

1           **Non-destructive measurement and real-time monitoring of apple hardness during**  
2           **ultrasonic contact drying via portable NIR spectroscopy and machine learning**

3   **Amir Malvandi<sup>a</sup>, Ragya Kapoor<sup>b</sup>, Hao Feng<sup>a,b</sup>, Mohammed Kamruzzaman<sup>a\*</sup>**

4   <sup>a</sup>Department of Agricultural and Biological Engineering, University of Illinois at Urbana-  
5   Champaign, Urbana, IL 61801, USA.

6   <sup>b</sup>Department of Food Science and Human Nutrition, University of Illinois at Urbana-  
7   Champaign, Urbana, IL 61801, USA.

8

9   **Abstract**

10   Portable near-infrared **spectrometer** in the spectral range of 900-1700 nm was evaluated for the  
11   first time to assess and monitor apple hardness in real-time during ultrasonic drying. Calibration  
12   models were developed using PLS and ANN, and their performances were evaluated by internal  
13   leave-one-out cross-validation and an external dataset. Several pre-treatments including standard  
14   normal variate (SNV), multiplicative scatter correction (MSC), Savitzky–Golay first and second  
15   derivatives were employed to examine the effects of spectral variations in hardness prediction.  
16   Seven important wavelengths were selected using weighted regression coefficients to develop a  
17   simple MLR model to facilitate the model interpretation and circumvent noise. The models using  
18   PLS, MLR, and ANN with selected wavelengths predicted the apple hardness with  $R^2_p$  of 0.91,  
19   0.91, 0.95, and RMSEP of 14.78, 14.85, and 12.46 N, respectively. The results indicate that  
20   **portable NIR spectrometers** are quite promising for real-time monitoring of apple hardness  
21   during ultrasonic drying.

22   **Keywords:** Near-infrared spectroscopy, Chemometrics, Multivariate analysis, Artificial neural  
23   network, Texture

---

\* *Corresponding author. E-mail: [mkamruz1@illinois.edu](mailto:mkamruz1@illinois.edu), Tel: + 1 217-265-0423, Website: <https://abe.illinois.edu/directory/mkamruz1>*

## 24        1. Introduction

25    Drying is an effective technology in preserving food, agricultural and biological products. It is  
26    also the most energy-intensive process in the food industry, ranked after distillation [1].  
27    Optimizing energy consumption and reducing the drying time along with producing a product  
28    with a better quality have led food engineers to look for advanced energy-efficient dryers with  
29    better control of drying parameters. Ultrasonic contact drying is one of the emerging advanced  
30    non-thermal drying techniques that utilize mechanical vibration instead of thermal energy to  
31    remove moisture and improve the nutrients and color of the final product [2]. However,  
32    continuous monitoring of food properties such as moisture content and texture profile during  
33    drying with conventional methods would be tedious or even impossible. Furthermore, many  
34    measurement techniques are destructive and incompatible for real-time application in advanced  
35    dryers, especially in ultrasonic drying, where a high-frequency vibrating sheet is in contact with  
36    products [3]. Thus, it would be beneficial to employ an accurate, non-destructive measurement  
37    technique for dryer control and production optimization.

38    Near-infrared spectroscopy (NIRS) is a process analytical technology for identifying and  
39    quantifying chemical constituents and is primarily based on the absorption of electromagnetic  
40    radiation in the wavelength range of 750 to 2500 nm. It is well-known for being a low-cost,  
41    rapid, real-time, non-destructive, non-invasive, precise, accurate, repeatable, reproducible, and  
42    sustainable analytical tool [4]. NIRS also provides a deeper understanding of physiochemical  
43    changes of food properties and directly integrates them into control strategies [5]. It primarily  
44    goes together with multivariate techniques and machine learning algorithms to obtain food  
45    quality components from a single sample measurement in a fraction of seconds. In principle, as  
46    the sample is irradiated, various chemical bonds in biological materials absorb or emit

47 electromagnetic waves at different wavelengths. Biological materials, thus, have different  
48 spectral fingerprints because of their precise vibrational frequencies of chemical bonds. The most  
49 prominent absorption bands in the NIR region are overtones and combinations of fundamental  
50 vibrations of hydrogen-based functional groups such as C–H, N–H, O–H, and S–H bonds [6].  
51 NIRS has been successfully applied for measuring the physical and chemical properties of food  
52 and agricultural products [7].

53 Expensive benchtop NIR instruments are commonly used to measure and analyze food and  
54 biological materials [8]. Although **portable NIR spectrometers** typically have lower spectral  
55 resolution and worse wavelength reproducibility than benchtop instruments, portability would be  
56 necessary when the parameters have to be measured in situ [9]. This portability will allow a  
57 spectrometer to be taken to the sample, in contrast to the sample being taken to the benchtop  
58 spectrometer in a laboratory, and will boost NIR application scope throughout the production  
59 process, from raw materials to finished product monitoring. In addition, benchtop instruments  
60 are complex, sizable, expensive, and sometimes difficult to apply in processes. Thus, **portable**  
61 **NIR spectrometers** with a dense and flexible structure have attracted considerable attention.  
62 Different researchers have successfully used the **portable NIR spectrometer** for predicting  
63 soluble solids content and maturity index [10], biodiesel content in biodiesel mixtures [11],  
64 authentication of fengdous and quantitative analysis of mulberry fruits [12], moisture content of  
65 edible coated apple chips [13], and quality parameters and dry matter of mango fruits [14].  
66 Recently, Cruz-Tirado [15] successfully applied NIR spectroscopy to predict eggs' freshness and  
67 concluded that **portable NIR spectrometers are** suitable for online monitoring. The applications  
68 of portable NIR have been summarized in a few recent review articles [4,16].

69 **The texture and flavor of the final product significantly correlate with drying processes** [17,18].  
70 Hardness is an important quality attribute related to the mechanical and microstructural  
71 properties of food [19]. It is also important from a sensory perspective as the overall organoleptic  
72 feeling of the food depends on hardness and firmness, which consumers can quickly identify.  
73 Drying methods and conditions significantly change the hardness and final quality of the  
74 product, so it is necessary to monitor and control these qualitative properties during drying [20].  
75 Texture profile analysis (TPA) is the conventional method for measuring the texture profile and  
76 has been used extensively for different foods, including potato and apple [21] ‘date flesh’ [22],  
77 and meat products [23]. However, these methods are expensive, destructive, and time-consuming  
78 and would not be suitable in food processing industries for rapid and non-destructive quality  
79 evaluation [24]. Li et al. [25] used NIR spectra to assess fracturability and chewiness of Fuji  
80 apples and concluded that NIRS coupled with chemometrics was suitable for non-destructive  
81 measurement of apple texture. McGlone and Kawano [26] investigated NIRS for kiwifruit  
82 firmness and concluded that low pectin concentration (<1% by weight) was the underlying  
83 reason behind the poor performance of the multivariate calibration model on independent data.  
84 Fan et al. [27] evaluated the firmness and soluble solids content of Red Fuji apples via Vis/NIR  
85 transmittance spectroscopy combined with PLS and reached  $R^2=0.81$ . Though their calibration  
86 model performance was not as perfect in predicting firmness as published research by Gómez et  
87 al. [28], their outcomes indicated that it is possible to use NIR spectrometer to monitor the  
88 firmness of apple in real-time. Pissard et al. [29] studied the use of NIRS to determine the dry  
89 matter and total phenolic content (TPC) of varieties of apples in peel and flesh and confirmed the  
90 potential of NIRS for TPC measurement and moisture content evaluation. A summary of NIRS

91 and other non-destructive methods for determining the firmness of apples can be found  
92 elsewhere [30].

93 Non-destructive measurements of moisture content, water activity and chroma during hot air  
94 drying were conducted using NIR spectrometer [31]. Moisture content evaluation of apple chips  
95 during ultrasonic drying was studied by NIR spectrometer and multivariate analysis [32]. NIR  
96 hyperspectral imaging technique was used to monitor melon drying (moisture content  
97 determination) for different drying processes [33]. However, to the best of the author's  
98 knowledge, no study has thus far been dedicated to applying either benchtop or **portable NIR**  
99 **spectrometer** for quantitative or qualitative monitoring of hardness during the ultrasonic contact  
100 drying process. The overall objective of this study was to evaluate the **portable NIR spectrometer**  
101 for monitoring the hardness of apples during the ultrasonic contact drying process. The specific  
102 objectives of this study were a) to develop a cost-effective and rapid approach for measuring the  
103 hardness during drying, b) to improve the prediction accuracy by using different pre-treatments  
104 and machine learning algorithms, and c) to introduce a closed-form equation for rapid and real-  
105 time monitoring of hardness based on few selected wavelengths. Future studies could be  
106 emphasized on the design and evaluation of cloud-based **portable NIR spectrometer** for real-time  
107 monitoring as it was done in [34] and to develop a multi-objective prediction model for  
108 evaluating the quality attributed during drying.

