ORIGINAL PAPER

A new honey adulteration detection approach using hyperspectral imaging and machine learning

Tessa Phillips[1] · Waleed Abdulla[1]

Abstract

This paper develops a new approach to fraud detection in honey. Specifically, we examine adulterating honey with sugar and use hyperspectral imaging and machine learning techniques to detect adulteration. The main contributions of this paper are introducing a new feature smoothing technique to conform to the classification model used to detect the adulterated samples and the perpetration of an adulterated honey data set using hyperspectral imaging, which has been made available online for the first time. Above 95% accuracy was achieved for binary adulteration detection and multi-class classification between different adulterant concentrations. The system developed in this paper can be used to prevent honey fraud as a reliable, low cost, data-driven solution.

Keywords Honey fraud detection · Honey adulteration · Hyperspectral imaging · Machine learning

Introduction

Honey fraud is a major global problem, with honey being the third most adulterated food product [4, 6, 18, 31]. Adulteration of any food product has severe implications for human health, and it is vital to detect any adulteration of food products [2].

There are existing measures to detect overall quality in premium honey types, such as New Zealand (NZ) Manuka honey [7]. Hyperspectral imaging has recently been developed to detect the botanical origins and quality of pure honey [23, 26, 27, 29]. This approach has not yet been thoroughly investigated for detecting adulterated honey.

Existing quality assurance techniques

Most honey botanical origins are classified using chemical measures or, more traditionally, honey experts taste and smell the honey. The chemical measurements include pollen analysis and tests for particular components that make up certain honey [25].

Several metrics have been developed using chemical processes to identify the quality of Manuka honey. Manuka honey is a premium honey product, so the higher quality types refer to a purer Manuka honey [18]. Unique Manuka Factor (UMF) is one such measure. This measure displays a quality rating resulting from several different quality tests. The second rating system is methylglyoxal concentration (MGO) which is directly the concentration of an active ingredient in Manuka honey. The rating systems are owned by competing Manuka honey companies and are misleading to the consumer as they display vastly different numerical values to indicate quality [18]. The key thing for consumers to understand is that the higher the UMF or MGO rating, the higher quality of the honey will be, and they should expect to pay more for it.

Spectroscopy has been investigated as a potential approach to classifying honey quality, both for mislabelling and adulteration detection, with some success on specified honey types, including Manuka honey [3, 15, 17]. The wavelength ranges used in these works were visible and near-infrared, ranging from 400 to 2500 nm. Using mid-infrared spectroscopy [15] determined that the peak wavelengths, between 1400 and 2000 nm, shifted when honey was adulterated.

✉ Tessa Phillips
  tphi709@aucklanduni.ac.nz

  Waleed Abdulla
  w.abdulla@auckland.ac.nz

1 Electrical, Computer, and Software Engineering, The University of Auckland, Auckland 1010, New Zealand

Several different types of sugar syrup were added to Manuka honey and captured using near-infrared spectroscopy, and aquaphotomics [41]. This work uses spectral bands between 1300*nm* and 1800*nm* and shows that the Manuka honey differs from adulterated Manuka honey at a wavelength of 1460*nm*. This work shows promise in using spectral approaches to classify adulteration with many different adulterants. Only one type of Manuka honey was used in the experiment, so it is uncertain if this approach will work for all honey or even other Manuka honey brands.

Several papers have investigated the use of Fourier transform infrared (FTIR) spectroscopy for detecting the adulteration of honey with cane sugar [12, 35]. These works evaluated adulteration between 0.5 and 25% cane sugar concentration. Only one type of honey was used in [12] which predicted the sugar concentration using statistical methods and artificial neural networks with an accuracy of 93.75%. The classification accuracy was below 80% when using three different honey types to classify adulteration [35]. These works show that it is possible to use spectroscopy with machine learning techniques to predict adulteration in honey; however, the ability to predict sugar concentration across a range of honey must be improved.

Hyperspectral imaging is a promising tool for food quality assurance and extends on spectroscopy, allowing spatial information to be used alongside spectral information [5]. Spatial information allows the image to indicate specific defects such as bruises on fruit in a particular location rather than just taking the spectrum at one point of the object [8]. Hyperspectral imaging has been used for many food quality applications, including meat, fish, fruit, vegetables, and grains [40]. Hyperspectral imaging has also been used to determine different properties of grapes [19, 32], and to predict properties of chocolate that indicate quality [9]. More recently, this technique has been applied to honey quality applications.

A Hyperspectral imaging approach has been developed for detecting the botanical origins of honey [17, 23, 24, 26, 27, 29]. In [27] the UMF grade of Manuka honey was predicted with 89% accuracy. The botanical origins of 21 different types of honey were predicted with 90% accuracy in [29]. These approaches used a hyperspectral imaging system to capture the data detailed in [21] and used a class embodiment autoencoder (CEAE) and support vector machines (SVM) for classification. In [17] a classification accuracy of 90% was achieved when classifying the botanical origins of a small data set containing just 52 samples from five honey types.

Previous research has been done on hyperspectral imaging to detect adulteration, although it is limited to a small data set and not publicly available [34]. This approach uses a classifier to calculate a percentage of pixels in the image as either sugar or honey. This approach indicates that the adulterated mixtures are not homogeneous, which most adequately adulterated honey would be.

The key advantage of using HSI over spectroscopy for this application is that many training examples can be captured in each hyperspectral image. This allows the use of machine learning models, which require large amounts of data to train. The disadvantage is that it is typically expensive compared to spectroscopy; however, it is still much cheaper than chemical techniques, which require experts to analyse each sample.

Spectral imaging for detecting adulterated substances

The theory states that the spectrums should add if the substances used are in unique wavelengths in a mixture [33]. Honey and sugar will not be so simple to determine adulteration, as honey contains many natural sugars. However, there is an example of this theory that seems to apply quite well when detecting added sugar in wine [36]. Promising studies [16, 33] are detecting the addition of different kinds of sugar in fruit juices. The problem they face is similar to ours, where there are already sugars in the juice, and they must detect the concentration of added sugar for the different juices. The spectral wavelengths used to determine adulteration of juice were in the range of visible to near-infrared [16]. The spectral responses showed that the primary response to the sugar adulteration happened at above 2000 nm. However, some wavelengths showed responses to the adulteration at 430 nm, 664 nm, 1154 nm, 1462 nm, and 1504 nm [16].

Aim and objectives

This work aims to develop a system to detect the concentration of sugar syrup in an adulterated sugar and honey solution.

The following objectives provide the pathway for achieving this aim:

- Develop a strategy for adulterating honey that could resemble what is done in the real-world
- Capture a data set of adulterated honey using hyperspectral imaging
- Analysis of the new data
- Develop a machine learning system to detect adulterated honey.

Our proposed solution is a machine learning and hyperspectral imaging system to classify the adulteration concentration in honey samples. This solution comprises a new data set of adulterated honey samples captured with a hyperspectral camera and a machine learning model trained on this data. A classifier can be trained on this data set using machine

learning to determine sugar concentration in honey. The idea for the model to be general to any honey type is valuable and can be deployed worldwide for different types of honey. The data-driven approach means that if new premium honey types are discovered, they can be easily added to the database making it more robust. Hyperspectral cameras are expensive; however, we can use spectroscopy [11], a lower cost option, for testing samples in a real-world setting. Hyperspectral imaging is advantageous over the spectroscopy techniques, as it provides complete spectral and spatial information of an entire sample in one scan, which we can use to train a machine learning model [21]. A testing example does not need many data points, meaning spectroscopy could easily capture a small number of spectral samples from a honey type for testing.

