Computers and Electronics in Agriculture. 2018; 145: 1-10

Determination of egg storage time at room temperature using a low-cost NIR spectrometer and machine learning techniques

Julian Coronel-Reyes[a], Ivan Ramirez-Morales[a,b] , Enrique Fernandez-Blanco[b], Daniel Rivero[b], Alejandro Pazos[b]

[a] *Universidad Técnica de Machala, Faculty of Agricultural & Livestock Sciences, 5.5 km Pan-American Av, Machala, El Oro, Ecuador*
[b] *Universidade A Coruña, Department of Computer Science, 15071 A Coruña (03082), A Coruña, Spain*

Abstract
Currently, consumers are more concerned about freshness and quality of food. Poultry egg storage time is a freshness and quality indicator in industrial and consumer applications, even though egg marking is not always required outside the European Union.
Other authors have already published works using expensive laboratory equipment in order to determine the storage time and freshness of eggs. This paper presents a novel alternative method based on low-cost devices for the rapid and non-destructive prediction of egg storage time at room temperature (23 ± 1 °C).
H&N brown flock with 49-week-old hens were used as a source for the sampled eggs. Samples were scanned for a period of 22 days beginning from the time the egg was laid. The spectral acquisition was performed using a low-cost near-infrared reflectance (NIR) spectrometer which has a wavelength range between 740 nm and 1070 nm. The resulting dataset of 660 samples was randomly split according to a 10-fold cross-validation in order to be used in a contrast and optimization process of two machine learning algorithms. During the optimization, several models were tested to develop a robust calibration model.
The best model used a Savitzky Golay pre-processing technique with a third derivative order and an artificial neural network with ten neurons in one hidden layer. Regressing the storage time of the eggs, tests achieved a coefficient of determination (*R-squared*) of 0.8319 ± 0.0377 and a root mean squared error in cross-validation test set (*RMSECV*) of 1.97 days.
Although further work is needed, this technique shows industrial potential and consumer utility to determine an egg's freshness using a low-cost spectrometer connected to a smartphone.

Keywords
Non-destructive; Chemometrics; Freshness; Poultry; Neural networks

1. Introduction

Many people find eggs as an affordable source of nutrients. However, the freshness and quality of that source are highly influenced by the storage time and conditions declining along the time. Variability in freshness might be perceived by consumers as lack of quality. Additionally, degradation can reach a point where the egg is unfit for human consumption. For this reason, it is very important to develop methods to better monitor egg storage time (Abdel-Nour et al., 2011, Akter et al., 2014, Akyurek and Okur, 2009, Mathew et al., 2016).

Important and complex changes occur in eggs during storage. Predicting these changes is critical in order to monitor egg freshness. These changes include the thinning of albumen, weakening of the vitelline membrane and an increase in the water content of the yolk. The foaming and emulsifying properties of the albumen and yolk, respectively, are affected by the protein concentration, pH and ionic strength (Karoui et al., 2006).

Storage time, temperature, humidity, air quality, and handling are external factors which can contribute to the degradation of eggs. In particular, the storage time is related to two major issues: the reduction of the nutritional value of eggs (Stadelman et al., 1995) and the decrease of freshness in a logarithmic relation (Silversides and Scott, 2001).

Akter et al. (2014) demonstrated that egg weight, pH, oxidation and Haugh Units are also adversely affected with increasing storage time at room temperature. In the same work, the authors propose a maximum storage time of 14 days at room temperature (28–31 °C).

The freshness can be assessed by physical, biochemical, microbial and sensory parameters. The Haugh Unit (HU) method is a widely used but destructive method to measure egg quality (Haugh, 1937). However, quality measurements based on HU are biased by the strain and age of the hen (Silversides and Scott, 2001). Liu et al. (2007) demonstrated a high correlation between HU and storage time with an R-squared value of 0.9868.

Sensor technologies are an attractive strategy for non-destructive determination of freshness of the egg, either at the production plant or at food industry sites (Galiş et al., 2012, Karoui et al., 2006).

In recent years, non-destructive techniques for assessing freshness and storage time at room temperature have emerged. These techniques include electronic nose (Yongwei et al., 2009), ultrasound (Aboonajmi et al., 2014), ultraviolet-visible spectroscopy (Liu et al., 2007), hyperspectral imaging (Suktanarak and Teerachaichayut, 2017), and near-infrared spectroscopy (Abdel-Nour et al., 2011, Aboonajmi et al., 2015, Lin et al., 2015, Zhao et al., 2010).

The food industry has used NIR spectroscopy for a long time (Stark, 1996) because it is an accurate, rapid, and non-destructive quality analysis technique (Kumaravelu and Gopal, 2015). Recent works have been published using NIR to predict storage time associated with freshness in Atlantic salmon (Kimiya et al., 2013), large yellow croaker (Gangying et al., 2015), snow crab (Lorentzen et al., 2016), pork (Chen et al., 2011), apples (Liu and Tang, 2015), Valerianella locusta (Giovenzana et al., 2014), and eggs (Abdel-Nour et al., 2011, Aboonajmi et al., 2015, Lin et al., 2015, Zhao et al., 2010).

In the past ten years, the evolution of small, hand-held instruments has seen considerable growth (Barton, 2016, Haughey et al., 2014). Recently, some low-cost NIR devices have appeared in the market, making NIRS applications affordable and, therefore, more accessible to the wider public (Haughey et al., 2014).

NIR spectra are the result of vibrational transitions associated with chemical bonds present in most organic compounds (dos Santos et al., 2013, Kumaravelu and Gopal, 2015, Teye et al., 2013). The resulting spectrum is a consequence of the modifications made simultaneously in all the properties of the sample, making the calibration process more complicated (Florkowski et al., 2009, Martens and Naes, 1992).

Chemometrics has become an essential technique aimed at developing NIR calibration models. Using this technique, it is possible to process numerous samples in a short time (Moros et al., 2010).

Multivariate analysis techniques are commonly used to process spectral data, and techniques such as principal component analysis (PCA) and partial least squares (PLS) have been widely used (Kumaravelu and Gopal, 2015). Recently, some machine learning techniques are being presented as good alternatives to the classic techniques because they are based on pattern recognition (Brereton, 2015).

The aim of this study was to assess the potential of a low-cost NIR spectrometer as a non-destructive and rapid technique for egg storage time assessment. A more specific objective was to develop and evaluate a chemometric NIR calibration model based on machine learning techniques for the determination of egg storage time at room temperature.

2. Materials and methods

The overall methodology, as seen in Fig. 1, consists of three segments: the acquisition of the data (Section 2.1), data partition using a cross-validation technique (Section 2.2) and the optimization of the chemometric model (Section 2.3). Each segment consists of several steps. In the following subsections, the methodology is described in detail.

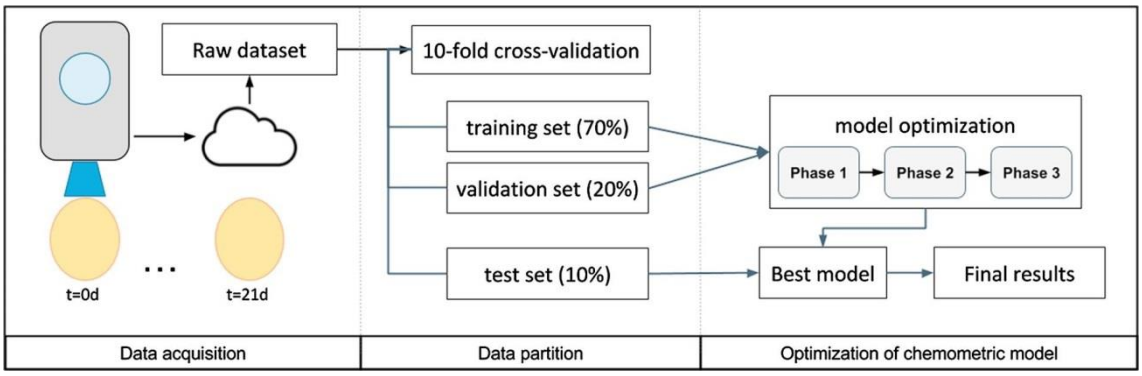

Fig. 1. Diagram of segments and steps in experimental methodology.

