a clear glass vial containing the whisky. All spectra were collected in the spectral range 270 – 2000  $\text{cm}^{-1}$  with a laser power of 450 mW, an integration time of 500 ms and an average of 5 scans. Each whisky sample had forty replicates collected, except for the methanol test samples which had twenty replicates collected. Continuous irradiation for 15 min resulted in a 3% reduction in the intensity of spectra, as can be seen in Fig. S3. During the measurement process, each sample was irradiated for about 1 min. This short exposure time had a negligible effect on photo-bleaching, as it only resulted in a 0.2% reduction in intensity. The supplementary material section S1 and Fig. S2 provide details on the experimental configuration and spectral characteristics of the Raman spectra from 28 whisky samples.

### 2.2.2. Through-bottle

Through the bottle Raman measurements were performed using a free-space system, consisting of a Spectra-Physics 3900s Ti:Sapphire tunable laser for excitation and an Andor Shamrock SR303i spectrometer for spectra collection. The experimental setup in this section was based on that previously demonstrated (Fleming et al., 2020; Shillito et al., 2022), and utilized an axicon-based focus-matched inverse spatially-offset Raman configuration. Two configurations were set up to allow switching between the axicon configuration and a conventional back-scattering (Gaussian beam profile) configuration through the use of flip mirrors. The system was aligned such that the two paths were collinear with the focal point optimized for maximum Raman signal collection as shown in Fig. S4. All spectra were collected in the spectral range 140 – 2700  $\text{cm}^{-1}$  with a laser power at the sample of 96 mW for the Gaussian beam and 105 mW for the Bessel, and an integration time of 5 s. Each whisky sample (see Table S2 for details) had 30 replicates collected. As for the through-vial analysis, a measurement was performed every second for 60 min to observe if the signal from the Whisky reduced over time due to photobleaching. Results here demonstrated that the signal remained within  $\pm 2\%$  of the initial peak value. The variation in intensity is notably larger when measuring through the bottle compared to the vial. This could be due to the inherent challenges of measuring through an uncontrolled medium like a commercial bottle, as opposed to a more controlled environment like a vial.

To thoroughly examine the performance of the machine learning methods, spectra were collected with this system both in the original (thick glass) bottles, as well as decanting the samples into thin-walled 4 mL vials similar to those used in the through-vial measurements.

## 2.3. Data processing

Deep learning networks were established using Python 3.9.15, sci-kit learn 1.1.3 and TensorFlow 2.11.0 on a computer equipped with an NVIDIA GeForce RTX 4090 GPU and an Intel Core i912900KS CPU. Conventional classification machine learning methods were performed using Python 3.11.3 and sci-kit learn 1.2.2 on a computer equipped with an NVIDIA GeForce GTX 1650 Ti mobile GPU and an AMD Ryzen 7 4800HS mobile CPU. PCR, PLSR, and ridge regression were processed using MATLAB R2021b software (MathWorks, Natick, USA) on a computer equipped with an NVIDIA GeForce GTX 1650 Ti mobile GPU and an AMD Ryzen 7 4800HS mobile CPU.

### 2.3.1. Deep learning models

Three different configurations of the deep learning model, including CNN, fully connected networks (FCN), and a hybrid parallel model (HPM), which is a combination of CNN and FCN, were applied to predict brand identification, ethanol concentration, and methanol concentration using the same analysis model. The spectral data were split into training (60%), validation (20%), and test (20%) sets for brand identification and ethanol quantification. The spectral data of Talisker, Cragganmore, and 40% ethanol/water spiked with methanol were divided into training (80%) and validation (20%) sets, while the spectra of Caol Ila and Cynelish spiked with methanol were used exclusively as a test set for methanol quantification as summarized in Table S3. The raw spectral data was either directly used as the input for deep learning models or the spectral data underwent PCA, and six PCA features were selected based on preliminary results summarized in Fig. S5. The performance of the ANN model, which is closely related to deep learning models, reached a maximum at six PCA features. In this case the artificial neural network (ANN) model has only a few layers and is hence not considered to be deep learning. As a result, six PCA features were determined to be the optimum value to use for deep learning. These selected features served as the input for the classification and regression process. The prediction accuracy was employed to evaluate the brand identification performance of different methods. The root mean square error (RMSE) and determination coefficient ( $R^2$ ) were used to evaluate the quality of quantification analysis. RMSE and  $R^2$  of the training, validation, and test sets are abbreviated as  $\text{RMSE}_T$  and  $R_T^2$ ,  $\text{RMSE}_V$  and  $R_V^2$ , and  $\text{RMSE}_P$  and  $R_P^2$ , respectively. For ethanol prediction, the RMSE and  $R^2$  values of completely new and unseen samples (e.g. different brands) were denoted as  $\text{RMSE}_U$  and  $R_U^2$ , respectively. The values of parameters for deep learning models are described in Table S4.

### 2.3.2. Conventional classification machine learning

The spectral data was divided into training (70%) and test (30%) sets for conventional classification machine learning. In this study, the unprocessed spectral data was either directly used as input for conventional machine learning models, or subjected to PCA which was used to reduce the dimensionality of the data and also provided an initial evaluation of data predictability (Shlens, 2014).

The raw spectral data was either directly used as the input for conventional machine learning models or the spectral data underwent PCA, and from one to nine PCA features were used as input for classification. PCA is an unsupervised learning method, which means that it does not require data that has been assigned a class or category to train the model. The lack of supervision allows the method to be applied to any dataset, as opposed to requiring a new method to be optimized each time (Shlens, 2014). In order to dramatically reduce the search space and minimize the associated computational time for each technique, a robust PCA model was constructed. This enables the retention of the most useful information in the data while efficiently discarding the excess. The parameter values for conventional classification machine learning are described in Table S5.

### 2.3.3. Conventional regression machine learning for methanol quantification

Three regression methods were evaluated: principal component regression (PCR), partial least squares regression (PLSR), and ridge regression. Table S3 describes how the experimental set was divided. As a training set, spectral data from Talisker, Cragganmore, and 40% ethanol/water spiked with methanol were used, and cross-validation was performed using venetian blinds with five crossvalidation groups. The spectra of Caol Ila and Cynelish spiked with methanol were used as a test set. The values of RMSE and  $R^2$  were used to evaluate the quality of quantification analysis.