Materials and methods

This section details the methodologies we use to detect the level of adulteration in a hyperspectral capture of adulterated honey. Figure 1 illustrates the procedure we follow for sample preparation and capturing.

Sample preparation and adulteration

The sample preparation follows the process defined in [21], with a few extra steps to adulterate the samples.

The adulterant we are working with is sugar syrup with a ratio of 57% sugar and 43% water. This ratio best mimics the consistency of most honey in our initial trials. Sugar syrup is the only adulterant used in this data set. In theory,

this system can be applied to detect other adulterants with labelled data of adulterated honey mixtures, although further experimentation would be required to confirm this.

Samples of honey and sugar syrup are heated to 40 degrees Celsius in an oven, so the honey sample becomes homogenous, as per previous work [21]. This temperature is hot enough to melt and mix the honey of all types but not hot enough to damage the active ingredients, particularly Manuka honey.

Samples of pure honey and adulterated mixtures at 5%, 10%, 25%, 50% sugar syrup concentration are prepared. Seven grams of each mix is poured into a small petri dish and captured with the hyperspectral imaging system. The capture system uses a dome lighting system with halogen bulbs to ensure even broadband lighting over the samples [21]. The images of honey have been captured with a hyperspectral imager SOC-710 from Surface Optic. The hypercubes captured are between $400\ and\ 1000\ nm$ in wavelength with a $5\ nm$ increment and have a $520x696$ spatial resolution. This wide range of wavelengths has been used in previous work in this application [20, 21]. This adulteration and capture process has been repeated a total of six times.

Figure 2 shows an RGB image of the adulterated honey samples from two classes: Manuka and Clover. This figure shows that it is not apparent from an RGB image of the adulterated samples. The 50% and 25% adulterated samples look different to the 5% and 10% adulterated honey samples. However, the 5% and 10% adulterated samples are hard to differentiate between and look very similar to pure honey. This figure demonstrates that the human eye alone cannot determine if a honey sample is adulterated, and there is a need for a more scientific approach. Even replacing 5% of
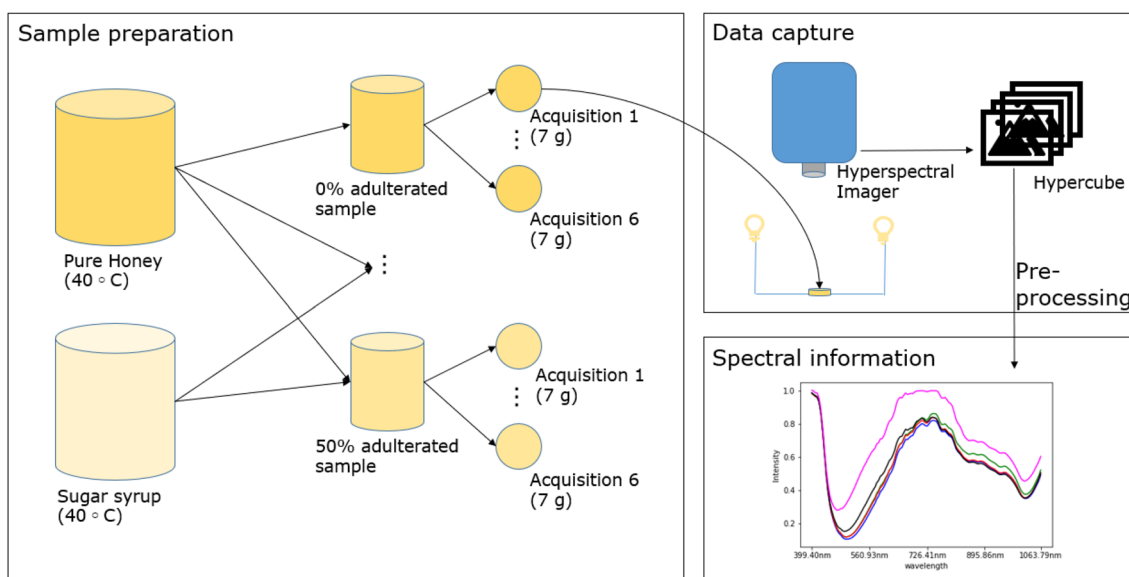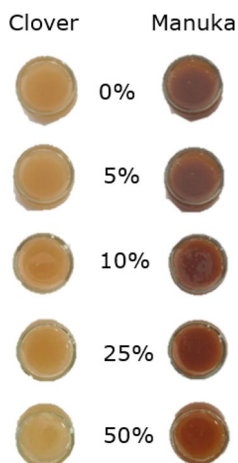

Fig. 1 Sample preparation process for one honey type in our database

Fig. 2 Adulterated honey samples of Manuka and Clover types

Fig. 3 Segmentation process for our data set, showing how the image is split into a five-by-five grid to obtain 25 samples

a premium honey type such as NZ Manuka honey with a cheaper substance will significantly reduce the cost of producing the honey. The approach needs to be sensitive enough to detect a minimal amount of adulteration.

Preprocessing and calibration

Because a hyperspectral imaging system measures the intensity of light reflected into the camera, the calibration process is crucial to obtaining consistent images. Calibration is performed using a dynamic white reference technique further detailed in [22].

Segmentation is a key preprocessing step in our work, and a point of difference between the two groups working in this area [21, 34]. The approach we take is to use segmentation rather than directly using the entire hyperspectral images, as this enables training of machine learning techniques with more training examples. The segmentation approach divides the hyperspectral image into a five-by-five grid, meaning it obtains 25 training/testing examples from each image captured. The segments are then averaged, excluding outliers, representing the overall honey sample. Figure 3 clearly illustrates this process.

Normalisation is applied to the final data set before using any machine learning techniques. The normalisation technique is a simple standard scalar, which scales most data to fit between 0 and 1. This technique has been used in previous work on honey botanical origins classification with hyperspectral imaging [23].

Feature reduction and classification

We use the most promising techniques from the existing work on honey botanical origins classification and consider the theory of adulteration from related applications. Feature reduction can reduce overfitting, therefore, improving performance by reducing the complexity of the problem [10].
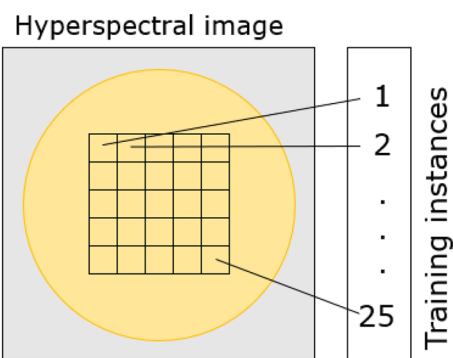
Suitable benchmark classifiers, including the k-nearest neighbour (KNN) classifier and SVMs using a linear and radial basis function (RBF) kernel, have been considered. The feature reduction techniques used are: principal component analysis (PCA), selecting the k best features based on feature ranking, and autoencoder techniques [26, 29].

We consider a new technique to smooth the features by summing together groups of several wavelengths. This technique is in line with the assumption that sugar concentration is best detected by taking the sum of a range of wavelengths rather than a single wavelength [33, 36].

