AGRICULTURAL UNIVERSITY OF ATHENS
School of Food, Biotechnology and Development
Department of Food Science and Human Nutrition
Laboratory of Microbiology and Biotechnology of Foods

ΓΕΩΠΟΝΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ
Σχολή Τροφίμων, Βιοτεχνολογίας και Ανάπτυξης
Τμήμα Επιστήμης Τροφίμων και Διατροφής του Ανθρώπου
Εργαστήριο Μικροβιολογίας και Βιοτεχνολογίας Τροφίμων

PhD Thesis

"DETERMINATION OF MINCED MEAT QUALITY USING MACHINE LEARNING"

«ΕΚΤΙΜΗΣΗ ΤΗΣ ΠΟΙΟΤΗΤΑΣ ΤΟΥ ΚΡΕΑΤΟΣ ΜΕ ΧΡΗΣΗ ΜΕΘΟΔΩΝ ΜΗΧΑΝΙΚΗΣ ΜΑΘΗΣΗΣ»

Athina I. Ropodi

Athens, 2017

Supervisor: Professor George-John Nychas

AGRICULTURAL UNIVERSITY OF ATHENS
School of Food, Biotechnology and Development
Department of Food Science and Human Nutrition
Laboratory of Microbiology and Biotechnology of Foods

ΓΕΩΠΟΝΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ
Σχολή Τροφίμων, Βιοτεχνολογίας και Ανάπτυξης
Τμήμα Επιστήμης Τροφίμων και Διατροφής του Ανθρώπου
Εργαστήριο Μικροβιολογίας και Βιοτεχνολογίας Τροφίμων

PhD Thesis

"DETERMINATION OF MINCED MEAT QUALITY USING MACHINE LEARNING"

«ΕΚΤΙΜΗΣΗ ΤΗΣ ΠΟΙΟΤΗΤΑΣ ΤΟΥ ΚΡΕΑΤΟΣ ΜΕ ΧΡΗΣΗ ΜΕΘΟΔΩΝ ΜΗΧΑΝΙΚΗΣ ΜΑΘΗΣΗΣ»

Athina I. Ropodi

Athens, 2017

Supervisor: Professor George-John Nychas

PhD Thesis

"DETERMINATION OF MINCED MEAT QUALITY USING MACHINE LEARNING"

Athina I. Ropodi

Members of Supervising Board

Supervisor

George-John Nychas, Professor

Members/ co-supervisors

Elias Manolakos, Professor

Efstathios Panagou, Assistant Professor

Members of Evaluation Board

George-John Nychas, Professor, Agricultural University of Athens

Elias Manolakos, Professor, National and Kapodistrian University of Athens

Efstathios Panagou, Assistant Professor, Agricultural University of Athens

Konstantinos Karantzalos, Assistant Professor, National Technical University of Athens

Eleftherios Drosinos, Professor, Agricultural University of Athens

Fady Mohareb, Senior Lecturer, Cranfield University

Panagiotis Skandamis, Associate Professor, Agricultural University of Athens

ΠΕΡΙΛΗΨΗ

Τη σημερινή εποχή, οι καταναλωτές απαιτούν συνεχή επιβεβαίωση της προέλευσης, της ποιότητας και της συμμόρφωσης με την ετικέτα των τροφίμων που αγοράζουν. Για το λόγο αυτό, οι βιομηχανίες τροφίμων, οι έμποροι και οι αρχές είναι αναγκαίο να αναπτύξουν προηγμένες, αποτελεσματικές και χαμηλού κόστους λύσεις για τη διασφάλιση της ποιότητας και τον εντοπισμό δόλιων πρακτικών. Σε αυτό το πλαίσιο, η μελέτη αυτή εστιάζει (α) στην πολυφασματική απεικόνιση (Multispectral Imaging-MSI), (β) την φασματοσκοπία υπέρυθρου με μετασχηματισμό Fourier (Fourier Transform Infrared -FTIR spectrometry) και (γ) την εφαρμογή προηγμένων μεθόδων ανάλυσης δεδομένων και μηχανικής μάθησης. Όσον αφορά την ποιότητα, εξετάστηκε η αποτελεσματικότητα των προαναφερόμενων μεθόδων σε σχέση με (α) τον εντοπισμό της μη-συμμόρφωσης με την ετικέτα ή/και δόλιων πρακτικών και (β) τη μικροβιολογική αλλοίωση. Οι ακόλουθες αναλύσεις, έλαβαν χώρα:

Στην 1^{η} περίπτωση, εξετάστηκε η νοθεία του μοσχαρίσιου κιμά με χοιρινό. Χρησιμοποιήθηκαν 220 πολυφασματικές εικόνες δειγμάτων από 4 ανεξάρτητες πειραματικές διαδικασίες (κομμάτια κρέατος διαφορετικής προέλευσης). Η νοθεία έγινε με βήμα 10% w/w, δημιουργώντας 11 κατηγορίες (συμπεριλαμβανομένων των ανόθευτων χοιρινών και μοσχαρίσιων δειγμάτων). Μετά από ένα στάδιο προεπεξεργασίας της εικόνας, εφαρμόστηκαν η Ιεραρχική Ανάλυση Συστάδων (Hierarchical Cluster Analysis - HCA) και Ανάλυση Κυρίων Συνιστωσών (Principal Component Analysis - PCA). Παρατηρήθηκαν δε σημαντικές διαφορές μεταξύ των διαφορετικών κομματιών κρέατος και των διαφορετικών κλάσεων όταν και τα τρία πρώτα ζευγάρια κομματιών κρέατος συμπεριλήφθηκαν στην ανάλυση. Μετά την

κατάτμηση των δεδομένων σε σετ εκπαίδευσης και επικύρωσης, τα δεδομένα του τέταρτου ζεύγους χρησιμοποιήθηκαν για ανεξάρτητη επικύρωση και εφαρμόστηκαν οι μέθοδοι Γραμμικής Διακριτικής Ανάλυσης και Μερικών Ελαχίστων Τετραγώνων Linear Discriminant Analysis - LDA, Partial least-squares discriminant analysis – PLSDA) για 11 και για 3 (ανόθευτα χοιρινά, μοσχαρίσια και νοθευμένα) κλάσεις. Στην περίπτωση των 11 κλάσεων, 98.48% και 96.97% των δειγμάτων κατηγοριοποιήθηκαν εντός μιας ±10% κατηγορίας για LDA και PLSDA αντίστοιχα, ενώ στην περίπτωση των τριών επιτεύχθηκε σωστή κατηγοριοποίηση 98.48%. Τα αποτελέσματα της ανεξάρτητης επικύρωσης ήταν λιγότερο ακριβή για την LDA, αλλά με την PLSDA όλα τα δείγματα κατηγοριοποιήθηκαν σωστά, αποδεικνύοντας ότι το ποσοστό 10% είναι εντός των ορίων ανίχνευσης.

Στην δεύτερη περίπτωση, 110 δείγματα κιμά τριών διαφορετικών κομματιών κρέατος από μοσχάρι και άλογο και επιπλέον εικόνες που ελήφθησαν μετά από 6, 24 και 48 ώρες χρησιμοποιήθηκαν για την ανίχνευση νοθείας. Η PCA χρησιμοποιήθηκε για οπτικοποίηση των δεδομένων, ενώ οι μέθοδοι PLSDA και Random Forest (RF) για κατηγοριοποίηση μεταξύ διαφορετικών ποσοστών νοθείας (4 κλάσεις), ανόθευτων μοσχαρίσιων, ανόθευτων αλογίσιων και νοθευμένων, ανόθευτων και νοθευμένων, και τέλος μεταξύ φρέσκων και συντηρημένων δειγμάτων. Τα μοντέλα κατά την ανεξάρτητη επικύρωση δεν είχαν υψηλή ακρίβεια. Στο τέλος, προτιμήθηκε η χρήση μηχανών διανυσμάτων υποστήριξης (Support Vector Machines – SVMs) σε δύο στάδια προκειμένου να διαχωριστούν τα φρέσκα από τα συντηρημένα δείγματα και μετά τα νοθευμένα από τα ανόθευτα. Έτσι, επιτεύχθηκε ποσοστό σωστής κατηγοριοποίησης 95.31% στο ανεξάρτητο σετ επικύρωσης.

Στην τρίτη περίπτωση, ελήφθησαν πολυφασματικές εικόνες και φάσματα FTIR από κιμά επτά διαφορετικών κομματιών και από αντίστοιχα αποψυγμένα δείγματα που είχαν καταψυχθεί στους -20°C για 7 και 32 μέρες (συνολικά 105 εικόνες και φάσματα). Η PCA χρησιμοποιήθηκε για τη διερεύνηση των δεδομένων, ενώ PLSDA και SVM πέτυχαν 100% σωστή κατηγοριοποίηση μεταξύ φρέσκων και αποψυγμένων κατά την επικύρωση και την ανεξάρτητη επικύρωση με χρήση πολυφασματικών εικόνων. Η FTIR ήταν λιγότερο ακριβής με 93.3 και 96.7% αντίστοιχα.

Στην 4^{η} και 5^{η} περίπτωση, διερευνήθηκε η αλλοίωση του βοδινού κιμά. Στην 4^{η} περίπτωση, χρησιμοποιώντας τη διαδικτυακή εφαρμογή “MeatReg”, χρησιμοποιήθηκαν επτά διαφορετικές μέθοδοι για την εκτίμηση του μικροβιακού πληθυσμού. Τα δεδομένα αποτελούνταν από 105 δείγματα συντηρημένα σε δυο διαφορετικές συσκευασίες -αέρας και modified air packaging (MAP - 20% CO_2/ 80% O_2)- και δύο θερμοκρασίες (4 και 10°C), μικροβιολογικές αναλύσεις (Pseudomonads, Lactobacilli, *B. thermosphacta* and Enterobacteriaceae, Ολική Μεσόφιλη Χλωρίδα - OMX). Τα δεδομένα πολυφασματικής απεικόνισης και FTIR συγκρίθηκαν με αυτά από ηλεκτρονική μύτη, υγρή χρωματογραφία υψηλής απόδοσης (HPLC) και αέρια χρωματογραφία/ φασματοσκοπία μάζας (GC-MS). Τα αποτελέσματα διαφοροποιήθηκαν αρκετά ανάλογα το είδος του οργάνου και της ομάδας μικροοργανισμών. Παρόλα αυτά υπήρξε καλή ακρίβεια, με την μέθοδο RF να δίνει τα καλύτερα αποτελέσματα.

Ομοίως στην 5^{η} περίπτωση, 168 δείγματα βοδινού κιμά αναλύθηκαν ως προς την OMX, ενώ παράλληλα έγιναν μετρήσεις FTIR. Τα δείγματα είχαν συντηρηθεί σε αέρα και MAP στους 4 και 10°C. Χρησιμοποιήθηκε μία προσέγγιση βασισμένη στην μεθοδολογία των ensemble μοντέλων, όπου η εκτίμηση της αλλοίωσης έγινε βάσει

μίας μίξης αποτελεσμάτων επιμέρους νευρωνικών δικτύων (artificial neural networks). Το μέσο τετραγωνικό σφάλμα της πρόβλεψης ήταν 0.16 (log CFU/g)^2.

Κύρια Επιστημονικά Πεδία: Γεωπονικές Επιστήμες - Μικροβιολογία Τροφίμων, Επιστήμες Υπολογιστών - Μηχανική Μάθηση.

Λέξεις-κλειδιά: ποιότητα κρέατος, ανάλυση δεδομένων, μηχανική μάθηση, πολυφασματική απεικόνιση, φασματοσκοπία υπέρυθρου με μετασχηματισμό Fourier (FTIR).

ABSTRACT

Currently, consumers expect constant reassurance of the origin, quality and compliance to label of the food products they purchase. Therefore, food industries, retailers and authorities have to develop advanced, effective and relatively low-cost solutions for quality assurance and detection of fraudulent practices. In this context, this study focuses on (a) Multispectral Imaging (MSI), (b) Fourier Transform Infrared (FTIR) spectrometry and (c) the application of advanced data analysis and machine learning methodologies. In terms of quality, the efficacy of the abovementioned methods concerning (a) non-compliance to label/ fraud detection and (b) microbiological spoilage is examined and the following analyses took place:

In the first case, minced beef adulteration with pork was investigated. MSI data from 220 meat samples from four independent experiments (different meat batches) were extracted. Adulteration was performed with a 10% w/w step, creating 11 classes of samples (including pure beef and pork). After an image preprocessing step, Hierarchical Cluster Analysis (HCA) and Principal Component Analysis (PCA) were applied. Meat batches displayed significant differences and different classes were less distinguishable when the first three batches were included in the analysis. After partitioning in training and validation sets, the fourth batch was retained for independent/ external validation and Linear Discriminant Analysis (LDA) and Partial Least Squares Discriminant Analysis (PLSDA) were performed for classification among 11 classes and pure *vs*. adulterated samples. For the case of 11 classes, 98.48% and 96.97% of the samples were classified within a ±10% category of adulteration for LDA and PLSDA respectively, whereas the 3-class case yielded 98.48% overall

correct classification (OCC). Results for the external validation set, proved LDA significantly less accurate compared to PLSDA where all samples were classified correctly, proving that 10% is within the method's detection limit.

In the second case, with 110 samples from three different batches of minced beef and horsemeat and additional images captured after 6, 24 and 48 h, model performance is investigated in terms of detection of adulteration. PCA was used for visualization purposes, while PLSDA and Random Forest (RF) for classification among different percentages of beef (4 classes), pure beef *vs.* pure horsemeat *vs.* adulterated samples, pure beef and horse *vs.* adulterated samples and freshly-ground *vs.* stored minced meat. Models significantly underperformed in independent validation. In the end, a two stage Support Vector Machine (SVM) methodology was utilized, where freshly-ground samples are separated from stored and then pure separated from adulterated. The OOC of the SVM model was equal to 95.31% for independent model validation.

In the third case, multispectral images and FTIR spectra from seven different batches of freshly-ground beef, along with MSI and FTIR spectral data after being frozen (-20°C) for 7 and 32 days and then thawed were acquired (in total, 105 measurements per sensor). In terms of data analysis methods, PCA was used for data exploration, while PLSDA and SVM yielded 100% correct classification between fresh and frozen-then-thawed samples for MSI test and external validation sets. FTIR proved less accurate, as PLSDA yielded 93.3 and 96.7% classification accuracy for the test and external validation set, respectively.

In cases 4 and 5, meat and specifically minced beef spoilage is explored. In case 4, using "MeatReg", a web-based application, seven methods were tested for the

prediction of bacterial counts. The dataset included 105 samples, stored in air or under modified air packaging (MAP) conditions (20% CO_2/ 80% O_2) at 4 and 10°C, microbiologically analyzed (Pseudomonads, Lactobacilli, *B. thermosphacta* and Enterobacteriaceae, as well as TVC). FTIR and MSI data were compared with electronic nose, High Performance Liquid Chromatography and Gas Chromatography coupled to Mass Spectrometry data. Results were mixed depending on the sensor and species counts, while RF regression yielded an overall good performance.

Similarly to case 4, in case 5, 168 minced beef samples were analyzed for TVC, while FTIR measurements took place. Samples were stored in air or under modified air packaging (MAP) conditions at 4 and 10°C. An ensemble-based approach was employed where spoilage estimation was a fusion of several artificial neural networks yielding a mean squared error equal to 0.16 $(\log CFU/g)^2$.

Main Scientific Disciplines: Agricultural Sciences - Food Microbiology, Computer Sciences - Machine Learning.

Related keywords: meat quality, data analysis, machine learning, multispectral imaging, Fourier Transform Infrared Spectroscopy (FTIR).

ACKNOWLEDGEMENTS

I would like to thank my supervisor Professor George-John Nychas for his excellent guidance, support and patience. I am grateful for the opportunity to carry out this PhD study.

Also many thanks to Professors Efstathios Panagou for his input and advice throughout this study and Elias Manolakos for his trust in me.

I would also like to thank my husband Thanassis for his support and encouragement, my parents, Mandy and Giannis, and my brother Kostas for their unconditional love and support all these years.

Lastly, I would like to express my gratitude to Dr. Alexandra Lianou for her valuable advice and friendship, my colleague Dimitrios Pavlidis for our excellent cooperation and assistance in our parallel studies, as well as all the members of the Laboratory of Microbiology and Biotechnology of Foods for their assistance.

TABLE OF CONTENTS

LIST OF TABLES[1]

[1] For tables not included in manuscripts, publications or supplementary material

LIST OF FIGURES[2]

[2] For figures not included in manuscripts, publications or supplementary material

LIST OF ABBREVIATIONS

FTIR	Fourier transform infrared spectroscopy
HSI	Hyperspectral imaging
MSI	Multispectral imaging
SSO	Specific spoilage organisms
E(S)SO	Ephemeral/ specific spoilage organisms
MAP	Modified atmosphere packaging
LAB	Lactic acid bacteria
TVC	Total viable counts
RMSE	Root mean squared error
MSE	Mean squared error
HCA	Hierarchical cluster analysis
PCA	Principal component analysis
PC	Principal component
DFA	Discriminant function analysis
MLR	Multiple linear regression
PCR	Principal components regression
PLSR	Partial least-squares regression
LDA	Linear discriminant analysis
PLSDA	Partial least-squares discriminant analysis

ANN	Artificial neural network
MLP	Multilayer Perceptron
RBF	Radial basis function
SVM	Support vector machine
SVR	Support vector regression
GA	Genetic algorithm
GP	Genetic programming
RF	Random Forest
kNN	k Nearest Neighbor
CV	Cross-validation
LOOCV	Leave-One-Out cross-validation
HPLC	High Performance Liquid Chromatography
GC-MS	Gas Chromatography coupled to Mass Spectrometry

CHAPTER 1:

MEAT QUALITY, RAPID METHODS & OBJECTIVES

1 PREAMBLE

Food quality, and meat quality especially, is of great importance to modern day consumers. While quality may be a subjective term depending on the consumer's cultural and/ or economic background, as well as his or hers sensory acuity, there are subjective criteria based on microbiological analyses that help determine a food commodity's quality especially in terms of microbiological spoilage (European Commission, 2005). Additionally, nowadays European consumers expect constant reassurance of the origin, quality and compliance to label of the food products they purchase, following recent food scandals and detected cases of fraud (European Commission, 2015). In order to safeguard consumer trust, food industries, retailers and authorities have to develop advanced, effective and relatively low-cost solutions for quality assurance and detection of fraudulent practices. In this context, rapid methods' potential for monitoring and controlling critical parameters of meat quality, as well as allowing traceability and detecting fraud needs to be explored.

The technological advances of the recent years have led to a plethora of instruments/ sensors such as (i) arrays of biomimetic sensors (e-nose, e-tongue), (ii) vibrational spectroscopy (Fourier transform infrared - FTIR, Raman) and (iii) surface chemistry (hyper/ multispectral imaging) (Ropodi, Panagou, & Nychas, 2016). These advances coupled with the advances in computer hardware and computer science disciplines, including machine learning algorithms and computational intelligence methodologies, have opened a new multi-disciplinary field where food science and conventional microbiology meet sensor technology, software development and case-specific data

analysis techniques in order to assist and/ or substitute traditional –and otherwise extremely reliable– time-consuming methods.

This chapter describes what meat quality is, as defined for this PhD study and presents some indicative applications of sensors used in previous studies. Lastly, the objectives and overall methodology of this study are presented.

1.1 MEAT QUALITY

This study focuses on the determination of minced meat -and specifically beef- quality. In fact, it focuses on quality in terms of (a) non-compliance to label/ fraud detection and (b) microbiological spoilage.

1.1.1 COMPLIANCE TO LABEL & FRAUD DETECTION

As mentioned above, the safety of meat products and minced beef in particular is extremely important following cases of microbiological outbreaks, dioxin contamination and other threats to human health (EFSA (European Food Safety Authority), 2015, 2017). Therefore, as consumer awareness increases, consumer trust becomes an important factor, since it may lead to dire economic consequences for the whole meat industry encompassing slaughterhouses, packers, importers, distribution and supply/ transportation operations, importers, as well as retailers.

Despite those consequences, minced beef is an attractive target for fraudulent and deceptive practices, due to its higher market price. Some practices involve adulteration and selling frozen-then-thawed meat as fresh.

- Meat Adulteration:

Indeed meat adulteration is a growing challenge for both the industry and the authorities, as the nature of the adulterants is many times unknown and unpredictable. As minced meat is the basic ingredient of beefburgers, adulteration is a serious problem involving economic, quality, safety and socio-religious issues (Alamprese, Casale, Sinelli, Lanteri, & Casiraghi, 2013).

Adulteration involves substitution or partial substitution of beef of obviously higher commercial value with cheaper meats or other ingredients, such as pork or offal or by adding proteins from several origins (Kamruzzaman, Sun, ElMasry, & Allen, 2013; Tian, Wang, & Cui, 2013).

Several standard analytical techniques, such as immunological and enzymatic techniques, DNA and protein-based assays and triacylglycerol analysis have been applied to authenticate food commodities (Ballin, 2010; Soares, Amaral, Mafra, & Oliveira, 2010). According to Ballin (2010), methods that have or could be used for meat adulteration (with other species, tissue/ fat, and proteins) include enzyme-linked immuno sorbent assay (ELISA), traditional and real time polymerase chain reaction (PCR), conformation sensitive gel electrophoresis (CSGE), liquid chromatography (LC), high performance liquid chromatography (HPLC), mass spectrometry (MS), random amplified polymorphic DNA (RAPD) and restriction fragment length polymorphism (RFLP). Specifically for species determination, rapidly evolving DNA-based techniques are considered superior compared to proteins, as DNA has a higher thermal stability and is present in the majority of cells. Furthermore, techniques like RFLP and RAPD are able to amplify few length polymorphisms and random

polymorphic DNA with little or no information on the sequence (Ballin, 2010), however RAPD reproducibility has been reported to be poor (Ballin, Vogensen, & Karlsson, 2009). In the case of sequencing, characterization of animal species depends on the availability of known sequences used for comparison. In addition, single-strand conformational analysis (SSCA) and conformation sensitive gel electrophoresis (CSGE) helps provide PCR speciation. Immunological methods take advantage of antigen-antibody interactions and are suitable for species determination. Antibodies in commercial kits used for heated/autoclaved meat samples and can be used on-site with results in about 15 min (Ballin, 2009). Substituting with cheap animal protein can also be detected with ELISA methods, while commercial kits don't necessarily meet the strict regulatory criteria (Ballin, 2010; Ballin, 2009). Lastly, LC methods based on protein profiles have been used for the detection of adulteration in meats. While -as mentioned previously- DNA-based methods are the most reliable methods, the limit of detection varies and not all of them are appropriate for quantification of adulteration (w/w). Unknown sample composition and processing procedures contribute to problems in correlating analytical results to meat content (w/w) (Ballin et al., 2009). For example, mitochondrial DNA in PCR provides a good limit of detection, but the number of mitochondria DNA copies varies in different tissues, making difficult to correlate content with PCR results. Other problems involve DNA extractability, DNA degradation and fat and water content. However, in real time PCR it is possible to extract quantitative results.

In all the above methods one or more of the following apply: they can be time-consuming, expensive, laborious, use harmful reagents, need expert laboratory staff and are dependent on rigorously following a standardized protocol to obtain accuracy

(Ding & Xu, 1999). As multispectral imaging requires only basic training in a user-friendly software, a few minutes for image acquisition and processing and no cost at all –excluding initial instrument and software purchase- it is only logical to explore the applicability of such a method. However, this approach does not reject the previous methods, but could work in conjunction with one of them for large scale and possibly on-line cases in order to reduce cost and increase the number of samples analyzed.

- Frozen-then-thawed meat:

In many cases, frozen-then-thawed meat is sold as fresh. This may not necessarily impact food safety *per se*; however it has an impact on the credibility of the food industries. Additionally, this case of fraud has obvious financial motives as the seller is profited doubly -first by extending a product's shelf-life and secondly by selling frozen meat at a higher price.

Various analytical methods have been proposed for the detection of frozen-then-thawed meat including enzymatic, DNA-based, microscopic and sensory techniques (Ballin & Lametsch, 2008). Again the most common are enzymatic and DNA-based techniques and the detection is based on the DNA damage occurring by freezing. In the first case, the most common method is the b-hydroxyacyl-CoA-dehydrogenase method (HADH) that takes advantage of the disruption of mitochondria, but is not appropriate for minced meat. In the second case, one method is the Comet assay on electrophoresis of lysed cells embedded in agarose on a microscopic slide which detects DNA damage or the more recent technology of real time PCR for measuring DNA fragmentation.

1.1.2 MEAT SPOILAGE & FACTORS AFFECTING SPOILAGE

Spoilage can be considered an ecological phenomenon where the proliferation of the microbial association of the stored meat -called specific spoilage organisms (SSO)- occurs. This phenomenon is accompanied by changes in the available components. It is actually the establishment of a maximum population called Ephemeral/ specific spoilage micro-organisms [E(S)SO] -an small fraction of the SSO- spoilage depends upon. The exponential nature of microbial growth and its resulting metabolism is associated with off-odors, slime etc., which are often described as meat spoilage (Nychas & Skandamis, 2005; Nychas, Skandamis, Tassou, & Koutsoumanis, 2008).

Five key factors determine the impact on quality/ spoilage -intrinsic, processing, extrinsic, implicit, and the emergent effect- and their combination contributes in spoilage.

Intrinsic factors are water activity, acidity, redox potential, available nutrients and natural antimicrobial substances, whereas extrinsic refer to environmental factors during storage, e.g. temperature, humidity and atmosphere composition. Physical or chemical treatments change the microbiota associated with the product. Implicit parameters are the result of the development of a microorganism which may have a synergistic or antagonistic effect on the microbial activity of other microbial communities present in the food product. Often, these individual factors interact to produce an combined effect on final meat quality (Argyri, 2010; Nychas & Skandamis, 2005; Nychas et al., 2008). From the above, it is clear that throughout the distribution chain the most important extrinsic factors are storage temperature and atmosphere composition connected to packaging:

- Temperature/ chill storage:

Temperature appears to be the most important factor that influences spoilage and chill storage conditions can effect significantly the type, composition and population of the microbiological association, resulting to the microbiota being dominated by psychrotrophs and delaying the onset of spoilage. No taxonomic restriction of phychrotrophic organisms is evident and mesophiles although cannot grow they are not necessarily killed (Koutsoumanis & Taoukis, 2005).

- Storage under aerobic conditions:

A consortium of bacteria responsible for spoilage in aerobic conditions is usually dominated by *Pseudomonas* spp. (especially *Ps. fragi, Ps. fluorescence, Ps. putida, and Ps. Ludensis*), causing slime and odor production. While *Brochothrix thermosphacta* and cold-tolerant *Enterobacteriaceae* (e.g. *Hafnia alvei*, *Serratia liquefaciens* and *Enterobacter agglomerans*), are known to occur on aerobically muscle foods they do not contribute in terms of population to the microbial associations. Lactic acid bacteria (LAB), although detected in aerobically spoiled chilled meat, are not considered to be important in beef (Nychas et al., 2008).

- Storage under vacuum or MAP (Modified atmosphere packaging):

The choice of atmosphere (usually a mix of oxygen, carbon dioxide and nitrogen) can increase product shelf life. Both vacuum and MAP conditions, change the microbiota, as the high carbon dioxide concentration inhibits pseudomonads and gram-positives, particularly LAB (e.g., *Lactobacillus*, *Leuconostoc*, *Lactococcus* and *Carnobacterium* spp.) typically develop on meat. Both LAB and *B. Thermosphacta* are the main,

causes of spoilage characterized by muscle souring (Argyri, 2010; Nychas et al., 2008).

1.2 SENSOR-BASED RAPID METHODOLOGIES

The product's history and origin, but also the by-products of the abovementioned metabolic activity of microorganisms, synthesize a unique biochemical profile/ fingerprint. The principle of rapid methods is that they could detect this fingerprint and apply this knowledge in order to indirectly evaluate quality and/ or safety (Ellis & Goodacre, 2001; Nychas et al., 2008).

A comprehensive review for the determination of food quality is presented in detail in section 1.2.4 regarding applications of MSI and FTIR and various types of meat and section 1.3.3 for the detailed presentation of various data analysis methodologies employed for various types of foodstuff. These methods refer to (i) arrays of biomimetic sensors (such as electronic noses), (ii) spectroscopy and particularly vibrational spectroscopy referring to Fourier transform infrared - FTIR and Raman spectroscopy and (iii) the so-called surface chemistry based on advanced imaging, i.e. hyper-/multi-spectral imaging – HSI/ MSI) (Argyri, Panagou, & Nychas, 2014; Kamruzzaman, Maniko, & Oshita, 2014; Loutfi, Coradeschi, Mani, Shankar, & Rayappan, 2015; Nunes, 2014; Qin, Chao, Kim, Lu, & Burks, 2013; Sun, Reddy Gangidi, & Proctor, 2009; Teena, Manickavasagan, Mothershaw, El Hadi, & Jayas, 2013; Velusamy, Arshak, Korostynska, Oliwa, & Adley, 2010; Xiong, Sun, Zeng, & Xie, 2014). This study focuses primarily on imaging and vibrational spectroscopy, particularly MSI and FTIR.

1.2.1 HYPERSPECTRAL & MULTISPECTRAL IMAGING

HSI and MSI principles are very similar and closely related with disciplines connected with Computer Vision. As with any optical technology, a sensor measures the interaction between light and molecules in the matter. The main advantage of this technique is that it is completely non-invasive, as image acquisition may theoretically take place without any interruption in the distribution chain since the sample is not destroyed. The main difference with simple every day images that imitate human vision and can be synthesized using three bands (red, green, blue - RGB colors) is that these types of images consist of tens (MSI) or hundreds (HSI) different wavelengths. Working in the visible range, the features obtained by computer vision include shape, color, size, and texture (Haralick & Shapiro, 1991). However, only occasionally is this method reported to be sufficient for detecting chemical and biological parameters. Indeed, various wavelengths are used by these instrument that go beyond visible and more commonly in the infrared (IR) spectrum.

The greatest difference between HSI and MSI is that HSI offers "continuous" for all purposes measurements from over a hundred wavelengths leading to high degree of collinearity in measurement values, whereas MSI offers measurements from tens of distinct wavelengths (usually up to 20), however missing the level of detail that HSI provides. This way, for every pixel (pixel: picture element) corresponding to an extremely small area in the actual sample surface, a distinct measurement is acquired. Therefore, the data acquired can be presented as a three dimensional (3D) data cube where a "spatial" and a "spectral" dimension are provided. Spatial meaning the measurement value for each pixel in a specific wavelength and spectral the intensity values of each pixel in the whole wavelength range (resembling a spectroscopic

profile), as shown in Figure 1.1(a). This way, they can also be used to generate "chemical maps" so as to show distributions of parameters of interest (Ropodi et al., 2016; Tsakanikas, Pavlidis, Panagou, & Nychas, 2016).

