

METHOD AND SYSTEM FOR PREDICTING OIL ADULTERATION OF AN EDIBLE OIL SAMPLE

Technical Field

5

[0001] Various embodiments relate to a method for predicting oil adulteration of an edible oil sample and an apparatus or system therefor.

Background

10

[0002] Prediction of oil adulteration is widely published. Prior publications described applications, for example, in extra virgin olive oil and sesame oil.

[0003] Oil adulteration may be determined using different techniques, for example, DNA-based marker screening, rapid screening using FT-NIR (Fourier transform near
15 infrared) and Raman spectroscopy, or prediction models.

[0004] In DNA-based marker screening, DNA from different plant origins are used to identify adulteration. The advantage of this technique is that DNA level markers provide a non-refutable evidence of adulteration. However, there are several unfavourable drawbacks to the DNA-based marker screening. For example, this screening technique is
20 a destructive process, and thus samples used for screening cannot be recovered. Further, DNA extraction is a slow process and requires large sample volume. Most importantly, the models constructed from this type of technology are predominantly qualitative as it is difficult to quantify spiked oils from DNA.

[0005] Rapid screening using FT-NIR and Raman spectroscopy uses near infrared light
25 energy measured by either transmission or reflection to detect chemical signatures embedded with a sample. In contrast to the DNA-based marker screening, this technique is a rapid and non-destructive process. Quantitative models may be built to mine insights from rich datasets. The main disadvantage is that features associated with FT-NIR or Raman spectroscopy are not direct evidence that a particular compound exists. Thus, the
30 levels of uncertainty and inaccuracy are relatively high.

[0006] Both DNA-marker based and rapid screening technologies offer a technological platform to provide a feature-rich dataset, however by these techniques do not offer a way to discriminate or predict outcomes. Thus, technologies that provide data acquisition are paired with modeling techniques so that features may be correlated with qualitative constitution of samples or quantitative composition of the samples.

[0007] The most common technique used for oil adulteration is a linear computational method called partial least squares (PLS). For example, PLS1 is the situation when PLS is modeled for a quantitative singular outcome.

[0008] A prior publication has described the use of PLS1 on multiple singular outcomes to simulate prediction of different adulterants. In this publication, four models were built to predict four different adulterants.

[0009] The disadvantage of using PLS1 is that multiple predicted outcomes have to be corroborated and the interpretation of the outcome is not easy or straightforward. For example, oil1 is adulterated 10% into oil-A, but there are four models to predict adulteration of oil1, oil2, oil3, and oil4. The prediction derived from model1 may give the correct 10% adulteration level, but so may models 2-4, each predicting a non-zero adulterant level. It is therefore difficult to establish the fact and distinguish which adulterant is spiked given that there are four models to evaluate.

[0010] The main reason as to why PLS works in general for single oil outcome is that adulterating one oil generates linear features in spectral data. However, when more than one oil is adulterated, or when there are more than one brand of oils, or when oils are obtained from different geographical locations, or when oils are obtained through different processing technologies, the signals generated may be non-linear, thus rendering PLS unsuitable for predicting more complex samples.

[0011] Developments also focused on rapid technologies such as FT-NIR and Raman spectroscopy typically with qualitative determination of adulteration. Quantitative determination of adulteration may be possibly modeled using partial least squares (e.g., PLS-DA and PLS1) and in a way that is specific to one adulterant. This process may require careful interpretation of several modeling results to decipher which specific adulterant is present and errors aggregated over multiple models make calling of

adulteration levels in industry use impractical because it entails a level of human intervention to interpret model results.

5 **[0012]** Thus, there is a need for a method and system of predicting quantitative measures of oil adulteration and capturing invariant adulteration information, thereby addressing at least the problems mentioned hereinabove and providing for industry needs that is generalizable to blind samples and brands.

Summary

10 **[0013]** According to an embodiment, a method for predicting oil adulteration of an edible oil sample is provided. The method may include receiving at least part of a spectral data of the edible oil sample; providing a single prediction model capable of generating a prediction of at least two adulterants in the edible oil sample; and processing the at least part of the spectral data using the single prediction model.

15 **[0014]** According to an embodiment, a computer readable storage medium including computer readable instructions operable when executed by a computer to predict oil adulteration of an edible oil sample is provided. The computer readable instructions may be configured to perform the method according to various embodiments.

