

YAN, Y., REN, J., TSCHANNERL, J., ZHAO, H., HARRISON, B. and JACK, F. 2021. Nondestructive phenolic compounds measurement and origin discrimination of peated barley malt using near-infrared hyperspectral imagery and machine learning. *IEEE transactions on instrumentation and measurement* [online], 70, article 5010715. Available from: <https://doi.org/10.1109/TIM.2021.3082274>

Nondestructive phenolic compounds measurement and origin discrimination of peated barley malt using near-infrared hyperspectral imagery and machine learning.

YAN, Y., REN, J., TSCHANNERL, J., ZHAO, H., HARRISON, B. and JACK, F.

2021

© 2021 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

Nondestructive Phenolic Compounds Measurement and Origin Discrimination of Peated Barley Malt using Near-infrared Hyperspectral Imagery and Machine Learning

Yijun Yan, Jinchang Ren, Julius Tschannerl, Huimin Zhao, Barry Harrison, Frances Jack

Abstract— Quantifying phenolic compound in peated barley malt and discriminating its origin are essential to maintain the aroma of high-quality Scottish whisky during the manufacturing process. The content of the total phenol varies in peated barley malts, which is critical in measuring the associated peatiness level. Existing methods for measuring such phenols are destructive and/or time consuming. To tackle these issues, we propose in this paper a novel nondestructive system for fast and effective estimating the phenolic concentrations and discriminating their origins with the near-infrared hyperspectral imagery and machine learning. First, novel ways of data acquisition and normalization are developed for robustness. Then, the principal component analysis (PCA) and folded-PCA are fused for extracting the global and local spectral features, followed by the support vector machine (SVM) based origin discrimination and deep neural network based phenolic measurement. In total 27 categories of peated barley malts from eight suppliers are utilized to form thousands of spectral samples for modelling. A classification accuracy up to 99.5% and a squared-correlation-coefficient up to 98.57% are achieved in our experiments, outperforming a few state-of-the-art. These have fully demonstrated the efficacy of our system in automated phenolic measurement and origin discrimination to benefit the quality monitoring in the whisky industry.

Index Terms— Phenolic compound measurement; origin discrimination; near infrared (NIR) Hyperspectral imagery; peated barley malt; machine learning.

I. INTRODUCTION

Hyperspectral imaging (HSI), or image spectroscopy, combines spectroscopy with conventional 2-D imaging to take the advantage of spectral characterization of each pixel in imaging an object. With the concept first derived in [1], HSI aims to identify the surface materials in a form of images. Afterwards, HSI has been successfully applied in many remote sensing tasks, including precision agriculture [4], land-cover analysis [5], military surveillance [6] and mineral exploration [7]. Owing to its additional radiance spectrum information for each pixel, HSI has become an emerging technique for nondestructive inspection and assessment in a wide range of lab-based new applications, e.g. food quality control [8],

medical diagnosis [9], forensics [10] and artwork authentication [11], etc. Moreover, HSI has also been successfully applied in other image processing and machine learning tasks, such as object detection and material identification [12].

As a high-quality spirit drink exclusively produced in Scotland, Scotch whisky has been exported to 175 countries and regions around the world, accounting for nearly a quarter of all UK food and drink exports [13]. Due to its high commercial value, maintaining high quality standards during production are of primary interest to the Scotch whisky industry. Flavor character is the key factor, which comes both naturally from raw materials and is generated through the production process. Peated Scotch whiskies differ from others in terms of smoky flavor that depends on the length of time that the malted barley is exposed to the smoke during the drying process in a kiln [14]. Volatile phenolic compounds in the peat smoke adhere to the surface of the malted barley and are carried through the production process, resulting in the characteristic smoky taste for many whiskies.

Large parts of the Scottish countryside are covered in peat bogs, where dried peat has a long history of utilization as a heating fuel. Peat from different regions introduces various taste notes and adds additional variety to the rich tastes of Scotch whiskies [15]. Depending on the flavor requirements, distilleries use peated barley malt from different origins with various peated levels. Distillers usually define a peated level in terms of Phenol Parts per Million (ppm), which often measures the collective concentration of a range of phenolic compounds by traditional chemical analytics methods e.g. spectrophotometric techniques and high-performance liquid chromatography (HPLC) [16]. Therefore, the combined total of all phenolic compounds is used as a marker to the degree of peatiness of the malted barley. However, those existing methods don't measure the phenols on the peated barley malt itself. Rather, they rely on a pre-distillation step to extract the phenols followed by analyzing the resulting distillate, which is destructive and time consuming. For the benefit of the Scotch Whisky industry, it is of value to develop a fast, accurate and nondestructive technique to determine the phenol levels in the

Date of submission: October 26 2020.

This work was partially funded by the Institute of Brewing and Distilling (2018/06 hyperspectral imaging), the Dazhi Scholarship of the Guangdong Polytechnic Normal University, the National Natural Science Foundation of China (Grant: 62072122), the Education Department of Guangdong Province (Grant: 2019KSYS009), the Scientific and Technological Planning Projects of Guangdong Province (2021A0505030074). Corresponding Author: Prof. J. Ren (jinchang.ren@ieee.org).

Y. Yan, J. Ren and H. Zhao are with School of Computer Sciences, Guangdong Polytechnic Normal University, Guangzhou, China. J. Ren is also affiliated with the National Subsea Centre, Robert Gordon University, Aberdeen, U.K.

J. Tschannerl is with Apeel Sciences, California, U.S.A.

B. Harrison and F. Jack are with Scotch Whisky Research Institute, Research Avenue North, Edinburgh, UK.

peated barley malt and trace the origin of the peated barley malt.

In recent years, a wide range of machine learning algorithms have been successfully applied for HSI based food quality inspection and grading applications. Noviyanto et al. [17] proposed a multi-stage model including noisy band elimination, spectral normalization and hierarchical classification, to classify the honey botanical origin with the k -nearest neighbor (KNN) and support vector machine (SVM) classifiers. Erkinbaev et al. [18] built an artificial neural network (ANN) to predict wheat hardness from HSI data. Han et al. [19] assessed the aflatoxin on peanut from HSI with the convolutional neural network (CNN). In [20], a random forest model was used to classify four types of bruising degrees of an apple. Comparing with the conventional chemical analysis, HSI and machine learning has two advantages, i.e. fast data acquisition and non-destructive testing. However, the analytics performance may be affected by the way of data acquisition and modelling.

In the past, HSI has been used to detect the concentration of phenolic compounds in skins/stems of grapes [21] and seeds [22]. However, applying HSI for nondestructive measurement of phenolic compound in peated barley malt has not been fully explored. Although some preliminary works [23, 24] have shown the potential of using HSI to analyze the peated barley malt, they have some major drawbacks, especially the limited category number of peated barley malt and impractical experimental settings to have both the training and testing data from the same hypercube. Moreover, the phenolic compounds show few salient features on their rich but noisy spectral profiles, which brings a fundamental challenge in this context.

To tackle the aforementioned issues, in this study, a novel nondestructive HSI-based system is developed for real-time origin discrimination and phenolic measurement of peated barley malt in accordance with industrial conditions. The proposed system has been successfully applied for estimating the total phenol levels and distinguishing 27 categories of peated barley malt provided by 8 suppliers. To the best of our knowledge, this is the first time that the sophisticated data acquisition and feature extraction methods are proposed and combined for extracting both global and local spectral features, followed by various machine learning models for classification of the origins and prediction of the phenolic levels. With the excellent results produced, the proposed system can be potentially applied in real production lines for benefiting the quality monitoring in the whisky industry.

The main contributions of this paper are highlighted below:

- 1) We propose a fast, accurate and nondestructive testing system for phenolic measurement and origin discrimination of peated barley malt, which is essential for accurate flavor control and traceability in the Scotch whisky industry;
- 2) A new way of data acquisition and a new data analysis framework are proposed for consistently spectral measurement of phenolic levels. After applying the joint bilateral filter to reduce the data noise, the principal component analysis (PCA) and folded PCA are fused to extract both the global and local spectral features;
- 3) We have carried out comprehensive experiments on 27 categories of 6750 spectral samples of the peated barley

malt from 8 suppliers, where useful discussions and conclusions are achieved. This includes the identified best routines i.e. support vector machines (SVM) based origin discrimination and deep neural network (DNN) based phenolic level prediction.

The rest of this paper is organized as follows. Section II introduces the proposed system. Section III describes the experimental settings including data analysis models and evaluation criteria. Section IV presents the experimental results and discussions. Finally, some concluding remarks are drawn in Section V with some future prospects.

II. MATERIAL AND METHODS

A. Imaging system

The hyperspectral data is acquired by an NIR-HSI camera, InnoSpec RedEye 1.7, which covers a spectral range of 950-1760 nm with a spectral resolution of 10 nm. The NIR-HSI camera operates in the push-broom mode, where the lens of the camera faces downwards and scans only a single line at a time. The light source is provided by 8 tungsten halogen light sources with the exposure time of 2.5ms and 3 watts each, which are equally split into two sets and hang on either side of the camera. The objects to be scanned are moved with a translational stage underneath the camera at an even speed. With a working distance of 24cm, the speed of the translational stage is set to 15mm/s and the length of path is 20cm. There are 320 pixels per line, and each pixel contains 256 spectral responses. 36 noisy bands in the wavelength ranging of 1680-1760nm are removed, resulting in 220 bands remained for spectral analysis.

B. Sample preparation

The peated barley malt grains used in this study are provided by 8 maltsters and distilleries in Scotland, forming 27 categories with the range of the total phenolic concentration varies from 0 ppm (unpeated) to 189.2 ppm (heavily peated) as detailed in Table I. The concentration of the total phenol, including phenol, guaiacol, o-cresol, m/p-cresol, methyl-guaiacol, ethyl-phenol and ethyl-guaiacol, differs in these categories. Although some categories have similar total phenols, the individual phenolic flavor compounds may vary significantly. As a result, origin discrimination is critical for peated barley malts selection and flavor control of Scotch whisky. In addition, as distilleries use the total phenol concentration as an indicator to the phenol level of the malted barley [11], this study focuses on estimating the concentration of total phenol rather than individual compounds.

During kilning, a bed of malted barley is exposed to the peated smoke. During this natural process, the amount of phenols on the surface of each grain is usually inconsistent. As a result, measuring the single spatial point with a plain NIR cannot fully represent the concentration of each grain. To this end, in our study, grains are spread out in a container to maximize their surface, where the average concentration is calculated to represent the spectra of the batch.

For each category, we prepare one container of peated barley malts for training dataset acquisition and another for

TABLE I
DESCRIPTION OF 27 CATEGORIES OF MALTED BARLEY IN PPM.