## 109 **2. Materials and Methods:**

110 Figure 1 shows the step-by-step procedure for developing calibration models and feature  
111 extraction from the sample set. In the following sections, each step is explained in detail with  
112 appropriate explanations.

## 113 **2.1 Sample preparation**

114 **A total of 8** Fresh Gala apples (*Malus Domestica*) were purchased from a local market in  
115 Urbana, IL, USA and were stored in a refrigerator at  $4\pm 2$  °C for 24 hours before the experiments.  
116 The apples were washed manually and peeled by a stainless-steel knife before slicing using an  
117 electric food slicer (FS-9001A, Mliter Co., Shenzhen, China) into  $3\pm 0.2$  mm thickness. A cork  
118 borer was used to cut the apple slices into circular discs having a 15 mm diameter. **About 18-20**  
119 **samples (depending on the size of the apple) from each apple were selected, comprising a total of**  
120 **148 samples.** Then, the apple slices were dried using a custom-designed ultrasonic contact dryer,  
121 as shown in Figure 2, at a uniform air temperature of 50°C and airspeed of 2 m/s. The samples  
122 were placed on the top surface of the contact dryer that was vibrated at a frequency of 20 kHz. A  
123 pulse generator has been used to drive the ultrasonic transducer, and digital signal processing  
124 was implemented to create a strong and uniform vibration over the sheet to dry (dehydrate) the  
125 samples. The mechanism of ultrasonic drying is mainly associated with cavitation and sponge  
126 effect due to rapid compression and dilation [35]. Hot air provided by a centrifugal air blower  
127 (Ebmpapst, Farmington, Connecticut, USA) and a heater (Intertek Inc., Model 1DKX3, 8900,  
128 China) flowed parallels over the dryer top surface to carry away moisture coming out of the  
129 sample. **The samples** were then collected at approximately 20 minutes intervals for spectral data  
130 collection and texture analysis.

## 131 **2.2 NIR spectra acquisition**

132 A portable NIR spectrophotometer based on a compact battery-operated evaluation module  
133 (EVM) (DLP2010NIR DLP® NIRscan™ Light, **Texas Instruments (TI), Dallas, Texas, United**  
134 **States**) was used to collect NIR spectra at 228 different wavelengths in the spectral range of 900–  
135 1700 nm. The NIR measurements were done rapidly after taking off the samples from the dryer.

136 To include the variability in apple, the sample rotated twice and averaged to provide a mean  
137 spectrum. The spectral acquisition was performed using a DLP NIRscan Nano GUI and 32 scans  
138 were completed instantaneously and averaged (total measurement time of 32 scans for each  
139 sample was 7.74s). The sample diameter (15 mm) was a little larger than the sensor size to cover  
140 the sensor completely during spectral acquisition.

## 141 **2.3 Texture analysis**

142 Textural properties of the apple samples were determined by a compression test using a TA-HD  
143 Plus Texture Analyzer (Stable Micro Systems Ltd., Godalming, UK) equipped with a 30 kg load  
144 cell. A 4 mm diameter probe (TA-54, Stable Micro Systems Ltd., Godalming, UK) and a Texture  
145 Exponent software (Stable Micro Systems Ltd., Godalming, UK) were used for the analysis. Pre-  
146 test, test, post-test speeds, and strain rate were adjusted to 3, 1, 10 mm/sec, and 90%,  
147 respectively. Due to the destructive nature of the experiment, it was impossible to have  
148 replications for each distinctive sample (or spectra). However, two different but close regions in  
149 each sample were tested, and the results for hardness were averaged.

## 150 **2.4 Multivariate spectral analysis**

### 151 **2.4.1 Spectral pre-treatment**

152 Spectral pre-treatment is considered an integral step in NIRS before calibration model  
153 development. NIR data usually contains some noises and background information in addition to  
154 the sample. Signal-to-noise ratio, baseline drift, wavelength accuracy, signal linearity, sample  
155 homogeneity, sample stability to temperature fluctuations are examples of possible undesired  
156 effects and errors in NIR spectral data [36]. Although there are no universal pre-treatment  
157 techniques to be applied to NIR spectral data, some procedures have been systematically offered

158 in literature to be used as trial and error [37]. In this study, SNV, MSC, Savitzky–Golay first  
159 derivative, and second derivatives were separately applied before calibration modeling. First and  
160 second derivatives are generally used to remove shifts in baseline (multiplicative effects) and  
161 non-linearities [38], while SNV and MSC remove the multiplicative interferences of scattering,  
162 particle size, and the change of light distance [39].

### 163 **2.4.2 Outliers detection**

164 Identification of potential outliers is crucial in any machine learning modeling technique.  
165 Outliers are inherently associated with experimental observations and data collection, primarily  
166 in NIR spectroscopy. Errors caused by human instrumentation or sudden change in  
167 environmental conditions could lead to incorrect measurements. Even one outlier can deteriorate  
168 the calibration model and affect the entire modeling process. Thus, outlier diagnostics algorithms  
169 are employed to identify the samples with spectra sufficiently different from the calibration  
170 spectra. In this study, z-score criteria for outlier detection was applied in a way that the outliers  
171 were detected and removed if the difference between the predicted value ( $y_{predicted}$ ) and the  
172 median of all predicted values ( $\hat{y}_{predicted}$ ) is greater than three standard deviations ( $\sigma(y_{predicted})$ ) of  
173 all predicted values ( $\frac{y_{predicted} - \hat{y}_{predicted}}{\sigma(y_{predicted})} > 3$ ) [40]. However, no outlier was detected in this study.

### 174 **2.4.3 Development of multivariate calibration models**

175 Multivariate analysis of NIR spectra was carried out with two linear models such as partial least  
176 squares (PLS) and multiple linear regression (MLR), and one nonlinear such as artificial neural  
177 network (ANN). PLS is a bilinear factor model which fits a linear regression by projecting the  
178 predictors and responses to a new space, called latent variables (LVs), with the best predictive

179 power. In PLS, overfitting can be avoided by using the leave-one-out cross-validation approach  
180 in the calibration step, and the number of LVs was selected based on the lowest root mean square  
181 error of cross-validation (RMSECV). In contrast to PLS regression which can be implemented to  
182 the whole spectra, MLR is used after feature extraction to find a closed-form equation between  
183 the predicted values and absorbance at the selected feature wavelength. In addition, PLSR finds  
184 the relationship between the reference measurements and the LVs (not physical parameters)  
185 while MLR finds the correlation with the absorption bands which is consequential for  
186 interpretation of the results.

187 In addition, an ANN combined with PLS has been employed to overcome possible shortcomings  
188 of linear regression methods to improve the calibration model accuracy. In this work, PLS was  
189 initially applied to the spectra to find the projected values of PLS components and reduce the  
190 dimension of variables. An appropriate neural network was then used to connect the projected  
191 variables to the reference values through two hidden layers (Fig S1, supplementary information).  
192 Adam optimizer as an adaptive learning rate gradient descent algorithm was employed to train  
193 the neural networks, and the weights of the model were updated by a back-propagation of errors  
194 algorithm. The PLS, MLR, and PLS-ANN along with pre-treatment algorithms were executed  
195 and programmed in Python 3.9.3, and neural network algorithms were implemented using  
196 PyTorch [41].

#### 197 **2.4.4 Model validation and evaluation**

198 After associating the spectra with the reference measurements, data were divided systematically  
199 into 98 calibration (~66%) and 50 validation sets (~34%). Calibration data were used for  
200 developing the calibration model, and the validation data were used as an external validation for  
201 assessing the model's performance. Consequently, the two mostly used validation criteria,

202 namely cross-validation, and external validation, were considered. Subsequently, the LVs were  
203 selected for each pre-treatment, and the calibration models were developed. The predictive  
204 capability of the model was evaluated based on root mean square error of calibration (RSMEC),  
205 root mean square error of prediction (RMSEP), coefficient of determination for calibration ( $R_c^2$ ),  
206 coefficient of determination for prediction ( $R_p^2$ ), and the residual performance deviation (RPD).

#### 207 **2.4.5 Feature selection**

208 One of the critical steps in analyzing a NIR spectrum is variable selection or feature extraction as  
209 the NIRS instruments often measure a spectrum at hundreds or a few thousands of different  
210 variables (wavelengths), and not all of them contain useful information. These variables are  
211 typically categorized as informative, non-informative, and interferential. Feature extraction aims  
212 to identify a subset of spectral wavelengths that capture the highest amount of important  
213 information (smallest possible errors) from the original spectra [42]. It is widely accepted that  
214 excluding the uninformative and interference variables will help to improve the stability,  
215 prediction capability, robustness of the model and decrease the model complexity [43,44]. Many  
216 different views on variable selection have been discussed in the literature [45]; however, no  
217 common strategy exists for variable selection. In this study, the weighted regression coefficients  
218 (BW) from the best calibration model were used to select feature wavelengths [46]. For this  
219 method, the spectra were normalized by their standard deviation to provide the same variability  
220 at each wavelength. Then, the PLS was used to find the regression coefficients of the calibration  
221 model at each wavelength. Due to the same variability of each wavelength, the magnitude of BW  
222 expresses the relative importance of each wavelength in the model [36]. In this paper, the

223 wavelengths with the highest absolute value of BW were used to find a robust and cost-effective  
224 calibration model.

