

2022-02-01

Learning fruit class from short wave near infrared spectral features, an AI approach towards determining fruit type

Ayesha Zeb

National University of Sciences and Technology, Islamabad

Waqar Shahid Qureshi

Technological University Dublin, waqar.shahid@gmail.com

Abdul Ghafoor

National University of Sciences and Technology, Islamabad

See next page for additional authors

Follow this and additional works at: <https://arrow.tudublin.ie/scschcomcon>

 Part of the [Computer Sciences Commons](#)

Recommended Citation

Zeb, A., Qureshi, W. S., Ghafoor, A., & O'Sullivan, D. (2022). Learning fruit class from short wave near infrared spectral features, an AI approach towards determining fruit type. Technological University Dublin. DOI: 10.21427/5G17-H620

This Conference Paper is brought to you for free and open access by the School of Computer Science at ARROW@TU Dublin. It has been accepted for inclusion in Conference papers by an authorized administrator of ARROW@TU Dublin. For more information, please contact arrow.admin@tudublin.ie, aisling.coyne@tudublin.ie, vera.kilshaw@tudublin.ie.



This work is licensed under a [Creative Commons Attribution 4.0 International License](#).

Authors

Ayesha Zeb, Waqar Shahid Qureshi, Abdul Ghafoor, and Dympna O'Sullivan

Learning fruit class from short wave near infrared spectral features, an AI approach towards determining fruit type

Ayesha Zeb¹, Waqar Shahid Qureshi², Abdul Ghafoor³, Dympna O' Sullivan⁴

¹Department of Electrical Engineering, MCS, National University of Sciences and Technology, H-12, Islamabad, Pakistan

²Department of Computer Science, Technological University Dublin, Ireland

³Department of Electrical Engineering, MCS, National University of Sciences and Technology, H-12, Islamabad, Pakistan

⁴Department of Computer Science, Technological University Dublin, Ireland

¹azeb.phdee@students.mcs.edu.pk, ²waqar.qureshi@tudublin.ie, ³abdulghafoor-mcs@nust.edu.pk, ⁴Dympna.O'Sullivan@tudublin.ie

Abstract— This paper analyzes the potential of using short-wave NIRS (near-infrared spectroscopy) for fruit classification problems. The research focuses on O-H and C-H overtone features of fruit and its correlation with NIRS and therefore opens a new dimension of fruit classification problems using NIRS. Eleven fruits, which include apple, cherry, hass, kiwi, grapes, mango, melon, orange, loquat, plum, and apricot, were used in this study to cover physical characteristics such as peel thickness, pulp, seed thickness, and size. NIR spectral data is collected using the industry-standard F-750 fruit quality meter (wavelength range 300-1100nm) for all fruit mentioned above. Different shallow machine learning architectures were trained to classify fruits using spectral feature vectors. At first, using 83 features vectors within the range of 725-975nm (3nm-resolution) and then using only four features of wavelength 770nm, 840nm, 910nm, and 960nm (corresponding to O-H and C-H overtone features). For the 83 spectral features range as an input, the QDA classifier achieved a cross-validation accuracy of 100% and a test data accuracy of 93.02%. For the four features vector as an input, the QDA classifier achieved a cross-validation accuracy of 97.1% and test data accuracy of 90.38%. The results demonstrate that fruit classification is mainly a function of absorptivity of short wave NIR radiation primarily with respect to O-H and C-H overtones features. An LED-based device mainly having 770nm, 840nm, 910nm, and 960nm range LEDs can be used in applications where automation in fruit classification is required.

Keywords— NIR Spectroscopy, Fruit classification, Machine learning

I. INTRODUCTION

Over the past decades, NIR spectroscopy has gained considerable attention for non-destructive fruit quality assessment due to its ease, detection speed, and precision. The Vis and NIR region of the light spectrum has the range 400-750 nm and 750-2500 nm, respectively. The short wave NIR (SWNIR) or Herschel region lies between 750-1100 nm and the extended NIR region lies between 1100-2500 nm. NIR spectroscopy involves the measurement of the absorbance of light linked with the vibration of molecular bonds. For intact fruits, this usually entails absorption linked with the stretching of O-H and C-H bonds [1-2], related principally with water and

storage reserves (the major macro constituents). The infrared region (>2500 nm) provides the fundamental absorption bands associated with these features with more fine and higher absorption peaks than linked with the overtones observed in the NIR region. NIR region gives lower absorptivity of overtones features than those of infrared (IR) region. Due to this characteristic, the effective pathlengths of NIR radiation through fruit are in the order of millimeters to centimeters rather than micrometers as in the case of IR radiation. Similarly, with the short-wave NIR (SWNIR) region longer effective pathlengths are achieved with higher overtone features as compared to the overtones in the NIR region. This is the reason that SWNIR radiation can be used to estimate fruit parameters of intact fruits [3]. In recent years, researchers have used NIR spectroscopy with machine learning regression algorithms to develop maturity indexes prediction models such as DM, Brix, color, chlorophyll, starch, and TA (only in high acid fruit like lemon and mandarin) of various fruits including apple, pear, nectarine, mango, banana, melon, mandarin, strawberry, apricot, kiwifruit, persimmon, grape, loquat and pineapple [4].