Category	Supplier	Phenol	Guaiacol	m/p-Cresol	o-Cresol	4-Methyl guaiacol	4-Ethyl phenol	4-Ethyl guaiacol	Total Phenols
1	Maltster A	45.6	1.8	20.2	3.9	3.1	8.2	3.0	85.9
2	Maltster A	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.1
3	Distillery A	14.7	1.8	6.7	2.5	0.7	3.6	1.6	31.6
4	Distillery A	33.0	3.3	15.2	5.7	1.2	8.0	4.7	71.0
5	Distillery A	25.1	2.2	10.7	3.7	1.3	5.4	3.0	51.4
6	Distillery A	19.6	1.7	7.5	2.8	0.6	3.0	1.4	36.6
7	Distillery A	16.7	1.5	6.4	2.3	0.4	3.6	2.0	32.9
8	Distillery A	27.4	2.3	12.4	4.7	0.8	6.1	2.5	56.2
9	Distillery A	20.6	1.9	8.3	3.1	0.0	3.9	2.1	39.9
10	Distillery A	24.8	2.3	10.2	3.7	1.0	4.2	1.0	47.2
11	Distillery A	11.3	0.1	3.2	1.1	6.2	2.4	1.4	25.6
12	Distillery A	26.3	2.0	11.6	3.6	0.0	5.3	6.2	55.0
13	Distillery B	18.3	1.5	6.1	1.5	2.2	2.8	0.0	32.4
14	Distillery B	12.9	0.0	4.4	0.9	0.0	1.4	0.0	19.6
15	Distillery C	4.7	0.0	0.2	0.0	0.0	0.3	0.0	5.3
16	Distillery D	37.0	2.9	11.4	3.3	0.0	4.5	1.8	61.0
17	Distillery D	40.4	4.4	17.7	4.4	3.8	6.9	2.4	80.1
18	Distillery D	40.7	4.5	19.0	5.1	2.9	9.6	3.5	85.3
19	Maltster D	51.3	5.2	24.4	6.4	4.4	11.7	7.0	110.4
20	Maltster D	0.1	0.0	0.0	0.0	0.0	0.0	0.2	0.2
21	Maltster E	0	0	0	0	0	0	0	0.0
22	Maltster E	66.5	4.6	31.2	8.6	1.4	12.6	5.2	130.0
23	Maltster F	11.1	0.5	3.3	1.0	0.3	1.4	0.0	17.6
24	Maltster F	25.2	1.3	8.3	1.7	1.9	3.4	2.1	43.8
25	Maltster F	60.0	3.7	24.0	5.9	3.0	9.0	5.4	111.1
26	Maltster F	83.8	6.2	34.1	7.7	6.0	12.7	4.8	155.4
27	Maltster F	101.3	7.3	42.9	10.2	4.6	15.8	7.0	189.2

testing data acquisition. To mitigate the effects of the ambient light and inconsistency between the training and testing dataset, each container of peated barley malts is scanned 4 times where with the container rotated in 4 different direction (0° , 90° , 180° , 270°) to form four hypercubes. The obtained 4 hypercubes will be pre-processed and stacked in a pixel-wise manner to produce a stacked image data pool for training or testing. The mean spectral samples are generated as follows. Let M denote the number of pixels in the selected region of interest (ROI), we randomly selected one pixel in every s pixels as representative spectral samples, for either training or testing. When $M=50000$ and $s=200$, we will have 250 spectral samples extracted from the hypercube. For 27 categories of peated barley malts, in total 6750 spectral samples are collected to form either the training or the testing data pool. The selection of s will affect the classification results, which is discussed in Section IV. The statistics of training and testing dataset used for three machine learning models with $s=200$ is given in Appendix: Table A.

C. Sample analysis

Fig. 1 shows an example spectrum of one category of the peated barley malt samples, in which the blue curve denotes the mean spectrum of all the valid pixels selected from a ROI in the green region of a pseudo-colored image of the peated barley malt (top right), and the red fill depicts the standard deviation (std). In the bottom-left, a normalized histogram of all pixels at a specified wavelength of 1424nm is also given. As seen, the standard deviation on all the bands is almost the same, and the

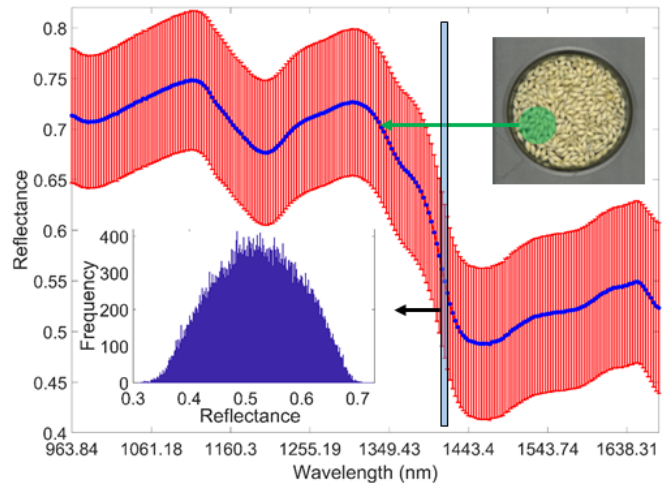


Fig. 1. Spectral profile of ROI with the histogram at certain wavelength

reflectance value of all pixels in a given band appears nearly a Gaussian distribution. As the noise follows a zero-mean normal distribution, the sigma of this Gaussian distribution is about 0.15, indicating a very low signal-to-noise ratio (SNR).

In addition, the main spectral characteristics of the peated barley malts in this study are visualized in Fig. 2 (a), where the highlighted important bands are explained according to some related studies^{1,2} [1, 3] as follows.

In [1], the phenol is shown in three absorption bands close to 880nm, 1130nm and 1660nm, with the highest absorptivity at 1130nm and lowest at 1660nm (Fig. 3(a)). Although our HSI

¹ <https://spectrabase.com/spectrum/Lqer4LFqcZq>

² <https://spectrabase.com/spectrum/CkVWlQpzCOW>

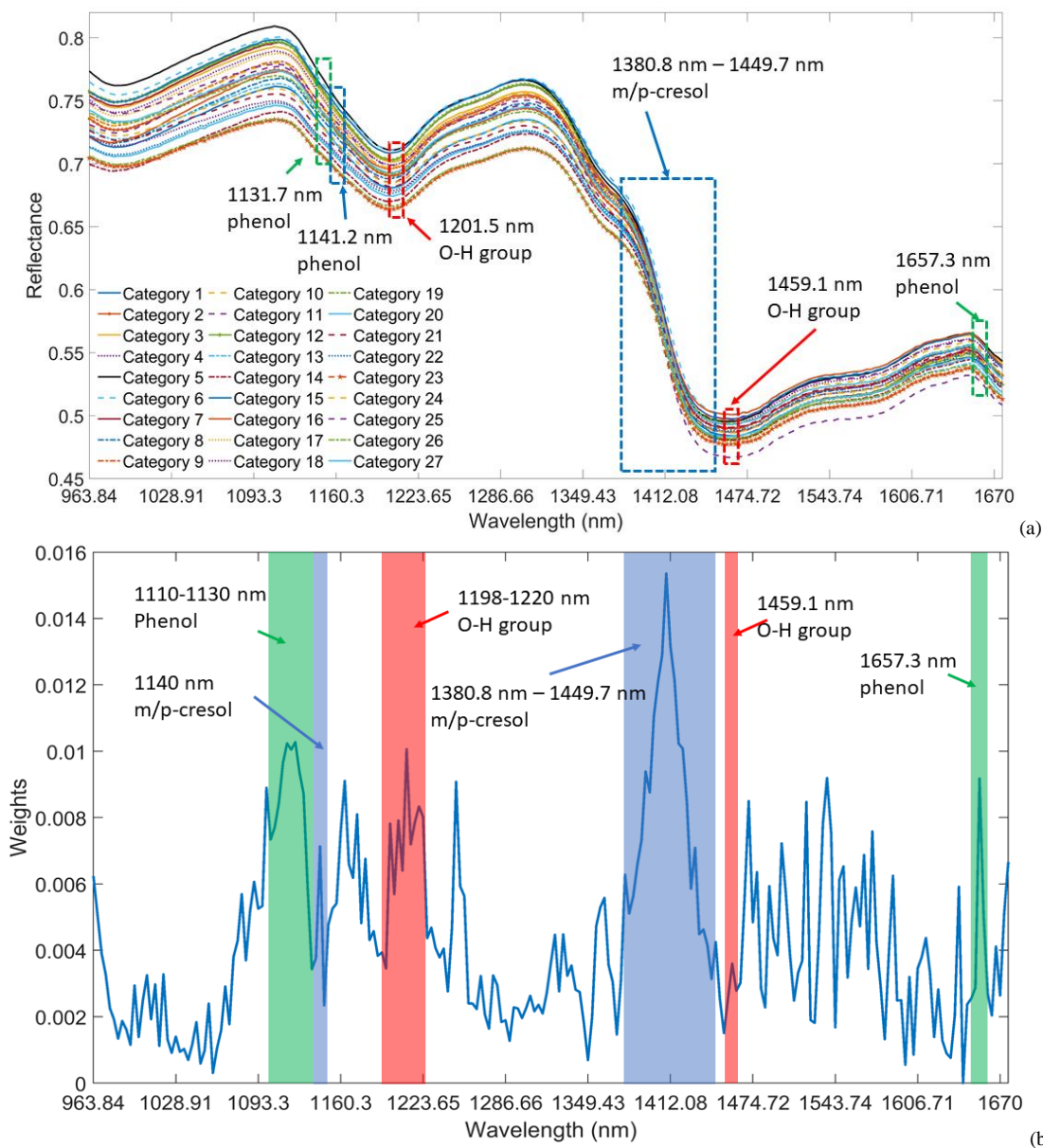


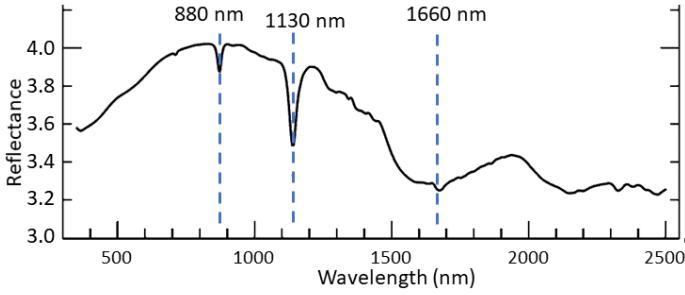
Fig. 2. Comparison of spectra of NIR reflectance of 27 categories of peated barley malt (a) and the corresponding band weights generated by ReliefF analysis using 10 nearest neighbor per classes (b).

camera cannot detect the spectral information below 880nm, the reflectance intensity drops at about 1131.7nm and 1657.3nm, as highlighted in green boxes. As indicated in [3], the O-H groups and C-H groups can represent the main characteristics in the phenolic compounds. Although our HSI camera cannot see the information above 1700nm, the absorption bands of two O-H groups, located close to 1200nm and 1459 nm (Fig. 3(b)) as highlighted in red boxes, match the findings in [3] (Fig. 2 (a)). According to the spectra data provided by a public database, PubChem^{1,2}, the main absorption bands of m/p-cresol occur at 1140nm and 1380-1450nm (Fig. 3(c)), which is highlighted in blue boxes with the quick dropping down of the reflectance intensity in Fig. 2 (a).