2.1. Spectral acquisition of eggs

A SCiO^{TM} handheld NIR Spectrometer was used to collect the sample, which has a spectral range between 740 nm and 1070 nm (Goldring et al., 2016). Its low price and reduced dimensions allow the researchers to perform rapid tests which could be developed and integrated for example into a smartphone (Cartwright, 2016, Das et al., 2015, Pügner et al., 2016). Previous works have already pointed to the high potential of this low-cost device (Haughey et al., 2014, Schulte et al., 2015).

Samples were collected from intact shell eggs, which were scanned twice from the top in the blunt end using the SCiO^{TM} shade accessory. The use of that accessory helps to avoid the influence of external light when the spectra were sampled and also it also helps to keep the same 10-mm distance in all the 660 collected spectral signals.

While the aim is to develop a non-destructive method to measure the freshness of the eggs, the sampled spectra represent the information of the shell and, consequently, no information of the inner parts are included, e.g., the yolk. A scheme of scanning method is shown in Fig. 2a, while Fig. 2b shows a raw signal of a sample along with the dark and white reference values taken with the spectrometer.

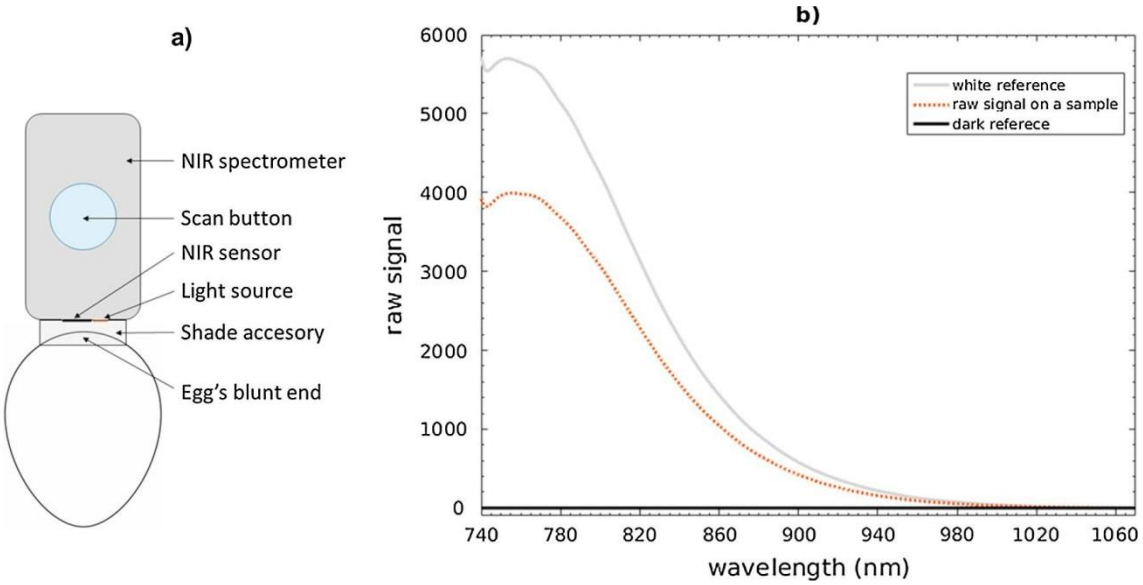

Fig. 2. (a) Scanning process, (b) Plot of raw signal of the sample along with dark and white references.

A 1-nm resolution between 740 nm and 1070 nm was used in the spectral data recording, which was lately stored in a cloud-based dataset with their corresponding reference values for time of storage. Using a research license of SCiO Lab, egg spectral signals were downloaded and imported into Matlab (The MathWorks Inc., Natick, MA) in order to develop and optimize the chemometric models.

Those 660 spectral curves are the result of 22 days of continuous monitoring of 30 shell-intact brown poultry eggs with weights between 55 g and 65 g (size M). Eggs used in the study were picked up from a flock of 20.000H&N strain hens between 49 and 52 weeks old. Those hens were housed in a stacked cage system and were fed with a standard ration without the use of egg-laying promoters.

The spectral data used for experimentation were obtained by averaging two duplicate measurements taken successively at the blunt end. Eggs were scanned in the poultry house immediately after being laid (day 0) and then transported to the laboratory in a thermally insulated container. Measurements from day 1 to day 21 were obtained in laboratory conditions monitored hourly at 23 ± 1 °C and a relative humidity of $90 \pm 2\%$. The interval between each measurement was exactly 24 h, and the employed procedure is simple with a very short time required to perform the measurements. A non-destructive technique was used in this experiment because the intention was to understand how the spectrum is modified in each of the eggs over time.

2.2. Data partition

The raw dataset was downloaded and then partitioned using a repeated 10-fold cross-validation technique in order to have training, validation and test subsets for optimization of the calibration model.

The model performance measures should be evaluated in a set of new data which have not been used for the training model. A good model should be able to make accurate estimations on this test data (Mucherino et al., 2009).

Cross-validation is one common technique applied in machine learning to maximize the use of available data. In this technique, the dataset is randomly divided into multiple subsets for training and testing the model. Cross-validation is used to avoid overfitting of the model. (Kuhn and Johnson, 2013, Refaeilzadeh et al., 2009).

In this work, spectral data were divided into training (calibration), validation and test subsets using a variation known as repeated cross-validation (Garcia and Filzmoser, 2015, Kuhn and Johnson, 2013). A repeated 10-fold cross-validation technique was chosen. Therefore, data were split into 10 groups in which 9 are used as calibration/validation sets, and the remaining one is used as a test set. This process was repeated 50 times.

The training and test set were changed until all folds had been tested. Data partition for each fold randomly divided the dataset, allotting 462 samples (70%) for training, 132 samples (20%) for validation and 66 samples (10%) for testing.

2.3. Optimization of calibration model

The relationship between the response in the spectral region of the NIR spectra and the target is often nonlinear (Bertran et al., 1999). The origin of these nonlinearities is difficult to identify and, for this reason, calibration is often performed using multivariate analysis (Martens and Naes, 1992). To develop a chemometric model, the NIR spectra, the reference values for calibration and an algorithm to link them are all required (Barton, 2016).

Parameters of the model were optimized in three consecutive phases. With the parameters identified by the best model, an evaluation of unseen data (the test set) was performed in order to estimate the future performance of the model on new data.

This step was performed in three consecutive phases of optimization. In Phase 1, two modelling algorithms (Section 2.3.1 Partial Least Squares (PLS), 2.3.2 Artificial neural networks (ANN)) were tested simultaneously with seven pre-processing techniques (Section 2.3.3). In Phase 2, the parameters of the selected model were tuned in order to optimize its performance. In Phase 3, the wavelength selection threshold (Section 2.3.4) was fine-tuned.

2.3.1. Partial Least Squares (PLS)

PLS was introduced by the Swedish statistician Herman Wold (Wold, 1985). In chemometrics, PLS-regression is used as a basic method for relating two data matrices using a linear multivariate model. However, this method goes beyond traditional regression since it has the ability to analyse data with incomplete, noisy and collinear variables (Wold et al., 2001).

This method is widely applied in NIR spectroscopy where multiple input variables are required. The accuracy of the model in PLS-regression improves when the number of relevant variables and the number of observations is increased (Hattori and Otsuka, 2017).

The aim of the PLS model is to find the direction of X in a multidimensional space, which explains the maximum variance direction in the Y. PLS regression is suited when the problem has more predictor variables than the number of observations and when there could be multicollinearity among X values (Yu et al., 2017).

2.3.2. Artificial neural networks (ANN)

Artificial neural networks (ANN) are data-modelling tools that aim to analyse complex relationships between inputs and outputs. In recent years, ANN have become a subject of much relevance in the scientific and research field and are inspired by the human central nervous system, which has numerous cells that work quickly and helps in decision making (Cascardi et al., 2017).

The Multilayer Perceptron (MLP) is a type of layered neural network with connections between consecutive forwarding layers. Fig. 3 shows the general scheme of an MLP, which consists of one input layer, one or more hidden layers and one output layer. The transfer function of neurons is commonly a sigmoid function, but other functions can also be used (Kruse et al., 2013, Ruck et al., 1990).

