RESEARCH

Convolutional Neural Networks for Estimating the Ripening State of Fuji Apples Using Visible and Near-Infrared Spectroscopy

Brahim Benmouna[1] · Ginés García-Mateos[1] · Sajad Sabzi[2] · Ruben Fernandez-Beltran[1] · Dolores Parras-Burgos[3] · José Miguel Molina-Martínez[4]

Abstract

The quality of fresh apple fruits is a major concern for consumers and manufacturers. Classification of these fruits according to their ripening stage is one of the most decisive factors in determining their quality. In this regard, the aim of this work is to develop a new method for non-destructive classification of the ripening state of Fuji apples using hyperspectral information in the visible and near-infrared (Vis/NIR) regions. Spectra of 172 apple samples in the range from 450 to 1000 nm were studied, which were selected from four different ripening stages. A convolutional neural network (CNN) model was proposed to perform the classification of the samples. The proposed method was compared with three alternative methods based on artificial neural networks (ANN), support vector machines (SVM), and k-nearest neighbors (KNN). The results revealed that the CNN method outperformed the alternative methods, achieving a correct classification rate (CCR) of 96.5%, compared with an average of 89.5%, 95.93%, and 91.68% for ANN, SVM, and KNN, respectively. These results will help in the development of a new device for fast and accurate estimation of the quality of apples.

Keywords Fuji apples · Spectroscopy · Deep learning · Neural networks

✉ Ginés García-Mateos
ginesgm@um.es

Brahim Benmouna
brahim.benmouna@um.es

Sajad Sabzi
s.sabzi@sharif.edu

Ruben Fernandez-Beltran
rufernan@um.es

Dolores Parras-Burgos
dolores.parras@upct.es

José Miguel Molina-Martínez
josem.molina@upct.es

1   Computer Science and Systems Department, University of Murcia, Murcia, Spain

2   Computer Engineering Department, Sharif University of Technology, Tehran, Iran

3   Department of Structures, Construction and Graphic Expression, Polytechnic University of Cartagena, Murcia, Spain

4   Agricultural and Marine Engineering Research Group, Polytechnic University of Cartagena, Cartagena, Murcia, Spain

Introduction

Harvesting at an appropriate ripening stage is the key factor in determining the compositional quality and storage life of fruits. Early harvesting reduces the taste and the quality of fruits, whereas harvesting fruits too late can reduce shelf life while causing bad appearance, off-flavors and odors. Various external and internal quality attributes of fruits have been measured in an attempt to provide an adequate estimation of their stage of ripeness including shape, size, texture, firmness, external color, internal color, concentration of starch, chlorophyll, acids, soluble solids content, oils, sugars, and internal ethylene concentration (Li et al., 2018).

Both destructive and non-destructive methods have been applied to measure the quality properties of fruits. However, non-destructive methods have demonstrated a great interest over traditional methods as they provide several advantages such as real-time assessment, multiple simultaneous measurements, and real-time decision-making. Many of these techniques have been proposed to determine the optimum harvest date of fruits, including fluorescence imaging (Cerovic et al., 2009), colorimetry (Baltazar et al., 2008), computed tomography scan (Kotwaliwale et al., 2012),

machine vision (Sabzi et al., 2019), visible and near-infrared (Vis/NIR) spectroscopy (Yang et al., 2011), multi-spectral imaging (Khodabakhshian et al., 2017), and hyperspectral imaging (Su et al., 2021). These non-destructive techniques, especially those based on optical and imaging properties, have proven to be a very useful and powerful tool for estimating the ripening stages of fruits.

In the context of non-destructive methods, NIR spectroscopy has become one of the most popular trends because it covers the spectral range related to the vibration of molecular bonds, that is, O–H and C–H bonds which mainly represent fruit macro-constituents. In this way, the low absorptivity of overtones features allows NIR to provide deeper penetration levels in fruits than other vibrational methods. Specifically, the short-wave NIR radiation, i.e., 750–1100 nm, is able to achieve the longest effective path lengths, which eventually makes this spectral range one of the most suitable for estimating fruit parameters from intact samples. The constant development of acquisition instruments also becomes an important factor to support this technology. For instance, the reduction in size of NIR spectrometers may certainly facilitate on-site measurements of fruits (Beć et al., 2021). Besides, the development of surface scanning systems, such as the automated rotation measurement system (Schmutzler & Huck, 2014), can also provide more robust acquisitions since individual fruit biological in-homogeneities can be averaged to achieve more accurate estimations. With these considerations in mind, many works in the literature exemplify the success of Vis/NIR spectroscopy.

For example, Wei et al. (2014) used hyperspectral imaging (HSI), covering the spectral range from 400 to 1000 nm, for estimating the ripeness stages (unripe, mid-ripe, ripe, and over-ripe) of persimmon fruits. Specifically, 192 HSI samples were classified using three types of classification models, linear discriminant analysis (LDA), soft independence modeling of class analogy (SIMCA), and least squares support vector machines (LS-SVM). The results showed that the best model was LDA, obtaining a classification accuracy of 95.3%. In this case, the input data was made from three predefined wavelengths (518, 711, and 980 nm). Giovenzana et al. (2014) used a portable commercial Vis/NIR spectrophotometer, covering the spectral range of 400–1000 nm, for the estimation of the total soluble solids and total polyphenols of grapes during ripening. To achieve this goal, partial least square regression coefficient analysis (PLS-RCA) was used to select three most effective wavelengths to discriminate grapes ready to be harvested. Then, principal component analysis (PCA) and multiple linear regression (MLR) were applied to verify the effectiveness of the selected wavelengths (670, 730, and 780 nm). Both qualitative and quantitative results confirmed a good fruit discrimination during ripening. More recently, Pourdarbani et al. (2020) developed a new non-destructive classification algorithm based on color properties and spectral data (from 450 to 1000 nm) to estimate the ripening stages of Fuji apples (unripe, half-ripe, ripe, and over-ripe). Specifically, this method involves the application of five different classifiers combined with a majority voting (MV) rule. They include k-nearest neighbors (KNN), support vector machine (SVM) and three hybrid classifiers that combine artificial neural networks (ANNs) together with metaheuristic algorithms. The input data consisted of color features extracted from the second channel of L*a*b* color space and multi-spectral information covering the ranges 465–485 nm, 675–700 nm, and 870–890 nm. The results showed that the proposed method was able to achieve the best performance when using color features with the 465–485-nm spectral range, obtaining a final accuracy of 99.37%. Gao et al. (2020) conducted a study to determine the ripening stage of strawberry fruits using HSI covering from 370 to 1015 nm under both field and laboratory conditions. Using the sequential feature selection (SFS) algorithm, two wavelengths were selected for field (530 and 604 nm) and other two for laboratory conditions (528 and 715 nm). The performance of the selected wavelengths was validated using a SVM classifier, obtaining a classification accuracy of 98.6%.