## 3. Results

### 3.1. Deep learning

In this section, we evaluate the performance of different deep learning approaches in identifying brands and quantifying ethanol/methanol. The deep learning models that were used in this study are described in Fig. S6 and the supplementary material section S2.

#### 3.1.1. Brand identification

We compared three different deep learning methods: CNN, FCN, and HPM. PCA was also combined with each of these methods (resulting in PCA + CNN/FCN/HPM). We trained each method with different numbers of epochs (100/200/500/1000/2000). The accuracy results depending on epochs are summarized in Fig. 1 and Table S6. Increasing the number of epochs used for training generally had a positive impact on accuracy, with many models achieving higher accuracy when the number of epochs was increased from 200 to 2000 (Fig. 1a). All models converged to a high accuracy on the training set before oscillating depending on the data in the set.

All models with more than 500 epochs demonstrated greater than 93% test accuracy, except for FCN without PCA (see blue dots in Fig. 1b). It is important to clarify that the accuracies mentioned here refer to the test set, not the training set. The use of PCA was observed to shorten the accuracy plateau time for the deep learning models (Fig. 1b). This suggests that reducing the dimensionality of the data using PCA before training could be an effective approach for improving the performance of deep learning models. Overall, the PCA + HPM combination outperformed all other deep learning models on this dataset, with training times of 89 s, and over 97% accuracy on all datasets when 200 or more epochs were used (Fig. 1b). This is likely due to the simpler FCN branch providing a shorter path to the classifier layers, assisting in training the deeper layers of the HPM model. Even when trained for only 100 epochs, this model achieved a test accuracy of 96% in just 46 s of training time.

#### 3.1.2. Chemical regression

The same deep learning algorithms and data sets used for brand identification were used to predict ethanol content in whisky samples. The results are summarized in Fig. 2a and b, and represented in more detail in Table S7. Without using PCA, the CNN method outperformed other models, achieving an  $R_p^2$  score of 0.994 and an  $RMSEP$  of 0.39% for ethanol content in the sample, as shown in Fig. 2a. The HPM method achieved competitive results compared to CNN, with an  $R_p^2$  score of 0.993 and  $RMSEP$  of 0.43%. All models with PCA performed significantly better than models without PCA. PCA + CNN and PCA + HPM displayed the best performance with a  $R_p^2$  scores of 0.998 and  $RMSEP$  of 0.24% and 0.25%, respectively.

The deep learning models were tested on samples that were not used during training to evaluate their ability to generalize. The unseen test set included twenty whisky and three gin samples. The best results were achieved using the PCA + FCN model when trained for 200 epochs, with an  $R_U^2$  of 0.863 and  $RMSE_U$  of 2.47% (Fig. 2b). The performance of ethanol content prediction on the unseen test set is shown in Table S7 (see values of  $R_U^2$  and  $RMSE_U$ ). This suggests that the simpler FCN

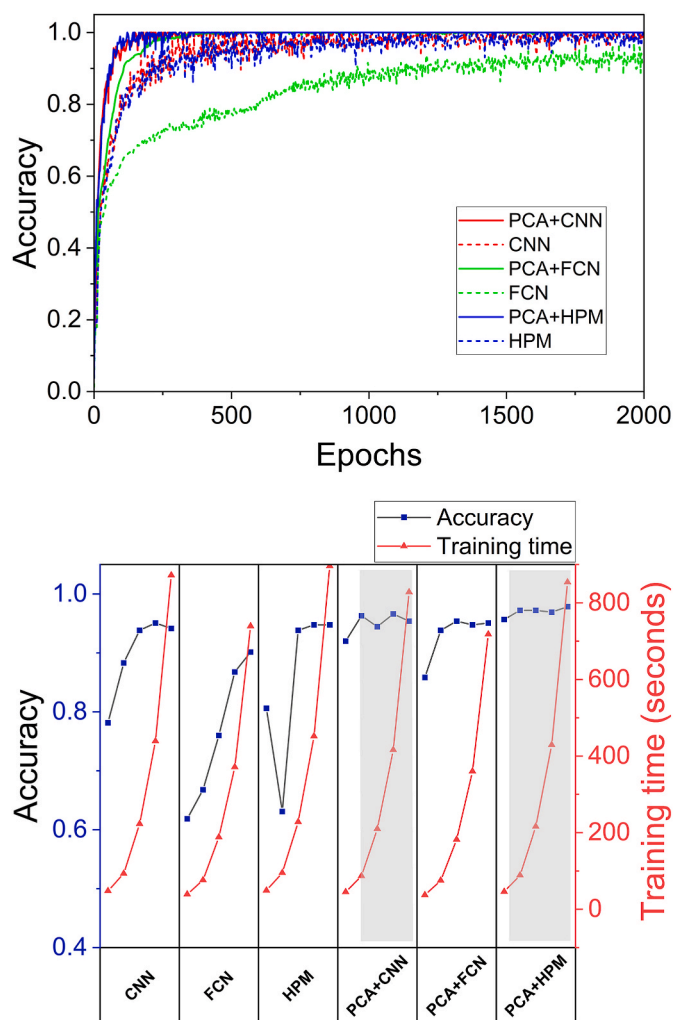
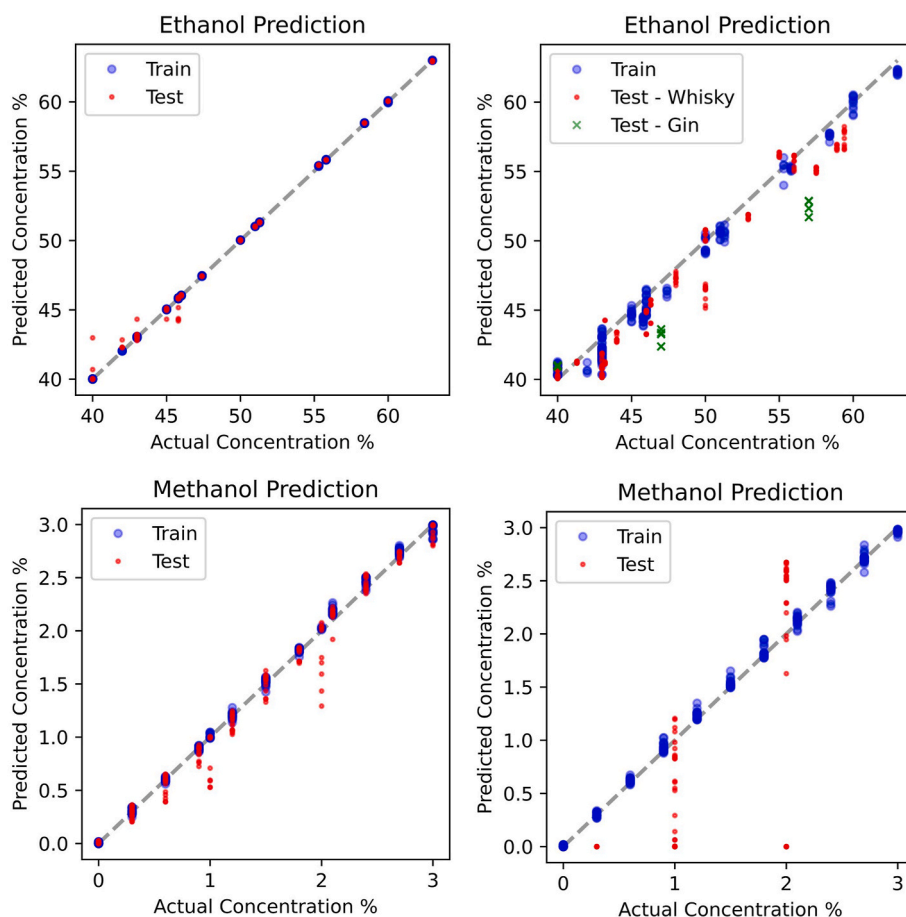