Feature smoothing

It has been validated that sugar concentration is best detected by taking the sum of a range of wavelengths rather than a single wavelength [33, 36]. Accordingly, feature smoothing is implemented before performing feature reduction and classification. Although this approach incurs data losses, it helps our classifiers to generalise by reducing complexity [10] in detecting sugar in a mixture of adulterated honey. Figure 4 shows this feature smoothing that we consider in this work. This smoothing works as a moving average across the original features.

The smoothing in our work uses a window size of 15, meaning that each new feature is the average of 15 original features. This window size was found empirically. Much larger sizes smooth the features too much and decrease classification accuracy, while smaller windows do not impact the classification accuracy.

Principal component analysis (PCA)

PCA is a standard method for feature extraction and reduction. The features are linearly combined, forming a set of orthogonal vectors. Only the high variance features are kept in this approach [39].
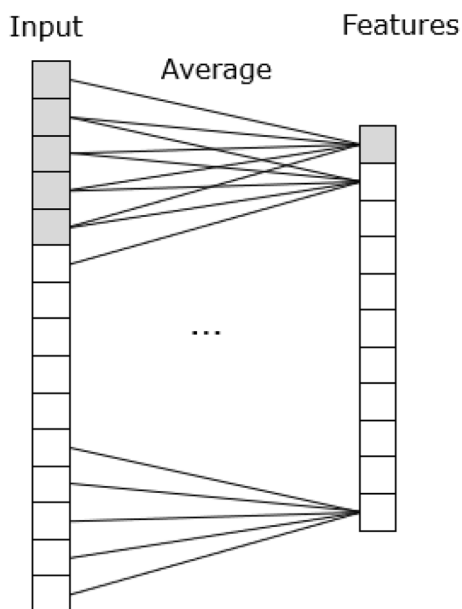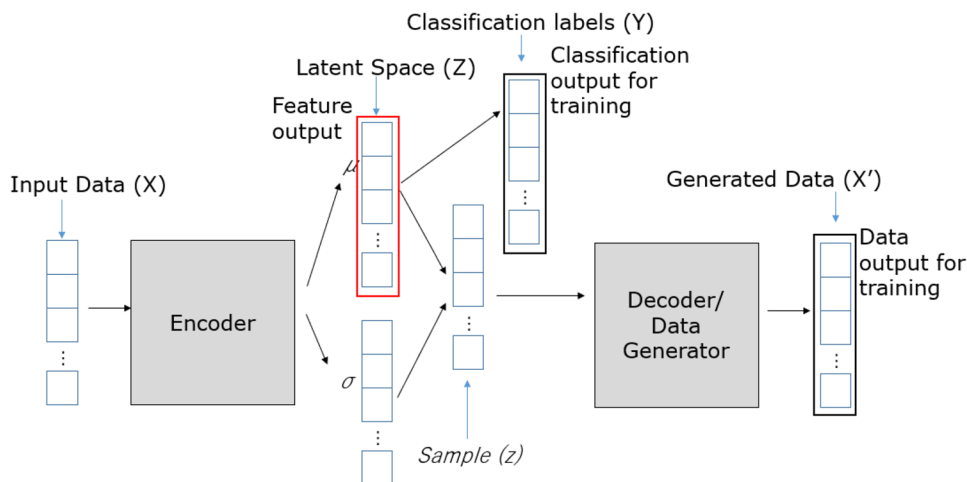
Fig. 4 Feature smoothing method. The input features are averaged in groups of five in this example to form the new features

PCA has been a high-performing benchmark method in our previous work on honey botanical origins classification [26, 27, 29].

Autoencoders

Autoencoders are another form of feature reduction; they transform the input space, similar to PCA, into a smaller set of features representing the same information [1]. Autoencoders are neural networks with two networks; encoder and decoder. The input data passes through the encoder to determine the features; then, the features pass through the decoder to recreate the input data.

The CEAE is a type of such structure but additionally includes the classification labels in the training process through using a weighted secondary output [29]. In this work, we use a variational class embodiment autoencoder (VCEAE) which combines the variational autoencoder (VAE) with the CEAE. Figure 5 shows the VCEAE structure we use in this paper. This structure improved the generalisation performance on the honey botanical origins classification [26].

K nearest neighbour classifier (KNN)

The KNN classifier is a standard machine learning classifier. In this work, we use KNN with $k = 5$, meaning that each example is classified based on the class labels of the five closest training examples. This method was used as a benchmark in our previous work on honey botanical origins classification [26, 27, 29]. KNN requires minimal parameter tuning, which is an advantage over more complicated classifiers.

Support vector machines (SVMs)

SVMs perform well in situations, where the data are of high dimensionality or has non-linear complexities [13, 14, 37, 38]. SVMs work by first transforming the feature space using a kernel function and then splitting the data by a hyperplane to minimise the risk of misclassifications. We investigated different SVM structures for Manuka honey classification [27, 29] with varied success. We consider these structures further in this experiment using linear and RBF kernel SVMS.

Data analysis

As this is a new database and the first publicly available database for adulterated honey mixtures, it is essential to

Fig. 5 Variational class embodiment autoencoder architecture, with four fully connected encoder and decoder layers

analyse the data to identify trends and evaluate the suitability of the data for the adulteration detection task. The data has been made available online [30].

Full data set analysis

The data set comprises 12 different honey products from seven brands with 11 different botanical origins labels. Each type of honey has been captured with six independent samples. Half of the samples are from Manuka honey, a premium NZ honey type, and the other half are from other NZ honey. Table 1 shows the makeup of the data set from these different kinds of honey. In creating the data set, we sampled and captured images of all the honey at each sugar concentration; however, some mixtures were of low quality, and we could not include the images in the final data set. Table 1 represents the number of training and testing examples available. Each sample contains 25 examples following segmentation. The 0% adulteration examples relate to samples of pure unadulterated honey. These are kinds of honey that have previously been captured for a botanical origins data set [28] and come from reputable sources. These pure honey samples are guaranteed pure honey of the botanical origin displayed on their label.

The mean spectral response of the different concentrations is given in Fig. 6, showing the general trend for how added sugar affects the spectral response of honey. Figure 6 shows that in the mid-range of wavelengths, the response is distinct depending on the concentration. The spectral responses follow a similar pattern across all concentrations, indicating that the overall shape of the spectral response is dominated by the honey, not the sugar syrup.

It is important to determine which features can give us more information about the concentration of the adulterated honey. Analysis of variance (ANOVA) calculates the F-score for each wavelength feature using the concentrations
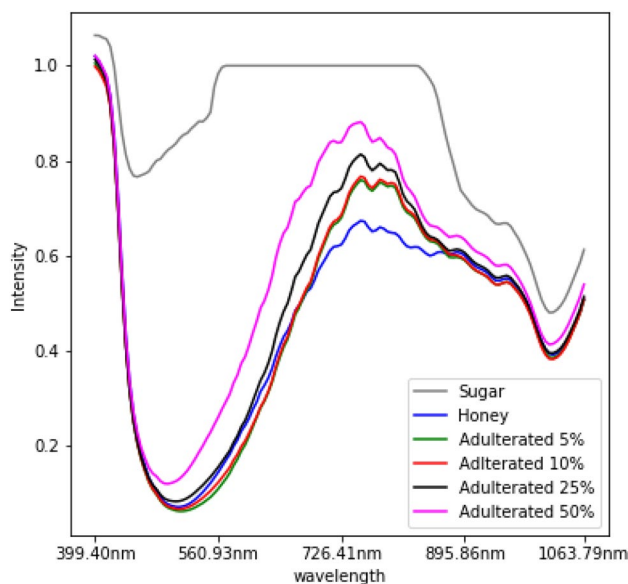

Fig. 6 Mean spectral response for all honey types at different concentrations of adulteration with sugar syrup

as the classes. Figure 7a shows ANOVA calculated F-scores. Another technique for analysing relevant features is the chi^2 statistic. Figure 7b shows the chi^2 results for all wavelengths.