On the other hand, deep learning (DL) is a subset of machine learning methods based on artificial neural networks consisting of multiple processing layers to automatically learn complex representations from data, without introducing hand-coded rules or human domain knowledge. Among DL models, convolutional neural networks (CNNs) are currently one of the most popular models since they do not require manual feature extraction. CNN is a class of advanced deep neural networks, inspired by the visual cortex of animal brain. CNN was discovered by Hubel and Wiesel (Hubel & Wiesel, 1962) and introduced in 1990. By the rapid development of these techniques, many models have been proposed in the literature. Some examples of such models are AlexNet, VGG Net, ZF Net, GoogLeNet, and fully convolutional networks (FCN).

CNN techniques have been applied in many fields and have particularly shown their crucial roles in the agricultural area. In this regard, Zeng et al. (2020) employed LeNet CNN architecture for the detection and classification of bruises of pears. The obtained results showed that the proposed architecture achieved great performance, producing an accuracy of 99.3%. Mohtar et al. (2019) employed V3 Inception CNN model for mangosteen fresh fruit ripeness classification. The obtained results showed that the CNN model achieved an accuracy of more than 90%. Pourdarbani et al. (2021) developed a new 1D-CNN architecture to estimate the nitrogen content of cucumber leaves. The obtained results showed that the proposed 1D-CNN model was able to estimate very accurately the nitrogen content.

In the literature, it is also possible to find other emerging trends that deserve to be mentioned for the sake of completeness. For instance, it is the case of theoretical NIR spectroscopy based on quantum chemistry (Beć & Huck, 2019). In more details, quantum methods rely on anharmonic models that describe in detail molecular vibrations for an accurate simulation of NIR spectra. Nonetheless, the high computational complexity of this type of models still constrains their applicability to practical scenarios, in contrast to the aforementioned technologies.

The aim of the present research is to develop a new nondestructive method based on Vis/NIR spectroscopy using a CNN classifier to estimate the four different ripeness states of Fuji apples (unripe, half-ripe, ripe, and over-ripe). To evaluate the effectiveness of the proposed method, the obtained results were compared with three alternative classifiers based on ANN, SVM, and KNN models.

Materials and Methods

The research methodology applied in this study for estimating the ripening states of Fuji apples consisted of four main steps: (i) apple samples collection; (ii) extraction of the spectral data for each apple sample; (iii) spectral data normalization; (vi) classification of the apples according to their ripening state. The development of these steps is described in the following sections.

Data Collection

A total of 172 of Fuji apple samples in four different states of ripening (43 samples per each harvested state) were taken from an orchard located in Kermanshah, Iran (34.3277° N, 47.0778° E). The samples were taken at intervals of 7 days around the ripening date, as indicated by expert farmers. More specifically, the first state of harvest (unripe) was done 14 days before ripening, the second state (half-ripe) was done 7 days before ripening, the third stage (ripe) on the ripening date given by the experts, and the fourth state (over-ripe) was done 7 days after ripening. The samples were immediately transferred to the laboratory in refrigerated transport for spectral analysis. For all the samples, the time elapsed between collection and analysis was always less than 6 h. Figure 1 shows a sample of apples in the trees at the four ripening stages defined.

Spectral Data Extraction

To extract the spectral data from the apple samples, a hardware system was configured. The components of this system were (1) a spectrometer; (2) a source of light; (3) an optical fiber; and (4) a laptop PC with AMD Quad Core A6, at 2.00 GHz, 4 GB of RAM, running Windows. A spectrometer of type EPP200NIR (StellarNet Inc., Tampa, Florida, USA) was used to detect the spectral reflectance in the range from 450 to 1000 nm, which includes visible light (from 450 to 750 nm) and a region of NIR (from 750 to 1000 nm). A tungsten halogen lamp with 20 W (StellarNet Inc., Tampa, Florida, USA) was used as the light source. An indium-gallium-arsenide (*InGaAs*) detector embedded was used to offer the highest sensitivity in the near-infrared regions. Two optical fibers were utilized to transmit the light from the lamp to the apple and also from the apples to the spectrometer. In this way, 1943 wavelengths were captured for each spectral band for each sample, with an average width of 0.28 nm per band.

Spectral Data Normalization

In this step, a preprocessing stage based on the mean and variance normalization method was used to reduce certain noise appearing in the spectral data, due to the effect of ambient light, the spectrometer used, the surface of the apple samples, and the type of lamp. This method works as follows:
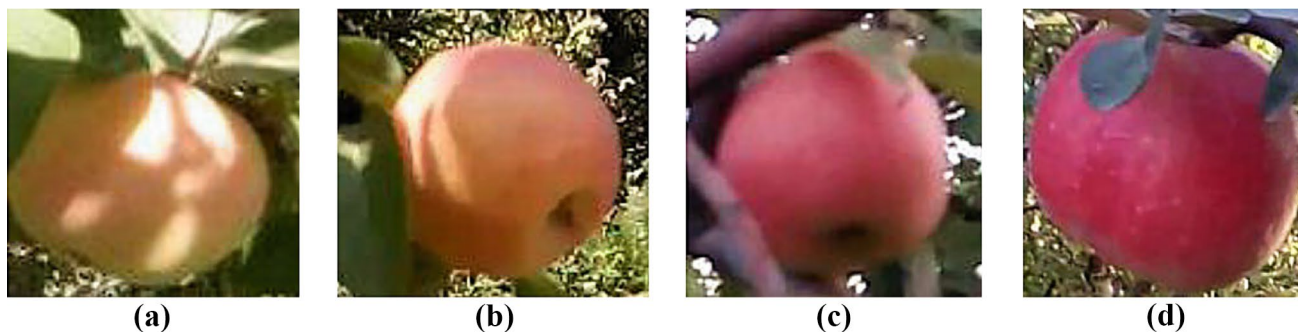