Fig. 1. Brand identification: (a) Deep learning model accuracies per epoch on the training set. (b) Accuracy of deep learning model on the test set and training time. The blue and red dots represent the test accuracy and training time results, respectively. Each dot represents an increase in accuracy and training time as the epochs increase by 100, 200, 500, 1000, and 2000. Shaded grey areas show where the model has > 96% accuracy. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

model can quickly fit to a particular spectrum, such as ethanol, while more complex CNN-based models are required for brand identification. Optimizing the number of epochs is necessary to avoid overfitting and to ensure the generalization of the model. As anticipated, the three deep learning models without PCA did not perform well on the unseen test set. Interestingly, while the performance of the PCA + HPM model improved with an increase in the number of epochs, the performance of the PCA + CNN and PCA + FCN models decreased. Increasing the number of samples in the training set or the number of epochs may improve the performance of the PCA + HPM model on the unseen test set. The predicted ethanol content for gin samples was slightly lower than the actual content, but the model still performed well, which is notable considering that gin was not included in the training data (Table S7). This suggests that the deep learning models were able to generalize well and accurately predict the ethanol content in samples considerably varied from those used in the training model.

The methanol prediction used the same deep learning algorithms as the brand identification and ethanol prediction. The results are presented in Table S8, and demonstrate the model performed poorly on the test set compared to the training and validation sets, indicating over-



**Fig. 2.** Ethanol content prediction: (a) Using PCA + HPM with 2000 epochs ( $R_p2 = 0.998$  and  $RMSE_p = 0.25\%$ ). (b) Using PCA + FCN with 200 epochs applied to a dataset with previously unseen whisky brands (red dot) and gin samples (green x) ( $R_p2 = 0.863$  and  $RMSE_{G} = 2.47\%$  for the test set). Methanol content prediction: (c) Using PCA + HPM with 1000 epochs. The spectral data of Talisker, Cragganmore, 40% ethanol/water, Caol Ila, and Cynelish spiked with methanol were divided into training (60%), validation (20%), and test (20%) sets. (d) Using PCA + HPM with 1000 epochs. The spectral data of Talisker, Cragganmore, and 40% ethanol/water spiked with methanol were split into training (80%) and validation (20%) sets. The spectra of Caol Ila and Cynelish spiked with methanol were only used as a test set. The blue and red dots represent the results of the training and test sets, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

fitting. The PCA + HPM model showed the best performance among all the models. It achieved the lowest  $RMSE_p$  of 0.98% (in terms of methanol in the sample). However, its performance ( $R_p2$  value of 0.300) was still unsatisfactory. The model showed a very good fit when it was trained and tested on the same samples. Fig. 2c shows the performance of the model when both the training set and test sets are comprised of identical whisky samples with the same methanol concentrations, achieving  $R^2$  values of 0.999 for the training set and 0.991 for the test set. Fig. 2d shows the model results when the test set is similar whiskies from different manufacturers, with methanol concentrations not included in the training set. The model was unable to accurately predict methanol concentrations in the samples.