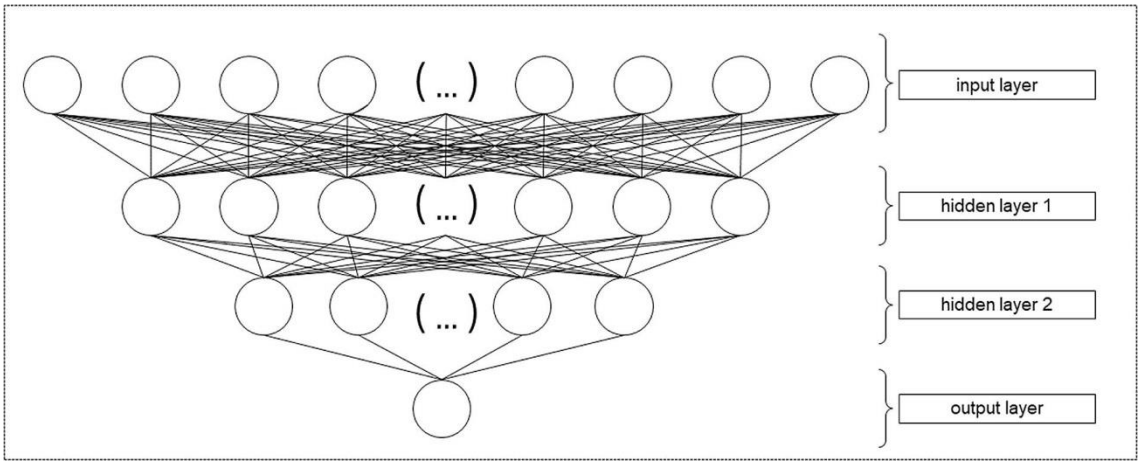

Fig. 3. Multilayer perceptron representation.

Each neuron receives the output signals of the neurons in the previous layer and provides an output for the next layer. The output layer receives, as input, the output of the last hidden layer and returns the output of the network (Gardner and Dorling, 1998, Kruse et al., 2013).

The number of neurons, layers and their connections is commonly known as the architecture of the neural network and is one of the key parameters to be optimized. The architecture depends on the complexity of the problem, and there is no general method for choosing the best structure. Choosing a good architecture is an empirical process, where multiple architectures are tested in order to find one that offers satisfactory results (Herrera et al., 2004, Rivero et al., 2011).

The use of artificial neural networks has been successfully applied to NIR spectrometry in the rapid quantification of wine compounds (Martelo-Vidal and Vázquez, 2015), evaluation of chemical components and properties of the jujube fruit (Guo et al., 2016) and characterization of blends containing refined and extra virgin olive oils (Aroca-Santos et al., 2016).

2.3.3. Pre-processing techniques

The data contained in the NIR spectra are the composition of several signals with overlapping information (Blanco and Villarroya, 2002, dos Santos et al., 2013, Rinnan et al., 2009). Therefore, NIR spectra are usually pre-processed to eliminate physical phenomena, which may alter raw spectra. This practice is an integral part of the chemometric modelling (Blanco and Villarroya, 2002) and is considered one of the most important steps.

Choosing the most suitable pre-processing technique is a quite difficult process to be assessed before the validation of the model. As a consequence, NIR spectra pre-processing is mainly performed by a trial and error approach (Rinnan et al., 2009, Xu et al., 2008; Xu et al., 2008). In this work, raw spectral signal and other six common pre-processing techniques were analysed:

(1) **Raw spectra:** NIR-reflectance measurement of a sample includes both the diffusively reflected and specularly reflected radiation (Rinnan et al., 2009).
(2) **Savitzky Golay:** includes a smoothing step for the numerical derivation of a vector. In this method, a p order polynomial is fitted in a symmetric window of w-width on the raw data, and then the d-order derivative is calculated at centre point i (Savitzky and Golay, 1964).
(3) **Beer-Lambert law:** This law suggests a linear relationship between the reflectance of the spectra and concentration of components (Rinnan et al., 2009).
(4) **Standard Normal Variate (SNV):** SNV is a method for scattering correction of NIR data (Barnes et al., 1989).
(5) **Multiplicative Scatter Correction (MSC):** MSC was first introduced by Martens et al. (1983). In this technique, undesirable scatter effects of spectra are removed from the data matrix prior to δατα modelling. The first step is comprised of the estimation of the correction coefficients and the second step consists of the correction of the recorded spectrum (Rinnan et al., 2009).
(6) **First Spectral Derivative (FSD):** This technique has been used in analytical spectroscopy for decades due to its ability to remove additive and multiplicative effects of the spectra (dos Santos et al., 2013). The finite difference is the basic method for spectral derivation.
(7) thus, the first derivative is calculated as the difference between two subsequent spectral points.
(8) **Second Spectral Derivative (SSD):** This technique is also based on the finite difference for spectral derivation.
(9) the second derivative is calculated as the difference between two subsequent points from the processed FSD signal. According to Rinnan et al. (2009) both FSD and SSD techniques should be avoided in practice since it is not feasible for most real measurements due to noise inflation.

2.3.4. Wavelengths selection threshold

Frequently, NIR spectra contain a large amount of information along the wavelength range. For this reason, it is important to perform a technique aimed at reducing this amount of data (Blanco and Villarroya, 2002). These techniques have now become a necessity and a requirement in chemometrics (Guyon et al., 2008, Saeys et al., 2007).

In machine learning techniques, the proper selection of wavelengths to obtain a small subset with lower sensitivity to non-linearities is usually effective in improving the performance of the models (Leardi et al., 1992, Saeys et al., 2007). Both selection of relevant wavelengths related to the compound of interest and avoiding interference of other wavelengths should be the goals in terms of building a robust predictive model.

Recent works on this topic point the importance of the selection of the most informative variables. For example, some methods applied to choose the informative features are successive projections algorithm (Liu et al., 2013), filter techniques (Prajapati et al., 2016), genetic algorithms (Anzanello et al., 2017, Cheng et al., 2016). In this study, a wavelengths selection method based on a correlation based filter (Hall, 1999, Mireei et al., 2017) was chosen to extract the informative wavelengths.

Wavelengths selection does not alter the original representation of the variables. This technique simply chooses a subset of the best wavelengths for the model (Saeys et al., 2007). A threshold is normally used together with a filter technique of either a univariate or multivariate model to evaluate which are the best wavelengths (Szymańska et al., 2015).

The present work used a multivariate filter described by Hall (1999) called Correlation-based Feature Selection (CFS). This filter is a simple algorithm that ranks feature importance according to its correlation function with the predicted variable. By using this method, the relevant informative wavelengths will be selected and, therefore, we expect to obtain an improvement of the model's performance.

2.3.5. Model evaluation

All models were evaluated using performance measures as the coefficient of determination (R-squared) in a validation set from the cross-validation (dos Santos et al., 2013, Viscarra Rossel, 2008).

Mean, and standard deviation of R-squared obtained from 500 models (50 repetitions by 10-fold cross-validation) were calculated for each model configuration in order to decide the best parameters of the model.

Final results are presented by a test set from the cross-validation. R-squared, mean absolute error (MAE) and root mean square error (RMSECV) were obtained according to Eqs. (1), (2), (3) respectively. MAE and RMSECV represent the difference between predictive values and the actual values (Armstrong and Collopy, 1992, Hyndman and Koehler, 2006).

$$R - squared = \frac{\sum_{i=1}^{i=n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{i=n} (y_i - \bar{y})^2} \quad (1)$$

$$MAE = \frac{\sum_{i=1}^{i=n} |y_i - \hat{y}_i|}{n} \quad (2)$$

$$RMSECV = \sqrt{\frac{\sum_{i=1}^{i=n} (y_i - \hat{y}_i)^2}{n}} \quad (3)$$

where y_i the real value of the i-th observation, $\hat{y}_i$ is the predicted value of the i-th observation, $\bar{y}$ s the average of real values and is the number of observations.

3. Results and discussion

Data acquisition and data partition were broadly covered in And 2.1 Spectral acquisition of eggs, 2.2 Data partition. Therefore, in this section, the results obtained in the three phases of the calibration model optimization are presented.