20 **[0015]** According to an embodiment, an apparatus or system is provided. The apparatus or system may include a receiving unit configured to receive at least part of a spectral data of an edible oil sample; a memory for storing a single prediction model capable of generating a prediction of at least two adulterants in the edible oil sample; and a processor configured to access the single prediction model stored in the memory to perform steps of the method in accordance with various embodiments for generating a
25 prediction of oil adulteration in the edible oil sample.

Brief Description of the Drawings

30 **[0016]** In the drawings, like reference characters generally refer to like parts throughout the different views. The drawings are not necessarily to scale, emphasis instead generally being placed upon illustrating the principles of the invention. In the following

description, various embodiments of the invention are described with reference to the following drawings, in which:

[0017] FIG. 1A shows a flow chart illustrating a method for predicting oil adulteration of an edible oil sample, according to various embodiments.

5 **[0018]** FIG. 1B shows a schematic view of an apparatus or system for predicting oil adulteration of an edible oil sample, according to various embodiments.

[0019] FIGS. 2A to 2N show graphs representing the final PLS2 experimental results, which reflect the predicted distribution for peanut oil and the adulterants of soybean oil and sunflower oil in terms of amount percentages.

10 **[0020]** FIGS. 3A to 3N show graphs representing the final deep learning experimental results, which reflect the predicted distribution for peanut oil and the adulterants of soybean oil and sunflower oil in terms of amount percentages.

[0021] FIG. 4 shows a graphical representation of data in a linear latent space based on different brands test, in accordance with one embodiment.

15 **[0022]** FIG. 5 shows a graphical representation of data in a non-linear transformed space based on different brands test, in accordance with one embodiment.

[0023] FIGS. 6A to 6J show graphs representing the final PLS2 results for the blind brands test, which reflect the predicted distribution for the peanut oil and the adulterants of soybean oil and sunflower oil in terms of amount percentages.

20 **[0024]** FIGS. 7A to 7J show graphs representing the final deep learning results for the blind brands test, which reflect the predicted distribution for the peanut oil and the adulterants of soybean oil and sunflower oil in terms of amount percentages.

[0025] FIG. 8 shows a schematic representation of the deep learning model, in accordance with various embodiments.

25

Detailed Description

[0026] The following detailed description refers to the accompanying drawings that show, by way of illustration, specific details and embodiments in which the invention
30 may be practiced. These embodiments are described in sufficient detail to enable those skilled in the art to practice the invention. Other embodiments may be utilized and

structural, logical, and electrical changes may be made without departing from the scope of the invention. The various embodiments are not necessarily mutually exclusive, as some embodiments can be combined with one or more other embodiments to form new embodiments.

5 **[0027]** Embodiments described in the context of one of the methods or devices/apparatus are analogously valid for the other methods or devices/apparatus. Similarly, embodiments described in the context of a method are analogously valid for a device, and vice versa.

10 **[0028]** Features that are described in the context of an embodiment may correspondingly be applicable to the same or similar features in the other embodiments. Features that are described in the context of an embodiment may correspondingly be applicable to the other embodiments, even if not explicitly described in these other embodiments. Furthermore, additions and/or combinations and/or alternatives as described for a feature in the context of an embodiment may correspondingly be applicable to the same or similar feature in the other embodiments.

15 **[0029]** In the context of various embodiments, the articles “a”, “an” and “the” as used with regard to a feature or element include a reference to one or more of the features or elements.

[0030] In the context of various embodiments, the phrase “substantially” may include “exactly” and a reasonable variance.

20 **[0031]** In the context of various embodiments, the term “about” or “approximately” as applied to a numeric value encompasses the exact value and a reasonable variance.

[0032] As used herein, the term “and/or” includes any and all combinations of one or more of the associated listed items.

25 **[0033]** As used herein, the phrase of the form of “at least one of A or B” may include A or B or both A and B. Correspondingly, the phrase of the form of “at least one of A or B or C”, or including further listed items, may include any and all combinations of one or more of the associated listed items.

[0034] Various embodiments may provide quantitative detection of oil adulterants using deep-learning models on FT-NIR and Raman spectra.