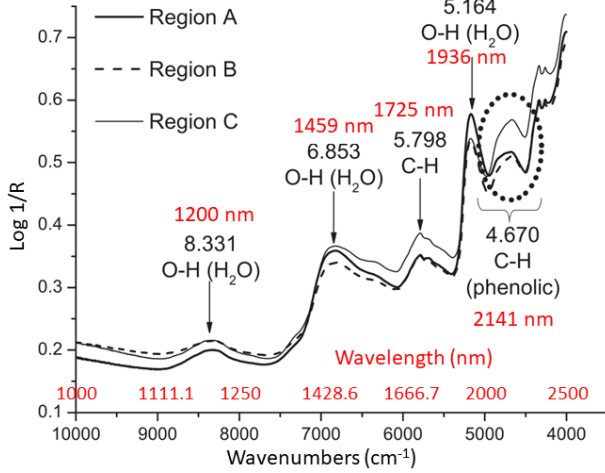
Moreover, the weight of each band is calculated using the ‘ReliefF’ analysis with 10 nearest neighbors per class [25]. For robustness, we randomly select 10 groups of spectral samples

for analysis, and the averaged weights are obtained as ranked importance of the corresponding bands and shown in Fig. 2 (b).

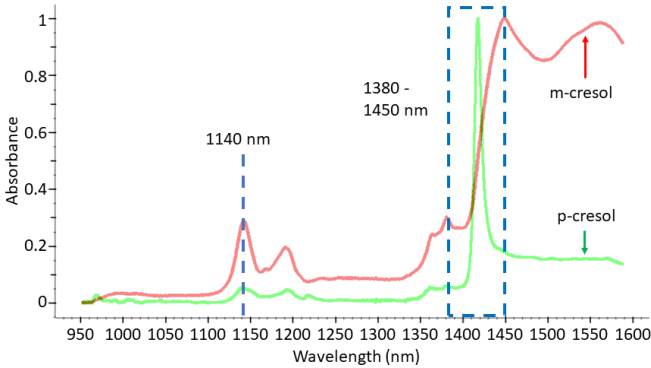
As seen in Fig. 2 (b), the important bands have been highlighted in different colors, where each of them corresponds to a peak in the plotted curve of band weights. Due to lack of detailed spectral information of all individual phenolic compounds, peaked weights at some bands cannot be fully explained. In fact, the chemical characteristics of peated barley malt is very complicated, as any band can contribute more or less to its chemical properties. Therefore, it is very difficult to select one or several representative bands for the purpose. For the data from various maltsters and distilleries, there is neither strictly linear relationship between the reflectance value and the associated phenol concentration nor salient characteristics for each category of peated barley malts. To this end, more sophisticated features are needed to extract the underlying



(a) Reflectance spectrum of phenol [1]



(b) Reflectance spectrum of O-H groups [3]



(c) Absorbance of m/p-cresol as in PubChem [12]

Fig. 3. Characterized spectra of phenol (a), O-H groups (b) and m/p-cresol (c), where the cited images have been modified for improved visual effects. In (b), we convert the original wavenumber to wavelength and insert the results as texts in red for clarity.

patterns of the data for characterizing the NIR spectrum of the scanned peated barley malt grains.

D. Multi-stage data processing

In this section, multi-stage data processing is applied in our proposed framework to extract the representative features for analyzing the spectral data. As illustrated in Fig. 4, this process includes four main modules, i.e. pre-processing, region of interest (ROI) extraction, data selection, feature extraction and fusion as detailed below.

1) Pre-processing

After data acquisition, three pre-processing steps are applied to the HSI for spectral calibration, spatial denoising and spectral

correction sequentially as detailed below.

Spectral calibration: During the data acquisition, the lighting conditions may vary slightly within a hypercube or apparently between different datasets across the scan lines. To tackle such inconsistency and keep the effect of light conditions constant, light calibration is utilized by transforming the raw radiance spectra s to the reflectance spectra r as follows:

$$r = \frac{s - d}{w - d} \quad (1)$$

where the dark reference spectra d is acquired by capturing without any light exposure to the NIR-HSI camera, which can estimate the shot noise of the camera. The white reference spectra w is acquired by capturing an optimally reflective white surface, e.g. Spectralon, aiming to estimate the light sensitivity to the current illumination and normalize the signal.

The spectra of the peated barley malt depends not only on the chemical absorption but also the physical light scattering on the surface of the objects. As the surface of barley grains is very rough, the light exposure of different regions varies. Therefore, shadow effects as well as diverse light scattering may appear and result in poor results of classification and regression. To tackle these issues, joint bilateral filtering (JBF) [2] and standard normal variate (SNV) [26] are employed to smooth and denoise the data in the spatial domain and also to eliminate the scatter distortion in the spectral domain, respectively.

Spatial denoising: As a non-linear method, bilateral filtering smooths images whilst preserving edges by fusing domain and range filtering together. In [2], an improved joint bilateral filtering (JBF) method is proposed for HSI data, which is also used in this paper to spatially smooth the data. Given a HSI data $D \in \mathcal{R}^{I \times J \times B}$, I and J are the spatial size of D and B is the number of spectral bands. The JBF result of the input data D at the location (i, j) of band b can be obtained by:

$$D_{JBF}(i, j, b) = \frac{1}{k(i, j)} \sum_{(p, q) \in w} (G_{\sigma_d}(i - p, j - q) \times G_{\sigma_r}(I_{PC1}(i, j) - I_{PC1}(p, q)) D(p, q, b)) \quad (2)$$

where w is a local window centered at (i, j) with a size of $(2\sigma_d + 1) \times (2\sigma_r + 1)$ pixels.

The normalization factor k is defined as

$$k(i, j) = \sum_{(p, q) \in w} \left(G_{\sigma_d}(i - p, j - q) \times G_{\sigma_r}(D_{PC1}(i, j) - D_{PC1}(p, q)) \right) \quad (3)$$

where D_{PC1} is the first principal components derived from principal component analysis (PCA) [27], and (p, q) represents the spatial location of a pixel in the local window w .

The kernels for domain and range filtering are given by:

$$G_{\sigma_d}(i - p, j - q) = \exp\left(-\frac{(i - p)^2 + (j - q)^2}{2\sigma_d^2}\right) \quad (4)$$

$$G_{\sigma_r}(D_{PC1}(i, j) - D_{PC1}(p, q)) = \exp\left(-\frac{(D_{PC1}(i, j) - D_{PC1}(p, q))^2}{2\sigma_r^2}\right) \quad (5)$$

where σ_d and σ_r determine the neighborhood window and the contributed weights of neighboring pixels, respectively [28].

Spectral correction: SNV is commonly used in HSI to

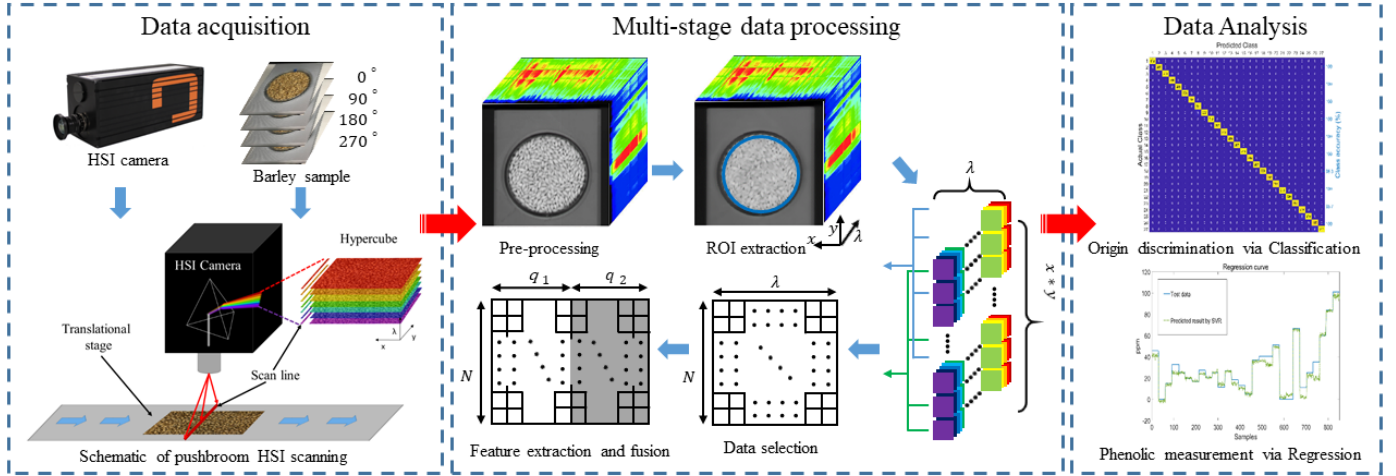


Fig. 5. The workflow of the proposed phenolic measurement and origin discrimination system.

compensate the scattering distortions, i.e. for spectral correction [26]. Here, it is used to reduce the variability between pixels due to scattering. For any pixel with a reflectance spectrum r_s in $D_{JBF} \in \mathcal{R}^{I \times J \times B}$ at the location (i, j) where $i \in 1: I, j \in 1: J$, the process of SNV can be defined by:

$$r_s = D_{JBF}(i, j, b_{1:B}) \quad (6)$$

$$r_s(SNV) = \frac{r_s - \mu}{\sigma} \quad (7)$$

$$\mu = \frac{1}{I \times J} \sum_{s=1}^{I \times J} r_s \quad (8)$$

$$\sigma = \sqrt{\frac{1}{I \times J} \sum_{s=1}^{I \times J} (r_s - \mu)^2} \quad (9)$$

where $b_{1:B}$ is a spectral vector at location (i, j) , μ and σ are the mean and standard deviation of all pixels in D_{JBF} .

2) ROI extraction

In data preparation process, the peated barley malt grains of each category are spread out in a round container. After data acquisition, the HSI data contains the spectra of peated barley malt grains (ROI) as well as the background. Fig. 5(a) illustrate the 100th band in the captured hypercube data and Fig. 5(b) is the 100th band after pre-processing. The background of each image data is pretty much the same, which is not helpful for classification and regression of peated barley malts. Therefore, in this section, we employ a robust circle detection method to extract the pixels of peated barley malt grains only, i.e. any pixels within the blue mask in Fig. 5(e). This will not only speed up the following on feature extraction and fusion but also improve the accuracy of classification and regression.