3.1. Spectra presentation

Response bands in NIR spectral region (720–1100 nm), are mainly the result of CH 3rd overtones, OH 2nd overtones and NH 2nd overtones of the fundamental molecular bonds in the Mid Infrared Region. In Fig. 4, the mean of raw spectral data is presented according to the storage time, and overtones of the fundamental molecular bonds are highlighted.

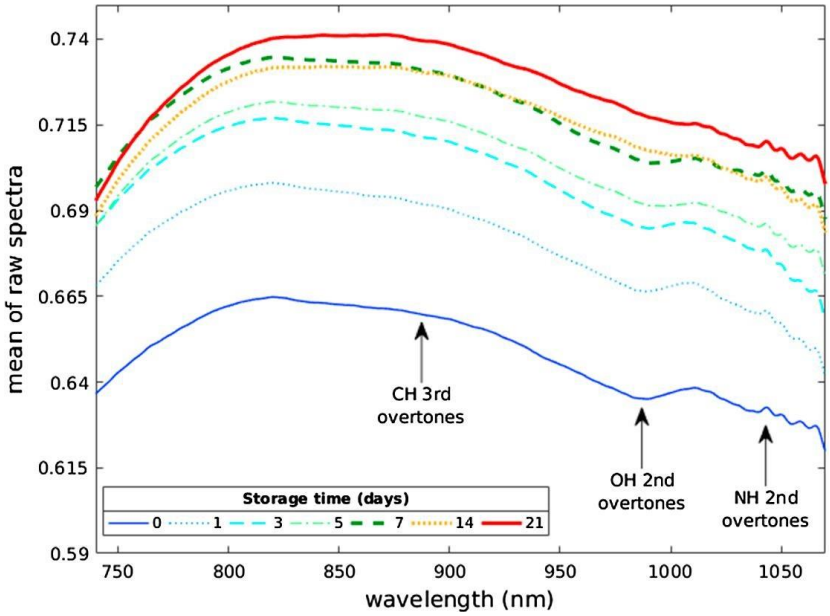

Fig. 4. Mean of raw spectra of eggs at different storage times.

According to Lammertyn et al. (2000) the penetration depth of light is directly dependent to wavelength. For example, a light in the range between 700 and 900 nm can penetrate up to 4 mm, while a light between 900 and 1900 nm sets it maximum penetration capacity between 2 and 3 mm. The spectral range of the measuring instrument, used in this work, allows a penetration between 3 and 4 mm which is enough to pass through eggshell and reach the interior. Therefore, changes in the spectra are directly related to the shell, cuticle and albumen of the egg even though there is no information of the yolk because penetration capacity is not enough to provide information about this one.

Focusing the attention on raw spectra of a freshly laid egg, they showed a very short OH band because the eggshell surface is moist shortly after an egg is laid. A gradual increment appeared in the intensity of the OH band through time, due to moisture loss as the cuticle of the egg dries. At the same time, that cuticle becomes thinner and the eggshell carbonate mineral becomes more exposed to the surface. This lastly process can be observed in Fig. 4 at 890 nm where CH overtones increase its value with time. Similar changes have also been observed in the behaviour of NH band, which can be the consequence due to a gradient in the chemical composition of the cuticle and the albumen, both rich in proteins. Therefore, with the passage of time, changes on molecular bonds CH, OH and NH can be observed as it is shown in Fig. 4.

It must be highlighted that the most obvious changes occur during the first 7 days, whilst changes between 7 and 14 days are less evident. Although, from 14 to 21 days of egg storage, changes are also noticeable again. This lastly change is probably the result of the decomposition process which has already begun at room temperature according to Akter et al. (2014) who propose a maximum storage time of 14 days in those conditions.

3.2. Phase 1: Selection of modelling algorithm and pre-processing technique

This phase was aimed to choose the modelling algorithm and the pre-processing techniques simultaneously. A grid search technique (Koch et al., 2012, Ma et al., 2015) was used to evaluate models.

PLS and ANN algorithms were trained using the same partition schema to ensure that both models received exactly the same input data and thus made the results comparable. Due to the influence of the wavelength selection threshold, the pre-processing techniques were evaluated at all thresholds using a correlation-based feature selection filter (Hall, 1999).

The results of the mean R-squared of 50 repeated 10-fold cross-validations, achieved using PLS with 10 latent variables, can be seen in Fig. 5a. Best results are obtained with pre-processing technique 4 (Standard Normal Variate) at a 70% wavelength selection threshold and pre-processing technique 6 (First Spectral Derivative) at a 50% wavelength selection threshold. In both cases, the results are below 0.8.

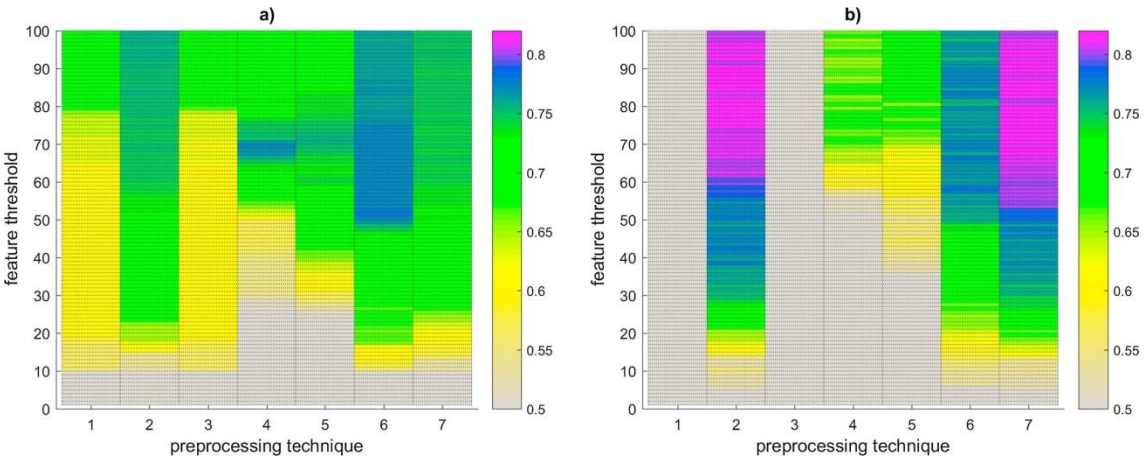

Fig. 5. R-squared in CV validation set obtained with the seven pre-processing techniques at different values of wavelength selection threshold. (a) Using PLS (b) Using ANN.

Fig. 5b shows the results of the mean R-squared of 50 repeated 10-fold cross-validations achieved using ANN with 10 neurons in one hidden layer. Best results are obtained by pre-processing technique 2 (Savitzky Golay) at a 90% wavelength selection threshold and pre-processing technique 7 (Second Spectral Derivative) at a 60% wavelength selection threshold. In both cases, the R-squared results are above 0.8. ANN models outperform the PLS models. Therefore, ANN will be used for optimization moving forward.

Despite the fact that the Savitzky Golay and Second Spectral Derivative techniques have similar results, the latter should be avoided in practice because it is not feasible due to the added noise (Rinnan et al., 2009). Additionally, the Savitzky Golay technique has tuning parameters which can be useful to optimize the model (Savitzky and Golay, 1964).

Once the pre-processing technique was chosen, optimization was executed again to optimize the wavelength selection threshold. The main reason for this is that Savitzky Golay's tuning parameters can transform the input pattern and therefore a reevaluation of the selected wavelengths is necessary. Therefore, this parameter was set to 100% for this part of the process.

A range of Savitzky Golay widths, specifically odd numbers between 3 and 101, were tested with different configurations of Polynomial order and Derivative order (between first and fifth). Table 1 shows the results of the mean, standard deviation, min and max, and p-value of a Tukey Honest Significant Difference (Tukey, 1949) test of R-squared of 50 repeated 10-fold cross-validations achieved at the best width for each configuration.

Table 1. R-squared of SavGol parameters at best width.

