



2021

Assessment of lemon juice quality and adulteration by ultra-high performance liquid chromatography/triple quadrupole mass spectrometry with interactive and interpretable machine learning

Follow this and additional works at: <https://www.jfda-online.com/journal>

 Part of the [Food Science Commons](#), [Medicinal Chemistry and Pharmaceutics Commons](#), [Pharmacology Commons](#), and the [Toxicology Commons](#)



This work is licensed under a [Creative Commons Attribution-Noncommercial-No Derivative Works 4.0 License](#).

Recommended Citation

Lyu, Weiting; Yuan, Bo; Liu, Siyu; Simon, James E.; and Wu, Qingli (2021) "Assessment of lemon juice quality and adulteration by ultra-high performance liquid chromatography/triple quadrupole mass spectrometry with interactive and interpretable machine learning," *Journal of Food and Drug Analysis: Vol. 29 : Iss. 2 , Article 7*.

Available at: <https://doi.org/10.38212/2224-6614.3356>

This Original Article is brought to you for free and open access by Journal of Food and Drug Analysis. It has been accepted for inclusion in Journal of Food and Drug Analysis by an authorized editor of Journal of Food and Drug Analysis.

Assessment of lemon juice quality and adulteration by ultra-high performance liquid chromatography/triple quadrupole mass spectrometry with interactive and interpretable machine learning

Weiting Lyu ^{a,b}, Bo Yuan ^{a,c}, Siyu Liu ^d, James E. Simon ^{a,b,*}, Qingli Wu ^{a,b,c,**}

^a New Use Agriculture and Natural Plant Products Program, Department of Plant Biology and Center for Agricultural Food Ecosystems, Institute of Food, Nutrition & Health, Rutgers University, 59 Dudley Road, New Brunswick, NJ, 08901, USA

^b Department of Medicinal Chemistry, Rutgers University, 160 Frelinghuysen Road, Piscataway, NJ, 08854, USA

^c Department of Food Science, Rutgers University, 65 Dudley Road, New Brunswick, NJ, 08901, USA

^d Citromax Flavors Group, Inc., 444 Washington Ave, Carlstadt, NJ, 07072, USA

Abstract

A total of 81 lemon juices samples were detected using an optimized UHPLC-QqQ-MS/MS method and colorimetric assays. Concentration of 3 organic acids (ascorbic acid, malic acid and citric acid), 3 saccharides (glucose, fructose and sucrose) and 6 phenolic acids (*trans-p*-coumaric acid, 3-hydroxybenzoic acid, 4-hydroxybenzoic acid, 3,4-dihydroxybenzoic acid, caffeic acid) were quantified. Their total polyphenol, antioxidant activity and Ferric reducing antioxidant power were also measured. For the prediction of authentic and adulterated lemon juices and commercially sourced lemonade beverages based on the acquired metabolic profile, machine learning models including linear discriminant analysis, Gaussian naïve Bayes, lasso-regularized logistic regression, random forest (RF) and support vector machine were developed based on training (70%)-cross-validation-testing (30%) workflow. The predicted accuracy on the testing set is 73–86% for different models. Individual conditional expectation analysis (how predicted probabilities change when the feature magnitude changes) was applied for model interpretation, which in particular revealed the close association of RF-probability prediction with nuance characteristics of the density distribution of metabolic features. Using established models, an open-source online dashboard was constructed for convenient classification prediction and interactive visualization in real practice.

Keywords: Interpretable machine learning, *Citrus limon*, Quality control, R shiny application, UHPLC-QqQ-MS/MS

Chemical Compounds: citric acid, PubChem CID: 311, sucrose, PubChem CID: 5988, ascorbic acid, PubChem CID: 54670067, glucose, PubChem CID: 5793, fructose, PubChem CID: 2723872, malic acid, PubChem CID: 525, *trans-p*-coumaric acid, PubChem CID: 637542, 3-hydroxybenzoic acid, PubChem CID: 7420, 4-hydroxybenzoic acid, PubChem CID: 135, 3, 4-dihydroxybenzoic acid, PubChem CID: 72, caffeic acid, PubChem CID: 689043, gallic acid, PubChem CID: 370, trolox, PubChem CID: 40634

1. Introduction

Applications of citrus, one of the most consumed fruits due to its pleasant flavor, aroma and multiple health benefits continues to be of increasing consumer and industry interest.

Lemon, classified as *Citrus limon* (L.) Burm. f., is the third most important citrus species after orange and mandarin, and has been associated with a great variety of health and therapeutic benefits relative to weight loss, skin, eyecare as well as against respiratory disorders and urinary

Received 14 January 2021; revised 12 April 2021; accepted 22 April 2021.
Available online 15 June 2021.

*, ** Corresponding author: Department of Plant Biology, Rutgers University, 59 Dudley Road, New Brunswick, NJ, 08901, USA. Tel: +1(848)-932-6239 (J.E. Simon), +1(848)-932-6238 (Q. Wu).
E-mail addresses: jimsimon@rutgers.edu (J.E. Simon), qlwu@sebs.rutgers.edu (Q. Wu).

<https://doi.org/10.38212/2224-6614.3356>

2224-6614/© 2021 Taiwan Food and Drug Administration. This is an open access article under the CC-BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

disorders, etc [1,2]. However, economically motivated adulteration (EMA) of lemon juice has been and remains a problem [3]. The EMA of juice may not only decrease the bioactive compounds but also poses potential health risks to the consumer [4]. Common adulterations in lemon juice include undeclared addition of sugar and water, peel and/or pulp wash, organic acid such as citric acid, ascorbic acid and malic acid, and juices from other botanically related citrus species [5].

Titrateable acidity (4.5% w/w as citric acid) and soluble solids, together refer to the refractometric sucrose value (6% by weight at 20 °C), are the only two regulatory specifications in America [6]. The titrateable acidity and soluble solids is intended for detection of water addition, but fails in cases of lemon juice dilution with simultaneous addition of inexpensive sugar and citric acid [4]. To improve adulteration detection, other common chemical and physical parameters are routinely analyzed and then compared with the reference data reported by the Association of the Industry of the Juices and Nectars (AJIN) from Fruits and Vegetables [4]. Apart from the traditional approaches, more sophisticated methods have been proposed, such as by means of analysis of carbon isotope ratios in citric acid and other marker compounds in the lemon juices [7]. This method, though appears effective in detection of sugar and water addition, requires highly trained staff and dedicated instrument for chemical analysis. Other approaches involve the calculation of amino acids ratio and the analysis of flavonoids profile [7–10] to judge authenticity but are unable to distinguish adulteration using juices from other botanically related citrus species such as grapefruit and limes due to lack of uniqueness of marker compounds.