Absorption at 770 nm and 960 nm is associated with a third and second overtone of O-H stretching, respectively and 840nm is associated with the O-H combination feature. While absorption at 910 nm and 1100 nm is associated with a third and second overtone of C-H stretching, respectively [2]. These overtone features have broad peaks with a full width half maximum (FWHM) of 20nm (the reported peaks are centers of the respective bands). However, the overall peak positions shift with temperature and solute concentration because the amount of H-bonding can change which influences the vibration of O-H bonds. In practice, it is hard to interpret short-wave NIR (being second and third overtones, so weak and broad) compared to extended NIR and IR regions. However, the features related to water can be interpreted, which is the main NIR active molecule in fruit. Around 80-90% of fleshy fruit is composed of water. This is the reason that any other parameter is measured with reference to the large absorption features of water. An increase in any other macroconstituent, e.g., solids soluble content (SSC) and dry matter (DM), causes a decrease in water content resulting in a negative correlation with water. The penetration depth of NIR in fruits is comparatively greater

in the 700–900 nm range [5] hence more information on internal quality attributes can be fetched by using wavelength absorption data of this region. Understanding the raw absorption spectra is challenging because all absorption characteristics are wide and overlapping. The use of NIRS over IR spectroscopy is successful due to chemometrics, which enabled useful information to be fetched out of the spectra.

Point spectroscopy delivers a sum of light absorption and scattering. The scattering characteristics of tissue affect the effective penetration depth of light into that tissue [6]. However, the quantity of light scattering differs between fruit types [6-7]. This fact can be used to classify different fruits. In this paper, the potential of SWNIR spectroscopy, primarily with respect to O-H and C-H overtone features, is analyzed for fruit type classification. Different shallow machine learning architectures were trained to classify fruits using spectral feature vectors obtained by the industry-standard F-750 fruit quality meter [8]. Two types of feature sets are used to compare the classification potential of both sets i.e. (1) 83 features within 725-975nm range and (2) only 4 features at wavelengths 770nm, 840nm, 910nm, and 960nm corresponding to O-H and C-H overtone features.

II. MATERIALS AND METHODS

A. Collection of Vis/NIR Spectra

Vis-NIR spectral data (wavelength range 400-1150nm) is collected using the industry standard F-750 fruit quality meter (Felix Instruments, Camas, WA, USA). The dataset for apple, grapes, mango, melon, orange, loquat, plum and apricot are collected for local cultivars of Pakistan while online available datasets of F-750 [7] for cherry, hass and kiwi have been used. Samples of apple, grapes, mango, melon, orange, loquat, plum and apricot were scanned from equator position with F-750. Table 1 shows the details of fruit datasets used.

TABLE I. DETAILS OF DATASETS COLLECTED FOR INVESTIGATED FRUITS

Fruit	Variety	Season	Number of Samples	Training and testing dataset samples
Apple	Golden Delicious, Red Delicious-Pak, Red Delicious-Turk	2019	35	Training
		2020	15	Testing
Grapes	Sundar Khani	2020	100	70 Training, 30 Testing
Mango	SB Chaunsa	2019	150	Training
		2020	50	Testing
Melon	Honey melons	2020	150	Training
		2021	50	Testing
Orange	Red-Blood	2020	74	54 Training, 20 Testing
	Mosambi		64	44 Training, 20 Testing
	Succari		54	44 Training, 10 Testing
Loquat	Tanaka	2020	150	Training
		2021	50	Testing
Plum	Fazle Manani	2020	200	150 Training, 50 Testing
Apricot	Badami	2020	200	150 Training,

				50 Testing
Cherry	Lapins	2016	200	150 Training, 50 Testing
Hass	unknown	2017	200	150 Training, 50 Testing
Kiwi	General Kiwis	2016	200	150 Training, 50 Testing

B. Chemometric Analysis

The basic idea behind finding fruit class from short wave NIR spectral features is to construct a linear predictor function that generates a score from a set of weights that are linearly combined with the set of spectral features of given observations as:

$$\text{Scores}(X_i, k) = \alpha_k \cdot X_i \quad (1)$$

where “.” is the dot operator, X_i is the spectral features vector corresponding to the i th observation, α_k is the coefficients vector corresponding to class k and $\text{Scores}(X_i, k)$ is the score associated with allocating observation i to class k . The predicted class is the one that has highest score. Machine learning classification algorithms determine the optimal scores and interpret them to assign class to observation i .