Polynomial degree	Derivative order	width	mean ± std	min	max	Tukey HSD
First	First	7	0.7644 ± 0.069	0,360	0.861	c
Second	First	7	0.7654 ± 0.067	0,445	0.894	c
Second	Second	29	0.8204 ± 0.047	0,610	0.909	ab
Third	First	19	0.7678 ± 0.065	0,472	0.880	c
Third	Second	21	0.8214 ± 0.045	0,675	0.910	ab
Third	Third	53	0.8249 ± 0.044	0,690	0.926	ab
Fourth	First	13	0.7679 ± 0.073	0,267	0.899	c
Fourth	Second	41	0.8222 ± 0.045	0,561	0.926	ab
Fourth	Third	51	0.8248 ± 0.043	0,618	0.914	ab
Fourth	Fourth	61	0.8199 ± 0.042	0,651	0.909	ab
Fifth	First	5	0.7684 ± 0.07	0.123	0.889	c
Fifth	Second	39	0.8252 ± 0.045	0,624	0.931	a
Fifth	Third	67	0.8276 ± 0.041	0,590	0.910	a
Fifth	Fourth	61	0.8189 ± 0.042	0,660	0.907	ab
Fifth	Fifth	67	0.8107 ± 0.045	0,607	0.913	b

Rows with different letters differ significantly according to Tukey HSD for a value of $p < .01$.

There is an intrinsic redundancy in the hierarchy of Savitzky Golay derivation. For each Polynomial order, the subsequent derivative order gave the same estimate of the coefficients. For example, for the first-degree polynomial, a first derivative and second derivative will give the same answer. The results of redundant configurations are not presented.

According to the Tukey HSD test, best configuration of Savitzky Golay is a fifth polynomial degree and a second or third derivative order. The latter was chosen since it had a greater mean and smaller standard deviation in test data.

Therefore, the optimized Savitzky Golay technique for this experiment consisted of 67 smoothing points (width), a fifth polynomial degree and a third order derivative.

There were evident differences in the spectra of the eggs as a function of storage time. Fig. 6a shows the result of applying the Savitzky Golay pre-processing technique. As the egg storage time increased, the obtained spectra show different characteristic values. These differences can also be seen in Fig. 6b, which shows an average of the signals corresponding to the eggs stored at 0, 7, 14 and 21 days.

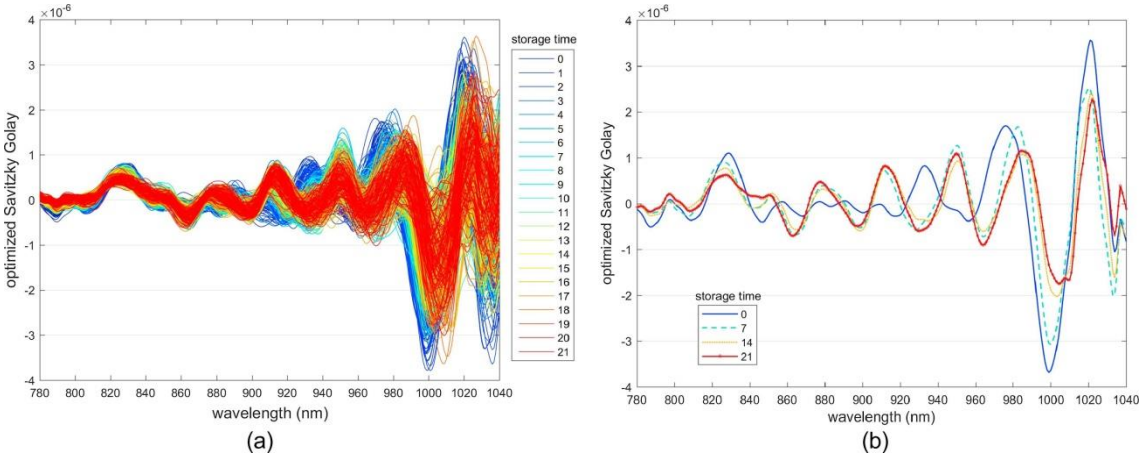

Fig. 6. Egg storage time represented by Savitzky Golay derivative spectra with optimized parameters. (5a) All spectral signals. (5b) mean spectra at 0, 7, 14, and 21 days of storage.

Visually exploring the dataset, some patterns emerge that indicate the spectra of eggs change as the storage time is increased. Notice in Fig. 6b, at wavelengths of 860–940 nm, that fresh eggs with zero days of storage have a Savitzky Golay reflectance near zero, while stored eggs oscillate between 1 and -1 in the same wavelength. Fresh eggs have a peak at 933 nm while stored eggs show a similar peak earlier at 913 nm.

At 983 nm and 999 nm, the difference between eggs stored up to 7 days and those stored longer is evident. In a visual exploration of data, it is difficult to find differences between eggs stored for more than 14 days. According to Akter et al. (2014), eggs can be stored a maximum of 14 days at room temperature. The small differences between the spectra of eggs stored for more than 14 days may be indicative of a deterioration in the freshness of the eggs.

This differences in spectra are related to the changes occurring in the shell and the cuticle of the egg (Rodríguez-Navarro et al., 2013). Molecular bonds CH, OH and NH change with the passage of time, as mentioned in Fig. 4.

3.3. Phase 2: Optimization of selected algorithm parameters

During phase 1, the model using an ANN algorithm obtained better results. For this reason, the ANN algorithm was selected. In this phase, we then optimized the architecture of the neural network. Architectures of both one and two layers were evaluated using a grid search (Koch et al., 2012, Ma et al., 2015). Values ranging from 0 to 200 neurons in each layer with intervals of 10 were evaluated.

Fig. 7b shows the standard deviation of 50 repeated 10-fold cross-validated models at each architecture. This measure indicates the stability and the repeatability of the proposed model. Best results are achieved in architectures of one hidden layer with 10 or 20 neurons and architectures of two hidden layers in which the second layer has 10 or 20 neurons.

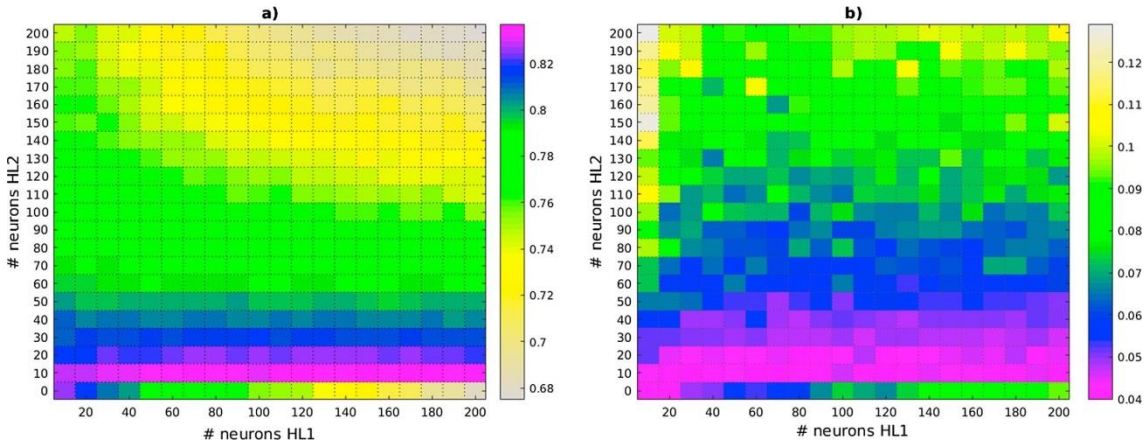

Fig. 7. R-squared in 50 repeated 10-fold CV validation set. (a) mean (b) standard deviation.

The best results are achieved with an architecture of two hidden layers with 180 and 10 neurons, respectively. We observed that using an architecture of 10 neurons in one hidden layer yields statistically equal results. According to Abu-Mostafa (1989), we should select the least complex system that produces the same results, which in this work is an architecture of 10 neurons in one hidden layer.

3.4. Phase 3: Fine tuning of wavelength selection threshold

The following experiment aims to find the best threshold value for wavelength selection (75, 150). Fig. 8 shows the influence of wavelength selection threshold on R-squared results.