Another major limit of past research on lemon juice adulteration detection is in the weak experimental design with those studies that only used lab- or home-made juices pressed straight from fresh lemons, as these processes unfortunately do not necessarily represent the commercial lemonade nor concentrated lemon juices. For example, extra water, sugar and citric acid were usually added to lemonade beverages for a better taste. For industrially sourced concentrated lemon juice, heat-based concentration and pasteurization generally increases the concentration of compounds by removal of water but could meanwhile destroy thermal-labile compounds. Besides, the common use of resin for juice filtration and clearing purpose could easily cause loss of phenolic compounds due to the retention of the resin.

To facilitate practical and robust identification of lemon juice adulteration, we propose a new methodology and provide validation for its application and use. A total of 81 lemon juice samples were analyzed, including industrially sourced authentic and adulterated lemon juice, home-made lemon juice and manually spiked one, as well as marketed lemonade beverages. A chemical fingerprint was established by quantitative analysis of 14 metabolic features using ultra-high performance liquid chromatography with triple quadrupole mass spectrometry (UPLC/QqQ-MS) as well as colorimetric assays for total polyphenols and antioxidants. Based on analyzing the outcomes from the analytical approaches, machine learning models were established for authenticity and identity prediction using the train-cross-validation, and an online open source dashboard was constructed to automate prediction for future new samples.

2. Materials and methods

2.1. Materials and chemical reagents

Citric acid (CTA), sucrose, ascorbic acid (AA), glucose, fructose, malic acid (MA), *trans-p*-coumaric acid (*p*-CA), 3-hydroxybenzoic acid (3-HBA), 4-hydroxybenzoic acid (4-HBA), 3, 4-dihydroxybenzoic acid (3, 4-diHBA), caffeic acid (CFA), gallic acid (GA), trolox standards were purchased from Sigma-Aldrich (St. Louis, MO, USA); ferulic acid (FA) was from ChromaDex Inc. (Irvine, CA, USA). Chemical reagents (all HPLC grade) including ammonium formate, ammonium acetate, ammonium hydroxide, formic acid, acetonitrile (ACN) and methanol were from Fisher Scientific (Fair Lawn, NJ, USA). Reagents (all ACS grade) including ethylenediaminetetraacetic acid (EDTA), Folin Ciocalteu's reagent, sodium carbonate, 2, 3, 5-triphenyltetrazolium chloride (TPTZ), 2, 2'-azino-bis(3-ethylbenzothiazoline-6-sulfonic acid) diammonium salt (ABTS), potassium persulfate, sodium acetate trihydrate, glacial acetic acid, hydrochloric acid were from Fisher Scientific (Fair Lawn, NJ, USA). Pierce™ LC-MS water was from Thermo Fisher Scientific (Waltham, MA, USA).

2.2. Preparation of reference standards

For reference solutions of CTA, *ca.* 10 mg of standard was accurately weighed and diluted to 10 mL using 0.1% formic acid in water. Aliquot of 100 μL stock solution was spiked into 10 mL 0.1% formic acid in water to make the first work solution.

Further serial dilutions were made using the same solvent to 1000–0.06 mg/mL.

For preparation of reference solutions of AA, MA and solutions of saccharides (fructose, glucose and sucrose), *ca.* 10 mg of each standard was accurately weighed and dissolved in 10 mL 0.1% formic acid with saturated EDTA. Aliquot of 100 μ L stock solution was spiked into 10 mL 0.1% formic acid in saturated EDTA solution to make the first work solution. Further serial dilutions were made using the same solvent to give work solutions having 1000–0.06 mg/mL standards.

For reference solutions of phenolic acids including FA, *p*-CA, 3-HBA, 4-HBA, 3,4-diHBA and CFA, each standard *ca.* 15 mg was accurately weighed and prepared in 10 mL 0.1% formic acid in 70% methanol solution. Aliquot of 40 μ L stock solution was spiked into 10 mL 0.1% formic acid in 70% methanol solution to make the first work solution, which was then serially diluted using the same dilution solvent to 4000–0.1 ng/mL as the work solutions.

2.3. Lemon juice and sample preparation

The commercially sourced authentic lemon juices (AULJ) [1–23] and adulterated lemon juices (ADLJ) (27–43) (either filtered and clear, or unfiltered and cloudy) were internally obtained from Citromax Group (Carlstadt, NJ, USA), including 23 different authentic samples and 17 identified adulterated ones. For preparation of hand-made juices, fresh lemon fruits purchased from local supermarkets were ground and squeezed and the collected juice was blended with water (1:1) as the self-made AULJ. To prepare hand-made ADLJ, lab made apple juice was spiked into random AULJ, with AULJ/apple juice volume ratio at 90/10, 70/30 and 50/50. A total of 3 hand-made AULJ (24–26) and 10 hand-made ADLJ (44–53) were prepared. In addition, 28 different commercially sourced lemon juice (54–57) and lemonade (LMND, 58–81) were collected from the commercial market. All juice samples were stored in -20°C and conditioned to room temperature before sample analysis.

For analysis of CTA using LC-MS method (a), all lemon juice samples were mixed with 0.1% formic acid in water (1:10) and sonicated for 15 min 100 μ L of each diluted sample was further diluted with 10 mL of the same solvent followed by sonication for 15 min. An aliquot of 100 μ L diluted sample was mixed with 10 mL 0.1% formic acid in water and vortexed for 30 s. Then 200 μ L solution was diluted to 1 mL, the final sample solution was now 500,000 times more diluted than the original sample. The prepared sample was centrifugated at 12,000 rpm

for 10 min and the supernatant was directly injected into UHPLC.

For instrumental analysis of organic acids and saccharides using LC-MS method (b), all lemon juice samples were mixed with 0.1% formic acid in saturated EDTA solution (1:10) and sonicated for 15 min. An aliquot of 20 μ L diluted sample was mixed with 10 mL of the same dilution solvent as above and sonicated for 15 min. Then 100 μ L solution was diluted to 1 mL, which was now 50,000 times more diluted than the original sample. The prepared sample was centrifugated at 12,000 rpm for 10 min and the supernatant was directly injected into UHPLC.

For analysis of phenolic acids using LC-MS method (c), all lemon juice samples were mixed with 0.1% formic acid in 70% methanol solution (1:1) and sonicated for 15 min. All samples were centrifugated at 12,000 rpm for 10 min and the supernatant was injected directly into UHPLC.

For colorimetric assays (sections 2.5, 2.6 and 2.7), each of the lemon juice samples were diluted using water by three times prior to the assay.

2.4. UHPLC-QqQ-MS/MS methods

The instrument used for chemical analysis was an Agilent 1290 Infinity II UHPLC (Agilent Technology, Palo Alto, CA, USA) hyphenated with 6470 triple quadrupole mass spectrometry with electrospray ionization source (ESI) (Santa Clara, CA, USA). Agilent MassHunter Optimizer (version B.07.00) was used for standard compound-related parameters optimization, and MassHunter Workstation software Data Acquisition (version B.08.00) and Quantitative Analysis (version B.07.01) were used for data processing. The columns used for compound separation were Prontosil C18 AQ (2.1 \times 150 mm, 3 μ m) (for method (a)) and Prontosil C18 AQ (2.1 \times 100 mm, 3 μ m) (for method (b) and method (c)) (Bischoff, Leonberg, Germany).