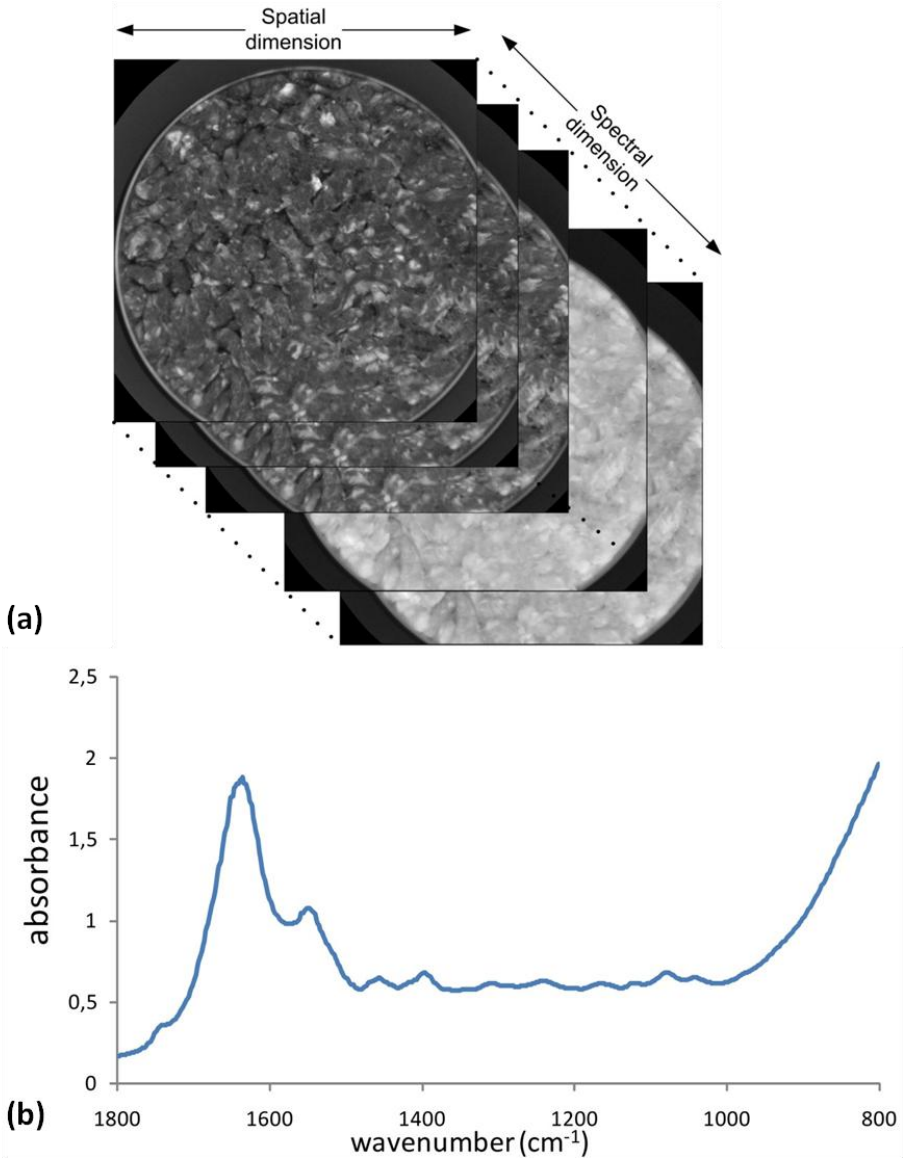


Figure 1.1 (a) Spectral and spatial dimensions of an indicative multispectral image of minced beef, (b) indicative FTIR spectrum of minced beef in the wavenumber range 1800-800cm^{-1}

Both these techniques have their own disadvantages: While HSI combines the merits of spectroscopy and computer vision, the rich information in HSI also results in difficulties in data processing, due to the high volume of data and the high degree of

collinearity observed. Advanced methodologies for image processing and wavelength selection have to be employed with HSI making it difficult for industrial online applications. In a sense, MSI is a simplified version where this problem is solved and is relatively cheaper to buy. However, the success of MSI is dependent on the actual wavelengths selected, and have to be chosen accordingly based on the application and the type of sample so as not to not to lose important information and/ or add bias in the measurement. In a way a successful MSI relies on previous knowledge and the efficiency of HSI for providing the important wavelengths. On the other hand, with HSI several options are available -near-infrared HSI, fluorescence HSI and Raman HSI- which provide great flexibility in finding solutions for all sorts of detection problems, versatility for creating wide applications in food inspection. Several review papers have been published regarding the applicability in different aspects of food quality (Gowen, Feng, Gaston, & Valdramidis, 2015; D. Liu, Zeng, & Sun, 2015).

1.2.2 VIBRATIONAL SPECTROSCOPY & INFRARED SPECTROSCOPY

Vibrational Spectroscopy (VS) is a collective term used to describe both analytical techniques; infrared and Raman spectroscopy. Infrared (IR) and Raman spectroscopy are considered non-destructive, slightly-invasive tools. The molecular composition, structure and interactions within a sample can be inferred by the absorbance/ intensity values using only a small percentage of the sample for spectroscopic measurements. Both techniques measure energy levels, which are associated with the chemical bonds in the sample and therefore characteristic, like a fingerprint. The original limitations in the low scanning process of IR instruments was overcome using Fourier Transform Infrared (FTIR) spectroscopy, which allows for measuring all of the infrared frequencies simultaneously, rather than individually. As infrared energy is emitted,

the beam is transmitted through or reflected off the surface of the sample and a detector measures the final signal.

IR and Raman spectroscopy have found a wide variety of applications in food quality analysis. In the case of IR, several studies have been applied in the evaluation of food spoilage, including not only meat but animal origin foods such as milk and cheese, and plant origin foods like wheat, fruit spirits and beer (Argyri et al., 2014; Damez & Clerjon, 2013). On the other hand, the studies reported for evaluating food spoilage through the use of Raman spectroscopy are rather limited, including food products such as meat and milk.

The resulting data for FTIR spectroscopy are two-dimensional, i.e. spectra consisting of a wavenumber and the value of the measured parameter. In Figure 1.1(b) an example of an FTIR spectrum of a minced beef sample is presented, where for a specific wavenumber a single value of absorbance is acquired.

1.2.3 IMAGE & FTIR PREPROCESSING

It is evident that MSI/ HSI data are extremely informative, but also very complex, so an image-preprocessing step has to be applied both in the spectral and in the spatial domain. The same applies for FTIR spectra.

Several methods for pre-processing before data analysis, and their combinations, have been proposed, as FTIR measurements are affected by noise and sometimes display baseline and scatter effects. Depending on the case, pre-processing methods have been known to vary, some of those being: first or second derivatives, Wavelet Transform (WT), detrending, Multiplicative Scatter Correction (MSC) or its extended form (EMSC), and Standard Normal Variate (SNV) transformation (Argyri et al., 2014;

Engel et al., 2013; Jarvis & Goodacre, 2005). Depending on the use-case, different methods provide different results (Biancolillo, Bucci, Magrì, Magrì, & Marini, 2014; Coppa et al., 2014). In fact, in a previous study, a genetic algorithm was employed in order to decide the appropriate method or combination of methods for FTIR spectra pretreatment (Jarvis & Goodacre, 2005).

Furthermore, as mentioned above, in the case of HSI the spectral bands are highly collinear and therefore have redundant information, while all images are susceptible to noise and other artifacts. Pre-processing includes -among others- radiometrics, spectral axis calibrations, removal of noise, blur and distortions. Noise in the spectra maybe decreased using median filtering and Savitzky Golay (SG) smoothing, which can also be applied in the spatial domain. Lastly, techniques described above for spectroscopy may be used to overcome unwanted spectral variation due to the natural morphology of food samples and/ or non-uniform lighting (Gowen et al., 2015; Kamruzzaman, Barbin, ElMasry, Sun, & Allen, 2012).

In the case of imaging however, an image-preprocessing step will be necessary for image segmentation and extraction of imaging features. The segmentation step is applied in order remove unwanted areas of an image, e.g. to remove image background and select a region of interest (ROI) (Teena et al., 2013). This step may include techniques ranging from simple thresholds to complex machine learning and artificial intelligence algorithms. After segmentation, basic features are extracted for further analysis. These consist of simple statistical measures, e.g. mean pixel intensity per wavelength, or more complex data, such as textural characteristics (Duchesne, Liu, & MacGregor, 2012; Haralick, Shanmugam, & Dinstein, 1973; Ma et al., 2015;

Panagou, Papadopoulou, Carstensen, & Nychas, 2014; Ropodi, Pavlidis, Mohareb, Panagou, & Nychas, 2015; Ropodi, Panagou, & Nychas, 2017).

1.2.4 LITERATURE REVIEW

In Table 1.1, the most relevant publications concerning meat in general and beef in particular using similar rapid analytical techniques for estimation of quality are presented along with the purpose of the study.

Table 1.1 List of representative studies using imaging or spectroscopy

Reference	Sensor type	Food type	Purpose
Alamprese et al., 2013	Spectroscopy	minced beef adulterated with turkey	Adulteration
Al-Jowder et al., 1997	Spectroscopy	minced chicken, pork & turkey	Frozen-then-thawed detection
Ammor et al., 2009	Spectroscopy	minced beef	Spoilage
Argyri et al., 2013	Spectroscopy	minced beef	Spoilage
Barbin et al., 2013	HSI	pork fillets	Frozen-then-thawed detection
Boyacı et al., 2014	Spectroscopy	beef adulterated with horsemeat	Adulteration
Ding & Xu, 1999	Spectroscopy	beef *vs*. kangaroo	Discrimination among types of meat
Dissing et al., 2013	MSI	minced pork	Spoilage
Downey & Beauchêne, 1997	Spectroscopy	beef fillets	Frozen-then-thawed detection
Ellis et al., 2005	Spectroscopy	chicken *vs*. turkey	Discrimination among types of meat
Feng & Sun, 2013	HSI	chicken fillet	Spoilage
Huang et al., 2013	HSI	pork fillets	Spoilage
Kamruzzaman et al., 2013	HSI	minced lamb adulterated with pork	Adulteration
Kamruzzaman et al., 2012	HSI	pork, beef and lamb, fillets & minced	Discrimination among types of meat

Ma et al., 2015	HSI	pork fillets	Frozen-then-thawed detection
Morsy& Sun, 2013	Spectroscopy	minced beef adulterated with pork, fat trimming & offal	Adulteration
Panagou et al., 2014	MSI	beef fillets	Spoilage
Puet al., 2015	HSI	pork fillets	Frozen-then-thawed detection
Rohman et al., 2011	Spectroscopy	beef meatballs adulterated with pork	Adulteration
Tsakanikas et al., 2016	MSI	beef fillets	Spoilage
Zhao et al., 2014	Spectroscopy	beefburgers adulterated with beef offal	Adulteration

In terms of spoilage prediction, both imaging and spectroscopy have been employed in recent years. In 2009, FTIR spectra were employed for spoilage estimation (total viable counts -TVC) using partial least-squares regression and for classification based on sensory quality (Ammor, Argyri, & Nychas, 2009). In addition to this, Argyri et al. (2013) utilized FTIR and Raman spectra for spoilage employing more advanced data analytics techniques. MSI has already been used in minced pork and beef fillets for spoilage estimation (Dissing et al., 2013; Efstathios Z. Panagou et al., 2014). Recently, MSI combined with advance computer vision technologies was employed for discrimination of meat samples based on microbiological criteria (TVC) and quantitative estimation of microbial counts during storage (Tsakanikas et al., 2016). Imaging (HSI) has also been applied in chicken as well as pork fillets (Feng & Sun, 2013; Huang, Zhao, Chen, & Zhang, 2013).

In the case of discrimination among different types of meat and adulteration some work has also been done apart from the studies mentioned in this thesis. In particular Vis-NIR spectroscopy was used for the discrimination between beef and kangaroo meat (Ding & Xu, 1999), Raman and FTIR spectra for chicken *vs*. turkey (Ellis,

Broadhurst, Clarke, & Goodacre, 2005) and HSI was used for the differentiation among pork, beef and lamb (fillets or minced) (Kamruzzaman et al., 2012).

In the case of adulteration, beef has been explored in several studies, as it is an attractive target: Using spectroscopy, Alamprese et al. (2013) explored the case of adulteration with turkey meat, while Morsy & Sun (2013) the case of adulteration with pork, fat trimming and offal. Adulteration with offal was also explored in beefburgers and meatballs by Zhao, Downey, & O'Donnell (2014) and Rohman, Sismindari, Erwanto, & Che Man (2011), respectively. One the other hand, Raman spectroscopy was used for beef adulteration with horsemeat (Boyacı et al., 2014), and HSI for lamb adulteration with pork (Kamruzzaman et al., 2013).

Furthermore, Al-Jowder et al. (1997) used spectroscopic data from minced turkey, pork and chicken to discriminate between fresh and frozen-then-thawed samples, whereas Downey & Beauchêne (1997) did the same for beef (*m. longissimus dorsi*) which was subjected to multiple (1-3) freeze-thawing cycles. The detection of frozen-then-thawed pork samples has been explored extensively with HSI (Barbin, Sun, & Su, 2013; Pu, Sun, Ma, & Cheng, 2015).

1.3 DATA ANALYSIS: CHEMOMETRICS, MACHINE LEARNING & COMPUTATIONAL INTELLIGENCE

With the development of sensors, a new type of "problem" emerges. Food scientists are now able to acquire a large amount of data, creating a data flood, but their analysis is exponentially complex. The new multivariate datasets now -simply represented- consist of a large number of variables (x-data) that correspond to an observed value or

category (y-data). In this study, y-data may be the microbiological measurements for a regression problem or a category for quality, (e.g. freshly-ground *vs.* frozen-then-thawed).

1.3.1 DEFINITIONS

As mentioned, the datasets display high complexity and are obviously difficult to inspect and visualize. For this reason, a deeper understanding of mathematics, statistics and computer science disciplines, disciplines beyond the scope of food scientists' and microbiologists' educations in the past, is necessary. The combination of these disciplines has brought the use of Chemometrics, machine learning, evolutionary computation/ computational intelligence methods in order to contribute in the visualization, dimensionality reduction, analysis, estimation of future results and presentation of each variable's contribution to the final result.

- Chemometrics:

It is defined as "the chemical discipline that uses mathematical and statistical methods, to (i) design or select optimal measurement procedures and experiments, and (ii) provide maximum chemical information by analyzing chemical or signal data generated by modern analytical instrumentation" (Otto, 2007). The term implies the data-driven extraction of information from chemical systems. In fact, due to continuously developed instrumentation, various chemometric methods were developed. Therefore, one of the initial aims of chemometrics was to make complicated mathematical methods practicable. Today many open-source and commercialized statistical and numerical software simplify this process, so that all important chemometric methods can be taught in appropriate computer

demonstrations (Brereton, 2014). In general, chemometrics is a highly interfacial discipline, using methods frequently employed in core data-analytic disciplines, such as multivariate statistics, applied mathematics, and computer science, to address problems not only in chemistry, but also to biochemistry, medicine, biology, chemical engineering and in this case-food science (Ropodi et al., 2016).

- Machine learning (ML):

The development and application of algorithms so as to "learn" from data consists the scientific discipline of machine learning. This way, inputs are used to construct a model in order to make predictions and/ or decisions (Kohavi & Provost, 1998). As a subfield of computer science, machine learning addresses problems in various fields and employs methodologies connected with fields, such as statistics and mathematics, which can go beyond explicit programming and overlap with computational intelligence. Machine learning has been used for various subfields and applications that could be exploited also in food science, e.g. computer vision.

- Computational intelligence/ evolutionary computation:

Computational intelligence is a well-established discipline, where theories inspired by sound biological examples have been evolving. The area is heterogeneous including technologies such as neural networks, fuzzy systems, evolutionary computation, swarm intelligence, probabilistic reasoning, multi-agent systems, etc. (Sumathi & Surekha, 2010).

Evolutionary computation is another subfield of computational intelligence where an iterative procedure based on the concept of evolution and natural genetics is used for the optimization of the applied algorithms. Specifically, algorithms deal with the

iterative growth of a population inspired by the mechanisms of evolution combined with a selection process, where a procedure is repeated multiple times until the stopping criteria are satisfied. Evolutionary computation has many applications in diverse fields, including food science, and provides highly optimized processes, but is also highly computationally intensive. Some of these algorithms include genetic algorithms (GAs), evolution strategies, evolutionary programming, and genetic programming (GP) (Argyri et al., 2013).

The aforementioned disciplines are very closely connected, sometimes overlapping and therefore difficult to separate. As, depending on the dataset, even the simplest solution may be enough, a data analyst should be familiar with all these approaches in order to offer the best approach. In this context, ML or computational intelligence should include even the simplest Chemometric methods.

1.3.2 CATEGORIZATION OF METHODS

All methods are divided in three major categories: supervised, unsupervised and reinforcement learning.

- Unsupervised learning consists of detecting data-driven underlying structures and groupings without any prior knowledge. It is based on cluster analysis (Everitt, Landau, Leese, & Stahl, 2011) and focuses on finding how similar is one sample to another using various approaches.
- In supervised learning, the objective is to model and/ or map the input variables (x-data) based on the output (y-data) (Goodacre, 2003). Supervised techniques, whether for regression or classification, tend to work well, as they take into account the actual knowledge of the target output. On the other hand,

model development and calibration are more complex and a number of hyper-parameters have to be adjusted, using e.g. cross-validation, grid search, based on criteria such as Root Mean Squared Error (RMSE). The process of developing the model is called calibration or training and may be simple or highly complicated depending on the method. After model training, the model can be used to predict the category or the output value of a new sample.

- Reinforcement learning is when the model interacts in the context of the dynamic environment and improves its performance based on goal without explicitly "knowing" if the goal is achieved (Sutton & Barto, 1998). However, this approach has not been used in the food industry yet and is not within the purposes of this study.

1.3.3 PRESENTATION OF METHODS

In the case of unsupervised techniques, undoubtedly Hierarchical Cluster Analysis (HCA) and Principal Component Analysis (PCA) have been utilized in several food applications (Argyri et al., 2014). In the first, a distance metric (not necessarily Euclidean) is employed in order to calculate the distance among objects. The results are further processed with an agglomerative distance algorithm leading to the construction of a dendrogram that connects the samples according to their similarities to each other (Everitt et al., 2011). PCA, on the other hand, is one of the most extensively employed methodologies. The reason for its popularity is that it can manipulate a multivariate dataset and "translate" it into a dataset of orthogonal components, called "Principal Components" (PCs). The order of the components is determined by the percentage of variability explained in a descending order (Jolliffe, 2002). One of the most common applications of PCA is the use of PCA score plots

(for the most important components) so as to help visualize the samples using a 2D or 3D projection of the samples. That way the analyst may observe groups of samples or outlying samples, indicating similarities or dissimilarities. Furthermore, it is a widely used method that enables a significant reduction of dimensionality, choosing uncorrelated variables, i.e. the PCs explaining most of the variance in the dataset. Furthermore, it can be used in outlier detection and highlighting important features to be used for further analysis, a process known as variable selection (Argyri et al., 2014; Goodacre, 2003).

Supervised techniques consist of developing models for qualitative or quantitative estimation based on previous data and their corresponding category or value respectively. In detail, the estimation of the categories refers to classification models, whereas value estimation/ prediction refers to regression models. For example, Discriminant Function Analysis (DFA) or Canonical Variate Analysis (CVA) is a cluster analysis based method that is used to classify individuals into two or more predetermined groups (Berrueta, Alonso-Salces, & Héberger, 2007). As far as linear regression methods are concerned, some extensively used in food applications are Multiple Linear Regression (MLR), Principal Component Regression (PCR) and Partial Least Squares Regression (PLSR) (Goodacre, 2003). Indeed, MLR is the simplest form of linear regression, based on least squares, but it has problems with collinearity. While PCR uses PCA to extract the new components (PCs), PLSR projects both observed and predicted values in a feature space and a linear regression model is established (de Jong, 1993; Wold, Sjöström, & Eriksson, 2001). It is also considered a linear method in the sense that the new components are linear combinations of the original variables (Boulesteix & Strimmer, 2007). The latter is a

very common approach and has many variations for variable selection and interval selection methodologies, e.g. interval partial least squares (iPLS), windows PLS and iterative PLS (Xiaobo, Jiewen, Povey, Holmes, & Hanpin, 2010). Lastly, coupled with Linear Discriminant Analysis (LDA), it can be extended to PLSDA for classification purposes (Barker & Rayens, 2003).

While the previous methods have been extensively implemented, lately methods based on more complicated algorithms involving machine learning and computational intelligence have been introduced. Some methodologies found in research articles and reviews are Artificial Neural Networks (ANNs), Support Vector Machines (SVMs) and evolutionary-based algorithms, including Genetic Algorithms (GAs) and Genetic Programming (GP). The latter algorithms are used for optimization purposes. ANNs are inspired from the biological paradigms of the human brain and the function of neurons and can be used for both classification and regression purposes. The building block of every neural network is a neuron where each of its inputs is multiplied by a connection weight and -in the simplest case- these products are simply added together, fed through a transfer function to generate an output. As there are various ways that these neurons can be connected, thus creating diverse network topologies, learning rules, transfer functions, summation functions is very important (Sumathi & Surekha, 2010). The most commonly applied ANNs are Multilayer Perceptrons (MLPs) and Radial Basis Function (RBF). Another ML method is SVMs that maps the input space in to a higher dimensional feature space using a kernel function so as to construct a maximum separation hyper-plane (Cortes & Vapnik, 1995). From the usual choices for kernel functions, namely linear, polynomial, sigmoid, and radial basis function (RBF), linear and RBF are the most common. There are also SVM regression (SVR)

models for value estimation purposes (Balabin & Lomakina, 2011). For a given regression problem, the goal of SVM is to find the optimal hyper-plane from which the distance to all the data points is minimum. The performance of SVMs depends on several factors including the kernel function type and its corresponding hyperparameter(s). Other approaches include ensemble modeling, the rationale of which is developing several of classifiers where the final prediction is a result of combining the individual prediction of the classifiers, e.g. Random Forest (RF) algorithm (Breiman, 2001; Liu, Wang, Wang, & Li, 2013). Lastly, the evolutionary-based algorithms are iterative processes mimicking growth or development in a population. GAs use Darwin's "survival of the fittest" strategy and reproduction operators to select the optimal set of features or parameters. A given GA can be used for feature selection and/ or the optimal adjustment of parameters, but these features or parameters are evaluated by a predetermined function, e.g. using a PLS method (Luke, 2003).

In Table 1.2, representative examples of machine learning methodologies combined with various sensors for different types of foodstuff are presented.

Vibrational spectroscopy, imaging, and e-nose have been used to predict microbial quality by Huang et al. (2014). The study attempted to measure total volatile basic nitrogen (TVB-N) content in pork meat and yielded a coefficient of determination equal to 0.9527 in the prediction set. Principal component analysis (PCA) was used to achieve data fusion based on these characteristic variables from 3 different sensors data and back-propagation artificial neural network (BP-ANN) was used to construct the model for TVB-N content prediction. Compared with single technique, integrating the three techniques, in this paper, has its own superiority.

Table 1.2 List of representative rapid method applications with their corresponding data analysis methodologies. Modified from Ropodi et al. (2016).

Reference	Sensor type	Food type	Purpose	Data analysis method
Alamprese et al., 2013	Spectroscopy	beef, turkey	Detection of adulteration	PCA, LDA, PLSR
Argyri et al., 2013	Spectroscopy	Minced beef	Spoilage/ sensory	PLSR,GA-GP, GA-ANN, SVR (various kernel function)
Coppa et al., 2014	Spectroscopy	Milk	Fatty acid composition	PLS and modified PLS (MPLS)
da Costa Filho, 2014	Spectroscopy	Edible oils	Trans-fatty acid determination	PLSR
Ellis et al., 2005	Spectroscopy	Chicken, pork, turkey, lamb and beef	Authentication of species and the distinct muscle groups within these species	PC-DFA, GA-MLR
Fu et al., 2014	Imaging	Milk powders	Adulteration	Spectral similarity measures
Gowen et al., 2008	Imaging	Mushrooms	Bruise detection	PCA
He, Sun, et al., 2014	Imaging	Salmon	Spoilage (LAB)	LS-SVM
Huang et al., 2014	Imaging, e-nose, spectroscopy	Pork	Freshness (TVB-N content)	PCA, BP-ANN
Liu et al., 2013	e-tongue	Orange beverage, Chinese vinegar	Authentication/ discrimination among brands	PCA, BP-ANN, SVM, RF
Mohareb et al., 2016	e-nose	Beef fillets	Spoilage/ sensory	Ensemble SVM & SVR
Pan et al., 2014	e-nose	strawberry	Detection of fungal disease	PCA, ANN (MLP)
Panagou et al., 2008	e-nose	Table olives	Sensory	PCA, HCA, DFA, MLP-NN
Panagou et al., 2014	Imaging	Beef fillets	Spoilage	HCA, PLSR, PLS-DA
Papadopoulou et al., 2013	e-nose	Beef fillets	Spoilage/ sensory	PCA-DFA SVM & SVR (RBF kernel)
Qiu et al., 2014	e-nose, e-tongue	Strawberry juice	Discrimination among processing approaches	LDA, PLSR, SVM, RF
Sharifzadeh et al., 2014	Imaging	Meat	Monitoring meat color	PCA *vs*. linear, non-linear and kernel-based regression methods (ANNs & SVM)

Tsakanikas et al., 2016	Imaging	Beef fillets	Spoilage (TVC)	SVR
Wilson et al., 2013	e-nose	Catfish fillets	Sensory/ off-flavor detection	PCA, ANN, QF
Wu et al., 2013	Imaging	prawn	Detection of adulteration	UVE-SPA-LS-SVM
Zakaria et al., 2012	e-nose, acoustic sensor	mangoes	Ripeness/ maturity	PCA, LDA, LDA- ANN
Zhang et al., 2012	e-nose	peach	Firmness, sugar content, acidity	PCA, LDA, PCR, PLSR
Zhao et al., 2014	Spectroscopy	beefburger	Detection of adulteration	PLS-DA, SIMCA, low & mid level fusion strategies based on PLS

A variety of methodologies was employed by Argyri et al. (2013) including PLSR, GP, GA, ANNs and SVM regression (SVR) with linear, polynomial, radial basis and sigmoid kernel functions for the comparison of FTIR and Raman spectroscopic techniques based on microbial and sensory data from minced beef samples. Results indicated a slightly better performance for FTIR models, while PLS and SVM models performed better in predicting microbial counts. Sensory categories were better estimated with the GA-GP model using the FTIR data, whereas the GA-ANN model performed better in predicting the sensory scores using the Raman data. Panagou et al. (2014) studied the potential of MSI in the visible and shortwave near infrared area (405–970 nm) in assessing the microbial quality of beef fillets stored at different isothermal conditions, using HCA, PLSR for prediction and PLSDA for discrimination among 3 microbial classes. It is interesting to note that models were validated with independent test samples stored at two new temperatures with satisfactory results for model calibration and validation. The same sensor was employed in order to discriminate between different microbial classes and estimate

TVC for beef fillets achieving good performance with overall correct classification rate for the two quality classes ranging from 89.2% to 80.8% for model validation and for the calculated regression results an R-square of 0.98 (Tsakanikas et al., 2016). In another case, imaging –specifically HSI- has also been employed to evaluate spoilage of farmed salmon by lactic acid bacteria (LAB) during cold storage applying the LS-SVM and Competitive adaptive reweighted sampling (CARS) algorithm in order to reduce spectral redundancy and identify the most informative wavelengths related with LAB prediction (He, Sun, & Wu, 2014). The derived LS-SVM model using 239 wavelengths yielded a regression coefficient of prediction Rp equal to 0.929 and RMSEP 0.515, very similar to the CARS-LS-SVM model using only 8 wavelengths (R_p= 0.925, RMSEP= 0.531). SVMs were used for both classification and regression of microbial and sensory characteristics respectively coupled with electronic nose measurements (Papadopoulou, Panagou, Mohareb, & Nychas, 2013). The resulting SVR model exhibited a mean correlation coefficient between observed and predicted counts in the test datasets equal to 0.863 for TVC, whereas in the case of the sensory categories fresh, semi-fresh, and spoiled, the average sensitivity values were 85.7, 87.3, and 88.9%, respectively. PCA followed by discriminant function analysis (DFA) were also employed but results were not satisfactory. The previous SVM methodology was further extended by Mohareb et al. (2016) where an ensemble learning method for SVMs was employed. Indeed, results were improved yielding an overall accuracy of 84.1% compared to 72.7% in the case of the single SVM model. E-nose has been applied for the discrimination of table olives' quality based on sensory score, using ANNs and specifically MLPs, where the classification accuracy for acceptable, marginal and unacceptable samples was ca. 90%, 78% and 52%, respectively (Panagou et al., 2008).

Imaging techniques have also been used in order to recognize and/or discriminate based on external features, e.g. for bruise detection in mushrooms Gowen et al. (2008) by two classification methodologies based on PCA: (a) applying PCA to entire hypercube and classify based on the 2nd PC, and (b) multiplying hypercube by loading function and classifying the resultant virtual image. Results showed that using virtual prediction image resulted in better classification in all cases except for the classification of one category, where the first and second method resulted in 100% and 90% correct classification, respectively. Furthermore, multispectral imaging has been used as a method for monitoring meat color (Sharifzadeh, Clemmensen, Borggaard, Støier, & Ersbøll, 2014). The authors compared PCA with a wide range of linear, non-linear, kernel-based regression and sparse regression methods coupled with variable selection methodologies, and linear ridge regression combined with the proposed elastic net-based feature selection strategy provided the best results. Other applications include assessment of meat tenderness and freshness, moisture and firmness of fruits, etc. (Dai, Sun, Xiong, Cheng, & Zeng, 2014), where the majority of these applications involve the use of PLSR, but also other methods mentioned above. PLSR and PCR were also used in order to predict quality indices of peaches, such as firmness, acidity and sugar content based on e-nose measurements (Zhang, Wang, Ye, & Chang, 2012). Models were carefully chosen so as to achieve high performance. In fact, PLS models displayed correlation coefficients ranging from 0.83 to 0.86. Additionally, a low-level fusion of e-nose and acoustic sensor data was also performed by Zakaria et al. (2012) in order to improve classification among different levels of ripeness and maturity of mangoes. Results were improved in the case of LDA after the fusion process, with 99.8% of grouped cases classified correctly after LOOCV. However, very promising testing results were achieved when an

unsupervised Competitive Learning ANN was employed as the prediction accuracy ranged from 66.7% to 84.4%. E-nose sensors have also been applied for the discrimination between good-flavor *vs.* off-flavor catfish meat samples (Wilson, Oberle, & Oberle, 2013) applying PCA for exploratory analysis and a statistical algorithm called Quality Factor (QF) analysis that determines statistical distances between profiles of classes measured using Euclidean distance. In addition, a profile measurement library was developed based on ANNs for class prediction and 91.4% of all catfish samples were classified correctly.

As fatty-acid composition is an important quality parameter, chemometric methods (PCA and PLS variants) have been applied to spectroscopic data in order to predict fatty-acid composition in milk and barley and trans-fatty acid in edible oils (Coppa et al., 2014; Cozzolino, 2014; da Costa Filho, 2014). In the case of fresh and thawed milk (Coppa et al., 2014) regressions were calculated with both partial least square (PLS) and modified partial least square (MPLS) with various preprocessing steps, while PCA was used for feature selection based on loadings. Most models showed high performance indicating that near and medium-IR spectroscopy could be used for routine milk FA composition estimation.

Some work has also been done in food quality assessment and MLPs. MLP networks have been applied for detection and classification of pathogenic fungal diseases in strawberries (Pan, Zhang, Zhu, Mao, & Tu, 2014), with three common pathogenic fungi. Decay was evident with PCA and MLP prediction accuracy of the fungal infection type for the four groups reached 96.6%. HSI has several other applications in food safety, e.g. detection of parasitic nematodes (cod worms, seal worms, whale worms) in fish fillets (Kamruzzaman, Makino, & Oshita, 2015), while various

methods have been presented for the detection of melamine in milk and milk products based on IR spectroscopy (Domingo, Tirelli, Nunes, Guerreiro, & Pinto, 2014), as well as HSI (Fu et al., 2014). Fu et al. (2014) compared milk powder and melamine samples with samples of milk-melamine mixtures (melamine concentrations ranging from 0.02% to 1%) using three different spectral similarity measures, (a) spectral correlation measure (SCM), (b) spectral angle measure (SAM) and (c) Euclidean distance measure (EDM). All measures proved to have similar performance and in most cases melamine particles were detected.

In addition, rapid methodologies have been widely used for authenticity claims (Oliveri & Downey, 2012). Specifically, there are several studies for authenticity and adulteration of edible oils and fats (Nunes, 2013). Furthermore, Ellis et al. (2005) applied PC-DFA between muscle foods using both Raman and FT-IR spectroscopy for the discrimination among closely related poultry species, chicken and turkey and distinct muscle groups. A GA-MLR methodology was also utilized and results showed very good discrimination. Alamprese et al. (2013) investigated minced beef adulteration with turkey meat by UV-visible (UV-vis), NIR and MIR spectroscopy and PLS techniques proved better compared to Linear Discriminant Analysis (LDA). In the first case, PLS regression was employed for the prediction of the percentage of turkey meat adulteration in minced beef, while better results were obtained with NIR and MIR spectroscopy. In addition, Wu et al. (2013) studied the adulteration of prawn with gelatin-like chemicals. The combination of uninformation variable elimination (UVE) and successive projections algorithm (SPA) followed by LS-SVM reduced the number of wavelengths drastically (from 462 to 13) and achieved a coefficient of determination of prediction equal to 0.965. Raman spectroscopy coupled with PCA

was employed for rapid determination of beef adulteration with 0, 25, 50,75, 100% w/w horsemeat presenting a good discrimination among adulteration levels (Boyacı et al., 2014). Offal-adulteration of fresh and frozen beefburger products has been investigated using MIR spectroscopy and chemometric data analysis (Zhao et al., 2014). Both discriminant (PLSDA) and class-modeling (SIMCA) methods were used and the former achieved 100% correct classification accuracies for fresh and frozen-then-thawed material. Other articles have explored the possible discrimination among different processing approaches, i.e. in strawberry juice using e-nose and e-tongue instruments coupled with LDA, PLSR, RF, and SVM. Lastly, Liu et al. (2013) uses an e-tongue instrument coupled with RF, SVM and BP-ANN for type and brand recognition of orange beverage and Chinese vinegar on 4 diverse datasets. For each data set, the performance parameters of RF are superior to those of SVM and BPNN.

A more extensive presentation of methods along with calibration techniques and validation criteria is presented in Chapter 5 in connection with the results of this thesis.