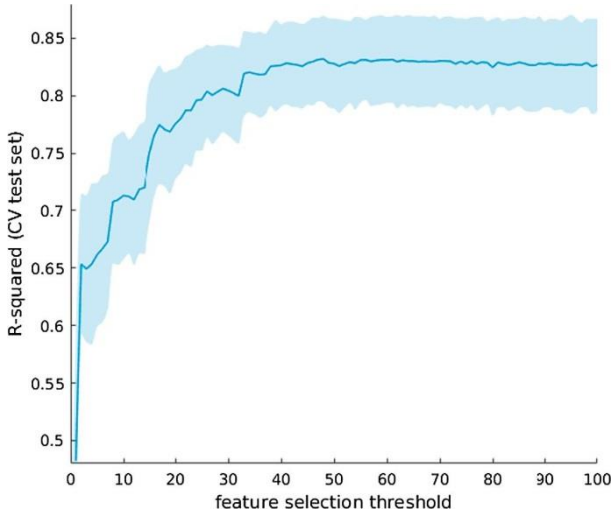

Fig. 8. Results of R-squared at different values of wavelength threshold.

Results above 0.8 are obtained with values of the wavelengths selection threshold from 30 onwards. Fig. 8 shows the importance of appropriate selection of wavelengths as inputs of the model. A wavelengths selection threshold of 48 was selected since the R-squared value obtained in test data is equal to 0.8319 ± 0.0377.

Although the evaluation of storage time of poultry eggs involves a complex process, this study shows that egg quality can be predicted using an ANN model with relevant wavelengths as input patterns. Our results confirm the conclusions stated by Guyon et al. (2008), that appropriate selection of relevant wavelengths is very important and must be utilized as a strategy to avoid the inclusion of uninformative and redundant wavelengths in the predictive model.

3.5. Evaluation of the model in unseen data (test set)

Once all parameters were optimized, the final configuration of the model involved a Savitzky Golay pre-processing technique with a fifth polynomial degree, a third derivative order and a width of 67. The ANN model has 10 neurons in one hidden layer and a wavelength selection threshold of 48. Those parameters are the consequence of the experimental results achieved during Phases 1–3, in which, optimal parameters were selected.

The calibration model using the previously mentioned parameters was tested using unseen data to evaluate its performance and generalization ability.

Fig. 9a shows a regression plot between the values of egg storage time predicted by the model with their corresponding actual values. It can be seen that most predictions present an absolute error of approximately 2 days, some predictions show an error of approximately 4 days, and one value has the highest error of 6 days. Fig. 9b shows the histogram of absolute errors.

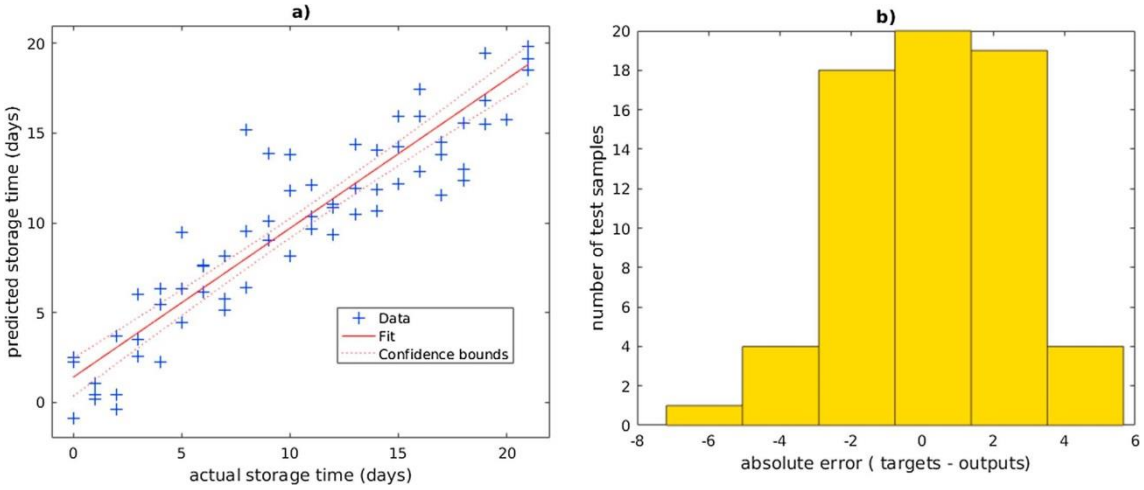

Fig. 9. Model performance in test set data. (a) Regression fit plot actual vs predicted. (b) Absolute error histogram.

To evaluate the performance of the model in unseen data, R-squared and RMSECV were calculated for the model and can be seen in Table 2. The proposed model has a high significance in all subsets.

Table 2. Performance metrics in training, validation and test set.

	Training	Validation	Test
n samples	462	132	66
R-squared	0.865	0.862	0.873
Mean Absolute Error	1.834	1.998	1.810
Root Mean Squared Error (days)	2.15	2.19	1.97
F-statistic vs. constant model	3.09e+03	673	449
p-value	<.01	<.01	<.01

The differences between the results of the test and training set performances could indicate the degree of overfitting (Stockwell and Peterson, 2002). However, Table 2 shows that these parameters are similar between the training set, validation set and test set. Therefore, we can affirm that the models are well fitted within the context of the calibration set.

Although the R-squared as a performance metric of the model is below 0.9, the RMSECV is about two days, which is an acceptable safety margin for the user. Some authors have evaluated predictive models for freshness assessment. Lin et al. (2011) obtained an R-squared value of 0.879 and an RMSECV of 2.443 using NIR spectroscopy for Haugh unit. Sun et al. (2015) used artificial vision and dynamic weighing to obtain an R-squared value of 0.8653 and an RMSECV of 3.745. Abdel-Nour et al. (2011) achieved an R-squared value of 0.89 and an RMSECV of 1.65 in their validation dataset for the prediction of storage days using a lab grade VIS/NIR spectroradiometer.

Our results are similar to those mentioned above, and this clearly indicates that there is potential for portable NIR instruments as an alternative to desktop lab instrumentation, which is in agreement with the conclusions of Haughey (2014). This is made increasingly possible due to the recent advances in sensor technology including Changhong (Rateni et al., 2017), who have developed an embedded NIR sensor. It is possible to create applications using the proposed model. This approach could enable consumers to predict the storage time of poultry eggs.

4. Conclusions and future work

Our findings show the potential of smartphone-connected devices using shortwave NIR as an effective method for the evaluation of storage time in poultry eggs. Spectral analysis of eggs is a rapid and non-destructive method for egg storage time determination.

Reflectance spectral data of the egg contains relevant information about its storage time, a parameter of freshness. There were noticeable differences in the processed spectral values of the eggs as a function of storage time.

Suitable predictive models were built with PLS and ANN regression techniques but the latter performed better, achieving an R-squared value of 0.873 and an RMSECV of 1.97 in test set data, suggesting that the spectra obtained with the smartphone-connected NIR spectrometer can be used as a non-destructive method for the assessment of egg storage time, a parameter of quality and freshness.

The use of a smartphone-connected NIR spectrometer is recommended for egg storage time assessment. However, further work is needed to assess the long-term reliability of the system. Currently, a combination of this method with traditional destructive techniques is recommended.

Future work will focus on exhaustive experimentation at both room and refrigeration temperatures using this low-cost spectrometer on eggs from hens of diverse ages and strains. The use of hyperspectral imaging for storage time prediction is also a potential technique to be studied.

Acknowledgments

The authors wish to thank Agrolomas CL for providing samples directly from their poultry houses. A special acknowledgement to CEDIA National Research and Education Network. Iván Ramírez-Morales and Enrique Fernández-Blanco would also like to thank the support provided by the NVIDIA Research Grants Program. This work is part of DINTA-UTMACH and RNASA-UDC research groups.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.compag.2017.12.030

References

Abdel-Nour, N., Ngadi, M., Prasher, S., Karimi, Y., 2011. Prediction of egg freshness andalbumen quality using visible/near infrared spectroscopy. Food Bioprocess Technol.4, 731–736.

Aboonajmi, M., Saberi, A., Abbasian Najafabadi, T., Kondo, N., 2015. Quality assessment of poultry egg based on Vis-NIR Spectroscopy and RBF networks. Int. J. Food Prop.2912 150831112742007.

Aboonajmi, M., Setarehdan, S.K., Akram, A., Nishizu, T., Kondo, N., 2014. Prediction of poultry egg freshness using ultrasound. Int. J. Food Prop. 17, 1889–1899.