Fig. 1 Example images of some of the Fuji apple samples at the different ripening stages. **a** Unripe. **b** Half-ripe. **c** Ripe. **d** Over-ripe

- Compute the mean and standard deviation of all the bands of the spectrum (1943 wavelengths) of each sample.
- Subtract the mean from each spectral band of the sample and divide by the standard deviation (SD).

This normalization method can be expressed using the following equation:

$$S_{norm}(w) = \frac{S(w) - \overline{S}}{SD} \tag{1}$$

where S represents the spectral information of the input sample, $\overline{S}$ is the average of the values of S, and w corresponds to the wavelength considered (from 450 to 1000 nm). The formula to calculate SD is as follows:

$$SD = \sqrt{\frac{1}{K-1} \sum_{w=1}^{K} \left(S(w) - \overline{S}\right)^2} \tag{2}$$

where $K = 1943$ is the number of wavelengths. This normalization process reduces the range of values produced for the different samples. It was manually programmed in Python using JupyterLab interface version 3.7.0. The resulting normalized spectra of all the apple samples are shown in Fig. 2.

Apple Ripeness Classification Using CNN

This paper presents a novel CNN-based architecture which has been specially designed to classify the ripening states of Fuji apples. In contrast to other existing models based on band assignment, the proposed CNN is a feed-forward model that employs 1-D convolutions in its convolutional layers. In this way, convolutional kernels can automatically learn the most discriminating spectral patterns to successfully identify ripening states depending on the input data without doing any specific assignment of the spectral bands. More specifically, the input of the CNN is the normalized spectral information in the ranges from 450 to 1000 and the outputs are the related classes, i.e., unripe, half-ripe, ripe, and over-ripe, which are encoded as one-hot-encoding vectors where only the position associated with one class is activated. To successfully learn such mapping, the proposed architecture includes the following sequential building blocks: two convolutional layers with 64 kernels, a max-pooling layer, a flatten layer, and two dense (or fully connected) layers with 100 units. Now, let us describe how we adjusted the optimal number of layers, filters, neurons, and other hyperparameters of the proposed network.

Basically, all the hyperparameters were adjusted following a grid search method. This method consists in defining the tentative values of each parameter, training the model with all the possible combinations, and selecting the optimal configuration according to the least mean squared error (MSE). In more detail, the number of convolutional layers was set between 1 and 10. Besides, the number of filters was fixed to 32 or 64 or 128. The number of dense layers was selected from 2 to 20 and the number of neurons for each fully connected layers was set between 100 and 300. The possible activation functions were selected from Keras open-source software deep learning library.

After choosing the optimal hyperparameters, the reliability of the CNN was tested using threefold cross-validation. In such validation method, the test data is randomly divided into three disjoint subsets; each subset, or partition, contains 56 samples, with 14 samples for each ripening class. The threefold cross-validation consists in using two partitions for training (112 samples) and the remaining partition for test (56 samples). This process is repeated three times, using different training
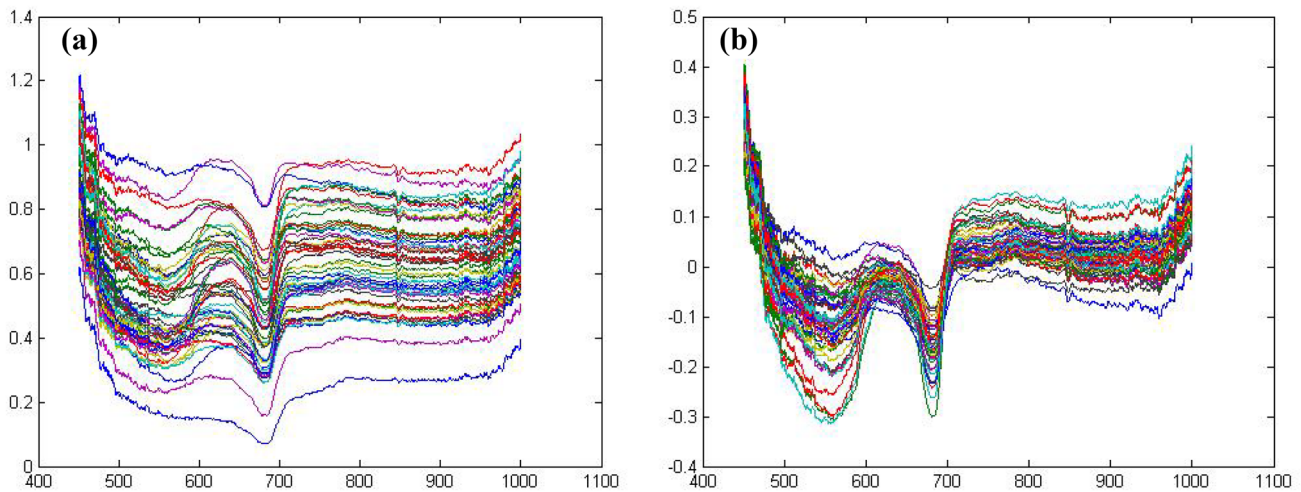

Fig. 2 Sample spectra of different Fuji apples: **a** reflectance spectra; **b** preprocessed spectra

and test subsets in each iteration. This validation method is very frequent in machine learning research (Fushiki, 2011), where it is commonly used when the dataset is not very large.