30 **[0035]** Detection of oil adulterants is a difficult task. Various embodiments may provide an application of integrating deep-learning and FT-NIR/Raman spectroscopy

technologies to detect if multiple adulterants are present in a sample and if so, the quantitative amounts of each adulterant, thus addressing the practical aspect of the industry requirement. Other advantages of deep-learning, for example, over partial least squares are that the technology is able to learn complex non-linear structures in the data which is otherwise not possible with linear models (as in partial least squares, thus addressing the errors associated with constructing multiple PLS1 models).

[0036] In contrast to PLS, deep learning models may handle non-linear structures in the data significantly well. Multiple adulterant oils may be handled as well. Deep learning requires training of datasets that are sufficiently large to at least minimize or avoid over-fitting.

[0037] Various embodiments provide a combination of technologies including FT-NIR, Raman and deep learning models to achieve accurate detection of the type of adulterated oil (qualitatively) under the situation that unknown oils may be adulterated, and accurate quantification of the adulterated matter in situations where two or more adulterants are added. In other words, accurate prediction of oil adulteration of more than one adulterant may be achieved.

[0038] FIG. 1A shows a flow chart illustrating a method for predicting oil adulteration of an edible oil sample 100, according to various embodiments. In FIG. 1A, at Step 102, at least part of a spectral data of the edible oil sample is received. At Step 104, a single prediction model capable of generating a prediction of at least two adulterants in the edible oil sample is provided. At Step 106, the at least part of the spectral data is processed using the single prediction model.

[0039] In the context of various embodiments, the term “processed” or “processing” may mean analyzing, or mathematically evaluating.

[0040] The method 100 may be a method of predicting more than one adulterant in the edible oil sample.

[0041] Various embodiments may provide modeling multiple quantitative adulterant outcomes in one single model. This advantageously addresses the problem faced in using PLS1 where multiple models were applied in predicting the adulteration, and further human interpretation of the results was necessary.

[0042] In one embodiment, the single prediction model may be a linear model. For example, a quantitative prediction of two variables may be realized using PLS2. PLS2 is relatively less common and even more so, apparently not known to be used in predicting oil adulteration. Generally, PLS2 may enable prediction of outcomes that generally do not require corroboration and may be more easily interpreted, as compared to PLS1.

[0043] In one embodiment, the single prediction model may be a non-linear model. The single prediction model may include a deep neural network (DNN). The DNN may be with multiple layers between an input layer and an output layer. The DNN may be a feedforward network in which data flows from the input layer to the output layer without looping back. The input layer may receive one or multiple inputs. The output layer may provide one or multiple outputs. For example, the deep neural network may be supervised.

[0044] The single prediction model may be for modeling complex traits and accounting for different oil brands, processing techniques, geographical locations and other variables. It is believed that these traits cause the (spectra) data to have non-linear structures. Deep learning models are purposed for non-linear pattern discovery, and thus provide a significantly good and useful tool to analyze and assess the non-linearity of the data.

[0045] Various embodiments may provide the development of additional model training strategies by influencing what the model learns from, e.g., complete information from a particular oil brand to predict a random oil brand. As such, the method 100 in accordance with various embodiments may enable the single prediction model to meet industry needs, including the prediction of adulteration generalized over different brand specific traits.

[0046] In various embodiments, the single prediction model may be trained by a matrix of features, wherein each feature corresponds to the intensity of the spectra data at a specific wavenumber. The spectra data may be based on training examples, which may include a processing technique of the edible oil sample, or a brand of the edible oil sample, or a geographical location from which the edible oil sample is obtained, or any combination thereof.

[0047] In various embodiments, the edible oil sample may be selected from the group consisting of a peanut oil sample, an olive oil sample, a corn oil sample, a coconut oil

sample, a cottonseed oil sample, a palm oil sample, a canola oil sample, a safflower oil sample, a sesame oil sample, a soybean oil sample, a sunflower oil sample, a camellia seed oil sample, a linseed (flaxseed) oil sample, and a sample of oil with a relatively higher proportion of certain fatty acids, for example, high erucic acid rapeseed oil, low erucic acid rapeseed oil, or high oleic acid sunflower oil.

[0048] For example, the edible oil sample may be the peanut oil sample.

[0049] In various embodiments, the method 100 may further include generating the prediction of at least two adulterants in the edible oil sample. In other words, the output layer of the single prediction model may provide multiple outputs of the at least two adulterants or variables.