### 3.2. Conventional machine learning

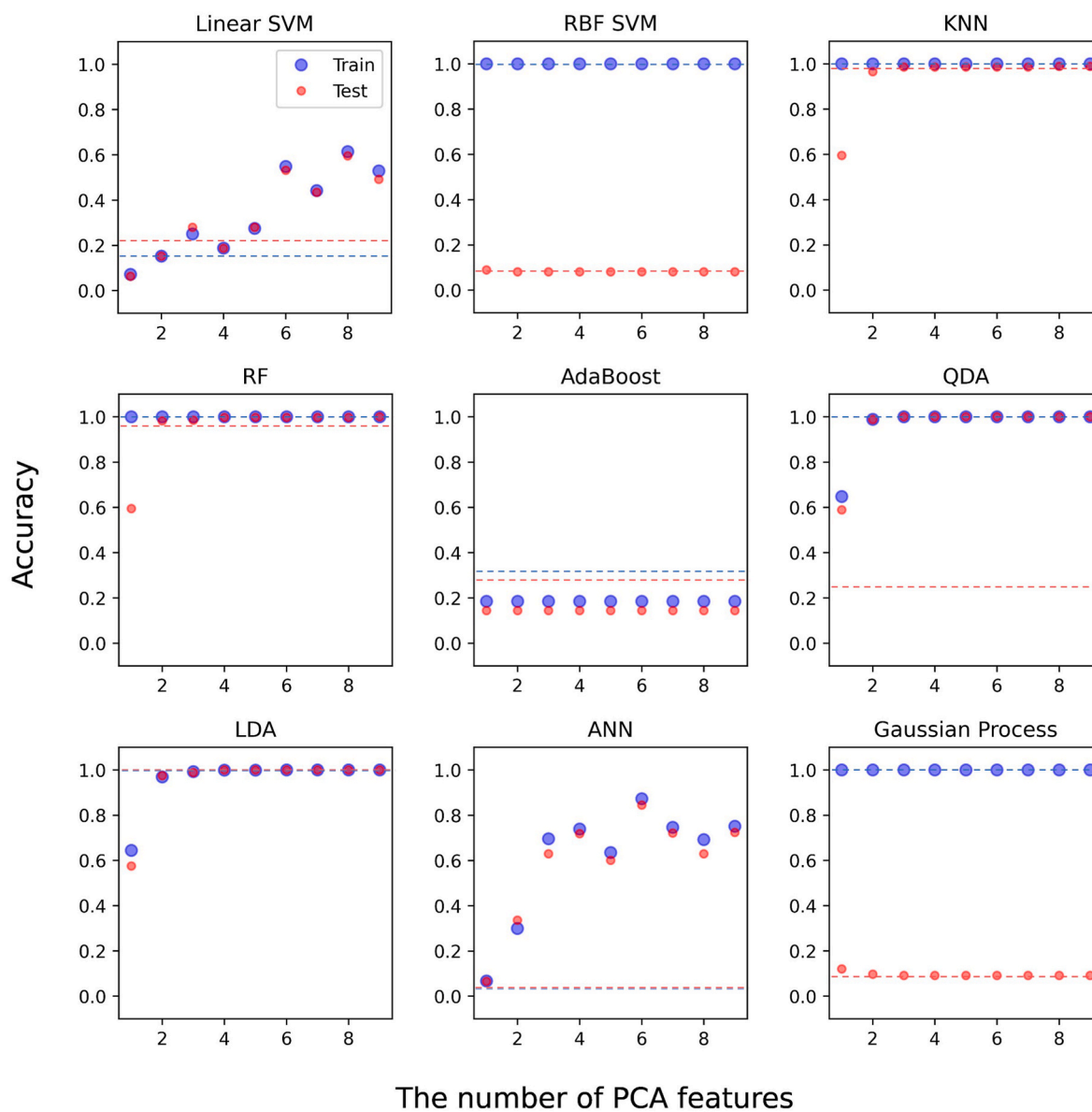
#### 3.2.1. Brand identification

Nine conventional machine learning techniques were applied to classify whisky brands using Raman spectra. All of the conventional machine learning models that were used in this study are described in the supplementary material section S3. Fig. 3 shows the accuracy results for conventional machine learning with and without PCA. Several machine learning algorithms including KNN, RF, and linear discriminant analysis (LDA), without using PCA, achieved a high level of test accuracy, exceeding 96%. Several machine learning methods had significantly improved performance when using PCA. In particular, the test

accuracy of quadratic discriminant analysis (QDA) increased from 25% to over 99% when using three or more PCA features while the test accuracy of artificial neural network (ANN) improved from 4% to 85% when using six PCA features. These findings support the fact that PCA can effectively reduce noise in the data and improve classifier performance. It is worth noting that although certain methods, such as radial basis function SVM and Gaussian Process, achieved high accuracy on the training set, this did not result in high accuracy on the test set, with values lower than 10%.

#### 3.2.2. Methanol regression

We tested machine learning regression models such as PCR, PLSR, and ridge regression to develop a model that could handle data from new whiskies with unknown methanol concentrations. PCR, PLSR, and ridge regression were used to create multivariate models to measure the levels of methanol. Machine learning regression models that were used in this study are described in the supplementary material section S4. The results are summarized in Table S9. The best results were achieved with PLSR using mean-centering (PLSR7), which had  $R_p2 = 0.997$  and  $RMSE_p = 0.05\%$  (in terms of methanol in the sample) (Fig. 4). In particular, the PLSR7 model exhibited high performance for both the training and test sets, with the same  $R^2$  (0.997) and RMSE (0.05%) values. The model demonstrated its effective learning and generalization capabilities by performing well on a different set of whisky samples that were not included in the training set, indicating the model can



**Fig. 3.** Brand identification accuracy of conventional machine learning models. The blue and red dots represent the results of the training and test sets, respectively, with PCA. The dashed lines show the respective accuracy levels without PCA. The red and blue dashed lines overlap each other on the LDA and ANN graphs due to the identical accuracy they achieved on both the training and test sets. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

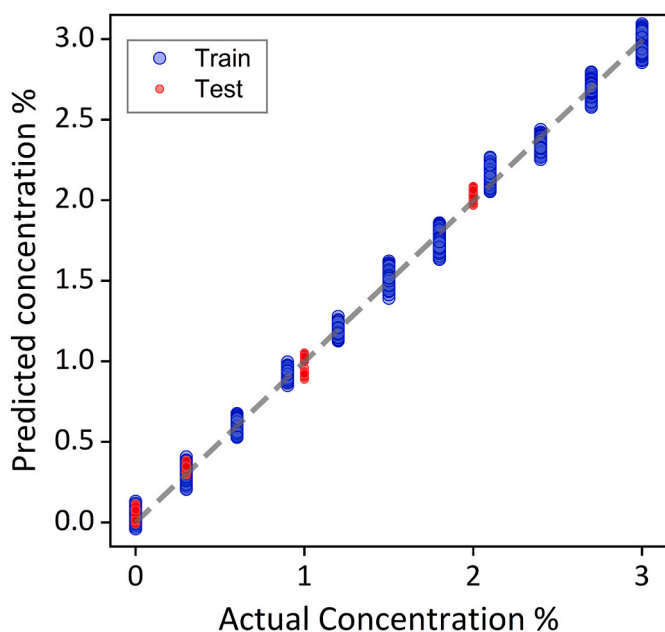
handle various types of whisky samples. The peak close to  $1020\text{ cm}^{-1}$  was assigned to the methanol C–O stretching vibration and was the most significant feature contributing to this model, according to comparison the PLSR loading plot and the Raman spectra of methanol (Fig. S7). It should be noted that the methanol peak intensity could not be tracked directly due to its relative proximity to the ethanol peaks, as shown in Fig. S7c.