At this point, it is important to note that the proposed CNN hierarchically propagates the features extracted by the considered convolutional layers. Hence, the corresponding feature maps do not have any physical interpretation since, once the first layer convolves the input spectra, all the information is combined by means of affine transformations and non-linear activations.

Apple Ripeness Classification Using ANN, SVM, and KNN

To compare the reliability of the proposed method, three alternative methods were applied on the same dataset of spectral
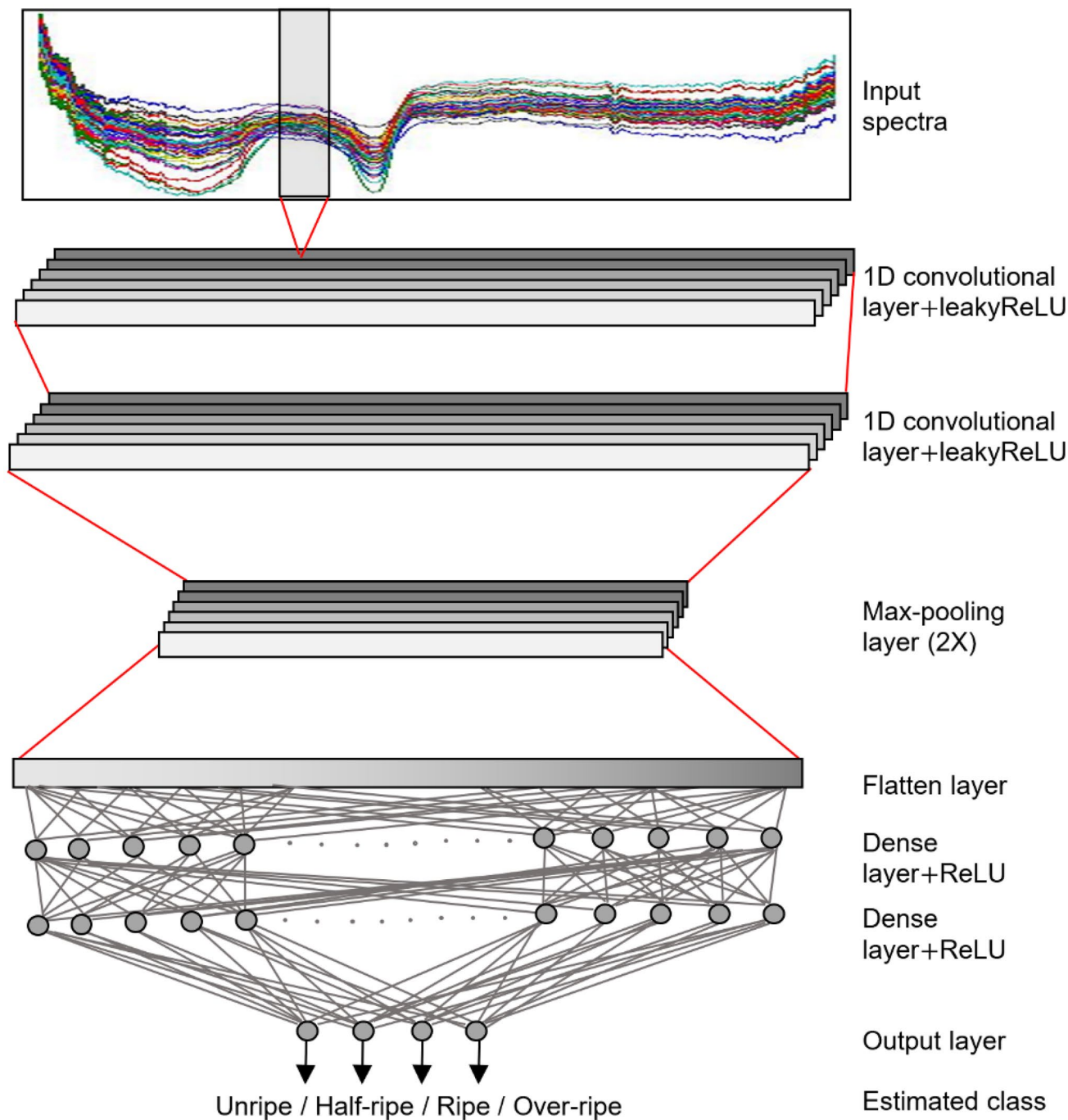

Fig. 3 Proposed convolutional neural network (CNN) architecture for estimating the ripening stage of Fuji apples based on Vis/NIR spectroscopy

data. In the first model, a classical artificial neural network (ANN) classifier, also known as multilayer perceptron (MLP), was developed. The ANN used in this study is a feed-forward, back-propagation neural network, whose hyperparameters were also selected by grid search. The input of the ANN is the spectral information in the range 450–1000 nm, and the outputs are the related classes (unripe, half-ripe, ripe, and over-ripe). The possible number of hidden layers was set between 2 and 20, and the number of neurons for each hidden layer was selected between 10 and 300. In this way, the cases with the highest number of layers (up to 20) can be considered as examples of deep neural networks, although they do not use convolutions.

Two other classification methods, not based on neural networks, were also applied for comparison, using support vector machines (SVM) and k-nearest neighbors (KNN) (Vapnik, 1999). Both of them are well-known classifiers in the field of machine learning. SVM finds the hyperplanes of maximum separability between the predefined classes in a high-dimensional space, using a kernel function for the distance measure and selecting the samples in the decision boundaries. On the other hand, KNN is also a sample-based method; given a new sample, it finds the class whose k-nearest samples in the train set are closest to the input sample. For the execution of SVM in the experiments, a linear kernel was used and C parameter (also known as the penalty parameter) was set to 1. In the case of KNN, the Euclidean distance was used, and the number of neighbors, k, was set to 3.

As in the CNN model, the reliability of all the classifiers were evaluated using threefold cross-validation with the same partitions of the dataset.