The process of ROI circle detection is summarized below:

- 1) Apply a thresholding on the 100th band image to extract a binary template. As seen in Fig. 5 (a-b), the grains have a higher intensity than the background. The threshold is selected as the first valley of the histogram going from bright to dark as shown in Fig. 5 (c). Here, the threshold value is determined as 140;

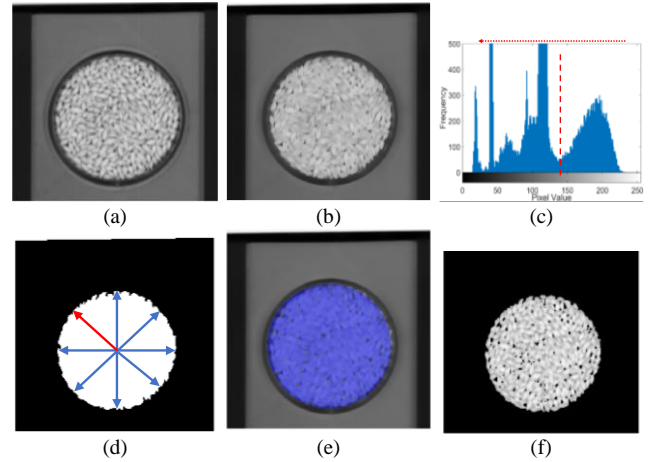


Fig. 4. Illustration of ROI processing. (a) 100th Band of HSI; (b) Pre-processed result of (a); (c) Step 1, histogram of (b); (d) Step 2-3, red arrow represents the minimum radius; (e) Step 4; (f) Step 5.

- 2) Fill the holes in the binary template for calculating its centroid using the central moment, i.e. the mean x and y values of all white pixels;
- 3) Determine the minimum radius in Fig. 5 (d) from the centroid to the boundary in eight directions, which is faster than the Hough transform;
- 4) Generate the circular mask and remove the outliers, the results are shown in Fig. 5 (e);
- 5) Apply the mask on the pre-processed hypercube to extract the pixels of peated barley malt grains as our extracted ROIs, see in Fig. 5 (f).

3) Data selection

As the distribution of phenols on the barley surface is uneven, the spectra of different pixels varies. In addition, the mean phenol level of the entire patch is desired in industrial applications. To this end, rather than taking individual pixels as samples, the average spectra over a subset of s pixels is used for analysis. For each acquired hypercube, the number of spectral samples depends on the number of the chosen subsets. How the number of subsets may affect the performance of classification and regression is discussed in Section IV.

4) Feature extraction and fusion

Considering high redundancy between neighboring spectral bands, feature extraction and data reduction are widely used in HSI data analysis [29-31]. PCA, also known as the Karhunen-Loeve Transform (KLT) [27], is a widely used unsupervised method for dimension reduction and feature extraction in HSI. By orthogonal projection and truncation of the transformed (feature) data, correlations among the data can be removed, resulting in a lower dimension of data as the global structure features of the hypercube. Given N spectral vectors $X = \{x_1, x_2, \dots, x_N\}$, $X \in \mathfrak{R}^{N \times B}$, we can have $PCA(X) \in \mathfrak{R}^{N \times q_{PCA}}$ as the extracted first q_{PCA} principal components.

As a modified version of PCA, Folded-PCA (FPCA) is proposed to extract mainly local structures in the spectral domain [32]. For a given spectral vector x_n , conventional PCA extracts the principal components from all the bands. In Folded-PCA, the spectral vector x_n is divided into H groups, each containing W bands, where $H \times W = B$. When $H = 1$, FPCA degrades to the PCA. The covariance matrix of FPCA is summed by H groups with a size of $W \times W$ in the main diagonal of the original covariance matrix. After calculating the covariance matrix, Eigen problem and data projection will be solved. Given N spectral vectors $X = \{x_1, x_2, \dots, x_N\} \in \mathfrak{R}^{N \times B}$, and q_{FPCA} principal components will be extracted. We can have $FPCA(X) \in \mathfrak{R}^{N \times q_{FPCA}}$. The major difference between PCA and FPCA is the way for calculation of the covariance matrix.

PCA can extract the global spectral features as a small number of low-order principal components, whilst FPCA can preserve local spectral features. To this end, PCA and FPCA features are quite supplementary to each other, which motivates us to fuse them together in the proposed approach. For a pixel x_n , the fused feature can be given by:

$$x_n(FF) = [PCA_{x_n(1)}, PCA_{x_n(2)}, \dots, PCA_{x_n(q_{PCA})}, FPCA_{x_n(1)}, FPCA_{x_n(2)}, \dots, FPCA_{x_n(q_{FPCA})}]^T \in \mathfrak{R}^{(q_{PCA} + q_{FPCA}) \times 1} \quad (10)$$

The fused feature has a smaller dimension ($q_{PCA} + q_{FPCA} < B$), and more representative than the raw data.

III. EXPERIMENTAL SETTINGS

To discriminate the origin of each category of peated barley malts, we consider it as a classification task, where the Support Vector Machine (SVM) is employed as it has been successfully applied in many HSI classification applications [33-35]. In this work, the SVM model is used to identify the 27 categories of peated barley malts provided by 8 suppliers.

To accurately measure the phenol concentration of the peated barley malts, we consider it as a regression task, where three machine learning tools i.e. Support Vector Regression (SVR) [36], Random Forest (RF) [37] and Deep Neural Network (DNN) are employed to train the prediction models. As DNN draws much attention recently and achieves very good results [38], a simple yet effective DNN model is introduced in our work. For efficient implementation, the cost and gamma of the Gaussian kernel in SVM are set to 1024 and 0.125, respectively. The number of decision trees is set to 100 in RF. For the DNN, the back-propagation feedforward neural

TABLE II
PARAMETERS AND FUNCTIONS OF THREE MACHINE LEARNING MODELS.

Machine learning models	Training/learning parameters	
SVM	No. of support vectors	1875
	Kernel function	Gaussian
	Cost	1024
	Gamma	0.125
RF	No. of decision trees	100
DNN	Neural network model	Feedforward
	Training algorithm	Scaled Conjugate Gradient backpropagation
	Performance function	Cross-Entropy
	Total number of layers	5
	Hidden layer neurons	64,32,16
	Input layer nodes	120
	Output layer nodes	1
	Hidden layer transfer function	Tansig
	Output layer transfer function	Logsig
	Data division	random
	Maximum no..of epochs	1000
	Validation check iterations	6
	Regularization	0.001
	Minimum performance gradient	1e-7
Maximum mu	1e10	

network (BP-FFNN) is widely used [8], where the number of hidden layers varies from one [39, 40] to 2-3 [41] [18], or even more [42, 43]. The number of neurons in the input layer is determined by the number of (spectral) features as input. For the output layer, the number of neurons will be one for regression analysis of the predicted phenol concentration, or more for classification of the concentration into multiple levels or classifying the origins of the input samples. The number of neurons in the hidden layer is usually decided through comprehensive trial and errors, though it is often suggested to be less than that of the input layer. In our designed DNN, BP-FFNN is used as the training model, where the numbers of hidden layers and neurons are empirically determined based on the minimized errors of the trained models. By taking the fused features as input, the DNN model is capable of predicting the total phenol level of 27 barley categories. Detailed comparison of the parameters and functions used in these three models is presented in Table II. The statistics of the dataset used for training and testing is tabulated in Appendix: Table A.

For quantitative performance evaluation, several commonly used metrics were adopted in our experiments, which include the overall accuracy (OA)[44], Kappa coefficient (KP) [23], root mean squared error (RMSE) [45], mean absolute error (MAE) [46] and correlation coefficient (r^2) [47]. The OA is the percentage of overall classification accuracy and defined by

$$OA = \frac{1}{N} * \sum_i^T C_i * 100\% \quad (11)$$

The Kappa coefficient below is used to measure the inter-rater reliability of data classification:

$$Kappa = (P_o - P_e) / (1 - P_e) \quad (12)$$

$$P_0 = OA, P_e = \frac{1}{N^2} \sum_{i=1}^T \left(\sum_{j=1}^T C_{ij} * \sum_{j=1}^T C_{ji} \right) \quad (13)$$

where N , T and C_i denote respectively the number of spectral samples, classes, and correctly classified spectral samples in class i . C_{ij} is the row element and C_{ji} is the column element in the confusion matrix.

The RMSE, MAE and r^2 are three popular metrics for evaluation the regression results. For N spectral samples, their definitions are given by:

$$RMSE = \sqrt{\frac{1}{N} * \sum_{i=1}^N (y_i - y'_i)^2} \quad (14)$$

$$MAE = \frac{1}{N} * \sum_{i=1}^N (|y_i - y'_i|) \quad (15)$$

$$r^2 = 1 - \frac{\sum_{i=1}^N (y_i - y'_i)^2}{\sum_{i=1}^N (y_i - y''_i)^2} \quad (16)$$

where y_i and y'_i represent the actual value and predicted value of the i^{th} Sample, respectively; y''_i is the mean of all y_i values. The smaller the value of RMSE and MAE is, the higher the prediction accuracy of the model is. The range of r^2 is within $[0,1]$, with 1 indicating 100% prediction accuracy.

IV. RESULTS AND DISCUSSION

In this section, comprehensive experiments are carried out for assessment of the phenolic level in peated barley malts. For robustness, data selection stage was repeated 10 times to randomly generate 10 groups of spectral samples for training and testing, and the averaged testing results are reported for evaluation. Detailed experimental results are presented below.

A. Results of classification

1) Data acquisition analysis

In the proposed work, each phenol level has one training data pool and one testing data pool. In our experiments, it is found that different rotation angles of the container may affect the classification results. As shown in Table III, with the size of pixel subsets $s=200$ using only the raw data, the classification results under various rotation angles vary, where the maximum difference between 270° and 90° is 10.31%. There are two possible reasons: 1) inconsistent ambient light and 2) uneven phenol distribution on the peated barley malts' surface in each container. To reduce these two negative factors, two major improvements have been introduced in this work. First, each container is scanned 4 times at 4 different rotation angles (i.e. 0° , 90° , 180° , 270°), resulting in 4 hypercubes. These hypercubes are stacked together to form the training dataset or the testing dataset. As shown in Table III, the classification accuracy has been much improved with the introduced stacking process, validating the value of our data acquisition strategy.

TABLE III

RESULTS FROM RAW DATA IN DIFFERENT WAYS OF DATA ACQUISITION.

Data strategy	0°	90°	180°	270°	Stacking
OA (%)	83.35±1.41	85.33±1.08	80.63±1.03	75.40±1.14	90.35±1.07
Kappa (%)	82.71±1.47	84.76±1.12	79.88±1.07	74.45±1.18	89.97±1.12

2) Key parameters setting

In our experiments, the selection of q_{PCA} and q_{FPCA} are determined by varying them within the range of 10 to 100 with a step of 10. The optimal values for them are found to both be 30. The selection of pixel subsets s are evaluated and see how s affects the classification results. The average over 50, 100, 200, 400, 600, 800, 1000, 2000 pixels is taken to generate spectral samples for classification, resulting in 27100, 13552, 6750, 3383, 2256, 1693, 1350, 675 spectral samples respectively for training or testing.