For fruit type classification, MATLAB R 2021a software was used. Classification was performed using MATLAB classification learner module with PCA enabled (first 7 principal components were used). Principle component analysis (PCA) has been widely used with spectroscopic data [4,9] to emphasize variation and bring out strong patterns in the data set. PCA was applied on spectral data and then several supervised and unsupervised machine learning classifiers are implemented and compared including tree, ensemble, K nearest neighbor (KNN), linear discriminant analysis (LDA), quadratic discriminant analysis (QDA), support vector machine (SVM), naïve bayes and artificial neural network (ANN).

III. RESULTS

A. Overview of spectra

The average absorbance spectra of all fruits within SWNIR range (Fig. 1) is dominated by small peaks around 770 nm, 840 nm and a strong peak around 960 nm associated with water absorption band [10].

B. Principal Component Analysis

PCA scores plot for the collected data set is shown in Fig. 2 (a and b). Observations that are similar together form a cluster in scores plot of PCA. Fig. 2 depicts that all the eleven investigated fruits form well defined clusters in both the cases i.e. (a) with 83 features from 725-975nm wavelength range (3nm resolution) and (b) with only 4 features at wavelengths 770nm, 840nm, 910nm and 960nm. Hence, the plots depict good potential for classification algorithms to distinguish fruits using SWNIR region or simply only the OH and CH overtones features.

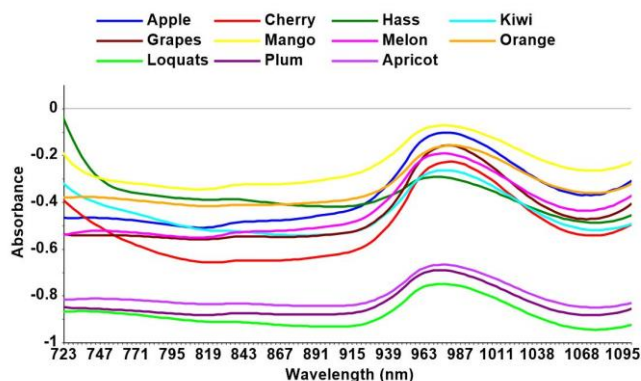


Fig. 1. Average raw absorbance spectra of collected datasets for the investigated fruits

C. Classification results

Table 2 shows the classification results in terms of cross validation (CV) accuracy and test data accuracy for 83 features (within range 725-975nm with 3nm resolution) as input to the classifier. Training dataset has 1507 observations for the eleven fruits while the test data has 531 observations (some of them are of independent season). QDA classifier outperformed other investigated classifiers with 100% CV accuracy and 93.02% test accuracy.

The classification results (Table 3) in terms of CV accuracy and test data accuracy for only 4 features at wavelengths 770nm, 840nm, 910nm and 960nm (OH and CH overtone features) as input to the classifier. QDA classifier outperformed other investigated classifiers with 97.1% CV accuracy and 90.38% test accuracy.