Selection of the Structure of CNN and ANN Networks

This section describes the experimental results obtained in the grid search processes to find the optimal structure of the CNN and ANN models. Both models were programmed on JupyterLab interface using Python version 3.7.0, Keras

Table 1 Optimal architecture of the proposed CNN for estimating the ripening stage of Fuji apples using Vis/NIR spectral data. *ReLU*, rectified linear unit; *LeakyReLU*, leaky rectified linear unit

Type of layer	Filter size	Number of filters	Number of neurons	Activation function
Convolutional	5×1	64	-	LeakyReLU
Convolutional	5×1	64	-	LeakyReLU
Max pooling	2×1	-	-	-
Flatten	-	-	-	-
Fully connected	-	-	100	ReLU
Fully connected	-	-	100	ReLU

Table 2 Optimal architecture of the ANN for estimating the ripening stage of Fuji apples based on Vis/NIR spectroscopy

Number of hidden layers	Number of neurons	Transfer function
2	200	1st layer: sigmoid
	200	2nd layer: sigmoid

version 2.2.0 (Gulli & Pal, 2017), and TensorFlow version 3.0 (Abadi et al., 2016).

Optimal Architecture of the CNN Classifier

As a result of the grid search process, the optimal setup found for the proposed CNN consists of six hidden layers: two convolutional layers, a max-pooling layer of size 2 × 1, a flatten layer, and two dense layers. Each convolutional layer contains 64 1D-filters of size 5 × 1, and each dense layer contains 100 neurons. Leaky rectified linear unit (LeakyReLU) was used as the activation function in the convolutional layers, and rectified linear unit (ReLU) was used as the activation function in the dense layers. The number of epochs, batch-size, and validation size were 30, 35, and 20, respectively. Figure 3 depicts the proposed architecture of the CNN resulting from this study, and Table 1 shows the corresponding optimal structure of the proposed CNN classifier.

Optimal Architecture of the ANN Classifier

Recall that the ANN architecture was also selected by grid search, where different options were assessed between 2 and 20 hidden layers. However, as the result of this process, the optimal configuration found for the ANN classifier consisted of only two dense layers, each of them with 200 neurons. In

Table 3 Performance measures of the classifiers associated with the confusion matrix

Criteria	Description
$Recall = \frac{TP}{TP+FN} \times 100$	How many positive samples are correctly detected
$Precision = \frac{TP}{TP+FP} \times 100$	How many detected outputs are correct
$Accuracy = \frac{TP+TN}{TP+FN+FP+TN} \times 100$	Total correct classification rate
$Specificity = \frac{TN}{TN+FP} \times 100$	How many negative samples are correctly detected
$F_{measure} = \frac{2 \times Recall \times Precision}{Recall+Precision}$	Recall and precision harmonic weighted average
TP: true positive	FP: false positive
TN: true negative	FN: false negative

Table 4 Classification performance obtained by the CNN model using threefold cross-validation, for estimating the ripening stage of Fuji apples

Real/estimated class	Unripe	Half-ripe	Ripe	Over-ripe	Number of samples	Classification error per class (%)	Correct classification rate (%)
Unripe	43	0	0	0	43	0	
Half-ripe	0	43	0	0	43	0	
Ripe	3	0	38	2	43	11.62	96.5
Over-ripe	0	0	1	42	43	2.32	

both layers, the activation function selected was the sigmoid function. Table 2 presents the optimal architecture of the ANN obtained with the grid search method.

Experimental Results and Discussion

As described in the "Materials and Methods" section, the CNN and ANN classifiers (after being adjusted with grid search) and the SVM and KNN methods were executed using the entire spectral data (450–1000 nm), applying a threefold cross-validation method to assess the predictive performance of the models. At each run of cross-validation, a confusion matrix was generated, and all such confusion matrices were added together to give the overall confusion matrix. Various common criteria were used to evaluate the performance of the proposed classifiers, including the performance parameters extracted from the confusion matrix, i.e., recall, accuracy, precision, specificity, and F-measure. The definition of these criteria is presented in Table 3. Additionally, the receiver operating characteristic curve (ROC) was also obtained for the two methods based on neural networks, CNN, and ANN, to show the effectiveness of these models for each class.

Ripeness Estimation Using the CNN Classifier

Table 4 contains the obtained classification performance of the proposed CNN method, including the overall confusion matrix, the error rate for each class, and the overall correct classification rate (CCR) or accuracy. As can be seen, only 6 of the 172 samples were misclassified, resulting in a CCR of 96.5%. The highest classification error of the CNN classifier was produced for the ripe class, with 11.62% error.

Table 5 presents the performance criteria which are obtained from the overall confusion matrix, namely recall, accuracy, specificity, precision, and F-measure. As shown in Table 5, the high recall (100%) obtained for unripe and half-ripe classes means that all the samples were correctly classified in these classes; thus, the low recall obtained for ripe class (88.37%) means that many ripe samples were misclassified in a different category. As demonstrated in Table 5, the highest accuracy is related to the class half-ripe (100%). The specificity was 100% for class half-ripe, which was higher than other classes. The precision was 93.47% for class unripe, which was the lowest value among all classes, indicating that many samples were incorrectly classified in other classes. Finally, since the F-measure is defined as weighted harmonic mean of recall and accuracy, it can be observed that the value for class half-ripe was higher than the others, which indicates that the classification of the samples of this class was perfect in this CNN model.

