POLITECNICO DI MILANO

SCHOOL OF INDUSTRIAL AND INFORMATION ENGINEERING

MASTER THESIS IN AUTOMATION AND CONTROL ENGINEERING

Hyperspectral Imaging and Machine Learning for Automatic Food Quality Inspection

Academic Supervisor: Prof. Marco TARABINI

Academic Co-Supervisor: Dott. Davide Maria FABRIS

Master Thesis by
Manuel Di Raimondo – 10390126

Academic Year 2020/21

Abstract

This master thesis investigates the coupled application of Machine Learning algorithms and Hyperspectral Imaging, with a focus on food quality analysis and control. Hyperspectral Imaging is a powerful instrument as it enables very detailed information to be acquired, but the potential applicability of this technology is mostly limited to remote sensing applications. Relying on Machine Learning algorithms, a fair compromise between the reduction of dimension of data-cubes from hyperspectral images and the loss of as minimal data was found. In particular, the Machine Learning algorithms on which the project focused on are Principal Component Analysis (PCA), Partial Least Squares-Discriminant Analysis (PLS-DA) and Spectral Angle Mapper (SAM).

This thesis considered the application of Hyperspectral Imaging and Machine Learning for the quality analysis of baked eggplants, according to different cooking conditions. Data was acquired by means of a push-broom Hyperspectral camera produced by Specim, FX-17. The gathered results testify the applicability of Hyperspectral Imaging (HSI) together with Machine Learning methods of dimensionality reduction and classification. Results were validated on test images. Confusion matrices and several statistical parameters were adopted in order to evaluate the classification performances. Among the considered algorithms, the most promising resulted being the Partial Least Squares-Discriminant Analysis.

Keywords: Machine Learning, Hyperspectral Image, Food Sorting

Table of Contents

List of Figures

1. Introduction

This master thesis has the purpose to investigate the coupled application of Machine Learning algorithms and Hyperspectral Imaging, with a focus on food quality analysis and sorting. Hyperspectral Imaging applications are mostly limited to remote sensing applications. The main relevant examples have been highlighted but the potential of Hyperspectral Imaging extends far beyond. The subject of this study has been parts of eggplants in order to classify several classes of them. In particular, the classes on which the analysis was based are those of the various cooking processes, together with references to fresh product and skin.

Several Machine Learning algorithms have been used. The fair compromise searched is between the reduction of dimension of data-cubes from hyperspectral images and the loss of as minimal data.

1.1. State of the Technique

This sub-chapter explores the state of the art of the principal technology on which the master thesis is focused: Food Sorting.

Food sorting is one of the most important phases in the horticultural, and all food, industry. When dealing with natural production processes, in fact, they are influenced by a wide range of factors and these must be analyzed in order to ensure quality and homogeneity. In food sorting there are primarily three categories to consider and to investigate: the external defects evaluation, the internal quality and maturity assessment, and food safety detection [1].

1.1.1. Food Sorting Technologies

Food is the basis of life for every human being and the adoption of a good product is linked to a better quality of life [2]. As already mentioned, horticultural product qualities are influenced by several factors. Besides, the parameters that define these products as food may be subject to evolutions during time leading to unavoidable deterioration. The resulting choice is to inspect the batches of products before they reach the consumer.

This kind of quality control started with the focus only on the appearance. Human operators used color, size, shape, and surface texture as discriminant during food sorting. Despite nowadays human operators are still widely used, this does not deny the fact that simply visual inspection with human eyes has its limitations. First, in terms of speed, volume, and subjectivity it is a technique that does not guarantee an optimal evaluation.

In this regard, a big step has been taken by replacing the human eye with the camera. The camera, in association with a computer as a brain and an actuation system as hands, made it to the first automatic sorting lines based on machine vision.

The traditional machine vision imitates the external quality inspection capturing images using three filters centered at red, green, and blue (RGB) wavelengths [2]. Unfortunately, also this first example of machine vision has its limitation. In fact, using only these three spectra as information, some unobvious defects, whose colors and textures can be very similar to the sound skin, are still hard or impossible to be detected.

To improve once again these performances and to get closer to the actual quality of the human eye, the subsequence step has been to increase the number of bands analyzed up to fifteen, obtaining the so-called Multi-Spectral Imaging (MSI). This type of choice has permitted to choose the most efficient wavelengths to detect detailed quality parameters. On the other hand, the disadvantages occur when different types of products are under observation. The only solution in MSI is to acquire different times the same images, each time with checks, new calibrations and debugging operations. The last step in nondestructive inspection is

Hyperspectral Imaging (HSI), overcoming the number of fifteen bands. This kind of scenario has permitted great strides not only in the external quality analysis but also in internal quality and maturity assessment thanks to the spectral analysis. In parallel with the growth of the number of bands used and increasingly larger hyperspectral data, there is an inexorable rising of the computational costs to significant limitations.

One of the leader companies in the food sorting sector is TOMRA. From 1970s the Norwegian company sells its machine in Europe and USA. In 2020 has reached the goal of over 100,000 installations in the entire worldwide. There are different technologies that lead TOMRA to be one of the reference points of the industrial automation network. FLUIDCOOL® LED technology, with the point scanning principle, causes a reduction till 70% in power consumption and enables steady and
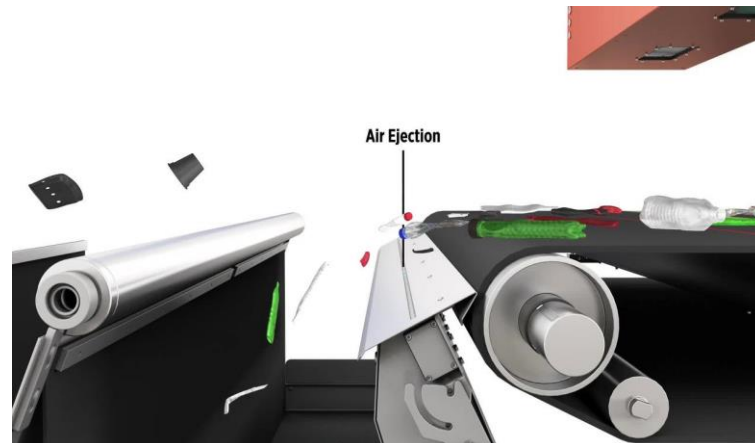
Figure 1 - LOD Operating in collaboration with air dejection

reproducible sorting results, mostly independent of ambient temperature changes. The Laser Object Detection (LOD) implemented ad hoc by TOMRA equipment helped in detection of waste products like black plastics and glass that cannot be seen from Near-InfraRed (NIR) sensors.

At the base of their installations there is their own deep learning-based algorithm. By collecting large amounts of data, the software detects objects with a specific form of texture supporting full belt-speed up to 3m/sec of the process chain. An Italian presence in the food sorting sector is Raytec Vision Spa. Besides it was created only in early 2000s, it works in close contact with Spanish, American, French companies and, recently, also with the Emirate States, Russia, and Algeria. The strength of the

Figure 2 - An installation dedicated to the control of peeled and unpeeled potatoes thanks to NIR technology

Raytec which in 2018 led it to reach the goal 28 million is the offer of a wide range of X-rays machines for solutions tailored to specific tasks. Recently, together with the Polytechnic of Milan, it has been developing installations involving the use of hyperspectral technologies.

1.2. State of the Art

The sub-chapter on the State of the Art focuses on the studies behind food sorting and, in general, automated analysis of objects: Hyperspectral Imaging and Machine Learning.

HSI led the computer vision to take many steps forward. The latter, although last twenty years it has reached unprecedented capabilities to measure, inspect, grade and sort in agricultural and food products, has shown some drawbacks in certain industrial applications. Based on reflectance, transmittance, absorbance, or scatter of poly-chromatic or monochromatic radiation, the fields where these optical properties have been studied are ultraviolet (UV), visible (VIS) and near-infrared (NIR) regions [4].

Hyperspectral Imaging is, as matter of fact, a combination of VIS and NIR spectroscopic techniques. It produces better results than simple Imaging or Spectroscopy. Imaging is good if there is the goal of acquiring spatial information, but it fails in learning spectral and multi-constituent information. On the other hand, Spectroscopy is focused in these two options, but it does not permit to catch spatial information.

HSI succeeds in getting both spatial and spectral information. The combination of these two properties helps in the success of the Chemical mapping. Chemical mapping allows to build images and labels of different entities in a unique sample and to show the gradients and spatial distributions of chemical compositions with a quickly access. HSI alone is not sufficient for an automated non-destructive quality and safety inspection, it needs a technology with classification and/or detection capabilities.

The introduction of Machine Learning is strictly connected with the development of computers. The opportunity of using calculators recently allowed the complexity of algorithms to be increased to solve problems of equal merit and, simultaneously, to store and work with large amount of data: impossible tasks for human capacities.

The term "learning" is very interesting because another goal of machine is to acquire experience through data. Experience permits to resolve very complex processes optimizing the path to follow. Therefore, Machine Learning has the task of replicating everything that is human experience. All this involves countless applications related to the artificial intelligence: recognizing speeches and handwriting, analyzing past data to improve retails, optimizing resources in tasks with robots just as examples.