1.4 OBJECTIVES OF THIS STUDY

The overall objective of this study is to determine the applicability of rapid methods in the determination of minced meat quality. For this reason, two different rapid analytical techniques will be applied, i.e. multispectral imaging and FTIR for various cases. In terms of quality, detection of non-compliance to label, whether accidental or in purpose, and estimation of the microbial spoilage are explored. The case of non-compliance is investigated in connection (a) two extensive experiments involving

minced beef adulteration, and (b) a study for the detection of frozen-then-thawed minced beef. Lastly, microbial spoilage of minced beef is investigated in two separate cases. In addition to this, different data analytics methodologies are explored and improved upon using advanced methodologies where necessary. Special care is taken to apply model validation and to avoid overoptimistic results utilizing the appropriate methods for calibration and independent validation where possible. This study includes:

i. Multispectral images in 18 different wavelengths of 220 meat samples in total from four independent experiments (55 samples per experiment) where the appropriate amount of -beef and pork- minced meat was mixed in order to achieve nine different proportions of adulteration and two categories of pure pork and beef.
ii. Multispectral images of 110 samples from three different batches of minced beef and horsemeat in 18 wavelengths, with additional images captured after 6, 24 and 48 hours to explore how model performance is affected by changes in meat color during storage.
iii. Multispectral images and FTIR spectra from seven different batches of freshly-ground beef, MSI and FTIR spectral measurements after being frozen (-20°C) for 7 and 32 days and then thawed. In total, 105 multispectral images and FTIR spectra.
iv. Multispectral images and FTIR spectra of minced beef samples stored under different temperature and packaging conditions (MAP *vs.* aerobic conditions) acquired at regular intervals and combined with microbiological

measurements (Pseudomonads, Lactobacilli, *B. thermosphacta* and Enterobacteriaceae, as well as TVC). In total, 105 samples.

v. FTIR spectra of 168 minced beef samples stored at under different temperature and packaging conditions (MAP *vs.* aerobic conditions) acquired at regular intervals and combined with microbiological measurements for TVC.

CHAPTER 2:

MINCED BEEF ADULTERATION WITH PORK AND HORSEMEAT

2 ABSTRACT

In this chapter, the use of rapid analytical techniques for the detection of minced beef adulteration is explored, which in turn has led to two publications, a modified version of which is presented in the following sections. The publications are mentioned below:

- Ropodi, A. I., Pavlidis, D. E., Mohareb, F., Panagou, E. Z., & Nychas, G.-J. E. (2015). Multispectral Image Analysis approach to detect adulteration of beef and pork in raw meats. *Food Research International*, *67*, *12–18*. http://doi.org/10.1016/j.foodres.2014.10.032, in which the author of this thesis performed the data analysis and model development, as well as contributed in the experimental design and preparation of the submitted manuscript.
- Ropodi, A. I., Panagou, E. Z., & Nychas, G.-J. E. (2017). Multispectral imaging (MSI): A promising method for the detection of minced beef adulteration with horsemeat. *Food Control*, *73*, 57–63. http://doi.org/10.1016/j.foodcont.2016.05.048, in which the author of this thesis performed the experiment, the data acquisition and analysis, as well as contributed in the experimental design and preparation of the submitted manuscript.

Both publications as well as their supplementary files are presented in Appendix I.

Briefly, the first study on minced beef adulteration involves the adulteration with pork. Emphasis is given on the use multiple meat batches and levels of adulteration,

as well as the validation of the developed model with the use of independent samples. The aim of this study was to investigate the potential of multispectral imaging supported by multivariate data analysis for the detection of minced beef fraudulently substituted with pork and vice versa. Multispectral images of 220 meat samples in total from four independent experiments were acquired for this work. The appropriate amount of beef and pork minced meat was mixed in order to achieve nine different proportions of adulteration and two categories of pure pork and beef. PLSDA and LDA were used so as to discriminate among all adulteration classes, as well as among adulterated, pure beef and pure pork samples. Results showed very good discrimination between pure and adulterated samples, for PLSDA and LDA, yielding 98.48% overall correct classification. Additionally, 98.48% and 96.97% of the samples were classified within a ±10% category of adulteration for LDA and PLSDA respectively. Lastly, the models were further validated using the data of the fourth experiment for independent testing, where all pure and adulterated samples were classified correctly in the case of PLSDA.

In the second publication, the case of detection of minced beef adulteration with horsemeat using multispectral imaging is introduced. Based on the results of the previous study, a new parameter was added in the analysis: the time in refrigerated storage of the meat samples. Multiple multispectral images per sample were acquired at different time intervals and the effectiveness of the algorithms was tested. Specifically, multispectral images of 110 samples from three different batches of minced beef and horsemeat were acquired. Images were taken again after samples were stored at 4°C for 6, 24 and 48h. Classification models (PLSDA, RF, SVM) based on the first two batches were developed while the third batch was set aside for

external/ independent validation. Results showed that freshly-ground and stored samples were clearly distinguishable, whereas classification model performance for detection of adulterated samples was significantly affected by changes in meat color during storage. Using a two-step SVM model however, all pure and freshly-ground samples were classified correctly and the overall correct classification was equal to 95.31% for independent batch validation.

2.1 INTRODUCTION

In recent years, European consumers are increasingly demanding information and reassurance not only on the origin but also on the content of their food. Protecting consumer rights and preventing fraudulent or deceptive practices, such as food adulteration, have become a major priority for food monitoring agencies and the food industry worldwide. While manufacturers are required to provide and confirm the authenticity and point of origin of food products and their components, adulterants are detected with great difficulty in the context of methods commonly applied in laboratories, since most adulterants are unknown and unpredictable (e.g., horsemeat). Several standard analytical techniques, such as immunological and enzymatic techniques, DNA and protein based assays and triacylglycerol analysis have been applied to authenticate food commodities (Ballin, 2010; Soares, Amaral, Mafra, & Oliveira, 2010). These methods are usually capable of detecting low levels of adulteration (Ballin, 2010), but they are expensive, invasive, sophisticated, laborious, and technically demanding (Ding & Xu, 1999).

Meat and meat products can be attractive targets for adulteration in many ways, including substitution or partial substitution of high commercial value meat with cheaper, such as pork or offal or by adding proteins from several origins (Kamruzzaman et al., 2013; Tian et al., 2013). With minced beef being the basic ingredient for burgers, adulteration of minced beef involves economic, quality, safety and socio-religious issues (Alamprese et al., 2013). For this reason, the meat industry needs methods that will screen food samples for contaminants in a rapid and cost efficient way for large-scale in-, on- or at-line applications in order to provide proof of origin and prevent deliberate or accidental undeclared admixture to food samples.

Hyperspectral and multispectral imaging have been used as rapid techniques to monitor quality attributes of food products (Wu & Sun, 2013b). The former has been used for the rapid detection of total viable counts in pork (Barbin, Sun, et al., 2013; Huang et al., 2013) and of the water-holding capacity of fresh beef (ElMasry, Sun, & Allen, 2011) and pork (Prevolnik, Čandek-Potokar, & Škorjanc, 2010). Meanwhile, multispectral image analysis has high potency for the evaluation of food quality systems during handling, processing and storage (Løkke, Seefeldt, Skov, & Edelenbos, 2013) and it has been previously used for the conversion of meat colour in L*, a*, b* values (Sharifzadeh et al., 2014) and for quality assessment of beef and pork (Dissing et al., 2013; Panagou et al., 2014).

Despite the fact that hyperspectral imaging has been used for the detection of minced lamb adulteration (Kamruzzaman et al., 2013) and gelatine adulteration in prawn (Wu et al., 2013), to the best of our knowledge the use of multispectral image analysis for meat adulteration, especially in the case of minced beef with pork, has never been previously explored before Ropodi et al. (2015). In terms of minced beef adulteration with horsemeat, Raman spectroscopy has been applied with promising results (Boyacı et al., 2014; Zając, Hanuza, & Dymińska, 2014), but MSI has not been used previously in the case of minced beef adulteration with horsemeat. Furthermore, no comparison has been performed so far between freshly-ground meat and meat stored in refrigerated conditions where changes in meat color naturally occur.

In both cases, multispectral imaging is introduced as a new approach in tandem with advanced statistical approaches, for the discrimination of raw minced beef, which has been fraudulently substituted or combined with raw minced pork or horsemeat. In the first case, the objective of the study was to (a) evaluate the potential use of

multispectral imaging to discriminate pork from beef, (b) identify if possible, the lowest percentage of minced pork adulteration in minced beef that can be safely detected, establishing a rapid and non-invasive technique for rapid and accurate results. In the second case and based on the acquired knowledge, the objective of the study was to (a) evaluate the potential of multispectral imaging in tandem with data analysis techniques to identify and/or quantify horsemeat in minced beef, and also (b) explore model performance under refrigerated storage of meat samples.

2.2 MATERIALS & METHODS

2.2.1 IMAGE ACQUISITION & SEGMENTATION

Multispectral images were captured using "VideometerLab", a system which acquires multispectral images in 18 -non uniformly distributed- different wavelengths ranging from 405 to 970 nm, i.e. 405, 430, 450, 470, 505, 565, 590, 630, 645, 660, 850, 870, 890, 910, 920, 940, 950 and 970 nm. The system has been developed by the Technical University of Denmark and commercialized by "Videometer A/S" (Carstensen & Hansen, 2003; http://www.videometer.com). A detailed description of the instrument has been reported previously (Dissing et al., 2013; Panagou et al., 2014). The advantage of this instrument is that it not only uses the information of visible and short-NIR spectral regions, but moreover uses the spatial information of each pixel. The acquisition system records surface reflections with a standard monochrome charge coupled device chip, nested in a Point Grey Scorpion camera. The object of interest is placed inside an Ulbricht sphere in which the camera is top-mounted. The sphere has its interior coated with a matt coating, which together with the curvature of the sphere ensures a uniform reflection of the cast light. At the rim of the sphere, light

emitting diodes (LEDs) are positioned side by side and are turned on successively for each wavelength.

The system was calibrated radiometrically and geometrically using well-defined standard targets, after a light setup based on the type of object to be recorded (Folm-Hansen, 1999) called "autolight". In autolight, it is always the brightest sections in the image that dictate the final result.

The resulting image includes redundant information, such as the Petri dish and its surrounding background, as well as the fat and connective tissue of the meat. For this reason an image-preprocessing step is needed that will result in an image mask where only meat tissue is included. This step, which includes transformation and segmentation procedures, was implemented using the respective routines of the VideometerLab software (version 2.12.39) that controls the operation of the instrument. Canonical Discriminant Analysis (CDA) was employed as a two-step supervised transformation building method to divide the images into regions of interest (Daugaard, Adler-Nissen, & Carstensen, 2010). Following this transformation, the separation was distinct and a simple threshold was enough to separate adipose from lean tissue and samples from background pixels. The result of this processing is a segmented image for each meat sample with the isolated part of the meat tissue as the main region of interest (ROI) to be used for the extraction of spectral data that were further employed in statistical analysis. The procedure is graphically presented in Figure 2.1.

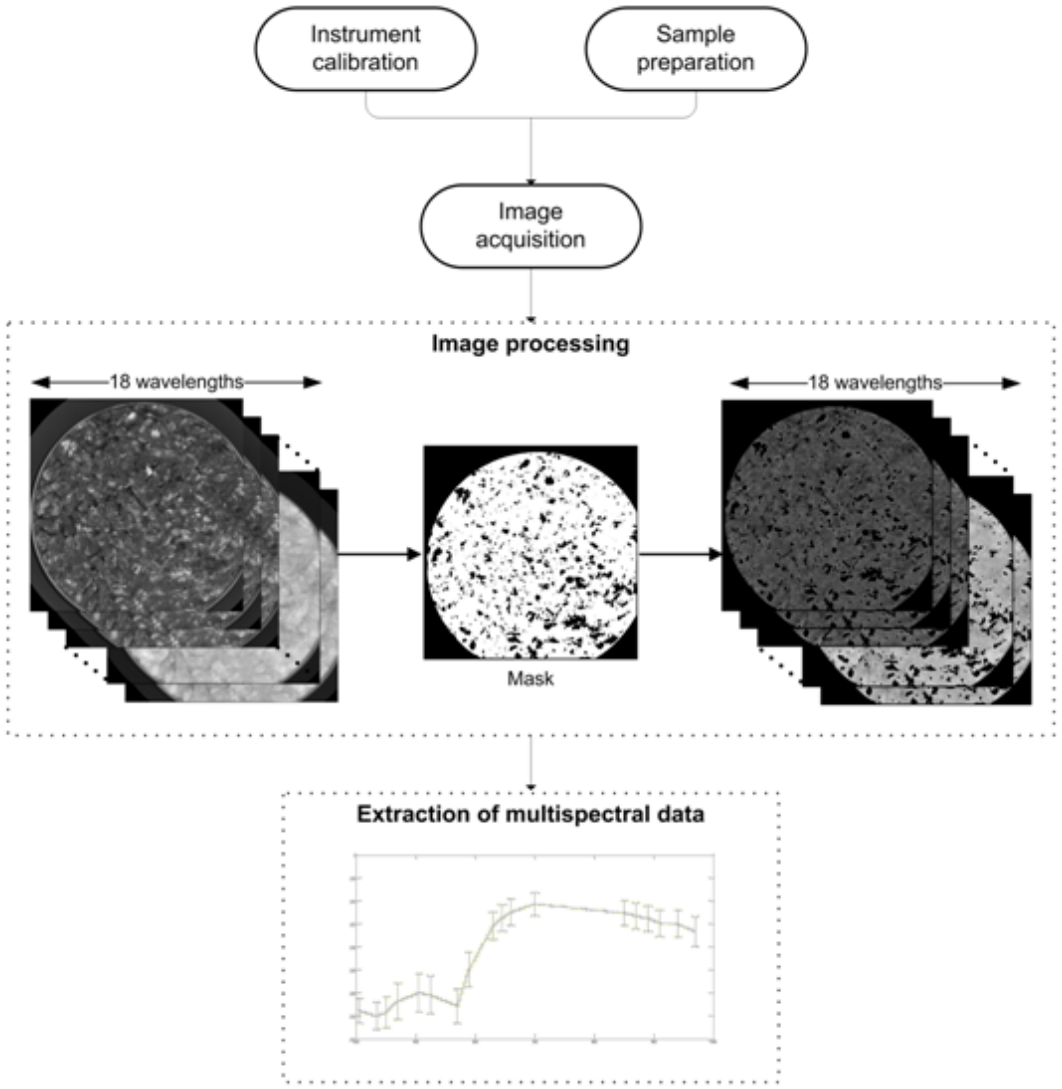


Figure 2.1 Process of multispectral image acquisition and extraction of data

2.2.2 DATA ANALYSIS

For each image, the mean reflectance spectrum was calculated by averaging the intensity of pixels within the ROI at each wavelength. Furthermore, the standard deviation of the pixels' intensity per wavelength was extracted. The resulting data consisted of 18 mean values and 18 standard deviations of the reflectance (36 variables) and were further analysed with various classification methods.

In the following sections, the data analysis methodologies employed in one or both cases include in the case of unsupervised techniques:

- Hierarchical Cluster Analysis (HCA) (Everitt et al., 2011) was performed as an unsupervised technique to explore the relationship between variables and adulteration classes, using Euclidean Distance and Ward's minimum variance agglomeration method.
- Principal Component Analysis - PCA (Jolliffe, 2002) was performed so as to visualise whether there were significant differences among samples from different batches, as well as among different classes.

Furthermore, various supervised classification techniques were employed in order to discriminate among different levels of adulteration and other classes, such as freshly-ground *vs.* stored minced meat samples and pure *vs.* adulterated samples. These techniques are presented below:

- The widely used Linear Discriminant Analysis (LDA) (Fisher, 1936) and partial least squares discriminant analysis (PLSDA) (Barker & Rayens, 2003; de Jong, 1993).
- Random forest (RF) (Breiman, 2001), a supervised learning algorithm which uses an ensemble of classification trees. Ensemble methodologies involve generating multiple classifiers and aggregating their results (e.g. bagging) (Breiman, 2001; M. Liu et al., 2013).
- Support Vector Machines (SVM) (Cortes & Vapnik, 1995), a method that maps the original data points from the input space into a higher dimensional

feature space using a kernel function, in order to construct a maximal separating hyper-plane. Various kernel functions were employed.

The details for model calibration and validation are presented per case in the following sections.

2.3 ADULTERATION WITH PORK

2.3.1 EXPERIMENTAL DESIGN

Different levels of adulteration of minced beef and pork were prepared as described below; Fresh beef and pork fillets *Longissimus* muscle of normal pH (5.6–5.8) were purchased from butcher shops in Athens and transported under refrigeration to the laboratory within 30 min. The fillets were cut into smaller pieces and grinded separately one at a time, using a domestic meat-mincing machine. The machine parts coming in contact with the meat were initially disinfected by washing with detergent and hot water, and rinsing with pure ethanol. To achieve different levels of adulteration, ranging from 10 to 90% with a 10% increment, the appropriate amount of each type of meat was used and mixed in conditions that simulate industrial processing. From each level of adulteration, five different portions of ca. 75-80 g were placed in Petri dishes and snapshots were taken using VideometerLab vision system (Videometer A/S, Hørsholm, Denmark). For every level of adulteration (nine categories of mixed meat and two categories of pure pork and beef), each Petri dish was considered as a replicate in the experiment (5 x 11 samples in total per experiment).

The experimental procedure took place aseptically and was repeated four times. For every random Petri dish in each level of adulteration, a different autolight procedure was employed. One hundred and sixty five (165) samples from three independent experiments (i.e., 55 samples per experiment, from this point on referred as samples from batches 1, 2, and 3) were used to develop the model and 55 samples from the fourth experiment (batch 4) were employed for the purpose of external validation. It should be noted that 220 samples from different batches were analysed in total.

2.3.2 DATA ANALYSIS

Two methods, PLSDA (Barker & Rayens, 2003; de Jong, 1993) and LDA (Fisher, 1936) were performed in order to discriminate among all adulteration classes (11 in total), as well as among adulterated, pure beef and pure pork samples. The dataset was partitioned in two sets: the training set used for model calibration and the test set used for validation. A 60-40% stratified partition was applied on the first three batches, meaning 60% of the dataset was chosen in a random way for calibration (99 samples out of 165) as long as all classes and batches were included and equally represented. The fourth batch was also reserved for independent model validation. Model performance was measured in terms of Recall (sensitivity) and Precision, as well as overall correct classification (OCC) (Sokolova & Lapalme, 2009). For PLSDA, the optimum number of PLS components was estimated using stratified three-fold cross-validation with maximum 20 components (i.e. 20 components were used for the 11-class case and 12 for the 3-class case).

HCA (Everitt et al., 2011) was performed per batch using Euclidean Distance and Ward's minimum variance agglomeration method, while PCA (Jolliffe, 2002) was performed per batch, as well as with all three batches so as to visualise whether there

were significant differences among samples from different batches, as well as among different classes.

The partitioning algorithms of the dataset and the LDA algorithm were implemented in MATLAB R2012a (The MathWorks, Inc., Natick, Massachusetts, United States), while HCA, PCA and PLSDA were implemented in R v.3.0.2 (RStudio, v. 0.97.551, RStudio, Inc., Boston, Massachusetts, United States), using the "plsgenomics" package (Boulesteix, 2004; Boulesteix & Strimmer, 2007; de Jong, 1993). Lastly, a heatmap was created using the MetaboAnalyst 2.0 software (Xia, Mandal, Sinelnikov, Broadhurst, & Wishart, 2012).

2.3.3 RESULTS

Selected spectra of minced beef in various adulteration levels are presented in Figure 2.2.

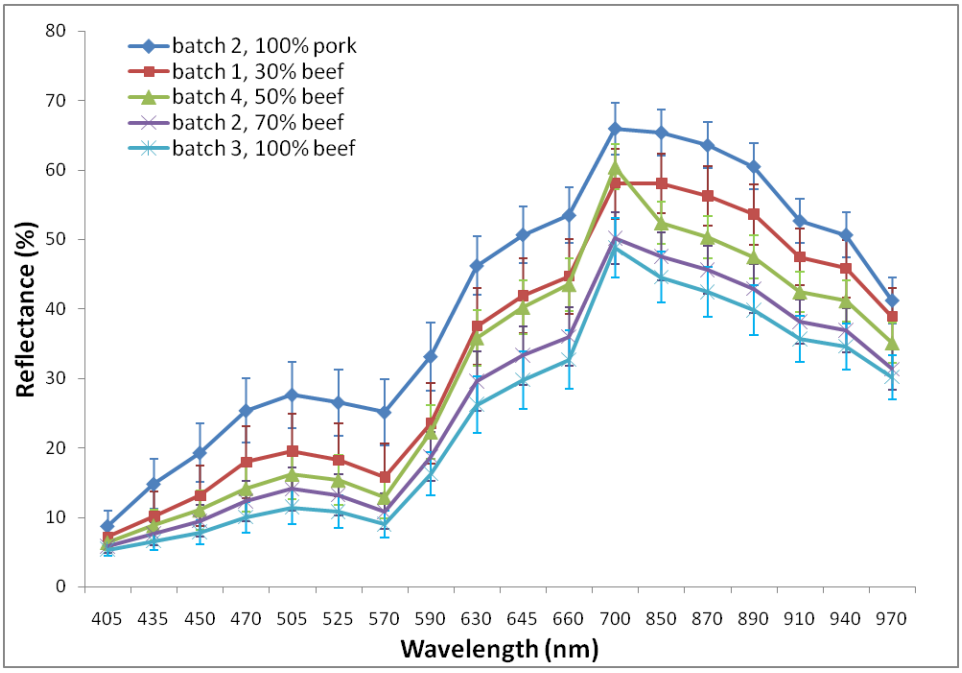


Figure 2.2 Selected spectra of the examined samples corresponding to different ratios of adulteration

It is characteristic that the reflectance of the sample increased in most wavelengths of the spectrum with increasing proportions of pork meat in the mixture, providing strong evidence for the effectiveness of multispectral imaging in discriminating meat adulteration.

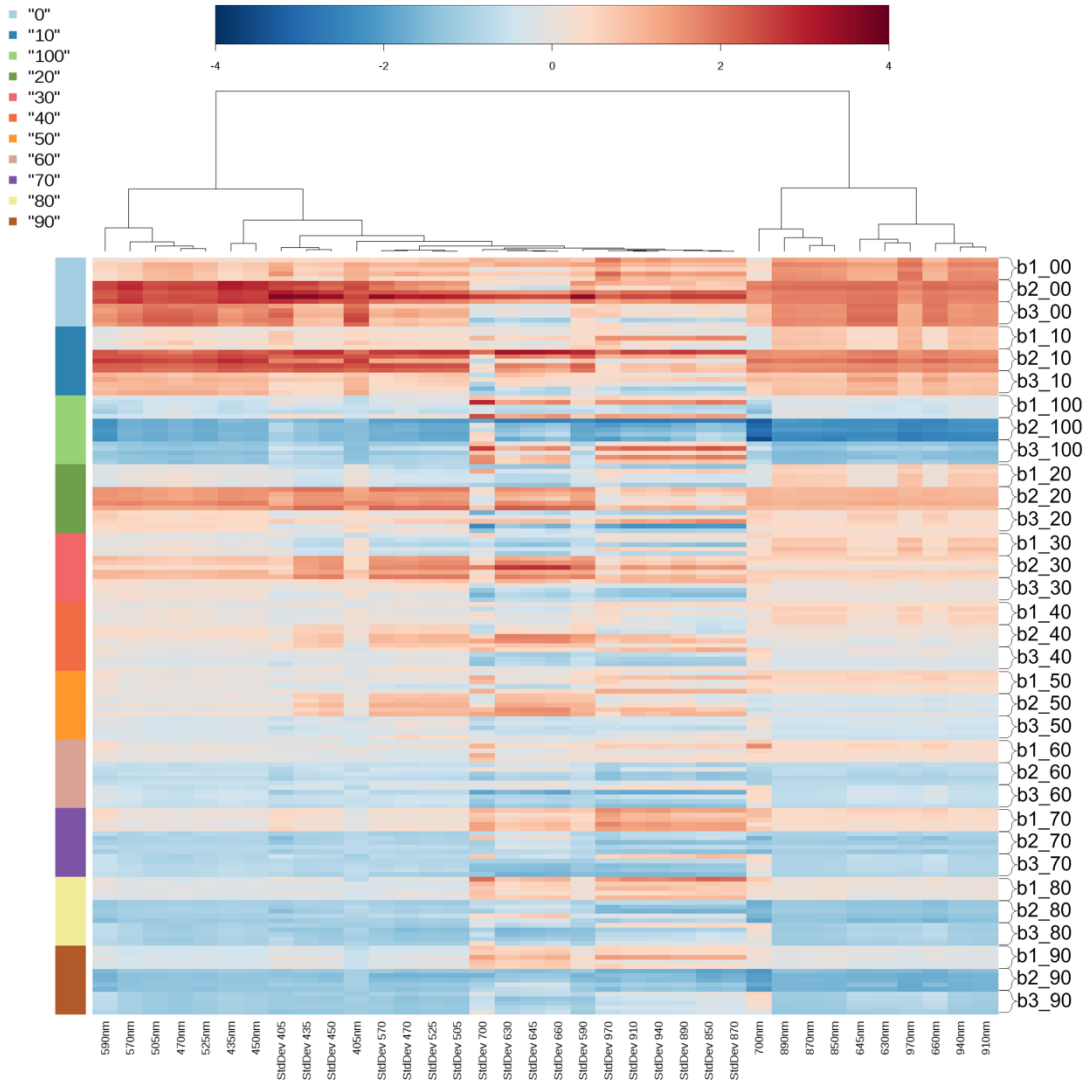


Figure 2.3 Heatmap of all samples from batches 1-3, different colours on the left side correspond to different ratios of adulteration. Samples named with "00" correspond to pure pork, all other categories correspond to the percentage of beef in the mix and consequently "100" refers to pure beef samples, whereas "b1", "b2", "b3" correspond to the number of batch.

Initial analysis with HCA and PCA showed that the use of different batches is critical, as we must take into account not only the variability within a batch (different samples

of the same class), but also the variability among batches. This was evident either when examining the results of HCA per batch (see Supplementary File, Appendix I) or the heatmap for all three batches (Figure 2.3).

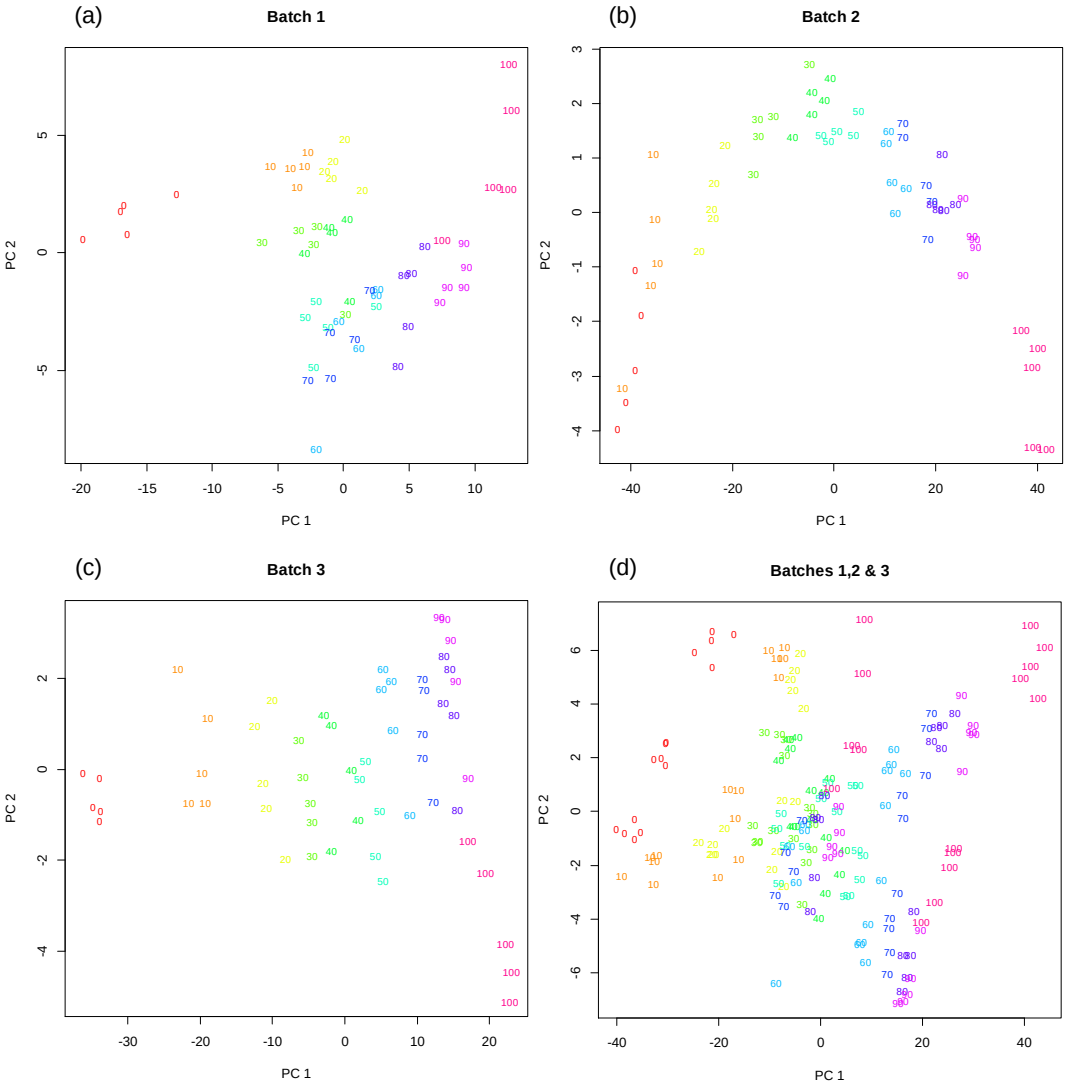


Figure 2.4 Principal Component Analysis scores for Batches 1(a), 2(b), 3(c) and all three batches (d). Samples named with "0" correspond to pure pork, all other categories correspond to the percentage of beef in the mix and consequently "100" refers to pure beef samples.

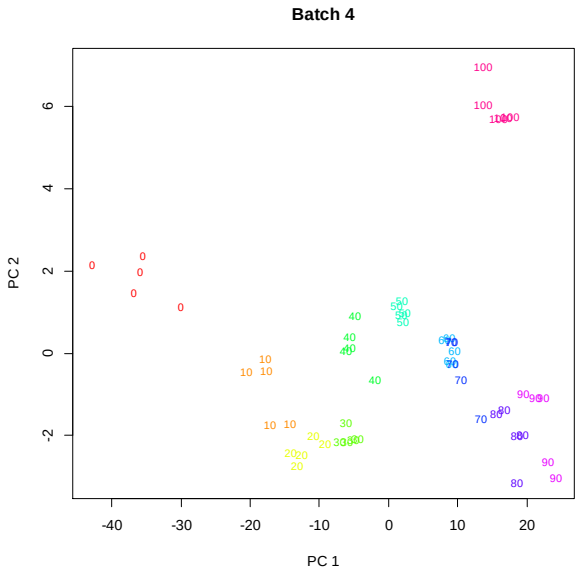


Figure 2.5 Principal Component Analysis scores for external validation batch. Samples named with "0" correspond to pure pork, all other categories correspond to the percentage of beef in the mix and consequently "100" refers to pure beef samples.

When all three batches were compared, samples belonging in the same category showed great differences among batches. Nevertheless, a potential for good discrimination could be concluded after PCA analysis. Different Principal Components (PCs) contributed differently in terms of the variability explained (results not shown), but the first two PC scores are presented in Figure 2.4(a)-(c) and Figure 2.5. In all cases, pure pork and pure beef were found on the far left and right of the plot respectively and the discrimination between classes was more evident. Only adjacent categories sometimes overlap. On the other hand, classes of mixed samples seem to be represented in a different way for each batch. Furthermore, when all three batches were included (Figure 2.4d), a definite trend to the right of the plot was seen as the percentage of beef in the mix increases, but the discrimination among classes was less evident.

The analysis of the data acquired for each class of adulteration is shown in Table 2.1, where the results for per-class Recall and Precision are presented.