Overall, the accurate quantification of methanol in whisky was made possible by PLSR and PCR with mean-centering, which also demonstrated acceptable expansion performance (see PLSR7 and PCR7 in Table S9). Depending on the outcomes, polynomial baseline correction may be included, whereas vector normalization is not necessary to create methanol prediction models using PCR and PLSR. The PLSR7 model was able to detect methanol concentrations in whiskies that were as low as 0.16%. This detection limit was based on the standard deviation of the response (ICH Harmonised Tripartite Guideline, 2005) and further details are provided in the supplementary material section S5. This detection limit for methanol is significantly lower than the

maximum tolerable concentration of 2% methanol in a spirit drink with 40% alcohol by volume, which is considered safe for human consumption (Paine and Dayan, 2001). Additionally, the detection limit is below the current EU general limit for naturally occurring methanol, which is 10 g methanol/L ethanol (equivalent to 0.4% (v/v) methanol at 40% alcohol) (Paine and Dayan, 2001).

### 3.2.3. Through-bottle whisky brand identification

To explore the broader applicability and robustness of these methods, the machine learning methods including KNN, LDA, ANN, RF, and RBF SVM were applied to a *through-bottle* dataset. Each dataset was evaluated using each model, both with and without the inclusion of a 6 component PCA model to determine its predictive power. The best outcomes of the experiments are shown in Table 1, with the full set of results in Table S10. The KNN model exhibited optimal performance when the training and testing datasets were similar, specifically when both were obtained either through vials (VV), bottles (TT), or a mix of vials and bottles (Mix). However, its performance declined when trained



**Fig. 4.** Methanol content prediction using a PLSR model (PLSR7). The spectral data of Talisker, Cragganmore, and 40% ethanol/water spiked with methanol were used as a training set. The spectra of Caol Ila and Cynelish spiked with methanol were only used as a test set. The blue and red dots represent the prediction results on the training and unseen test sets, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

**Table 1**

The best outcomes of through-bottle brand identification.

Dataset <sup>a</sup>	Best Model	Accuracy (%)
VV	KNN, RF, PCA + KNN, PCA + LDA, PCA + RF	100.0
TT	KNN, PCA + KNN	100.0
VT	RF	39.7
TV	PCA + ANN	52.3
Mix <sup>b</sup>	KNN, PCA + KNN, PCA + RF	100.0

<sup>a</sup> The datasets are named using the format "<Train><Test>". Raman spectra were obtained either through vials (denoted as 'V') or through bottles (denoted as 'T'). For example, if the model was trained using measurements obtained through vials and tested using measurements obtained through bottles, the dataset would be named 'VT'.

<sup>b</sup> The training and test datasets include measurements obtained through both vials and bottles.

on data obtained through vials and tested on the data obtained through bottles (VT) or vice versa (TV). It should be noted that while the KNN model achieved perfect prediction in certain instances, other models also demonstrated high accuracy, exceeding 98% in some cases.

The results obtained with this dataset provide valuable insights into the relative performance and predictability of the techniques used. Deep learning models were not applied in this case due to their need for larger volumes of data to produce meaningful results.

#### 4. Discussion

For the primary analysis we chose a diverse set of twenty-eight whisky samples in terms of distilleries, flavours, cask types, ages, and ethanol contents in order to ensure that the training set is representative of real-world scenarios and collected data across different days to reduce environmental variability. Three deep learning methods were used to classify whisky brands and to quantify ethanol or methanol, and nine conventional machine learning techniques were used for whisky brand

identification, and three regression models for methanol quantification. The accuracy of these approaches was compared when the spectral data was either unprocessed or preprocessed using PCA. The deep learning models and three regression models for chemical regression were also compared. The results on the test set are summarized in Table 2.

Several traditional machine learning algorithms such as KNN, RF, and LDA, as well as deep learning algorithms such as CNN and HPM, demonstrated excellent performance of >94% using unprocessed data without PCA. It appears that the use of PCA improved the test accuracy of several conventional machine learning models, including QDA and ANN. Conventional machine learning models such as KNN, RF, and LDA, as well as deep learning models like CNN and HPM, achieved the best performance both with and without the use of PCA. The QDA and FCN models also performed well when PCA was used. While PCA can be beneficial for reducing the dimensionality of data and potentially improving the performance of some machine learning models, it is important to note that PCA is not always necessary. Certain machine learning algorithms, such as KNN and LDA, can handle high-dimensional data effectively and find the optimal features for classification without the need for PCA. However, it is worth noting that while these conventional classification models are effective for classification tasks, they are not applicable for regression analysis. On the other hand, the deep learning models used in this study, such as CNN and HPM, have the flexibility to be applied to both classification and regression tasks, including chemical regression. This versatility makes deep learning models a powerful tool in spectroscopic analysis. On the unseen ethanol test set, only the PCA + FCN model demonstrated good performance, achieving an  $R^2$  value of 0.863 and an RMSE of 2.47%. In contrast, none of the deep learning models exhibited good performance on the unseen methanol test set while the three conventional regression models excelled on the unseen methanol test set, with  $R^2$  values greater than or equal to 0.995 and RMSEs greater than or equal to 0.07%.

The findings of this study can be used to develop new methods for the quantification of ethanol in whisky, and to improve the quality and efficiency of whisky production. In particular, it is useful to note that the conventional machine learning and deep learning techniques are advantageous in different applications, particularly when there are varied quantities of training data available. For example, if the goal is to perform both brand identification and chemical regression simultaneously, a deep learning model may be the most suitable choice. However, if the task only requires either brand identification or chemical regression, then conventional machine learning models may result in equivalent or higher accuracy. It is also important to note that in order to achieve optimal performance with conventional machine learning models, it is necessary to carefully optimize them based on the results of spectral preprocessing and the number of PCA or PLS features used.