In Table IV, the overall classification accuracy (OA) with the standard deviation of our system with the two different settings (i.e. RAW and PRE+FF) is presented. As can be seen, a smaller s leads to a lower classification accuracy, which fits to some parts of our previous findings [23]. As the phenol distribution on the barley surface is not a constant, taking the mean spectrum over a number of subsets tends to represent the average concentration and introduce sufficient statistic of the whole batch of peated barley malts. In general, larger subsets are highly likely to represent more sufficient statistics and more precise the overall concentration of the patch and result in better classification. However, the accuracy of our previous method [23] starts to decrease when the number of subsets is larger than 200, due possibly to two main reasons. The first is different experimental settings, as we choose the training and testing spectral samples from two hypercubes rather than one. The second is lack of effective feature extraction, as the proposed fusion strategy can help to extract more representative features for better data classification.

In our system, when the number of subsets is too large ($s = 2000$), the classification accuracy with two settings also drops. The possible reason is that a larger s may lead to less spectral samples, which may cause the SVM classifier under-trained. As a result, the OA becomes lower. In addition, when s is 600, 800 or 1000, the OA of PRE+FF is not as consistently increasing as that of RAW[23]. This is because the accuracy has already reached the peak with $s = 600$, hence the fluctuation at $s = 800$ or $s=1000$ is caused by random factor in SVM.

3) Key stage analysis

As the proposed data processing framework of phenolic

TABLE IV

OA (%) OF PROPOSED FRAMEWORK WITH TWO SETTINGS. RAW, PRE AND FF DENOTES THE RAW DATA, PRE-PROCESSED DATA AND FEATURE FUSION. THE ABBREVIATION WILL BE USED IN THE REST TABLES.

Number of subsets (s)	RAW	PRE+FF
50	85.68±0.44	92.52±0.86
100	89.44±0.76	95.96±0.59
200	90.35±1.07	98.32±0.55
400	89.46±0.86	98.72±0.65
600	89.58±1.02	99.50± 0.62
800	89.12±1.64	99.16±0.66
1000	88.47±1.45	99.20±0.69
2000	87.19±1.53	96.29±2.05

level measurement is a multi-stage approach, the contributions of the five major stages are assessed as follows. These will include the analysis of 8 different settings, i.e. raw imaged data with SNV (RAW) [23], pre-processed data (PRE), individual PCA feature extracted from raw data (RAW+PCA), individual FPCA feature extracted from raw data (RAW+FPCA), and feature fusion for raw data (RAW+FF), individual PCA feature extracted from pre-processed data (PRE+PCA), individual FPCA feature extracted from pre-processed data (PRE+FPCA), and feature fusion for pre-processed data (PRE+FF). For the RAW setting, it is also the same as the one in [23].

Table V shows the OA and Kappa with standard deviation obtained from eight settings with $s = 200$. For the raw data, the results with OA at 90.35% seems undesirable. After applying PCA or FPCA directly on the raw data, the OA improves by 2.98% and 1.87%, and reaches 93.33% and 92.22% respectively. By further feature fusion of PCA and FPCA, the OA can be increased to 93.59%. As seen, feature fusion $X(FF) \in \mathcal{R}^{n \times (q_{PCA} + q_{FPCA})}$ gives better classification performance than using either PCA or FPCA individually, because fusing $X(PCA)$ and $X(FPCA)$ together enables

extraction of both global and local spectral features for more effective data modelling and prediction. When applying the pre-processing for the raw data, the gain on OA varies. For PRE, PRE+PCA, PRE+FPCA and PRE+FF, the gains are 0.96%, 3.22%, 4.10%, and 4.73% more than RAW, RAW+PCA, RAW+FPCA, RAW+FF, respectively. This has validated the

TABLE V
CLASSIFICATION RESULTS OF THE PROPOSED FRAMEWORK WITH
VARIOUS SETTINGS.

Stages	OA (%)	Kappa (%)
RAW	90.35±1.07	89.97±1.11
PRE	91.31±1.10	90.97±1.14
RAW+PCA	93.33±0.99	93.07±1.03
RAW+FPCA	92.22±0.64	91.92±0.66
RAW+FF	93.59±0.82	93.35±0.75
PRE+PCA	96.55±0.68	96.42±0.71
PRE+FPCA	96.32±0.60	96.18±0.62
PRE+FF	98.32±0.55	98.26±0.57

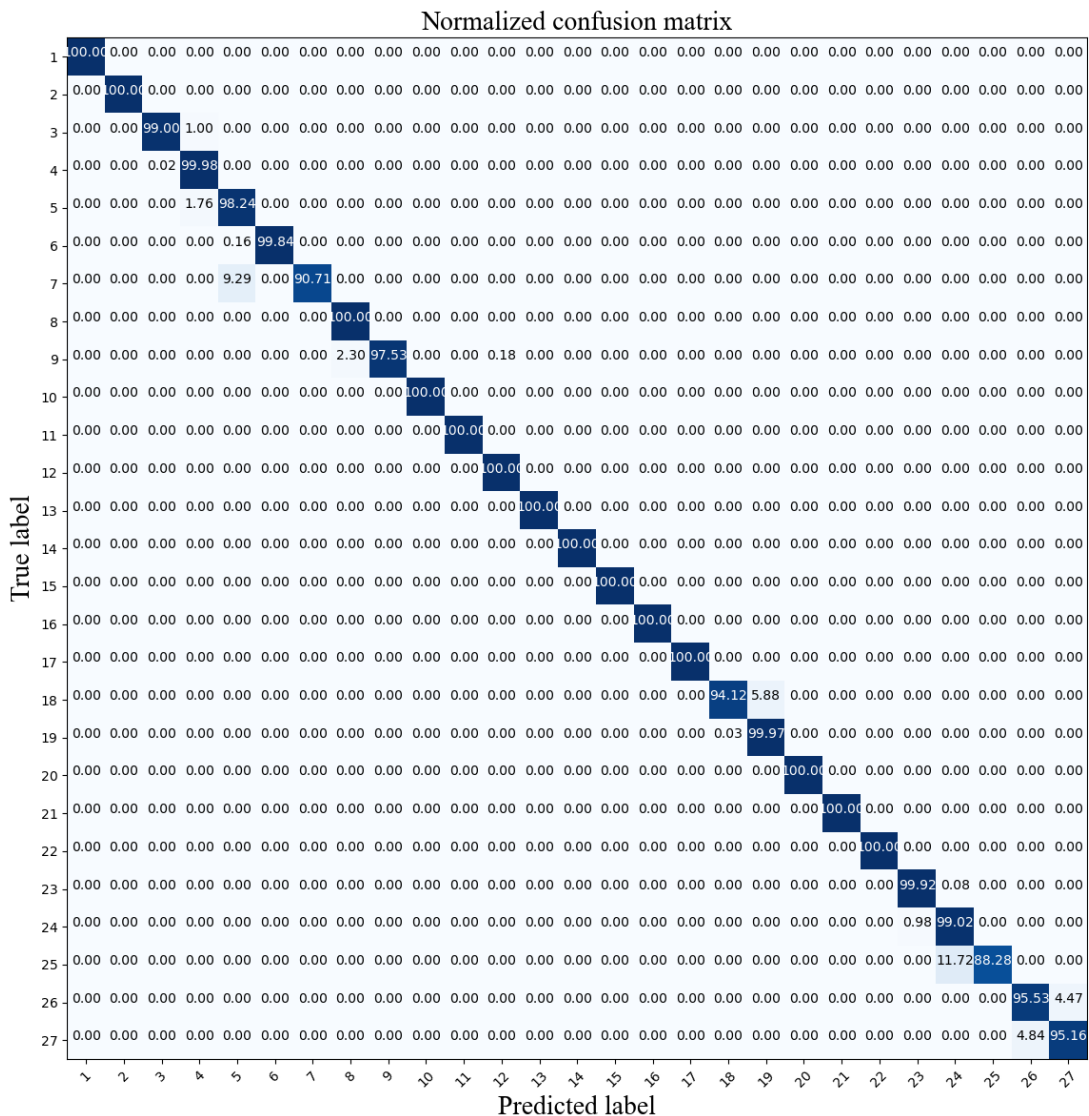


Fig. 6. Confusion matrix of classification results with $s=200$ and the stage of PRE+FF.

importance of our pre-process. Finally, with $s = 200$, the classification accuracy can reach 98.32% with a standard deviation of 0.55%, i.e. an additional gain of 7.97% (0.53%) against the raw data. To this end, the contribution of each key stages has been clearly demonstrated.

To gain a more detailed insight, Fig. 6 shows the confusion matrix of the classification results at the stage of PRE+FF with $s=200$. The percentage of correctly classified pixels are listed on the anti-diagonal, and all other values indicate misclassified pixels. As shown in Table I, the categories of 2, 20 and 21 are from three suppliers, but their total phenol levels are very similar to each other, although they are different in terms of the concentrations of individual phenolic compounds. As seen in Fig. 6, the classification results of these three categories can reach 100%, i.e. no misclassification, which means our model can well distinguish the peated barley malts from different suppliers, which can be very useful for more accurate flavour control in the whisky industry. Similarly, the categories 1 and 18 are from Maltster A and Distillery D respectively. The difference of their total concentration is only 0.6ppm, but the differences of their 7 individual measurements are 4.9, 2.7, 1.3, 1.3, 0.2, 1.4, 0.5 ppm, respectively. Therefore, they have very unique peat flavour. For the category 1, it can be well classified with an accuracy of 100%. For the category 18, the classification accuracy is 94.12%, where 5.88% of the pixels are misclassified into the category 17, which is the same Distillery D. This has indicated that different categories of peated barley malts from the same supplier may have been misclassified each other. However, these will not affect the results of origin discrimination, which can be also observed for classifications of the categories 23-27 from Maltster F and the categories 5-7 from Distillery A et al.

In summary, our model can accurately discriminate the origins or suppliers of the peated barley malts despite of the very similar levels of the total phenols. Although there are misclassified cases, they are tended to be from the same supplier. To this end, the request of origin discrimination can still be fully satisfied.