2.4.1. LC-MS method of CTA

Method (a) was developed and applied for analysis of CTA. For the chromatographic part, the mobile phase A was water, and mobile phase B was 65% ACN in water, both mobile phases modified with 10 mM ammonium acetate with pH adjusted to 9 using ammonium hydroxide. The flow rate was 0.3 mL/min. The gradient was 100% A from 0 to 4 min, 100%–0% A from 4 to 4.5 min, and 0% A from 4.5 to 7.5 min. The column was equilibrated with 100% A for 3 min between injections. The column was thermostatted at 40 $^{\circ}\text{C}$. The autosampler was set to 4 $^{\circ}\text{C}$. The injection volume was 1 μ L. For the MS

part, high-purity nitrogen was used as the nebulizing, drying and sheath gas in the ionization chamber. The nebulizer was set at 30 psi, and the drying gas at 300 °C with flow rate of 13 L/min. The sheath gas was 250 °C with a flow rate of 12 L/min. Negative polarity was applied. The scan mode was dynamic multiple reaction of monitoring (dMRM) optimized using MassHunter Optimizer as prior reported [11], with parameters presented in Fig. 1A.

2.4.2. LC-MS method of MA, AA and saccharides

Method (b) was applied for analysis of MA, AA and saccharides (glucose, fructose and sucrose). The mobile phase A was 30% ACN in water, and mobile phase B was 65% ACN in water, both mobile phases modified with 10 mM ammonium acetate with pH adjusted to 9 using ammonium hydroxide. The flow rate was 0.5 mL/min. All other instrument conditions remained the same as method (a). Dynamic MRM parameters were shown in Fig. 1B.

2.5. LC-MS method of phenolic acids

Method (c) was applied for analysis of phenolic acids. The mobile phase A was 0.1% formic acid in

water, and mobile phase B was 0.1% formic acid in ACN. The gradient was 100%A from 0 to 2 min, 100% to 60%A from 2 to 5 min, and kept at 60%A from 5 to 6 min. The column was equilibrated with 100% A for 1 min between injections. The flow rate was 0.5 mL/min. The column was thermostatted at 30 °C. All other instrumental conditions remained the same as method (a). Dynamic MRM parameters were expressed in Fig. 1C.

2.6. Total polyphenol (TPP) test

The method of TPP assay was based on the previously reported methods [12–14] with modification. 50 mL Folin Ciocalteu's reagent was diluted with distilled water to 500 mL, then 900 µL diluted Folin reagent was mixed with 80 µL diluted juice samples, followed by addition of 400 µL saturated sodium carbonate solution. The reaction system was fully vortexed and let stand still for 1 h. 200 µL supernatant was transferred to a 96 multiple well plate and subject to absorption measurement under 765 nm. As to the calibration curve, *ca.* 30 mg gallic acid was dissolved in 5 mL 70% methanol to make the stock solution, which was then diluted into a

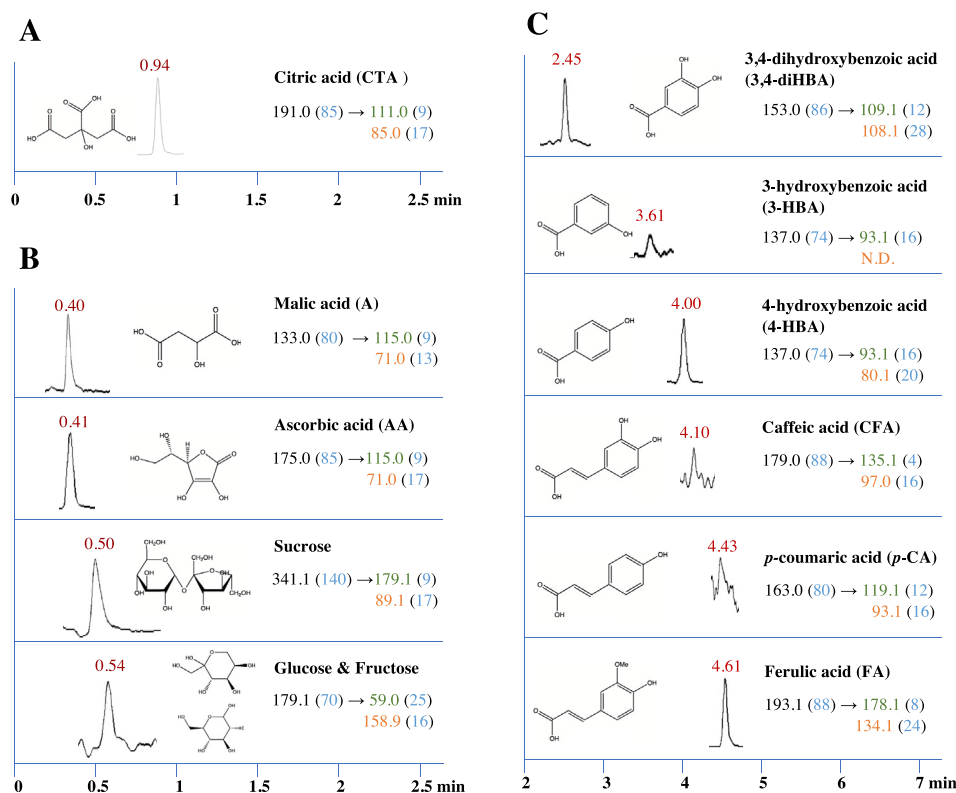


Fig. 1. Representative chromatograms of Sample No. 8 obtained under dynamic multiple reaction monitoring (MRM) mode with analytes' chemical structures. (A), chromatograms obtained using method (a), (B) by method (b) and (C) by method (c). The retention time was shown in red above each peak. MRM transitions are noted in form of: precursor ion (fragmentor voltage) → product ion (collision energy). The quantifier ion is noted in green and the qualifier ion in orange, and the associated voltages are noted in blue.

series of work solutions with concentrations ranging from 0.02 mg/mL to 1.6 mg/mL. The TPP content in samples was expressed as the amount of gallic acid equivalent (GAE)/mg·mL⁻¹.

2.7. Antioxidant activity

Antioxidant activity assay was based on the method described by Re et al. [15] with modification. All the samples were diluted using water by three times prior to the assay. 31.7 mg ABTS and 8.6 mg potassium persulfate were dissolved in 10 mL distilled water. The mixture was then kept in darkness under room temperature for 12–16 h to form stable radical and diluted using water to an absorption of ~1.3 at 734 nm. Next, 200 µL ABTS solution was mixed with 20 µL diluted juice samples and reacted for 15 min under room temperature. The reaction system was fully vortexed and 200 µL supernatant was transferred to a 96 multiple well and the absorption was measured at 734 nm. For the calibration curve, *ca.* 12.5 mg standard Trolox was dissolved in 5 mL pure ethanol as the stock solution and diluted to series of work solutions with concentrations ranging from 0.04 mg/mL to 1.5 mg/mL. Antioxidant activity was expressed as the amount of trolox equivalent (TE)/mg·mL⁻¹.