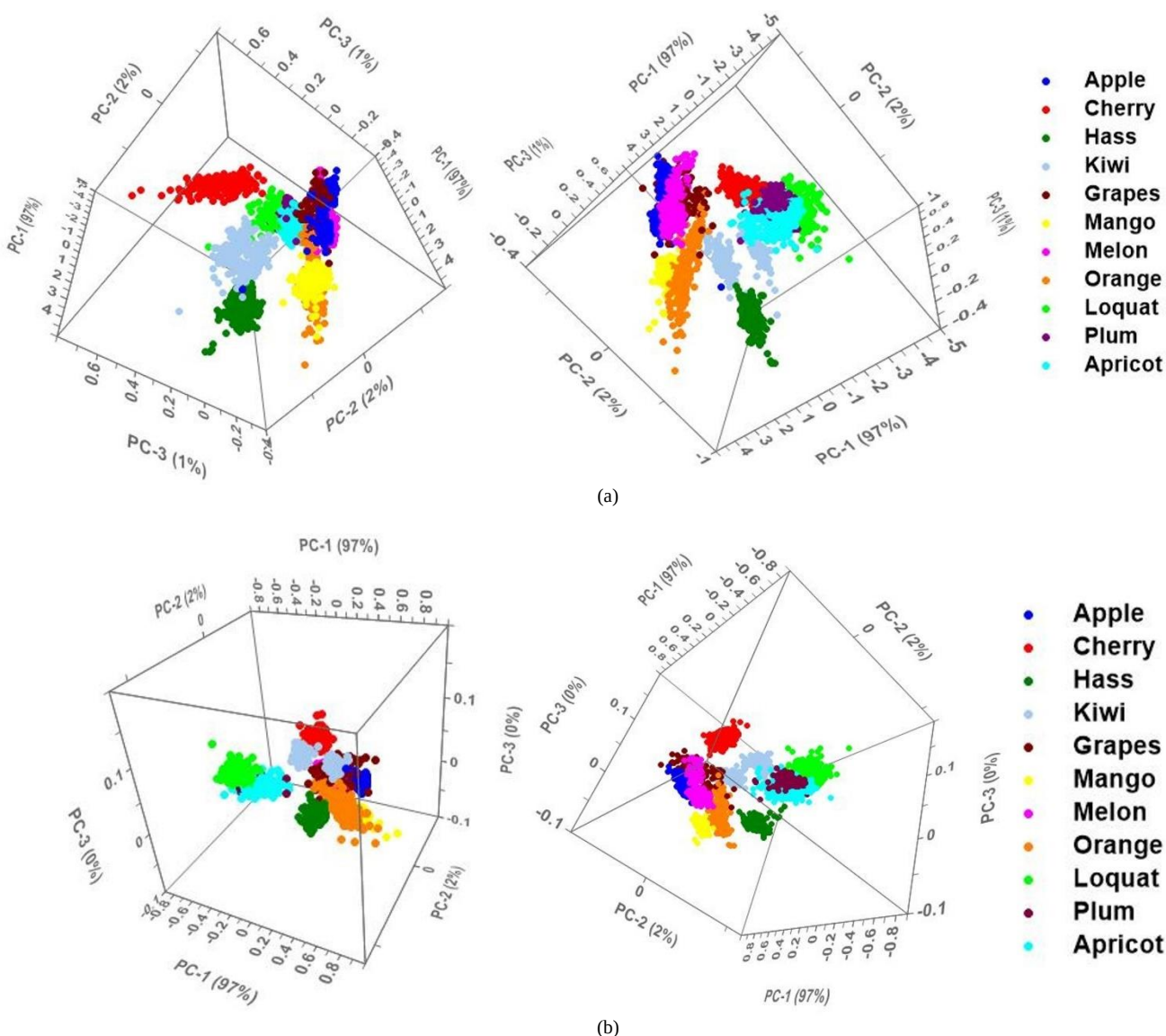


Fig. 2. 3D scores plot for first three principal components of spectra of the investigated fruits (a) 725-975 nm wavelength range, and (b) only 4 features at wavelengths 770nm, 840nm, 910nm and 960nm

TABLE II. CLASSIFICATION ACCURACIES USING 83 FEATURES WITHIN WAVELENGTH RANGE 725-975NM

Classifier	CV accuracy	Test accuracy
Tree	97.9	88.68
LDA	99.3	90.94
QDA	100	93.02
Naïve Bayes	98.9	91.89
SVM	99.8	92.26
KNN	99.7	93.21
Ensemble	99.3	90.57
ANN	99.4	91.89

TABLE III. CLASSIFICATION ACCURACIES USING 4 FEATURES AT WAVELENGTHS 770NM, 840NM, 910NM AND 960NM

Classifier	CV accuracy	Test accuracy
Tree	94.8	87.55
LDA	95.7	87.92
QDA	97.1	90.38
Naïve Bayes	94	81.32
SVM	98.3	90
KNN	97.7	89.81
Ensemble	96.6	89.43
ANN	97.5	89.81

IV. OBSERVATIONS AND CONCLUSION

The results obtained from PCA, and classification suggest that fruit classification is mainly a function of OH and CH overtone features. PCA results depict that both the feature sets, i.e., 83 features from the 725-975nm wavelength range and only 4 features representing the OH and CH overtones peaks within the 725-975nm range, form well-defined clusters in the PCA space. The PCA scores plot depicts a good intra-cluster correlation for all the eleven clusters in both cases (Fig. 2). The three-stone fruits i.e. loquat, plum, and apricot have less inter-cluster distance between them (see Fig. 2) and also show similar absorbance behaviour in Fig. 1. This is because all the three investigated stone fruits have similar physical structures i.e. thin peel, a thin pulp (and hence lesser NIR radiation absorbance by pulp) and a big stone/stones in the center of the fruit. Apple, mango, grapes, melon, and orange also have less inter-cluster distances amongst each other (Fig. 2) and also it shows a similar trend in the absorbance spectra (see Fig. 1) i.e. top five peaks at 960nm wavelength related to water absorption band. All these five fruits have thick pulp and small seeds (the investigated grapes variety is 'Sundar khani' which has no seeds). Hass, cherry, and kiwi show good inter-cluster distances from all other clusters. Kiwi's spectra form two clusters with a less inter-cluster distance between them, the reason is unknown as kiwi's dataset was taken from an online source with no information available about the cultivars that were scanned.