4) Extended experiment on background analysis

In this section, we further investigate how the background may affect the classification of the peated barley malts in various phenolic concentrations from different suppliers. Here, the background is treated as another class rather than being eliminated from the scene. Followed by the same procedure of data selection, in total 24000 background samples and 6750 barley samples are generated in the training and testing dataset with $s=200$. In addition, three conventional and four deep learning models are used for benchmarking, including PCA [27], Folded PCA (FPCA) [32], 1-D Singular Spectrum Analysis (1DSSA) [48], Deep convolutional neural network (CNN) [49], Stacked auto-encoder (SAE) [50], Deep recurrent neural network (RNN) [51], and Auto-CNN [52]. These state-of-the-art approaches are selected for two main reasons, i.e. highly related to our proposed method, and/or widely cited spectral feature extraction approaches. For the four deep learning models, we fine-tune the model on our data. For PCA

TABLE VI
CLASSIFICATION RESULTS OF OUR METHOD AND SEVEN BENCHMARKING APPROACHES BY TAKING THE BACKGROUND AS A NEW CLASS

Methods	AA (%)	OA (%)	KP (%)
PCA	90.68±0.60	96.02±0.26	93.85±0.40
FPCA	90.26±0.30	95.86±0.13	93.61±0.20
1DSSA	89.37±0.45	95.48±0.19	93.02±0.30
CNN	54.11±10.89	80.47±4.65	69.84±7.18
SAE	56.96±1.26	81.68±0.54	71.70±0.84
RNN	82.76±1.50	92.67±0.63	88.67±0.98
Auto-CNN	78.49±11.04	90.85±4.72	85.87±7.29
Proposed	93.73±0.88	97.32±0.37	95.86±0.58

and FPCA, the number of principal components is set to 30. For the rest parameters, the suggested default settings are used.

From the results presented in Table VI, three important findings can be highlighted below.

- 1) Although deep learning models may have better classification performance on some public HSI datasets, they show inferior results than conventional models in our experiments. In addition, deep learning models usually rely on heavy computation source and long training time, which seem inefficient for our specific problem. This may also explain why many food inspection works [17, 18, 20-22] prefer to use conventional machine learning models rather than deep learning.
- 2) After introducing the background as another class per your constructive suggestion, more misclassification cases occur inevitably, leading to degraded OA, KP and average accuracy (AA) of our proposed method which are reduced from 98.32%, 98.26% and 98.32% to 97.32%, 95.86% and 93.73%, respectively. Due to a larger number of spectral samples in the background, i.e. unbalanced pixels in different classes, the misclassification causes few negative effects to the OA but a much reduced AA, resulting in poor performance of origin discrimination. This also explains the necessity of eliminating the background from the acquired data through extraction of the ROIs.
- 3) Comparing with three conventional classification models, our proposed model has the best classification performance, which validates the effectiveness of our model even without removing the background.

B. Predicted Results of Phenolic Level from Regression

For the benefit of the whisky industry, it is always interested to estimate the actual phenol level of peated barley malts from different maltsters/distilleries. Therefore, the performance of regression is also assessed on the same data. Likewise, with $s = 200$, the result of SVR and key stage analysis is performed first, followed by analysis of the results of DNN and RF as follows.

TABLE VII
PERFORMANCE OF DIFFERENT DATA ACQUISITION STRATEGY USING RAW DATA ONLY.

Data strategy	0°	90°	180°	270°	Stacking
r^2 (%)	89.02±0.55	90.79±0.33	83.79±1.33	82.65±1.62	91.98±0.10
MAE/ppm	6.56±0.17	5.99±0.11	6.70±0.13	7.04±0.17	5.64±0.04
RMSE/ppm	8.17±0.20	7.48±0.13	9.98±0.41	10.30±0.47	6.99±0.04

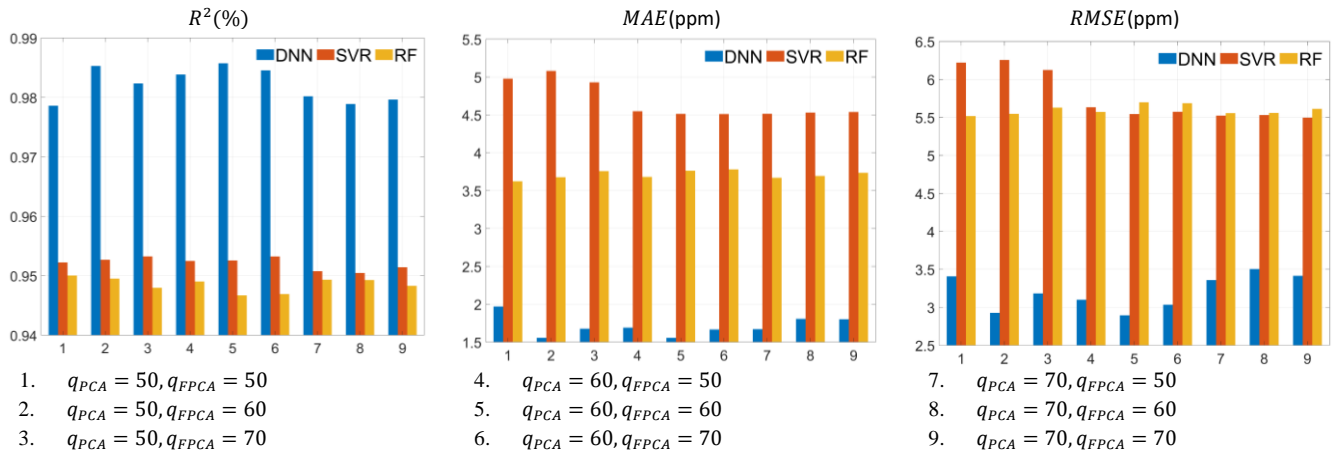


Fig. 8. Comparison of the selection of q_{PCA} and q_{FPCA} in FF for three regressors.

1) Data acquisition analysis

When only using the raw data, the stacked hypercubes show better performance than individual hypercube (Table VII), indicating that stacked hypercubes can help to mitigate the issues of inconsistent ambient light and uneven phenol distribution to some extent, leading to a higher r^2 and lower MAE and RMSE. Meanwhile, the standard deviation of each criteria is also reduced, which validates again the robustness of proposed data acquisition method.

2) Key stage analysis

For the two key parameters q_{PCA} and q_{FPCA} , they are experimentally determined in [10,100] with a step of 10. Eventually, the optimal values are found to be 60 for both of them, where SVR can produce the best prediction results.

In Table VIII, the contribution of each stage in the proposed framework to regression is compared. As seen, pre-processing and feature extraction both bring the benefit to the estimated phenol level. By fusion of PCA and FPCA [32] ($q_{PCA} = 60$ and $q_{FPCA} = 60$), the best regression in terms of r^2 , MAE and RMSE can reach 95.32%, 4.51ppm and 5.57ppm, respectively. It is worth noting that, the standard deviation of each stage is very low, showing the robustness of the process in each stage.

To further explore the best regression results, two more regression methods, i.e. deep neural network (DNN) and random forest (RF), are employed in this study. In addition, the

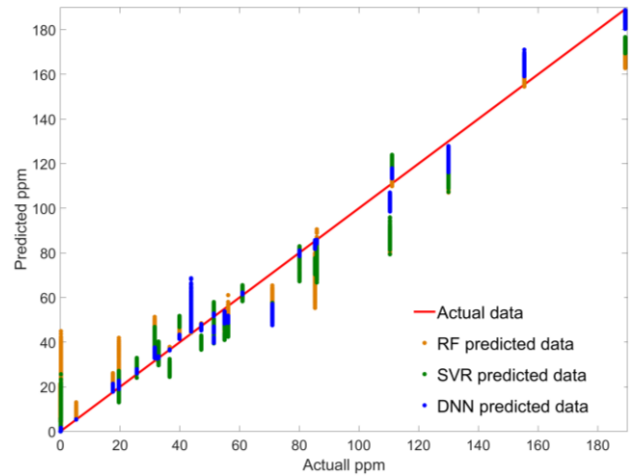


Fig. 7. Scatter plot of actual vs. predicted phenol level.

selection of q within PCA and FPCA for these regression methods is also investigated. There are three hidden layers in DNN where the number of each layer is set to 64, 32 and 16. The number of nodes in RF is set to 100. As seen in Fig. 7, the SVR is not the best regression model in this study. Compared with SVR, RF has a less MAE but also a lower r^2 . SVR and RF have a quite comparable RMSE, though it varies with the selected q_{PCA} and q_{FPCA} in FF. For DNN, it significantly outperforms the other two regressors regardless of parameter settings. As also shown in Fig. 8, the predicted ppm value of from DNN is better than those of SVM and RF. The best-performed DNN is the 5th combination, which has a r^2 value of 98.57%, a RMSE of 2.89ppm and a MAE of 1.57ppm.

Furthermore, the spatial map of different phenolic levels is shown in a compact manner (Fig. 9). We cropped 27 regions of various predicted phenolic levels in Fig. 9 (a) and stack them together along the Y-axis in Fig. 9 (b). As there are certain gaps between the pieces of peated barley malt, we use the black color to represents these gaps. As can be seen, different phenolic levels in the testing data can be well predicted.

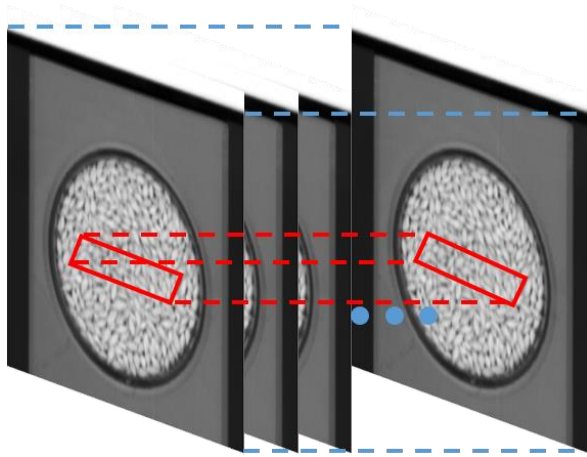
C. Computational complexity

In the proposed NIR-HSI phenolic measurement and origin discrimination system, spectral-spatial fusion based feature extraction and prediction is actually the key. In the pre-

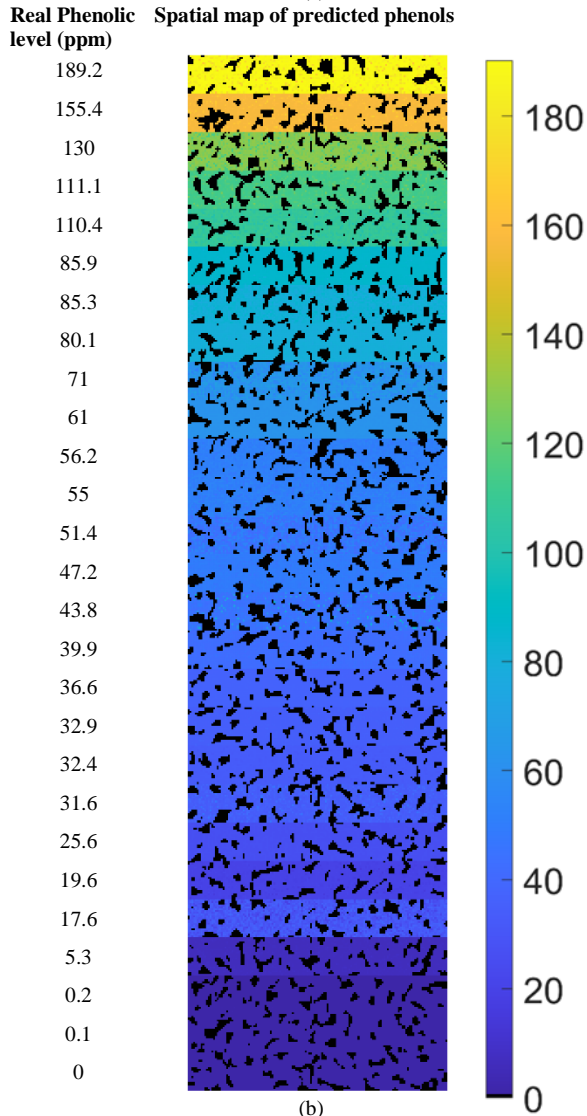
TABLE VIII

REGRESSION RESULTS OF THE PROPOSED FRAMEWORK WITH VARIOUS SETTINGS.