2.8. Ferric reducing antioxidant power (FRAP) assay

The FRAP assay was based on method reported by Benzie & Strain [16] with modification. Each sample was diluted three times using water. First, 3.1 g sodium acetate trihydrate was dissolved in 16 mL glacial acetic acid and pH was adjusted to 3.6 using sodium hydroxide. Then, the solution was diluted to 1 L by distilled water to make the 300 mM acetate buffer. Second, 27 mg ferric chloride was dissolved in 5 mL water to make the ferric chloride solution. Then 31.2 mg TPTZ was dissolved in 4 mL 40 mM hydrochloric acid solution to prepare the TPTZ solution. Third, 10 mL acetate buffer, 1 mL ferric chloride solution, 1 mL TPTZ solution and 1.2 mL water were mixed to prepare the FRAP reagent. Fourth, 500 µL FRAP reagent, 225 µL water, and 25 µL sample were reacted in 1.5 mL microcentrifuge tube. Next, the tube was left in dark for 30 min. The reaction system was fully vortexed, 200 µL supernatant was transferred to a 96 multiple well and the absorption was measured at 593 nm. For the calibration curve, *ca.* 30 mg vitamin C was dissolved in 5 mL water as the stock solution and was then serially diluted to 0.05–1.7 mg/mL as the

work solutions. The FRAP was expressed as the amount of Vitamin C equivalent (VCE)/mg·mL⁻¹.

2.9. Machine learning

The acquired dataset containing 14 metabolic features across 77 samples, including 26 AULJ (sample codes 1–26) (four lemon juice samples with codes 54–57 purchased from the market were not included in the category of AULJ), 27 ADLJ (27–53) and 24 commercially sourced LMND of different brands (58–81), were subjected to classification study using machine learning methods [17,18]. Principle component analysis (PCA) and linear discriminant analysis (LDA) were first applied on the entire dataset as part of an exploratory data analysis (EDA) procedure. Then the performance of classification prediction was more vigorously tested using the train-test procedure using five different models, including LDA, lasso-regularized logistic regression (LR), Gaussian naïve Bayes (NB), random forest (RF) and support vector machine (SVM). The entire dataset was randomly split with 70/30 ratio using stratified sampling into the training set and testing set. The training set was normalized into z-score prior to modelling, and the testing set was normalized based on the means and variances of the training set. Five-fold cross-validation was applied for hyper-parameter tuning of LR, RF and SVM. The five models were then trained using the whole training set with optimized hyper-parameter if any, and then tested using the prior held-out testing set. In addition, to facilitate model application and prediction in practice, an online open-source dashboard at <https://boyuan.shinyapps.io/LemonClassification/> was constructed with models trained on the entire dataset. R and associated markdown and Shiny web application were applied for statistical computation and visualization, script presentation and online dashboard construction [19]. The associated R script refers to https://yuanbofaith.github.io/Lemon_Juice_Classification2/index.html.

3. Result and discussion

3.1. UHPLC-QqQ-MS/MS method optimization

The solvent used for preparation of lemon juice samples and reference standards was selected based on preliminary study critical to compound's response linearity as well as the chromatographic peak shape, especially for the three organic acids AA, CTA and MA. As such, a factorial design was conducted testing the effect of 6 different sample solvents, including water with or without 0.1%

respective addition of ammonium hydroxide or formic acid, and the respective counterparts with saturated EDTA (Fig. S1). Briefly, MA and CTA presented *ca.* 5–20 times higher signal intensity in basic solvent with 0.1% ammonium hydroxide than otherwise in acidic sample solvent with 0.1% formic acid. For the CTA, the linearity and sensitivity was best in solvent with 0.1% formic acid, and as a result, it was selected as the optimal sample preparation solvent of method (a). The AA presented nonetheless compromised sensitivity and barely any response linearity at all basic solvent due to accelerated degradation. In acidic solvent with 0.1% formic acid, the linearity and sensitivity of AA was much improved, but still *ca.* 30% loss was noted after 4 h storage under 4 °C. The quick loss of AA could be attributed to the reducing substance-initiated Haber-Weiss oxidation cycle, where metal ions (Cu^{2+} or Fe^{3+} etc.) in the sample reduce oxygen and emanate the Fenton's reagent, and produce highly reactive hydroxyl radical [20,21]. Chelation of the metal ions disrupt the oxidation cycle and preserves AA stability. Therefore, EDTA as a metal chelator, was added to the sample solvent and was found highly effective in maintaining AA stability as well as significantly restored the response linearity. Considering the effect on AA and other compounds, 0.1% formic acid in saturated EDTA solutions was finally selected as the optimal sample preparation solvent of method (b). In addition, the mobile phase composition was also optimized by separate examination of the effect of modifiers ammonium formate and ammonium acetate at 5, 10 and 25 mM, respectively with or without pH adjusted to 9 using ammonium hydroxide. And 10 mM ammonium

acetate with pH adjusted to 9 was selected as the best mobile phase for method (a).

3.2. Calibration, linearity and sensitivity

Calibration curve parameters, coefficient of determination (R^2), linear range, lower limit of detection (LLOD) and lower limit of quantification (LLOQ) of all target analytes are shown in Table S1. For quantification of organic acids of CTA, MA and AA, and saccharides of sucrose, glucose and fructose using method (a), great calibration linearity was achieved with R^2 above 0.993. The LLODs were in the range of 0.06–10 ng/mL and LLOQs in the range of 0.2–39.9 ng/mL. For analysis of phenolic acids using method (b), the calibration R^2 was above 0.995 for target compounds. The LLOD and LLOQ were in the ranges of 0.4–36.1 and 1.2–72.2 ng/mL, respectively.

3.3. Juice chemical profile and exploratory data analysis

A representative chromatogram of the target compounds analyzed in an AULJ (Sample No. 8) is presented in Fig. 1. The acquired all metabolic features for 81 lemon juice and lemonade samples were visually displayed in Fig. 2 with original data in Appendix C [22]. We observed that in AUJI samples the concentration of most phenolic compounds in the clear type was lower than in the cloudy type which is presented in Fig. S2. This is likely a result of using resin for juice filtration to make the clear type, where phenolic acids can be easily lost by adsorption to the resin.