[0050] In the context of various embodiments, the term “generating” may mean “determining”.

[0051] The prediction of the at least two adulterants may be substantially simultaneously generated. No further corroboration of data and no human intervention are required.

[0052] The prediction may include at least a type or a percentage amount of each of the at least two adulterants.

[0053] In various embodiments, the method 100 may further include prior to receiving the at least part of the spectral data, obtaining the spectral data using Fourier transform near-infrared (FT-NIR) or Raman spectroscopy.

[0054] For example, the spectral data may include FT-NIR or Raman spectral data. The spectral data may be representative of tocopherols/tocotrienols band at specific wave numbers, or acid values for fatty acids, or peroxide values, or polymer compounds, amongst others. The specific wave numbers may be from 4476 to 9008.

[0055] The spectral data may be preprocessed, for example, by correction of baseline shifts, standard normal variate transformation for scaling the data within a sample, or multiplicative scattering correction for scaling the data across samples, and a Savitzky-Goly filter. The filter may be further configured to emulate a smoother 1st or 2nd order derivative signal.

[0056] While the method described above is illustrated and described as a series of steps or events, it will be appreciated that any ordering of such steps or events are not to be interpreted in a limiting sense. For example, some steps may occur in different orders

and/or concurrently with other steps or events apart from those illustrated and/or described herein. In addition, not all illustrated steps may be required to implement one or more aspects or embodiments described herein. Also, one or more of the steps depicted herein may be carried out in one or more separate acts and/or phases.

5 **[0057]** Various embodiments further provide a computer readable storage medium including computer readable instructions operable when executed by a computer to predict oil adulteration of an edible oil sample. The computer readable instructions may be configured to perform the method 100, in accordance with various embodiments.

[0058] FIG. 1B shows a schematic view of an apparatus or system 120 for predicting oil
10 adulteration of an edible oil sample, according to various embodiments. In FIG. 1B, the apparatus or system 120 includes a receiving unit 122 configured to receive at least part of a spectral data of an edible oil sample; a memory 124 for storing a single prediction model capable of generating a prediction of at least two adulterants in the edible oil sample; and a processor 126 configured to access the single prediction model stored in
15 the memory 124 to perform steps of the method 100 (FIG. 1A) for generating a prediction of oil adulteration in the edible oil sample, in accordance with various embodiments. The receiving unit 122, the memory 124 and the processor 126 may be in communication with one another, as depicted by lines 128, 130. The communication may be bi-directional.

[0059] The apparatus or system 120 may include the same or like elements or
20 components as those described in the method 100 of FIG. 1A, and as such, the like elements may be as described in the context of the method 100 of FIG. 1A, and therefore the corresponding descriptions are omitted here.

[0060] Examples will be described below in forms of experiments conducted to provide a better understanding of the method 100 and the apparatus or system 120.

25 **[0061]** *Samples and sample preparation*

[0062] Pure peanut oil samples were spiked with adulterant soybean and adulterant sunflower oil. The ratios of adulteration were in 5% increments from 5% to 50%. Binary and ternary blends were allowed, thus increasing the complexity of the experiment but also increasing the feature-rich dataset. For example, ternary blends permit up to the
30 order of $O(n^2)$ combinations.

[0063] Each blend was measured in replicates, using up to two different glass vials and up to six different rotations of the glass vials to obtain example spectra for a given sample. Twenty peanut oil samples blended against soybean and sunflower were being profiled, with a total of 977 spectra generated. The spectra, more specifically FT-NIR spectra, was preprocessed using first derivative and transformed using signal normal variate.

[0064] *Deep learning model*

[0065] A deep learning model is constructed to model the quantitative adulteration of the soybean and sunflower oil in peanut oil. The model includes multiple layers and each layer is associated with a mathematical manipulation. The models find a correct mathematical manipulation to turn an input (e.g., the spectra or spectral data) into an output. The model may identify the correct mathematical manipulation by going through the layers with the calculation of the probability of the output.

[0066] In a simple form, for mathematical manipulation, each the deep learning model includes a matrix of features (f_1 to f_x) and each feature is manipulated by a weight. The weight may represent or may be associated with a level of influence on the output. The features may correspond to the intensity of the FT-NIR spectra at different wave numbers. It should be appreciated that other features, not mentioned herein, may also be considered.