Table 2.1 LDA vs. PLSDA for both validation set and external validation batch with 11 classes ranging from pure pork (0%) to beef (100%)

	Validation set						External validation set					
	LDA			PLSDA			LDA			PLSDA		
Sample	Recall	Precision	±10% error	Recall	Precision	±10% error	Recall	Precision	±10% error	Recall	Precision	±10% error
Is 0%	83.33%	100.00%	100.00%	66.67%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%
Is 10%	100.00%	85.71%	100.00%	100.00%	75.00%	100.00%	20.00%	16.67%	100.00%	0.00%	0.00%	20.00%
Is 20%	83.33%	100.00%	100.00%	83.33%	100.00%	100.00%	20.00%	7.14%	20.00%	0.00%	0.00%	20.00%
Is 30%	83.33%	71.43%	83.33%	50.00%	75.00%	100.00%	0.00%	0.00%	100.00%	0.00%	0.00%	40.00%
Is 40%	66.67%	80.00%	100.00%	66.67%	50.00%	100.00%	20.00%	7.69%	80.00%	20.00%	14.29%	20.00%
Is 50%	83.33%	71.43%	100.00%	100.00%	75.00%	100.00%	0.00%	0.00%	0.00%	0.00%	0.00%	60.00%
Is 60%	66.67%	80.00%	100.00%	33.33%	50.00%	83.33%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Is 70%	66.67%	57.14%	100.00%	83.33%	50.00%	100.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Is 80%	83.33%	83.33%	100.00%	16.67%	100.00%	83.33%	0.00%	0.00%	0.00%	0.00%	0.00%	20.00%
Is 90%	100.00%	100.00%	100.00%	83.33%	71.43%	100.00%	0.00%	0.00%	0.00%	100.00%	33.33%	100.00%
Is 100%	100.00%	100.00%	100.00%	100.00%	85.71%	100.00%	0.00%	0.00%	0.00%	100.00%	55.56%	100.00%
Mean per-class	83.33%	84.46%	98.48%	71.21%	75.65%	96.97%	14.55%	11.95%	36.36%	29.09%	18.47%	43.64%

Table 2.2 LDA and PLSDA (12 PLS components) for both validation set and external validation batch with 3 classes (pork - adulterated - beef)

LDA					**PLSDA**				
Validation set									
	classified as					**classified as**			
	pork	**adulterated**	**beef**	**Recall**		**pork**	**adulterated**	**beef**	**Recall**
is pork	5	1	0	83.33%	**is pork**	5	1	0	83.33%
is adulterated	0	54	0	100.00%	**is adulterated**	0	54	0	100.00%
is beef	0	0	6	100.00%	**is beef**	0	0	6	100.00%
Precision	100.00%	98.18%	100.00%		**Precision**	100.00%	98.18%	100.00%	
External validation Batch									
	classified as					**classified as**			
	pork	**adulterated**	**beef**	**Recall**		**pork**	**adulterated**	**beef**	**Recall**
is pork	4	0	1	80.00%	**is pork**	5	0	0	100.00%
is adulterated	0	35	10	77.78%	**is adulterated**	0	45	0	100.00%
is beef	0	0	5	100.00%	**is beef**	0	0	5	100.00%
Precision	100.00%	100.00%	31.25%		**Precision**	100.00%	100.00%	100.00%	

The overall correct classification, mean per-class Recall and Precision were 83.33%, 83.33% and 84.46% respectively, the classification error for 98.48% of the samples was at most 10%, for LDA. The classification of each sample of the validation set is presented along with the per-class Precision and Recall in Table 1 in the supplement. Very good results were also acquired when classification among pure pork,

adulterated and pure beef was tested (Table 2.2), where the overall correct classification, mean per-class Recall and Precision was over 94% (Mean Recall: 94.44%, Precision: 99.39% and overall correct classification: 98.48%). In fact, an only one out of 66 sample was misclassified.

Application of PLSDA for the three-class case, yielded the same results (98.48% correct classification). PLSDA for all categories gave similar results to the LDA. The calibration of the model was done with cross-validation, as described previously, based on overall correct classification criterion (Table 1, and Table 2 in the supplement). It is interesting to note that, although the overall correct classification dropped considerably, this method classified 96.97% of the samples within the ±10% error of prediction.

In the case of external validation, PLSDA performed well, classifying all samples correctly in the three-class problem of pure *vs.* adulterated samples (Table 2.2). In the case of 11 categories, all pure samples were classified correctly, but the prediction of adulteration levels in the samples was less accurate (Table 2.1). However, only 4 out of 45 adulterated samples were classified as pure (see Table 4 in the supplement).

The LDA model was less successful in predicting pure beef samples (Table 2.1), as well as the adulteration level for the case of 11 categories (Table 3 in the supplement), whereas in the case of pure vs. adulterated samples, the results were better yielding an overall correct classification of 80% and mean per-class Recall and Precision of 85.93% and 77.08%, respectively

2.3.4 DISCUSSION

Since various standard analytical methods are now available to identify meat's adulteration at a very low level (Ballin, 2010; Ballin, Vogensen, & Karlsson, 2009), this method showed great potential. Previously reported methods (Ballin, 2010) are time-consuming, expensive, use harmful reagents, need expert laboratory staff and are strongly dependent on rigorously following a standardized protocol to obtain accuracy, where multispectral imaging requires only basic training in a user-friendly software, a few minutes for image acquisition and processing and no cost at all – excluding initial instrument and software purchase.

Compared to other similar studies on rapid techniques, in most cases, the main objective is the differentiation among different types of meat, e.g. beef *vs.* kangaroo (Ding & Xu, 1999), pork *vs.* beef *vs*. lamb (Kamruzzaman et al., 2012), beef *vs.* horsemeat (Boyacı et al., 2014). Few studies has been published on the adulteration of poultry with pork (Soares et al., 2010), pork in minced mutton (Tian et al., 2013), pork in beef meatball (Rohman et al., 2011), pork meat in raw beef burger (Giaretta, Di Giuseppe, Lippert, Parente, & Di Maro, 2013), minced lamb (Kamruzzaman et al., 2013), gelatin in prawn (Wu et al., 2013). Some have been reported in the case of classification between adulterated and pure samples with different percentages of adulteration (Alamprese et al., 2013; Kamruzzaman et al., 2013; Tian et al., 2013), and the case of beef adulterated with pork (Morsy & Sun, 2013; Rohman et al., 2011). In this study, pure beef, pure pork and nine levels of adulteration were employed in order not only to discriminate but also to quantify the minimum possible level of adulteration detected.

As in most of the above cases, HCA, PCA, LDA and PLSDA were the predominant methods used for data analysis. Results showed that multispectral imaging has the potential to identify adulterated beef samples with pork and vice versa in a rapid, non-invasive way. Furthermore, the variability between meat batches was taken into account – an important issue that is not always presented in the available literature- by using three different batches for model training and testing, and a fourth external batch for validation. Results showed that a 10% adulteration with pork in beef and vice versa could be successfully identified and could thus be considered as a detection limit of the applied method, which can be related to the results by Morsy and Sun (2013) using NIR spectroscopy, although no external validation was performed in this work, and the results by Alamprese et al. (2013) in the case of adulteration with turkey.

The quantification of the level of adulteration was proved to be more difficult task. A large number of adulteration classes were used (11 classes in total), whereas in other studies discriminant analysis was performed with fewer categories. For example, Alamprese et al. (2013) used 5 classes for cross-validation, grouping very low adulteration with pure samples. However, in this study the applied method was found to provide additional information on the detection limit of 10% and as such can be considered as an advantage. Finally, while very few of the abovementioned studies use external batch validation, results demonstrated -especially for LDA- the necessity of such an approach in order to exclude cases of overoptimistic results.

2.4 ADULTERATION WITH HORSEMEAT

2.4.1 EXPERIMENTAL DESIGN

The experimental procedure consisted of three distinct stages. In the first stage, different levels of adulteration with a 20% step were prepared based on the procedure described by Ropodi et al. (2015) and Section 2.3. Briefly, fresh beef and horsemeat *Longissimus* muscle fillets were purchased, cut into smaller pieces and ground separately using a domestic meat-mincing machine. The appropriate portions of each meat were mixed in order to achieve four levels of adulteration, 20-80%, 40-60%, 60-40% and 80-20% (w/w), as well as pure beef and horsemeat. From each level of adulteration, five different portions of ca.75–80 g (5×6=30 samples in total) were placed in Petri dishes and snapshots were taken using VideometerLab vision system (Videometer A/S, Hørsholm, Denmark) using a random sample for the autolight procedure.

After preliminary analysis of the data, it was decided to create more samples per category and focus on the pure samples, and the levels of 60-40% and 80-20% (w/w) for beef and horsemeat, respectively, as well as add a 90-10% level for the second stage in order to explore smaller levels of adulteration. Indeed, eight samples per level (8×5=40 samples) were prepared and multispectral images were acquired at the time, as well as after the samples were stored in high-precision incubators at 4°C for 6, 24 and 48 hours. A graphical representation of the experimental design is shown in Figure 2.6. Lastly, in the third stage, the previous procedure was repeated for validation purposes. From now on, meat batches from each experimental stage will be referred to as batch 1, 2 or 3 (b1, b2 or b3).

In total, 110 samples were prepared and 350 images were acquired (i.e., 30 images from batch 1, 40×4=160 images from batch 2, and 40×4=160 images from batch 3).

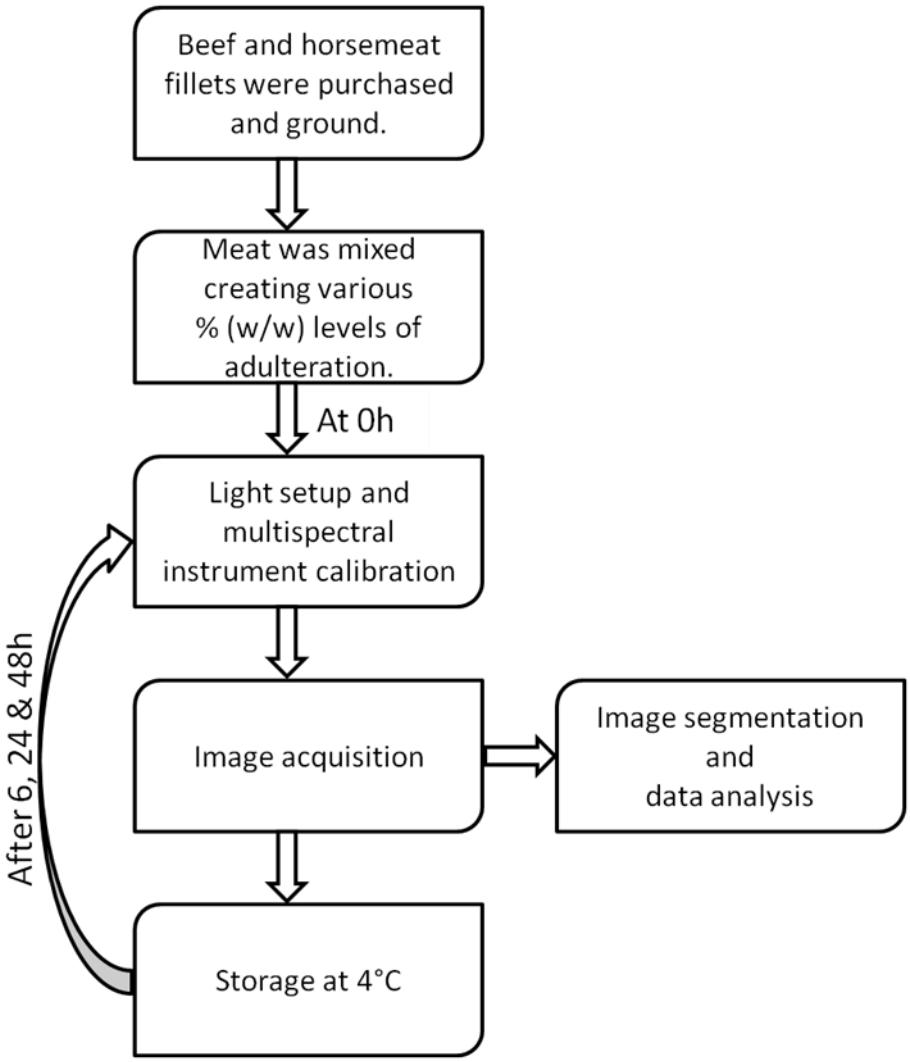


Figure 2.6 Graphical representation of experimental design.

2.4.2 DATA ANALYSIS

Following image segmentation and extraction of MSI data, PCA (Jolliffe, 2002) was used in order to visualize and interpret data compared to previous works as well as among different experimental stages. Additionally, classification techniques were employed in order to discriminate among different levels of adulteration and other

classes, such as freshly-ground *vs.* stored minced meat samples and pure *vs.* adulterated samples. These techniques are presented below:

- PLSDA (Barker & Rayens, 2003; de Jong, 1993) where the optimum number of PLS components was estimated based on the overall correct classification (OCC) using cross-validation (CV) results of 100 random partitions (80% for training, 20% for testing).
- RF (Breiman, 2001), where various parameters were explored and models were chosen based on Out-Of-Bag (OOB) classification error, as a subset of the training instances is left out in order to obtain an unbiased estimate of the error.
- SVMs (Cortes & Vapnik, 1995) where different kernel functions were employed (results not shown) and training as well as model selection was performed based on the OOC criterion for 3-fold CV coupled with a grid search for the optimal hyper-parameters. For the two-step model, a SVM model (SVM-1) with b2 training data (including classes 0, 60, 80, 90, and 100% beef) and a radial basis function (RBF) kernel was calibrated using grid search coupled with 3-fold cross validation for calculating the optimal parameters (Capacity=3, gamma=0.028). Next, a new SVM model (SVM-2) was developed using as input all freshly-ground data (common levels of adulteration and pure samples of b1 & 2) and a linear kernel (Capacity=1).

In all supervised methods, while b1 and/or b2 samples were used for model development, b3 samples were reserved for independent model validation, as proposed by Ropodi et al. (2015). This was done so that the models could be tested with an unknown and unbiased dataset in order to exclude overoptimistic results. In

addition, b1 and b2 were used for model calibration to take into account the variability within (replicate samples of the same adulteration level) and among meat batches. Furthermore, only the common levels of adulteration were used during model development.

Different classification problems were explored including classification among different percentages of beef (4 classes), pure beef *vs.* pure horsemeat *vs.* adulterated samples, pure meat (both beef and horse) *vs.* adulterated samples and freshly-ground *vs.* stored minced meat. Model performance was measured mainly in terms of OCC, as well as recall (sensitivity) and precision (Sokolova & Lapalme, 2009).

PCA was performed using MetaboAnalyst 3.0 software (Xia, Sinelnikov, Han, & Wishart, 2015) and PLSDA was implemented in R v.3.0.2 (http://www.r-project.org/, The R Foundation for Statistical Computing, Vienna, Austria) and Rstudio v.0.97.551 interface (RStudio, Inc., Boston, Massachusetts, USA), using the "plsgenomics" package (Boulesteix, 2004; Boulesteix & Strimmer, 2007; de Jong, 1993). RF was performed in MATLAB 2012a software (The MathWorks, Inc., Natick, Massachusetts, USA) and lastly SVMs were employed using Statistica v. 8.0 software (Statsoft Inc., Tulsa, OK, USA).

2.4.3 RESULTS

In Figure 2.7, the mean reflectance values in 18 wavelengths is presented for a pure horsemeat (a) and a pure beef (b) sample before and after storage for 6, 24 and 48h. It is evident that, although stored samples are more difficult to distinguish among themselves, freshly-ground samples are easily differentiated.

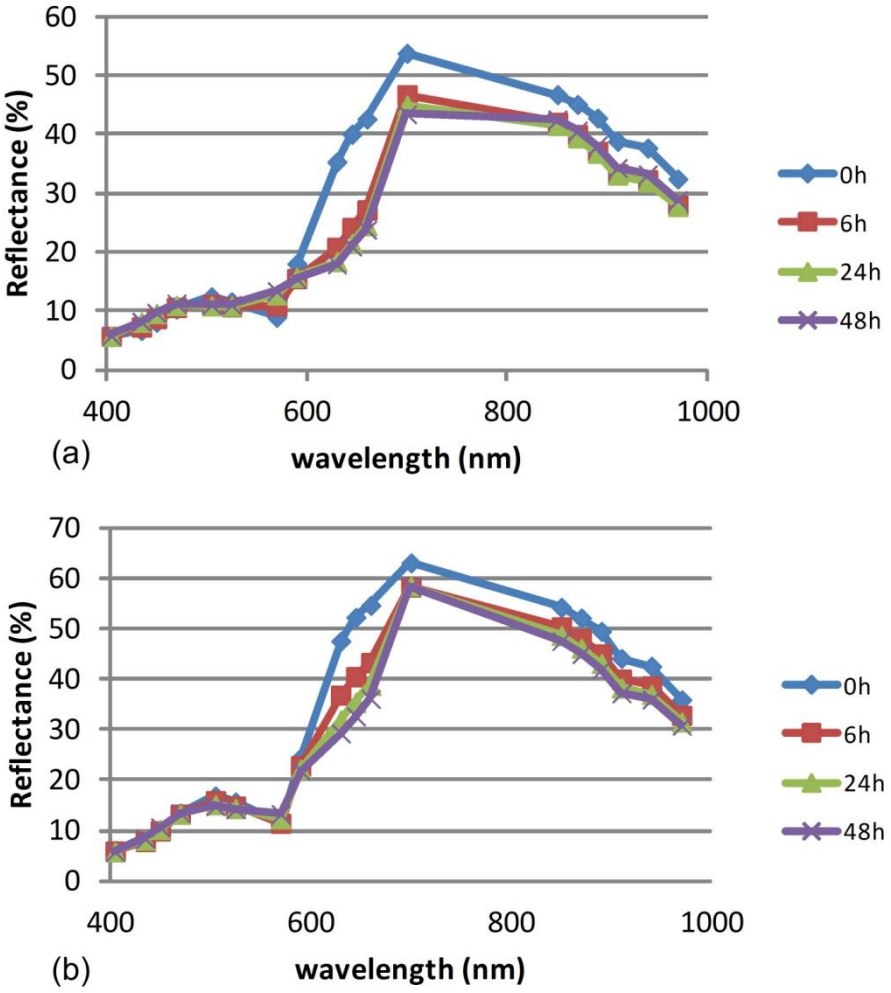


Figure 2.7 Mean reflectance (%) values of (a) a minced beef and (b) a minced horsemeat sample for 18 distinct wavelengths ranging from 405 to 970nm. Samples are freshly-ground (0h) and stored in refrigerated conditions for 6, 24 & 48h.

Additionally, PCA scores for batch 1 are presented for principal components (PCs) 1 and 3 in Figure 2.8. Indeed, PC1 vs. PC3 scores displayed a distinct separation of pure *vs.* adulterated samples. Pure beef and horsemeat samples were located on the top right of the plot and although they seemed close, discrimination was more evident in the PC2 *vs.* PC5 plot (Supplementary, Fig. 1). Furthermore, various levels of adulteration were easily distinguishable with only adjacent categories sometimes overlapping.

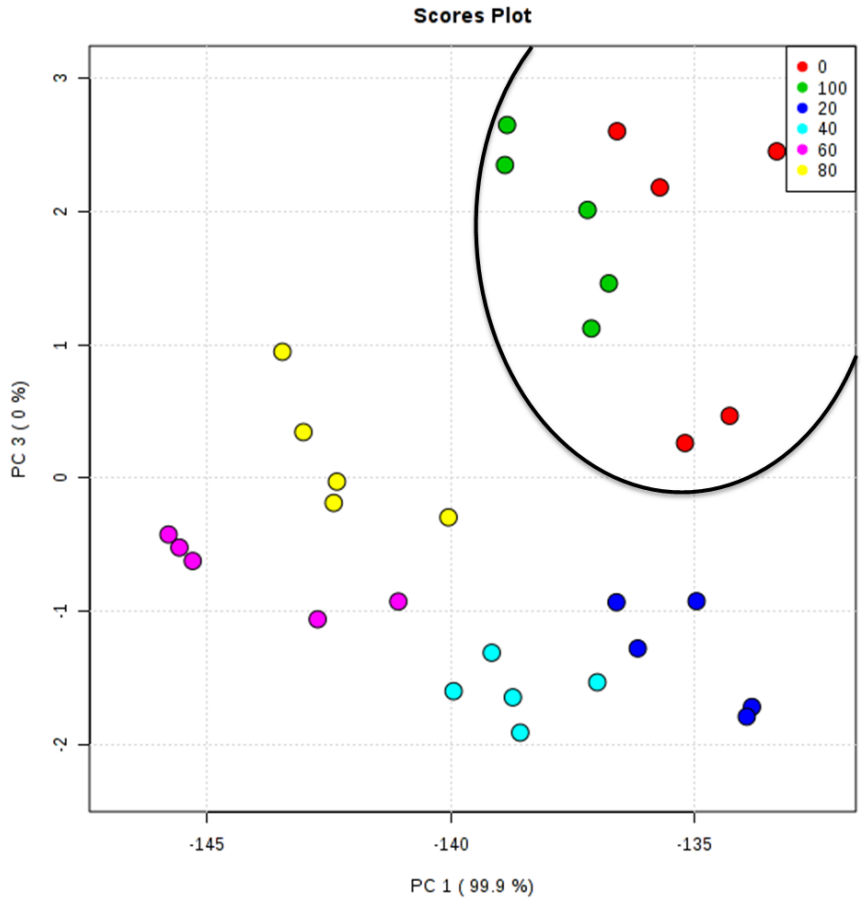


Figure 2.8 Principal component analysis scores (PC1 *vs.* PC3) for batch 1. Samples indicated with "0" correspond to pure horse, all other categories correspond to the percentage of beef in the mix and consequently "100" refers to pure beef samples.

These results are in good agreement with PCA analysis performed on Raman spectra of fat samples with a 25% step among different levels of adulteration, where different categories were evident in the PC1 *vs.* PC2 plot (Boyacı et al., 2014). They are also consistent with the results in the case of adulteration with pork where the same experimental design (with a 10% step) was implemented (Ropodi et al., 2015). However, it should be noted that, while in the latter two cases the differences were apparent using the first two components and therefore the PCs with the highest variance explained (%), this was not the case in b1 data. In fact, PCs 1 *vs.* 3 and PCs 2 *vs.* 3 (Supplementary, Fig. 2) displayed higher discriminatory power than PC1 *vs.* PC2

(Supplementary, Fig. 3), especially in the case of distinguishing between pure and adulterated samples, denoting a significant difference between the datasets.

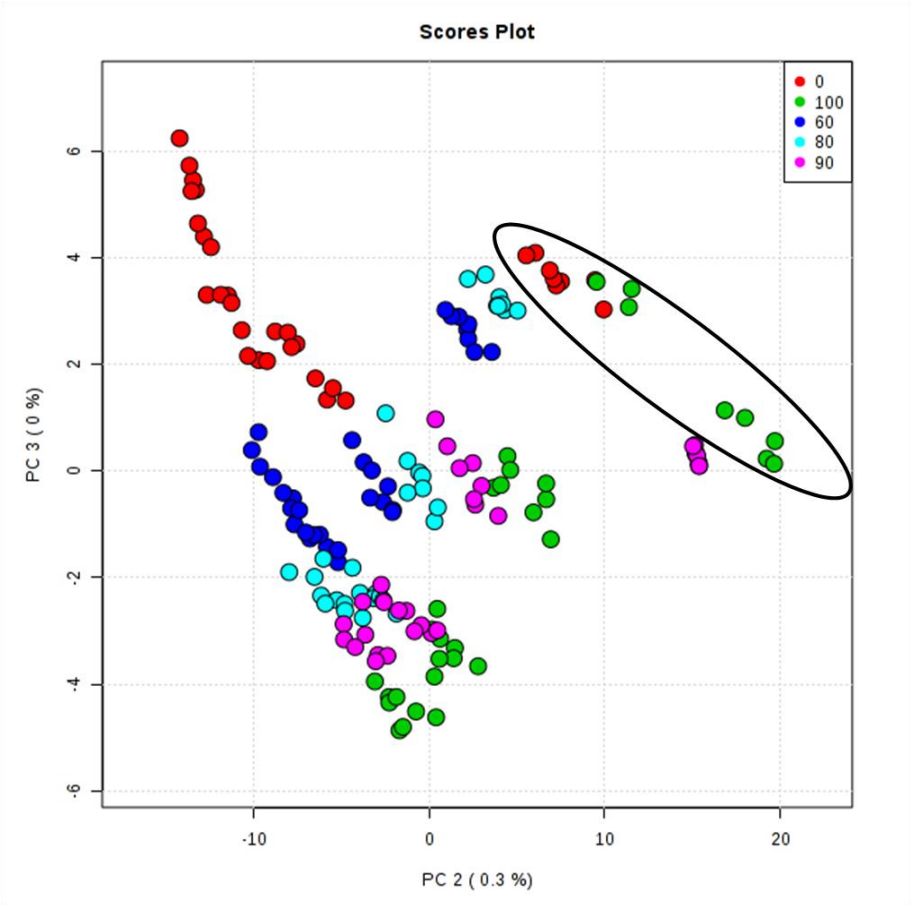


Figure 2.9 Principal component analysis scores (PC2 *vs.* PC3) for batch 2. Samples indicated with "0" correspond to pure horse, all other categories correspond to the percentage of beef in the mix and consequently "100" refers to pure beef samples. Samples within the ellipse are freshly-ground pure beef and horsemeat.

Different levels of adulteration were also evident when PCA was applied to b2 freshly-ground image data with the 90-10% (w/w) to seem more difficult to separate from pure beef samples (Supplementary, Fig. 4). Furthermore, freshly-ground and especially pure samples were clearly noticeable indicating the major changes occurring in meat color during storage (Figure 2.9). Lastly, as storage time elapsed, less image data from different storage times and percentages of adulteration could be distinguishable.

In Table 2.3, OCC classification results for PLSDA and RF are presented.

Table 2.3 Overall correct classification results for training and validation sets for partial least squares discriminant analysis and random forest.

	Overall correct classification (%)			
	PLSDA		RF	
	training set (b1&b2)	validation set (b3)	OOB prediction	validation set (b3)
freshly ground vs. stored (0, 60, 80, 100)	100.00	66.41	96.62	75.00
pure vs. adulterated (0, 60, 80, 100)	97.30	52.34	95.27	50.00
pure horse vs. pure beef vs. adulterated (0, 60, 80, 100)	100.00	50.00	95.27	25.00
4 classes 0, 60, 80&100	98.65	37.50	92.57	25.00

In the case of PLSDA for the different classification problems explored including classification among different percentages of beef (4 classes), pure beef *vs.* pure horsemeat *vs.* adulterated samples, pure meat (both beef and horse) *vs.* adulterated samples and freshly-ground *vs.* stored minced meat. Results showed that the majority of the models significantly underperformed in the stage of independent validation, although they exhibited very good performance during calibration. PLSDA yielded a 66.41% OCC in the case of freshly-ground *vs.* stored meat samples, however only one freshly-ground sample was misclassified as stored.

Furthermore, when the proportion of observations to be included in the training set for each cross-validation iteration was set to 0.2, results improved and mean per-class Recall (sensitivity) and Precision were 89.06 and 81.00%, respectively (Supplementary, Table 1). While this may imply that another methodology should be used, lack of generalization in the former models may also be attributed to a case of

overfitting due to inadequate representative training samples. Indeed, b1 and b2 datasets include two batches of freshly ground samples, but only one batch of stored samples in refrigerated conditions. This may have led to a lack of sufficient data for adulteration detection. However, the improved results of the latter PLSDA model and PCA scores for b2 (Figure 2.9) suggest that data were sufficient for the "freshly-ground *vs.* stored" classification problem and possibly "freshly-ground and pure *vs.* other samples.

SVM models were calibrated in order to solve this problem and yielded very good results. The SVM model (SVM-1) with b2 training data and a radial basis function (RBF) kernel classified all training samples correctly (OCC=100%) in the subset retained for internal model validation. Additionally, independent model validation yielded an OCC equal to 99.38%, as well as 99.58% and 98.78% mean per-class recall and precision respectively. In fact, only one testing sample, a minced horsemeat sample stored for 6h, was misclassified (Supplementary, Table 2). The above classification results led to considering a new two-step approach in order to identify pure as well as freshly-ground samples, namely (a) classifying samples as freshly-ground or stored, and (b) classifying the former as pure or adulterated.

In Figure 2.10, a graphical presentation of the classification results is presented. Using SVM-1, only one "stored" sample was misclassified. Applying the SVM-2 model, all pure samples were classified as pure, but 5 adulterated samples were misclassified yielding an OCC of 84.38%. However, the misclassified samples were all adulterated with 20% horsemeat, therefore the largest (w/w) percentage of horsemeat where samples were misclassified was 20% after independent validation. The misclassified sample of SVM-1 was classified as pure and as a consequence the

final results yielded an OOC equal to 95.31%. The performance of the two-step model is presented in Table 2.4.

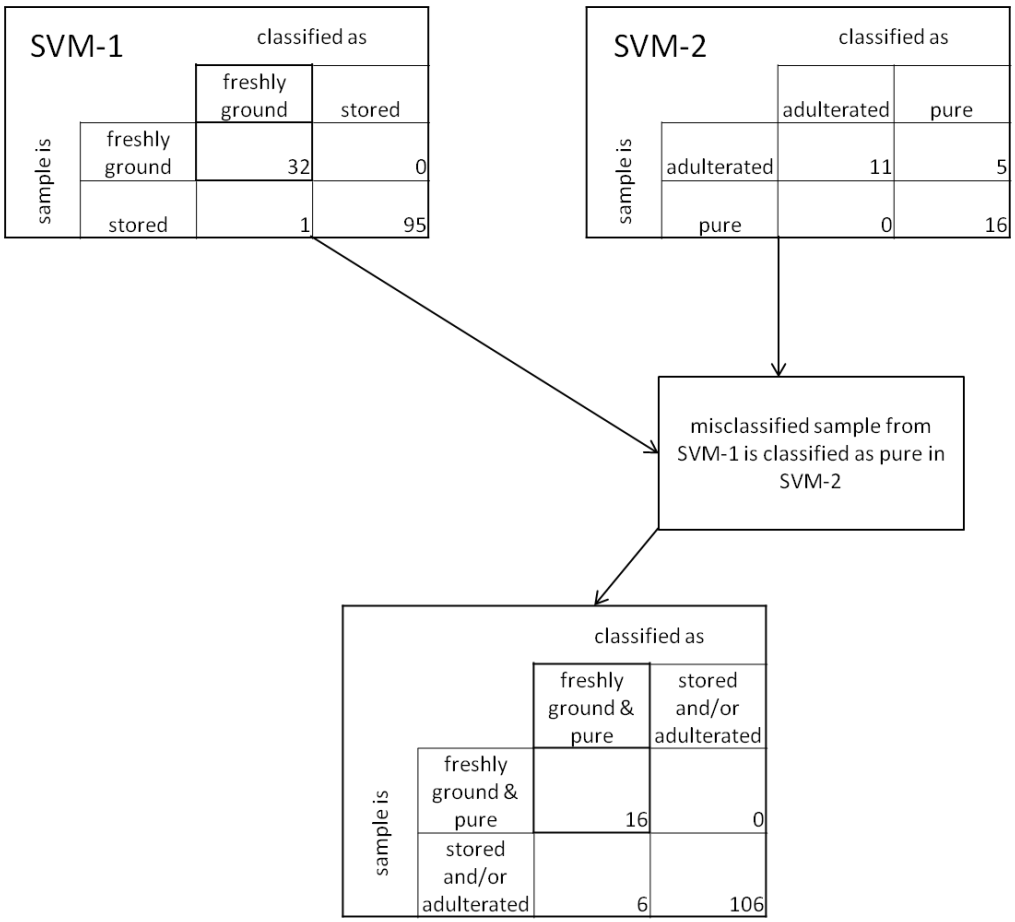


Figure 2.10 SVM two-step classification scheme and classification results for the identification of freshly-ground and non-adulterated samples.

Table 2.4 Final confusion matrix for 2-step SVM model and classification performance indices for validation batch (b3) and samples of common levels of adulteration (100-0, 80-20, 60-40, 0-100% w/w beef and horsemeat respectively).

		To			
		freshly-ground & pure	**stored and/or adulterated**	Recall	Mean Recall
From	**freshly-ground & pure**	16	0	100.00%	97.32%
	stored and/or adulterated	6	106	94.64%	
	Precision	72.73%	100.00%		
	Mean precision	86.36%			
				OOC=	95.31%

2.4.4 DISCUSSION

Similar studies on meat authenticity have been published previously, but they usually focus on classification among different types of meat such as chicken *vs.* turkey (Ellis et al., 2005), beef *vs.* kangaroo (Ding & Xu, 1999) and beef *vs.* horsemeat (Boyacı et al., 2014; Ebrahim, Sowoidnich, & Kronfeldt, 2013) using FTIR, visible/NIR and Raman spectroscopy. In addition, hyperspectral imaging has been employed for the differentiation of various types of meat such as pork *vs.* beef *vs.* lamb (Kamruzzaman et al., 2012). MSI has only been used in the detection of adulteration of minced beef with pork (Ropodi et al., 2015). In fact, little work has been undertaken in terms of multiple levels of meat adulteration with horsemeat (Boyacı et al., 2014), but without extended model development and validation using different batches of meat. On the other hand, published studies have presented cases of adulteration of poultry with pork (Soares et al., 2010), pork in beef meatball and in minced mutton (Rohman et al., 2011; Tian et al., 2013), beef offal in fresh and frozen beefburger (Zhao et al., 2014), pork meat in raw beef burger and minced lamb (Giaretta et al., 2013; Kamruzzaman et al., 2013). Although some of these studies displayed good results, they cannot be compared to this study due to the different type of meat and/ or adulterant. While the maximum error was 20%, lower detection limits were found by Morsy and Sun (2013) using NIR spectroscopy, but with no external validation. Alamprese et al. (2013) used independent samples for validation with very good results in the case of beef adulteration with turkey, but pure samples were grouped together with low levels of adulteration for classification. Lastly, none of the above studies has taken into consideration changes occurring during refrigerated storage.