Abu-Mostafa, Y.S., 1989. The vapnik-chervonenkis dimension: information versus complexity in learning. Neural Comput. 1, 312–317.

Akter, Y., Kasim, A., Omar, H., Sazili, A.Q., 2014. Effect of storage time and temperature on the quality characteristics of chicken eggs. J. Food Agric.

Akyurek, H., Okur, A.A., 2009. Effect of storage time, temperature and hen age on egg quality in free-range layer hens. J. Anim. Vet. Adv. 8, 1953–1958.

Anzanello, M.J., Yamashita, G., Marcelo, M., Fogliatto, F.S., Ortiz, R.S., Mariotti, K.,Ferrão, M.F., 2017. A genetic algorithm-based framework for wavelength selection on sample categorization. Drug Test. Anal.

Armstrong, J.S., Collopy, F., 1992. Error measures for generalizing about forecasting methods: empirical comparisons. Int. J. Forecast. 8, 69–80.

Aroca-Santos, R., Cancilla, J.C., Pariente, E.S., Torrecilla, J.S., 2016. Neural networks applied to characterize blends containing refined and extra virgin olive oils. Talanta 161, 304–308.

Barnes, R.J., Dhanoa, M.S., Lister, S.J., 1989. Standard normal variate transformation and detrending of near-infrared diffuse reflectance spectra. Appl. Spectrosc. AS 43,772–777.

Barton II, F., 2016. Near infrared equipment through the ages and into the future. NIRNews 27, 41.

Bertran, E., Blanco, M., Maspoch, S., Ortiz, M.C., Sánchez, M.S., Sarabia, L.A., 1999.Handling intrinsic non-linearity in near-infrared reflectance spectroscopy. Chemomet. Intellig. Lab. Syst. 49, 215–224.

Blanco, M., Villarroya, I., 2002. NIR spectroscopy: a rapid-response analytical tool. Trends Analyt. Chem. 21, 240–250.

Brereton, R.G., 2015. Pattern recognition in chemometrics. Chemomet. Intellig. Lab. Syst.149, Part B, 90–96.

Cartwright, J., 2016. Technology: smartphone science. Nature 531, 669–671.

Cascardi, A., Micelli, F., Aiello, M.A., 2017. An Artificial Neural Networks model for the prediction of the compressive strength of FRP-confined concrete circular columns. Eng. Struct. 140, 199–208.

Cheng, J.-H., Sun, D.-W., Pu, H., 2016. Combining the genetic algorithm and successive projection algorithm for the selection of feature wavelengths to evaluate exudative characteristics in frozen–thawed fish muscle. Food Chem. 197, 855–863.
Chen,Q., Cai, J., Wan, X., Zhao, J., 2011. Application of linear/non-linear classification algorithms in discrimination of pork storage time using Fourier transform near infrared (FT-NIR) spectroscopy. LWT Food Sci. Technol. 44, 2053–2058.
Das, A., Swedish, T., Wahi, A., Moufarrej, M., Noland, M., Gurry, T., Aranda-Michel, E., Aksel, D., Wagh, S., Sadashivaiah, V., Zhang, X., Raskar, R., 2015. Mobile phone based mini-spectrometer for rapid screening of skin cancer. In: SPIE Sensing Technology + Applications. International Society for Optics and Photonics. p.94820M-94820M-5.
dos Santos, C.A.T., Lopo, M., Páscoa, R.N.M.J., Lopes, J.A., 2013. A review on the applications of portable near-infrared spectrometers in the agro-food industry. Appl.Spectrosc. 67, 1215–1233.
Florkowski, W.J., Prussia, S.E., Shewfelt, R.L., Brueckner, B., 2009. Postharvest Handling: A Systems Approach, Food Science and Technology. Elsevier Science.
Galiş, A.-M., Dale, L.M., Boudry, C., Théwis, A., 2012. The potential use of near-infraredspectroscopy for the quality assessment of eggs and egg products. Sci. Works C Ser.Veterinary Med. 58, 294–307.
Gangying, Z., Yangyang, G., Wei, L., Jie, H., Minmin, W., Guohua, H., 2015. Large yellow croaker (pseudosciaena crocea) storage time detecting method by visible-nir spectroscopy combined with aperiodic stochastic resonance. J. Chin. Inst. Food Sci. Technol.1, 050.
Garcia, H., Filzmoser, P., 2015. Multivariate Statistical Analysis using the R Package Chemometrics. Vienna, Austria.
Gardner, M.W., Dorling, S.R., 1998. Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences. Atmos. Environ. 32,2627–2636.
Giovenzana, V., Beghi, R., Buratti, S., Civelli, R., Guidetti, R., 2014. Monitoring of fresh-cut Valerianella locusta Laterr. shelf life by electronic nose and VIS–NIR spectroscopy. Talanta 120, 368–375.
Goldring, D., Sharon, D., Brodetzki, G., Ruf, A., 2016. Withdrawn Patent as per the Latest Uspto Withdrawn List. US Patent.
Guo, Y., Ni, Y., Kokot, S., 2016. Evaluation of chemical components and properties of the jujube fruit using near infrared spectroscopy and chemometrics. Spectrochim. Acta AMol. Biomol. Spectrosc. 153, 79–86.
Guyon, I., Gunn, S., Nikravesh, M., Zadeh, L.A., 2008. Feature Extraction: Foundation sand Applications, Studies in Fuzziness and Soft Computing. Springer, Berlin, Heidelberg.
Hall, M.A., 1999. Correlation-Based Feature Selection for Machine Learning. The University of Waikato.
Hattori, Y., Otsuka, M., 2017. Modeling of feed-forward control using the partial least squares regression method in the tablet compression process. Int. J. Pharm. 524,407–413.
Haughey, S.A., Galvin-King, P., Malechaux, A., Elliott, C.T., 2014. The use of hand held near-infrared reflectance spectroscopy (NIRS) for the proximate analysis of poultry feed and to detect melamine adulteration of soya bean meal. Anal. Methods 7,181–186.
Haugh, R.R., 1937. The Haugh Unit for Measuring Egg Quality.
Herrera, F., Hervas, C., Otero, J., Sánchez, L., 2004. Un estudio empírico preliminar sobre los tests estadísticos más habituales en el aprendizaje automático. Tendencias de la Mineıa de Datos en España, Red Española de Minería de Datos y Aprendizaje (TIC2002-11124-E), pp. 403–412.
Hyndman, R.J., Koehler, A.B., 2006. Another look at measures of forecast accuracy. Int. J.Forecast.22, 679–688.
Karoui, R., Kemps, B., Bamelis, F., De Ketelaere, B., Decuypere, E., De Baerdemaeker, J.,2006. Methods to evaluate egg freshness in research and industry: a review. Eur. Food Res. Technol. 222, 727–732.
Kimiya, T., Sivertsen, A.H., Heia, K., 2013. VIS/NIR spectroscopy for non-destructive freshness assessment of Atlantic salmon (Salmo salar L.) fillets. J. Food Eng. 116,758–764.
Koch, P., Bischl, B., Flasch, O., Bartz-Beielstein, T., Weihs, C., Konen, W., 2012. Tuningand evolution of support vector kernels. Evol. Intell. 5, 153–170.
Kruse, R., Borgelt, C., Klawonn, F., Moewes, C., Steinbrecher, M., Held, P., 2013. Multi-layer perceptrons. In: Computational Intelligence, Texts in Computer Science. Springer, London, pp. 47–81.
Kuhn, M., Johnson, K., 2013. Applied Predictive Modeling. Springer, New York.
Kumaravelu, C., Gopal, A., 2015. A Review on the Applications of Near-Infrared Spectrometer and Chemometrics for the Agro-food Processing Industries, pp. 8–12.
Lammertyn, J., Peirs, A., De Baerdemaeker, J., Nicolaı, B., 2000. Light penetration properties of NIR radiation in fruit with respect to non-destructive quality assessment. Postharvest Biol. Technol. 18, 121–132.
Leardi, R., Boggia, R., Terrile, M., 1992. Genetic algorithms as a strategy for feature selection. J. Chemom. 6, 267–281.
Lin, H., Zhao, J., Sun, L., Chen, Q., Zhou, F., 2011. Freshness measurement of eggs using near infrared (NIR) spectroscopy and multivariate data analysis. Innov. Food Sci.Emerg. Technol. 12, 182–186.
Lin, H., Zhao, J., Sun, L., kun Bi, X., Cai, J., 2015. Effective variables selection in eggs freshness graphically oriented local multivariate analysis using NIR spectroscopy. In: International Conference on Chemical, Material and Food Engineering. pp. 13–18.
Liu, D., Sun, D.-W., Zeng, X.-A., 2013. Recent advances in wavelength selection techniques for hyperspectral image processing in the food industry. Food BioprocessTechnol. 7, 307–323.
Liu, F., Tang, X., 2015. Fuji apple storage time rapid determination method using Vis/NIR spectroscopy. Bioengineered 6, 166–169.
Liu, Y., Ying, Y., Ouyang, A., Li, Y., 2007. Measurement of internal quality in chicken eggs using visible transmittance spectroscopy technology. Food Control 18, 18–22.
Lorentzen, G., Rotabakk, B.T., Olsen, S.H., Skuland, A.V., Siikavuopio, S.I., 2016. Shelf life of snow crab clusters (Chionoecetes opilio) stored at 0 and 4 C. Food Control 59,454–460.