### 225 **3. Results and discussion**

#### 226 **3.1 Spectral features and reference measurements**

227 The collected spectra in this study covered some important wavelengths in the range of 900-1000  
228 nm (water and C-H third overtones), 1150-1250 nm (C-H second overtone), and 1400-1500 nm  
229 (C-H and water second overtone). Although the range of spectra collected from the NIR  
230 equipment was from 900 nm to 1700 nm, a small proportion of signals at both ends of spectra  
231 were removed to eliminate the noise. As a result, the spectral range from 913-1670 nm with 215  
232 spectral variables was used for spectral analysis. Figure 3a illustrates the absorbance spectra of  
233 all 148 collected samples at different drying times. There are three absorption peaks at spectral  
234 windows of around 960-1020 nm, 1150-1250 nm, and 1430-1470 nm. The first peak and the  
235 third peaks are associated with the third and second overtone of water, respectively, while the  
236 second peak represents the second overtone of C-H molecule bonds [47]. So, it is not surprising  
237 that fresh apples with high moisture content have higher peaks in absorbance spectra, and as the  
238 drying continues, the absorbance magnitude decreases. Indeed, NIR spectra are sensitive to  
239 organic molecule bonds (C-H, N-H, and O-H), and their concentration changes the absorbance  
240 spectra. In addition, it is recognized that pectin, containing functional groups of C-H and O-H, is  
241 associated with the texture [25]. Hence, measuring hardness is potentially feasible using NIRS.

242 Table 1 shows the overview of data distribution by the statistical parameters of mean, standard  
243 deviation, minimum and maximum values among calibration and validation datasets. The mean  
244 and standard deviation between the two datasets was around the same, but the range of data for

245 calibration was kept higher than the validation dataset. The large variations in the hardness range  
246 are beneficial for developing a more stable and efficient model. More details on the number of  
247 samples collected at each time interval (20 min) and the distribution of hardness value were  
248 provided in Fig. 3b where the measured hardness value of all the samples during drying at each  
249 time step was shown. The samples were selected at six different time steps, fresh (green), 20min  
250 (light blue), 40min (orange), 60min (gray), 80min (yellow), 100min (dark blue), and taken out  
251 for hardness measurements. A corresponding dry basis moisture content at each time step is also  
252 provided in Fig. 3c. It can be seen that the total variations of moisture content at the final stages  
253 of drying (time>80 min) decreases significantly. However, the standard deviation of hardness  
254 values for fresh samples was small while it increased during drying. For example, at t=100min  
255 where the dry basis moisture content was almost constant in the range of 0.07-0.1, the hardness  
256 ranges from 65N to 190N. Since each time step is associated with a specific range of moisture  
257 content, especially at the final drying stages (Fig. 3c), the high standard deviation of hardness at  
258 the corresponding time signifies that hardness value could not be precisely estimated based on  
259 just the water content. In essence, hardness estimation based on NIRS would be more than a  
260 simple pseudo-correlation measurement based on the water content.

### 261 **3.2 Calibration models at full wavelengths**

262 The key step in developing a robust PLS model is to find the optimum number of LVs to reduce  
263 the dimension of data. A large number of LVs will lead to an over-fitted model while selecting  
264 insufficient LVs would not result in an accurate calibration model (under-fitting problem). The  
265 optimum number of LVs for all PLS models was selected based on the minimum value of  
266 RMSECV; thus, the calibration model was obtained for a wide range of LVs and an objective  
267 function is defined to select the optimum LVs based on the criteria (Fig. S3, supplementary

268 information). The optimum number of LVs, the statistical parameters of interest for calibration,  
269 cross-validation, and prediction set including  $R_c^2$ , RMSEC,  $R_v^2$ , RMSECV,  $R_p^2$ , and RMSEP  
270 were calculated and shown in Table 2. Greater values of RPD,  $R_c^2$ ,  $R_p^2$ , and  $R_v^2$  and lower values  
271 RMSEC, EMSECV, and RMSEP represent a calibration model with stronger predictive  
272 performance. The PLS model for hardness with 11 LVs resulted in an  $R_c^2$  of 0.94, RMSEC of  
273 11.74 N,  $R_v^2$  of 0.91, RMSECV of 14.85 N,  $R_p^2$  of 0.91, and RMSEP of 15.22 N. In general,  $R_p^2$   
274 in the range of 0.82-0.90 is considered to be good and acceptable whereas a value of more than  
275 0.91 would be excellent [48,49]. The  $R_p^2$  obtained in this study was similar to the one obtained for  
276 hardness in apple using a benchtop FT-NIRS [25]. This is an important outcome to show that  
277 **portable NIR spectrometer** is as competitive as the costly benchtop equipment. The proximity of  
278 the values of  $R_v^2$  and  $R_p^2$  along with RMSECV and RMSEP signify that PLS is representative and  
279 can be applied accurately to the unknown data. The results also indicated that the pre-treatments  
280 on the raw data were not effective in improving the prediction capability of the model and the  
281 raw spectral data produced a more accurate calibration model. The SNV pre-treatments, for  
282 example, increased the accuracy of the calibration dataset (higher  $R_c^2$  and lower RMSEC for raw  
283 data) while it slightly decreases the prediction capability of the validation dataset (lower  $R_p^2$  and  
284 higher RMSEP for raw data). In fact, the highest  $R_p^2$  and the lowest RMSEP corresponded to  
285 data without any pre-treatments. This outcome is in line with the ones pertained in literature [36].  
286 NIR spectra follow the principles of Beer-Lambert law which states that there is a linear  
287 relationship between the concentration and absorbance of spectra [50]. However, non-linear

288 features may appear in spectra due to scattering for varying degrees [51]. This happens more  
289 often for food and biological materials, composed of multiple layers with different densities  
290 [52,53]. Several authors also emphasized that there might be non-linear relation between spectra  
291 and hardness of different food products [25,54–56]. However, PLS models essentially select  
292 linear features in the data, which is probably very simplistic compared to the complexity of  
293 hardness. Therefore, an attempt was made to use non-linear regression such as ANN to capture  
294 the possible non-linearity between the absorption spectra and hardness. PLS-ANN uses the same  
295 procedure as PLS except that a nonlinear regression based on ANN is applied instead of linear  
296 regression at the final stage. Figures 4a and 4b illustrate the predicted levels of hardness for  
297 calibration and validation data obtained by PLS and PLS-ANN, respectively. The PLS-ANN  
298 improves the prediction ability of the calibration model with respect to the linear model obtained  
299 using PLS. The statistical variables of PLS-ANN are also summarized in Table 2. Like the PLS  
300 models, pre-treatments had no significant influence on the prediction ability of the PLS-ANN  
301 model and raw data led to a calibration model with better prediction capability. Thus, all the  
302 subsequent analyses will be performed on raw spectral data. With PLS-ANN, the prediction  
303 ability of the calibration model has significantly improved with increasing  $R_p^2$  from 0.91 to 0.95  
304 and decreasing RMSEP from 15.22 N to 11.49 N, a 25% reduction in root mean square error. In  
305 essence, comparing the statistical variables for PLS-ANN and PLS based on raw data, it is clear  
306 that the difference between the RMSEC and RMSEP in PLS-ANN are significantly reduced  
307 from around 30% to almost 0.5%, indicating that the fitted model based on PLS-ANN had better  
308 prediction performance. Since the only difference between PLS-ANN and PLS is the last step in  
309 linear regression analysis, it can be concluded that NIR spectra has a slightly nonlinear  
310 correlation with hardness factor and nonlinear multivariate calibration approaches would be a

311 suitable choice for enhancing the prediction capability. Adoption of a nonlinear relation between  
312 LVs and ANN has also been observed in a firmness prediction of ‘Mazafati’ date by [54].

313 RPD is another important statistical parameter that can be considered to evaluate model  
314 performance. It is defined as the ratio of standard deviation (SD) of the reference values to  
315 RMSEP. A high RPD means that the RMSEP is low compared to SD, indicating that the  
316 calibration model is robust and accurate. For agricultural products, specifically, it has been  
317 suggested that the RPD should be at least 3 or higher. An RPD of between 1.5 and 2 is  
318 considered to have a good prediction ability, while an RPD value of less than 1.5 represents a  
319 poor prediction quality [57]. A low RPD means that the calibration model could barely  
320 discriminate response variables close to each other, leading to only a rough screening between  
321 low and high values [58]. Although the RPD value of PLS shows an excellent prediction  
322 performance (Table 2), PLS-ANN enhances the RPD from 3.26 to 4.32 and provides a perfect  
323 calibration model for hardness prediction.