Although the deep learning model has the ability to classify and predict many categories and contaminant levels simultaneously to a high precision, it takes vastly more data and computation to develop this knowledge base than traditional techniques. For example, when it comes to detecting methanol concentrations, an inherently challenging task given the low concentrations and the small changes to the spectra, this technique in its current form struggles compared to a more hands-on technique. Manual spectral windowing would likely improve performance, however this requires prior knowledge of the expected spectral changes which we have generally avoided with the goal of developing a flexible analysis approach. As such, it is recommended that if this algorithm detects a low quantity of methanol a sample should be sent off for conventional laboratory analysis.

Machine learning methods have shown strong performance in identifying brands even when applied to datasets obtained through bottles, as summarized in Table 1. The through-vial Raman measurement system differs from the through-vial measurement system in terms of spectral range and laser power. These differences underscore the potential for generalizing machine learning methods. Our findings indicate that when the training and testing sets are consistent, such as

**Table 2**  
Comparison table of brand identification and chemical regression on the test set.

Model	without PCA			with PCA <sup>a</sup>		
	Accuracy (%) for brand	R <sup>2</sup> (RMSE) for ethanol	R <sup>2</sup> (RMSE) for methanol	Accuracy (%) for brand	R <sup>2</sup> (RMSE) for ethanol	R <sup>2</sup> (RMSE) for methanol
CNN	95.1	0.342 (6.04)	0.092 (1.30)	96.6	0.574 (4.71)	0.246 (1.13)
FCN	90.2	0.032 (7.40)	0.007 (1.06)	95.1	0.863 (2.47)	0.601 (1.00)
HPM	94.8	0.253 (7.06)	0.163 (1.30)	97.9	0.645 (4.34)	0.300 (0.98)
Linear SVM	20.7	N/A	N/A	53.2	N/A	N/A
RBF SVM	8.1	N/A	N/A	8.1	N/A	N/A
KNN	98.9	N/A	N/A	98.6	N/A	N/A
RF	96.0	N/A	N/A	99.7	N/A	N/A
AdaBoost	27.6	N/A	N/A	14.4	N/A	N/A
QDA	24.7	N/A	N/A	100.0	N/A	N/A
LDA	100.0	N/A	N/A	100.0	N/A	N/A
ANN	3.5	N/A	N/A	84.5	N/A	N/A
Gaussian Process	9.2	N/A	N/A	9.2	N/A	N/A
PCR	N/A	A	0.997 (0.05)			
PLSR	N/A	A	0.997 (0.05)			
Ridge regression	N/A	A	0.995 (0.07)			

<sup>a</sup>Six PCA features were used as input for classification and regression.

N/A (pink) - not applicable.

A - applicable but not tested.

Green - excellent performance.

Red - poor performance.

not being trained on through-bottle data and tested on through-vial data, the prediction problem can be largely solved using the techniques outlined in this work. Notably, even when applied to different types of containers, some level of predictability was still observed. This suggests that the key features of the whisky spectra remain prominent when compared between beam and vessel types. These results highlight the robustness of the models and methods described here, although further training and a more diverse dataset may be required for practical applications as although the through-the-bottle method reduces the glass background it does not eliminate it completely. Future work will investigate methods to separately record the glass fluorescence and Raman spectra, such that this can be integrated within the data analysis process to more accurately separate the contents Raman from the container.

## 5. Conclusions

In this study, conventional machine learning and deep learning effectively discriminated between twenty-eight whisky brands. Spectra were obtained with a small volume and, aside from decanting from the bottles into vials, no sample preparation was required. Despite previous publications in the literature addressing the application of Raman spectroscopy and chemometric methods for whisky analysis, our approach does not require a manual preprocessing and performed extremely well in terms of brand identification and ethanol prediction. We utilized PCA to reduce the dimensionality of the Raman spectra and retain the most relevant features for classification. Following PCA, we applied various machine learning algorithms for classification and

regression. KNN and LDA performed well with or without PCA, while RF and QDA excelled with PCA. Therefore, the use of PCA depends on the machine learning algorithm and the data. For deep learning, the PCA + HPM was the best for both brand identification (test accuracy of 98%) and ethanol content prediction ( $R_p^2$  of 0.998 and  $RMSE_p$  of 0.25%). The PCA model accurately predicted the ethanol content in new samples, which were not used for training the model, with an  $R_u^2$  of 0.863 and  $RMSE_U$  of 2.47%. The PLSR effectively predicted methanol contents ( $R_p^2$  of 0.997,  $RMSE_p$  of 0.05%, and the detection limit of 0.17%).

Machine learning and deep learning models have shown strong performance on raw data, indicating their adaptability to new applications without the need for spectral preprocessing. The new HPM model can identify brands and quantify ethanol simultaneously, unlike conventional machine learning methods. The ability of deep learning models to quantify ethanol and methanol is expected to improve with larger training datasets. Our methods were applied to a through-bottle configuration, allowing spectra to be obtained from unopened samples. Even with a small preliminary dataset, the accuracy of brand identification remained high, demonstrating the practical potential of our methods. Our chemical regression models for ethanol and methanol have demonstrated their ability to be applied to new samples. Specifically, the ethanol quantification model can predict the concentration of ethanol in gin, and the methanol quantification model can effectively generalize to unseen samples.

Here is a summary of when to use each model.

- Deep learning models: Ideal for brand identification and ethanol quantification, but demand substantial data and resources.



- Conventional machine learning models (KNN, LDA, RF): Great for brand identification, not for ethanol quantification, require careful setup.
- PLSR: Remarkably effective for methanol quantification, even on new samples.
- PCA: Useful for simplifying Raman spectra and enhancing performance of some models, especially deep learning models, though not always necessary.

Our technique has demonstrated its potential in detecting counterfeit spirits and assessing the quality of other high-value liquid samples. Our machine learning method was successful in accurately identifying whisky brands with subtle differences in ingredients, providing evidence of its effectiveness. The methods and techniques we developed should be applicable to detecting adulteration or substitution by training the model with a new dataset. Additionally, the absence of manual pre-processing should enable this model to be applied directly to other high-value products such as edible oil, honey, or caviar. This represents a significant advancement in the field and opens up new possibilities for the detection of counterfeit and adulterated products.