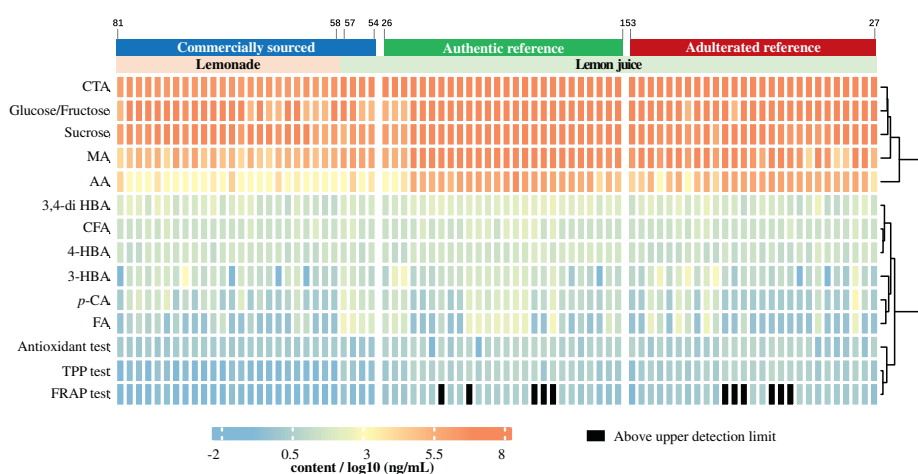


Fig. 2. Heatmap of target compounds concentration and chemistry test data in lemon juice. All data are expressed as \log_{10} (ng/ml). Sample codes are noted on top of the heatmap. Compound abbreviations refer to Fig. 1.

A primal goal of this work is to separate and predict the authenticity of lemon juices, AULJ vs. ADLJ, using machine learning methods based on the acquired metabolic feature so as to fulfill the quality control purpose. Meanwhile, it is also of intellectual interest to investigate how LMND, a most popular derivative product of lemon juices, are chemically and statistically differentiated from their precursors. As part of an exploratory analysis, PCA were applied to the entire dataset (prior to train-test split) for visualization of potential separation (Fig. S3; 3D interactive PCA at https://rpubs.com/Boyuan/lemon_juice_3D_PCA). Briefly, LMND as a palatable and drinkable product were most clearly distinguished from their raw material counterparts by occupying the far positive direction of the first principle component (PC1, 38.1% total variance explained), and also presented more within-group homogeneity. The two lemon juice groups AULJ and ADLJ were smeared across the PC1, reflecting large sample-to-sample variance. Combining both PC1 and PC2 (the latter counting for 13.4% total variance), AULJ and ADLJ were roughly separated by respective occupation of the center and the peripheral region of the two-dimensional PC space. Additional separation was also visible along direction of PC3 (9.9% data variance explained). Different from PCA which seeks maximum presentation of total data variance and somewhat passive in group separation per se, LDA actively seeks the maximum group separation in terms of Mahalanobis distance of the group mean, or equivalently the simpler Euclidean distance on the discriminant scale after LDA transformation. And unlike PCA which had nearly half of the information loss in 2D visualization and 40% loss in 3D plot in this work, LDA presented convenient visualization on a 2D plot without any information loss, and visually rendered better separation of AULJ, ADLJ and LMND. In addition, a linear decision boundary could also be conveniently computed and visualized as shown in Fig. 3. In sum, both PCA and LDA conveyed positive message about the separability of the three classes AULJ, ADLJ and LMND.

3.4. Classification models

Following the exploratory analysis, the classification prediction of the three classes was then vigorously tested using the train-cross-validation-test technique (Fig. 4-A-D). While LDA is an elegant model for dimension reduction and visualization as applied above, as a quick and yet simple classifier it rendered only $63.6 \pm 12.9\%$ cross-validation (CV) accuracy within the training set. The undesirable CV

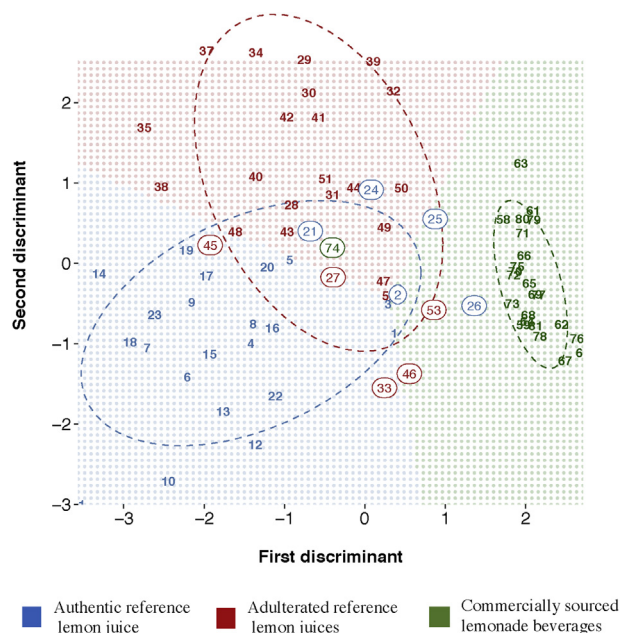


Fig. 3. Linear discriminant analysis (LDA) of authentic and adulterated reference lemon juices and commercially sourced lemonade beverages (from the entire dataset without train-test split). The number on the plot refers to the sample codes, with font color denoting associated identity class. The ellipse marks the boundary containing ca. 80% data points of corresponding class. The grid-background color marks the prediction boundary. Samples with incorrect prediction are highlighted with a circle. Sample No. 53, for instance, was actually an adulterated (red font) sample, but would be incorrectly predicted as a lemonade since it falls into the green-grids region. Note that predicted green region of lemonade beverages is rather wide despite the compact distribution of lemonade samples, less because of the outlier No. 74, but because of the pooled covariance matrix from authentic and adulterated lemon juices, which leads to incorrect predictions at the proximity of prediction boundaries.

accuracy to a great extent was caused by the unwarranted assumption of equal variance-covariance matrix among all three classes, which is clearly untrue as manifested by the compact gathering of LMND but wide dispersion of AULJ and ADLJ on the LDA plot (Fig. 3) (and also easily diagnosed by the univariate density distribution plot in Fig. S4). This directly caused overestimated variance of LMND, and thus AULJ and ADLJ are particularly prone to incorrect prediction as LMND when falling into proximity of the decision boundary with LMND. The model constructed using the entire training set showed prediction accuracy on the testing set of 72.7%. An immediate improvement of LDA would be to evaluate the covariance matrix separately for each class, i.e., using quadratic discriminant analysis (QDA), though this method was not mathematically feasible given the limited size of the available dataset. NB is similar to LDA by both being a generative learning algorithm and use