1.2.1. Hyperspectral Imaging

A hyperspectral image consists in over and over congruent images representing of composed vector pixels (voxels) at different wavelength bands [4]. It contains two-dimension spatial information corresponding to m rows and n columns plus a third
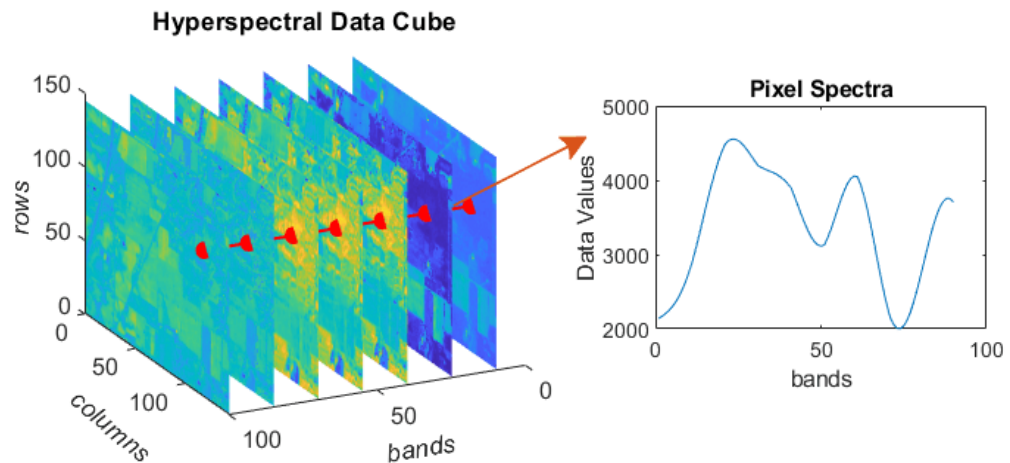
Figure 3 - Representation of Hypercube and of the Spectral Signature of a pixel

dimension that explores the spectral information through k wavelengths. This three-dimensional data (m x n x k) is defined as hyperspectral cube, or hypercube.

It is easy to imagine a hyperspectral cube as the same image taken with different filters: in each there is a representation of the spectral sensitivity in that band.

Consequently, considering the same spatial pixel and analyzing it along the bands, a spectral signature will be obtained, hence the relative reflectance with the respect to the wavelengths.

The process of acquiring a hypercube starts when a light source illuminates the elements to frame. The camera lens collects the radiation from the objects crossing first the spectrograph thanks to a 2D detector array named Charge-Coupled Device (CCD). The prismatic elements inside spectrograph permit to acquire different 2-D images of the same spatial dimension with a continuous spectrum.

As already mentioned, the introduction of HSI has had its pros and cons.

The most logical option is to reduce the dimension of data to limit the computation costs, even if it seems nonsense given the road that led to the development to HSI itself. The reduction approach is useful as it makes a kind of summary of the given hypercube with the most important information.

There are three main approaches that can be used for acquiring HSI images.

The first one is called point scanning and this method consists in the acquisition of a single point along the two spatial dimensions by moving the detector or the scanning sample. At the same time the spectrophotometer has the function to collect the spectrum of that pixel.

The line-scanning method can be considered the evolution of the previous one: in line-scanning there is a parallel acquisition of a segment of spatial and spectral information. The result is a two-dimensional image with one spatial dimension and one spectral dimension. It is necessary to acquire all the samples in the direction of motion to complete the hypercube.

The last method is called area-scanning. It differs in that it acquires the entire 2D spatial information at once. There is not any movement between samples or detector and the hypercube is given from the different images at different wavelengths.

For the nature of product movement applied in food sorting the method preferred is line-scanning. Still according to Da-Wen Sun [4] HSI systems need to be
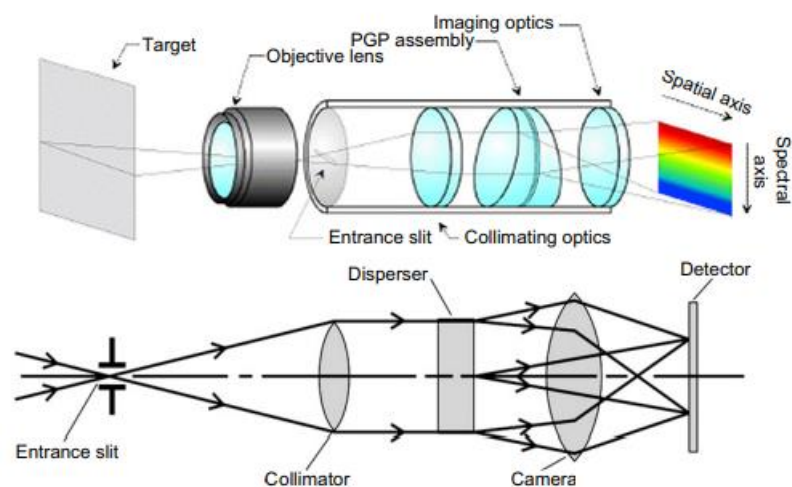
Figure 4 - Working principle of a prism-grating-prism (PGP) spectrograph

associated to some software to gain higher performances in acquisition, controlling and analyses.

Fundamental to these results is the use of a hyperspectral camera. At this point it is possible to extract data from different Region of Interests (ROIs) that present different quality features starting first with a calibration of dark and white hyperspectral images and then preprocessing data in a suitable way according to purposes,

One of the first applications of HSI systems were related to aerospace research. In fact, itself "hyperspectral imaging" derives from the development of NASA's Airborne Imaging Spectrometer (AIS). The goal is to analyze portions of land areas for mining and oil industries. From 1980s researchers have developed the application of HSI in different fields. Some examples must be mentioned: military industry uses hyperspectral technology for surveillance purposes, especially in facial recognition and scanning of areas.

1.2.2. Machine Learning

The most famous definition of Machine Learning comes from Tom M. Mitchell as follows: *"A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P if its performance at tasks in T, as measured by P, improves with experience E"* [5].

In this way, the computer (the machine) has the task to automatically extract the algorithm for its task [6]. There are three types of models: predictive to make predictions in the future, descriptive as they gain knowledge from data, or both.

One of the main advantages of using computer technology is to work with large amounts of data, optimizing time and resourcing costs. This kind of application of machine learning is called data mining. Starting from raw data, the computer analysis leads to a simple model with high predictive accuracy. The general approach suggests that data mining has rich fields of interest: finance banking, medicine, manufacturing are only few examples.

Machine learning, however, is much more than simple database problem. It is a part of artificial intelligence since it has the ability to learn from a changing environment,

it finds solutions to problem which every day maybe human eyes process, but minds may not easily express.

Regression is another kind of application of Machine Learning. It consists in predicting a continuous or a discrete outcome basing the analysis on one or more predictor variables. A practical example of regression can be the prediction of the price of a house given the features of house like size, position, nearby facilities etc. and the applications of this process are feasible to almost any context.

As summery, Machine Learning is programming computers in building mathematical models using the theory of statistics.

In Machine Learning there are unsupervised algorithms, supervised algorithms, and semi-supervised algorithms [7].

The supervised methods are based on the determination of discriminant criteria due to the pre-knowledge of sample categories. Examples of this kind of classification are Regression, Support Vector Machine (SVM), Decision Tree, Partial Least Squares–Discriminant Analysis (PLS-DA), Artificial Neural Network and Spectral Angle Mapper. Otherwise, unsupervised classification focuses on one aspect: clustering with no prior knowledge. Based on the different clustering strategies, the main ones to name are Principal Component Analysis (PCA) and K-Means.

Halfway, semi-supervised method works with both labeled and unlabeled data to train the classifier. It has the main disadvantage that its accuracy depends all on the number of training datasets and high timing and computational costs are needed.

1.3. Existing Applications

Approaching these studies, the implications in agricultural sector of HSI are countless.

Spectrometers and Machine Learning have been able to realize warning systems for disease outbreaks [8]. Other agriculture applications are about detecting chemical composition of plants and animal meat, to monitor and investigate on nutrients and proteins.

Our case, finally, focuses on food sorting processing. The HSI combined with intelligent software can do sorting operations in a non-destructive way.

The earliest application of HSI in surface detection in fruit were Mehl et al [9].

They performed their studies around 153 samples and three typologies of apples: Golden Delicious, Red Delicious, and Gala. The goal was to isolate those spectral bands which should correspond to scabs, fungal, soil contaminations, and bruises. Using the Principal Component Analysis (PCA) they first made a hyperspectral analysis to find spectral features of apples for the specific selection of filters to design the multispectral imaging system. The hyperspectral analysis led to the determination of three spectral bands capable of separating normal samples from contaminated samples. The implementation with the multispectral imaging system served to enable a rapid detection of apple contaminations. Good results have been achieved for the first two kind of apples: 95% of correct detections for Gala apples and 85% for Golden Delicious. Limited results for Red Delicious (75%).

In 2014 Mascetti et al [10] realized a multi-class PLS-DA classifier for hazelnuts classification. They used four quality classes (Class Extra, Class I, Class II and Waste) and two different exposure times (5 and 8 ms) to optimize the improvement of intensity and uniformity of the light along the field of view of the camera. The PLS-DA model, with those terms, has managed to ensure over 80% of sensitivity and

selectivity in all the classes and proved that themself spatial dimensions can be used to improve the accuracy of the classification.

The interests of this master thesis will focus on the classification of vegetables according to various firings, but, to date, no studies have been found that elaborate on it.

1.4. Scheme of the Thesis

The following thesis is articulated in 4 main chapters: Setup, Methods, Results and Conclusion. All of them are fundamental to fully illustrate all the aspects, functions, and results that we have encountered on this route.

More deeply, the first main chapter, called Setup, has inside on it three sub-chapters in such a way as to obtain a very high level of representation of data according to our algorithms: Experimental Setup, Dataset Samples and Software Setup.

Experimental Setup gives a description of considerations taken about the conveyor belt, light conditions and the HSI Camera used to acquire data. In the Dataset Samples section, you can find the reference classes to be classified and in which way dataset has been decided to be considered. Software Setup, instead, contains a brief introduction to functions and tools used during programming with MATLAB.

The second chapter is Methods. It starts from algorithms involving into programs, it passes throw the pre-processing of dataset and it ends by showing how different outputs have been evaluated among themselves.