Ripeness Estimation Using ANN, SVM, and KNN Classifiers

Tables 6, 7, and 8 show the confusion matrices and the CCRs achieved by the ANN, SVM, and KNN classifiers, respectively. The experiments show that the effectiveness of the ANN and KNN classifiers are significantly lower than that of the CNN classifier. The obtained CCR is over 89.5% for ANN and 91.7% for KNN, which are 7% and 4.8% less than that obtained by the CNN. In other terms, the relative percent difference (RPD) is 7.5% for ANN and 5.1% for KNN. It should be noted the high misclassification rate of ANN classifier for the ripe class, with an error of 32.56%; although the classification of the half-ripe samples was perfect, the errors in the unripe and over-ripe classes are large (6.97%). Although the grid search tested different configurations of the ANN, the best structure was not able to achieve

Table 5 Performance measure obtained by the CNN model using threefold cross-validation, for estimating the ripening stage of Fuji apples

Class	Recall (%)	Accuracy (%)	Specificity (%)	Precision (%)	F-measure (%)
Unripe	100	98.25	97.67	93.47	96.62
Half-ripe	100	100	100	100	100
Ripe	88.37	96.51	99.22	97.43	92.68
Over-ripe	97.67	98.25	98.44	95.45	96.55

Table 6 Classification performance obtained by the ANN model using threefold cross-validation, for the estimation of the ripening stage of Fuji apples

Real/estimated class	Unripe	Half-ripe	Ripe	Over-ripe	Number of samples	Classification error per class (%)	Correct classification rate (%)
Unripe	40	1	0	2	43	6.97	
Half-ripe	0	43	0	0	43	0	
Ripe	7	0	29	7	43	32.56	89.5
Over-ripe	2	0	1	40	43	6.97	

the high accuracy obtained by the CNN. This indicates a clear superiority of the convolutional model, which is able to work correctly with spectral data.

On the other hand, the SVM classifier was able to achieve very accurate results, with a CCR of 95.93%. This is only 0.57% less than the CCR of the CNN classifier or an RPD of 0.6% in relative terms. As in the rest of methods, the highest error of SVM is found in the ripe class, with an error of 9.3%. The problem of this SVM method is that, since it is based on a selection of samples from the training set, it could present generalization problems when the dataset is changed. Table 9 presents the performance measures computed from the confusion matrices for the four models.

Again, the two best methods are CNN and SVM, with accuracies per class always above 98%. As it could be expected from the confusion matrices, recall is always lower for the ripe class, which is mostly confused with over-ripe and unripe classes. Figure 4 shows the receiver operating characteristic curves (ROC) for the CNN and ANN models for the four defined classes, using threefold cross-validation. As shown in Fig. 4a, the ROC curves of all the classes are very close to the ideal curve, indicating a high performance of the proposed CNN method.

Comparison with Other Works

After evaluating the performance of the proposed CNN classifier, the results obtained were compared with other studies reported by different researchers. Table 10 presents the results of these methods using the reported values of classification accuracy (CCR) for problems of ripeness classification of several types of fruits using color and spectral information. Although these works were selected among the most similar to our case in the field of ripeness classification, this comparison must be viewed in context since they refer

to different types of fruits, different datasets, and different types of classifiers. Only the first five works of this table are specific to ripening of Fuji apples. In any case, it can be seen that our method is able to reach a precise result in the context of the state of the art, with an accuracy of more than 96.5%.

In general, the methods based on recent convolutional and deep learning techniques are able to produce better results than those based on traditional machine learning models. Although these methods can be computationally expensive for the training tasks, after training the parameters of the network, the execution can be done in an inexpensive computing device, such as a smartphone. This is the case of the proposed technique, which requires only two convolutional layers and two dense layers. So, the proposed method could be applied for the development of a portable device for estimating the ripening stage of fruits in the field.

In the context of Fuji apples, the current classification methods range from 85.7% for (Mulyani & Susanto, 2017), which is based on fuzzy logic and RGB images, to 99.4% of (Pourdarbani et al., 2020) which uses color and spectral data. Since the datasets of these works are not the same, the obtained values cannot be directly compared. For example, Zhang et al. (2020) also used Vis/NIR spectroscopy, but the dataset consisted of 846 apple samples divided into three classes: immature, harvest maturity, and eatable maturity. Mulyani and Susanto (2017) used three classes: raw, half-ripe, and ripe, but the input data only contains RGB images. Although their obtained CCR is lower than the rest of methods, the advantages of using inexpensive and highly portable capture devices cannot be obviated since it can be applied with standard cameras instead of spectrometers.

Compared with other predefined networks, such as the commented AlexNet, VGG Net, ZF Net, and GoogLeNet, the specifically designed network was the benefit of adapting better to the problem of interest at a reduced cost. However,

Table 7 Classification performance obtained by the SVM model using threefold cross-validation, for the estimation of the ripening stage of Fuji apples

Real/estimated class	Unripe	Half-ripe	Ripe	Over-ripe	Total data	Classification error per class (%)	Correct classification rate (%)
Unripe	42	1	0	0	43	2.32	
Half-ripe	0	43	0	0	43	0	
Ripe	1	0	39	3	43	9.3	95.93
Over-ripe	0	0	2	41	43	4.65	

Food and Bioprocess Technology (2022) 15:2226–2236

Table 8 Classification performance obtained by the KNN model, with $K = 3$, using threefold cross-validation, for the estimation of the ripening stage of Fuji apples

Real/estimated class	Unripe	Half-ripe	Ripe	Over-ripe	Total data	Classification error per class (%)	Correct classification rate (%)
Unripe	39	1	3	0	43	9.3	
Half-ripe	2	41	0	0	43	4.65	
Ripe	4	0	38	1	43	11.63	91. 68
Over-ripe	0	0	3	40	43	6.98	