Stages	r^2 (%)	MAE (ppm)	RMSE (ppm)
RAW	91.98±0.10	5.64±0.04	6.99±0.04
PRE	93.72±0.09	5.14±0.03	6.19±0.04
RAW+PCA	91.79±0.06	5.68±0.04	7.07±0.03
RAW+FPCA	92.42±0.06	5.43±0.02	6.80±0.03
RAW+FF	92.88±0.08	5.60±0.04	6.99±0.04
PRE+PCA	94.06±0.07	4.94±0.04	6.03±0.04
PRE+FPCA	94.34±0.05	4.79±0.03	5.88±0.03
PRE+FF	95.32±0.06	4.51±0.06	5.57±0.06



(a)



(b)

Fig. 9. Illustration of phenolic spatial map

processing stage, the acquired data is rectified in both spectral and spatial domains using SNV and JBF, respectively. In the feature extraction stage, FPCA and PCA has helped to extract both the local and global spectral features, respectively. In this subsection, we briefly analyze the computational complexity of

TABLE IX
COMPUTATIONAL COMPLEXITY OF DIFFERENT APPROACHES

Stage	Computational complexity	Detailed description Note: I, J and B denote the spatial width, height and spectral bands of the hypercube, respectively.
JBF	$IJB^2 + B^3 + 2IJB$	Derived from fast implementation of Bilateral filter and PCA [2]
SNV	IJ	Pixel-wise normalization
PCA	$NB^2 + B^3 + NBq_{PCA}$	PCA on N selected samples, where $N \ll IJ$. q_{PCA} is the number of principal components in PCA
FPCA	$NBW + W^3 + NWq_{FPCA}$	FPCA on N selected samples, where $N \ll IJ$. q_{FPCA} is the number of principal components in FPCA, W is the number of bands in H groups, where $H \times W = B$

these approaches in Table IX, where I, J and B represent respectively the spatial width, height and number of bands of the hypercube, respectively. As seen, the most time-consuming part is the JBF, and the least part is SNV. The latency caused by the computation may affect the real-time response speed of the system. Nevertheless, our proposed system can still be considered to be near real-time, as the processing time is much less than the operation time of data acquisition especially with a push-broom HSI imaging system.

V. CONCLUSION

In this paper, we have developed a workable solution for origin discrimination and phenolic measurement of the peated barley malts using NIR hyperspectral imagery and machine learning. In total, 27 categories of peated barley malt from 8 suppliers covering a concentration range of 0 to 189.2 ppm are used for testing. As the original spectra features cannot explain the concentration of peated barley malt, we proposed a multi-stage data processing framework to project the whole spectral profile into the PCA and folded-PCA domains to extract the local and global spectral features, followed by various machine learning models for classification of the origins of the barley samples and prediction of the phenolic levels. The proposed new way of data acquisition is found to effectively overcome the inconsistency caused by varying lighting conditions and uneven phenol distributions on the peated barley malt surface. In addition, the fusion of PCA and folded-PCA has significantly improved the efficacy of feature extraction and data analysis, where the classification accuracy has been dramatically improved with a much reduced regression error. Specifically, the overall classification accuracy is up to 99.50% and the r^2 value of regression analysis has achieved 98.57% with a RMSE of 2.89ppm and MAE of 1.57ppm. With these promising results, the proposed system can be applied for traceability and quality control of peated barley malt and further benefit the Scotch whisky as the experiments are designed in accordance with real industrial conditions.

There are still some challenges and limitations of the proposed system as summarized below, which will be the focus of our future work. First, the sample selection process will affect the performance of our system. As seen in Table IV, a smaller number of subsets can yield sufficient number of

sampled pixels, but they can potentially fail to fully represent the spectral information of the peated barley malt. A larger number of subsets can extract the global spectral information, resulting in insufficient spectral samples though. Both these two situations will cause inferior performance. To address this limitation, some state-of-the-art data argumentation techniques [53] can be used to enhance the quantity and quality of the data and further improve the anti-overfitting ability, stability and robustness of the system. Some deep-learning-based feature extraction methods [54] can also be employed to extract the deep spectral features for improved performance. Second, accurately predicting of unseen categories of peated barley malt is needed in Whisky industry, which is crucial for practical deployment. Due to the limited data quantity and uneven distribution of the phenol concentrations, this challenge cannot be fully addressed without expanding the data samples. Two possible solutions can be applied to tackle this particular issue, one is to enrich the phenol concentrations of the collected barley samples, and the other is to further refine the developed machine learning models, such as the incremental learning [55, 56], for more robust modelling and prediction even with new categories of data.

ACKNOWLEDGEMENTS

We thank Mr. Ming Gong from University of Strathclyde for his assistance in modelling some benchmarking methods.

REFERENCE

- [1] R. F. Kokaly and A. K. Skidmore, "Plant phenolics and absorption features in vegetation reflectance spectra near 1.66 μm ," *Int. J. of Applied Earth Observation and Geoinformation*, vol. 43, pp. 55-83, 2015.
- [2] T. Qiao *et al.*, "Joint bilateral filtering and spectral similarity-based sparse representation: A generic framework for effective feature extraction and data classification in hyperspectral imaging," *Pattern Recognition*, vol. 77, pp. 316-328, 2018.
- [3] C. N. Frizon *et al.*, "Determination of total phenolic compounds in yerba mate (*Ilex paraguariensis*) combining near infrared spectroscopy (NIR) and multivariate analysis," *LWT-Food Sci. and Tech.*, vol. 60, no. 2, pp. 795-801, 2015.
- [4] T. Adão *et al.*, "Hyperspectral imaging: A review on UAV-based sensors, data processing and applications for agriculture and forestry," *Remote Sensing*, vol. 9, no. 11, p. 1110, 2017.
- [5] A. Vali, S. Comai, and M. Matteucci, "Deep learning for land use and land cover classification based on hyperspectral and multispectral earth observation data: A review," *Remote Sensing*, vol. 12, no. 15, p. 2495, 2020.
- [6] M. Shimoni, R. Haelterman, and C. Perneel, "Hyperspectral imaging for military and security applications: Combining myriad processing and sensing techniques," *IEEE Geosci. and Remote Sens. Magaz.*, vol. 7, no. 2, pp. 101-117, 2019.
- [7] E. Bedini, "The use of hyperspectral remote sensing for mineral exploration: A review," *J. of Hyperspectral Remote Sensing*, vol. 7, no. 4, pp. 189-211, 2017.
- [8] D. Saha and M. Annamalai, "Machine learning techniques for analysis of hyperspectral images to determine quality of food products: A review," *Current Research in Food Science*, 2021.
- [9] A. ul Rehman and S. A. Qureshi, "A review of the medical hyperspectral imaging systems and unmixing algorithms' in biological tissues," *Photodiagnosis and Photodynamic Therapy*, p. 102165, 2020.
- [10] R. Qureshi, M. Uzair, etc, "Hyperspectral document image processing: Applications, challenges and future prospects," *Pattern Recognition*, vol. 90, pp. 12-22, 2019.
- [11] C. Cucci and A. Casini, "Hyperspectral imaging for artworks investigation," in *Data Handling in Science and Technology*, vol. 32: Elsevier, 2020, pp. 583-604.
- [12] M. J. Khan, H. S. Khan, A. Yousaf, K. Khurshid, and A. Abbas, "Modern trends in hyperspectral image analysis: A review," *IEEE Access*, vol. 6, pp. 14118-14129, 2018.
- [13] S. W. Association, "Facts & Figures about Scotch Whisky," 2018. [Online]. Available: <https://www.scotch-whisky.org.uk/insights/facts-figures/>.
- [14] B. M. Harrison and F. G. Priest, "Composition of Peats Used in the Preparation of Malt for Scotch Whisky Production Influence of Geographical Source and Extraction Depth," *Journal of Agricultural and Food Chemistry*, vol. 57, no. 6, pp. 2385-2391, 2009.
- [15] B. Harrison, et al, "Differentiation of peats used in the preparation of malt for Scotch whisky production using Fourier transform infrared spectroscopy," *Journal of the Institute of Brewing*, vol. 112, no. 4, pp. 333-339, 2006.
- [16] T. A. Bringhurst and J. Brosnan, "Scotch whisky: raw material selection and processing," in *Whisky*: Elsevier, 2014, pp. 49-122.
- [17] A. Noviyanto and W. H. Abdulla, "Honey botanical origin classification using hyperspectral imaging and machine learning," *Journal of Food Engineering*, vol. 265, p. 109684, 2020.
- [18] C. Erkinbaev, K. Derksen, and J. Paliwal, "Single kernel wheat hardness estimation using near infrared hyperspectral imaging," *Infrared Phys. & Tech.*, vol. 98, pp. 250-255, 2019.
- [19] Z. Han and J. Gao, "Pixel-level aflatoxin detecting based on deep learning and hyperspectral imaging," *Computers and Electronics in Agriculture*, vol. 164, p. 104888, 2019.
- [20] W. Tan *et al.*, "Study on bruising degree classification of apples using hyperspectral imaging and GS-SVM," *Optik*, vol. 154, pp. 581-592, 2018.
- [21] N. Zhang *et al.*, "Determination of total iron-reactive phenolics, anthocyanins and tannins in wine grapes of skins and seeds based on near-infrared hyperspectral imaging," *Food Chemistry*, vol. 237, pp. 811-817, 2017.
- [22] J.-Y. Choi, S. Heo, etc, "Discriminating the origin of basil seeds (*Ocimum basilicum* L.) using hyperspectral imaging analysis," *LWT*, vol. 118, p. 108715, 2020.
- [23] J. Tschannerl *et al.*, "Potential of UV and SWIR hyperspectral imaging for determination of levels of phenolic flavour compounds in peated barley malt," *Food Chemistry*, vol. 270, pp. 105-112, 2019.
- [24] J. Tschannerl, J. Ren, etc, "Employing NIR-SWIR hyperspectral imaging to predict the smokiness of scotch whisky," in *Proc. OCM, 2017: KIT Scientific Publish.*, p. 1.