### 324 **3.3 Important feature extraction**

325 Feature extraction is vital in most machine learning algorithms and model developments,  
326 especially in NIR spectroscopy, where contiguous spectral bands contain redundant and highly  
327 correlated information. Decreasing the number of variables has additional advantages, such as a)  
328 improving the calibration performance by removing noise and uninformative information [59], b)  
329 facilitating the interpretation of the results [60], and c) accelerating the modeling process and  
330 response time [61]. In this study, the weighted regression coefficients (BW) were used to select  
331 the feature wavelengths. The BW of the PLS model at each wavelength is depicted in Figure 4c.  
332 The wavelengths centered at peaks and troughs of the BW curve were selected where the

333 magnitude of regression coefficients was maximum (1042.04, 1118.86, 1245.62, 1288.51,  
334 1345.34, 1387.67, and 1659.31 nm). The other peaks and troughs were found to have an  
335 insignificant effect on the predicted ability of the model (less than 0.5%). The PLS and ANN  
336 models were developed with these seven wavelengths, and the results were also summarized in  
337 Table 2. Due to only seven input features, ANN was applied directly without implementing PLS  
338 projections. Based on these selected wavelengths the developed models had a good performance  
339 for both PLS ( $R_p^2 = 0.91$ , RMSEP=14.78 N, and RPD=3.36) and ANN ( $R_p^2 = 0.94$ , RMSEP=12.54  
340 N, and RPD=4.01). On the other side, the performance of the models developed using seven  
341 feature wavelengths was comparable with the models developed using a full spectral range of  
342 215 wavelengths, indicating that the proper selection of wavelength and feature extraction  
343 decreases the complexity of the model without significant influence on the prediction ability. The  
344 PLS calibration model based on the selected wavelength had a slightly better prediction  
345 performance than the one with full wavelength. This improvement by feature extraction was  
346 demonstrated by different researchers as well [62,63].

### 347 **3.5 Closed-form equation for real-time application**

348 Rapid and cost-effective calculation together with easy interpretation of calibration models  
349 attracted considerable attention in smart industrial systems [36,64]. Consequently, an MLR  
350 model was developed to find a closed-form equation using the selected feature wavelength, and  
351 the results of MLR are shown at the end of Table 2. The MLR has a similar prediction  
352 performance as the PLS since both models use linear regression to connect the feature  
353 wavelengths to the reference value. However, the MLR is simple and easy to interpret and  
354 implement than the latent variable-based PLS model. Using MLR, a closed-form equation for  
355 hardness prediction was obtained in the following form:

$$\hat{H} = -10.55 + 10474.11 \times \lambda_{1042.04} - 14784.99 \times \lambda_{1118.86} - 16409.42 \times \lambda_{1245.62} + 29820.43 \times \lambda_{1288.51} - 13689.55 \times \lambda_{1345.34} + 5279.88 \times \lambda_{1387.67} - 822.20 \times \lambda_{1659.31} \quad (1)$$

$R_p^2=0.91, RMSEP=14.85, RPD=3.34$

356 where  $\lambda$  is the absorbance data at a specific wavelength corresponding to the footnotes and  $\hat{H}$   
 357 is the predicted hardness (N). It is interesting to see that all those overtone bands related to the  
 358 first overtone of C-H bonds (CH, CH<sub>2</sub>, and CH<sub>3</sub>), namely, 1387, 1345, and 1659 nm, have direct  
 359 contributions to the prediction of hardness. The second overtone bands of C-H, e.g., between  
 360 1050 to 1300nm, also directly impacted predicted values of hardness. This is not surprising as  
 361 apple is composed of around 14% carbohydrates (CH<sub>2</sub>O)<sub>n</sub> including pectin, fructose, glucose,  
 362 and fibers.

#### 363 4. Conclusion

364 This is the first reported study for a rapid and cost-effective, non-destructive measurement and  
 365 real-time monitoring approach of apple hardness during ultrasonic contact drying using a  
 366 **portable NIR spectrometer**. Linear PLS and MLR algorithms and nonlinear PLS-ANN were used  
 367 for multivariate analysis of NIR spectra and developing the calibration model for hardness. It  
 368 was shown that the correlation between absorbance spectra and hardness was improved  
 369 significantly using non-linear PLS-ANN hybrid methods. The best linear PLS model could  
 370 predict the hardness of apple with  $R_p^2$  of 0.91 and RMSEP of 15.22 N and the best nonlinear PLS-  
 371 ANN calibration model reached  $R_p^2$  of 0.95 and RMSEP of 11.49 N. Furthermore, weighted  
 372 regression coefficients (BW) were employed to select feature wavelengths and reduce the  
 373 dimensions of the variables. The hardness of the apple was predicted with  $R_p^2$  of 0.91 and  
 374 RMSEP of 14.78 N using only seven wavelengths and a closed-form equation for prediction of

375 hardness based on absorbance spectra was introduced. Using only seven feature wavelengths, a  
376 low-cost filter-based spectroscopic system can be developed for rapid determination and real-  
377 time monitoring of apple hardness during ultrasonic contact drying. It is worth noting that this  
378 study initiates a low-cost non-destructive measurement of the physical property of food material  
379 during ultrasonic drying by a **portable NIR spectrometer**. More studies on other quantities and  
380 qualities of interest including nutrients, water activity, and moisture content have yet to be  
381 investigated. Developing ultrasonic smart dryers using NIRS would be an accessible aim in the  
382 near future.

## 383 **5. Acknowledgments**

384 This study was supported by Agriculture and Food Research Initiative (AFRI) awards no. 2018-  
385 67017-27913 from the USDA National Institute of Food and Agriculture (NIFA) and by DOE  
386 AMO award no. DE-EE0009125. The corresponding author acknowledges funding from the  
387 USDA National Institute of Food and Agriculture, Hatch project ILLU-741-334.

## 388 **References**

- 389 [1] A.S. Mujumdar, Handbook of Industrial Drying, CRC Press, 2014.
- 390 [2] G. Musielak, D. Mierzwa, J. Kroehnke, Food drying enhancement by ultrasound – A review,  
391 Trends in Food Science and Technology. 56 (2016) 126–141.  
392 <https://doi.org/10.1016/j.tifs.2016.08.003>.
- 393 [3] O. Kahraman, A. Malvandi, L. Vargas, H. Feng, Drying characteristics and quality attributes of  
394 apple slices dried by a non-thermal ultrasonic contact drying method, Ultrasonics Sonochemistry.  
395 73 (2021) 105510. <https://doi.org/https://doi.org/10.1016/j.ultsonch.2021.105510>.
- 396 [4] R. Deidda, P.-Y. Sacre, M. Clavaud, L. Coïc, H. Avohou, P. Hubert, E. Ziemons, Vibrational  
397 spectroscopy in analysis of pharmaceuticals: Critical review of innovative portable and handheld  
398 NIR and Raman spectrophotometers, TrAC Trends in Analytical Chemistry. 114 (2019) 251–259.  
399 <https://doi.org/https://doi.org/10.1016/j.trac.2019.02.035>.
- 400 [5] G.J.E. von Gersdorff, B. Kulig, O. Hensel, B. Sturm, Method comparison between real-time  
401 spectral and laboratory based measurements of moisture content and CIELAB color pattern

- 402 during dehydration of beef slices, *Journal of Food Engineering*. 294 (2021) 110419.  
403 <https://doi.org/https://doi.org/10.1016/j.jfoodeng.2020.110419>.
- 404 [6] Y. Ozaki, C.W. Huck, K.B. Beć, Chapter 2 - Near-IR Spectroscopy and Its Applications, in: V.P.B.T.-  
405 M. and L.S. Gupta (Ed.), Elsevier, 2018: pp. 11–38. [https://doi.org/https://doi.org/10.1016/B978-](https://doi.org/https://doi.org/10.1016/B978-0-12-849883-5.00002-4)  
406 [0-12-849883-5.00002-4](https://doi.org/https://doi.org/10.1016/B978-0-12-849883-5.00002-4).
- 407 [7] J.U. Porep, D.R. Kammerer, R. Carle, On-line application of near infrared (NIR) spectroscopy in  
408 food production, *Trends in Food Science & Technology*. 46 (2015) 211–230.  
409 <https://doi.org/https://doi.org/10.1016/j.tifs.2015.10.002>.
- 410 [8] Y. Ozaki, C. Huck, S. Tsuchikawa, S.B. Engelsen, *Near-Infrared Spectroscopy: Theory, Spectral*  
411 *Analysis, Instrumentation, and Applications*, Springer Singapore, 2020.
- 412 [9] T. Temma, K. Hanamatsu, F. Shinoki, Development of a Portable near Infrared Sugar-Measuring  
413 Instrument, *Journal of Near Infrared Spectroscopy*. 10 (2002) 77–83.  
414 <https://doi.org/10.1255/jnirs.324>.
- 415 [10] J.A. Cayuela, C. Weiland, Intact orange quality prediction with two portable NIR spectrometers,  
416 *Postharvest Biology and Technology*. 58 (2010) 113–120.  
417 <https://doi.org/https://doi.org/10.1016/j.postharvbio.2010.06.001>.
- 418 [11] E.M. Paiva, J.J.R. Rohwedder, C. Pasquini, M.F. Pimentel, C.F. Pereira, Quantification of biodiesel  
419 and adulteration with vegetable oils in diesel/biodiesel blends using portable near-infrared  
420 spectrometer, *Fuel*. 160 (2015) 57–63.  
421 <https://doi.org/https://doi.org/10.1016/j.fuel.2015.07.067>.
- 422 [12] H. Yan, Y.-C. Xu, H.W. Siesler, B.-X. Han, G.-Z. Zhang, Hand-held near-infrared spectroscopy for  
423 authentication of fengdous and quantitative analysis of mulberry fruits, *Frontiers in Plant*  
424 *Science*. 10 (2019) 1548.
- 425 [13] R. Kapoor, A. Malvandi, H. Feng, M. Kamruzzaman, Real-time moisture monitoring of edible  
426 coated apple chips during hot air drying using miniature NIR spectroscopy and chemometrics,  
427 *LWT*. (2021) 112602.
- 428 [14] E.J.N. Marques, S.T. de Freitas, M.F. Pimentel, C. Pasquini, Rapid and non-destructive  
429 determination of quality parameters in the ‘Tommy Atkins’ mango using a novel handheld near  
430 infrared spectrometer, *Food Chemistry*. 197 (2016) 1207–1214.  
431 <https://doi.org/https://doi.org/10.1016/j.foodchem.2015.11.080>.
- 432 [15] J.P. Cruz-Tirado, M. Lucimar da Silva Medeiros, D.F. Barbin, On-line monitoring of egg freshness  
433 using a portable NIR spectrometer in tandem with machine learning, *Journal of Food Engineering*.  
434 306 (2021) 110643. <https://doi.org/https://doi.org/10.1016/j.jfoodeng.2021.110643>.
- 435 [16] C.A.T. dos Santos, M. Lopo, R.N.M.J. Páscoa, J.A. Lopes, A Review on the Applications of Portable  
436 Near-Infrared Spectrometers in the Agro-Food Industry, *Applied Spectroscopy*. 67 (2013) 1215–  
437 1233.