Table 9 Performance measures obtained by the four classification models using threefold cross-validation, for estimating the ripening stage of Fuji apples

| Method	Class	Recall (%)	Accuracy (%)	Specificity (%)	Precision (%)	*F*-measure (%)
CNN	Unripe	100	98.25	97.67	93.47	96.62
	Half-ripe	100	100	100	100	100
	Ripe	88.37	96.51	99.22	97.43	92.68
	Over-ripe	97.67	98.25	98.44	95.45	96.55
ANN	Unripe	93.02	93.02	93.02	81.63	86.95
	Half-ripe	100	99.41	99.22	97.72	98.95
	Ripe	67.44	91.37	99.22	96.66	79.45
	Over-ripe	93.02	93.02	93.02	81.63	86.95
SVM	Unripe	97.67	98.83	99.22	97.67	97.67
	Half-ripe	100	99.41	99.22	97.72	98.85
	Ripe	90.69	96.51	98.44	95.12	92.85
	Over-ripe	95.93	97.09	97.67	93.18	94.25
KNN	Unripe	90.70	90.70	96.74	86.67	88.63
	Half-ripe	95.34	95.35	98.32	97.61	96.47
	Ripe	88.37	88.37	96.00	86.36	87.36
	Over-ripe	93.02	93.02	97.75	97.56	95.23
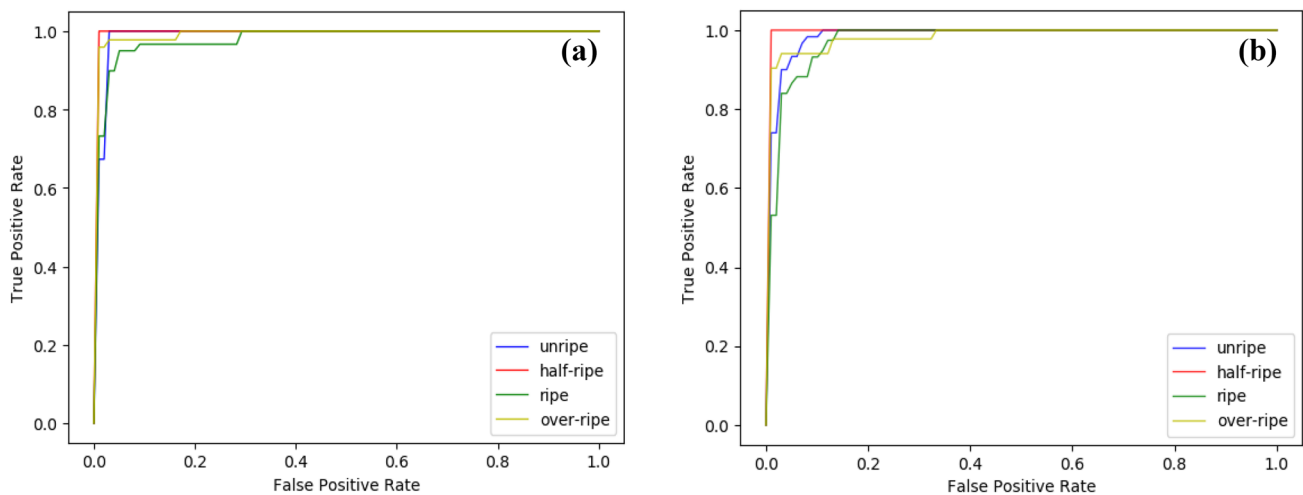
Fig. 4 ROC curves of the CNN and ANN models for estimating the ripening stage of Fuji apples, using threefold cross-validation: **a** CNN classifier; **b** ANN classifier

Table 10 Comparison of the correct classification rate (CCR) obtained by other works in the classification of ripening stage of different types of fruit using machine learning and hyperspectral data

Researchers	Fruit type	Method	CCR (%)
Proposed method	Fuji apple	CNN	96.5%
Pourdarbani et al. (2020)	Fuji apple	Ensemble classifiers with MV	99.37%
Pourdarbani et al. (2020)	Fuji apple	Fully connected ANN	92.12%
Zhang et al. (2020)	Fuji apple	Least squares support vector machine	89.05%
Mulyani and Susanto (2017)	Fuji apple	Fuzzy logic	85.71%
Mazen and Nashat (2019)	Banana	ANN	97.75%
Wei et al. (2014)	Persimmon	Linear discriminant analysis (LDA)	95.30%
Gao et al. (2020)	Strawberry	CNN	98.6%
Indrabayu et al. (2019)	Strawberry	SVM	85%
Khodabakhshian and Emadi (2017)	Pear	Partial least square-discriminant analysis (PLS-DA)	87.86%

for a practical application of the system in the field, the problems derived from the effect of natural lighting conditions in spectral data should be addressed, requiring more experimentation.

Conclusions

Non-destructive evaluation of fruit ripeness has received increasing attention as it offers many advantages over traditional destructive methods. Hyperspectral imaging has been proposed as a promising non-destructive and rapid technique for assessing fruit quality and safety. Therefore, in this paper, we have presented a new non-destructive method for fast and accurate estimation of the ripening stages of Fuji apples using visible and near-infrared spectroscopy (Vis/NIR), in the range of 450–1000 nm. To reach an effective estimation, a new convolutional neural network (CNN) architecture has been proposed using grid search. Then, the reliability of the CNN was measured applying threefold cross-validation.

In order to evaluate the validity of the proposed method, the CNN classifier was compared with three alternative classifiers based on artificial neural networks (ANN), support vector machines (SVM), and *k*-nearest neighbors (KNN). The experimental results indicate that the correct classification rate (CCR) of the proposed method achieved 96.5%, while the CCR of the ANN and KNN methods reached an average of 89.5% and 91.68%. Only SVM achieved a comparable CCR of 95.93%. Thus, the proposed CNN classifier proves to be a feasible method for fast and accurate classification of apple fruits for harvesting or post-harvesting operations. The robustness of the presented model against different crop seasons, locations, fruit varieties, and capture equipment should also be demonstrated by further experimentation under different capture conditions. As future work, it will be interesting to verify the obtained results with large datasets using the most effective spectral intervals. Data augmentation techniques can be used to increase the amount of data to get more artificial samples, which could be useful for improving the performance and outcomes of deep learning models. Another line of research would be to develop a non-destructive method using other properties of the Fuji apples as features to estimate their stages of ripening.

Declarations

References

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., ... & Zheng, X. (2016). {TensorFlow}: a system for {large-scale} machine learning. In *12th USENIX Symposium on Operating Systems Design and Implementation,* (OSDI 16), 265–283.

Baltazar, A., Aranda, J. I., & González-Aguilar, G. (2008). Bayesian classification of ripening stages of tomato fruit using acoustic impact and colorimeter sensor data. *Computers and Electronics in Agriculture, 60*(2), 113–121.