Intra honey-type analysis

The variation within the different adulterated honey types is essential. This section displays the average spectral responses for each honey type at different concentrations. A few of the honey types do not contain samples from all concentrations due to quality control of our data, where some concentrations were contaminated or incorrectly captured.

Figure 8 shows the mean spectra for each honey type separately. The adulterated spectra are closely related in some

Table 1 Overall makeup of the adulterated honey data set from each brand and botanical origins label of honey

Brand	Class	Adulteration concentration					
		0%	5%	10%	25%	50%	Sum
C1	Clover	150	150	300	300	300	1200
C10	MultiFloral	150	150			150	450
	ManukaUMF5		150	150	150	150	600
	ManukaUMF15		150	150	150	150	600
	ManukaUMF20		150	150	150	150	600
C4	ManukaUMF10		150	150	150	125	575
C5	ManukaBlend		150		150	150	450
C7	BorageField	150	150	150	150	150	750
	Kamahi	150	150	150	150	150	750
	Rewarewa	150	150	150			450
C8	ManukaBlend	150	150	150	150	150	750
C9	Manuka	150	300	300	300	300	1350

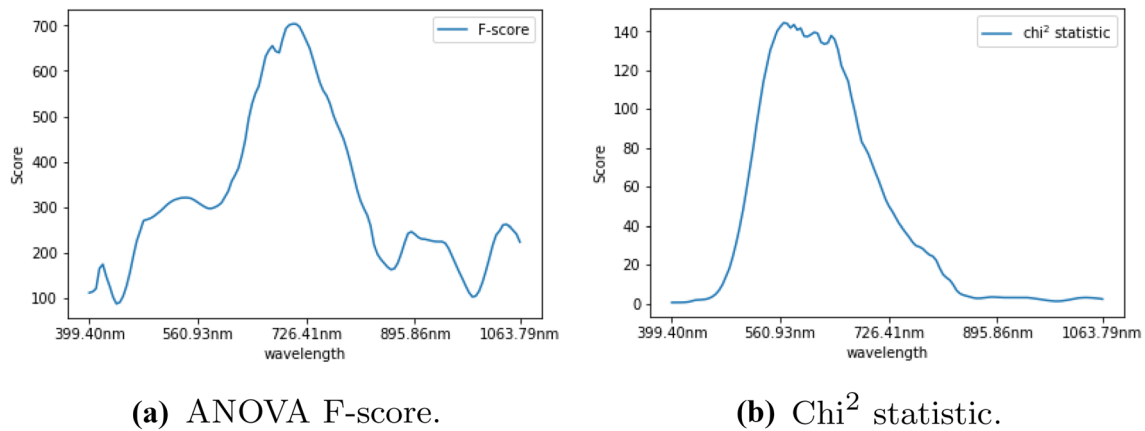**(a)** ANOVA F-score.

(b) Chi^2 statistic.

Fig. 7 ANOVA F-score and chi^2 statistics calculated for each wavelength using the adulteration concentration as the classification label

honey types. This finding could indicate a high level of sugar naturally occurring in these kinds of honey. Detecting the adulterant generally across all types of honey is not straightforward as honey shares some attributes with the diluted sugar solutions.

Some types of honey are very similar across all adulteration concentrations, such as Rewarewa honey and Manuka UMF15+ and UMF20+. Figure 8 shows these spectral responses are, on average, very similar for these honey types. Other types of honey have apparent differences between the different adulteration concentrations, such as Borage Field, Manuka, and Manukablend, where higher classification accuracy is expected.

Experiment design

This section describes how the experiments are designed to classify adulterated honey.

We evaluate the classification accuracy in two different scenarios:

- Binary classification between pure honey and all concentrations of adulterated honey.
- Classification of adulteration % into categories of 0%, 5%, 10%, 25%, and 50%.

These two scenarios represent a realistic real-world need, which is to, most importantly, be able to detect if a honey sample is fraudulent. Once adulteration is detected, estimating how much adulterant is used would become crucial.

Our concentration detection is limited to a classification problem with four possible adulteration concentrations due to the data being heavily skewed toward these four concentrations. This data skew was by design when we created the data set, as we aim to evaluate how well the classifiers could detect different ranges of concentrations. Once it is proven

to be possible to classify between these four classes, future work can capture more data and train a regression algorithm to predict the sugar concentration.

The feature reduction techniques adopted are PCA and the VCEAE. For PCA, we use the first 20 principal components. The network parameters for the VCEAE are kept consistent with what was used in previous work for honey botanical origins classification, detailed in Sect. 2.3.3. The VCEAE structure is a six hidden layer encoder and decoder network with layers [128(*input*), 128, 110, 92, 74, 56, 38, 20(*output*)], and the reverse for the decoder. There are 20 features in the latent space, and it has a classification weight of 0.4. We used 50 epochs with a batch size of 32 and a learning rate of 0.001. The rectified linear unit (ReLU) activation function was used throughout the entire network, aside from the classification output, which used sigmoid. We used dropout for all our network layers with a dropout rate of 0.0005. These parameters align with what was used for honey botanical origins classification, where the VCEAE had the best generalisation performance [26].

In addition to these feature reduction techniques, we also apply the feature smoothing method detailed in Sect. 2.3.1 with a window size of 15.

The classifiers we consider are the KNN classifier with $k = 5$, RBF and linear SVMs, where γ and C are tuned parameters on the cross-validation set using seven values on a log scale between 10^{-3} and 10^3