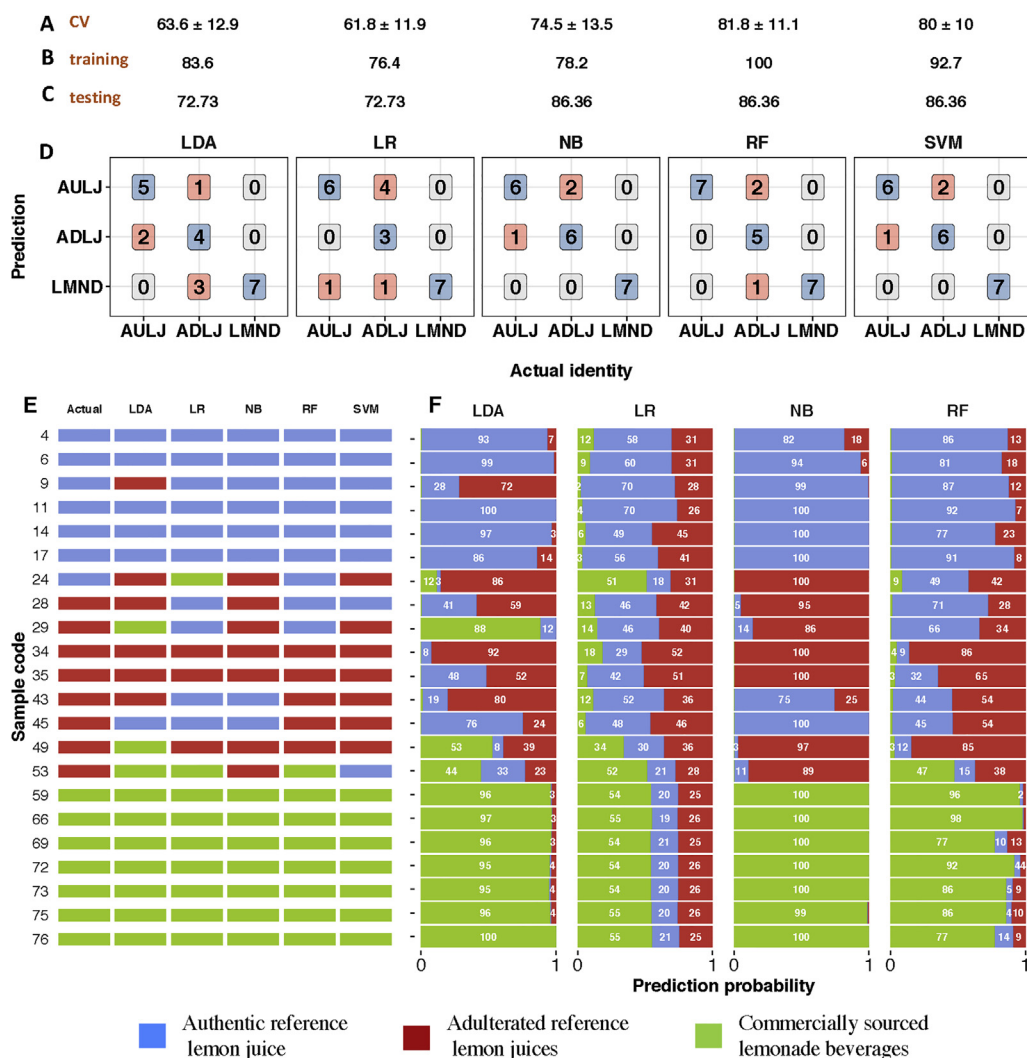


Fig. 4. Classification prediction of authentic lemon juice (AULJ), adulterated lemon juices (ADLJ) and commercially sourced lemonade beverages (LMND) in the testing set. (A), five-fold cross-validation accuracy within the training set. (B), accuracy on the training set. (C), accuracy of the testing set. (D), confusion matrix. (E) sample-wise actual and predicted class label; (F) the predicted percentage probability associated with each class. For model abbreviations, LDA, linear discriminant analysis; LR, lasso-regularized logistic regression; NB, Gaussian naïve Bayes; RF, random forest; SVM, support vector machine. Sample codes and associated information refers to [Appendix C](#).

of Bayesian theory, but unlike LDA based on joint multivariate normal distribution, NB estimates conditional probability on a univariate basis and separately for each class. And NB achieved improved CV accuracy by ca. 10% than LDA, and 86.4% accuracy on the testing set. While NB overcame the aforementioned variance inequality among classes, the assumed conditional independence among all feature variables is less substantiated (Fig. S5). LR is distinguished from the NB and LDA by being a discriminative algorithm, and learns and computes the conditional probability directly without reliance on Bayesian theory, and is robust to deviation from normal distribution. In addition, lasso-regularization helps avoid overfitting. And yet it only scored 72.7% accuracy on the testing set, a

result comparable to LDA. RF is distinct from all the other models of this work by being a classification-tree based ensemble technique, highly efficient in avoidance of over-fitting, and allows for flexible characterization of highly non-linear features. RF achieved accuracy of 81.8 ± 11.1% on the CV set, correctly predicted all training set when fed back to the algorithm, and achieved 86.4% accuracy on the testing set. Radial-kernel based SVM is another powerful tool, distinguished from the prior ones by focused optimization of the decision boundary. It had comparable prediction accuracy on the CV and testing set as RF. Gauged by prediction performance on both the validation sets and the testing set, RF and SVM are two of the most consistent models balanced with high accuracy.

Apart from prediction of the class label per se, a shared nice feature of LDA, LR, NB and RF yet lacking in SVM is their direct prediction of the probability for each class. Such direct association with probability prediction is of value of reference for decision making in real practice. Sample No. 24, for example, while actually being AULJ, was incorrectly predicted by LR as lemondade (Fig. 4-E). The predicted probability of being lemonade, however, was only 51% (Fig. 4-F), with a large Gini index (subtraction of sum of squared probabilities from one) of 0.61. Such prediction is not a decisive one given that a confident correct prediction of the class label is usually supported by a predominant probability of one class over the others, or equivalently a small Gini index value measured on the scale of 0–0.67 in case of three-group separation of this work. Therefore, the prediction result should only be taken lightly instead of being allowed to play a role in decision making. While scrutiny of the probability is an informative way to judge prediction confidence, it is also of importance to interpret the probability in context of characteristics of each model, since the same probability value could mean differently for varied models. For instance, LR correctly predicted 6 out of 7 AULJ and all 7 LMND in the testing set, while the associated probabilities for the correctly predicted samples were only 49–70% and 54–55%, respectively, a probability level too conservative and obscure for NB to achieve confident prediction. Regarding the assignment of probability to each class, NB is probabilistically the most sensitive and aggressive, followed by LDA and RF, with LR being the most conservative (Fig. S6).

3.5. Individual conditional expectation (ICE) analysis

To make the machine learning black box more transparent and intuitive, ICE analysis was constructed to shed light on the association between the metabolic features and the model predicted outcome. ICE analysis is comparable to a simulated single-factor experiment, where predicted probabilities are made in response to a range of values of the feature of interest, while keeping the magnitude of all the other variable unchanged. Such simulation intuitively reveals how the magnitude of the feature of interest impacts predicted likelihood. In addition, the simulation is made for each individual sample in the training set using their own original values of unrelated features, hence one simulation trace for each training example. The multiplicity of simulation on one hand constitutes the many *pseudo*-replicates, which checks consistency of the investigated

feature, while meantime also unveils possible interaction between the investigated feature and other unrelated features [23].