The Result Chapter, together with the Conclusion part, shows performances of Machine Learning methods used and which are more confidential with respect to the dataset considered.

2. Setup

2.1. Experimental Setup

The experimental data were acquired in the ImageS laboratory through a specific setup chosen to better fit to this master thesis. Two were the technologies involved in the experiment: the Specim FX17 Camera and the Specim LabScanner 20x20. The former was harnessed to acquire the large amount of data required, the latter was used as a supporting equipment. Since the HSI camera was a line scanning camera, it was required to synchronize the conveyor speed of the LabScanner with the exposure time of the camera.

2.1.1. Specim FX17 Camera

In this kind of food processing thesis, the selection of the camera is fundamental: the analysis of spectra of object pixels is strictly connected with very performing instrumentation and there are not so many companies that produce industrial HSI camera. We had the opportunity to work with the equipment from one of the leading companies in the sector in Finland: Specim.

In the world of imaging systems, Specim has for years been one of the benchmarks for remote sensing application with its HSI cameras. The company has different products for a wide range of implementation. For example, their Thermal Infrared Hyperspectral Camera is the first hyperspectral camera used efficiently for outdoor surveillance and for UAV (Unmanned Aerial Vehicle) applications without any kind of external light source, whether sun or moon.

Figure 5 - The logo of Specim Company
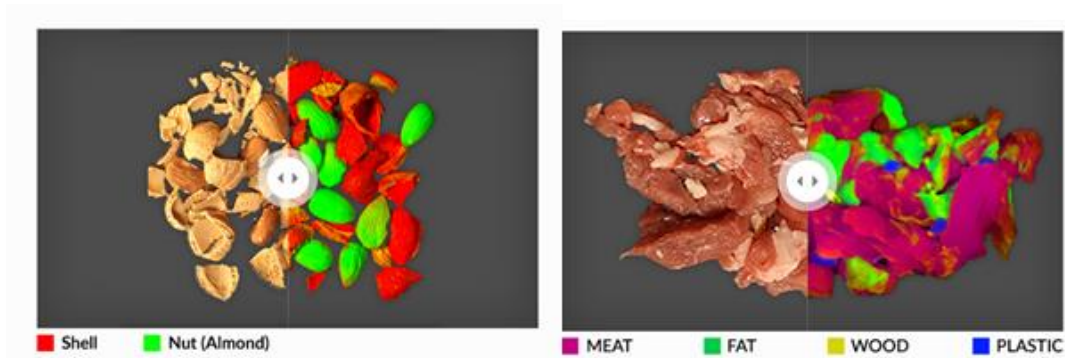
Figure 6 - Two different approaches in food industrial sorting

Particularly in the last twenty-five years, Specim has improved its business in producing HSI camera specialized for industrial sorting applications. The FX17 Camera is able to scan, inspect and acquire chemical footprints of a sample in a good and user-friendly way.

The camera operates in the near-infrared region, from 950 to 1700 nm. This ensures the possibility to sense a wider spectrum compared to RGB cameras and human eyes. The Specim camera can be applied to investigate the chemical composition of a target, making a superior introspection in the sorting sector. The linear resolution is of 640 pixels, with an image speed of 527 fps for GigE version and 670 fps for CameraLink version. This specification is suitable with this master thesis.

Figure 7 - Specim FX17, the Hyperspectral camera supplied

Camera focus Configuration

Some steps are necessary to properly setup the camera. First, the camera must be focused on the working plane. This process begins with the opening of the aperture. After that, the focus is set by rotating the focusing ring, since the FX17 is a manual focus camera. With the targets on the working table, it is possible to connect the camera to a PC and to have a preview of the picture in real time. Other adjustments in terms of exposure and focus are possible by setting camera parameters.

With the open wide aperture, the focusing plane would be at a unique distance from the lens. For this reason, to increase the depth of focus, the aperture opening is reduced. As result, the image results darker since a lower amount of light is reaching the sensor. This effect is counterbalanced by increasing the exposure time. The longer time the shutter is open, the more light will enter inside the lens. It is not recommended to fail the setting of the focus of a Hyperspectral Camera. There are several consequences. The principal one is a resulting image with blurry spatial features and wrong spectral features, too [11]. The result from this procedure is a sharp, focused image ready to be used for data acquisition.

In the Figure 9 it is showed the image acquisition setting for the master thesis: the FX17 Camera is positioned on the top the sliced eggplants that are on the conveyor. In the conveyor, it can be seen the white reference at the edge of the conveyor. This bar is needed for the dark current calibration that will be explained in the dedicated sub-chapter.

Figure 9 - Dark current calibration and image acquisition setting
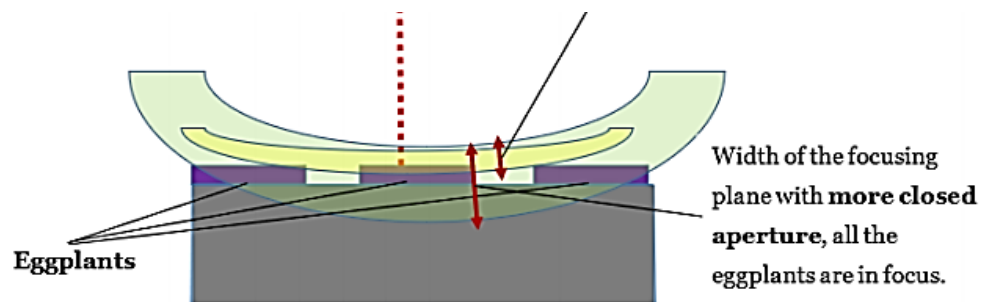
Figure 8 - The above scheme shows the importance of a correct set-up in terms of aperture and focusing distance.

2.1.2. Supporting System Tools

As made for the camera point, it has been chosen the Specim's LabScanner 40x20 as main supporting system for our setup. It is a small scanner system, good for laboratory use. LabScanner has 400 x 200 mm sample tray, a mount of a camera, halogen illumination and optional camera height adjustment.

The Figure 10 shows how the conveyor is mounted right on the bottom with respect to the camera and the option to move the latter up or down.

The scanning speed range can vary from 0.1 mm/s to 99 mm/s, but the choice of the value is non arbitrary. This one depends on the camera parameters, one among all is the exposure time. A different setting of speed property may affect the spatial data quality, causing a worsening.

Figure 10 - Specim LabScanner 400 x 200 made by Specim Spectral Imaging Oy Ltd

Conveyor Belt Speed Configuration
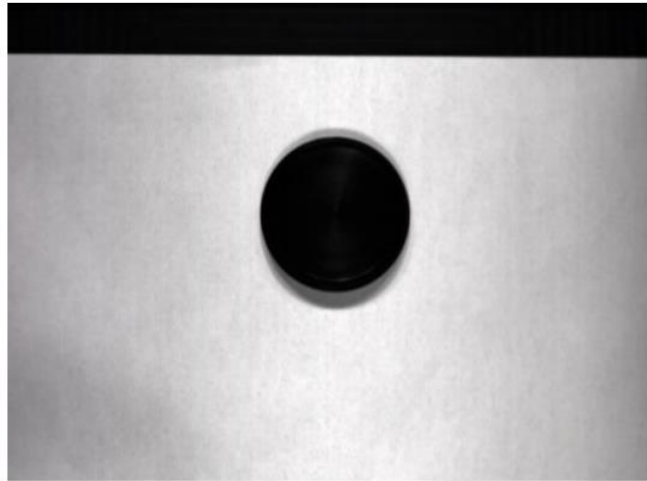
Figure 11 - Elongated circle sample due to the too low conveyor speed

It is necessary to make an in-depth study about the setting of the conveyor speed. In fact, the FX17 camera works by capturing line by line the spatial and spectral features of the samples, so what is set is the relative motion between camera and samples. Hence it is simple to point out the importance of fixing the camera to the ground and the correct moving of samples with respect to the ground.

The setting of the relative speed is performed by putting a circle-shape sample above the conveyor belt that is continuously moving when the data acquisition process starts. If the relative speed is right, then the resulting picture will not change the shape of the samples.
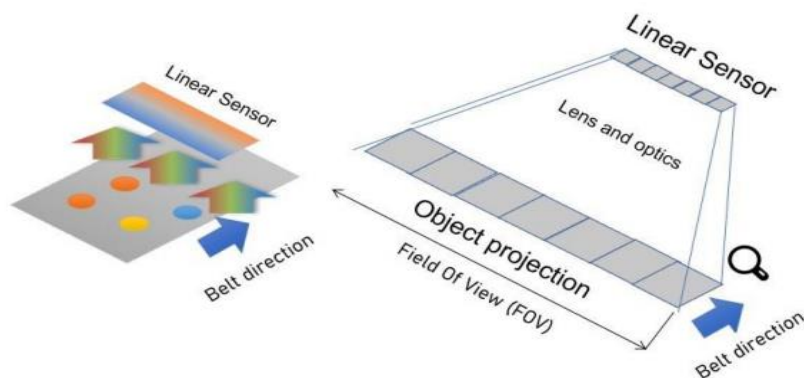
Figure 12 - Image acquisition of line-scanning camera

Other issues may come with a non-matching exposure time with the speed of the samples. A higher speed than necessary will correspond to a wider resulting picture in the orthogonal direction of motion. On the other hand, setting the speed too low will give as output a wider picture in the moving direction of motion.

Tuning the speed is possible if we consider three parameters: the sampling rate, the exposition time and processing time. The sampling rate (f_{rate}) is the frequency whereby the sensor of the HSI camera must work, the exposition time ($t_{exposure}$) shows how much time pixels are exposed to the light and the processing time ($t_{processing}$) contains how long it needs to transmit and elaborate data.