In this work, we deal with unbalanced data sets for the binary classification case, so it is important to measure accuracy and understand how each class is performing. We use the F1 score for this measure, which calculates a balanced average between recall and precision. The F1 scores are then averaged with a macro average technique that weights each class equally, not based on the class size. The F1 score can be calculated by Eq. 3, where precision is defined by equation 1, and recall is defined by Eq. 2. True positive refers to
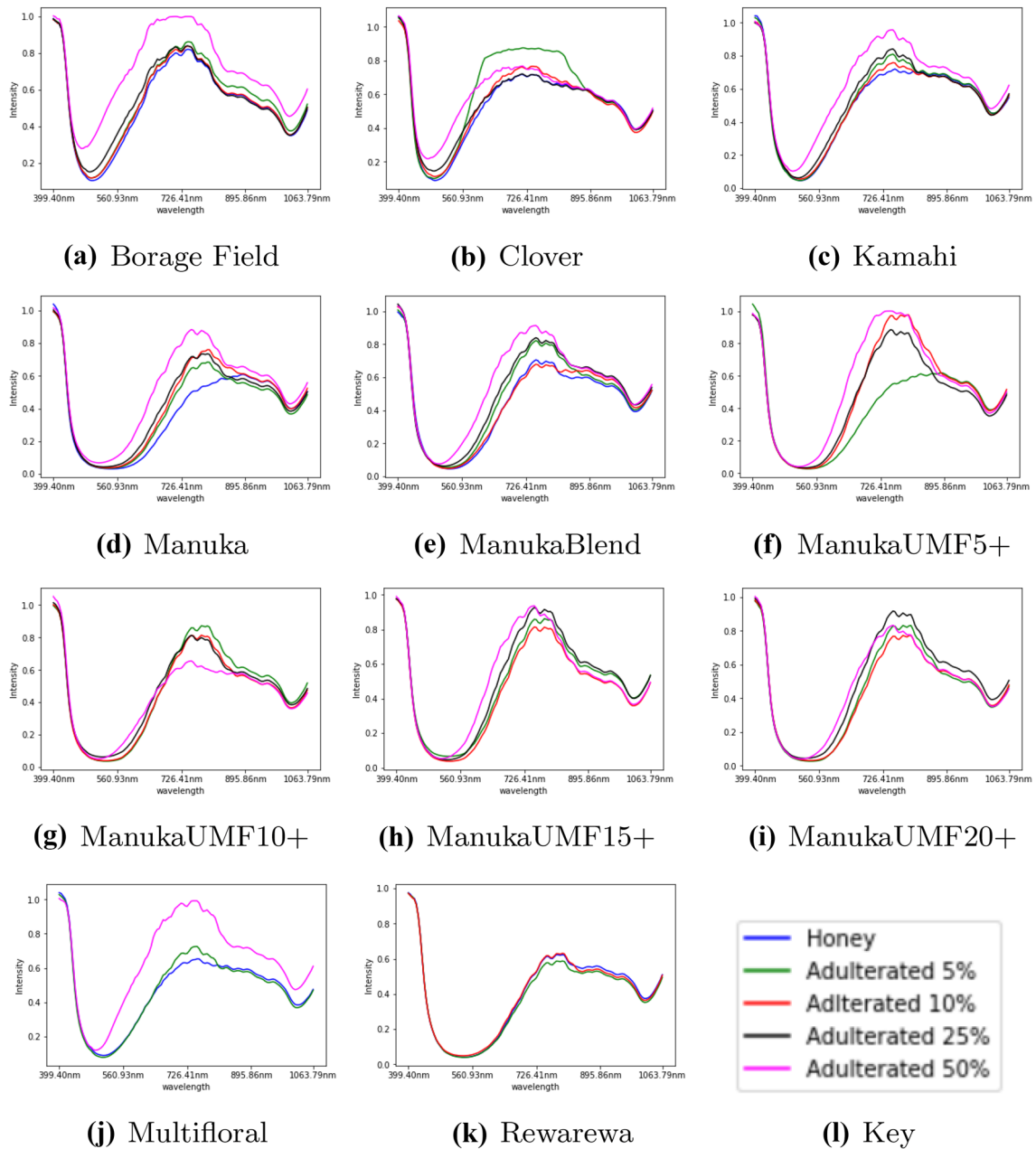
Fig. 8 Mean spectral response of each honey type at different adulteration concentrations

the positive examples from each class that were correctly classified, whereas false positive is the negative examples incorrectly classified as positive. False-negative are the examples that were classified as negative but are positive. This metric is calculated per class and then averaged after the final F1 score calculation in Eq. 3. The F1 score metric, along with standard accuracy, is a useful tool to evaluate the performance of a classifier and ensure it is not biased toward the larger class:

$$Precision = \frac{True\ positive}{True\ positive + False\ positive} \qquad (1)$$

$$Recall = \frac{True\ positive}{True\ positive + False\ negative} \qquad (2)$$

$$F1 = \frac{2 * (Recall * Precision)}{Recall + Precision} \qquad (3)$$

Results

This section discusses the results of feature smoothing, feature reduction, and classification of the adulteration concentrations. There are two classification approaches in this paper; binary and multi-class classifications, as described in Sect. 2.6.

Binary classification results

This section details the results of binary classification between adulterated and non-adulterated honey. Table 2 refers to a classifier trained to generally classify adulterated vs non-adulterated honey. The training and testing set consists of all the different concentrations of sugar adulteration we have captured as adulterated honey and the pure honey captured as non-adulterated. The classifiers used have been used in previous work on honey botanical origins classification. As detailed in Sect. 2.3, we use a feature smoothing approach, along with feature reduction and classification. The feature reduction techniques we use are PCA and VCEAE and the original features. The classifiers used are a

KNN classifier, a linear SVM, and an RBF SVM. The SVM classifiers have been tuned to this particular problem.

Table 2 shows the results of binary classification between non-adulterated honey and all concentrations of adulterated honey. Acc is the performance accuracy between zero and one. Std is the standard deviation of the cross-validation results. F1 refers to the average F1 score for all classes calculated by Eq. 3. The results show that even classical techniques such as KNN can perform well. The best classifiers achieve above 97% accuracy on this problem, making them a valuable addition to fraud protection measures already in place for honey adulteration. A quick and reliable method of detecting adulterated honey, even at low percentages, is beneficial for exporting and importing honey products with reliable quality assurance.

The results from Table 2 show that there is some improvement from reducing the number of features and using support vector machine classifiers. However, the KNN classifier with the smoothed features can perform very well. The linear SVM using the VCEAE feature reduction technique was the best performing classifier. These more complex classifiers significantly impacted when classifying honey botanical origins, but the improvement is only small for this adulteration classification problem. For the next set of results, we consider using only the best classifier, the linear SVM with the VCEAE. In general, the feature smoothing technique improved the results for most classifiers. When comparing the same classifiers and feature reduction techniques with

Table 2 Results of binary classification experiments between non-adulterated and adulterated honey samples

Smoothing	Features	Classifier	Cross validation			Training Set		Testing Set	
			Acc	Std	F1	Acc	F1	Acc	F1
None	All	KNN	0.998	0.001	0.998	0.999	1.000	0.963	0.900
		LinearSVM	0.964	0.006	0.918	0.959	0.910	0.921	0.810
		RBFSVM	0.973	0.002	0.942	0.977	0.950	0.961	0.910
	PCA	KNN	0.998	0.001	0.996	0.999	1.000	0.961	0.900
		LinearSVM	0.819	0.048	0.660	0.870	0.710	0.876	0.600
		RBFSVM	0.968	0.004	0.934	0.972	0.940	0.958	0.900
	VCEAE	KNN	1.000	0.000	1.000	1.000	1.000	0.969	0.920
		LinearSVM	**0.999**	**0.001**	**1.000**	**1.000**	**1.000**	**0.976**	**0.940**
		RBFSVM	0.998	0.002	0.998	0.999	1.000	0.971	0.930
Sliding Window	All	KNN	0.999	0.001	0.998	1.000	1.000	0.967	0.920
		LinearSVM	0.918	0.055	0.842	0.903	0.830	0.903	0.790
		RBFSVM	0.963	0.002	0.924	0.967	0.930	0.960	0.910
	PCA	KNN	0.999	0.001	1.000	1.000	1.000	0.966	0.910
		LinearSVM	0.892	0.024	0.764	0.870	0.720	0.879	0.670
		RBFSVM	0.962	0.005	0.922	0.966	0.930	0.959	0.910
	VCEAE	KNN	1.000	0.001	1.000	1.000	1.000	0.946	0.870
		LinearSVM	0.999	0.001	1.000	0.998	1.000	0.957	0.890
		RBFSVM	0.996	0.003	0.992	0.996	0.990	0.960	0.900

Acc is the classification accuracy between zero and one, Std is the standard deviation, and F1 is the F1 score

and without the feature smoothing, the classification systems with the smoothing performed better in both accuracy and F1 score in most cases. However, the feature smoothing did not improve the results for our best classifier and feature reduction combination.