- 438 [17] Y. Sun, M. Zhang, A. Mujumdar, Berry Drying: Mechanism, Pretreatment, Drying Technology,  
439 Nutrient Preservation, and Mathematical Models, *Food Engineering Reviews* 2019 11:2. 11  
440 (2019) 61–77. <https://doi.org/10.1007/S12393-019-9188-3>.
- 441 [18] L.E. Kurozawa, M.D. Hubinger, K.J. Park, Glass transition phenomenon on shrinkage of papaya  
442 during convective drying, *Journal of Food Engineering*. 108 (2012) 43–50.  
443 <https://doi.org/10.1016/J.JFOODENG.2011.07.033>.
- 444 [19] J. Chen, A. Rosenthal, 1 - Food texture and structure, in: J. Chen, A.B.T.-M.F.T. Rosenthal (Eds.),  
445 Woodhead Publishing Series in Food Science, Technology and Nutrition, Woodhead Publishing,  
446 2015: pp. 3–24. <https://doi.org/https://doi.org/10.1016/B978-1-78242-333-1.00001-2>.
- 447 [20] A. Martynenko, M.A. Janaszek, Texture Changes During Drying of Apple Slices, *Drying*  
448 *Technology*. 32 (2014) 567–577. <https://doi.org/10.1080/07373937.2013.845573>.
- 449 [21] M. Alvarez, W. Canet, M. López, Influence of deformation rate and degree of compression on  
450 textural parameters of potato and apple tissues in texture profile analysis, *European Food*  
451 *Research and Technology*. 215 (2002) 13–20. <https://doi.org/10.1007/s00217-002-0515-0>.
- 452 [22] M.S. Rahman, S.A. Al-Farsi, Instrumental texture profile analysis (TPA) of date flesh as a function  
453 of moisture content, *Journal of Food Engineering*. 66 (2005) 505–511.  
454 <https://doi.org/https://doi.org/10.1016/j.jfoodeng.2004.04.022>.
- 455 [23] O. Martínez, J. Salmerón, M.D. Guillén, C. Casas, Texture profile analysis of meat products treated  
456 with commercial liquid smoke flavourings, *Food Control*. 15 (2004) 457–461.  
457 [https://doi.org/https://doi.org/10.1016/S0956-7135\(03\)00130-0](https://doi.org/https://doi.org/10.1016/S0956-7135(03)00130-0).
- 458 [24] S. Abasi, S. Minaei, B. Jamshidi, D. Fathi, Dedicated non-destructive devices for food quality  
459 measurement: A review, *Trends in Food Science & Technology*. 78 (2018) 197–205.  
460 <https://doi.org/https://doi.org/10.1016/j.tifs.2018.05.009>.
- 461 [25] G. Li, Y. Ren, X. Ren, X. Zhang, Non-destructive measurement of fracturability and chewiness of  
462 apple by FT-NIRS, *Journal of Food Science and Technology*. 52 (2015) 258–266.  
463 <https://doi.org/10.1007/s13197-013-0990-2>.
- 464 [26] V.A. McGlone, S. Kawano, Firmness, dry-matter and soluble-solids assessment of postharvest  
465 kiwifruit by NIR spectroscopy, *Postharvest Biology and Technology*. 13 (1998) 131–141.  
466 [https://doi.org/https://doi.org/10.1016/S0925-5214\(98\)00007-6](https://doi.org/https://doi.org/10.1016/S0925-5214(98)00007-6).
- 467 [27] G. Fan, J. Zha, R. Du, L. Gao, Determination of soluble solids and firmness of apples by Vis/NIR  
468 transmittance, *Journal of Food Engineering*. 93 (2009) 416–420.  
469 <https://doi.org/https://doi.org/10.1016/j.jfoodeng.2009.02.006>.
- 470 [28] A.H. Gómez, Y. He, A.G. Pereira, Non-destructive measurement of acidity, soluble solids and  
471 firmness of Satsuma mandarin using Vis/NIR-spectroscopy techniques, *Journal of Food*  
472 *Engineering*. 77 (2006) 313–319.
- 473 [29] A. Pissard, V. Baeten, P. Dardenne, P. Dupont, M. Lateur, Use of NIR spectroscopy on fresh apples  
474 to determine the phenolic compounds and dry matter content in peel and flesh, *BASE*. (2018).

- 475 [30] Z. Fathizadeh, M. Aboonajmi, S.R. Hassan-Beygi, Nondestructive methods for determining the  
476 firmness of apple fruit flesh, *Information Processing in Agriculture*. (2021).  
477 <https://doi.org/https://doi.org/10.1016/j.inpa.2020.12.002>.
- 478 [31] R. Moschetti, F. Raponi, S. Ferri, A. Colantoni, D. Monarca, R. Massantini, Real-time monitoring of  
479 organic apple (var. Gala) during hot-air drying using near-infrared spectroscopy, *Journal of Food*  
480 *Engineering*. 222 (2018) 139–150.  
481 <https://doi.org/https://doi.org/10.1016/j.jfoodeng.2017.11.023>.
- 482 [32] A. Malvandi, H. Feng, M. Kamruzzaman, Application of NIR spectroscopy and multivariate  
483 analysis for Non-destructive evaluation of apple moisture content during ultrasonic drying,  
484 *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*. 269 (2022) 120733.  
485 <https://doi.org/10.1016/J.SAA.2021.120733>.
- 486 [33] J.M.S. Netto, F.A. Honorato, P.M. Azoubel, L.E. Kurozawa, D.F. Barbin, Evaluation of melon drying  
487 using hyperspectral imaging technique in the near infrared region, *LWT*. 143 (2021) 111092.  
488 <https://doi.org/https://doi.org/10.1016/j.lwt.2021.111092>.
- 489 [34] T. Wang, J. Chen, Y. Fan, Z. Qiu, Y. He, SeeFruits: Design and evaluation of a cloud-based ultra-  
490 portable NIRS system for sweet cherry quality detection, *Computers and Electronics in*  
491 *Agriculture*. 152 (2018) 302–313.
- 492 [35] Y. Zhang, N. Abatzoglou, Review: Fundamentals, applications and potentials of ultrasound-  
493 assisted drying, *Chemical Engineering Research and Design*. 154 (2020) 21–46.  
494 <https://doi.org/https://doi.org/10.1016/j.cherd.2019.11.025>.
- 495 [36] M. Kamruzzaman, Y. Makino, S. Oshita, Parsimonious model development for real-time  
496 monitoring of moisture in red meat using hyperspectral imaging, *Food Chemistry*. 196 (2016)  
497 1084–1091. <https://doi.org/https://doi.org/10.1016/j.foodchem.2015.10.051>.
- 498 [37] H. Cen, Y. He, Theory and application of near infrared reflectance spectroscopy in determination  
499 of food quality, *Trends in Food Science & Technology*. 18 (2007) 72–83.  
500 <https://doi.org/https://doi.org/10.1016/j.tifs.2006.09.003>.
- 501 [38] Å. Rinnan, F. van den Berg, S.B. Engelsen, Review of the most common pre-processing techniques  
502 for near-infrared spectra, *TrAC Trends in Analytical Chemistry*. 28 (2009) 1201–1222.  
503 <https://doi.org/https://doi.org/10.1016/j.trac.2009.07.007>.
- 504 [39] D.W. Sun, *Modern Techniques for Food Authentication*, Elsevier Science, 2018.
- 505 [40] P.J. Rousseeuw, M. Hubert, Robust statistics for outlier detection, *Wiley Interdisciplinary*  
506 *Reviews: Data Mining and Knowledge Discovery*. 1 (2011) 73–79.
- 507 [41] A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga, A.  
508 Lerer, *Automatic differentiation in pytorch*, (2017).
- 509 [42] Z. Xia, C. Zhang, H. Weng, P. Nie, Y. He, Sensitive Wavelengths Selection in Identification of  
510 *Ophiopogon japonicus* Based on Near-Infrared Hyperspectral Imaging Technology, *International*  
511 *Journal of Analytical Chemistry*. 2017 (2017) 6018769. <https://doi.org/10.1155/2017/6018769>.