The formula is the following:

$$\frac{1}{f_{rate}} \geq t_{exposure} + t_{processing} \tag{1}$$

2.2. Dataset Samples

Using the techniques explained above, the dataset was acquired by taking several classes of eggplants. Different types of experiments were performed with data. From those experiments, some Machine Learning algorithms have been performed to test predictions. The choice of dataset started considering various eggplants (*Solanum melongena*) with a similar shape. The first hyperspectral data-cube was the entire

Figure 13 - The false colored image of the entire eggplant

eggplant with no damages. The importance of this image is strictly connected to the recognition of the eggplant skin and of the background.

The same eggplant sample was cut into pieces to have another kind of data-cube. In fact, the second hyperspectral image is a composition of sections of the same eggplant. The first section is a piece of the eggplant without changes. Then, two pieces were selected and were baked for 10 and 20 minutes, respectively. One last section was burnt in the middle using a heat lamp. These samples were finally on finally the conveyor belt and acquired with the FX17 camera. After all the data-cubes were acquired, the two images were joined together as shown in Figure 15. This was performed so to have a unique hyperspectral image to be analyzed and to speed up the data processing. This last hyperspectral data-cube was rather memory demanding, approximately 3.5 GB. Two further steps resulted being fundamental in order to work with a lower amount of data. The first was to choose a Region of Interest (ROI) for each class in the dataset. The second step was to create a train

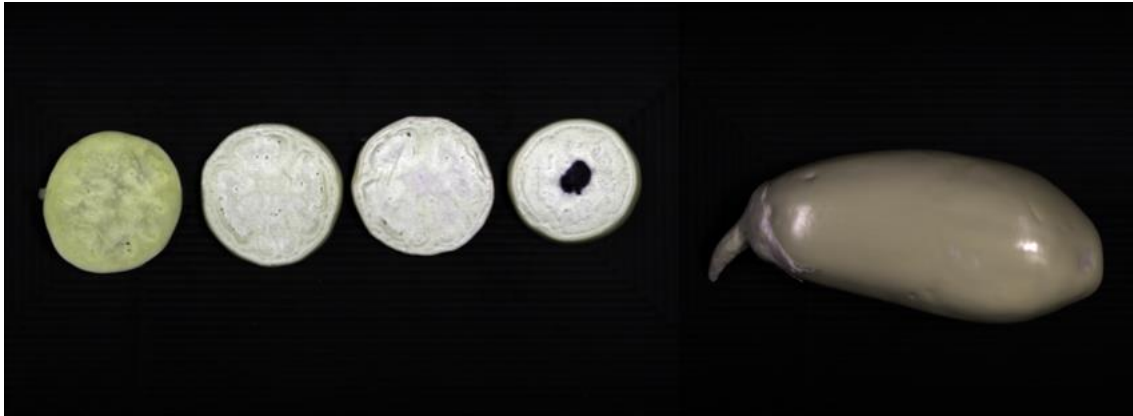**Figure 14 – The false colored image of the different classes of the cut eggplants**

Figure 15 - The false colored image of resulting new dataset

image to calibrate algorithms and a test image to perform predictions from the dataset. The train image groups all odds pixels and the test one all the even pixels.

3. Methods

In Chapter 3 the selected techniques are going to be introduced. Namely, the main sections are Data pre-processing, Algorithms, Metrics and Experiments. The first reports the manipulation applied on the hyperspectral images. The second is devoted to the algorithms' description. The third highlights the metrics that were chosen in order to objectively compare the algorithms' performances. The latter reports the experiments that were carried out.

3.1. Data pre-processing

Dataset pre-processing is an essential step before training and testing the algorithms on the dataset. Hyperspectral images contain a considerable amount of additional information with respect to simple images. This statement leads to a considerable increase in the amount of time invested in analyzing examples, their characteristics, and their differences in nature or quality. Without pre-processing it would be difficult, in some cases impossible, to perform functions with MATLAB. This also due to the high dimension of images and due to signal noise.

3.1.1. Dark Current Calibration

In general, when Hyperspectral Camera acquires data, there is always electronic current flowing in the detector arrays even when there is no light shining on it [12]. This one, called dark current, is generated by thermally induced electron hole pairs. To minimize it, a calibration of the output of the camera is required.

The calibration consists in the normalization of the acquired image with the white reference value and the black reference value. The white reference value was obtained by framing a white bar put on top of our conveyor belt. On the other hand, the black reference value is the result of the data acquisition in the dark mode, i.e., when the shutter is close.

To obtain the normalized data and to have the dark current calibration, the following formula was applied to each pixel of the data cube.

$$pixel_{norm} = \frac{pixel_{raw} - ref_{dark}}{ref_{white} - ref_{dark}} \qquad (2)$$

3.2. Algorithms

This section contains the algorithms of Machine Learning adopted to achieve a reduction of dimensionality of the given dataset and the prediction of the classes. As already discussed, basing on the chosen approach, a first distinction of methods is made for supervised classification and unsupervised classification. The discriminant is the presence or not of labeled data. The Principal Component Analysis algorithm was selected as an unsupervised classification method. Whereas the PLS-DA and the SAM were chosen as supervised ones. The codes used in this master thesis are completely written in MATLAB because it is one of the most used programming and numerical calculation platform in university with its technical processing tasks, statistics, and Machine Learning. In addition, it has hundreds and hundreds of libraries optimized and the codes are compiled fast.

3.2.1. Hypertools

Hypertools3 is a Graphical User-friendly Interface (GUI) for the analysis of multispectral and hyperspectral images. It has powerful visualization and interaction tools implemented [13]. It integrates a lot of features. For example, the analysis of spectral and spatial matrices or images with different preprocessing methods (binning, cropping, spectral range selection, etc.) and data mining with the techniques of exploratory, clustering, and spectral unmixing. In our case, we used this tool in the classification with a Machine Learning algorithm: the Partial Least Squares-Discrimination Analysis.

CountEndmembersHFC and Fippi

Around the concept of endmembers, it was used three functions to determine the ones with the richest information and to extract them from the raw data.

- numEndmembers = countEndmembersHFC (inputData, Name, Value)

    From an inputData, the function returns the number of endmembers present in the hypercube by using the Noise-Whitened Harsanyi-Farrand-Chang (NWHFC) method [14]. Specifying Name with 'PFA' and Value with '10^{-7}', there were considered only those endmembers with the probability of false alarm of that entity.

- endmembers = fippi (inputData, numEndmembers)

    The function fippi extracts endmember signatures from hyperspectral data inputData by using the fast iterative pixel purity index (FIPPI) algorithm. numEndmembers is the number of endmember signatures to be extracted using the fippi algorithm.

3.2.2. Principal Component Analysis

Principal Component Analysis (PCA) is a method within the linear transformation problems that is widely used in various fields, especially for feature extraction and dimensionality reduction. PCA allows one to find the directions of maximum variance in high-dimensional data. These can then be used to project the data onto a new subspace, with equal or lower dimensions than the original one. By means of mathematical projection, the original data set, which may have involved many variables, could be represented by only a few variables. The output of PCA are precisely these principal components, the number of which is lower or equal to the number of original variables.

Ultimately, the Principal Component Analysis allows to identify patterns in the data, based on the correlation between features. In fact, it is often the case that a study of the smaller data set allows the user to identify trends, patterns, and outliers in the data. This is generally done much more easily than it would have been possible without performing principal component analysis.

The principal components have some useful properties which are listed below:
- they are essentially linear combinations of the original variables, the vector of weights in this combination is the eigenvector found which in turn satisfies the least squares principle.
- they are orthogonal to each other.
- the variation present between them decreases as we move from the first component to the last, hence the importance.

The PCA steps are the following:
1. Matrix standardization (or subtracting the average of columns or dimensions from the original matrix).
2. After subtracting the mean what is done is the creation of a covariance/correlation matrix

3. Calculate the eigenvalues and eigenvectors, the eigenvectors only define the directions of the new axis as they all have the same unit length 1.
4. Form the principal components
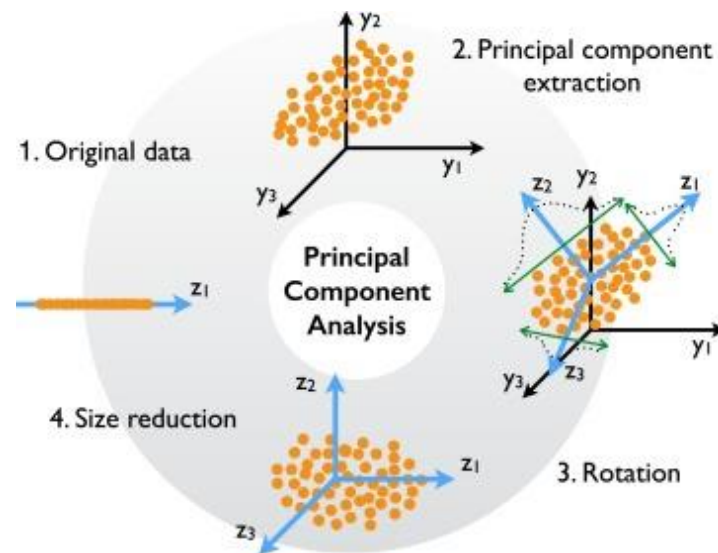
Figure 16 - Schematic overview of principal component reduction for correlated variables in the state-space.

$$CP = W^T \; x \; B^T \qquad\qquad (3)$$

Where:

- CP is the matrix formed by the principal components.
- W^T is the matrix we formed using the eigenvectors we chose to keep from the previous step.
- B^T is the rescaled/standardized version (in step 1) of the original dataset.

The "^{T}" in the superscript indicates the transpose of a matrix that is formed by swapping rows into columns and vice versa.

Figure 17 - Structure of a hypercube. On the left: depiction of a hypercube obtained for the plastics. Several pixels belonging to different plastics plus the background are selected and, on the right, their corresponding classification output.