We also consider training a classifier on binary classification between each concentration of adulterant and non-adulterated honey. The results of this are shown in Table 3. These results are useful to get a picture of how the different concentrations of honey are misclassified. The results in Table 3 are obtained using a linear SVM classifier with the VCEAE feature reduction method.

The results in Table 3 show that, as expected, the lower the sugar concentration, the harder it is to classify the adulterated honey. The best testing accuracy on 5% honey was 87%, with an F1 score of 0.850. The best accuracy for the 50% and 25% adulterated honey was 100% with an F1 score

of one. This accuracy shows that most misclassification is happening for the very low concentrations of adulteration.

Multi-class classification results

We also perform multi-class classification between all adulteration concentrations; 0%, 5%, 10%, 25%, 50%. This problem relates to a real-world scenario, where we want to understand how much sugar is added to adulterated honey. Table 4 shows the results of multi-class classification using our set of benchmark feature reduction techniques and classifiers, as well as the new feature smoothing technique.

The results in Table 4 show that it is possible to achieve over 95% accuracy in this classification. The best classifier is the KNN with all features, and the feature smoothing method improves it slightly. The multi-class classification results are consistent with the binary classification results regarding the

Table 3 Binary classification between pure honey (0% adulteration) and each concentration of adulteration (5%, 10%, 25%, 50%)

Problem	Cross validation			Training set		Testing set	
	Acc	Std	F1	Acc	F1	Acc	F1
5% vs 0%	0.998	0.005	0.998	1.000	1.000	0.870	0.850
10% vs 0%	1.000	0.001	1.000	0.999	1.000	0.954	0.950
25% vs 0%	1.000	0.000	1.000	1.000	1.000	1.000	1.000
50% vs 0%	0.999	0.002	0.998	1.000	1.000	1.000	1.000

Acc is the classification accuracy between zero and one, Std is the standard deviation, and F1 is the F1 score

Table 4 Results of multi-class classification experiments of adulteration concentration

Smoothing	Features	Classifier	Cross validation			Training set		Testing set	
			Acc	Std	F1	Acc	F1	Acc	F1
None	All	KNN	0.991	0.003	0.990	0.996	1.000	0.950	0.940
		LinearSVM	0.727	0.010	0.730	0.748	0.750	0.644	0.640
		RBFSVM	0.956	0.006	0.956	0.969	0.970	0.899	0.880
	PCA	KNN	0.991	0.003	0.992	0.995	1.000	0.948	0.940
		LinearSVM	0.501	0.009	0.486	0.507	0.490	0.474	0.440
		RBFSVM	0.944	0.003	0.946	0.959	0.960	0.891	0.880
	VCEAE	KNN	0.996	0.001	0.998	0.999	1.000	0.945	0.930
		LinearSVM	0.995	0.002	0.996	0.999	1.000	0.946	0.930
		RBFSVM	0.993	0.002	0.992	0.998	1.000	0.946	0.930
Sliding Window	**All**	**KNN**	**0.995**	**0.002**	**0.994**	**0.999**	**1.000**	**0.951**	**0.940**
		LinearSVM	0.638	0.008	0.634	0.657	0.650	0.574	0.560
		RBFSVM	0.932	0.005	0.932	0.945	0.950	0.885	0.870
	PCA	KNN	0.995	0.002	0.994	0.999	1.000	0.950	0.940
		LinearSVM	0.580	0.009	0.576	0.590	0.580	0.527	0.490
		RBFSVM	0.930	0.006	0.930	0.944	0.940	0.883	0.870
	VCEAE	KNN	0.997	0.001	1.000	0.999	1.000	0.943	0.930
		LinearSVM	0.996	0.002	0.998	0.998	1.000	0.940	0.930
		RBFSVM	0.991	0.006	0.990	0.995	1.000	0.939	0.930

Acc is the classification accuracy between zero and one, Std is the standard deviation, and F1 is the F1 score

high-performing classifiers. However, the best classifier is different for this problem. The accuracy of the multi-class classifiers is slightly less than binary classification; the F1 scores are higher than the binary case. This difference in F1 score is likely a result of minor class imbalance in the multi-class scenario.

Overall the results achieved above 95% using a feature smoothing method and a KNN classifier for binary and multi-class classification tests. The misclassifications should be investigated to analyse if a pattern or some aspect of the data is causing an adulterated sample to be classified as pure honey or vice versa.

Discussion

In this section, we discuss the impact of the results reported in Sect. 3 and perform further analysis to give insight into the misclassifications.

The main result of over 95% classification accuracy for binary and multi-class classification shows that hyperspectral imaging and machine learning can accurately detect if honey is adulterated with sugar syrup. This system would be a valuable addition to the existing quality assurance methods used for honey. The performance evaluation is limited to honey in our data set. Detecting fraud with this level of accuracy in a known honey set is a valuable tool, and as this data set builds up over time, a more general fraud detection system can be developed.

The misclassifications in the binary classifier are occurring mainly with the 5% and 10% adulterated honey samples. Evaluating the individual concentration classification shows that the misclassifications mainly impact C1 Clover honey. Table 5 shows the misclassification percentage of each honey type when using the best classifier and feature

reduction techniques. This analysis is coresponding to the results shown in Tables 2, 3, and 4.

Overall we can see in Table 5 that some honey types are more often misclassified than others. One-third of the samples were misclassified for the C1 Clover honey type in the binary classification between 0 and 5% adulterated honey. This misclassification might be to do with the constituents of clover honey compared to the other honey. It is typically quite a light coloured sweet honey, so perhaps the spectrum is not changed much when small amounts of sugar are added. The Manuka honey performed exceptionally well, with few misclassification percentages in the multi-class problem. This high performance could be because Manuka is very rich in flavoured and coloured honey, and adding sugar syrup does not mimic the constituents of Manuka honey very well. This result is positive for our application as Manuka is the most expensive honey and has been the main target for fraud.

The other form of analysis is to analyse which classes are being misclassified. This is particularly important for the multi-class problem; however, we will analyse all the classification problems. Table 6 shows the confusion matrices for all of the classification scenarios. For the binary classification problem in Table 6a, pure honey means the unadulterated honey samples. These unadulterated samples are referred to as 0% adulteration in Table 6b–f.

Table 6 shows that the misclassifications are commonly the case that the pure, 0% adulterated honey is being misclassified as adulterated. In Table 6f, the multi-class case, these are mostly misclassified as 5% honey. Based on Table 6f, it is clear that the misclassifications occur between similar classes, which is positive for our results. A 25% adulterated honey is only confused with 10% and 50% adulterated honey.