#### CRedit authorship contribution statement

**Kwang Jun Lee:** Data curation, Formal analysis, Methodology, Investigation, Visualization, Writing – original draft, Writing – review & editing. **Alexander C. Trowbridge:** Data curation, Investigation, Software, Methodology, Formal analysis, Writing – original draft, Writing – review & editing. **Graham D. Bruce:** Methodology, Writing – review & editing. **George O. Dwapanyin:** Methodology, Visualization, Writing – review & editing. **Kylie R. Dunning:** Resources, Writing – review & editing. **Kishan Dholakia:** Conceptualization, Funding acquisition, Methodology, Supervision, Project administration, Writing – original draft, Writing – review & editing. **Erik P. Schartner:** Conceptualization, Funding acquisition, Resources, Project administration, Supervision, Writing – review & editing.

#### Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

#### Data availability

Data will be made available on request.

#### Acknowledgements

The authors acknowledge funding support from an ARC Laureate Fellowship FL210100099, an NHMRC Ideas Grant APP2003786, the European Union's Horizon 2020 research and innovation programme "Proscope" (871212), funding through the Innovation Connections program partnering with Beston Global Foods Pty Ltd and from the University of St Andrews Impact and Innovation Fund. The authors acknowledge useful discussions with Alistair McFarlane, Yining Hu & Andrew Richardson. This work was performed in part at the OptoFab node of the Australian National Fabrication Facility utilizing Commonwealth and SA State Government funding.

#### Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.crfs.2024.100729>.

#### References

- Ashok, Praveen C., Praveen, Bavishna B., Dholakia, K., 2011. Near infrared spectroscopic analysis of single malt Scotch whisky on an optofluidic chip, 19 (23), 22982–22992.
- Ashok, Praveen C., Praveen, Bavishna B., Dholakia, K., 2013. Optofluidic Raman sensor for simultaneous detection of the toxicity and quality of alcoholic beverages. *J. Raman Spectrosc.* 44 (6), 795–797.
- Bzdok, D., Altman, N., Krzywinski, M., 2018. Statistics versus machine learning. *Nat. Methods* 15 (4), 233–234.
- Chaudhry, P.E., Zimmerman, A., Peters, J.R., Cordell, V.V., 2009. Preserving intellectual property rights: managerial insight into the escalating counterfeit market quandary. *Bus. Horiz.* 52 (1), 57–66.
- Chollet, F., 2017. *Deep Learning with Python*. Simon and Schuster.
- CNN, 2019. Web page. url: <https://edition.cnn.com/2019/02/24/asia/india-alcoholpoisoning/index.html>. (Accessed 20 March 2023). visited on.
- Dos Santos, C.A.T., Páscoa, R.N., Porto, P.A., Cerdeira, A.L., González-Sáiz, J.M., Pizarro, C., Lopes, J.A., 2018. Raman spectroscopy for wine analyses: a comparison with near and mid infrared spectroscopy. *Talanta* 186, 306–314.
- Ellis, D.I., Eccles, R., Xu, Y., Griffen, J., Muhamadali, H., Matousek, P., Goodall, I., Goodacre, R., 2017. Through-container, extremely low concentration detection of multiple chemical markers of counterfeit alcohol using a handheld SORS device. *Sci. Rep.* 7, 12082.
- Ellis, D.I., Muhamadali, H., Xu, Y., Eccles, R., Goodall, I., Goodacre, R., 2019. Rapid through-container detection of fake spirits and methanol quantification with handheld Raman spectroscopy. *Analyst* 144 (1), 324–330.
- Fan, X.Q., Ming, W., Zeng, H.T., Zhang, Z.M., Lu, H.M., 2019. Deep learning-based component identification for the Raman spectra of mixtures. *Analyst* 144 (5), 1789–1798.
- Flack, A., 2022. Whisky analysis by Raman spectroscopy. Application note url: [https://www.spectroscopyeurope.com/system/files/pdf/AN\\_R11\\_Whisky-Analysis-by-Raman-Spectroscopy.pdf](https://www.spectroscopyeurope.com/system/files/pdf/AN_R11_Whisky-Analysis-by-Raman-Spectroscopy.pdf). (Accessed 20 March 2023). visited on.
- Fleming, H., Chen, M., Bruce, G.D., Dholakia, K., 2020. Through-bottle whisky sensing and classification using Raman spectroscopy in an axicon-based backscattering configuration. *Anal. Methods* 12 (37), 4572–4578.
- Fraud Advisory Panel, 2022. *What's your poison? The true cost of fake alcohol and how to catch the culprits*. Report. url: <https://www.fraudadvisorypanel.org/wp-content/uploads/2023/01/FAP-Special-Report-on-Alcohol-Fraud%5C.WEB.pdf>. (Accessed 20 March 2023). visited on.
- Green, R.T., Smith, T., 2002. Executive insights: countering brand counterfeiters. *J. Int. Market.* 10 (4), 89–106.
- ICH Harmonised Tripartite Guideline, 2005. *Validation of Analytical Procedures: Text and Methodology*, 1.20, p. 5. Q2 (R1).
- Izake, Emad L., 2010. Forensic and homeland security applications of modern portable Raman spectroscopy. *Forensic Sci. Int.* 202 (1–3), 1–8.
- Janiesch, C., Zscheck, P., Heinrich, K., 2021. Machine learning and deep learning. *Electron. Mark.* 31 (3), 685–695.
- Jimenez-Carvelo, A.M., Osorio, M.T., Koidis, A., Gonzalez-Casado, A., Cuadros-Rodriguez, L., 2017. Chemometric classification and quantification of olive oil in blends with any edible vegetable oils using FTIR-ATR and Raman spectroscopy. *LWT—Food Sci. Technol.* 86, 174–184.
- Khan, S., Ullah, R., Khan, A., Sohail, A., Wahab, N., Bilal, M., Ahmed, M., 2017. Random forest-based evaluation of Raman spectroscopy for dengue fever analysis. *Appl. Spectrosc.* 71 (9), 2111–2117.
- Kiefer, J., Cromwell, A.L., 2017. Analysis of single malt Scotch whisky using Raman spectroscopy. *Anal. Methods* 9 (3), 511–518.
- Kotsiantis, S.B., 2007. Supervised machine learning: a review of classification techniques. *Informatica-Journal of Computing and Informatics* 31.3, 249–268.
- Kotsiantis, S.B., Zaharakis, I.D., Pintelas, P.E., 2006. Machine learning: a review of classification and combining techniques. *Artif. Intell. Rev.* 26 (3), 159–190.
- LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. *Nature* 521 (7553), 436–444.
- Lednev, I.K., Sikirzhitski, V., 2012. Raman spectroscopy and advanced statistics for biochemical research and analytical purposes. *Am. Pharmaceut. Rev.* 15 (4).
- Liland, K.H., Almøy, T., Mevik, B.H., 2010. Optimal choice of baseline correction for multivariate calibration of spectra. *Appl. Spectrosc.* 64 (9), 1007–1016.
- Limm, W., Karunathilaka, S.R., Yakes, B.J., Mossoba, M.M., 2018. A portable mid-infrared spectrometer and a non-targeted chemometric approach for the rapid screening of economically motivated adulteration of milk powder. *Int. Dairy J.* 85, 177–183.
- Lussier, F., Thibault, V., Charron, B., Wallace, G.Q., Masson, J.F., 2020. Deep learning and artificial intelligence methods for Raman and surface-enhanced Raman scattering. *TrACTrends in Analytical Chemistry* 124, 115796.
- Mandriole, L., Zeppa, G., Giovannozzi, A.M., Rossi, A.M., 2016. Controlling protected designation of origin of wine by Raman spectroscopy. *Food Chem.* 211, 260–267.
- Moore, David S., Scharff, R Jason, 2009. Portable Raman explosives detection. *Anal. Bioanal. Chem.* 393, 1571–1578.
- Nordon, A., Mills, A., Burn, R.T., Cusick, F.M., Littlejohn, D., 2005. Comparison of noninvasive NIR and Raman spectrometries for determination of alcohol content of spirits. *Anal. Chim. Acta* 548 (1–2), 148–158.
- Oliveira, L. P. de, Rocha, D.P., de Araujo, W.R., Munoz, R.A.A., Paixao, T.R.L.C., Salles, M.O., 2018. Forensics in hand: new trends in forensic devices (2013–2017). *Anal. Methods* 10 (43), 5135–5163.
- Paine, A.J., Dayan, A.D., 2001. Defining a tolerable concentration of methanol in alcoholic drinks. *Hum. Exp. Toxicol.* 20 (11), 563–568.
- Pan, L.R., Pipitsunthonsan, P., Daenggam, C., Channumsin, S., Sreesawat, S., Chongcheawchannan, M., 2021. Identification of complex mixtures for Raman