- [25] M. Robnik-Šikonja and I. Kononenko, "Theoretical and empirical analysis of ReliefF and RReliefF," *Machine Learning*, vol. 53, no. 1-2, pp. 23-69, 2003.
- [26] Å. Rinnan, F. Van Den Berg, and S. B. Engelsen, "Review of the most common pre-processing techniques for near-infrared spectra," *TrAC Trends in Analytical Chemistry*, vol. 28, no. 10, pp. 1201-1222, 2009.
- [27] H. Abdi and L. J. Williams, "Principal component analysis," *Wiley Interdisciplinary Reviews: Computational Statistics*, vol. 2, no. 4, pp. 433-459, 2010.
- [28] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in *Sixth International Conference on Computer Vision*, 1998: IEEE, pp. 839-846.
- [29] J. Zabalza *et al.*, "Novel segmented stacked autoencoder for effective dimensionality reduction and feature extraction in hyperspectral imaging," *Neurocomputing*, vol. 185, pp. 1-10, 2016.
- [30] J. Ren, J. Zabalza, S. Marshall, and J. Zheng, "Effective feature extraction and data reduction with hyperspectral imaging in remote sensing," *IEEE Signal Processing Magazine*, vol. 31, no. 4, pp. 149-154, 2014.
- [31] J. Zabalza, J. Ren, J. Ren, Z. Liu, and S. Marshall, "Structured covariance principal component analysis for real-time onsite feature extraction and dimensionality reduction in hyperspectral imaging," *Applied Optics*, vol. 53, no. 20, pp. 4440-4449, 2014.
- [32] J. Zabalza *et al.*, "Novel folded-PCA for improved feature extraction and data reduction with hyperspectral imaging and SAR in remote sensing," *ISPRS Journal of Photo. and Rem. Sen.*, vol. 93, pp. 112-122, 2014.
- [33] T. Qiao, J. Ren, C. Craigie, *et al.*, "Quantitative prediction of beef quality using visible and NIR spectroscopy with large data samples under industry conditions," *J. of Applied Spectroscopy*, vol. 82, no. 1, pp. 137-144, 2015.
- [34] H. Sun, J. Ren, H. Zhao, Y. Yan, *et al.*, "Superpixel based feature specific sparse representation for spectral-spatial classification of hyperspectral images," *Remote Sens.*, vol. 11, no. 5, p. 536, 2019.
- [35] C.-l. Wang, J. Ren, H.-w. Wang, *et al.*, "Spectral-spatial classification of hyperspectral data using spectral-domain local binary patterns," *Multimedia Tools and Applications*, vol. 77, no. 22, pp. 29889-29903, 2018.
- [36] C.-C. Chang, C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. on Intelligent Systems and Technology (TIST)*, vol. 2, no. 3, pp. 1-27, 2011.
- [37] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5-32, 2001.
- [38] S. Ma *et al.*, "Image and video compression with neural networks: A review," *IEEE Trans. on Circuits and Systems for Video Techno.*, vol. 30, pp. 1683-1698, 2019.
- [39] N. Zhu, K. Wang, S.-L. Zhang, B. Zhao, J.-N. Yang, and S.-w. Wang, "Application of artificial neural networks to predict multiple quality of dry-cured ham based on protein degradation," *Food Chemistry*, vol. 344, p. 128586, 2021.
- [40] N. González-Viveros, P. Gómez-Gil, J. Castro-Ramos, and H. Cerecedo-Núñez, "On the estimation of sugars concentrations using Raman spectroscopy and artificial neural networks," *Food Chemistry*, p. 129375, 2021.
- [41] Y. Cao *et al.*, "Identification of species and geographical strains of *Sitophilus oryzae* and *Sitophilus zeamais* using the visible/near - infrared hyperspectral imaging technique," *Pest Management Sci.*, vol. 71, no. 8, pp. 1113-1121, 2015.
- [42] H. Rastegarfar *et al.*, "TCP flow classification and bandwidth aggregation in optically interconnected data center networks," *J. Optical Comm. and Networking*, vol. 8, no. 10, pp. 777-786, 2016.
- [43] D. A. Pamplona, L. Weigang, A. G. de Barros, E. H. Shiguemori, and C. J. P. Alves, "Supervised Neural Network with multilevel input layers for predicting of air traffic delays," in *Int. Joint Conf. on Neural Networks (IJCNN)*, 2018: IEEE, pp. 1-6.
- [44] Y. Sun *et al.*, "Hyperspectral reflectance imaging combined with chemometrics and successive projections algorithm for chilling injury classification in peaches," *LWT*, vol. 75, pp. 557-564, 2017.
- [45] J. Li, "Support vector regression for the determination of the nutritional components of edible oil by terahertz spectroscopy," *IEEE Trans. Instrum. and Measur.*, vol. 59, no. 8, pp. 2094-2098, 2009.
- [46] I. Ahmad, A. Singh, M. Fahad, M. Waqas, "Remote sensing-based framework to predict and assess the interannual variability of maize yields in Pakistan using Landsat imagery," *Computers and Electronics in Agriculture*, vol. 178, p. 105732, 2020.
- [47] J. Chen *et al.*, "Multivariate regression model for industrial process measurement based on double locally weighted partial least squares," *IEEE Trans. Instrum. and Measur.*, vol. 69, no. 7, pp. 3962-3971, 2019.
- [48] J. Zabalza, *et al.*, "Fast implementation of singular spectrum analysis for effective feature extraction in hyperspectral imaging," *IEEE J. Selected Topics in Applied Earth Observ. Remote Sens.*, vol. 8, no. 6, pp. 2845-2853, 2014.
- [49] W. Hu, Y. Huang, L. Wei, F. Zhang, and H. Li, "Deep convolutional neural networks for hyperspectral image classification," *Journal of Sensors*, vol. 2015, 2015.
- [50] L. Windrim, R. Ramakrishnan, A. Melkumyan, R. J. Murphy, and A. Chlingaryan, "Unsupervised feature-learning for hyperspectral data with autoencoders," *Remote Sensing*, vol. 11, no. 7, p. 864, 2019.
- [51] L. Mou, P. Ghamisi, and X. X. Zhu, "Deep recurrent neural networks for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 7, pp. 3639-3655, 2017.
- [52] Y. Chen, K. Zhu, L. Zhu, *et al.*, "Automatic design of convolutional neural network for hyperspectral image classification," *IEEE Trans. Geosci. and Remote Sens.*, vol. 57, no. 9, pp. 7048-7066, 2019.
- [53] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *Journal of Big Data*, vol. 6, no. 1, pp. 1-48, 2019.
- [54] B. Rasti *et al.*, "Feature Extraction for Hyperspectral Imagery: The Evolution From Shallow to Deep: Overview and Toolbox," *IEEE Geoscience and Remote Sensing Magazine*, vol. 8, no. 4, pp. 60-88, 2020.
- [55] J. Bai *et al.*, "Class Incremental Learning With Few-Shots Based on Linear Programming for Hyperspectral Image Classification," *IEEE Transactions on Cybernetics*, 2020.
- [56] Y. Liu, Y. Su, A.-A. Liu, B. Schiele, and Q. Sun, "Mnemonics training: Multi-class incremental learning

without forgetting," in *Proc. IEEE/CVF Conf. on Computer Vision and Pattern Recognition*, 2020, pp. 12245-12254.

BIOGRAPHY



Yijun Yan received his M.Sc. and Ph.D. degrees in electrical and electronic engineering from University of Strathclyde in 2013 and 2018, respectively. He is currently a research fellow with Robert Gordon University, Aberdeen, UK. His research interests include hyperspectral imaging, pattern recognition, computer vision and machine learning.



Jinchang Ren (Senior Member, IEEE) received the B.E. in computer software, in 1992, M.Eng. in image processing, in 1997, D.Eng. in computer vision, in 2000, all from Northwestern Polytechnical Univ., Xi'an, China, and the Ph.D. in electronic imaging & media communication, in 2009, from the Univ. of Bradford, Bradford, U.K.

He is currently a Professor of Computing Science, Robert Gordon University, Aberdeen, UK. He has published 300+ peer-reviewed articles, and acts as an Associate Editor for several international journals including IEEE TGRS and Journal of the Franklin Institute et al. His research interests include hyperspectral imaging, image processing, computer vision, big data analytics, and machine learning.



Julius Tschannerl was born in Munich, Germany. He received his Ph.D. degree in electrical and electronic engineering from University of Strathclyde in 2019. He is currently a design engineer in Apeel, Log Angeles, California, United States. His research interests include hyperspectral imaging, and machine learning.



Huimin Zhao was born in Shaanxi, China, in 1966. He received the B.Sc. and M.Sc. degrees in signal processing from NWPU, Xi'an, China, in 1992 and 1997, respectively, and the Ph.D. in electrical engineering from Sun Yat-sen University, Guangzhou, China, in 2001.

He is currently a Professor and Dean with the School of Computer Sciences, Guangdong Polytechnic Normal University, Guangzhou, China. His research interests include image/video and information security technologies, and applications.



Barry Harrison received his BSc degree in biochemistry from Edinburgh University and Ph.D. degree in whisky flavour chemistry from Heriot-Watt University. He is currently a senior scientist at the Scotch Whisky Research Institute. His research interests include food science, and analytical chemistry.



Frances Jack received her BSc degree in food science and Ph.D. degree in bioscience and biotechnology, all from University of Strathclyde. She is currently a whisky scientist and flavour/ sensory specialist at the Scotch Whisky Research Institute. Her research interests include food science, microbiology, and brewing.

APPENDIX

TABLE A
STATISTICS OF THE DATASET USED FOR TRAINING AND TESTING.

Category	DNN		SVM/RF	Testing	Total Phenols (ppm)
	Training	Validation	Training		
1	194	49	243	255	85.9
2	203	51	254	251	0.1
3	189	47	236	247	31.6
4	197	49	246	248	71.0
5	198	49	247	252	51.4
6	202	50	252	250	36.6
7	197	49	246	245	32.9
8	196	49	245	255	56.2
9	197	49	246	248	39.9
10	202	50	252	254	47.2
11	202	51	253	250	25.6
12	198	50	248	244	55.0
13	204	51	255	245	32.4
14	201	50	251	250	19.6
15	203	51	254	255	5.3
16	195	49	244	249	61.0
17	201	50	251	251	80.1
18	204	51	255	250	85.3
19	201	50	251	244	110.4
20	196	49	245	250	0.2
21	204	51	255	252	0.0
22	202	51	253	263	130.0
23	207	52	259	251	17.6
24	203	51	254	256	43.8
25	202	51	253	251	111.1
26	201	50	251	249	155.4
27	203	51	254	256	189.2
Mean	200	50	250	250	58.3
Median	201	50	251	250	49.3
1 st Qu	197	49	246	248	28.6
3 rd Qu	203	51	254	253	82.7
Min	189	47	236	244	0.0
Max	207	52	259	263	189.2
StdDev	3.9	1.0	4.9	4.2	47.1