In Tables 5 and 6, we can see that the majority of misclassification is happening on a small group of honey and

Table 5 Table showing misclassification % for each honey type on each problem

Brand	Class	Mislassification Percentage					
		Binary	5%	10%	25%	50%	Multi
C1	Clover	12.50%	50.00%	32.00%			12.50%
C10	MultiFloral	5.33%	48.00%				1.33%
	ManukaUMF5						
	ManukaUMF15						
	ManukaUMF20						8.00%
C4	ManukaUMF10						14.00%
C5	ManukaBlend						
C7	BorageField	0.80%		2.00%			0.80%
	Kamahi	2.40%	48.00%	6.00%			15.20%
	Rewarewa						
C8	ManukaBlend						0.80%
C9	Manuka						

The classification problems labeled 5%, 10%, 25%, and 50% refer to the binary classification between the % and pure honey

Table 6 Confusion matrices for binary and multi-class classification problems using the VCEAE and linear SVM for binary problems, and the KNN classifier with smoothed features for multi-class classification

(a) Binary

	Pure	Adulterated
Pure	143	32
Adulterated	1	1224

(b) 5% vs 0%

	0%	5%
0%	125	50
5%	23	302

(c) 10% vs 0%

	0%	10%
0%	147	28
10%	0	300

(d) 25% vs 0%

	0%	25%
0%	175	0
25%	0	300

(e) 50% vs 0%

	0%	50%
0%	175	0
50%	0	300

(f) Multi-class

	0%	5%	10%	25%	50%
0%	129	45	1	0	0
5%	0	320	5	0	0
10%	0	2	284	14	0
25%	0	0	0	298	2
50%	0	0	0	0	300

are typically from nearby classes. This result indicates that generalising the classifiers to all types of honey is causing errors. The classifier can train a good solution for most of the data set, especially where there are many similar training examples, such as with Manuka honey. The solution to this problem is to add data from many more types of honey to have a broader representation of botanical origins labels. We saw high misclassification from Clover honey which has few samples, and particularly low misclassification for Manuka honey which has many samples from several different brands. As discussed in Sects. 1.1 and 1.1.1, previous work on adulteration has often focused on tiny data sets of honey or only one type of honey. This paper has solved some issues with classifying several honey types at once. Still,

future work is needed to develop a more broadly representative database to classify all honey types accurately.

Compared to existing work on honey adulteration detection, we use the largest data set with the most variety of honey types. The accuracy of 97.6% for binary adulteration detection, as detailed in Table 2, and 95% for multi-class adulteration classification, as detailed in Table 4 is higher than any known existing work that uses multiple honey types. Higher accuracy of 99.2% for adulteration detection has been reported in [15], where near-infrared spectroscopy was used to detect adulteration in four different honey types. This work, however, used only 16 samples for four different concentrations and four honey types, and the test set did not use independent samples from the training set, so this result is not statistically valid.

Fourier transform infrared spectroscopy was used to detect the adulteration of honey with cane sugar in [12, 35]. This paper used only three types of honey that were adulterated to a concentration of cane sugar between 1% and 25%. The average test accuracy in [35] was 92.5% when the model was trained on each honey type separately but dropped to 78.4% when the model was trained and tested on all three honey types simultaneously. Statistical methods of canonical variate analysis are used as the classification model. Our methods are superior to this approach, as we have trained our model across all honey types simultaneously and achieved over 95% for multi-class classification. This work was extended for one honey type (Clover) in [12] which achieved 93.95% using a neural network and linear discriminant analysis. The conclusion of this work stated that it needed to be extended to more honey types.

A single honey type (Manuka) was adulterated with five different types of sugar syrup (corn syrup, sucrose syrup, high fructose corn syrup, beet syrup, and rice syrup) and captured with near-infrared spectroscopy [41]. The honey was adulterated with all the adulterants at concentrations 10%, 20%, 30%, 40%, and 50%. This work showed a clear difference in the spectrum between 1300 and 1800 nm that can be used to detect the adulteration of Manuka honey. This work was limited by only having one honey type. However, it has been demonstrated that spectral imaging can detect adulteration with many different adulterants.

Hyperspectral imaging was used to detect adulteration between sugar syrup and honey using a neural network with 95% accuracy in [34]. The approach used was different to our method, as they did not use segmentation on the hyperspectral images. Instead, they used an entire hyperspectral image as a single training/testing example. The accuracy achieved is lower than our accuracy for binary classification, and the data set is much smaller than the one we provide. Multi-class classification is mentioned, but the results are not quantified and, therefore, cannot be compared.

Conclusions

This paper has introduced new adulteration detection techniques and a suitable adulterated database, where honey is adulterated with different percentages of sugar syrups. This database is the first to be made publicly available online. The process of capturing honey samples is also made available to the research community. We also introduced a new method of feature smoothing before applying standard feature extraction and classification techniques. The proposed smoothing feature method improved the classification accuracy across all classifiers and testing scenarios while reducing the complexity of the algorithms. This paper shows that using hyperspectral imaging; adulterated honey can be detected with above 95% accuracy for both binary and multi-class classification, as shown in Tables 2 and 4. This accuracy is higher than comparable work on adulteration detection in honey. As honey, especially Manuka, increases in value globally, honey fraud will become more common. Thus, we propose a quick, non-invasive, and reliable method to detect adulteration with sugar syrup. The main limitation of this work is that although the data set is the largest captured for honey adulteration, it is still far too small and only covers a minimal set of honey types at five different adulteration concentrations. Future work involves: extending the data set to a broader range of honey, evaluating and improving the generalisation to unknown honey types of the current data set, extending the data set to include other adulterants, such as rice sugar or cheaper honey, and extending the data set to cover more concentrations of adulteration particularly concentrations below 5%. Compared to the existing work on detecting sugar adulteration in honey with hyperspectral imaging, this approach is more accurate and tested on a broader publicly available data set.

Declarations

References

1. Baldi P (2012) Autoencoders, unsupervised learning, and deep architectures. In: Proceedings of ICML workshop on unsupervised and transfer learning, pp 37–49
2. Bansal S, Singh A, Mangal M et al (2017) Food adulteration: sources, health risks, and detection methods. Criti Rev Food Sci Nutr 57(6):1174–1189
3. Bong J, Loomes KM, Schlothauer RC et al (2016) Fluorescence markers in some New Zealand honeys. Food Chem 192:1006–1014. https://doi.org/10.1016/j.foodchem.2015.07.118
4. Canadian Food Inspection Agency (2019) Report: enhanced honey authenticity surveillance (2018–2019). Tech. rep., Canadian Food Inspection Agency, https://www.inspection.gc.ca/about-cfia/science-and-research/our-research-and-publications/report/eng/1557531883418/1557531883647
5. El Masry G, Sun DW (2010) Principles of hyperspectral imaging technology. In: Hyperspectral imaging for food quality analysis and control. Elsevier, p 3–43. https://doi.org/10.1016/b978-0-12-374753-2.10001-2
6. García NL (2018) The current situation on the international honey market. Bee World 95(3):89–94. https://doi.org/10.1080/0005772x.2018.1483814
7. Girma A, Seo W, She RC (2019) Antibacterial activity of varying umf-graded manuka honeys. PloS One 14(10):e0224495. https://doi.org/10.1371/journal.pone.0224495
8. Gowen A, O'Donnell C, Cullen P et al (2007) Hyperspectral imaging—an emerging process analytical tool for food quality and safety control. Trend Food Sci Technol 18(12):590–598. https://doi.org/10.1016/j.tifs.2007.06.001
9. Gunaratne Viejo G, Gunaratne (2019) Chocolate quality assessment based on chemical fingerprinting using near infra-red and machine learning modeling. Foods 8(10):426. https://doi.org/10.3390/foods8100426
10. Guyon I, Gunn S, Nikravesh M et al (2008) Feature extraction: foundations and applications, vol 207. Springer, Berlin
11. Hollas JM (2004) Modern spectroscopy. Wiley, Hoboken
12. Irudayaraj J, Xu R, Tewari J (2003) Rapid determination of invert cane sugar adulteration in honey using ftir spectroscopy and multivariate analysis. J Food Sci 68(6):2040–2045. https://doi.org/10.1111/j.1365-2621.2003.tb07015.x
13. Kecman V (2004) Support vector machines basics. School of Engineering. University of Auckland, Auckland
14. Kecman V (2005) Support vector machines–an introduction. In: Support vector machines: theory and applications. Springer, p 1–47. https://doi.org/10.1007/10984697_1
15. Kumaravelu C, Gopal A (2015) Detection and quantification of adulteration in honey through near infrared spectroscopy. Int J Food Proper 18(9):1930–1935. https://doi.org/10.1080/10942912.2014.919320
16. León L, Kelly JD, Downey G (2005) Detection of apple juice adulteration using near-infrared transflectance spectroscopy. Appl Spectrosc 59(5):593–599. https://doi.org/10.1366/0003702053945921
17. Minaei S, Shafiee S, Polder G et al (2017) Vis/nir imaging application for honey floral origin determination. Infrar Phys Technol 86:218–225. https://doi.org/10.1016/j.infrared.2017.09.001
18. New Zealand Consulate-General Los Angeles (2015) The New Zealand honey phenomenon in the USA. New Zealand