3.2.3. Partial Least Squares-Discriminant Analysis

Partial Least Squares-Discriminant Analysis (PLS-DA) is a very versatile algorithm on a statistical basis. This is used for predicting and describing patterns, as well as for discriminant selection of variables. Versatility is good when dealing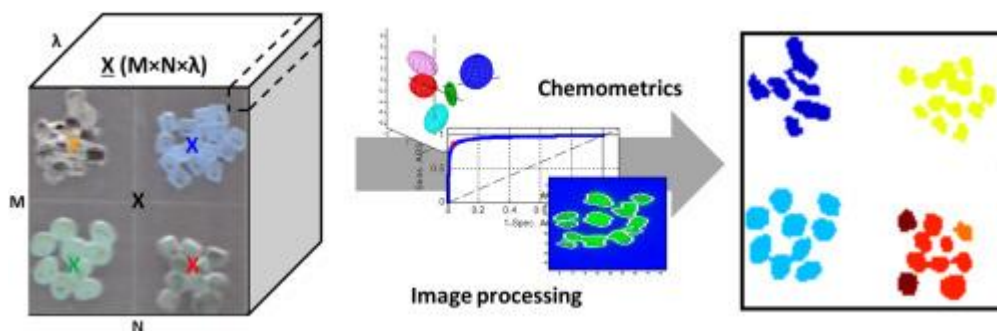 with large datasets, but at the same time the user must motivate several parameters before achieving reliable and valid results.

In turn, PLS-DA is an evolution of PLS-Regression where the expected output is a discrete value. In our case what the algorithm estimates are matrices (W, T, P and Q) and then obtains, when a prediction is made, a categorical value.

The prediction ability of PLS-DA is highly dependent on the number (k) of latent vectors (LV) used in the model. The optimal value of k is determined by predicting the results for independent groups of samples (test sets) for different values of k. By determining the value of k for which the highest percentage of correct classification was found in the test set, the optimal k is chosen. Having a large hyperspectral image at disposal and choosing a very low k would cut out several elements essential for good classification. On the contrary, choosing more elements than necessary would

lead to a result that might entail substantially equivalent classification but certainly heavier data to be carried.

Modelling with PLS-DA permits to classify each chosen class, it also calculates the prediction probability and a classification threshold. These are calculated by taking the values predicted by the PLS-DA model, applying a normal distribution, and then obtaining the probability of observing a given value (Bayesian method).

The method determines the percentage of correct classification and the weight of each species in each LV. The modelling efficiency is indicated by the sensitivity and specificity of the parameters. Sensitivity is the percentage of species in the category accepted by the class model. Specificity is the percentage of species from categories different from the model, therefore rejected by the class model. In PLS-DA, the model includes a calibration (training) phase and a test phase. For both phases, the percentage of correct classification can be determined by calculating the residual errors.

3.2.4. Spectral Angle Mapper

Spectral Angle Mapper (SAM) is a spectral classification algorithm that uses an n-dimensional angle to match pixels with reference spectra. This algorithm determines the spectral similarity between two spectra. This is done by calculating the angle between the spectra and treating them as vectors in a space with dimension equal to the number of bands. Smaller angles represent a better match with the reference spectrum. The calculation of the angle α of a given pixel is based on this trigonometric function:

$$\alpha = cos^{-1}\left(\frac{\sum_{i=1}^{C} t_i r_i}{\sqrt{\sum_{i=1}^{C} t_i^2}\ \sqrt{\sum_{i=1}^{C} r_i^2}}\right) \tag{4}$$
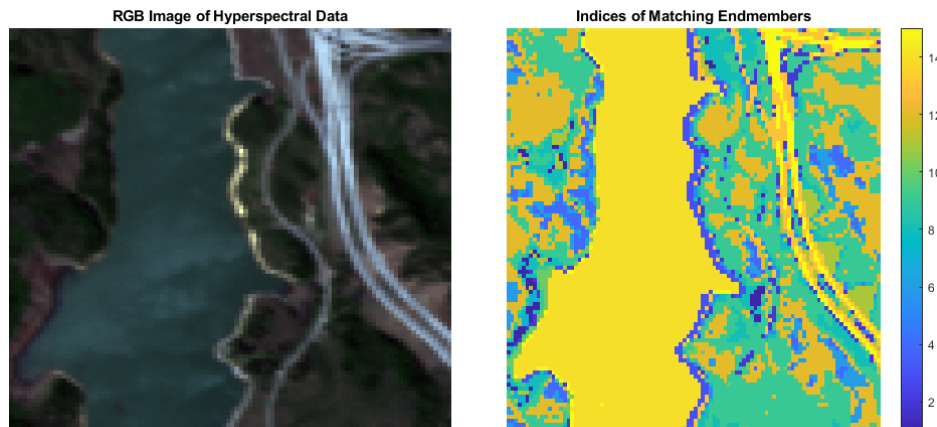
Figure 18 - On the left: a geographical area used as input. On the left: how SAM algorithm assigns a class using the reference spectra

Where:

- t_i is the i-th test spectrum.
- r_i is the i-th reference spectrum.
- C the length of vectors of spectra.

3.3. Metrics

To have objective feedbacks on performances and not to rely solely on graphical evidence, the measure of the Confusion Matrix was introduced. It is a commonly considered method both in binary and multiclass classification problems.

Confusion matrices represent counts from predicted and expected values. There are different outputs to represent the various combination of labels: True Positive (TP), False Positive (FP), True Negative (TN), False Negative (FN).

The term TP represents the positive examples that were classified as positive. FP are the negative examples that are classified as positive. Vice versa, FN are the positive examples that are classified as negative. TN are the negative elements classified correctly.

These four values are strictly connected to other three parameters: Recall, Precision and F1 Score.

The precision is the ratio of correctly predicted positive observations with respect to the total predictions. The considered algorithm predicts the tested image better as this parameter is closer to the value of one. The formula is the following:

$$Precision = \frac{TP}{TP + FP}$$

(5)

The recall parameter provides an overview about the sensitivity of predictions. It is calculated as the ratio among the number of correctly predicted positive observations and all the occurrences of the expected class.

Figure 19 - The corresponding table of confusion

$$Recall = \frac{TP}{TP + FN} \tag{6}$$

The F1 Score is the weighted average of Precision and Recall.

Despite this score is not so easy to understand with respect to Accuracy, it is very useful in our case. First, Accuracy works best if FP and FN have similar costs (in the next chapter we will be able to observe that it is not our case), else it is better to consider both Precision and Recall together. Related to this reason, it is preferred to use F1 Score because of the uneven class distribution.
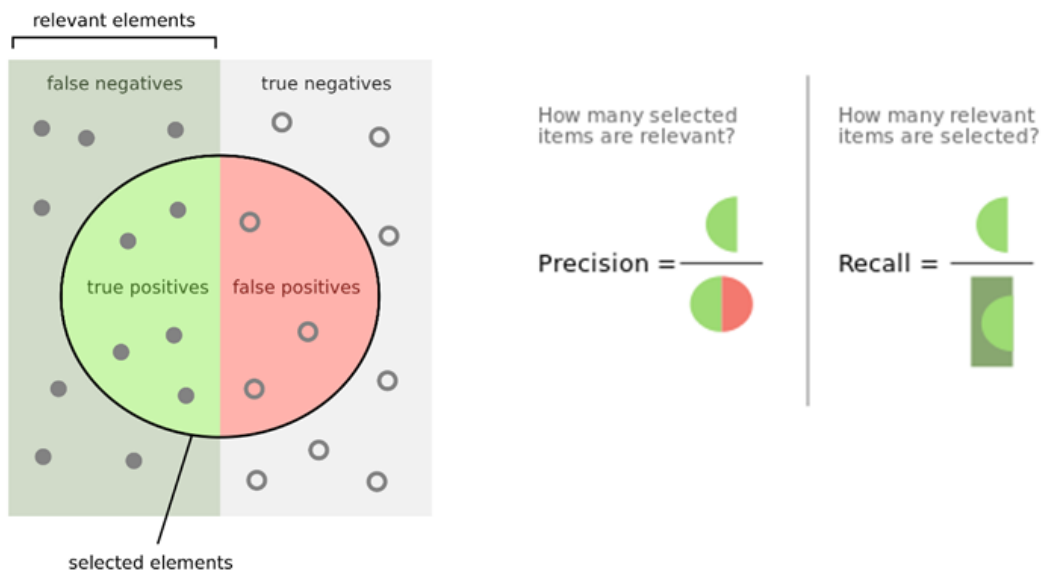
Figure 20 - Graphic Meaning of Precision and Recall parameters

$$F1\ Score = \frac{2 * Precision * Recall}{Precision + Recall} \tag{7}$$

3.4. Experiments

As already mentioned, simple PLS-DA, PCA plus PLS-DA, and SAM have been used for the following classification of pieces of eggplants of different cooking. The PLS-DA was implemented with the MATLAB's tool Hypertools [13].