2.5 CONCLUSIONS

In conclusion, multispectral imaging was used for the first time as a rapid method for food authentication and detection of adulteration of raw meat, illustrating a clear separation of pure *vs.* adulterated samples. While in case 1, PLSDA performed better (10% adulteration could be considered as a limit of detection) compared to LDA, a more complex modelling scheme (2 SVMs) was necessary in the case of stored samples and adulteration with horsemeat (case 2). Moreover, the quantification of the percentage of adulteration, as well as the detection of adulteration in the cases of stored samples was proved to be more challenging (a maximum error of 20%). However, it also emphasized the necessity of independent sample validation and the significant changes occurring in color during storage. Although these changes affected the model performance greatly, both studies have proven that MSI could be used for large scale quality control applications in the future.

Acknowledgements

This work has been supported by the project "Intelligent multi-sensor system for meat analysis – iMeatSense550" co-financed by the European Union (European Social Fund – ESF) and Greek national funds through the Operational Program "Education and Lifelong Learning" of the National Strategic Reference Framework (NSRF) - Research Funding Program: ARISTEIA-I.

CHAPTER 3:

DETECTION OF FROZEN-THEN-THAWED MINCED MEAT

3 ABSTRACT

In this chapter, the ability of both MSI and FTIR to detect frozen-then-thawed minced meat is explored. The work is presented in a manuscript format. Supplementary material is provided in Appendix I.

In short, freshly-ground beef was purchased from seven separate shops at different times, divided in fifteen portions and placed in Petri dishes. Multi-spectral images and FTIR spectra of the first five were immediately acquired while the remaining were frozen (-20°C) and stored for seven and 32 days (5 samples for each time interval). Samples were thawed and subsequently subjected to similar data acquisition. In total, 105 multispectral images and 105 FTIR spectra were collected, which were further analyzed using partial least-squares discriminant analysis and support vector machines. Two meat batches (30 samples) were reserved for independent validation and the remaining five batches were divided in training and test set (75 samples). Results showed 100% overall correct classification for test and external validation MSI data, while FTIR data yielded 93.3 and 96.7% overall correct classification for FTIR test set and external validation set respectively.

3.1 INTRODUCTION

The safety of meat and meat products is currently very much in focus, owing to calamities with microbiological outbreaks, dioxin contamination and other threats to human health (EFSA (European Food Safety Authority), 2015, 2017). At a time when consumer awareness of nutrition and health is increasing, it is important that meat products should be safe for all consumers. This is the case particularly with beef, which is a large part of human diet and therefore its quality is of importance not only to consumers, but also to food authorities and the meat industry. Meat production encompasses slaughterhouses, packers, company- owned distribution and supply operations and importers. Wholesaling includes the nationwide network of meat and meat product wholesalers and related warehouse and transportation units. Retailing includes locations where meat is consumed "on-premise", such as restaurants. "Off-premise" retail outlets are supermarkets, butchers, warehouse stores, and similar locations. Therefore non-compliance to labels may lead to loss of consumer trust and subsequently loss of revenue for food retailers, as well as have other economic and safety consequences (Alamprese et al., 2013). Nevertheless, meat can be an attractive target for fraudulent and deceptive practices, such as selling adulterated with cheaper ingredients meat and thawed meat as fresh due to its higher market price.

Various analytical methods have been proposed for the detection of frozen-then-thawed meat including enzymatic, DNA-based, microscopic and sensory techniques (Ballin & Lametsch, 2008). However, in order to ensure constant large-scale and effective monitoring, rapid, low-cost, non-invasive methods or methods requiring a small sample have to be employed. These rapid methods involve various sensors such as hyperspectral (HSI) and multispectral imaging (MSI), NIR and Raman

spectroscopy. Such sensors have been previously used in tandem with various data analysis methods (Dai, Sun, Xiong, Cheng, & Zeng, 2014; Ropodi, Panagou, & Nychas, 2016) for spoilage estimation (Argyri et al., 2013, 2014; Dissing et al., 2013; Panagou et al., 2014; Wu & Sun, 2013a), as well as fraud detection e.g. adulteration (Alamprese et al., 2013; Kamruzzaman, Sun, ElMasry, & Allen, 2013; Morsy & Sun, 2013; Ropodi, Pavlidis, Mohareb, Panagou, & Nychas, 2015; Ropodi, Panagou, & Nychas, 2017). In terms of detection of frozen-then-thawed meat some work has been done with fish fillets and HSI (Cheng et al., 2015), minced chicken, pork and turkey and MIR spectroscopy (Al-Jowder et al., 1997), beef and NIR spectroscopy (Downey & Beauchêne, 1997), as well as pork and HSI (Barbin, Sun, et al., 2013; Ma et al., 2015; Pu et al., 2015). To our knowledge, two or more sensors including multispectral imaging and vibrational spectroscopy have not been used previously either in combination or separately for frozen-then-thawed minced beef detection.

Thus, the objective of this study was to (a) evaluate the potential of multispectral imaging and FTIR spectroscopy in tandem with data analysis techniques to identify frozen-then-thawed minced beef, and (b) compare –and combine if possible– sensors and models.

3.2 MATERIALS & METHODS

3.2.1 EXPERIMENTAL DESIGN/ SAMPLE PREPARATION

In this study, freshly-ground beef was purchased at seven separate dates from different butcher shops and supermarkets in Athens, Greece, with no prior knowledge as to the packaging and storage conditions before purchase. Each purchase

corresponding to a unique meat batch was then transported to the laboratory within approximately 30 minutes (seven batches in total). Each batch of ground beef was divided in fifteen samples of 70-75g each and placed in Petri dishes, while a small sample was retained for microbiological analyses for the purpose of acquiring background information. Then, multi-spectral images of the first five samples were acquired followed by Fourier Transform InfraRed (FTIR) spectroscopy measurements using ~3g portions from each Petri dish.

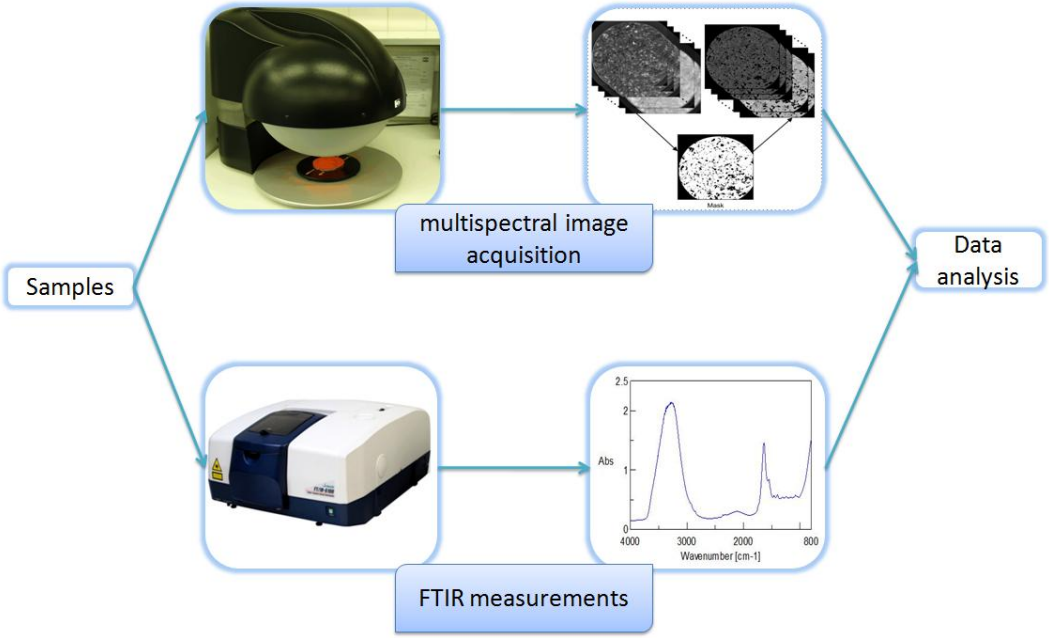


Figure 3.1 Graphical presentation of the data acquisition process.

The remaining samples were frozen at -20°C and stored for seven and 32 days (5 samples/ batch). Samples were then thawed for close to 5h at 4°C so as there was no evidence of the sample having been frozen in the surface and subsequently subjected to similar image acquisition. FTIR measurements were then conducted within the next hour.

In total, 105 multispectral images and 105 FTIR spectra were collected (7 batches × 15 Petri dishes). In Figure 3.1, a schematic presentation of the data acquisition process is presented.

3.2.2 IMAGE ACQUISITION & SEGMENTATION

The procedure for image acquisition and subsequent image segmentation has been described extensively in previous studies (Panagou, Papadopoulou, Carstensen, & Nychas, 2014; Ropodi, Pavlidis, Mohareb, Panagou, & Nychas, 2015; Ropodi, Panagou, & Nychas, 2017), and the previous chapter. Briefly, multispectral images were captured in 18 non-uniformly distributed wavelengths ranging from 405 to 970nm, including the visible (VIS) and near-infrared (NIR) spectrum using the VideometerLab instrument commercialized by "Videometer A/S" (Carstensen & Hansen, 2003). Instrument calibration was performed with: (a) a light setup procedure called "autolight" which takes into account the type of object to be recorded and in this case, the first sample of the first meat batch, and (b) a geometrical and radiometrical calibration (Folm-Hansen, 1999).

The acquired images were then segmented so that redundant information (e.g. sample background, Petri dish) can be excluded. During this image-processing stage the respective routines of the VideometerLab software (version 2.12.39) which also controls the operation of the instrument were used and canonical discriminant analysis (CDA) was employed as a two-step supervised method to divide the images into regions of interest by using a simple threshold to separate between pixels of lean tissue and other pixels. This way, image background and the Petri dish was removed from the actual sample, as well as adipose from lean tissue was separated.

3.2.3 FOURIER TRANSFORM INFRARED (FTIR) SPECTROMETRY MEASUREMENTS & PREPROCESSING

Small portions (approximately 3g) of meat were placed on the surface of a ZnSe 45° HATR (Horizontal Attenuated Total Reflectance) crystal (PIKE Technologies, Madison, Wisconsin, United States) and using FTIR 6200 JASCO spectrometer, FTIR measurements were performed. With the Spectra Manager software version 2 (Jasco Corp.), spectra were collected from 4000 to 400 cm^{-1} (100 scans, resolution of 4 cm^{-1}) within a period of 2 min. A reference spectrum was acquired at regular intervals using the crystal with no added meat and the crystal's surface was cleaned after each acquisition with detergent and distilled water and then with analytical grade acetone. Lastly, it was dried with lint-free tissue. Spectra ranging approximately from 1800 to 800 cm^{-1} were exported.

3.2.4 DATA ANALYSIS

After image segmentation, the average reflectance values and their standard deviation per wavelength based on selected pixel intensity values were extracted resulting in 36 variables (18 mean values and 18 standard deviations) per image. Additionally, standard normal variate (SNV) transformation was performed on the acquired spectra and then the FTIR spectra were autoscaled.

The unsupervised method of principal component analysis (PCA) (Jolliffe, 2002) was used in order to visualize and provide insight. Furthermore, two widely-spread and robust supervised classification techniques were employed in order to discriminate between fresh and frozen-then-thawed samples: Partial least squares discriminant analysis (PLSDA) and support vector machines (SVM).

Analytically, The PLSDA implementation used is described by Boulesteix and Strimmer (2007). Model development was performed with maximum number of PLS components equal to 10 and the optimum number of components was estimated based on the overall correct classification (OCC) criterion using the results of cross-validation (CV) with 10 random partitions (80% for PLS model development, 20% for CV). On the other hand, SVMs (Cortes & Vapnik, 1995) is a very popular machine learning methodology where the original data points are mapped from the input space into a higher dimensional feature space using a kernel function, in order to construct a maximal separating hyper-plane. In this case, the linear kernel was employed. Grid search for the optimal capacity (C) parameter coupled with 10-fold CV was employed for model development. The capacity parameter varied from 0.5 to 10, using a 0.5 step for both MSI and FTIR data.

In both methods, a scheme for robust model training and validation was employed. Specifically, the first five batches were partitioned in a stratified manner, 60% for model training and 40% of testing/validation, i.e. three fresh, three frozen for 7 days and three frozen for 32 days samples per batch for training (45 samples for training and 30 for testing). The last two batches (30 samples) were retained for independent batch validation, thus testing the generalization ability of the models when meat batches are not previously known and assuming batch variability. Models were developed based on the training set (60% of batches 1-5, 45 samples), a test set (40% of batches 1-5, 30 samples) used for validation, while batches 6 and 7 (30 samples) were reserved for external validation.

In terms of implementation, PCA as well as PLSDA figures were made using MetaboAnalyst3.0 and software (Xia et al., 2015). PLSDA was implemented in R

v.3.0.2 (http://www.r-project.org/, The R Foundation for Statistical Computing, Vienna, Austria) and Rstudio v.0.97.551 interface (RStudio, Inc., Boston, Massachusetts, USA), using the "plsgenomics" package (Boulesteix, 2004; Boulesteix & Strimmer, 2007; de Jong, 1993). SVMs were developed using Statistica v.8.0 software (Statsoft Inc., Tulsa, OK, USA). Lastly, FTIR preprocessing was performed in MATLAB 2012a software (The MathWorks, Inc., Natick, Massachusetts, USA).

3.3 RESULTS & DISCUSSION

3.3.1 MULTISPECTRAL IMAGE DATA VS. FTIR DATA

Typical MSI and FTIR profiles of fresh and frozen-then-thawed samples, are given in Figure 3.2a and b, respectively. In particular, red lines correspond to fresh minced beef, whereas green and purple to frozen-then-thawed samples, i.e. green for 7 days and purple for 32 days of storage. Differences in spectra were observed among fresh and frozen-then-thawed MSI spectra (Figure 3.2a) both in the visible (mostly for wavelengths equal and greater than 505nm) and the NIR area. The range of 510–650 nm has previously been associated among others with myoglobin, metmyoglobin and oxymyoglobin, and the spectra between 750 and 970 nm are connected with O–H stretching relating to water content/ moisture as suggested in the literature (Liu et al., 2003; Ma et al., 2015; Pu et al., 2015). While the visible area provides information likely suggesting a conversion and degradation for a number of myoglobin derivatives and discoloration due to the thawing process, NIR spectra might provide information such as water content and denaturation of proteins (910 nm band might indicate a contribution of proteins) (Liu et al., 2003).

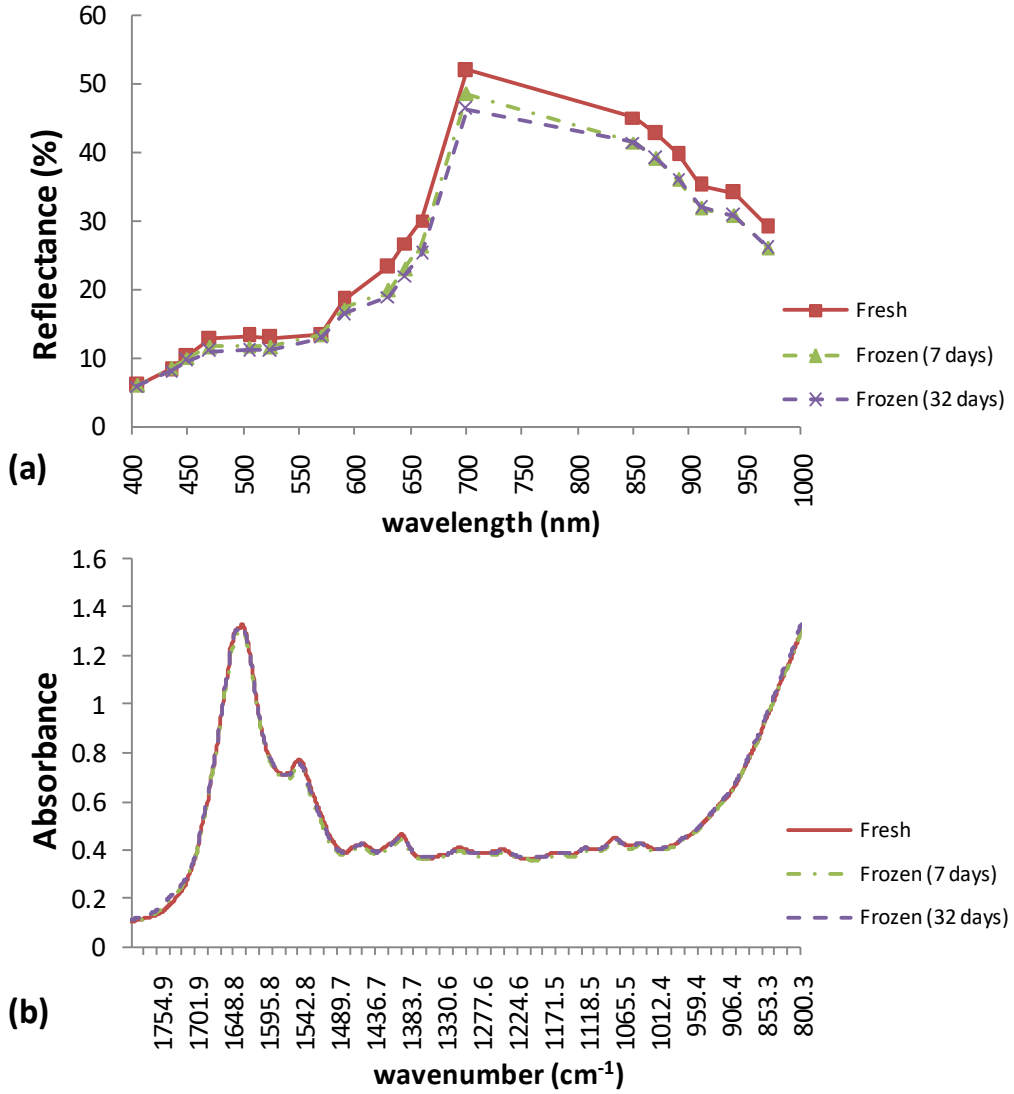


Figure 3.2 Typical (a) MSI and (b) FTIR spectra of minced beef for fresh, as well as frozen-then-thawed minced beef after 7 days and 32 days of storage at -20°C.

Similarly, differences in FTIR measurements are less evident, as small changes in absorbance in certain wavenumber ranges can be observed (Figure 3.2b). These changes are mostly observed in the area 1650-1620 cm^{-1} associated with water/ moisture and therefore relevant to the thawing process and 1560-1530 cm^{-1} assigned with the amide II band and previously related to the spoilage process (Ammor et al., 2009; Argyri et al., 2014).

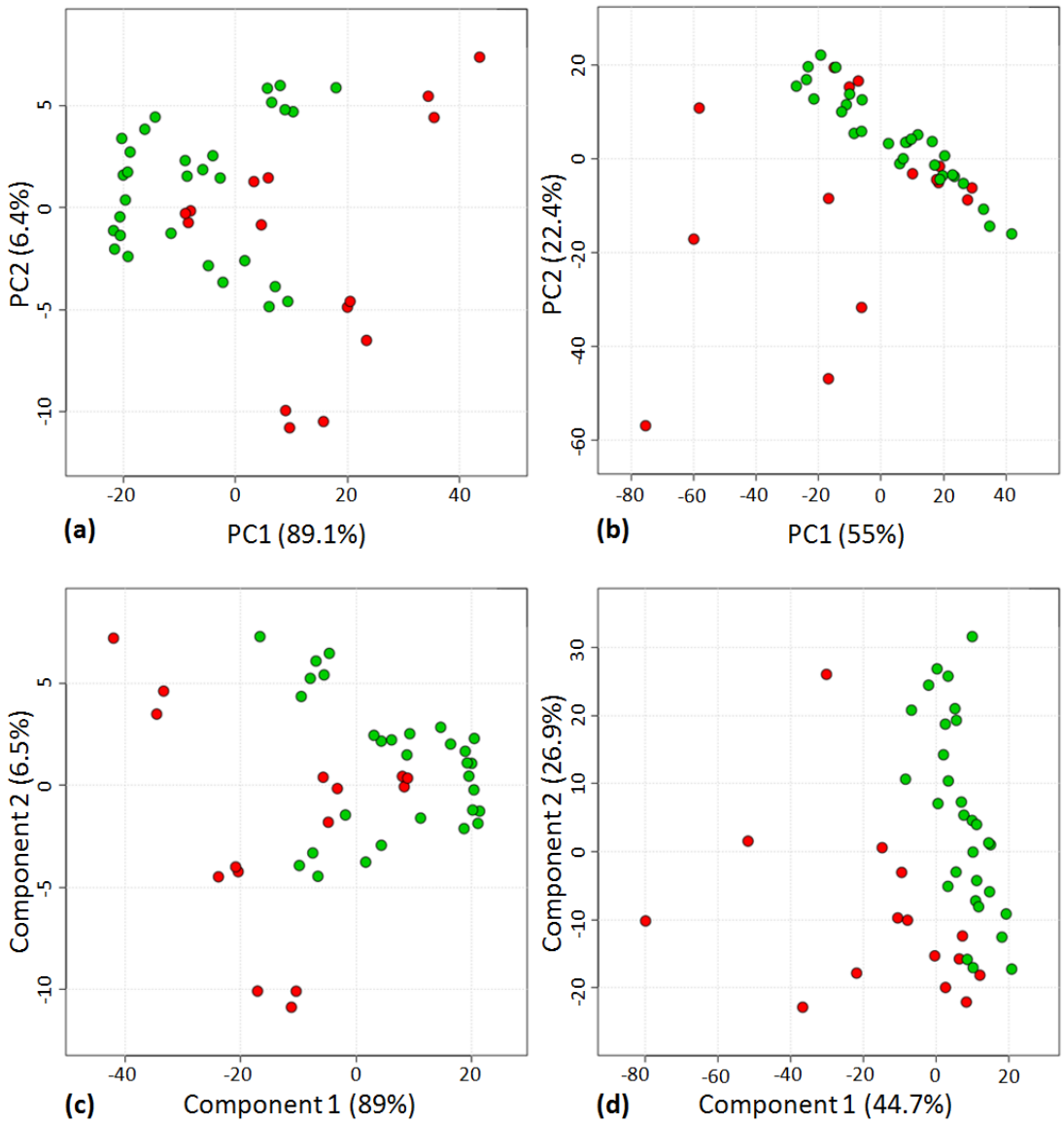


Figure 3.3 Principal component analysis scores for (a) multispectral imaging and (b) FTIR training data (Principal Components - PCs 1 vs. 2) and partial least-squares discriminant analysis scores for (c) multispectral imaging and (d) FTIR training data (PLS components 1 vs. 2). Red and green dots correspond to fresh and frozen-then-thawed samples.

The use of PCA revealed that in the case of multispectral data the separation, although not absolute, is more evident among fresh and frozen-then-thawed samples using only the first two principal components (Figure 3.3a and b). This can be attributed to the fact that the variance explained using only two components is more than 95% for MSI, whereas in the case of FTIR data is close to 77%. However, while these show

that there are some differences among the types of samples, it does not provide any conclusions, as more than two components may be used to describe the data. As a supervised technique, the PLSDA model is developed in order for components to sharpen the separation between groups, which is not the case for PCA. Therefore it is logical to assume that the PLSDA components of the training data will display a better discriminatory power for FTIR data, as is the case in Figure 3.3c and d. VideometerLab data are displayed similarly in PCA and PLSDA score plots (Figure 3.3a and c).

It should be noted that total viable counts (TVC) ranged from 4.85 to 8.06 log CFU/g, displaying variability not only in terms of meat batch but also in terms of microbial quality. These microbiological measurements are also available in the supplementary material (Table S1, Appendix I).

3.3.2 MODEL VALIDATION & DISCUSSION

As mentioned in Section 3.2.4, models were developed based on the training set (45 samples), a test set (30 samples) used for validation, while batches 6 and 7 (30 samples) were reserved for external validation.

In the case of PLSDA, the optimal number of components selected was 8 in both datasets, whereas in SVM with linear kernel optimal capacity was equal to 7 and 0.5 for MSI data and FTIR data, respectively.

The overall correct classification results are shown in Table 3.1, for both instruments.

Table 3.1 Overall correct classification (%) results for training, test and external validation sets for partial least squares discriminant analysis (PLSDA) and support vector machines (SVM) with linear kernel (multispectral imaging and FTIR spectra).

	Overall correct classification (%)		
	training set	*test set*	*external validation set*
Multispectral imaging			
PLSDA (8 components)	100.00%	100.00%	100.00%
SVM (Linear kernel, C=7.0)	100.00%	100.00%	100.00%
FTIR			
PLSDA (8 components)	100.00%	93.33%	96.67%
SVM (Linear kernel, C=0.5)	100.00%	86.67%	83.33%

As shown, in all cases the classification accuracy for the training set is 100%. This is not uncommon, as models were trained based on these samples, however in the case of multispectral data, all samples are classified correctly (Table 3.1, Table 3.2) in both testing and external validation sets. PLSDA also yielded a high correct classification rate of, 93.33 and 96.67% for testing and external validation, respectively, in the case of FTIR data (Table 3.1). SVM performance for FTIR data was inferior compared to PLSDA but still satisfactory presenting 86.67 and 83.33% correct classification rates for testing and external validation, respectively. This can be attributed to the large number of variables compared to the number of observations, as well as the high degree of collinearity observed among absorbance values for different wavenumbers of the FTIR spectra that may influence model performance. It is interesting to note that all fresh samples from batch 7 were misclassified as frozen-then-thawed (Table

3.2), while at the same time this batch had by far the lowest TVC (4.85 log CFU/g) as shown in the supplementary material (Table S1). Given that spoilage affects meat color and can be detected with multispectral imaging (Dissing et al., 2013; Panagou et al., 2014), it is likely that the SVM model performance was influenced by the lack of minced beef samples with a similar spoilage profile.

Table 3.2 Confusion matrices for testing and external validation samples, for both multispectral imaging and FTIR data.

VideometerLab - test samples		
	PLSDA & SVM - predicted class	
true class	fresh	frozen-then-thawed
fresh	10	0
frozen-then-thawed	0	20
VideometerLab - validation samples		
	PLSDA & SVM - predicted class	
true class	fresh	frozen-then-thawed
fresh	10	0
frozen-then-thawed	0	20
FTIR -test samples		
	PLSDA - predicted class	
true class	fresh	frozen-then-thawed
fresh	8	2
frozen-then-thawed	0	20

FTIR -validation samples		
	PLSDA - predicted class	
true class	fresh	frozen-then-thawed
fresh	9	1
frozen-then-thawed	0	20

FTIR -test samples		
	SVM - predicted class	
true class	fresh	frozen-then-thawed
fresh	10	0
frozen-then-thawed	4	16

FTIR -validation samples		
	SVM - predicted class	
true class	fresh	frozen-then-thawed
fresh	5	5
frozen-then-thawed	0	20

Indeed, PLSDA proved more robust in the FTIR case, as this method is good at dealing with data matrices having a large number of variables, which can also be highly correlated/ collinear (Biancolillo et al., 2014; Dai et al., 2014). Furthermore, the validation scheme employed has showed a balance among test and external validation results and therefore a case of model overfitting can be excluded. On the contrary, this balance is indicative of the models' ability to generalize and identify frozen-then-thawed samples regardless of within-batch and between-batch variability. Lastly, fresh samples were separated from frozen samples, regardless of the number of days stored and their microbiological profile.

While multiple instruments and minced beef have not been investigated before, the current results are in good agreement with previously published studies. Specifically, Al-Jowder et al. (1997) used MIR data from minced turkey, pork and chicken to discriminate between fresh and frozen-then-thawed samples. Principal component analysis scores shown for chicken displayed very good separation among samples, and while it was applied for different types of meat, the results suggest that the thawing process alters meat significantly. Furthermore, Downey & Beauchêne (1997) used NIR and factorial discriminant analysis (FDA) for beef (*m. longissimus dorsi*) subjected to multiple (1-3) freeze-thawing cycles. While the model was unable to differentiate among samples subjected to different number of cycles, all thawed samples were classified as thawed. In total, only three of the 16 fresh samples were misclassified and the OCC was 95.3%, which is similar to the PLSDA results obtained in this work for the FTIR data. In the case of HSI, OCC was over 97%, reaching in some cases 100% after wavelength selection using various modeling techniques (Barbin, Sun, et al., 2013; Ma et al., 2015; Pu et al., 2015). PLSDA was also employed by Barbin et al. (2013) and Ma et al. (2015). The former yielded 100% OCC after wavelength selection, which is consistent with the current PLSDA results. Optimal wavelengths were selected based on regression coefficients. The latter yielded similar results (97.73%) for the three-class case (fresh, thawed once & thawed twice) while incorporating textural features in the analysis. Textural characteristics were also employed by Pu et al. (2015) and OOC was 100% after variable selection (40 wavelengths). The MSI sensor used for this work was therefore appropriate as it yielded similar results with the HSI sensors after wavelength selection. However, it should be noted that in these cases -even when different meat batches were employed- samples from similar animals (age, weight, feeding environment) and/or from specific

retailers, local markets (Barbin, Sun, et al., 2013; Ma et al., 2015; Pu et al., 2015), or commercial slaughter-house were acquired (Downey & Beauchêne, 1997).

In conclusion, rapid methods coupled with data analysis methodologies can be utilized in monitoring compliance to label and specifically the detection of frozen-then-thawed minced beef. The generalization ability of the models utilizing meat batches with no prior knowledge of quality and product history is also highly significant. MSI proved slightly superior and left no room for a combination of sensor data at a low, mid- or high- level as applied in other cases (Biancolillo et al., 2014). The validation scheme proved effective in avoiding overfitted models, which in turn could be used for screening purposes provided that the original dataset is extended.

Acknowledgements

This work has been supported by the project "Intelligent multi-sensor system for meat analysis - iMeatSense_550" co-financed by the European Union (European Social Fund-ESF) and Greek national funds through the Operational Program "Education and Lifelong Learning" of the National Strategic Reference Framework (NSRF) Research Funding Program: ARISTEIA-I.

CHAPTER 4:

SPOILAGE ESTIMATION WITH MACHINE LEARNING

4 ABSTRACT

This chapter explores the modeling of spoilage using various machine learning methodologies and two different datasets; based on two separate studies. The first has been accepted in the scientific journal "Food Research International" and both the manuscript and supplementary files are available in Appendix I, while the second was accepted as a poster presentation. The studies and the author's contribution are presented below:

- Estelles-Lopez, L., Ropodi, A., Pavlidis, D., Fotopoulou, J., Gkousari, C., Peyrodie, A., Panagou, E., Nychas, G.-J., Mohareb, F. (2017). An automated ranking platform for machine learning regression models for meat spoilage prediction using multi-spectral imaging and metabolic profiling. *Food Research International, In press*. http://doi.org/10.1016/j.foodres.2017.05.013, where the author of this thesis was involved in the laboratory work (data acquisition from instruments).
- Ropodi A.I., Panagou E.Z. and Nychas G.-J.E. (2015) "Assessment of minced beef spoilage using Fourier Transform InfraRed (FTIR) spectroscopy, ensemble learning and Artificial Neural Networks (ANNs)", *9th International Conference on Predictive Modelling in Food (ICPMF9), Rio de Janeiro, Brazil, 8-12 September 2015*, [P.104], where the author was responsible for the data analysis employed and also contributed in the data acquisition process and submission.

The following sections present a modified version of the submitted version focusing mostly on the author's involvement and results of the machine learning methodologies.

Briefly, in Section 4.1 using "MeatReg" -a web-based application- regression models for the estimation of spoilage have be tested and compared. The results are visualized for various sensors and methodologies. The dataset includes samples stored in different temperatures and under aerobic and MAP conditions.

In Section 4.2, standalone artificial neural networks (ANNs) and combined ANNs in an ensemble-based approach, where multiple models are employed for the final prediction result, are used for spoilage estimation of minced beef. This method is tested for TVC prediction with FTIR data.

4.1 ESTIMATION OF SPOILAGE WITH "MEATREG"

4.1.1 INTRODUCTION

As mentioned in Chapter 1, spoilage is subjective among consumers, usually associated with the presence of gross discoloration, strong off-odours and development of slime due to intrinsic or extrinsic factors (Nychas et al., 2008) and involving very different microbial groups depending on the storage conditions e.g. packaging, temperature (Doulgeraki, Ercolini, Villani, & Nychas, 2012). However, there are some objective criteria, such as those set by EU authorities (European Commission, 2005) where the quality of fresh meat is evaluated only by viable counts of bacteria able to grow on very generic medium (Total viable counts) or on counts of the Enterobacteriaceae family. In addition, it is well established that pseudomonads are the major cause of spoilage in aerobic conditions producing slime and off-odours, while other microorganisms like *B. thermosphacta*, Enterobacteriaceae and lactic acid bacteria are dominating under modified atmosphere packaging (Nychas et al., 2008). While counting colonies is certainly time-consuming and costly, recently, rapid, non-invasive methods relying on processing large datasets using computational analysis are gaining popularity (Ropodi et al., 2016). On the other hand, while research involving these data with machine learning techniques is extensive, guidelines to choose the machine learning method that provides the best results for a specific type of data are still needed.