- 512 [43] R.M. Balabin, S. V Smirnov, Variable selection in near-infrared spectroscopy: Benchmarking of  
513 feature selection methods on biodiesel data, *Analytica Chimica Acta*. 692 (2011) 63–72.  
514 <https://doi.org/https://doi.org/10.1016/j.aca.2011.03.006>.
- 515 [44] H. Pu, M. Kamruzzaman, D.-W. Sun, Selection of feature wavelengths for developing  
516 multispectral imaging systems for quality, safety and authenticity of muscle foods-a review,  
517 *Trends in Food Science & Technology*. 45 (2015) 86–104.  
518 <https://doi.org/https://doi.org/10.1016/j.tifs.2015.05.006>.
- 519 [45] Z. Xiaobo, Z. Jiewen, M.J.W. Povey, M. Holmes, M. Hanpin, Variables selection methods in near-  
520 infrared spectroscopy, *Analytica Chimica Acta*. 667 (2010) 14–32.  
521 <https://doi.org/https://doi.org/10.1016/j.aca.2010.03.048>.
- 522 [46] G. Elmasry, M. Kamruzzaman, D.-W. Sun, P. Allen, Principles and applications of hyperspectral  
523 imaging in quality evaluation of agro-food products: a review, *Critical Reviews in Food Science  
524 and Nutrition*. 52 (2012) 999–1023.
- 525 [47] D.F. Barbin, A.L. de S.M. Felicio, D.-W. Sun, S.L. Nixdorf, E.Y. Hirooka, Application of infrared  
526 spectral techniques on quality and compositional attributes of coffee: An overview, *Food  
527 Research International*. 61 (2014) 23–32.  
528 <https://doi.org/https://doi.org/10.1016/j.foodres.2014.01.005>.
- 529 [48] J.-H. Cheng, D.-W. Sun, Partial Least Squares Regression (PLSR) Applied to NIR and HSI Spectral  
530 Data Modeling to Predict Chemical Properties of Fish Muscle, *Food Engineering Reviews*. 9 (2017)  
531 36–49. <https://doi.org/10.1007/s12393-016-9147-1>.
- 532 [49] T. Næs, T. Isaksson, T. Fearn, T. Davies, A user-friendly guide to multivariate calibration and  
533 classification, *NIR Chichester*, 2002.
- 534 [50] E.W. Ciurczak, B. Igne, J. Workman Jr, D.A. Burns, *Handbook of near-infrared analysis*, CRC press,  
535 2021.
- 536 [51] B.G. Osborne, *Near-infrared spectroscopy in food analysis*, *Encyclopedia of Analytical Chemistry:  
537 Applications, Theory and Instrumentation*. (2006).
- 538 [52] H. Cen, R. Lu, F. Mendoza, R.M. Beaudry, Relationship of the optical absorption and scattering  
539 properties with mechanical and structural properties of apple tissue, *Postharvest Biology and  
540 Technology*. 85 (2013) 30–38.  
541 <https://doi.org/https://doi.org/10.1016/j.postharvbio.2013.04.014>.
- 542 [53] Z. Wang, R. Van Beers, B. Aernouts, R. Watté, P. Verboven, B. Nicolai, W. Saeys, Microstructure  
543 affects light scattering in apples, *Postharvest Biology and Technology*. 159 (2020) 110996.  
544 <https://doi.org/https://doi.org/10.1016/j.postharvbio.2019.110996>.
- 545 [54] S.A. Mireei, S.S. Mohtasebi, M. Sadeghi, Comparison Of Linear And Non-Linear Calibration  
546 Models For Non-Destructive Firmness Determining Of ‘Mazafati’ Date Fruit By Near Infrared  
547 Spectroscopy, *International Journal of Food Properties*. 17 (2014) 1199–1210.  
548 <https://doi.org/10.1080/10942912.2012.678533>.

- 549 [55] M. Kamruzzaman, G. ElMasry, D.-W. Sun, P. Allen, Non-destructive assessment of instrumental  
550 and sensory tenderness of lamb meat using NIR hyperspectral imaging, *Food Chemistry*. 141  
551 (2013) 389–396. <https://doi.org/https://doi.org/10.1016/j.foodchem.2013.02.094>.
- 552 [56] K.M. Morisseau, C.T. Rhodes, Near-Infrared Spectroscopy as a Nondestructive Alternative to  
553 Conventional Tablet Hardness Testing, *Pharmaceutical Research*. 14 (1997) 108–111.  
554 <https://doi.org/10.1023/A:1012071904673>.
- 555 [57] L.P. D’Acqui, A. Pucci, L.J. Janik, Soil properties prediction of western Mediterranean islands with  
556 similar climatic environments by means of mid-infrared diffuse reflectance spectroscopy,  
557 *European Journal of Soil Science*. 61 (2010) 865–876.  
558 <https://doi.org/https://doi.org/10.1111/j.1365-2389.2010.01301.x>.
- 559 [58] D. Cozzolino, M. Parker, R.G. Damberg, M. Herderich, M. Gishen, Chemometrics and visible-near  
560 infrared spectroscopic monitoring of red wine fermentation in a pilot scale, *Biotechnology and  
561 Bioengineering*. 95 (2006) 1101–1107.
- 562 [59] M. Marois, S.L. Jacques, K.D. Paulsen, Optimal wavelength selection for optical spectroscopy of  
563 hemoglobin and water within a simulated light-scattering tissue, *Journal of Biomedical Optics*. 23  
564 (2018) 41202. <https://doi.org/10.1117/1.JBO.23.4.041202>.
- 565 [60] A. Vellido, J.D. Martín-Guerrero, P.J.G. Lisboa, Making machine learning models interpretable., in:  
566 *ESANN, Citeseer*, 2012: pp. 163–172.
- 567 [61] X. Guan, J. Liu, K. Huang, J. Kuang, D. Liu, Evaluation of moisture content in processed apple chips  
568 using NIRS and wavelength selection techniques, *Infrared Physics & Technology*. 98 (2019) 305–  
569 310. <https://doi.org/https://doi.org/10.1016/j.infrared.2019.01.010>.
- 570 [62] C. Abrahamsson, J. Johansson, A. Sparén, F. Lindgren, Comparison of different variable selection  
571 methods conducted on NIR transmission measurements on intact tablets, *Chemometrics and  
572 Intelligent Laboratory Systems*. 69 (2003) 3–12. [https://doi.org/https://doi.org/10.1016/S0169-  
573 7439\(03\)00064-9](https://doi.org/https://doi.org/10.1016/S0169-7439(03)00064-9).
- 574 [63] E. Bonah, X. Huang, J.H. Aheto, R. Yi, S. Yu, H. Tu, Comparison of variable selection algorithms on  
575 vis-NIR hyperspectral imaging spectra for quantitative monitoring and visualization of bacterial  
576 foodborne pathogens in fresh pork muscles, *Infrared Physics & Technology*. 107 (2020) 103327.  
577 <https://doi.org/https://doi.org/10.1016/j.infrared.2020.103327>.
- 578 [64] H.-J. He, D. Wu, D.-W. Sun, Non-destructive and rapid analysis of moisture distribution in farmed  
579 Atlantic salmon (*Salmo salar*) fillets using visible and near-infrared hyperspectral imaging,  
580 *Innovative Food Science & Emerging Technologies*. 18 (2013) 237–245.  
581 <https://doi.org/https://doi.org/10.1016/j.ifset.2013.02.009>.

582

583

584

585

586 **Figure Captions:**

587 **Figure 1.** Full procedure for developing the calibration models and feature extraction for  
588 selecting important wavelengths.

589 **Figure 2.** Schematic of ultrasonic contact dryer.

590 **Figure 3.** (a) Spectra of apple samples during drying. Blue, red, and yellow lines represent the  
591 spectrum of one sample for fresh, semi-dried, and fully dried, respectively. (b) Measured  
592 hardness value (N) at each time interval. (c) **Dry basis moisture content at each time interval**  
593 **(drying curve).**

594 **Figure 4.** Levels of predicted hardness versus measured hardness for calibration and validation  
595 data via (a) PLS, (b) PLS-ANN, and (c) plot of the weighted regression coefficients (BW)  
596 versus wavelength (Savitzky–Golay filter is used to smooth and remove the unnecessary noise in  
597 the regression coefficient curve).