- spectroscopy using a novel scheme based on a new multi-label deep neural network. *IEEE Sensor. J.* 21 (9), 10834–10843.
- Power, A.C., Ni Neill, C., Geoghegan, S., Currivan, S., Deasy, M., Cozzolino, D., 2020. A brief history of whiskey adulteration and the role of spectroscopy combined with chemometrics in the detection of modern whiskey fraud. *Beverages* 6.3, 49.
- Sarker, I.H., 2021. Deep learning: a comprehensive overview on techniques, taxonomy, applications and research directions. *SN Computer Science* 2 (6), 420.
- Shillito, G.E., McMillan, L., Bruce, G.D., Dholakia, K., 2022. To focus-match or not to focusmatch inverse spatially offset Raman spectroscopy: a question of light penetration. *Opt Express* 30 (6), 8876–8888.
- Shlens, J., 2014. A tutorial on principal component analysis. *arXiv preprint arXiv:1404.1100*.
- Singh, A., Thakur, N., Sharma, A., 2016. A review of supervised machine learning algorithms. In: *Proceedings of the 10th Indiacom - 2016 3rd International Conference on Computing for Sustainable Global Development*, pp. 1310–1315.
- Singh, P., Singh, N., Singh, K.K., Singh, A., 2021. Diagnosing of disease using machine learning. In: *Machine Learning and the Internet of Medical Things in Healthcare*. Elsevier, pp. 89–111.
- Soon, J.M., Manning, L., 2019. Developing anti-counterfeiting measures: the role of smart packaging. *Food Res. Int.* 123, 135–143.
- Taye, Mohammad Mustafa, 2023. Understanding of machine learning with deep learning: architectures, workflow, applications and future directions. *Computers* 12.5, 91.
- Tech Times, 2020. Web page. url: <https://www.techtimes.com/articles/248428/20200329/social-media-misinformation-that-led-to-300-deaths-in-iran-claimed-drinking-methanol-was-a-cure-for-covid-19.htm>. (Accessed 20 March 2023). visited on.
- The British Broadcasting Corporation, 2018. url: Web Page <https://www.bbc.com/news/uk-scotland-scotland-business-46566703>. (Accessed 20 March 2023). visited on.
- The Guardian, 2018. url: Web Page <https://www.theguardian.com/uk-news/2018/dec/20/rare-whisky-market-flooded-with-fakes-says-dealer>. (Accessed 20 March 2023). visited on.
- The Times, 2018. Web page. url: <https://www.thetimes.co.uk/article/dram-and-blasthit-rd-of-vintage-scotch-whisky-found-to-be-fake-frhscnlx0>. (Accessed 20 March 2023). visited on.
- Vankeirsbilck, Tineke, Vercauteren, Ann, Baeyens, Willy, Van der Weken, G., Verpoort, Francis, Vergote, Geert, Remon, Jean Paul, 2002. Applications of Raman spectroscopy in pharmaceutical analysis. *Trac. Trends Anal. Chem.* 21 (12), 869–877.
- Wang, C.-H., Huang, C.-C., Chen, W., Lai, Y.-S., 2020. The chemical aspects of Raman spectroscopy: statistical structure-spectrum relationship in the analyses of bioflavonoids. *J. Food Drug Anal.* 28 (2), 239.