Therefore, the aim of this work is (i) to develop spoilage prediction models from data derived from different analytical instruments, and (ii) to implement an accuracy ranking system through a platform (MeatReg), which assesses the suitability of machine learning methods to specific type of metabolic data provided by a certain

analytical process. For this study, metabolomics data from minced beef samples stored under aerobic and modified atmosphere packaging were collected using five different analytical and imaging instruments: electronic nose (e-nose), High Performance Liquid Chromatography (HPLC), Fourier Transformed Infrared Spectroscopy (FTIR), Gas Chromatography coupled to Mass Spectrometry (GC-MS) and multi-spectral imaging (MSI).

4.1.2 MATERIALS & METHODS

Experimental design

Fresh minced meat was obtained from a central butcher shop in Athens and transported under refrigeration to the laboratory within 30 min. Portions of approximately 75-80 g were placed on styrofoam trays, were stored in aerobic or under modified air packaging (MAP) conditions (20% CO_2/ 80% O_2) at 4 and 10°C. For aerobic storage, samples were covered with plastic food membrane for domestic use and for MAP storage, samples were packed into plastic pouches of gas permeability at 20°C and 50% relative humidity of *ca*. 25 and 90 cm^3/m^2 per day/10^5Pa for CO_2 and O_2 respectively, using a HenkoVac 1900 Machine. At appropriate time intervals (approximately every 24 and 12 hours for the case of 4°C and 10°C respectively), multispectral images of duplicate samples were captured and samples were analyzed microbiologically until spoilage was pronounced and sub-samples were stored (-20°C) for FTIR, HPLC, GC-MS and e-nose measurements. Additionally, three more samples at 0 hours (control samples) were analyzed. In total, 105 samples (three control samples and 11-14 duplicate samplings per packaging condition per storage temperature) were analyzed.

Microbiological analysis

Twenty-five gram-portions from each meat sample were weighted aseptically in 400ml sterile stomacher bags (Seward Medical, London, United Kingdom), containing 225ml of sterile quarter Ringer's solution (LabM Limited, Lancashire, United Kingdom) and were homogenized for 60 sec (Lab Blender 400, Seward Medical). Appropriate serial dilutions were prepared with the same Ringer's solution and duplicate 0.1 or 1 mL samples of the appropriate dilutions were spread or mixed on the following media: plate count agar (PCA, Biolife 4021452, Milano, Italy) for total viable counts (TVC), incubated at 30 °C for 48–72 h; Pseudomonas agar base (PAB, Biolife 401961,Milano, Italy) for *Pseudomonas* spp., incubated at 25 °C for 48–72 h; streptomycinthallous acetate-actidione agar (STAA, Biolife 402079, Milano, Italy) for *B. thermosphacta*, incubated at 25 °C for 72 h; and deMan–Rogosa–Sharpe medium (MRS, Biolife, 4017282, Milano, Italy) with pH adjusted to 5.7 with 10 N HCl, for lactic acid bacteria overlaid with the same medium and incubated at 30 °C for 48–72 h. All plates were examined visually for typical colony types and morphological characteristics that were associated with each growth medium. Moreover, the selectivity of each medium was routinely checked by Gram staining and microscopic examination of smears prepared from randomly selected colonies.

Data acquisition

MSI and FTIR measurements were acquired using VidometerLab (VM) and Jasco 6200, as described in Chapters 2 and 3 previously.

Since HPLC, GC-MS and e-nose measurements are beyond the scope of this thesis, the data acquisition and data extraction for these instruments are presented in detail in the accepted manuscript (Estelles-Lopez et al., 2017), in Appendix I.

Data analysis

In particular in this study, seven machine learning methods; namely Ordinary Least Squares regression (OLS-R), Stepwise Linear regression (SL-R), Principal Component regression (PCR) , Partial Least Squares Regression (PLSR), Support Vector Regression (SVR), Random Forests Regression (RF-R) and k-Nearest Neighbours' Regression (kNN-R) were used to predict bacterial counts for Pseudomonads, Lactobacilli, *B. thermosphacta* and Enterobacteriaceae, as well as for TVC (Figure 4.1). This way, the most suitable analytical platforms to predict bacterial counts for each type of bacteria present in meat stored under aerobic or modified atmosphere conditions were identified and machine-learning methods were ranked for each scenario according to their performance.

OLS-R and SL-R are linear regression methods (Hastie, Tibshirani, & Friedman, 2009) and were performed using the "lm" function present in the package "stats" in R base. PCR, and PLSR were implemented using the "pcr" and "pls" functions with 3 components from the "pls" package (Mevik & Wehrens, 2007), Random forest was implemented with 200 trees using the "randomForest" function from the "randomForest" package (Liaw & Wiener, 2002). For SVR the implementation was performed using the "svm" function from the "e1071" package (Dimitriadou et al., 2011) and the kernel selected was the radial basis kernel (RBF). The values for grid search were 0, 0.03, 0.06, 0.09, 0.20, 0.30, 0.60 and 0.90 for gamma; 0, 0.02, 0.04,

0.06, 0.10, 0.20, 0.30 and 0.60 for epsilon; and 1, 4, 7, 10, 13, 16, 19, 30, 50, 70, 90 and 110 for the cost. Finally, kNN, predicts the value of an unknown sample based on the values of the *k* samples closer in distance. It was implemented using the function "knn.reg" from the "FNN" package (Beygelzimer, Kakadet, Langford, Arya, & Mount, 2013).The best *k* was selected using grid search from k=4 to 10. It should be noted that no data pre-treatment was performed prior to analysis.

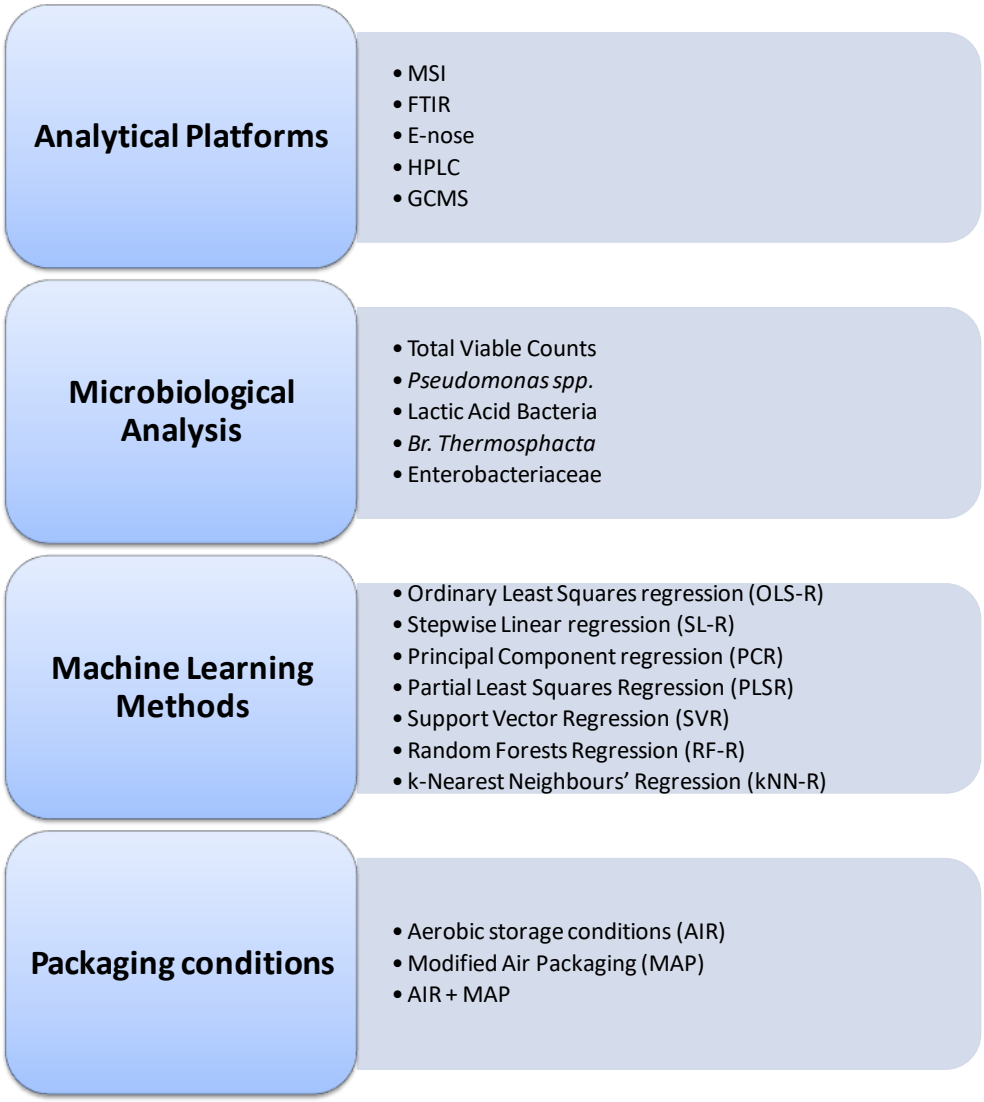


Figure 4.1 All possible combinations for model development in MeatReg.

For the methods that required grid search the dataset was split into training and testing. The training dataset was used to perform a model for each one of the possibilities of the grid search. The performance of each one of the models was calculated using the testing dataset and the parameters, which provided the model with the lowest RMSE, were selected. Furthermore, to assess the performance of the machine learning algorithms, Monte Carlo cross validation (Xu & Liang, 2001) was used. For K=20, the samples from each dataset were randomly distributed into training and testing datasets in 20 different splits. Then, the performance is calculated as an average of the performance of the 20 models.

Results were evaluated by calculating:

- the root-mean-square error (RMSE), quantifying the difference between predicted and observed values
- Δ_{max}, the maximum difference between predicted and observed values
- the bias factor (B_f) and the accuracy (A_f) factor, indicating the difference between observed and predicted values and the systematic bias (Argyri, Panagou, Tarantilis, Polysiou, & Nychas, 2010; Ross, 1996)
- the "accuracy" metric, i.e. the percentage of samples correctly predicted, where a sample is considered correctly predicted if the difference between predicted and observed values is less than 1 log CFU/g (Mohareb et al., 2016).

4.1.3 RESULTS & DISCUSSION

In Figure 4.2- Figure 4.4, an intuitive mapping of the results is presented based on RMSE.

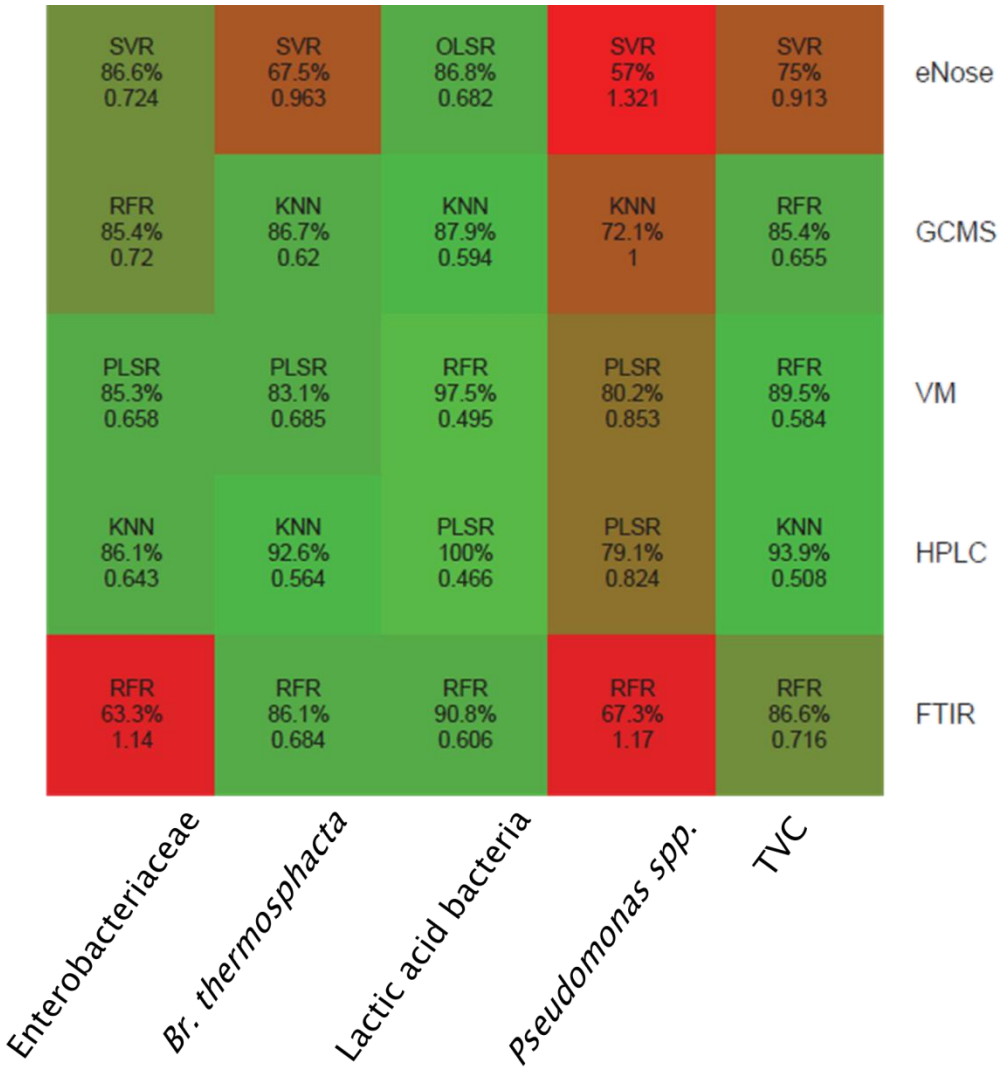


Figure 4.2 MeatReg results for samples stored in aerobic conditions**.**

In the plots both RMSE and accuracy values are mentioned based on the best machine learning method, per microorganism and analytical instrument. Light green tones represent that a good prediction of the bacterial count whilst red tones illustrate that no machine learning method provided a good performance.

In particular, Figure 4.2, Figure 4.3, Figure 4.4 refer to results concerning aerobic storage conditions, MAP and both respectively. The analytical performance per platform and machine learning method can be found in the extracted MeatReg reports (Estelles-Lopez et al., 2017, supplementary material), also provided in Appendix I.

However, the most relevant results will be mentioned below, focusing mostly on MSI and FTIR.

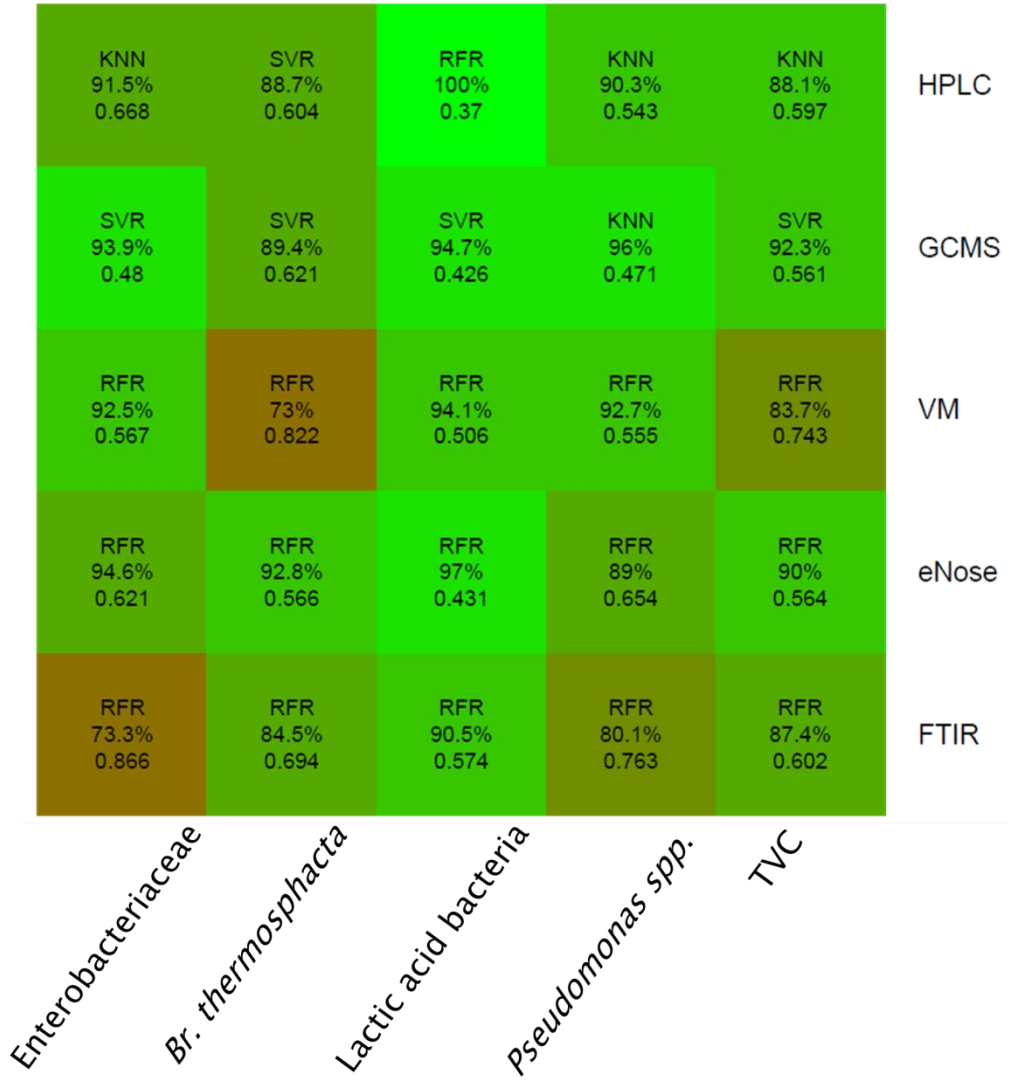


Figure 4.3 MeatReg results for samples stored in MAP.

Model's performance validation revealed that RMSE values ranged from 0.370 to 1.321 across different best machine learning methods to predict each of the species counts for AIR, MAP conditions separately (Figure 4.2, Figure 4.3). Furthermore, for mixed storage conditions (AIR + MAP) showed comparable results to storage-specific models; with RMSE values ranging from 0.388 to 1.343 (Figure 4.4).

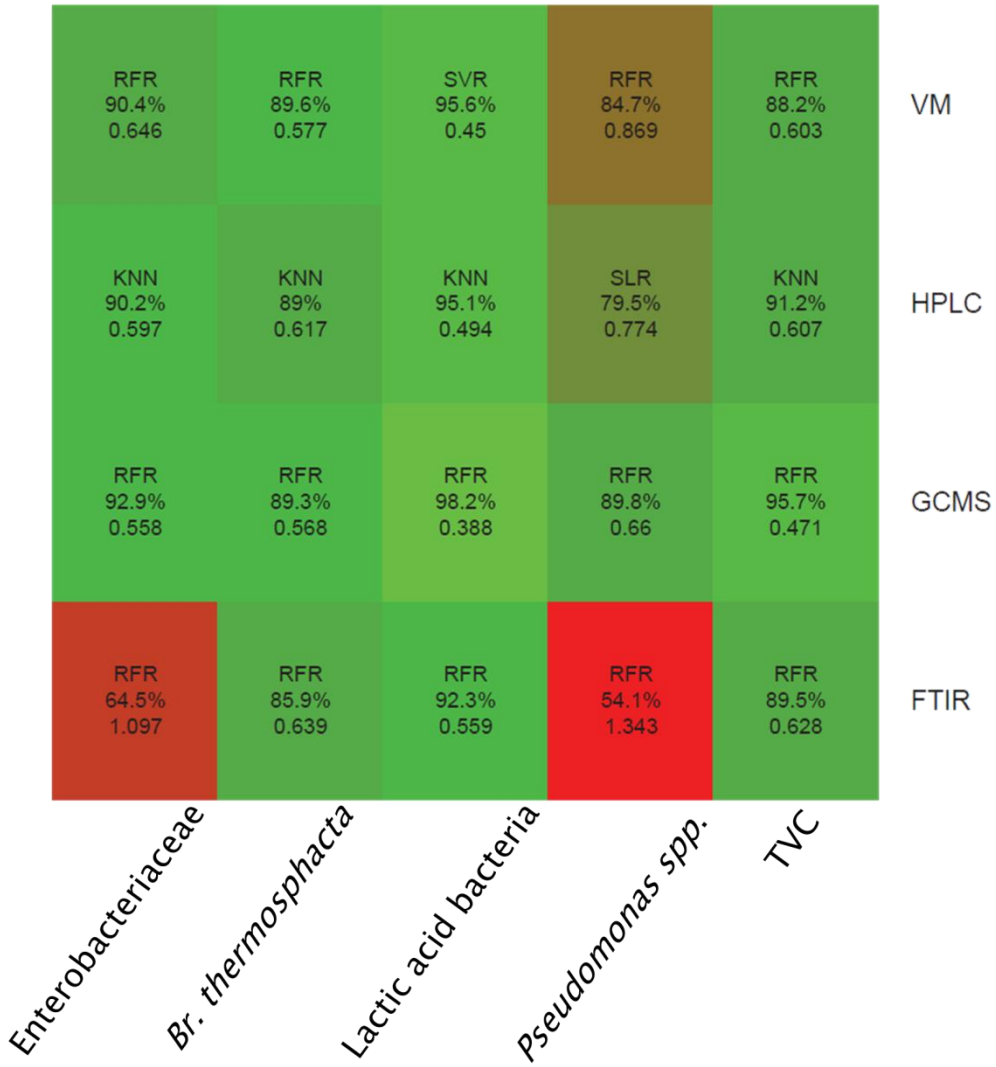


Figure 4.4 MeatReg results for samples stored in aerobic and MAP conditions.

Under AIR packaging storage (Figure 4.2), the best overall prediction outcomes were obtained from data derived with HPLC and MSI. Pseudomonads, were best predicted using MSI combined with PLSR (RMSE = 0.853), while HPLC achieved 100% prediction accuracy for Lactobacilli (RMSE = 0.466) combined with RF-R, 93.9% and 92.6% combined with kNN-R for TVC (RMSE = 0.508) and *B. thermosphacta* (RMSE = 0.564) respectively. MSI proved very good (accuracy = 97.5%, RMSE = 0.495) for Lactobacilli estimation. E-nose achieved over 86% for Enterobacteriaceae and lactic acid bacteria respectively.

For MAP storage (Figure 4.3), models achieved a better overall performance as the RMSE ranged from 0.370 to 0.866, with RF-R providing the best predictions in most cases. In the case of MSI, RF-R yielded 94.1, 92.7 and 92.5% accuracy for Lactobacilli, Pseudomonads and Enterobacteriaceae respectively. In addition to this, the RMSE ranged from 0.506 for Lactobacilli to 0.822 for *B. thermosphacta* both using RF-R. While FTIR results were slightly inferior, again RF-R was the best method and RMSE ranged from 0.574 for Lactobacilli to 0.866 for Enterobacteriaceae.

For e-nose, the RMSE ranged from 0.431 for Lactobacilli to 0.654 for Pseudomonads. For GC-MS data, the RMSE ranged from 0.426 for Lactobacilli to 0.621 for *B. Thermosphacta* and the accuracy was above 89% for all bacterial counts.

The best predictions for Enterobacteriaceae, were obtained using GC-MS, with an RMSE of 0.480 and an accuracy of 93.9%. For Lactobacilli the best RMSE was obtained using HPLC in tandem with RF-R. For the total viable counts, GC-MS and e-nose in tandem with SVR and RF-R provided predictions with an RMSE of 0.561 and 0.564 respectively. The best predictions for pseudomonads, were obtained using GC-MS with kNN-R (RMSE = 0.471 and 96.0% accuracy) and the best predictions for *B. Thermosphacta*, was obtained using e-nose combined with RF-R (RMSE = 0.566 and 92.8% accuracy).

In the case of mixed samples (AIR + MAP), the packaging type was not included as a model input variable, in order to evaluate the performance of the developed pipeline in achieving a good prediction accuracy regardless of the packaging system applied (Figure 4.4). MSI yielded good accuracy (over 88%) for all cases, except for

pseudomonads and RMSE ranged from 0.45 for lactic acid bacteria to 0.869 for pseudomonads. While FTIR data had the worst prediction performance, it did well in the prediction of lactid acid bacteria (RMSE = 0.559 and accuracy = 92.3%) and TVC (RMSE = 0.628 and accuracy = 89.5%) combined with RF-R. For this case, GC-MS coupled with RF-R achieved the best prediction accuracy throughout all growth media, followed by HPLC coupled with KNN-R (Figure 4.4). The best predictions for Enterobactriaceae, *B, thermosphacta*, Lactobacilli, Pseudomonads and total viable counts were all achieved using random forest with the RMSE at 0.558, 0.568, 0.368, 0.66, and 0.471 respectively.

In total, this work explored in detail the concept of using rapid analytical techniques based on sensors requiring a limited amount of sample and non-destructive/ non-invasive sensors for rapid microbial quality estimation. Indeed, the implementation of these rapid techniques coupled with data analysis methods has given promising results in several food products (Ropodi et al., 2016). MeatReg is the first attempt to compare machine-learning methods to determine the suitable instrument and machine learning methodology and visualize the results in an intuitive way. This analysis could be used to evaluate (i) spoilage regardless the storage conditions e.g. temperature, packaging etc., and (ii) the feasibility of using in the food sector specific instruments/ sensors.

More specifically, MSI proved a suitable choice for the prediction of Pseudomonads and Enterobacteriaceae counts, as they are well known to be related with meat spoilage under aerobic packaging conditions (Nychas et al. 2008). FTIR results were less accurate, but satisfactory nonetheless.

Another important factor that has to be taken into account is the convenience in terms of portability of the instruments and their ability to be used for in- or on-line applications. For example, due to technological advances MSI is quite convenient due to its portability and can be part of the food industry production line, while providing very fast and accurate results. On the other hand, although HPLC and GC-MS usually provide accurate predictions, they are not portable and cannot be deployed within the production chain. They could, however, be used by authorities in their analytical laboratories for supplementary analyses. Electronic nose and FTIR are also convenient methods and provide good prediction (>86%) for TVC, but the lack of accuracy in the case of pseudomonads in meat stored under aerobic packaging, can be a drawback.

In terms of validation approaches, Monte Carlo cross-validation was used in order to decrease the risk of overfitting (Xu & Liang, 2001). Other common strategies involve LOOCV methodologies and randomly splitting the original dataset into training and testing datasets (Westad & Marini, 2015), but the calculation of the statistical parameters highly depends on which samples fell into each dataset. In contrast to splitting the dataset into training and testing datasets only once, this method is more reliable to assess the performance as it does not depend that much in which samples randomly fall into the training and testing subsets (Mohareb et al., 2016).

As far as the models performance for each of the applied analytical instruments is concerned, MSI, HPLC and GC-MS provide good predictions for all types of counts. The predictions regarding Enterobacteriaceae and *B. thermosphacta* are limited for FTIR and e-nose respectively, while both failed to predict pseudomonads counts under aerobic packaging. Additionally, results show that kNN-R and RF-R could

provide high prediction accuracy and be promising methods. FTIR's less accurate performance could be attributed to (a) no data preprocessing was performed, e.g. for baseline correction, and (b) the fact that a huge amount of data are derived in comparison with HPLC (18 peaks), e-nose (12-20 peaks), MSI (18 wavelengths) and GC-MS (130 peaks). It should be taken into account whether data mining should be applied before the implementation of 'MeatReg'.

To sum up, HPLC data is very suitable to predict bacterial counts in most cases. MSI and GC-MS data are suitable combined with various methods. However, GC-MS is an unlikely candidate to be introduced in the industry due to its high cost, high maintenance and size, compared to more portable MSI platforms. For lactic acid bacteria all the methods provided accurate predictions, while for Enterobacteria support vector machines, linear regression and stepwise linear regression provided the best results. For FTIR data, aside from Pseudomonads and Enterobacteria under aerobic conditions, RF-R provided the best predictions. Additionally, the results rank PLSR, kNN-R, PCR and SVR next, which work generally fine for all methods with some exceptions.

4.2 Ensemble Neural Networks for the prediction of microbial spoilage

In this subsection, standalone artificial neural networks (ANNs) and combined ANNs in an ensemble-based approach, where multiple models are employed for the final prediction result are employed. This method is tested for TVC prediction with FTIR data.

4.2.1 Introduction

Lately, rapid methods based on spectroscopy techniques have been applied in various food products. FTIR spectroscopy and ANNs have been applied previously indicating promising results (Argyri et al., 2013; Panagou, Mohareb, Argyri, Bessant, & Nychas, 2011). Panagou et al. (2011) used a multilayer-perceptron (MLP) for classification among fresh, semi-fresh and spoiled beef fillets based on sensory data, as well as prediction of TVC, whereas Argyri et al. (2013) used ANNs -combined with a genetic algorithm for fine-tuning- for the case of minced beef.

One common problem often associated with machine learning prediction models and ANNs in particular is overfitting, i.e. the case of good model performance in the training phase, but poor performance in unseen/ independent data. This can be attributed in various factors, for example the quality and size of the training set and how representative the samples are. Since ANNs are often very difficult to fine-tune (e.g. determining learning rate, number of layers and neurons per layer, choosing transfer function, etc.) training these models given the relatively smaller amount of samples/ measurements found in food applications -compared to other machine learning applications- can be very complicated. Models with similar performance in known datasets will display very different performance when tested on new data. To

avoid these failures, the ensemble-based methodologies have been proposed. This methodology is based on combining different models and their corresponding results using bagging and boosting (for creating random data repartitions and data resampling), therefore developing multiple classifiers or regression models and using a fusion method to determine the final result (e.g. majority voting in the case of classifiers) (Mohareb et al., 2016). Mohareb et al. (2016) employed this concept with e-nose data and SVMs for beef fillets with good results.

The purpose of with this work was two-fold; (a) to develop improved ANN models based on spectroscopic measurements for spoilage assessment and (b) explore the efficacy of an ensemble learning methodology.

4.2.2 MATERIALS & METHODS

Experimental design

Fresh minced beef of two different batches was purchased from a local retailer. Portions of 70-75g were placed onto styrofoam trays, packaged in air or under modified air packaging conditions (20% CO_2/ 80% O_2), and stored at 4 and 10°C. At appropriate intervals, four samples were analyzed; total viable counts (TVC) were determined and sub-samples were stored (-20°C) for FTIR measurements. In total, 168 samples were analyzed over 13 days.

The FTIR measurements and microbiological analysis methodology are described in detail in Section 4.1.2.

Model development

FTIR measurements were pre-processed applying Standard Normal Variate (SNV) transformation at the approximate wavelength range of 1800-800 cm^{-1} and subjected to PCA for dimensionality reduction, as ANN training can be computationally demanding. The PCs whose eigenvalues were equal to or greater than one; were selected. The resulting PCA scores were then used for model development.

The dataset was partitioned randomly and approximately 20% of the measurements was retained for testing and the remaining was used for model calibration. After some initial experimentation, the ANNs used included an input layer (18 neurons), two hidden layers with hyperbolic tangent sigmoid transfer function and a single output neuron with linear (identity) transfer function. To avoid overfitting, early stopping criteria were employed and a small subset was used for internal validation. The calibration set was then split anew and ~80% was used for model training and the rest for internal validation. Each feed-forward ANN was trained for up to 200 epochs, with early stopping if the internal validation MSE showed no improvement for 5 epochs. A 0.01 learning rate was employed and momentum was set to 0.9, while the BFGS quasi-Newton back-propagation algorithm was employed for training (Dennis Jr. & Schnabel, 1996; Sumathi & Surekha, 2010). For the ensemble approach, various combinations of neurons per layer were investigated with maximum neurons being 30. The "fused" decision was the average TVC of all ANNs that followed certain criteria based on MSE and Coefficient of correlation (R).

All models were developed using MATLAB software ("MATLAB 2012a, The MathWorks, Inc., Natick, Massachusetts, United States").

4.2.3 RESULTS & DISCUSSION

Based on both mean squared error (MSE) and correlation coefficient (R) during calibration and internal validation, 10 ANNs were selected for the final model. Criteria consisted of: (a) R being greater than 0.9, (b) MSE ≤ 0.15 $(\log CFU/g)^2$ for both training and internal validation, (c) the absolute difference in MSE (|ΔMSE|) being smaller or equal to 0.1, indicating a balance in results and (d) the total number of neurons in the hidden layers being no more than 30.

Table 4.1 MSE values for the selected ANNs for training, internal validation and testing set.