Results obtained from PCA and classification in case 2 (90.38% test accuracy for 11 classes) i.e. only 4 features at wavelengths 770nm, 840nm, 910nm and 960nm (OH and CH overtone features) provide good motivation for application of low cost LED based devices in fruit classification problems. Spectrometer based portable devices are also available e.g. Felix instruments F-750 [8], Consumer Physics' SciO [11], Sunforest H-100C [12] and Atago Hikari [13], however these

devices are expensive compared to LED based portable devices for assessment of specific pigment in fruit (at two to five wavelengths) i.e. DA meter [14], kiwi meter [14], cherry meter [14], Multiplex 330 [15] and FIORAMA [16].

ACKNOWLEDGMENT

This research was supported by Pakistan Agriculture Research Council, Agriculture Linkage Program (AE-007).

REFERENCES

- [1] S. Kawano, "Non-destructive NIR quality evaluation of fruits and vegetables in Japan," NIR news, vol. 5(6), pp. 10-12, 1994.
- [2] M. Golic, K. Walsh, and P. Lawson, "Short-wavelength near-infrared spectra of sucrose, glucose, and fructose with respect to sugar concentration and temperature." Applied spectroscopy, vol. 57(2), pp.139-145, 2003.
- [3] K. B. Walsh, J. Blasco, M. Zude-Sasse, and X. Sun, "Visible-NIR 'point' spectroscopy in postharvest fruit and vegetable assessment: the science behind three decades of commercial use." Postharvest Biology and Technology, vol.168, pp.111246, 2020.
- [4] S. S. A. Shah, et al., "Towards fruit maturity estimation using NIR spectroscopy." Infrared Physics & Technology, vol. 111, pp.103479, 2020.
- [5] J. Lammertyn, A. Peirs, J. De Baerdemaeker, and B. Nicolai, "Light penetration properties of NIR radiation in fruit with respect to non-destructive quality assessment." Postharvest Biol. Technol. Vol. 18 (2), pp. 121-132, 2000.
- [6] R. Cubeddu, et al., "Time-resolved reflectance spectroscopy applied to the nondestructive monitoring of the internal optical properties in apples." Applied spectroscopy, vol. 55(10), pp.1368-1374, 2001.
- [7] R. Lu, R. Van Beers, W. Saeys, C. Li, and H. Cen, "Measurement of optical properties of fruits and vegetables: A review." Postharvest Biology and Technology, vol.159, pp.111003, 2020.
- [8] Felix Instruments, USA, Retrieved Nov 7, 2021, from website: <https://felixinstruments.com/food-science-instruments/portable-nir-analyzers/f-750-producequality-meter/>
- [9] W. Guo, J. Gu, D. Liu, and L. Shang, "Peach variety identification using near-infrared diffuse reflectance spectroscopy." Computers and Electronics in Agriculture, vol. 123, 297-303, 2016.
- [10] G. Giovanelli, N. Sinelli, R. Beghi, R. Guidetti, and E. Casiraghi, "NIR spectroscopy for the optimization of postharvest apple management." Postharvest Biology and Technology, vol. 87, pp. 13-20, 2014.
- [11] SciO, Israel, Retrieved on Nov 7, 2021, from website: <https://www.consumerphysics.com/>
- [12] SunForest, Korea. Retrieved on Nov 7, 2021, from website: http://www.sunforest.kr/category_main.php?sm_idx=168
- [13] Atago, Hikari. Japan, Retrieved Nov 7, 2021, from website: http://atago.net/product/?l=ue&f=products_pal_hikari.html
- [14] T.R. Turoni Srl. Forli, Italy, Retrieved Nov 7 2021, from website: <https://www.trturoni.com/en/fruit-veg-ripeness-quality-control/da-meter>
- [15] Force-A, Paris, France, Retrieved Nov 7, 2021 from website: <https://www.force-a.com/fr/produits/multiplex>
- [16] CP, Potsdam-Golm, Germany, Retrieved Nov 7, 2021 from website: <http://cp-info.de/Our-Products/>