598

**Table 1.** Summary of statistics of Gala apple data in calibration and validation dataset

Dataset	No of samples	Mean (N)	Standard deviation (N)	min-max (N)
Calibration	98	81.02	49.88	1.42-189.91
Validation	50	81.76	49.65	10.90-173.21
Total samples	148	81.27	49.80	1.42-189.91

599

600

**Table 2.** The calibration, cross-validation, and prediction results of PLS, MLR and ANN

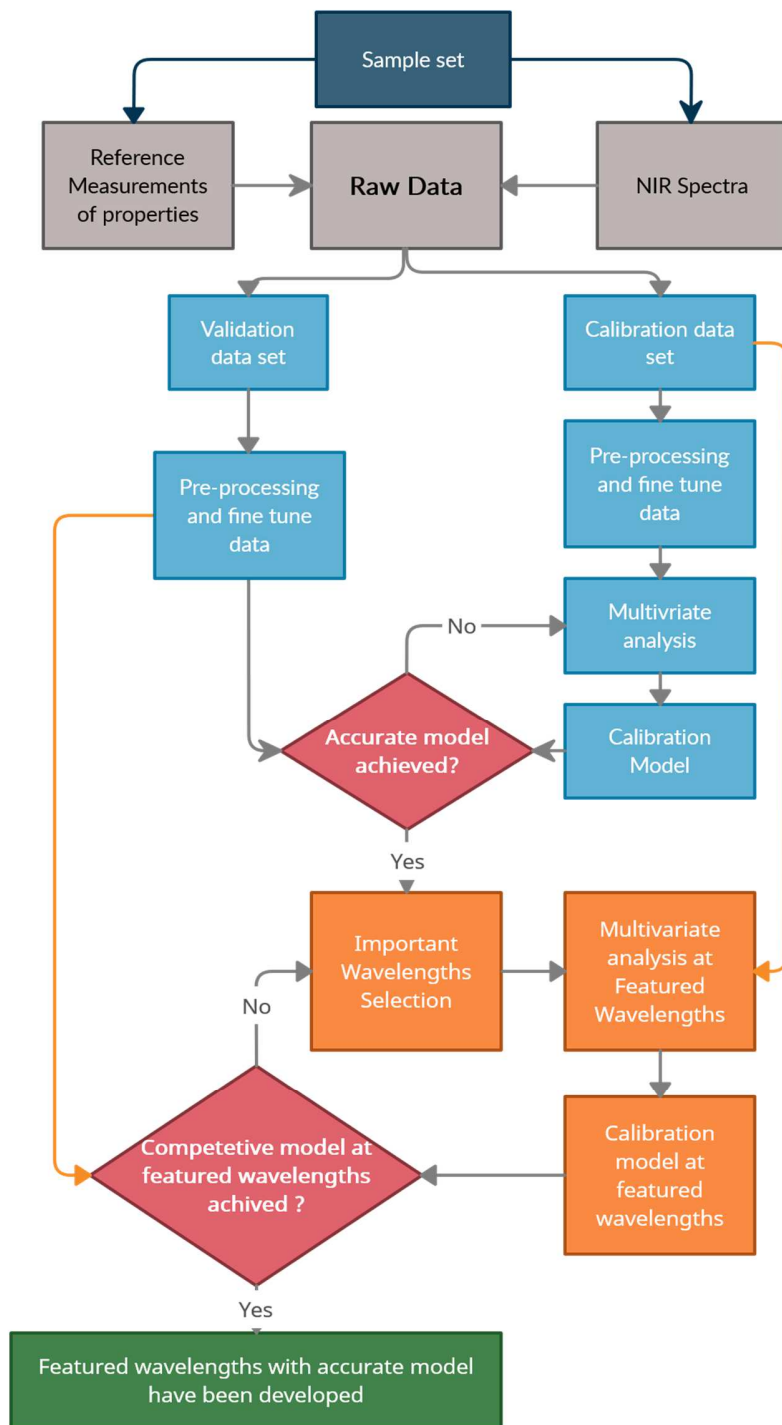
Model	Pre-treatment	LVs	Calibration		Cross-Validation		Prediction		
			$R_c^2$	RMSEC (N)	$R_{cv}^2$	RMSECV (N)	$R_p^2$	RMSEP (N)	RPD
PLS	Raw	11	0.94	11.74	0.91	14.85	0.91	15.22	3.26
	SNV	10	0.97	9.14	0.92	13.82	0.90	15.53	3.20
	MSC	8	0.81	21.87	0.68	28.32	0.79	22.52	2.20
	1st derivative	11	0.94	11.97	0.90	15.70	0.90	15.98	3.11
	2nd derivative	10	0.95	10.85	0.90	15.70	0.90	15.49	3.21
PLS-ANN	Raw	N.A	0.95	11.54	0.95	10.47	0.95	11.49	4.32
	SNV	N.A	0.97	8.56	0.95	10.18	0.93	13.14	3.78
	MSC	N.A	0.96	10.57	0.95	11.10	0.92	13.68	3.63
	1st derivative	N.A	0.93	12.82	0.93	12.25	0.93	12.76	3.89

	2nd derivative	N.A	0.95	11.99	0.95	11.36	0.94	12.65	3.93
<b>PLS</b>	Selected wavelength	7	0.93	13.66	0.90	14.92	0.91	14.78	3.36
<b>ANN</b>	Selected wavelength	7	0.94	12.60	0.94	12.54	0.94	12.55	4.01
<b>MLR</b>	Selected wavelength	7	0.93	13.66	0.90	14.92	0.91	14.85	3.34