[0067] More specifically, the deep learning model, or interchangeably referred to as deep neural network, accepts as input the spectra. Each of the neural network layer is represented by a number of nodes, initialized with random weights. Data inputs flow through this network by multiplication with the weights associated with a node, through a process of forward propagation. An suitable activation function transforms this result so that non-linearity may be modeled. The activation function may be a sigmoid function or a tanh function. This process may be repeated over the number of layers in the network until finally reaching the output layer where the results are cumulated to represent the quantitative proportions of the oil and its adulterant(s). FIG. 8 shows a schematic representation of the deep learning model 800, in accordance with various embodiments. In FIG. 8, the inputs, i_m , include m number of features, e.g., the intensity of spectra used for modelling. The hidden nodes, $h_{r,k}$ are of size r and k denoting the depth and

width of the hidden nodes 804. Each hidden node 804 (as expanded and depicted in a dotted-line area 806) is computing the sum of weights, w , represented by Σ 808 and applying an activation function f 810 to estimate non-linear relationships in the data. The outputs, o_s , include s number of outputs to estimate, quantification of oils and the adulterants. The initial estimate of the output generally have large errors as the nodes (e.g., 804) of the network 800 was initially assigned random weights. A back propagation algorithm readjusts these weights, w , so that the errors in the outputs 812 may be reduced. This iterative process of forward and backward propagations, changes the weights, w , in the deep neural network 800 until the error is sufficiently small, or if the errors have stopped decreasing. To prevent overfitting of data, a hold-out set of data may usually be reserved for forward propagation to substantially simultaneously check if errors on an independent dataset not trained on the deep neural network 800 also has sufficiently small errors.

[0068] In the event where the spectral is not preprocessed, for example, as described hereinabove, it may be possible to include 1-Dimensional convolutional neural network nodes configured in front of the network so that a suitable filter may be learned to “automatically” preprocess the data for higher quality discriminatory signals.

[0069] *PLS2 model*

[0070] PLS2 model is a linear model. PLS2 corresponds to the case where there are several dependent variables. This is different from PLS1 which corresponds to the case where there is only one dependent variable. For PLS1, features X are projected into a latent variable space T and one dependent variable y . For PLS2, both the features X and dependent variables Y are matrices, and are both projected into latent variable spaces T and U , where a regression model may be used to ensure that the covariances are maximized.

[0071] *Experiments*

[0072] Each of the deep learning model and the PLS2 model were trained by cross-validation and performed in the following two ways:

- (i) partitioning 90% of the data for training and 10% of the samples for blind testing;

(ii) using 90% of the oil brands for training and 10% of the blind brands for testing.

[0073] FIGS. 2A to 2N show graphs representing the final PLS2 results (based on way (i) above), which reflect the predicted distribution for the peanut oil and the adulterants of soybean oil and sunflower oil in terms of amount percentages. The actual ratio for each case is presented in Table 1 below.

[0074] Table 1

FIG.	Actual ratio of peanut oil to soybean oil to sunflower oil
2A	100 : 0 : 0
2B	60 : 0 : 40
2C	70 : 15 : 15
2D	80 : 10 : 10
2E	90 : 10 : 0
2F	50 : 0 : 50
2G	60 : 40 : 0
2H	70 : 30 : 0
2I	80 : 20 : 0
2J	90 : 5 : 5
2K	50 : 50 : 0
2L	70 : 0 : 30
2M	80 : 0 : 20
2N	90 : 0 : 10

[0075] FIGS. 3A to 3N show graphs representing the final deep learning results, which reflect the predicted distribution for the peanut oil and the adulterants of soybean oil and sunflower oil in terms of amount percentages. The actual ratio for each case is presented in Table 2 below.

[0076] Table 2

FIG.	Actual ratio of peanut oil to soybean oil to sunflower oil
3A	100 : 0 : 0

3B	60 : 0 : 40
3C	70 : 15 : 15
3D	80 : 10 : 10
3E	90 : 10 : 0
3F	50 : 0 : 50
3G	60 : 40 : 0
3H	70 : 30 : 0
3I	80 : 20 : 0
3J	90 : 5 : 5
3K	50 : 50 : 0
3L	70 : 0 : 30
3M	80 : 0 : 20
3N	90 : 0 : 10

[0077] The final results from the two models were compared against each other to reflect the contrast in accuracy.