Consulate-General Los Angeles. http://www.honeynetwork.com/media/1322/honey-report-la-consulate.pdf

19. Nogales-Bueno J, Hernández-Hierro JM, Rodríguez-Pulido FJ et al (2014) Determination of technological maturity of grapes and total phenolic compounds of grape skins in red and white cultivars during ripening by near infrared hyperspectral image: a preliminary approach. Food Chem 152:586–591. https://doi.org/10.1016/j.foodchem.2013.12.030

20. Noviyanto A (2018) Honey botanical origin classification using hyperspectral imaging and machine learning. PhD thesis. The University of Auckland. https://doi.org/10.1016/j.jfoodeng.2019.109684

21. Noviyanto A, Abdulla WH (2017) Honey dataset standard using hyperspectral imaging for machine learning problems. In: 2017 25th European Signal Processing Conference (EUSIPCO), IEEE, pp 473–477, https://doi.org/10.23919/eusipco.2017.8081252

22. Noviyanto A, Abdulla WH (2019) Segmentation and calibration of hyperspectral imaging for honey analysis. Comput Electron Agric 159:129–139. https://doi.org/10.1016/j.compag.2019.02.006

23. Noviyanto A, Abdulla WH (2020) Honey botanical origin classification using hyperspectral imaging and machine learning. J Food Eng 265(109):684. https://doi.org/10.1016/j.jfoodeng.2019.109684

24. Noviyanto A, Abdulla WH (2021) Signifying the information carrying bands of hyperspectral imaging for honey botanical origin classification. J Food Eng 292(110):281. https://doi.org/10.1016/j.jfoodeng.2020.110281

25. Noviyanto A, Abdullah W, Yu W, et al (2015) Research trends in optical spectrum for honey analysis. In: Signal and Information Processing Association Annual Summit and Conference (APSIPA), 2015 Asia-Pacific, IEEE, pp 416–425. https://doi.org/10.1109/apsipa.2015.7415305

26. Phillips T, Abdulla W (2020) Generalisation techniques using a variational ceae for classifying manuka honey quality. In: 2020 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), IEEE, pp 1631–1640

27. Phillips T, Abdulla W (2021) Developing a new ensemble approach with multi-class svms for manuka honey quality classification. Appl Soft Comput: 107710. https://www.sciencedirect.com/science/article/pii/S1568494621006311

28. Phillips T, Noviyanto A, Abdulla W (2020) Hyperspectral imaging honey database. https://doi.org/10.17608/k6.auckland.12170475.v1. https://figshare.com/s/25afe30ff531b8f1e65f

29. Phillips T, Abdulla W (2019) Class embodiment autoencoder (CEAE) for classifying the botanical origins of honey. In: 2019 International Conference on Image and Vision Computing New Zealand (IVCNZ). IEEE. https://doi.org/10.1109/ivcnz48456.2019.8961004

30. Phillips T, Coleman B, Takano S, et al (2021) Hyperspectral imaging adulterated honey dataset. https://doi.org/10.17608/k6.auckland.16441686.v1, https://auckland.figshare.com/articles/dataset/Hyperspectral_Imaging_adulterated_honey_dataset/16441686/1

31. Phipps R (2020) International honey market. CPNA International Ltd. https://www.apiservices.biz/en/articles/sort-by-popularity/2533-international-honey-market-report-september-2020

32. Rodríguez-Pulido FJ, Barbin DF, Sun DW et al (2013) Grape seed characterization by nir hyperspectral imaging. Postharv Biol Technol 76:74–82. https://doi.org/10.1016/j.postharvbio.2012.09.007

33. Rodriguez-Saona LE, Fry FS, McLaughlin MA et al (2001) Rapid analysis of sugars in fruit juices by FT-NIR spectroscopy. Carbohydr Res 336(1):63–74. https://doi.org/10.1016/s0008-6215(01)00244-0

34. Shafiee S, Polder G, Minaei S, et al (2016) Detection of honey adulteration using hyperspectral imaging. IFAC-PapersOnLine 49(16):311–314. https://doi.org/10.1016/j.ifacol.2016.10.057 (5th IFAC Conference on Sensing, Control and Automation Technologies for Agriculture AGRICONTROL 2016)

35. Sivakesava S, Irudayaraj J (2001) Prediction of inverted cane sugar adulteration of honey by fourier transform infrared spectroscopy. J Food Sci 66(7):972–978. https://doi.org/10.1111/j.1365-2621.2001.tb08221.x

36. Sutcu Y (2014) Detection of added sugar in red wine using visual light spectroscopy. Available at https://publiclab.org/notes/ygzstc/07-23-2014/detection-of-added-sugar-in-red-wine-using-visual-light-spectroscopy

37. Tang Y (2013) Deep learning using linear support vector machines. arXiv preprint arXiv:1306.0239

38. Vapnik V (1998) The support vector method of function estimation. In: Nonlinear Modeling. Springer, p 55–85, https://doi.org/10.1007/978-1-4615-5703-6_3

39. Wold S, Esbensen K, Geladi P (1987) Principal component analysis. Chemomet Intell Lab Syst 2(1–3):37–52. https://doi.org/10.1016/0169-7439(87)80084-9

40. Wu D, Sun DW (2013) Advanced applications of hyperspectral imaging technology for food quality and safety analysis and assessment: a review-part ii: Applications. Innov Food Sci Emerg Technol 19:15–28. https://doi.org/10.1016/j.ifset.2013.04.016

41. Yang X, Guang P, Xu G et al (2020) Manuka honey adulteration detection based on near-infrared spectroscopy combined with aquaphotomics. LWT 132(109):837. https://doi.org/10.1016/j.lwt.2020.109837