601

602

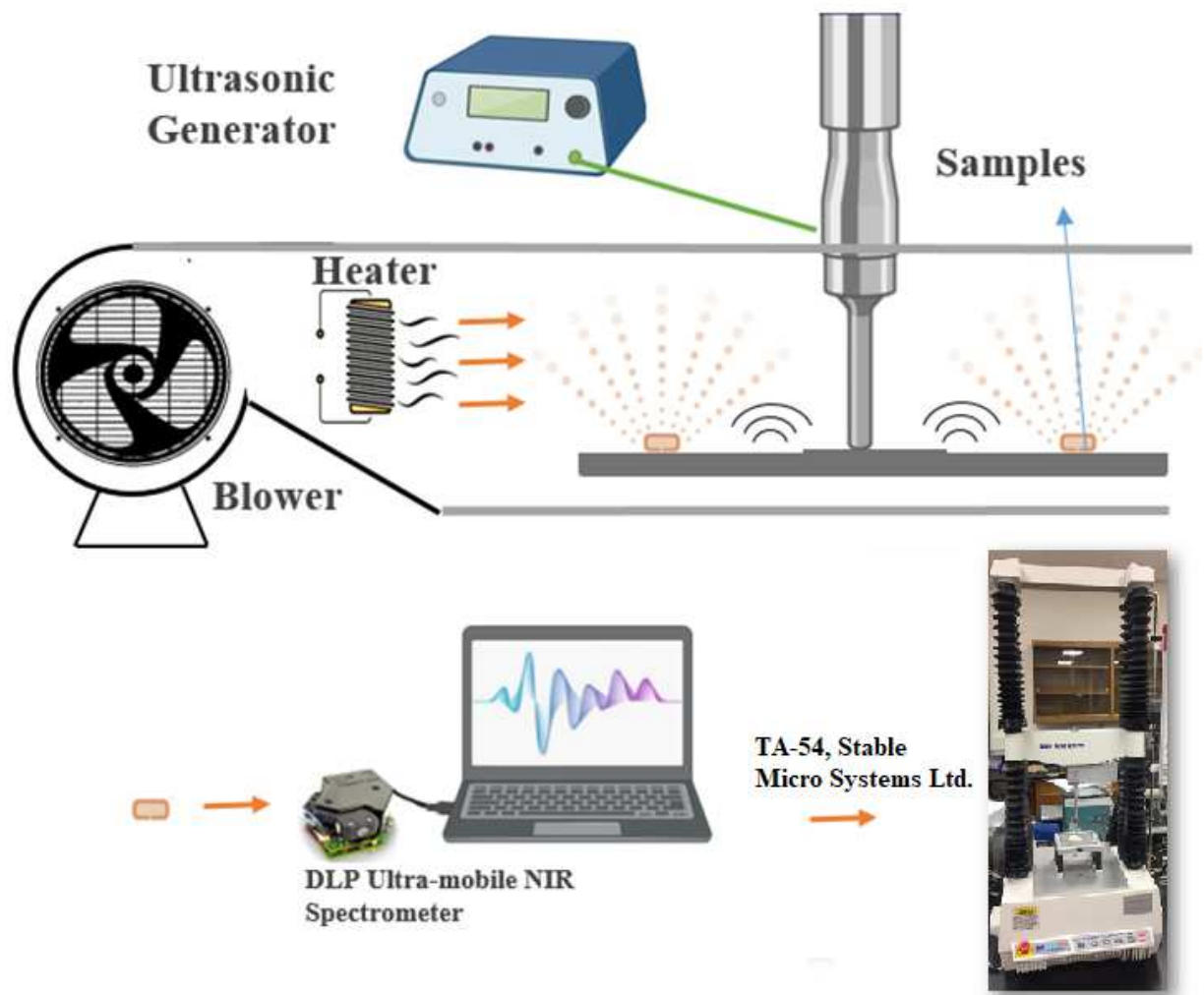
603



604

605

**Figure 1.**

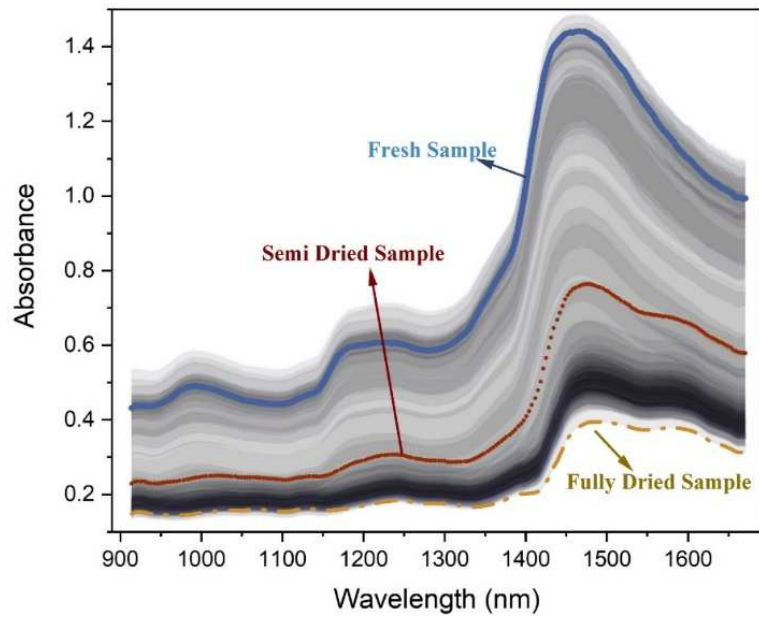


606

607

608

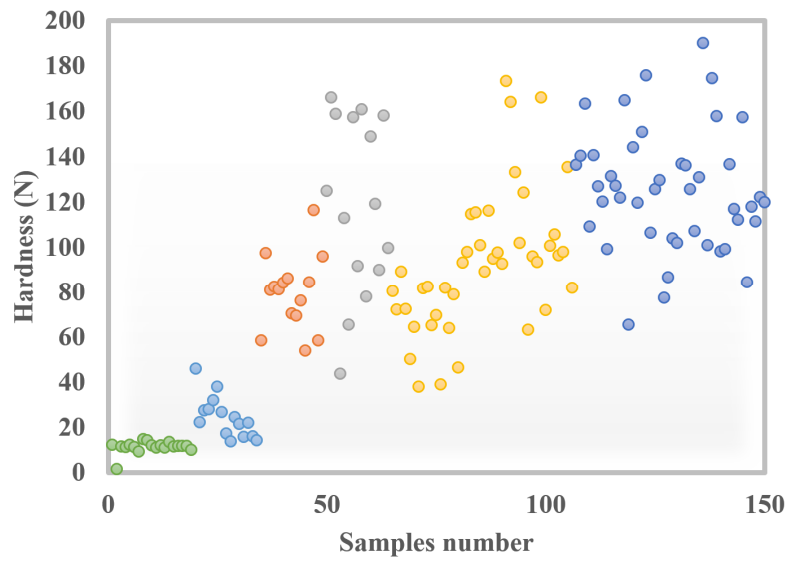
Figure 2



609

610

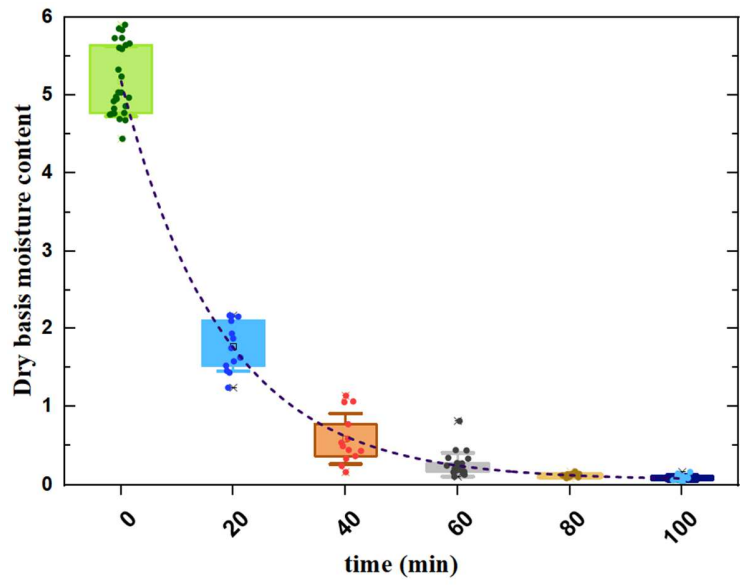
(a)



611

612

(b)



613

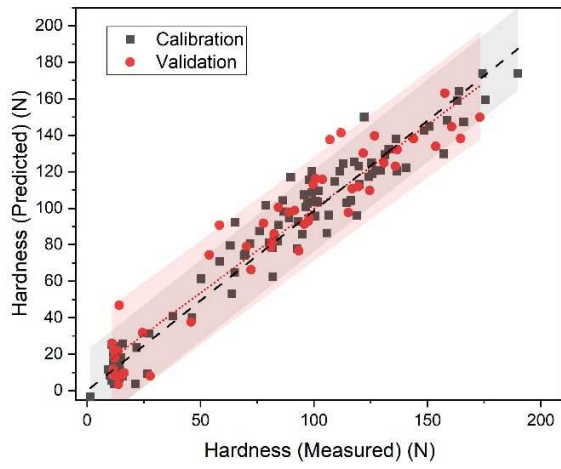
614

(c)

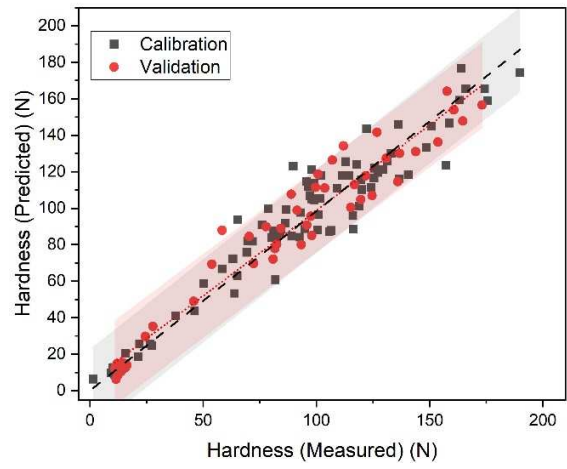
615

Figure 3

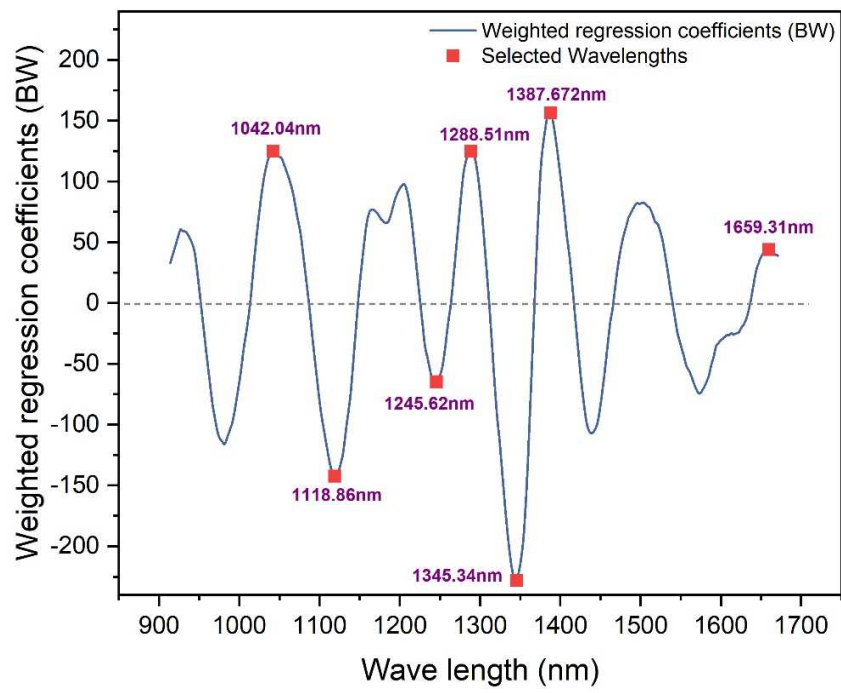
616



(a)



(b)



(c)

617

Figure 4

618

619