In this work, ICE analysis was conducted to interpret the RF and LR model, shown in Fig. 5 and Appendix A2. Malic acid, for instance, was found an effective marker compound in distinguishing LMND from AULJ and ADLJ. When MA content increased above ca. 1–2 mg/mL, the RF-predicted probability for simulated training examples being LMND rapidly decreased regardless of the condition of all other features, manifested by the almost unanimous L-shape curve in A1-3 (of Fig. 5, same figure for following discussion). Accompanied with the decrease in LMND probability, the ADLJ and AULJ probability rapidly increased (A1-1 and A1-2). Similar pattern is also seen in citric acid and ascorbic acid. This was expected since LMND as a pleasant beverage tends to contain less sourness-inducing compounds than the lemon juice, which as a raw material of LMND is too sour to drink in its original form. Mathematically, the RF predicted probability reflects the density distribution of the feature magnitude of each class. RF not only reflected the pattern of significant changes in the density distribution but also captured nuance details and adjusted the probability prediction accordingly. For example, when the MA content was higher than 20 mg/mL, there were more AULJ samples than ADLJ (A2, region b), and accordingly the predicted probability of ADLJ stepped down and the probability of AULJ simultaneously jumped up (A1-1 and A1-2, region b). For compound 4-HBA, in like manner, the falling cascade of ADLJ probability in region d and e (B1-1) respectively mirrored the valley d and sliding slope e of the density distribution (B2). This sensitive adjustment in probability prediction shows that RF possess highly flexible non-linear capability, a desirable feature lacking in some of other models such as LR, which modeled the probability as a more defined sigmoid curve (Fig. S7).

Feature interaction was observed and this can complicate model interpretation. In the range of MA content of 5–20 mg/mL and for the predicted probability of ADLJ, majority of the simulation traces derived from actual ADLJ samples presented much flat or slightly concave-down profile, while a bunch of traces from actual AULJ slightly concaved up (noted with asterisk, A1-1). Such differences were caused by the difference in the other variables separately characteristic of ADLJ and AULJ, which altered the way in which prediction probability would change in response to alteration of the MA content.

Another informative phenomenon unveiled by RF-ICE plot is that the identity of all simulated

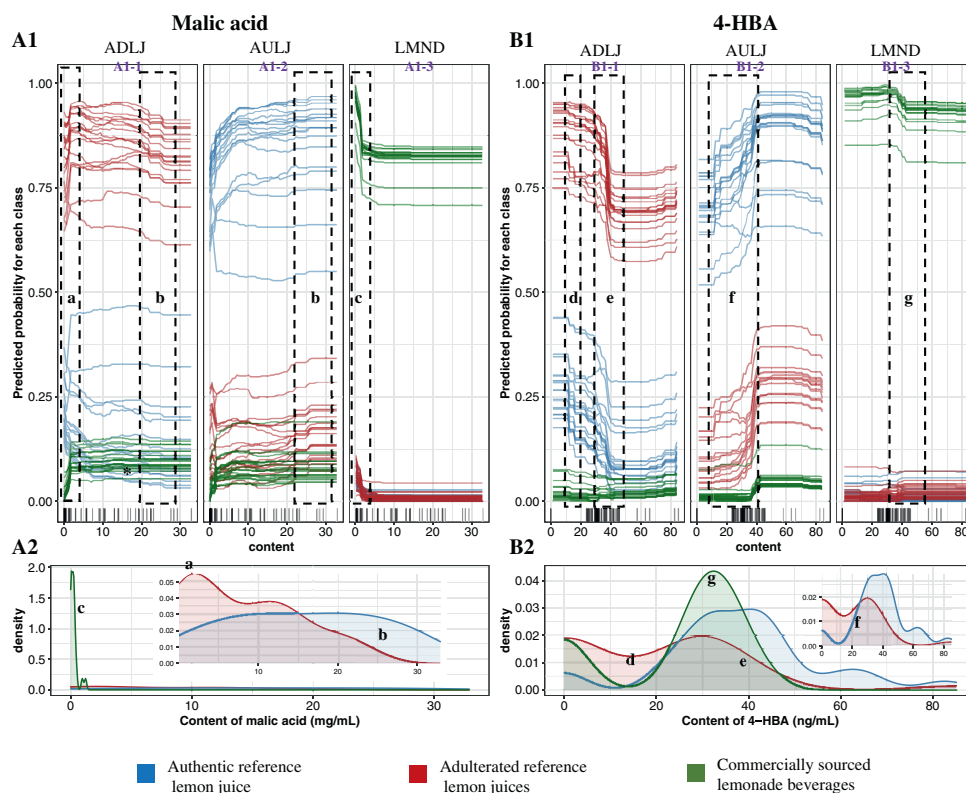


Fig. 5. Interpretation of random forest model with individual conditional expectation (ICE) analysis (A1 and B1) and associated density plots (A2 and B2). Shown is the effect of two metabolic features of malic acid (MA) and 4-hydroxybenzoic acid (4-HBA) on the predicted probability of authentic lemon juices (AULJ), adulterated lemon juices (ADLJ) and commercially sourced lemonade beverages (LMND). In ICE plot, each trace is derived from one sample in the training set. The content magnitude of MA (A1) and 4-HBA (B1) is forced to iterate over a range of grid values ($n = 100$) while keeping all other variable unchanged. The derived new samples are fed into the trained model to predict the probability associated with each class. The probability of being ADLJ, AULJ and LMND is displayed in separated facets/panels (e.g., A1-1, A1-2 and A1-3), and within each facet, the trace color refers to the actual identity of the training sample from which the simulation trace is derived. The barcode on the abscissa of the ICE plots represent the feature density distribution in the original training set. The density plot (A2 and B2) shows the feature distribution density of the entire dataset prior to train-test split. The density plot inset displays a closeup for AULJ and ADLJ. Associations between the density profile and probability prediction is marked with small letters (a–g). Additional ICE analysis refers to [Appendix A2](#).

samples derived from the actual ADLJ were predicted as ADLJ (the associated probabilities were above at least 50%), regardless of the tested feature and associated magnitude. The same rule also applied to AULJ and LMND. It showed that all samples in the original training set were correctly predicted, showing the high flexibility and learning capacity of RF model. More importantly, it also suggested that there is no single metabolic feature that solely determines the identity of a sample, and that identification of an unknown sample must be based on the complete profile of all metabolic features investigated.

3.6. R shiny online dashboard

To facilitate convenient application of machine learning for identity prediction among AULJ, ADLJ and LMND without coding, an online open-source

dashboard was constructed based on R Shiny application. While all aforementioned modelling was based on the training-testing technique, for practical prediction the prior training and testing sets were merged together to train the final models. Built upon the trained models, the R Shiny application automatically predicts the identity and associated probabilities of unknown samples. The training database could be downloaded as template and then updated and uploaded back to the application for prediction of new samples.