Martelo-Vidal, M.J., Vázquez, M., 2015. Application of artificial neural networks coupled to UV–VIS–NIR spectroscopy for the rapid quantification of wine compounds inaqueous mixtures. CyTA J. Food 13, 32–39.
Martens, H., Jensen, S.A., Geladi, P., 1983. Multivariate linearity transformation for near-infrared reflectance spectrometry. In: Proceedings of the Nordic Symposium on Applied Statistics. Stokkand Forlag Publishers Stavanger, Norway, pp. 205–234.
Martens, H., Naes, T., 1992. Multivariate Calibration, Wiley Series in Probability and Mathematical Statistics. Wiley.
Mathew, A.O., Olufemi, A.M., Foluke, A., 2016. Relationship of temperature and length of storage on ph of internal contents of chicken table egg in humid tropics. Biotechnol.Anim.
Ma, X., Zhang, Y., Wang, Y., 2015. Performance evaluation of kernel functions based on grid search for support vector regression. In: 2015 IEEE 7th International Conference on Cybernetics and Intelligent Systems (CIS) and IEEE Conference on Robotics, Automation and Mechatronics (RAM). ieeexplore.ieee.org. pp. 283–288.
Mireei, S.A., Amini-Pozveh, S., Nazeri, M., 2017. Selecting optimal wavelengths for detection of insect infested tomatoes based on SIMCA-aided CFS algorithm. Postharvest Biol. Technol. 123, 22–32.
Moros, J., Garrigues, S., de la Guardia, M., 2010. Vibrational spectroscopy provides a green tool for multi-component analysis. Trends Analyt. Chem. 29, 578–591.
Mucherino, A., Papajorgji, P.J., Pardalos, P.M., 2009. Data Mining in Agriculture, Springer Optimization and Its Applications. Springer, New York, New York, NY.
Prajapati, P., Solanki, R., Modi, V., Basuri, T., 2016. A brief review on NIR spectroscopy and its pharmaceutical applications. Int. J. Pharmaceut. Chem. Anal. 3, 117–123.
Pügner, T., Knobbe, J., Grüger, H., 2016. Near-infrared grating spectrometer for mobile phone applications. Appl. Spectrosc. 70, 734–745.
Rateni, G., Dario, P., Cavallo, F., 2017. Smartphone-based food diagnostic technologies: a review. Sensors 17. http://dx.doi.org/10.3390/s17061453.
Refaeilzadeh, P., Tang, L., Liu, H., 2009. Cross-validation. In: Liu, L., Tamer Özsu, M.(Eds.), Encyclopedia of Database Systems. Springer, US, pp. 532–538.
Rinnan, Å., Berg, F., van, D., Engelsen, S.B., 2009. Review of the most common pre-processing techniques for near-infrared spectra. Trends Analyt. Chem. 28,1201–1222.
Rivero, D., Fernández-Blanco, E., Dorado, J., Pazos, A., 2011. Using recurrent ANNs for the detection of epileptic seizures in EEG signals. In: 2011 IEEE Congress of Evolutionary Computation (CEC). pp. 587–592.
Rodríguez-Navarro, A.B., Domínguez-Gasca, N., Muñoz, A., Ortega-Huertas, M., 2013.Change in the chicken eggshell cuticle with hen age and egg freshness. Poult. Sci. 92,3026–3035.
Ruck, D.W., Rogers, S.K., Kabrisky, M., Oxley, M.E., Suter, B.W., 1990. The multilayer perceptron as an approximation to a Bayes optimal discriminant function. IEEE Trans. Neural Netw. 1, 296–298.
Saeys, Y., Inza, I., Larrañaga, P., 2007. A review of feature selection techniques in bioinformatics. Bioinformatics 23, 2507–2517.
Savitzky, A., Golay, M.J.E., 1964. Smoothing and differentiation of data by simplified least squares procedures. Anal. Chem. 36, 1627–1639.
Schulte, H., Brink, G., Gruna, R., Herzog, R., Gruger, H., 2015. Utilization of spectral signatures of food for daily use. In: OCM 2015-Optical Characterization of Materials-Conference Proceedings. KIT Scientific Publishing, pp. 39.
Silversides, F.G., Scott, T.A., 2001. Effect of storage and layer age on quality of eggs from two lines of hens. Poult. Sci. 80, 1240–1245.
Stadelman,W.J., Newkirk, D., Newby, L., 1995. Egg Science and Technology. Stark, E., 1996. Near infrared spectroscopy past and future. Near Infrared Spectrosc.Future Waves 701–713.
Stockwell, D.R.B., Peterson, A.T., 2002. Effects of sample size on accuracy of species distribution models. Ecol. Modell. 148, 1–13.
Suktanarak, S., Teerachaichayut, S., 2017. Non-destructive quality assessment of hens'eggs using hyperspectral images. J. Food Eng. 215, 97–103.
Sun, L., Yuan, L.-M., Cai, J.-R., Lin, H., Zhao, J.-W., 2015. Egg freshness on-line estimation using machine vision and dynamic weighing. Food Anal. Methods 8, 922–928.
Szymańska, E., Gerretzen, J., Engel, J., Geurts, B., Blanchet, L., Buydens, L.M.C., 2015.Chemometrics and qualitative analysis have a vibrant relationship. Trends Analyt.Chem. 69, 34–51.
Teye, E., Huang, X.-Y., Afoakwa, N., 2013. Review on the potential use of near infrared spectroscopy (NIRS) for the measurement of chemical residues in food. Am. J. FoodSci. Technol. 1, 1–8.
Tukey, J.W., 1949. Comparing individual means in the analysis of variance. Biometrics 5,99–114.
Viscarra Rossel, R.A., 2008. ParLeS: software for chemometric analysis of spectroscopic data. Chemomet. Intellig. Lab. Syst. 90, 72–83.
Wold, H., 1985. Partial Least Squares. Encyclopedia of Statistical Sciences.
Wold, S., Sjöström, M., Eriksson, L., 2001. PLS-regression: a basic tool of chemometrics. Chemomet. Intellig. Lab. Syst. 58, 109–130.
Xu, L., Zhou, Y.-P., Tang, L.-J., Wu, H.-L., Jiang, J.-H., Shen, G.-L., Yu, R.-Q., 2008. Ensemble preprocessing of near-infrared (NIR) spectra for multivariate calibration. Anal. Chim. Acta 616, 138–143.
Yongwei, W., Wang, J., Zhou, B., Lu, Q., 2009. Monitoring storage time and quality attribute of egg based on electronic nose. Anal. Chim. Acta 650, 183–188.
Yu, S., Xiao, X., Ding, H., Xu, G., Li, H., Liu, J., 2017. Weighted partial least squares based on the error and variance of the recovery rate in calibration set. Spectrochim. Acta AMol. Biomol. Spectrosc. 183, 138–143.
Zhao, J., Lin, H., Chen, Q., Huang, X., Sun, Z., Zhou, F., 2010. Identification of egg's freshness using NIR and support vector data description. J. Food Eng. 98, 408–414.