Beć, K. B., Grabska, J., & Huck, C. W. (2021). Principles and applications of miniaturized near-infrared (NIR) spectrometers. *Chemistry–A European Journal, 27*(5), 1514–1532.

Beć, K. B., & Huck, C. W. (2019). Breakthrough potential in near-infrared spectroscopy: Spectra simulation. A review of recent developments. *Frontiers in chemistry*, 7, 48.

Cerovic, Z. G., Goutouly, J. P., Hilbert, G., Destrac-Irvine, A., Martinon, V., & Moise, N. (2009). Mapping winegrape quality attributes using portable fluorescence-based sensors. *Proc INIA, FRUTIC 09, Conception, Chile, 09,* 301–310.

Fushiki, T. (2011). Estimation of prediction error by using K-fold cross-validation. *Statistics and Computing, 21*(2), 137–146.

Gao, Z., Shao, Y., Xuan, G., Wang, Y., Liu, Y., & Han, X. (2020). Real-time hyperspectral imaging for the in-field estimation of strawberry ripeness with deep learning. *Artificial Intelligence in Agriculture, 4*, 31–38.

Giovenzana, V., Beghi, R., Malegori, C., Civelli, R., & Guidetti, R. (2014). Wavelength selection with a view to a simplified handheld optical system to estimate grape ripeness. *American Journal of Enology and Viticulture, 65*(1), 117–123.

Gulli, A., & Pal, S. (2017). *Deep learning with Keras*. Packt Publishing Ltd.

Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the Cat's Visual Cortex. *The Journal of Physiology*, 160(1), 106–54.

Indrabayu, I., Arifin, N., & Areni, I. S. (2019). Strawberry ripeness classification system based on skin tone color using multi-class support vector machine. In *2019 International Conference on Information and Communications Technology (ICOIACT)* (pp. 191–195). IEEE.

Khodabakhshian, R., & Emadi, B. (2017). Application of Vis/SNIR hyperspectral imaging in ripeness classification of pear. *International Journal of Food Properties, 20*(sup3), S3149–S3163.

Khodabakhshian, R., Emadi, B., Khojastehpour, M., & Golzarian. (2017). Determining quality and maturity of pomegranates using multispectral imaging. *Journal of the Saudi Society of Agricultural Sciences, 16*(4), 322–331.

Kotwaliwale, N., Kalne, A., & Singh, K. (2012). Monitoring of mango (Mangifera indica L.)(Cv.: Chousa) ripening using X-ray computed tomography. In *2012 Sixth International Conference on Sensing Technology (ICST)* (pp. 326–330). IEEE.

Li, B., Lecourt, J., & Bishop, G. (2018). Advances in non-destructive early assessment of fruit ripeness towards defining optimal time of harvest and yield prediction—A review. *Plants*, 7(1), 3.

Mazen, F. M., & Nashat, A. A. (2019). Ripeness classification of bananas using an artificial neural network. *Arabian Journal for Science and Engineering, 44*(8), 6901–6910.

Mohtar, I. A., Ramli, N. S. S., & Ahmad, Z. (2019). Automatic classification of mangosteen ripening stages using deep learning. *2019 1st International Conference on Artificial Intelligence and Data Sciences (AiDAS) (IEEE)* (pp. 44–47).

Mulyani, E. D. S., & Susanto, J. P. (2017). Classification of maturity level of fuji apple fruit with fuzzy logic method. In *2017 5th International Conference on Cyber and IT Service Management (CITSM)* (pp. 1–4). IEEE.

Pourdarbani, R., Sabzi, S., Kalantari, D., Paliwal, J., Benmouna, B., García-Mateos, G., & Molina-Martínez, J. M. (2020). Estimation of different ripening stages of Fuji apples using image processing and spectroscopy based on the majority voting method. *Computers and Electronics in Agriculture, 176*, 105643.

Pourdarbani, R., Sabzi, S., Rohban, M. H., Hernández- Hernández, J. L., Gallardo-Bernal, I., Herrera-Miranda, I., & García-Mateos, G. (2021). One-dimensional convolutional neural networks for hyperspectral analysis of nitrogen in plant leaves. *Applied Sciences*, 11(24), 11853.

Sabzi, S., Abbaspour-Gilandeh, Y., García-Mateos, G., Ruiz-Canales, A., Molina-Martínez, J. M., & Arribas, J. I. (2019). An automatic non-destructive method for the classification of the ripeness stage of red delicious apples in orchards using aerial video. *Agronomy 9*(2), 84.

Schmutzler, M., & Huck, C. W. (2014). Automatic sample rotation for simultaneous determination of geographical origin and quality characteristics of apples based on near infrared spectroscopy (NIRS). *Vibrational Spectroscopy, 72*, 97–104.

Su, Z., Zhang, C., Yan, T., Zhu, J., Zeng, Y., Lu, X., ... & Fan, L. (2021). Application of hyperspectral imaging for maturity and soluble solids content determination of strawberry with deep learning approaches. *Frontiers in Plant Science*, 1897.

Vapnik, V. N. (1999). An overview of statistical learning theory. *IEEE Transactions on Neural Networks / a Publication of the IEEE Neural Networks Council, 10*(5), 988–999.

Wei, X., Liu, F., Qiu, Z., Shao, Y., & He, Y. (2014). Ripeness classification of astringent persimmon using hyperspectral imaging technique. *Food and Bioprocess Technology, 7*(5), 1371–1380.

Yang, H. Q., Kuang, B. Y., & Mouazen, A. M. (2011). Size estimation of tomato fruits based on spectroscopic analysis. *In Advanced Materials Research 225*, 1254–57.

Zhang, M., Zhang, B., Li, H., Shen, M., Tian, S., Zhang, H., ... & Zhao, J. (2020). Determination of bagged 'Fuji' apple maturity by visible and near-infrared spectroscopy combined with a machine learning algorithm. *Infrared Physics & Technology, 111*, 103529.