[0078] From the final PLS2 results, it can be observed that the prediction of some samples may not reflect the actual percentages correctly. For example, in FIG. 2B, some predicted values for actual 40% sunflower oil may exceed 55%, which is closer to the actual percentage of peanut oil (at 60%) instead. This may cause some uncertainty in the prediction of the type and amount of adulteration in the sample. However, the PLS2 model may be applicable for oil samples where the adulteration levels are comparatively low and where the distinction of different adulterants are not crucial.

[0079] In contrast, FIG. 3B shows consistent predicted values of about 40% close to the actual 40% sunflower oil using the deep learning model.

[0080] In accordance with the final results, the deep learning model predicts to a much higher accuracy. This is even more apparent when the amount of adulteration is comparable to the amount of the oil sample under test.

[0081] The deep learning model is purposed for non-linear pattern discovery and works on a non-linear transformed space, while the PLS2 model works on a linear latent space.

[0082] FIG. 4 shows a graphical representation of the data in a linear latent space (based on way (ii) above, more specifically, on different brands test). The legend on the right side of FIG. 4 shows the respective percentage amounts of peanut oil to soybean oil to sunflower oil, which have been grouped in Groups a, b, and c to facilitate the identification of the different plot points in FIG. 4. The rectangular labels with white background represent the oil samples of different brands, which possibly involved different processing techniques, and/or possibly obtaining from different geographical locations, amongst other variables. For example, the label SG may represent the oil brand obtained from Santa Gloria (Spain).

[0083] It is observed from Groups a, b, and c that the PLS2 model (which is a linear method) may be able to sufficiently distinguish the different levels of oil adulteration. However, it is further observed from Group a in FIG. 4 that the plot points representing the different oil brands containing low levels of adulterants form over a large spread or distribution in the linear latent space. This suggests that the PLS2 model may not be applicable in determining the oil adulteration if the oil brand under test differs from the oil brand used to built the PLS2 model. In other words, in the linear latent space, the data may discriminate adulteration patterns which is desirable. However, the effects of processing techniques, brands, geographical locations and other variables are not treated as an invariant pattern.

[0084] FIG. 5 shows a graphical representation of the data in a non-linear transformed space (based on way (ii) above, more specifically, on different brands test). The legend on the right side of FIG. 5 shows the respective percentage amounts of peanut oil to soybean oil to sunflower oil, which have been grouped in Groups d, e, and f to facilitate the identification of the different plot points in FIG. 5. Similar to FIG. 4, the rectangular labels with white background in FIG. 5 represent the oil samples of different brands, which possibly involved different processing techniques, and/or possibly obtaining from different geographical locations, amongst other variables.

[0085] It is observed from Groups d, e, and f that the deep learning model may be able to sufficiently distinguish the different levels of oil adulteration. More importantly, it is observed from Group d in FIG. 5 that the plot points of the different oil brands containing low levels of adulterants are consolidated in the non-linear transformed space. This

suggests that the deep learning model is capable of modeling complex traits and accounting for different oil brands, processing techniques, geographical locations and other variables. In other words, in the non-linear transformed space, the effects of brands (geographical location and processing type) become invariant.

5 **[0086]** The deep learning model also accounts for blind brands which is not possible with the PLS2 model.

[0087] FIGS. 6A to 6J show graphs representing the final PLS2 results (based on way (ii) above, more specifically, the blind brands test), which reflect the predicted distribution for the peanut oil and the adulterants of soybean oil and sunflower oil in terms of amount percentages. In FIGS. 6A to 6J, each of the peanut oil and the adulterants is represented by three different groups of shadings 602, 604, 606. Each group 602, 604, 606 may be associated with a different brand, which may be representative of a different geographical location or a different processing type. The actual ratio for each case is presented in Table 3 below.

15 **[0088]** Table 3

FIG.	Actual ratio of peanut oil to soybean oil to sunflower oil
6A	50 : 0 : 50
6B	60 : 40 : 0
6C	80 : 0 : 20
6D	90 : 10 : 0
6E	50 : 50 : 0
6F	70 : 0 : 30
6G	80 : 20 : 0
6H	60 : 0 : 40
6I	70 : 30 : 0
6J	90 : 0 : 10

[0089] FIGS. 7A to 7J show graphs representing the final deep learning results for the blind brands test, which reflect the predicted distribution for the peanut oil and the adulterants of soybean oil and sunflower oil in terms of amount percentages. In FIGS. 7A to 7J, each of the peanut oil and the adulterants is represented by three different groups of

shadings 602, 604, 606. Each group 602, 604, 606 may be associated with a different brand, which may be representative of a different geographical location or a different processing type. The actual ratio for each case is presented in Table 4 below.