Apart from making predictions per se, this application could also serve as a convenient analytical tool to investigate the model performance. Under the functionality of single sample prediction (tab 1), for example, incremental adjustment of the input feature values via the slider-bars animatedly updates the predicted probabilities associated with each class, which generates a dynamic version of the

static ICE analysis. Meanwhile, the dynamic realization of ICE analysis can be readily extended to multiple features as well. For another instructive usage, this application helps simulate how robust models are to the measurement error sourced from upstream data acquisition. Under the batch prediction functionality (tab 2), prior to upload of new data, the default presents the predicted result when the training database is fed back to the built models. The training set can be downloaded and updated and then fed back as new data to the models. When $\pm 20\%$ uniformly distributed random noise is added to the training database as simulation of the error occurring during data collection, *ca.* 94.8–98.7% of prediction remained the same; when $\pm 50\%$ aggressive noise is added, *ca.* 85.7–97.4% prediction remains unchanged. RF and SVM were the least susceptible to noise addition, and 98.7 and 97.4% of prediction remained untouched with $\pm 50\%$ noise (Fig. S8). This simulation suggests that the developed models especially RF and SVM are rather robust to the error sourced from data acquisition.

4. Conclusion

In this study, sensitive and reproducible UHPLC-QqQ-MS/MS methods were developed, and quantitative analysis of 14 metabolic features using the optimized LC-MS methods as well as colorimetric assays was achieved. Parameters in relation to dilution solvent, LC and MS conditions were optimized. For quantification of organic acids of CTA, MA and AA, calibration linearity was greatly improved when using 0.1% formic acid in saturated EDTA solution to prepare and dilute standard solution. The possible mechanism can be proposed to explain the improvement is EDTA can inhibit the Haber-Weiss oxidation cycle as a metal-chelator to protect those organic acids from degradation. The analytical methods developed were then applied to 81 lemon juices samples. For classification of AULJ, ADLJ and LMND, five different models LDA, NB, LR, RF and SVM were constructed, with RF and SVM being two of the most robust models. In particular, RF with its ensemble algorithm flexibly captured the characteristic nuances of the density distribution of the metabolic features of each class and associates the density distribution with the predicted probability as revealed by ICE analysis. Applying the established models, the constructed online dashboard would serve as a convenient and useful tool to facilitate identification of lemon juice authenticity and distinction with the lemonade beverages and provides a robust new approach to detect adulteration and improve product quality.

Declaration of competing interest

The authors declare that there are no known competing financial interests or personal relationships that could influence the work reported in this work.

Authors contribution

Weiting Lyu co-designed and executed the experiment, performed data analysis and co-drafted the manuscript. Bo Yuan performed data analysis, designed the R-shiny application webpage and co-drafted the manuscript. Siyu Liu, James E. Simon and Qingli Wu co-designed and supervised the overarching project, provided funds to conduct the research and revised the manuscript.

Acknowledgement

We thank the New Use Agriculture and Natural Plant Products Lab and the New Jersey Agriculture Experiment Station (Simon's HATCH project 12170) for their support.

References

- [1] Gironés-Vilaplana A, Mena P, García-Viguera C, Moreno DA. A novel beverage rich in antioxidant phenolics: maqui berry (*Aristotelia chilensis*) and lemon juice. *LWT-Food Sci Technol* 2012;47:279–86.
- [2] Mohanapriya M, Ramaswamy L, Rajendran R. Health and medicinal properties of lemon (*Citrus limonum*). *Int J Ayur Herb Med* 2013;3:1095–100.
- [3] Everstine K, Spink J, Kennedy S. Economically motivated adulteration (EMA) of food: common characteristics of EMA incidents. *J Food Protect* 2013;76:723–35.
- [4] Wang Z, Jablonski JE. Targeted and non-targeted detection of lemon juice adulteration by LC-MS and chemometrics. *Food Addit Contam* 2016;33:560–73.
- [5] Cautela D, Laratta B, Santelli F, Trifiro A, Servillo L, Castaldo D. Estimating bergamot juice adulteration of lemon juice by high-performance liquid chromatography (HPLC) analysis of flavanone glycosides. *J Agric Food Chem* 2008;56:5407–14.
- [6] Food U, Administration D. CFR-code of federal regulations title. 2017. p. 21.
- [7] Doner LW. Carbon isotope ratios in natural and synthetic citric acid as indicators of lemon juice adulteration. *J Agric Food Chem* 1985;33:770–2.
- [8] Vandercook CE, Mackey BE, Price RL. New statistical approach to evaluation of lemon juice. *J Agric Food Chem* 1973;21:681–3.
- [9] Grandi R, Trifiro A, Gherardi S, Calza M, Saccani G. Characterization of lemon juice on the basis of flavonoid content. *Fruit processing (Germany)*. 1994.
- [10] Mouly PP, Arzouyan CR, Gaydou EM, Estienne JM. Differentiation of citrus juices by factorial discriminant analysis using liquid chromatography of flavanone glycosides. *J Agric Food Chem* 1994;42:70–9.
- [11] Yuan B, Zhao D, Du R, Kshatriya D, Bello NT, Simon JE, et al. A highly sensitive ultra-high performance liquid chromatography/tandem mass spectrometry method with in-source fragmentation for rapid quantification of raspberry ketone. *J Food Drug Anal* 2019;27:778–85.

- [12] Kao F-J, Chiu Y-S, Chiang W-D. Effect of water cooking on antioxidant capacity of carotenoid-rich vegetables in Taiwan. *J Food Drug Anal* 2014;22:202–9.
- [13] Singleton VL, Rossi JA. Colorimetry of total phenolics with phosphomolybdic-phosphotungstic acid reagents. *Am J Enol Vitic* 1965;16:144–58.
- [14] Yuan B, Byrnes D, Giurleo D, Villani T, Simon JE, Wu Q. Rapid screening of toxic glycoalkaloids and micronutrients in edible nightshades (*Solanum* spp.). *J Food Drug Anal* 2018; 26:751–60.
- [15] Re R, Pellegrini N, Proteggente A, Pannala A, Yang M, Rice-Evans C. Antioxidant activity applying an improved ABTS radical cation decolorization assay. *Free Radic Biol Med* 1999; 26:1231–7.
- [16] Benzie IF, Strain JJ. The ferric reducing ability of plasma (FRAP) as a measure of “antioxidant power”: the FRAP assay. *Anal Biochem* 1996;239:70–6.
- [17] Friedman J, Hastie T, Tibshirani R. *The elements of statistical learning*. New York: Springer series in statistics; 2001.
- [18] Johnson RA, Wichern DW. *Applied multivariate statistical analysis*. Upper Saddle River, NJ: Prentice hall; 2002.
- [19] Ihaka R, Gentleman RR. A language for data analysis and graphics. *J Comput Graph Stat* 1996;5:299–314.
- [20] Mahoney JR, Graf E. Role of alpha-tocopherol, ascorbic acid, citric acid and EDTA as oxidants in model systems. *J Food Sci* 1986;51:1293–6.
- [21] Takano S, Ishiwata S, Nakazawa M, Mizugaki M, Tamai M. Determination of ascorbic acid in human vitreous humor by high-performance liquid chromatography with UV detection. *Curr Eye Res* 1997;16:589–94.
- [22] ComplexHeatmap GZ. Making complex heatmaps. R package version 1.6.0. 2015.
- [23] Molnar C. A guide for making black box models explainable. <https://christophm.github.io/interpretable-ml-book>. 2021.