# ANN	# neurons in layer1	# neurons in layer 2	MSE_{tr}	MSE_{intval}	|ΔMSE|	MSE_{test}
1	3	1	0.09	0.10	0.00	0.17
2	9	9	0.10	0.11	0.01	0.25
3	6	9	0.10	0.12	0.02	0.18
4	12	4	0.09	0.13	0.04	0.24
5	4	15	0.11	0.14	0.03	0.20
6	5	2	0.08	0.14	0.06	0.23
7	16	11	0.10	0.14	0.04	0.16
8	9	8	0.10	0.14	0.04	0.26
9	5	5	0.06	0.14	0.08	0.21
10	6	17	0.06	0.15	0.08	0.22

* MSE_{tr}, MSE_{intval}, MSE_{test}: Mean squared error for training, internal validation and testing sets respectively, |ΔMSE|: absolute difference of training and internal validation MSEs.

Models were then tested using the original test set. Results showed that ANNs with two hidden layers performed well in terms of mean squared error (MSE), however testing results were mixed. In Table 4.1, the MSE values for training, internal validation and testing of the selected ANNs are presented.

Their standalone performance of the best ANNs for the assessment of TVC in terms of MSE_{test} $(\log CFU/g)^2$ varied from 0.16 to 0.26.

Additionally, the ensemble model which was created combining the former ANNs yielded an improved MSE equal to 0.16 compared to the majority of ANNs explored,

while R_{test} was approximately 0.91.While the "best" ANN (#7) seems equally good, based on MSE values, it is interesting to note that another ANN would have been chosen, therefore yielding worse results.

Furthermore, these results in terms of MSE/ RMSE are improved compared to previous works, as GA-ANNs yielded RMSE = 0.71 log CFU/ g (MSE=0.50) for only one storage temperature (5°C) and the error in the case of beef fillets was larger.

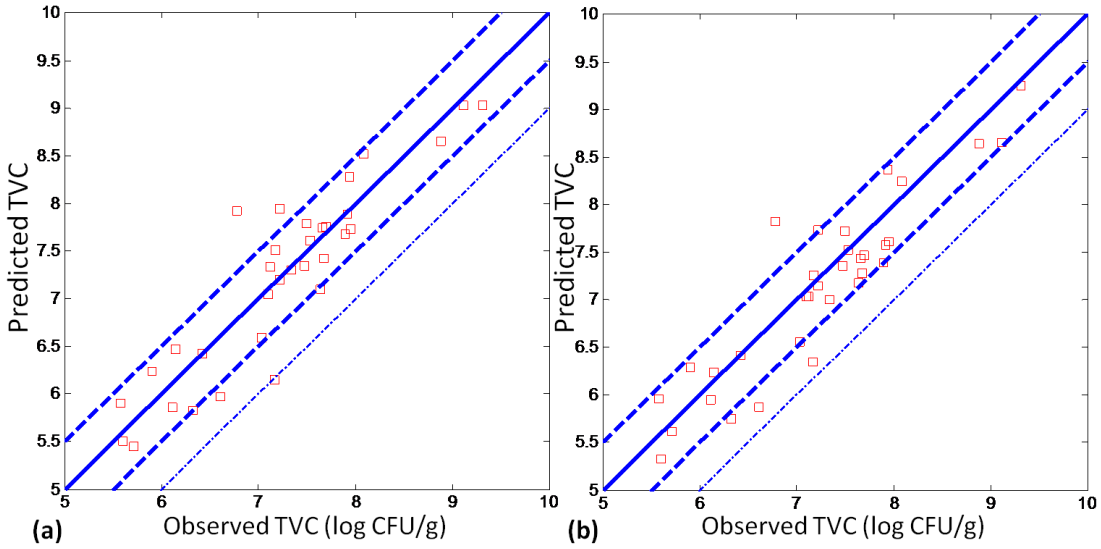


Figure 4.5 Observed *vs.* predicted TVC values for Ensemble (a) and best-case ANN (b).

Lastly, in Figure 4.5 the observed *vs.* predicted TVC values for the ensemble model (a) and the best -in terms of testing MSE- ANN (b) are presented. The error (Δv = observed-predicted values) is within the ±0.5 log CFU/g area (dashed blue lines) for most cases, and there is only one sample where marginally Δv is greater than 1 log CFU/g (dash-dotted line) and specifically $\Delta v \approx 1.02$, in the case of the ensemble approach.

In conclusion, FTIR spectroscopy, ensemble learning and ANNs can be used for the assessment of microbiological quality of minced meat. In fact, by combining more

than one ANN, more robust predictions have been achieved. This approach avoids choices based on overoptimistic criteria or unfavorable data partitions.

In the future, this model is can be further validated using an independent dataset in the same and/or dynamic temperature conditions.

Acknowledgements

This work has been supported by the project "Intelligent multi-sensor system for meat analysis - iMeatSense_550" co-financed by the European Union (European Social Fund-ESF) and Greek national funds through the Operational Program "Education and Lifelong Learning" of the National Strategic Reference Framework (NSRF) Research Funding Program: ARISTEIA-I.

CHAPTER 5:

DISCUSSION, CONCLUSIONS AND NEXT STEPS

5 INTRODUCTION

In the previous chapters (Chapter 2-4) various applications of multispectral imaging and FTIR spectroscopy in tandem with advanced data analysis methodologies have been explored. These applications involve:

I. Detection of minced meat adulteration using MSI (Chapter 2);
    a. Detection and quantification of minced beef adulteration with pork and vice versa (from now on "Case 1")
    b. Detection of minced beef adulteration with horsemeat combined with storage time ("Case 2")
II. Detection of frozen-then-thawed *vs.* freshly-ground minced beef using MSI and FTIR, presented in Chapter 3 and for our purposes called "Case 3".
III. Spoilage estimation (Chapter 4);
    a. Using multiple sensors and comparing results from various machine learning methodologies (from this point on referred as "Case 4")
    b. Using FTIR, ANNs and ensemble-based approaches ("Case 5").

Other cases related to the work in this thesis involving different datasets and/or other methodologies are presented in Appendix III.

Analyzing the abovementioned cases, the efficacy of developing machine learning approaches for different applications and datasets using imaging and spectroscopy instruments is explored. Results showed satisfactory model performance in general terms, but also indicated the difficulties both in model development and validation, and possible real-life applications and scenarios. In the following sections, the results

per case are presented along with an extended discussion on the methodologies employed, the experimental design, the model development approaches used and their validation.

5.1 RESULTS PER CASE

In this section results are shown per case:

Case 1:

The aim of this study was to investigate the potential of multispectral imaging for the detection of minced beef fraudulently substituted with pork and vice versa coupled with data analysis techniques. In four independent experiments, a different batch of beef and pork meat were acquired, ground and the appropriate amount of minced beef and pork was mixed creating nine different proportions of adulteration (10% step) and two categories of pure pork and beef. Multispectral images in 18 different wavelengths -55 samples per experiment and 220 in total- were acquired for this work.

After an image processing step, as an unsupervised technique hierarchical cluster analysis - HCA (Everitt et al., 2011) was employed per batch to explore the relationship among adulteration classes using Euclidean Distance and Ward's minimum variance agglomeration method. In addition to this, PCA was applied in order to visualize any significant possible differences and groupings among samples, as well as among classes, both per batch as well as using the first three batches. The supervised techniques employed were LDA (Fisher, 1936) and PLSDA so as to

discriminate (a) among the eleven adulteration classes and (b) among the "pure beef", "pure pork" and "adulterated" samples.

Initial analysis with HCA and PCA showed that different meat batches display significant differences, a parameter that must be taken into account. This was evident when the batches were compared, as well as when the PCA was applied to the first three batches. In the latter case, while different classes per experiment were obvious, the PCA of three batches showed that classes were less evident and more difficult to differentiate.

On the other hand for the supervised techniques, a 60–40% stratified partition was applied on the first three batches, meaning 99 samples out of 165 of the three-batch dataset were chosen in a random way for model calibration the rest were used as a test set. The fourth batch (55 samples) was also reserved for independent model validation, in order to determine the models' ability to generalize regardless of meat batches. Model performance was measured in terms of recall (sensitivity) and precision, as well as overall correct classification (OCC) (Sokolova & Lapalme, 2009).

Results showed very good discrimination between pure and adulterated samples, for PLSDA and LDA, yielding 98.48% overall correct classification. For the 11-class case, OCC was slightly lower; however 98.48% and 96.97% of the samples were classified within a±10% category of adulteration for LDA and PLSDA respectively. When the models were further validated using the data of the fourth experiment for independent testing, all pure and adulterated samples were classified correctly in the case of PLSDA. While in the case of 11 categories, all pure samples were classified

correctly, and only 4 out of 45 adulterated samples were classified as pure. However, LDA was proved to be less accurate, where, in the case of pure *vs*. adulterated samples, the results yielded overall correct classification of 80% and mean per-class recall and precision of 85.93% and 77.08%, respectively. Lastly, this work proved that (a) a 10% adulteration with pork in beef and vice versa could be successfully identified and could thus be considered as a detection limit of the applied method and (b) the variability among meat batches should be taken into account -not always the case in previous works.

Case 2:

Following the results of case 1, the case of minced beef adulteration with horsemeat was explored with an added factor -storage under refrigerated conditions- being taken into account. In the first stage, similarly to case 1, different levels of adulteration with a 20% step were prepared, creating four levels of adulteration, as well as pure beef and horsemeat (b1). For every adulteration class, five samples were prepared and 30 images in total were acquired. Since the dataset displayed similar results compared to case 1 after preliminary analysis, for the second batch (b2) eight samples were created per adulteration level focusing on 60-40% and 80-20% and adding a 90-10% level (w/w) for beef and horsemeat respectively. The resulting 40 samples were prepared and multispectral images were acquired at the time, as well as after the samples were stored in high-precision incubators at 4°C for 6, 24 and 48 h. The same procedure was repeated with a third meat batch (b3) and samples were reserved for independent

validation purposes. In total, 110 samples were prepared and 350 images were acquired.

PCA analysis for b2 showed that freshly-ground and unadulterated samples were clearly separated from the rest of the samples, indicating a difficulty to differentiate adulterated and pure samples when stored for hours in refrigerated conditions. In terms of supervised approaches, PLSDA, RF and SVMs were employed. PLSDA and RF were used for classification among different percentages of beef (4 classes), pure beef *vs.* pure horsemeat *vs.* adulterated samples, pure meat (both beef and horse) *vs.* adulterated samples and freshly-ground *vs.* stored minced meat. While displaying high OCCs for training sets, they significantly underperformed in testing and independent validation. While this may imply that another methodology should be used, lack of generalization in the former models may also be attributed to a case of overfitting due to inadequate representative training samples. Indeed, b1 and b2 datasets include two batches of freshly ground samples, but only one batch of stored samples in refrigerated conditions, possibly leading to a lack of sufficient data for adulteration detection.

On the other hand, the improved results for freshly-ground *vs.* stored samples and PCA scores for b2 suggested that the classification problem could evolve as "freshly-ground and pure *vs.* other samples". The SVM model (SVM-1) with b2 training data and a radial basis function (RBF) kernel classified all training samples correctly (OCC=100%) for internal model validation, while for independent model validation OCC was equal to 99.38%, with 99.58% and 98.78% mean per-class recall and precision. Indeed, only one testing sample, a minced horsemeat sample stored for 6 h, was misclassified. Applying the SVM-2 model, five adulterated samples were

misclassified (OCC=84.38%) and all pure samples were classified as pure. The largest (w/w) percentage of horsemeat where samples were misclassified was 20% after independent validation. In addition, the misclassified sample of SVM-1 was classified as pure and as a consequence in the two-step SVM model employed all pure and freshly-ground samples were classified correctly and the overall correct classification was equal to 95.31% for independent batch validation.

Results show that pure and freshly-ground samples can be differentiated with a maximum error of 20%. Furthermore, the necessity of independent sample validation was again proven, while a significant parameter -changes occurring in color during storage- has been emphasized. Despite these difficulties, pure and freshly-ground samples were clearly distinguishable, proving that MSI could be used for large scale detection of adulteration.

Case 3:

In this case, the efficacy of FTIR spectroscopy and MSI data was explored and compared with the purpose of identifying frozen-then-thawed minced beef labeled as fresh. For this reason, freshly-ground beef was purchased from seven separate shops at different times; therefore providing seven different meat batches so as to improve model generalization and validation. For each batch, minced beef was divided in fifteen portions and placed in Petri dishes. For the first five samples, multispectral images were acquired, followed by FTIR spectroscopy measurements using~3g portions from each Petri dish. The remaining ten were frozen (-20°C) and stored for 7 and 32 days (5 samples per time interval). Then, samples were thawed and subjected

to similar data acquisition. All in all, 105 multispectral images and 105 FTIR spectra were collected. TVC ranged from 4.85 to 8.06 log CFU/g displaying variability in terms of microbial quality.

In terms of data analysis methods, PCA was used for data exploration, while PLSDA and SVM as a supervised approach. In the latter methods, a scheme for robust model training and validation was employed. Specifically, the first five batches were partitioned in a 60-40% stratified manner for training and testing respectively, so that , three fresh, three frozen for seven days and three frozen for 32 days samples per batch were used for training. In total, 45 samples were reserved for training and 30 for testing. The last two batches, consisting of 30 samples, were retained for independent batch validation, and ensuring model robustness.

The use of PCA revealed a case of differentiation, although not absolute, among fresh and frozen-then-thawed samples using only the first two principal components for MSI. This was not so evident with FTIR data, as the variance explained by the first two components was significantly lower. PLSDA and SVM yielded 100% correct classification for MSI test and external validation sets. FTIR proved less accurate, as PLSDA yielded 93.33 and 96.67% classification accuracy for the test and external validation set. SVM performance was inferior to PLSDA with 86.67 and 83.33% correct classification rates for testing and external validation, respectively. This can be attributed to the large number of variables compared to the number of observations, as well as the high degree of collinearity observed among absorbance values for different wavenumbers of the FTIR spectra that may have influenced model performance. In conclusion, results showed that rapid methods and especially MSI and data analysis methodologies can be utilized in monitoring compliance to label and detecting frozen-then-thawed minced beef. This was evident even in the cases of meat batches where

no prior knowledge of quality and product history was known.

Case 4:

In this case, meat spoilage with the use of multiple sensors was investigated. Fresh minced meat was purchased and. portions of approximately 75-80 g were placed on styrofoam trays, were stored in air or under modified air packaging (MAP) conditions (20% CO_2/ 80% O_2) at 4 and 10°C. Three samples at 0 hours (control samples) were analyzed and at appropriate time intervals (approximately every 24 and 12 hours for the case of 4°C and 10°C respectively), multispectral images of duplicate samples were captured and samples were analyzed microbiologically until spoilage was pronounced. Sub-samples were stored (-20°C) for FTIR, High Performance Liquid Chromatography (HPLC), Gas Chromatography coupled to Mass Spectrometry (GC-MS) and e-nose measurements. As a result, 105 samples were analyzed. Microbiological analyses included Pseudomonads, Lactobacilli, *B. thermosphacta* and Enterobacteriaceae counts, as well as TVC.

In terms of data analysis methods, seven were tested (Ordinary Least Squares regression - OLS-R, Stepwise Linear regression - SL-R, Principal Component regression - PCR , PLSR, SVR, RF regression - RF-R and k-Nearest Neighbours' Regression - kNN-R) for the prediction of bacterial counts. In total, 50 combinations were explored (two types of packaging – AIR & MAP, five types of instruments/ measurements and five species counts) using "MeatReg" a web-based application and were ranked according to the RMSE for each scenario and summary plots were exported. Results were also presented based on accuracy, i.e. percentage of sampled predicted within a ±1 log CFU/ g (Mohareb et al., 2016).

Results showed that RMSE ranged from 0.370 to 1.321 across different best machine

learning methods under AIR and MAP. It is interesting to note that when both AIR and MAP samples were included in the model results were similar to storage-specific models; ranging from 0.388 to 1.343. Under AIR packaging storage the best overall prediction outcomes were obtained HPLC and MSI data. HPLC achieved 100% prediction accuracy for Lactobacilli when using RF-R and 93.9% when combined with kNN-R for TVC. Pseudomonads were best predicted using MSI combined with PLSR. For MAP samples, the RMSE ranged from 0.370 to 0.866 and RF-R provided the best predictions in most cases. For FTIR, the RMSE ranged from 0.574 for Lactobacilli to 0.866 for Enterobacteriaceae, while kNN-R and RF-R provided the best results for HPLC and MSI. Specifically, for MSI measurements, the RMSE ranged from 0.506 for Lactobacilli, to 0.822 for *B. thermosphacta*. In general, GC-MS coupled with RF-R achieved the best prediction accuracy throughout all growth media, followed by HPLC coupled with KNN-R, while data derived from FTIR had the worst prediction performance, however the large amount of variables should be taken into account. The best predictions for Enterobactriaceae, *B. thermosphacta*, Lactobacilli, Pseudomonads and TVC, were all achieved using random forest.

Case 5:

Fresh minced beef of two different batches was purchased and portions of 70-75g were placed onto styrofoam trays, packaged in air or under modified air packaging conditions (20% CO_2/ 80% O_2), and stored at 4 and 10°C. At appropriate intervals, four samples were analyzed; total viable counts (TVC) were determined and sub-

samples were stored for FTIR measurements. In total, 168 samples were analyzed over 13 days.

In terms of data analysis, FTIR measurements were subjected to PCA for dimensionality reduction and multiple ANNs were trained that included an input layer, two hidden layers with hyperbolic tangent sigmoid transfer function and a single output neuron with linear (identity) transfer function. Based on both MSE and R, the best ANNs were selected and the final TVC prediction was the average predicted value per sample.

The best ANNs' standalone performance in terms of MSE $(\log \text{CFU/g})^2$ for the testing dataset varied from 0.16 to 0.26, whereas the ensemble model which was created combining the former ANNs yielded an improved MSE equal to 0.16 compared to the majority of ANNs explored. In fact, only one sample was marginally larger than 1 log CFU/g, proving the ability of ensemble techniques in providing robust results compared to standalone models.

5.2 MODEL DEVELOPMENT AND VALIDATION

Data preprocessing for MSI and FTIR data

As mentioned in Chapter 1, an image-preprocessing step is necessary in order to explore and analyze MSI data. This involves an image segmentation step that can be provided using standard instrument software as was done for this work (cases 1-4) where the results are part of a semi-supervised procedure, or by applying an automated process, as shown in previous study (Tsakanikas, Pavlidis, & Nychas,

2015). After the useful ROI is determined, various features/ variables may be extracted and used for model development. In fact, in several imaging applications measures based on statistics have been used previously with the most common being the pixel measurement, sometimes preprocessed using SNV, MSC, 1st and 2nd derivatives etc. (Barbin, Sun, et al., 2013; Feng & Sun, 2013). For cases 1-4, the mean reflectance values and their standard deviation per wavelength was employed thus incorporating in a simple way the differences among the pixel spectra in the spatial dimension and exporting 36 variables (18 mean values and 18 standard deviations) used previously with promising results (Panagou et al., 2014). Since the number of wavelengths is significantly lower compared to HSI cases, no variable/ wavelength selection scheme was employed, but in the past HSI applications have used various approaches, including GAs, PCA, using PLS regression coefficients and using uninformative variable elimination and successive projections algorithm (UVE-SPA) (Barbin, Sun, et al., 2013; Feng & Sun, 2013; Pu et al., 2015). Lately, however, other studies have explored the usage of more advanced image features, such as textural characteristics or textural characteristics combined with histogram statistics providing promising results, while other times achieving similar results with simpler image features (Huang et al., 2013; Ma et al., 2015; Pu et al., 2015).

As far as spectroscopy is concerned, the preferred approach used for preprocessing employed in cases 3 & 5 was SNV combined with autoscaling. SNV has been used and often been presented with other similar methods for comparison purposes extensively (Alamprese et al., 2013; Argyri et al., 2013; Barbin, ElMasry, Sun, Allen, & Morsy, 2013; Zhao et al., 2014). However, it should be noted that all these

preprocessing algorithms depend on the actual dataset/ use-case providing different results per case (Biancolillo et al., 2014; Coppa et al., 2014).

Experimental design; number of samples, dataset partitioning, cross-validation methodologies and calibration/ validation approaches

The quantitative and qualitative characteristics of the dataset, such as the number of samples, the size of the dataset (number of variables to be used), distribution of samples based on specific characteristics to be explored (e.g. TVC) or background knowledge information (different meat batches), are of at-most importance and effect greatly all aspects of model calibration, validation and overall performance.

Specifically, while in food applications it is common to have a relatively small number of samples, this could lead to random results (Nunes, Alvarenga, de Souza Sant'Ana, de Sousa Santos, & Granato, 2015). For this reason, it would be better to validate models using a test set, usually a small percentage of the original dataset to further validate the models (Westad & Marini, 2015), which was the approach -among other validation approaches- chosen in all cases (1-5). According to Westad and Marini (2015), if the former is not possible due to the number of samples, cross-validation (CV) and most often Leave-One-Out CV (LOOCV) and k-fold CV can be applied. The first method leaves one sample out of the calibration process, which is used for validation and the process is performed for all samples in the dataset. In the second method, the calibration data are separated in k "equal" subsets and each subset is then used iteratively for testing the models built from the rest of the data (Ropodi et al., 2016). The average value of the chosen performance metric is then used for model evaluation (e.g. average MSE). However, both approaches leave a lot to be desired,

especially for small sized datasets as LOOCV tends to give over-optimist results and k-fold partitioning could result in "biased" subsets where all categories/ types of samples are not represented in an equal manner. The latter has to be addressed not only in CV folds, but also in the calibration/ validation splitting of the dataset. To improve on this, usually a stratification strategy is employed based on background information, i.e. stratification across replicates, treatment, etc. Alternatively, methods that span the sample space as uniformly as possible e.g. the Kennard-Stone (KS) algorithm (Kamruzzaman et al., 2012) or methods like the Duplex algorithm, selection based on D-optimal criterion or even use of clustering techniques like k-means and Kohonen mapping may be employed as presented by Westad and Marini (2015). Multiple random partitions if the dataset is large enough can also give a better assessment of the model performance as was the case in Papadopoulou et al. (2013) for e-nose data where the model was evaluated on three train/ test partitions (Papadopoulou, Panagou, Mohareb, & Nychas, 2013).

For the abovementioned reasons, a large number of samples was used for model development in all cases. Specifically in case 5, 168 samples were utilized which also displayed variability in terms of microbiological quality (TVC) and storage conditions (two different temperatures, aerobic *vs.* MAP conditions). Therefore sufficient data were available for training and testing (33 samples for testing, approximately 20% of the dataset). In addition to this, while in case 4 the samples are 105 in total, the samples displayed variability as described previously and they included five types of measurements (GC-MS, HPLC, FTIR, MSI, e-nose), providing a rare source for comparison of those methods. To prove the robustness of the results, the "MeatReg" platform offers the ability of random train/ test splits (70-30%) performed in 20

iterations. Lastly, in the cases of 1-3, an even more complex and robust validation scheme was employed. This scheme involved the stratified partitioning for the creation of the train and test datasets where samples from multiple different batches, different quality/ adulteration level, replicates of the same category were included, as well as the use of one (cases 1, 2) or more (case 3) completely independent meat batches with multiple replicates for independent / external validation. In detail, in case 1, 220 samples from 4 batches were used in total. From the first three meat batches, 99 and 66 were selected for training and testing respectively, while the fourth batch (55 samples) was retained for external validation. In case 2, 110 samples and 350 images in total were utilized for the analysis, out of which 40 samples (with their corresponding 120 images) were used for validation. Lastly, in case 3, seven batches (105 samples) were purchased, out of which two batches (30 samples) were used for independent validation. From the remaining five batches, 45 samples were used for training and 30 for testing.

In case of a successful implementation of a machine learning/ chemometric algorithm, results will have to display a certain balance between training and testing results. While calibration error is expected to be smaller than validation error, since the model is trained on the calibration data, a large difference would imply a case of overfitting. This term is used when a model is overly-optimistic on the data used for calibration and lacks ability to generalize when applied in test dataset. This may occur due to lack of a correct calibration scheme but also due the complexity of an algorithm and the need for calibration of multiple hyperparameters (e.g. number of PLS components, number of neurons in ANN). It is possible that the dataset itself may not be enough in order to avoid overfitting, as is suggested in case 2. Lack of generalization in the case

of PLSDA and RF may also be connected to the fact that b1 and b2 datasets include two batches of freshly ground samples, but only one batch of stored samples in refrigerated conditions. This may have led to a lack of sufficient data for adulteration detection in stored samples. However, it was enough to separate clearly freshly-ground and pure samples from the rest using SVMs.

The calibration scheme employed involves (a) techniques such as cross-validation, multiple partitions, internal validation datasets, early stopping criteria, etc., and (b) some performance measure or measures according to which the hyperparameters are calculated. The most common performance criteria and their definitions for regression and classification models are described in various articles (Argyri, Panagou, Tarantilis, Polysiou, & Nychas, 2010; Ross, 1996; Sokolova & Lapalme, 2009). In regression problems those are usually some residual-based metric (e.g. MSE or RMSE, Predicted Residuals Sum of Squares - PRESS) for calibration, cross-validation and prediction as well as other criteria, such as the bias (B_f) and accuracy (A_f) factor, coefficient of correlation (R) and determination (R-squared, R^2). Other measures may be the percentage of predicted values within an acceptable error range etc. Similarly in classification problems, the percentage of correctly classified samples (overall correct classification - OCC), recall/ sensitivity, precision, specificity and F-score are usually employed.

In the available literature concerning food products, regardless of its disadvantages, LOOCV is the one of the most common CV methods, as it can be applied in small datasets and uses all samples for model development (Al-Jowder et al., 1997; Argyri et al., 2013; Feng & Sun, 2013; Kamruzzaman et al., 2013; Morsy & Sun, 2013; Zhao et al., 2014). Panagou et al. (2014) and Argyri et al. (2013) used LOOCV for PLSR

calibration, while the former chose PRESS as a calibration performance measure. da Costa Filho, (2014) also used PRESS as a PLSR calibration criterion, where a component was included only if it improved PRESS by at least 2% in a food application unrelated to meat. Alamprese et al. (2013) applied 5-fold cross-validation in PLSDA, while the model was further validated with independent meat batches. In the case of SVM models (classification or regression), a grid-search methodology coupled with CV or random partitions is usually applied, whereas for other methods such as ANNs, internal validations sets, R, R^2 and/or residual-based criteria have been deployed (Argyri et al., 2013; Papadopoulou et al., 2013).

In this study, several of the machine learning methods described in Chapter 1, as well as the calibration schemes have been deployed. Specifically, PLSDA calibration was performed based on the OCC criterion coupled with k-fold CV in case 1 and multiple random partitions in case 2 and 3 for the determination of the number of PLS components. In the case of SVM development, grid search coupled with 3-fold and 10-fold CV was employed in case 2 and 3 respectively. In case 4, calibration was based on the default schemes and parameters of the "MeatReg" platform, which included grid search for SVR and for kNN-R with k (number of neighbors) ranging from 4 to 10. PLSR and PCR components were set to 3 and RFR trees to 200. Lastly, in case 5, the training set was repartitioned and approximately 20% of the samples were used for early stopping (terminating ANN calibration process after a number of epochs with no positive change in MSE) and internal validation. Models were chosen based on R, calibration and internal validation MSE. It is also interesting to note that PCA was applied in all cases for visualization purposes, except for case 5 where it

was used for dimensionality reduction, with the PCs retained being those whose eigenvalues were equal to or greater than one.

5.3 CONCLUSIONS AND NEXT STEPS

Extended discussion on the results and the subsequent conclusions can be found in the Chapters 2-4. However, the overall results of this study allow for broader conclusions on the machine learning applications and rapid methodologies applied.

As expected, in all cases MSI and FTIR showed great potential for utilization in a fast, accurate way, whether the purpose was fraud detection or spoilage estimation. What is interesting to note is that various methods worked well for one dataset but not so well for another, as was observed for PLSDA in cases 1 and 2. On the other hand, different sensors may work well for one application but not for another. As shown in case 4, there was no sensor that displayed superior performance in all cases and that concerns only one dataset. Furthermore, other approaches (e.g. HPLC or GC-MS) have displayed in some occasions better results combined with machine learning approaches, however MSI and FTIR could more easily be used on-, in- or at-line (Nychas et al., 2008) within a Process Analytical Technology (PAT) context in the food industry.

As far as data handling techniques are concerned, preprocessing, model selection and calibration is application-, data- and sensor-specific. Adding to this the availability of several methodologies that may be deployed (Dai, Sun, Xiong, Cheng, & Zeng, 2014; Ropodi et al., 2016), it is imperative to improve on the calibration and validation schemes in order to achieve good performance and avoid overfitting. Extended testing

procedures and independent validation sets should be included where possible to explore the models' generalization ability.

On the other hand, it is established that while traditional microbiological analysis, DNA-based approaches and other methods currently used in food science for detecting spoilage or monitoring compliance to label are highly effective, their main drawback is that they are time-consuming, destructive, providing retrospective information and requiring highly-trained personnel (Papadopoulou, Panagou, Tassou, & Nychas, 2011). Therefore they are not suitable for on-line monitoring applications in the food industry and for large-scale inspections by food authorities. What remains to be seen is the possibility of combining multiple sensors, as some initial work has been presented (Alamprese et al., 2013; Biancolillo et al., 2014).

Undeniably, sensors can provide rapid, reagent-less and non-destructive techniques for the quality estimation of food products and specifically beef. MSI/ HSI and FTIR spectroscopy have been evolving steadily and combined with the rapid development of computer hardware and software, Internet of Things (IoT) applications and Big Data technologies, food scientists, industries and regulatory authorities have a great opportunity to harness these advances for assurance of food safety and quality. For this reason, the following steps should be taken:

a) Food scientists, industries and regulatory authorities should adopt a more interdisciplinary approach, working with data analysts, bioinformaticians, IT experts, etc., bringing together people with different backgrounds for the benefit of science.

b) Existing datasets and methods should be enhanced with more data, adding sample variability. Open large data repositories should be made available to scientists to improve on model development and extraction of knowledge.

c) Algorithms and mechanisms that will enable these sensors to automate the data acquisition, quality estimation and/ or decision process should be developed.

CHAPTER 6:

REFERENCES

Al-Jowder, O., Kemsley, E. K., & Wilson, R. H. (1997). Mid-infrared spectroscopy and authenticity problems in selected meats: a feasibility study. *Food Chemistry*, *59*(2), 195–201. http://doi.org/10.1016/S0308-8146(96)00289-0

Alamprese, C., Casale, M., Sinelli, N., Lanteri, S., & Casiraghi, E. (2013). Detection of minced beef adulteration with turkey meat by UV–vis, NIR and MIR spectroscopy. *LWT - Food Science and Technology*, *53*(1), 225–232. http://doi.org/10.1016/j.lwt.2013.01.027

Ammor, M. S., Argyri, A., & Nychas, G.-J. E. (2009). Rapid monitoring of the spoilage of minced beef stored under conventionally and active packaging conditions using Fourier transform infrared spectroscopy in tandem with chemometrics. *Meat Science*, *81*(3), 507–514. http://doi.org/10.1016/j.meatsci.2008.10.015

Argyri, A. A. (2010). *Quantifying meat spoilage with an array of biochemical indicators, PhD Thesis*.