[0090] Table 4

FIG.	Actual ratio of peanut oil to soybean oil to sunflower oil
7A	50 : 0 : 50
7B	60 : 40 : 0
7C	80 : 0 : 20
7D	90 : 10 : 0
7E	50 : 50 : 0
7F	70 : 0 : 30
7G	80 : 20 : 0
7H	60 : 0 : 40
7I	70 : 30 : 0
7J	90 : 0 : 10

5

[0091] The final results for the blind brands test from the two models were compared against each other to reflect the contrast in accuracy.

[0092] The deep learning model was found to be able to provide consistent results despite the oil being of a different brand, while the PLS2 model results show that a different oil brand affects the predicted values.

10

[0093] Table 5 shows a summary of the exemplary samples used in each of the three representative groups of shadings 602, 604, 606.

[0094] Table 5

Sample ID	Batch Date(s)	Source	Notes
Noncommercial product - by Wilmar factories (Group 602)			
Xingongyi (QD) (201702)	2017.02	Wilmar Kerry Grain Oil (Qingdao) Co., Ltd.	by enzymatic process
XGY #1 (2017.07.05)	2017.07.05		by microwave roasting
XGY #3 (2017.07.05)	2017.07.05		by microwave roasting
XGY #2 (2017.07.05)	2017.07.05		by microwave roasting
Traditional	2017.07.05		Traditional

(2017.07.05)			processing
Xiaozha (05.07.2018)	2018.07.05		Traditional processing
Xiaozha (QD) SP1 (20170508)	2017.05.08		Traditional processing
High oleic peanut oil (SG) (04.04.2018)	2018.04.04		Traditional processing
HMA	2018.03.23		by mixing process
HMB	2018.03.23		by mixing process
HMC	2018.03.23		by mixing process
HMD	2018.03.23		by mixing process
HME	2018.03.23		by mixing process
HMF	2018.03.23		by mixing process
Soyabean (RSBO) (A)	2018.03.14,2018.03.27, 2018.04.02		Shanghai Wilmar Kerry Grain and Oil Industry Co. Ltd.
Soyabean Oil (RSBO)	2018.03.14,2018.03.27, 2018.04.02		
Sunflower Oil (RSFO) (B)	2018.02.07,2018.03.14, 2018.04.04	3 RSFO samples used	
Sunflower Oil (RSFO)	2018.02.07,2018.03.14, 2018.04.04		
Rapeseed oil (RRSO)	2018.01.26,2018.03.01, 2018.04.01	Wilmar Kerry (Guanghan) Grain and Oil Feeding Co., Ltd.	3 RRSO samples used
Rapeseed Oil (RRSO)	2018.01.26,2018.03.01, 2018.04.01		
Commercial peanut oil (Group 604)			
Hute (QD) (16.01.2017)	2017.01.16	Wilmar Kerry Grain Oil (Qingdao) Co., Ltd.	
Hute (QD, New formula) (11.2.2018)	2018.02.11		
Tuzha (QD) (20170208)	2017.02.08		
Hute 4 (SJZ) (20170106)	2017.01.06	Wilmar Kerry (Shijiazhuang) Grain and Oil Industry Co. Ltd.	
Hute (SJZ) (20170320)	2017.03.20		
Funanpai (BB) (20170603)	2017.06.03	China Oil & Foodstuffs Corporation	
Diyifang (QD) (11.2.2017)	2017.02.11	Qingdao Changshou Food Co. Ltd.	
Duonong (TC) (20170221)	2017.02.11	Jiage Investment (China) Limited	
Leleng (DY)	2017.05.24	Shandong Leyouyou Peanut Oil Technology Co. Ltd.	
Leleng (DY)	2017.05.20		