It was applied a mask of predetermined areas to cover all the pre-established classes starting from the loading of the training hypercube containing the different pieces of eggplant plus the entire one. The order was in this way:

1. Fresh
2. Baked 10'
3. Baked 20'
4. Burnt
5. Skin
6. Background
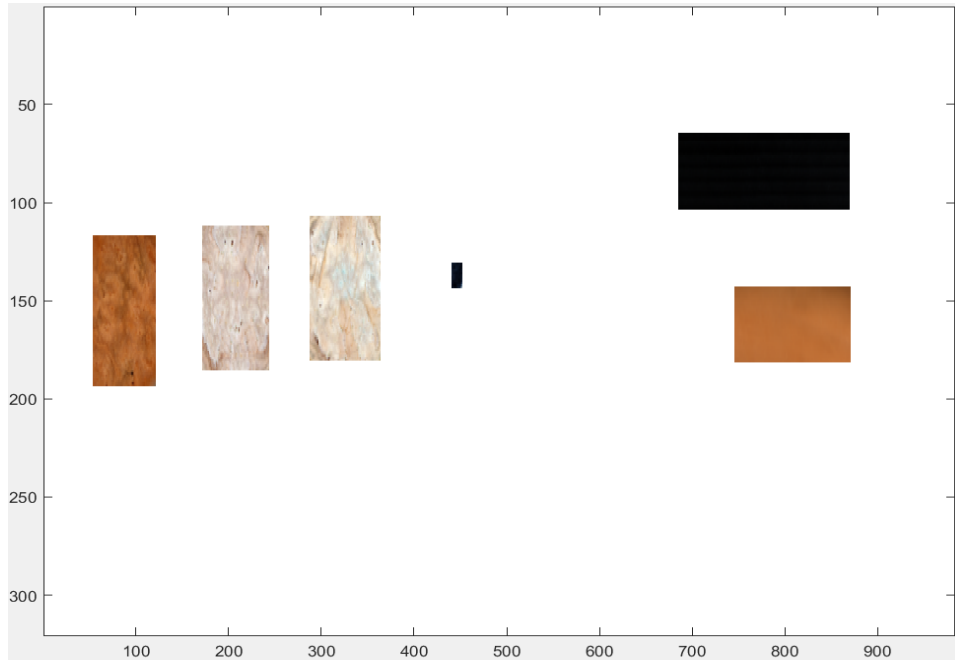
Figure 21 - The dataset with the mask applied. X and Y axis represent the spatial dimension of pixels

This choice has been mandatory in the forward view to compare the results of this algorithm with the SAM. In fact, the tool permits only the manual selection of the different classes and the presence of the mask allow an identical choice of pixels in all the tests.
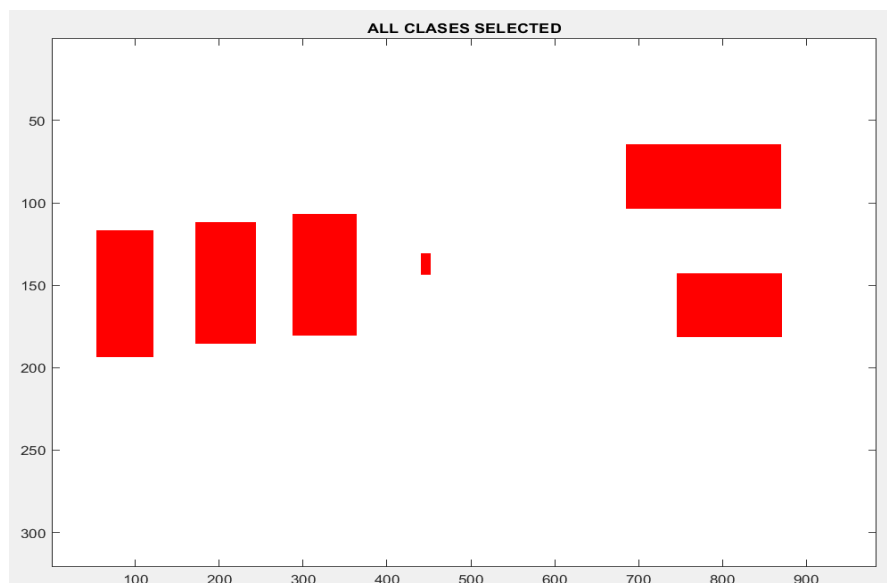
Figure 22 - Selection of the six classes of reference

Before the training it was possible to make some pre-processing transformations, like Smoothing, to clean the image from noises but none of them have really improved the scenario or the results.

We have a confirmation of that thanks to Figure 23, that shows no relevant differences between pre and after Smoothing. This is a further confirmation of the goodness of the experimental setup.
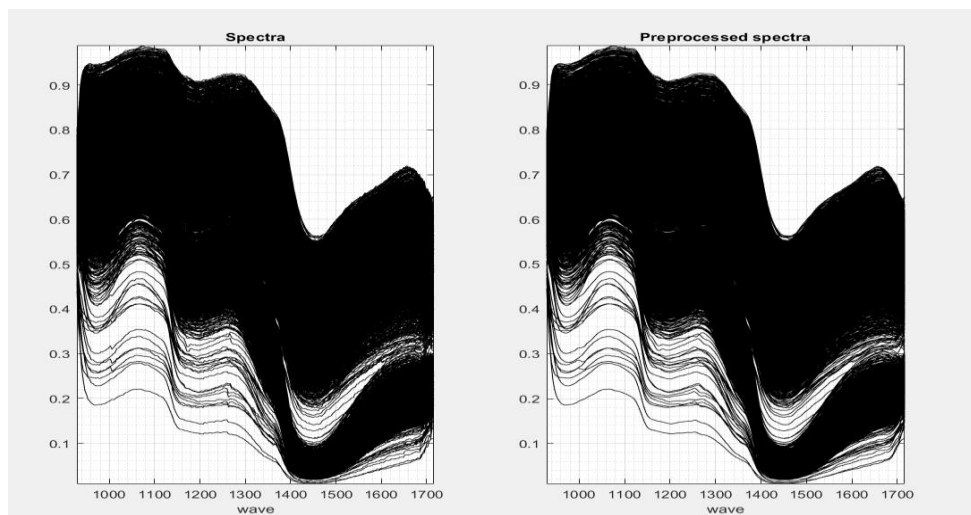
Figure 23 - The two graphics show the raw spectra of the image (on the left) and the spectra after Smoothing transformation (on the right). X axis represents wavelengths in nm, the Y axis the intensity in arbitrary unit

The next step is the calibration of the PLS-DA model. There are several options to select to start the cross validation (c.v.), a resampling procedure used to evaluate machine learning models on a limited data. Data scaling, assignation criterion, kind of validation and number of c.v. groups are the options and, respectively, there were selected the mean centering scaling, the Bayesian method, the Venetian blinds c.v. and three c.v. groups (to minimize timing cost). After that, the tool shows the behavior of the PLS-DA model with the reference to the error rate of c.v. according to the latent variables chosen. Considering the selected areas and the previous options, the Figure 24 gives a perception of the progression of the error rate and, hence, the decision to take ten components. Finally, the tool shows the results of the calibration in fitting and cross-validation (Table 1).

Figure 24 - The Hypertools PLS-DA Setting window

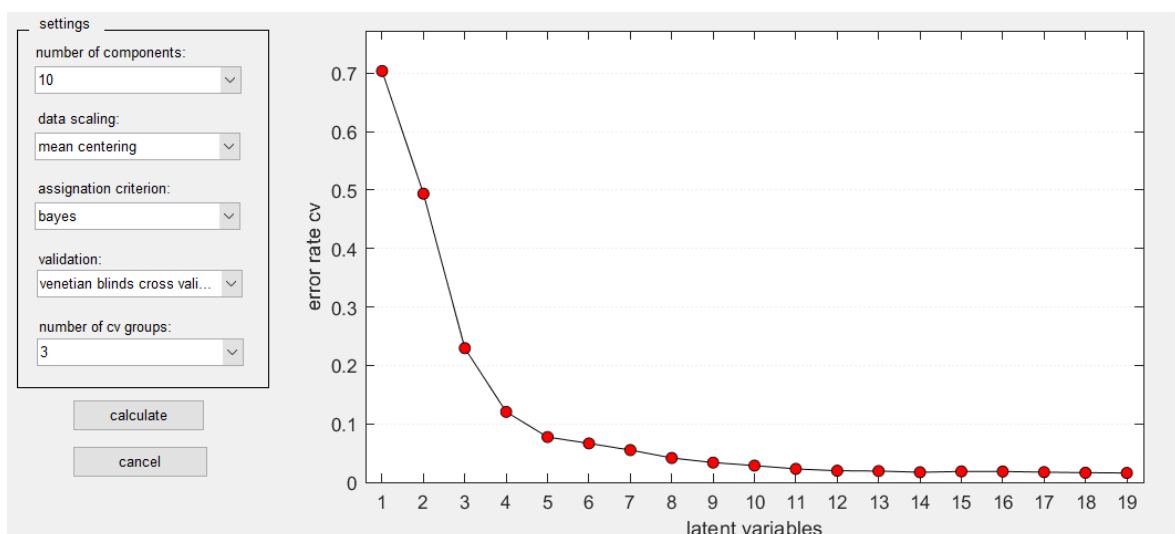| Fitting					
Error Rate	0.01		**Class**	**Specificity**	**Sensitivity**	**Precision**
Non-error Rate	0.99		1	1	1	1
Accuracy	0.99		2	0.99	0.96	0.97
Not assigned	0.07		3	0.99	0.97	0.96
			4	1	0.99	0.99
			5	1	1	1
			6	1	1	1

Validation					
Error Rate	0.01		**Class**	**Specificity**	**Sensitivity**	**Precision**
Non-error Rate	0.99		1	1	1	1
Accuracy	0.99		2	0.99	0.96	0.97
Not assigned	0.07		3	0.99	0.97	0.96
			4	1	0.99	0.99
			5	1	1	1
			6	1	1	1

Table 1- Results of PLS-DA calibration in fitting and validation

The second case involves a similar process with respect to the previous case but not identical. What it changes is the hypercube fed to the calibration of the PLS-DA model. The idea behind this process is to utilize data already compressed quickly to further reduce the compilation time of PLS-DA.

In this experiment it has been thought to reduce the original dataset with 224 spectral bands to a reduced one containing the first twelve components according to the algorithm of the Principal Component Analysis. From these assumptions, in the calibration of this second model was done in the same way as written above.

The last classification process is totally different. In MATLAB the application of SAM algorithm is very easy to use as there is a direct comparison between a test image and n different reference spectra. The final output is a 2-D image with pixels corresponding to a different reference spectrum calculated by similarity.

Usually with a lighter dataset it is used to get an algorithm to extract the most significant spectra from the training image. In this case such a choice would not be productive since it would generate not very sensitive predictive images. On the other hand, the choice of a too high number of reference spectra would bring out ghostly noises and minor signatures. For this reason, several tests have been carried out to get a good differentiation of the various areas of analysis and therefore of the six classes. In Figure 25 and Figure 26 it is possible to note the two problems already exposed, as in Figure 27 the found compromise of 40 reference spectra for our analysis. The last step to make these results comparable with others involves applying the same image mask and assigning a given reference spectrum to one of the six study classes, assigning unclassified to the others.
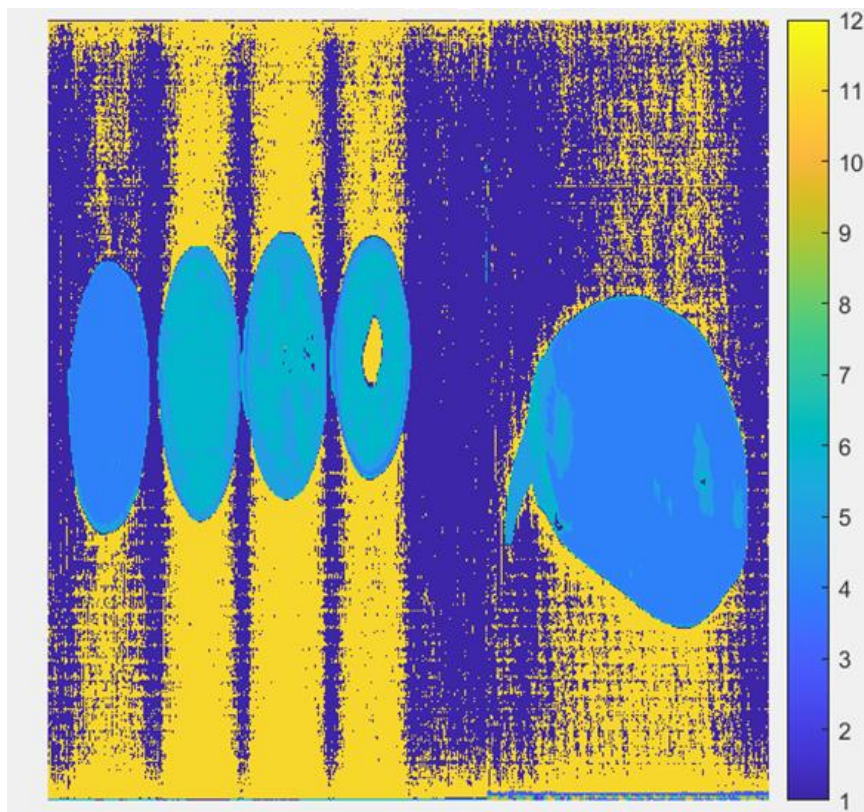

Figure 25 - Output of comparison of test image with first 12 reference spectra according to SAM
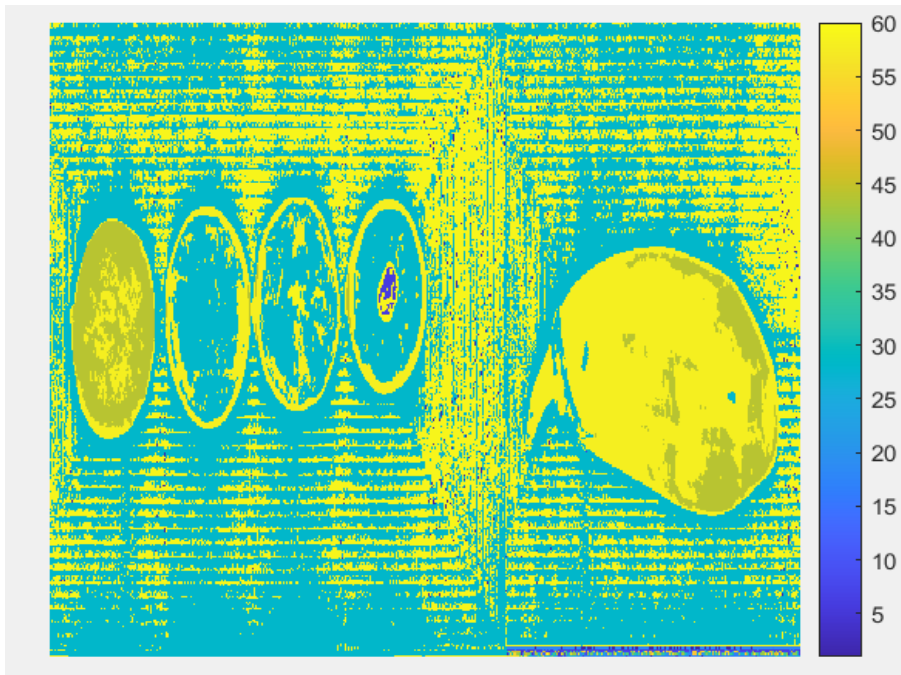
Figure 26 - Output of comparison of test image with first 60 reference spectra according to SAM
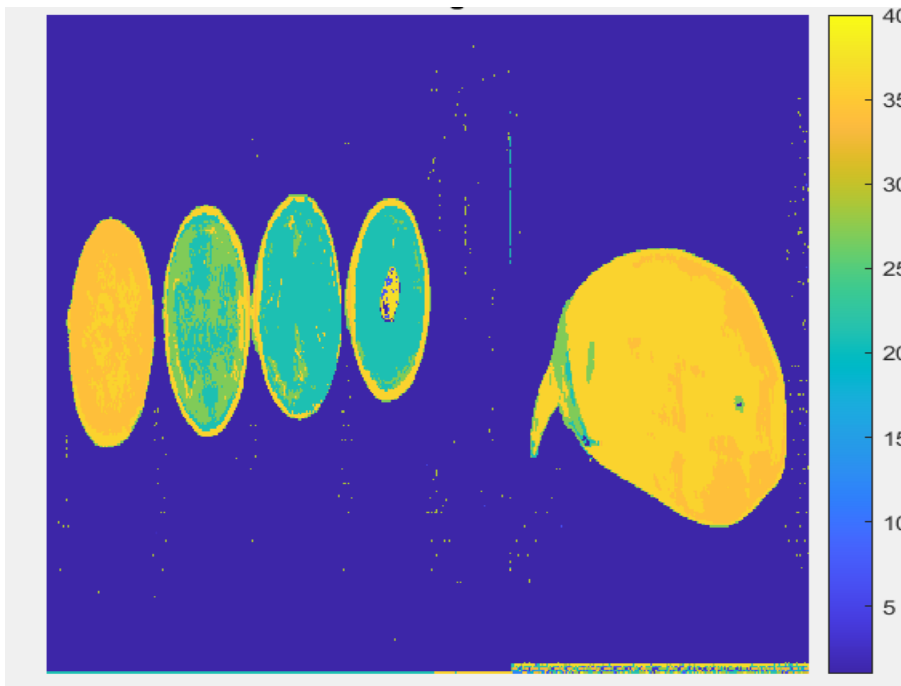
Figure 27 - Output of comparison of test image with first 40 reference spectra according to SAM

42

4. Results

In this chapter will be exposed the results in the chosen parameters discussed in the previous sub-chapter of Metrics.

To calculate the Confusion matrices of the three prediction images, for each case all six labeled areas were taken and placed on a one-dimensional vector.

Having done this, the final prediction vector was compared to the vector of classes we expected.

As remind, each label corresponds to a particular class:

1. Fresh
2. Baked 10'
3. Baked 20'
4. Burnt
5. Skin
6. Background

The class 0 means that a given pixel does not belong to any of the previous classes.

PLS-DA

In terms of Confusion Matrix, the Figure 28 shows us graphically how the PLS-DA model behaved with the test image. First, the background (class 6) is perfectly predicted without any kind of exception and, more importantly, there is no predictive error with the burnt area (class 4). Similarly, no criticalities come up on the Skin part (class 5) and the Fresh part of eggplant (class 1). There are few misunderstandings in the classifications between class 2 and 3. In these cases of different cooking, we have good predictions although it is fair to point out that the model fails to catalogue some of the pixels in these two categories.

The time for the elaboration of this PLS-DA model amounted to 588 s.

Figure 28 - Confusion Matrix of prediction from the simple PLS-DA model

PCA + PLS-DA

Even in this experiment there are no serious problems in correctly predicting the skin, the fresh part, or the background (Figure 29). The result of applying an already reduced dataset to the PLS-DA model is a difficulty in correctly predicting the burnt eggplant pixels as they are not even assigned to a class, also since it is the area with the least data to work with. Also, in this case there were several non-assignments of class in the baked areas, a slight increase compared to the case of the simple PLS-DA certainly due to the reduced dataset. The focal point of this test is the creation of a prediction model that could work with already smaller datasets in order to achieve computational savings, and these goals have been achieved. In fact, 23 s are needed to reduce the dataset with PCA method and X sec to create this PLS-DA model.

Figure 29 - Confusion Matrix of prediction from the PLS-DA model in combination with a dataset reduced with PCA

SAM

The last experiment with SAM is reported below.

In general, what comes out is a good prediction model with a significant decrease of unassigned pixels. At the same time, however, the incorrect classification of Fresh and Baked 10' pixels as, respectively, Skin and Baked 20' pixels increases.

The biggest weakness of classification with SAM is that, as there is a direct comparison between test images and reference spectra without any kind of structure or model creation, the time the algorithm takes to compare such different spectral signatures is very high each time: about 1162 seconds with respect o few seconds of other predictions.