(20170520)			
Luhua (SZ)	2016.12.23	Shandong Luhua Group Co., Ltd.	
Luhua (sample 27, DG, New Tech)	2017.11.22		
Luhua (QZ) (09.05.2017)	2017.05.09		
Luhua (YT) (14.05.2017)	2017.05.14		
Luhua (DG New Formula) (1.2.18)	2018.02.01		
Peanut oil collected from small workshop (Group 606)			
S1 (20170508)	2017.05.08	Small workshop from Guangxi province	Purchased from non-brand producer.
Wuzhou (WZ) S1 (20170508)	2017.05.08	Small workshop from Guangxi province	
Zao Zhuang (ZZ) (11.02.2017)	2017.02.11	Small workshop from Shandong province	

[0095] To analyze the effects of different oil samples, a separate set of samples containing camellia seed oil and the adulterant of high-oleic sunflower (HOSF) oil was subjected to the two models for a blind brands test. The results obtained (not presented here) lead to the same conclusion as for the peanut oil sample and the adulterants of soybean oil and sunflower oil, thereby reflecting the ability of the deep learning model to accurately predict the oil adulterants in different oil type samples and regardless of the oil brands.

[0096] While the invention has been particularly shown and described with reference to specific embodiments, it should be understood by those skilled in the art that various changes in form and detail may be made therein without departing from the spirit and scope of the invention as defined by the appended claims. The scope of the invention is thus indicated by the appended claims and all changes which come within the meaning and range of equivalency of the claims are therefore intended to be embraced.

CLAIMS

1. A method for predicting quantitative measures of oil adulteration of an edible oil sample, the method comprising:

5 receiving at least part of a spectral data of the edible oil sample, wherein the spectral data comprises Fourier transform near-infrared (FT-NIR) spectral data or Raman spectral data;

providing a single prediction model capable of generating a prediction of at least two adulterants of different types in the edible oil sample, wherein the single prediction
10 model is a non-linear model comprising a deep neural network; and

processing the at least part of the spectral data using the single prediction model.

2. The method of claim 1, wherein the single prediction model is constructed from a matrix of features, wherein each feature corresponds to an intensity of the spectra data at
15 a specific wave number.

3. The method of claim 1 or 2, wherein the edible oil sample is selected from the group consisting of a peanut oil sample, an olive oil sample, a corn oil sample, a coconut oil sample, a cottonseed oil sample, a palm oil sample, a canola oil sample, a safflower
20 oil sample, a sesame oil sample, a soybean oil sample, a sunflower oil sample, a camellia seed oil sample, a linseed (flaxseed) oil sample, and a sample of oil with a relatively higher proportion of certain fatty acids, including high erucic acid rapeseed oil, low erucic acid rapeseed oil, or high oleic acid sunflower oil.

25 4. The method of claim 3, wherein the edible oil sample is the peanut oil sample.

5. The method of any one of the preceding claims, wherein the step of processing the at least part of the spectral data using the single prediction model comprises generating the prediction of the at least two adulterants, each adulterant being of a different oil type
5 from the other adulterant and the edible oil sample.

6. The method of any one of the preceding claims, wherein
the deep neural network comprises an input layer, an output layer comprising
multiple outputs, and multiple layers between the input layer and the output layer, and
10 the prediction of the at least two adulterants are substantially simultaneously
generated at the output layer.

7. The method of any one of the preceding claims, wherein the prediction comprises
a type and a percentage amount of each of the at least two adulterants.

15

8. The method of any one of the preceding claims, further comprising prior to the
step of receiving the at least part of the spectral data, obtaining the spectral data using
Fourier transform near-infrared (FT-NIR) or Raman spectroscopy.

20 9. A computer readable storage medium comprising computer readable instructions
operable when executed by a computer to predict quantitative measures of oil
adulteration of an edible oil sample, the computer readable instructions configured to
perform a method of any one of claims 1 to 8.

25 10. An apparatus or system comprising:
a receiving unit configured to receive at least part of a spectral data of an edible
oil sample, wherein the spectral data comprises Fourier transform near-infrared (FT-NIR)
spectral data or Raman spectral data;
a memory for storing a single prediction model capable of generating a prediction
30 of at least two adulterants of different types in the edible oil sample, wherein the single
prediction model is a non-linear model comprising a deep neural network; and

a processor configured to access the single prediction model stored in the memory to perform steps of a method of any one of claims 1 to 8 for generating a prediction of quantitative measures of oil adulteration in the edible oil sample.