Figure 30 - Confusion Matrix of predictions from SAM algorithm

Discussion

Having given a first impression of the results, it is essential to use the various parameters of Precision, Recall and F1 Score to give a more accurate judgement of what has been achieved. In parallel to this master thesis, a thesis on Deep Learning

Experiment	Class	Precision	Recall	F1 Score
PLS-DA	1	1.0	0.99	1.0
	2	0.97	0.82	0.89
	3	0.96	0.78	0.86
	4	0.97	0.97	0.97
	5	1.0	1	1.0
	6	1.0	1	1.0
PCA+PLS-DA	1	1.0	0.94	0.97
	2	0.93	0.74	0.82
	3	0.96	0.69	0.80
	4	0.86	0.12	0.20
	5	0.98	0.98	0.98
	6	1.0	1.0	1.0
SAM	1	0.92	0.64	0.75
	2	0.89	0.52	0.65
	3	0.66	0.84	0.74
	4	1.0	0.89	0.94
	5	0.69	0.94	0.80
	6	1.0	1.0	1.0
LucasNNN (with 200 Epochs and 0.001 Learning Rate)	1	1.0	1.0	1.0
	2	1.0	1.0	1.0
	3	0.79	1.0	0.88
	4	1.0	0.82	0.90
	5	1.0	1.0	1.0
	6	1.0	1.0	1.0
Hu et al. (with 200 Epochs and 0.001 Learning Rate)	1	1.0	1.0	1.0
	2	1.0	1.0	1.0
	3	0.98	0.99	0.99
	4	0.99	0.98	0.99
	5	1.0	1.0	1.0
	6	1.0	1.0	1.0

Algorithm of Machine Learning

Algorithm of Deep Learning

Table 2 - Performance parameters of Master Thesis Machine Learning algorithms and two Deep Learning algorithms applied to the same dataset

algorithms was carried out by other colleagues of Politecnico di Milano on the same dataset and with the same choice of areas and classes.

Above there is the Table 2 summarizing the various models taken into consideration with their respective parameters.

Just a glance one can note that the prediction of the Fresh, Burnt, Skin and Background class is good for almost all the algorithms considered.

Exceptions are PCA + PLS-DA with Burnt (F1-Score 0.20) and SAM with Fresh (F1-Score 0.75). The first problem is the fact that we have class 5 with a very small selected range in itself and much smaller than the others, and at the same time we have significantly reduced the number of band components. This is confirmed if you look not only at the weighted performance value of F1-Score, but also at the first two. It comes out that although there is good precision (0.86), the sensitivity value is very low (0.12).

As for SAM, here too the value of Recall is decisive for the lowering of performance (Precision 0.92, Recall 0.64), due to a predictive difficulty between Fresh class and Skin class (as it can see in Figure 30). The most obvious differences between the Machine Learning algorithms under study and the best Deep Learning algorithms compared here are in the prediction of cooked eggplants, in both 10' and 20'.

In fact, except for the F1-Score value of class 3 of PLS-DA and LucasNN (0.86 and 0.88 respectively), the other Deep Learning algorithms values do not go below 0.99 while PLS-DA stops at 0.89 with class 10'. The good fortune of Deep Learning algorithms lies in the fact that much more complex architectures are created with Convolutional Neural Networks than with Machine Learning algorithms [15], resulting in much longer training times than those used in this master thesis.

Coming back with a focus on the selected machine learning algorithms, Table 2 shows that the choice of 10 components with PLS-DA is the best choice if we want a correct prediction of all 6 classes considered.

Certainly, the algorithm that creates the most problems in class assignment is SAM. Although there is an easier and more straightforward way to assign a given class to a pixel, what limits this method is the use of many (40) reference spectra that obviously do not all match the six needed, generating also an important computational cost every time a prediction is made. On the other hand, as already written, a choice of a lower number of spectra would lead to whole areas being confused with each other, given the similar nature of the objects.

A fair compromise can be found with PLS-DA with an already reduced dataset thanks to the PCA technique. The speed with which the two methods are applied together in this case and the small amount of data used are the key words that summarize the focus of this experiment. In a world where more information is needed for increasingly detailed analysis, methods against data wastage are being studied at the same time, and this is the way forward. In our case, these advantages are reflected by reducing performance, except for the already discussed class 4, by a few percent compared to the PLS-DA reference.

5. Conclusions

This Master Thesis brought important first insights into the classification of vegetables exposed to different cooking times for Food Quality investigations using Machine Learning in association with Hyperspectral Imaging.

1. To date, among the various methods analyzed, the one that provides the best performance across all six classes is PLS-DA.

2. The greatest difficulty lies in correctly predicting an eggplant with 10' baking versus one with 20'.

3. If you need to work quickly with a small dataset, the PCA + PLS-DA solution is the most suitable, unless you necessarily must consider burnt parts (F1-score 0.20).

4. In the prediction of the Burnt class, an excellent alternative is the SAM method, even though the selected reference area is much smaller than the others.

5. In all the algorithms considered, the background is always correctly predicted with no margin of error.

6. For an excellent Fresh and Skin class assessment PLS-DA and PCA + PLS-DA are both recommended, thanks to F1-scores (1.0, 1.0) and (0.97, 0.98).

There are some suggestions for the next research for this thesis topic:

1. Upgrade the program into real time classification.

2. Increase the class area of Burnt, due to have a better prediction of it with reduced datasets.

3. Increase the semantic segmentation with other possible useful classes (such as rotten eggplants).

4. Subdivision of class analyses into several models to achieve greater sensitivity and accuracy (e.g., Fresh, Baked 10' and Baked 20', or Fresh, Burnt and Background) without necessarily using Deep Learning analysis techniques

Finally, it is legitimate to write that we are only at the beginning of a process of studies and applications in the world of Food Sorting technologies that will certainly lead to enormous progress in the field of non-destructive and ever faster analysis, thanks mainly to the Hyperspectral Image Systems. Personally, I am proud to have been able to take part in this master thesis, which focused on aspects that had never been analyzed before and which may serve as a reference for future studies.

Bibliography

[1] Y. Lu, W. Saeys, M. Kim, Y. Peng, and R. Lu, "Hyperspectral imaging technology for quality and safety evaluation of horticultural products: A review and celebration of the past 20-year progress," Postharvest Biol. Technol., vol. 170, no. August, p. 111318, 2020, doi: 10.1016/j.postharvbio.2020.111318.

[2] Saha, D., & Manickavasagan, A. (2021). Machine learning techniques for analysis of hyperspectral images to determine quality of food products: A review. *Current Research in Food Science*, *4*(December 2020), 28–44. https://doi.org/10.1016/j.crfs.2021.01.002.

[3] B. Zhang et al., "Principles, developments and applications of computer vision for external quality inspection of fruits and vegetables: A review," Food Res. Int., vol. 62, pp. 326–343, 2014, doi: 10.1016/j.foodres.2014.03.012.

[4] D. Sun, G. Elmasry, and D. Sun, Hyperspectral Imaging for Food Quality Analysis and Control Principles of Hyperspectral Imaging Technology. Dublin: Academic Press, 2010.

[5] *Mitchell, T. (1997). Machine Learning. McGraw Hill. p. 2.*

[6] Baştanlar, Y., & Ozuysal, M. (2014). Introduction to Machine Learning Second Edition. In *Methods in molecular biology (Clifton, N.J.)* (Vol. 1107). https://doi.org/10.1007/978-1-62703-748-8_7.

[7] W. Lv and X. Wang, "Overview of Hyperspectral Image Classification," J. Sensors, vol. 2020, 2020, doi: 10.1155/2020/4817234.

[8] Lacar, F.M.; et al. (2001). "Use of hyperspectral imagery for mapping grape varieties in the Barossa Valley, South Australia". Geoscience and Remote Sensing Symposium (IGARSS'01) – IEEE 2001 International. 6: 2875–2877. doi:10.1109/IGARSS.2001.978191.

[9] P. M. Mehl, K. Chao, M. Kim, and Y. R. Chen, "Detection of defects on selected apple cultivars using hyperspectral and multispectral image analysis," J. Agric. Saf. Health, vol. 18, no. 2, pp. 219–226, 2012, doi: 10.13031/2013.7790.

[10] R. Moscetti et al., "Hazelnut Quality Sorting Using High Dynamic Range Short-Wave Infrared Hyperspectral Imaging," Food Bioprocess Technol., vol. 8, no. 7, pp. 1593–1604, 2015, doi: 10.1007/s11947-015-1503-2.

[11] Y. E. Esin, O. Ozdil, S. Ozturk, and B. Demirel, "Practical Focus Adjustment Method for Hyperspectral Cameras," 2019.

[12] D. Sun, G. Elmasry, and D. Sun, Hyperspectral Imaging for Food Quality Analysis and Control Principles of Hyperspectral Imaging Technology. Dublin: Academic Press, 2010.

[13] Mobaraki, N., & Amigo, J. M. (2018). HYPER-Tools. A graphical user-friendly interface for hyperspectral image analysis. *Chemometrics and Intelligent Laboratory Systems*, *172*(October 2017).

[14]   Chang, C.-I., and Q. Du. "Estimation of Number of Spectrally Distinct Signal Sources in Hyperspectral Imagery." IEEE Transactions on Geoscience and Remote Sensing 42, no. 3 (March 2004): 608–19. https://doi.org/10.1109/TGRS.2003.819189.

[15]   L. Zhu, P. Spachos, E. Pensini, and K. N. Plataniotis, "Deep learning and machine vision for food processing: A survey," *Curr. Res. Food Sci.*, vol. 4, no. December 2020, pp. 233–249, 2021, doi: 10.1016/j.crfs.2021.03.009.