Argyri, A. A., Jarvis, R. M., Wedge, D., Xu, Y., Panagou, E. Z., Goodacre, R., & Nychas, G.-J. E. (2013). A comparison of Raman and FT-IR spectroscopy for the prediction of meat spoilage. *Food Control*, *29*(2), 461–470. http://doi.org/10.1016/j.foodcont.2012.05.040

Argyri, A. A., Panagou, E. Z., & Nychas, G.-J. (2014). Monitoring Microbial Spoilage of Foods by Vibrational Spectroscopy (FT-IR & Raman). In *Novel Food Preservation and Microbial Assessment Techniques* (pp. 386–434). CRC Press. http://doi.org/doi:10.1201/b16758-17

Argyri, A. A., Panagou, E. Z., Tarantilis, P. A., Polysiou, M., & Nychas, G.-J. E. (2010). Rapid qualitative and quantitative detection of beef fillets spoilage based on Fourier transform infrared spectroscopy data and artificial neural networks. *Sensors and Actuators B: Chemical*, *145*(1), 146–154. http://doi.org/10.1016/j.snb.2009.11.052

Balabin, R. M., & Lomakina, E. I. (2011). Support vector machine regression (LS-SVM)--an alternative to artificial neural networks (ANNs) for the analysis of quantum chemistry data? *Physical Chemistry Chemical Physics : PCCP*, *13*(24), 11710–8. http://doi.org/10.1039/c1cp00051a

Ballin, N. Z. (2010). Authentication of meat and meat products. *Meat Science*, *86*(3), 577–87. http://doi.org/10.1016/j.meatsci.2010.06.001

Ballin, N. Z., & Lametsch, R. (2008). Analytical methods for authentication of fresh vs. thawed meat - A review. *Meat Science*, *80*(2), 151–8. http://doi.org/10.1016/j.meatsci.2007.12.024

Ballin, N. Z., Vogensen, F. K., & Karlsson, A. H. (2009). Species determination - Can we detect and quantify meat adulteration? *Meat Science*, *83*(2), 165–74. http://doi.org/10.1016/j.meatsci.2009.06.003

Barbin, D. F., ElMasry, G., Sun, D.-W., Allen, P., & Morsy, N. (2013). Non-destructive assessment of microbial contamination in porcine meat using NIR

hyperspectral imaging. *Innovative Food Science & Emerging Technologies*, *17*, 180–191. http://doi.org/10.1016/j.ifset.2012.11.001

Barbin, D. F., Sun, D.-W., & Su, C. (2013). NIR hyperspectral imaging as non-destructive evaluation tool for the recognition of fresh and frozen–thawed porcine longissimus dorsi muscles. *Innovative Food Science & Emerging Technologies*, *18*, 226–236. http://doi.org/10.1016/j.ifset.2012.12.011

Barker, M., & Rayens, W. (2003). Partial least squares for discrimination. *Journal of Chemometrics*, *17*(3), 166–173. http://doi.org/10.1002/cem.785

Berrueta, L. A., Alonso-Salces, R. M., & Héberger, K. (2007). Supervised pattern recognition in food analysis. *Journal of Chromatography. A*, *1158*(1–2), 196–214. http://doi.org/10.1016/j.chroma.2007.05.024

Biancolillo, A., Bucci, R., Magrì, A. L., Magrì, A. D., & Marini, F. (2014). Data-fusion for multiplatform characterization of an Italian craft beer aimed at its authentication. *Analytica Chimica Acta*, *820*, 23–31. http://doi.org/10.1016/j.aca.2014.02.024

Boulesteix, A.-L. (2004). PLS dimension reduction for classification with microarray data. *Statistical Applications in Genetics and Molecular Biology*, *3*(14), Article33. http://doi.org/10.2202/1544-6115.1075

Boulesteix, A.-L., & Strimmer, K. (2007). Partial least squares: a versatile tool for the analysis of high-dimensional genomic data. *Briefings in Bioinformatics*, *8*(1), 32–44. http://doi.org/10.1093/bib/bbl016

Boyacı, İ. H., Temiz, H. T., Uysal, R. S., Velioğlu, H. M., Yadegari, R. J., & Rishkan, M. M. (2014). A novel method for discrimination of beef and horsemeat using Raman spectroscopy. *Food Chemistry*, *148*, 37–41. http://doi.org/10.1016/j.foodchem.2013.10.006

Breiman, L. (2001). Random Forests. *Machine Learning*, *45*(1), 5–32. http://doi.org/10.1023/A:1010933404324

Brereton, R. G. (2014). A short history of chemometrics: a personal view. *Journal of Chemometrics*, *28*(10), 749–760. http://doi.org/10.1002/cem.2633

Carstensen, J. M., & Hansen, J. F. (2003). An apparatus and a method of recording an image of an object, Patent family EP1051660.

Cheng, J. H., Sun, D. W., Pu, H. Bin, Chen, X., Liu, Y., Zhang, H., & Li, J. L. (2015). Integration of classifiers analysis and hyperspectral imaging for rapid discrimination of fresh from cold-stored and frozen-thawed fish fillets. *Journal of Food Engineering*, *161*, 33–39. http://doi.org/10.1016/j.jfoodeng.2015.03.011

Coppa, M., Revello-Chion, A., Giaccone, D., Ferlay, A., Tabacco, E., & Borreani, G. (2014). Comparison of near and medium infrared spectroscopy to predict fatty acid composition on fresh and thawed milk. *Food Chemistry*, *150*, 49–57. http://doi.org/10.1016/j.foodchem.2013.10.087

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, *20*(3),

273–297. http://doi.org/10.1007/BF00994018

Cozzolino, D. (2014). An overview of the use of infrared spectroscopy and chemometrics in authenticity and traceability of cereals. *Food Research International*, *60*, 262–265. http://doi.org/10.1016/j.foodres.2013.08.034

da Costa Filho, P. A. (2014). Developing a rapid and sensitive method for determination of trans-fatty acids in edible oils using middle-infrared spectroscopy. *Food Chemistry*, *158*, 1–7. http://doi.org/10.1016/j.foodchem.2014.02.084

Dai, Q., Sun, D.-W., Xiong, Z., Cheng, J.-H., & Zeng, X.-A. (2014). Recent Advances in Data Mining Techniques and Their Applications in Hyperspectral Image Processing for the Food Industry. *Comprehensive Reviews in Food Science and Food Safety*, *13*(5), 891–905. http://doi.org/10.1111/1541-4337.12088

Damez, J.-L., & Clerjon, S. (2013). Quantifying and predicting meat and meat products quality attributes using electromagnetic waves: an overview. *Meat Science*, *95*(4), 879–96. http://doi.org/10.1016/j.meatsci.2013.04.037

Daugaard, S. B., Adler-Nissen, J., & Carstensen, J. M. (2010). New vision technology for multidimensional quality monitoring of continuous frying of meat. *Food Control*, *21*(5), 626–632. http://doi.org/10.1016/j.foodcont.2009.09.007

de Jong, S. (1993). SIMPLS: An alternative approach to partial least squares regression. *Chemometrics and Intelligent Laboratory Systems*, *18*(3), 251–263. http://doi.org/10.1016/0169-7439(93)85002-X

Dennis Jr., J. E., & Schnabel, R. B. (1996). *Numerical Methods for Unconstrained Optimization and Nonlinear Equations (Classics in Applied Mathematics, 16)*. Soc for Industrial & Applied Math.

Dimitriadou, E., Hornik, K., Leisch, F., Meyer, D., Weingessel, A., & Leisch, M. F. (2011). *Misc Functions of the Department of Statistics (e1071), TU Wien*. *R Software package, avaliable at http://cran. rproject. org/web/packages/e1071/index. html*. Retrieved from https://cran.r-project.org/web/packages/e1071/index.html

Ding, H. B., & Xu, R. J. (1999). Differentiation of beef and kangaroo meat by visible/near-infrared reflectance spectroscopy. *Journal of Food Science*, *64*(5), 814–817. http://doi.org/10.1111/j.1365-2621.1999.tb15918.x

Dissing, B. S., Papadopoulou, O. S., Tassou, C., Ersbøll, B. K., Carstensen, J. M., Panagou, E. Z., & Nychas, G.-J. (2013). Using Multispectral Imaging for Spoilage Detection of Pork Meat. *Food and Bioprocess Technology*, *6*(9), 2268–2279. http://doi.org/10.1007/s11947-012-0886-6

Domingo, E., Tirelli, A. A., Nunes, C. A., Guerreiro, M. C., & Pinto, S. M. (2014). Melamine detection in milk using vibrational spectroscopy and chemometrics analysis: A review. *Food Research International*, *60*, 131–139. http://doi.org/10.1016/j.foodres.2013.11.006

Doulgeraki, A. I., Ercolini, D., Villani, F., & Nychas, G.-J. E. (2012). Spoilage microbiota associated to the storage of raw meat in different conditions. *International Journal of Food Microbiology*, *157*(2), 130–41. http://doi.org/10.1016/j.ijfoodmicro.2012.05.020

Downey, G., & Beauchêne, D. (1997). Discrimination between fresh and frozen-then-thawed beef m. longissimus dorsi by combined visible-near infrared reflectance spectroscopy: A feasibility study. *Meat Science*, *45*(3), 353–363. http://doi.org/10.1016/S0309-1740(96)00127-1

Duchesne, C., Liu, J. J., & MacGregor, J. F. (2012). Multivariate image analysis in the process industries: A review. *Chemometrics and Intelligent Laboratory Systems*, *117*, 116–128. http://doi.org/10.1016/j.chemolab.2012.04.003

Ebrahim, H. Al, Sowoidnich, K., & Kronfeldt, H.-D. (2013). Raman spectroscopic differentiation of beef and horse meat using a 671 nm microsystem diode laser. *Applied Physics B*, *113*(2), 159–163. http://doi.org/10.1007/s00340-013-5677-x

EFSA (European Food Safety Authority). (2015). The European Union summary report on trends and sources of zoonoses, zoonotic agents and food-borne outbreaks in 2013. *EFSA Journal*, *13*(1). http://doi.org/10.2903/j.efsa.2015.3991

EFSA (European Food Safety Authority). (2017). Dioxins and PCBs. Retrieved March 13, 2017, from http://www.efsa.europa.eu/en/topics/topic/dioxins-and-pcbs

Ellis, D. I., Broadhurst, D., Clarke, S. J., & Goodacre, R. (2005). Rapid identification of closely related muscle foods by vibrational spectroscopy and machine learning. *The Analyst*, *130*(12), 1648–54. http://doi.org/10.1039/b511484e

Ellis, D. I., & Goodacre, R. (2001). Rapid and quantitative detection of the microbial spoilage of muscle foods: current status and future trends. *Trends in Food Science & Technology*, *12*(11), 414–424. http://doi.org/10.1016/S0924-2244(02)00019-5

ElMasry, G., Sun, D.-W., & Allen, P. (2011). Non-destructive determination of water-holding capacity in fresh beef by using NIR hyperspectral imaging. *Food Research International*, *44*(9), 2624–2633. http://doi.org/10.1016/j.foodres.2011.05.001

Engel, J., Gerretzen, J., Szymańska, E., Jansen, J. J., Downey, G., Blanchet, L., & Buydens, L. M. C. (2013). Breaking with trends in pre-processing? *TrAC Trends in Analytical Chemistry*, *50*, 96–106. http://doi.org/10.1016/j.trac.2013.04.015

Estelles-Lopez, L., Ropodi, A., Pavlidis, D., Fotopoulou, J., Gkousari, C., Peyrodie, A., … Mohareb, F. (2017). An automated ranking platform for machine learning regression models for meat spoilage prediction using multi-spectral imaging and metabolic profiling. *Food Research International*, *In press*. http://doi.org/10.1016/j.foodres.2017.05.013

European Commission. (2005). Commision Regulation (EC) No 2073/2005 on microbiological criteria for foodstuffs. *Official Journal of the European Union*,

OJ L 338, 1–26.

European Commission. (2015). 2014 - Food Fraud Network Activity Report. Retrieved from http://ec.europa.eu/food/safety/docs/official-controls_food-fraud_network-activity-report_2015.pdf

Everitt, B., Landau, S., Leese, M., & Stahl, D. (2011). *Cluster analysis* (5th ed.). Wiley.

Feng, Y.-Z., & Sun, D.-W. (2013). Near-infrared hyperspectral imaging in tandem with partial least squares regression and genetic algorithm for non-destructive determination and visualization of Pseudomonas loads in chicken fillets. *Talanta*, *109*, 74–83. http://doi.org/10.1016/j.talanta.2013.01.057

Fisher, R. A. (1936). The Use of Multiple Measurements in Taxonomic Problems. *Annals of Eugenics*, *7*(2), 179–188. http://doi.org/10.1111/j.1469-1809.1936.tb02137.x

Folm-Hansen, J. (1999). *On Chromatic and Geometrical Calibration (Ph.D. thesis)*. Technical University of Denmark (DTU), Department of Mathematical Modeling.

Fu, X., Kim, M. S., Chao, K., Qin, J., Lim, J., Lee, H., … Ying, Y. (2014). Detection of melamine in milk powders based on NIR hyperspectral imaging and spectral similarity analyses. *Journal of Food Engineering*, *124*, 97–104. http://doi.org/10.1016/j.jfoodeng.2013.09.023

Giaretta, N., Di Giuseppe, A. M. A., Lippert, M., Parente, A., & Di Maro, A. (2013). Myoglobin as marker in meat adulteration: a UPLC method for determining the presence of pork meat in raw beef burger. *Food Chemistry*, *141*(3), 1814–20. http://doi.org/10.1016/j.foodchem.2013.04.124

Goodacre, R. (2003). Explanatory analysis of spectroscopic data using machine learning of simple, interpretable rules. *Vibrational Spectroscopy*, *32*(1), 33–45. http://doi.org/10.1016/S0924-2031(03)00045-6

Gowen, A. A., Feng, Y., Gaston, E., & Valdramidis, V. (2015). Recent applications of hyperspectral imaging in microbiology. *Talanta*, *137*, 43–54. http://doi.org/10.1016/j.talanta.2015.01.012

Gowen, A. A., O'Donnell, C. P., Taghizadeh, M., Cullen, P. J., Frias, J. M., & Downey, G. (2008). Hyperspectral imaging combined with principal component analysis for bruise damage detection on white mushrooms (Agaricus bisporus). *Journal of Chemometrics*, *22*(3–4), 259–267. http://doi.org/10.1002/cem.1127

Haralick, R. M., Shanmugam, K., & Dinstein, I. (1973). Textural Features for Image Classification. *IEEE Transactions on Systems, Man, and Cybernetics*, *3*(6), 610–621. http://doi.org/10.1109/TSMC.1973.4309314

Haralick, R. M., & Shapiro, L. G. (1991). *Computer and Robot Vision* (1st ed.). Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning*

(2nd ed.). Springer New York.

He, H.-J., Sun, D.-W., & Wu, D. (2014). Rapid and real-time prediction of lactic acid bacteria (LAB) in farmed salmon flesh using near-infrared (NIR) hyperspectral imaging combined with chemometric analysis. *Food Research International*, *62*, 476–483. http://doi.org/10.1016/j.foodres.2014.03.064

Huang, L., Zhao, J., Chen, Q., & Zhang, Y. (2013). Rapid detection of total viable count (TVC) in pork meat by hyperspectral imaging. *Food Research International*, *54*(1), 821–828. http://doi.org/10.1016/j.foodres.2013.08.011

Huang, L., Zhao, J., Chen, Q., & Zhang, Y. (2014). Nondestructive measurement of total volatile basic nitrogen (TVB-N) in pork meat by integrating near infrared spectroscopy, computer vision and electronic nose techniques. *Food Chemistry*, *145*, 228–36. http://doi.org/10.1016/j.foodchem.2013.06.073

Jarvis, R. M., & Goodacre, R. (2005). Genetic algorithm optimization for pre-processing and variable selection of spectroscopic data. *Bioinformatics (Oxford, England)*, *21*(7), 860–8. http://doi.org/10.1093/bioinformatics/bti102

Jolliffe, I. T. (2002). *Principal Component Analysis* (2nd ed.). Springer New York. http://doi.org/10.1007/b98835

Kamruzzaman, M., Barbin, D., ElMasry, G., Sun, D.-W., & Allen, P. (2012). Potential of hyperspectral imaging and pattern recognition for categorization and authentication of red meat. *Innovative Food Science & Emerging Technologies*, *16*, 316–325. http://doi.org/10.1016/j.ifset.2012.07.007

Kamruzzaman, M., Makino, Y., & Oshita, S. (2015). Non-invasive analytical technology for the detection of contamination, adulteration, and authenticity of meat, poultry, and fish: A review. *Analytica Chimica Acta*, *853*, 19–29. http://doi.org/10.1016/j.aca.2014.08.043

Kamruzzaman, M., Maniko, Y., & Oshita, S. (2014). Non-invasive analytical technology for the detection of contamination, adulteration, and authenticity of meat, poultry, and fish: A review. *Analytica Chimica Acta*. http://doi.org/10.1016/j.aca.2014.08.043

Kamruzzaman, M., Sun, D.-W., ElMasry, G., & Allen, P. (2013). Fast detection and visualization of minced lamb meat adulteration using NIR hyperspectral imaging and multivariate image analysis. *Talanta*, *103*, 130–6. http://doi.org/10.1016/j.talanta.2012.10.020

Kohavi, R., & Provost, F. (1998). Glossary of Terms. *Machine Learning.*, *30*(2–3), 271–274. Retrieved from http://dl.acm.org/citation.cfm?id=288808.288815

Koutsoumanis, K., & Taoukis, P. S. (2005). 23 – Meat safety, refrigerated storage and transport: modeling and management. In *Improving the Safety of Fresh Meat* (pp. 503–561). http://doi.org/10.1533/9781845691028.2.503

Liaw, a, & Wiener, M. (2002). Classification and Regression by randomForest. *R News*, *2*(December), 18–22. http://doi.org/10.1177/154405910408300516

Liu, D., Zeng, X.-A., & Sun, D.-W. (2015). Recent Developments and Applications of Hyperspectral Imaging for Quality Evaluation of Agricultural Products: A Review. *Critical Reviews in Food Science and Nutrition*, *55*(12), 1744–1757. http://doi.org/10.1080/10408398.2013.777020

Liu, M., Wang, M., Wang, J., & Li, D. (2013). Comparison of random forest, support vector machine and back propagation neural network for electronic tongue data classification: Application to the recognition of orange beverage and Chinese vinegar. *Sensors and Actuators B: Chemical*, *177*, 970–980. http://doi.org/10.1016/j.snb.2012.11.071

Liu, Y., Lyon, B. G., Windham, W. R., Realini, C. E., Pringle, T. D. D., & Duckett, S. (2003). Prediction of color, texture, and sensory characteristics of beef steaks by visible and near infrared reflectance spectroscopy. A feasibility study. *Meat Science*, *65*(3), 1107–1115. http://doi.org/10.1016/S0309-1740(02)00328-5

Løkke, M. M., Seefeldt, H. F., Skov, T., & Edelenbos, M. (2013). Color and textural quality of packaged wild rocket measured by multispectral imaging. *Postharvest Biology and Technology*, *75*, 86–95. http://doi.org/10.1016/j.postharvbio.2012.06.018

Loutfi, A., Coradeschi, S., Mani, G. K., Shankar, P., & Rayappan, J. B. B. (2015). Electronic noses for food quality: A review. *Journal of Food Engineering*, *144*, 103–111. http://doi.org/10.1016/j.jfoodeng.2014.07.019

Luke, B. T. (2003). *Nature-insprired Methods in Chemometrics: Genetic Algorithms and Artificial Neural Networks*. *Data Handling in Science and Technology* (Vol. 23). Elsevier. http://doi.org/10.1016/S0922-3487(03)23001-X

Ma, J., Pu, H., Sun, D. W., Gao, W., Qu, J. H., & Ma, K. Y. (2015). Application of Vis-NIR hyperspectral imaging in classification between fresh and frozen-thawed pork Longissimus Dorsi muscles. *International Journal of Refrigeration*, *50*, 10–18. http://doi.org/10.1016/j.ijrefrig.2014.10.024

MATLAB 2012a, The MathWorks, Inc., Natick, Massachusetts, United States. (n.d.). The MathWorks, Inc., Natick, Massachusetts, United States.

Mevik, B.-H., & Wehrens, R. (2007). The pls Package: Principal Component and Partial Least Squares Regression in R. *Journal of Statistical Software; Vol 1, Issue 2 (2007)* . Retrieved from https://www.jstatsoft.org/v018/i02

Mohareb, F., Papadopoulou, O., Panagou, E., Nychas, G.-J., & Bessant, C. (2016). Ensemble-based support vector machine classifiers as an efficient tool for quality assessment of beef fillets from electronic nose data. *Anal. Methods*, *8*(18), 3711–3721. http://doi.org/10.1039/C6AY00147E

Morsy, N., & Sun, D.-W. (2013). Robust linear and non-linear models of NIR spectroscopy for detection and quantification of adulterants in fresh and frozen-thawed minced beef. *Meat Science*, *93*(2), 292–302. http://doi.org/10.1016/j.meatsci.2012.09.005

Nunes, C. A. (2014). Vibrational spectroscopy and chemometrics to assess

authenticity, adulteration and intrinsic quality parameters of edible oils and fats. *Food Research International*, *60*, 255–261. http://doi.org/10.1016/j.foodres.2013.08.041

Nunes, C. A., Alvarenga, V. O., de Souza Sant'Ana, A., de Sousa Santos, J., & Granato, D. (2015). The Use of Statistical Software in Food Science and Technology: Advantages, Limitations and Misuses. *Food Research International*. http://doi.org/10.1016/j.foodres.2015.06.011

Nychas, G.-J. E., & Skandamis, P. N. (2005). 22 – Fresh meat spoilage and modified atmosphere packaging (MAP). In *Improving the Safety of Fresh Meat* (pp. 461–502). http://doi.org/10.1533/9781845691028.2.461

Nychas, G.-J. E., Skandamis, P. N., Tassou, C. C., & Koutsoumanis, K. P. (2008). Meat spoilage during distribution. *Meat Science*, *78*, 77–89. http://doi.org/10.1016/j.meatsci.2007.06.020

Otto, M. (2007). *Chemometrics* (2nd ed.). WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim.

Pan, L., Zhang, W., Zhu, N., Mao, S., & Tu, K. (2014). Early detection and classification of pathogenic fungal disease in post-harvest strawberry fruit by electronic nose and gas chromatography–mass spectrometry. *Food Research International*, *62*, 162–168. http://doi.org/10.1016/j.foodres.2014.02.020

Panagou, E. Z., Mohareb, F. R., Argyri, A. A., Bessant, C. M., & Nychas, G. J. E. (2011). A comparison of artificial neural networks and partial least squares modelling for the rapid detection of the microbial spoilage of beef fillets based on Fourier transform infrared spectral fingerprints. *Food Microbiology*, *28*(4), 782–790. http://doi.org/10.1016/j.fm.2010.05.014

Panagou, E. Z., Papadopoulou, O., Carstensen, J. M., & Nychas, G.-J. E. (2014). Potential of multispectral imaging technology for rapid and non-destructive determination of the microbiological quality of beef filets during aerobic storage. *International Journal of Food Microbiology*, *174*, 1–11. http://doi.org/10.1016/j.ijfoodmicro.2013.12.026

Panagou, E. Z., Sahgal, N., Magan, N., & Nychas, G.-J. E. (2008). Table olives volatile fingerprints: Potential of an electronic nose for quality discrimination. *Sensors and Actuators B: Chemical*, *134*(2), 902–907. http://doi.org/10.1016/j.snb.2008.06.038

Papadopoulou, O., Panagou, E. Z., Tassou, C. C., & Nychas, G.-J. E. (2011). Contribution of Fourier transform infrared (FTIR) spectroscopy data on the quantitative determination of minced pork meat spoilage. *Food Research International*, *44*(10), 3264–3271. http://doi.org/10.1016/j.foodres.2011.09.012

Papadopoulou, O. S., Panagou, E. Z., Mohareb, F. R., & Nychas, G.-J. E. (2013). Sensory and microbiological quality assessment of beef fillets using a portable electronic nose in tandem with support vector machine analysis. *Food Research International*, *50*(1), 241–249. http://doi.org/10.1016/j.foodres.2012.10.020

Prevolnik, M., Čandek-Potokar, M., & Škorjanc, D. (2010). Predicting pork water-holding capacity with NIR spectroscopy in relation to different reference methods. *Journal of Food Engineering*, *98*(3), 347–352. http://doi.org/10.1016/j.jfoodeng.2009.11.022

Pu, H., Sun, D. W., Ma, J., & Cheng, J. H. (2015). Classification of fresh and frozen-thawed pork muscles using visible and near infrared hyperspectral imaging and textural analysis. *Meat Science*, *99*, 81–88. http://doi.org/10.1016/j.meatsci.2014.09.001

Qin, J., Chao, K., Kim, M. S., Lu, R., & Burks, T. F. (2013). Hyperspectral and multispectral imaging for evaluating food safety and quality. *Journal of Food Engineering*, *118*(2), 157–171. http://doi.org/10.1016/j.jfoodeng.2013.04.001

Rohman, A., Sismindari, Erwanto, Y., & Che Man, Y. B. (2011). Analysis of pork adulteration in beef meatball using Fourier transform infrared (FTIR) spectroscopy. *Meat Science*, *88*(1), 91–5. http://doi.org/10.1016/j.meatsci.2010.12.007

Ropodi, A. I., Panagou, E. Z., & Nychas, G.-J. E. (2016). Data mining derived from Food analyses using non-invasive/non-destructive analytical techniques; Determination of Food authenticity, quality & safety in tandem with Computer Science Disciplines. *Trends in Food Science & Technology*, *50*, 11–25. http://doi.org/10.1016/j.tifs.2016.01.011

Ropodi, A. I., Panagou, E. Z., & Nychas, G.-J. E. (2017). Multispectral imaging (MSI): A promising method for the detection of minced beef adulteration with horsemeat. *Food Control*, *73*, 57–63. http://doi.org/10.1016/j.foodcont.2016.05.048

Ropodi, A. I., Pavlidis, D. E., Mohareb, F., Panagou, E. Z., & Nychas, G.-J. E. (2015). Multispectral Image Analysis approach to detect adulteration of beef and pork in raw meats. *Food Research International*, *67*, 12–18. http://doi.org/10.1016/j.foodres.2014.10.032

Ross, T. (1996). Indices for performance evaluation of predictive models in food microbiology. *Journal of Applied Bacteriology*, *81*(5), 501–508. http://doi.org/10.1111/j.1365-2672.1996.tb03539.x

RStudio, v. 0.97.551, RStudio, Inc., Boston, Massachusetts, United States. (n.d.).

Sharifzadeh, S., Clemmensen, L. H., Borggaard, C., Støier, S., & Ersbøll, B. K. (2014). Supervised feature selection for linear and non-linear regression of L*a*b* color from multispectral images of meat. *Engineering Applications of Artificial Intelligence*, *27*, 211–227. http://doi.org/10.1016/j.engappai.2013.09.004

Soares, S., Amaral, J. S., Mafra, I., & Oliveira, M. B. P. P. (2010). Quantitative detection of poultry meat adulteration with pork by a duplex PCR assay. *Meat Science*, *85*(3), 531–6. http://doi.org/10.1016/j.meatsci.2010.03.001

Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures

for classification tasks. *Information Processing & Management*, *45*(4), 427–437. http://doi.org/10.1016/j.ipm.2009.03.002

Sumathi, S., & Surekha, P. (2010). *Computational Intelligence Paradigms*. *Computational Intelligence Paradigms*. CRC Press. http://doi.org/10.1201/9781439809037

Sun, D.-W., Reddy Gangidi, R., & Proctor, A. (2009). *Infrared Spectroscopy for Food Quality Analysis and Control*. *Infrared Spectroscopy for Food Quality Analysis and Control*. Elsevier. http://doi.org/10.1016/B978-0-12-374136-3.00008-0

Sutton, R. S., & Barto, A. G. (1998). *Introduction to Reinforcement Learning* (1st ed.). Cambridge, MA, USA: MIT Press.

Teena, M., Manickavasagan, a., Mothershaw, a., El Hadi, S., & Jayas, D. S. (2013). Potential of Machine Vision Techniques for Detecting Fecal and Microbial Contamination of Food Products: A Review. *Food and Bioprocess Technology*, *6*(7), 1621–1634. http://doi.org/10.1007/s11947-013-1079-7

Tian, X., Wang, J., & Cui, S. (2013). Analysis of pork adulteration in minced mutton using electronic nose of metal oxide sensors. *Journal of Food Engineering*, *119*(4), 744–749. http://doi.org/10.1016/j.jfoodeng.2013.07.004

Tsakanikas, P., Pavlidis, D., & Nychas, G.-J. (2015). High Throughput Multispectral Image Processing with Applications in Food Science. *PloS One*, *10*(10), e0140122. http://doi.org/10.1371/journal.pone.0140122

Tsakanikas, P., Pavlidis, D., Panagou, E., & Nychas, G. J. (2016). Exploiting multispectral imaging for non-invasive contamination assessment and mapping of meat samples. *Talanta*, *161*, 606–614. http://doi.org/10.1016/j.talanta.2016.09.019

Velusamy, V., Arshak, K., Korostynska, O., Oliwa, K., & Adley, C. (2010). An overview of foodborne pathogen detection: in the perspective of biosensors. *Biotechnology Advances*, *28*(2), 232–54. http://doi.org/10.1016/j.biotechadv.2009.12.004

Videometer A/S (official website). (n.d.). Retrieved August 29, 2014, from http://www.videometer.com

Westad, F., & Marini, F. (2015). Validation of chemometric models - a tutorial. *Analytica Chimica Acta*, *893*, 14–24. http://doi.org/10.1016/j.aca.2015.06.056

Wilson, A., Oberle, C., & Oberle, D. (2013). Detection of Off-Flavor in Catfish Using a Conducting Polymer Electronic-Nose Technology. *Sensors*, *13*(12), 15968–15984. http://doi.org/10.3390/s131215968

Wold, S., Sjöström, M., & Eriksson, L. (2001). PLS-regression: a basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems*, *58*(2), 109–130. http://doi.org/10.1016/S0169-7439(01)00155-1

Wu, D., Shi, H., He, Y., Yu, X., & Bao, Y. (2013). Potential of hyperspectral imaging and multivariate analysis for rapid and non-invasive detection of gelatin

adulteration in prawn. *Journal of Food Engineering*, *119*(3), 680–686. http://doi.org/10.1016/j.jfoodeng.2013.06.039

Wu, D., & Sun, D.-W. (2013a). Advanced applications of hyperspectral imaging technology for food quality and safety analysis and assessment: A review — Part I: Fundamentals. *Innovative Food Science & Emerging Technologies*, *19*, 1–14. http://doi.org/10.1016/j.ifset.2013.04.014

Wu, D., & Sun, D.-W. (2013b). Advanced applications of hyperspectral imaging technology for food quality and safety analysis and assessment: A review — Part II: Applications. *Innovative Food Science & Emerging Technologies*, *19*, 15–28. http://doi.org/10.1016/j.ifset.2013.04.016

Xia, J., Mandal, R., Sinelnikov, I. V, Broadhurst, D., & Wishart, D. S. (2012). MetaboAnalyst 2.0: a comprehensive server for metabolomic data analysis. *Nucleic Acids Research*, *40*(Web Server issue), W127-33. http://doi.org/10.1093/nar/gks374

Xia, J., Sinelnikov, I. V., Han, B., & Wishart, D. S. (2015). MetaboAnalyst 3.0--making metabolomics more meaningful. *Nucleic Acids Research*, *43*(W1), W251-257. http://doi.org/10.1093/nar/gkv380

Xiaobo, Z., Jiewen, Z., Povey, M. J. W., Holmes, M., & Hanpin, M. (2010). Variables selection methods in near-infrared spectroscopy. *Analytica Chimica Acta*, *667*(1–2), 14–32. http://doi.org/10.1016/j.aca.2010.03.048

Xiong, Z., Sun, D.-W., Zeng, X.-A., & Xie, A. (2014). Recent developments of hyperspectral imaging systems and their applications in detecting quality attributes of red meats: A review. *Journal of Food Engineering*, *132*, 1–13. http://doi.org/10.1016/j.jfoodeng.2014.02.004

Xu, Q.-S., & Liang, Y.-Z. (2001). Monte Carlo cross validation. *Chemometrics and Intelligent Laboratory Systems*, *56*(1), 1–11.

Zając, A., Hanuza, J., & Dymińska, L. (2014). Raman spectroscopy in determination of horse meat content in the mixture with other meats. *Food Chemistry*, *156*, 333–8. http://doi.org/10.1016/j.foodchem.2014.02.002

Zakaria, A., Md Shakaff, A. Y., Masnan, M. J., Saad, F. S. A., Adom, A. H., Ahmad, M. N., … Kamarudin, L. M. (2012). Improved maturity and ripeness classifications of Magnifera Indica cv. harumanis mangoes through sensor fusion of an electronic nose and acoustic sensor. *Sensors (Switzerland)*, *12*(5), 6023–6048. http://doi.org/10.3390/s120506023

Zhang, H., Wang, J., Ye, S., & Chang, M. (2012). Application of Electronic Nose and Statistical Analysis to Predict Quality Indices of Peach. *Food and Bioprocess Technology*, *5*, 65–72. http://doi.org/10.1007/s11947-009-0295-7

Zhao, M., Downey, G., & O'Donnell, C. P. (2014). Detection of adulteration in fresh and frozen beefburger products by beef offal using mid-infrared ATR spectroscopy and multivariate data analysis. *Meat Science*, *96*(2 Pt A), 1003–11. http://doi.org/10.1016/j.meatsci.2013.10.015

APPENDIX I

Published studies and supplementary material for chapters 2-4.

Food Research International 67 (2015) 12–18

Contents lists available at ScienceDirect

Food Research International

journal homepage: www.elsevier.com/locate/foodres

Multispectral image analysis approach to detect adulteration of beef and pork in raw meats

A.I. Ropodi [a,1], D.E. Pavlidis [a,1], F. Mohareb [b], E.Z. Panagou [a], G.-J.E. Nychas [a,*]

[a] Agricultural University of Athens, School of Food, Biotechnology & Development, Dept Food Science & Human Nutrition, Lab of Microbiology & Biotechnology of Foods, Iera Odos 75, Athens 11855 Greece

[b] The Bioinformatics Group, Biomedical Engineering Centre, Cranfield University, College Road, Bedford, MK43 0AL, UK

ARTICLE INFO

Article history:
Received 2 September 2014
Accepted 31 October 2014
Available online 8 November 2014

Keywords:
Meat adulteration
Multispectral image analysis
Discriminant Analysis
Minced beef/pork
External validation

ABSTRACT

The aim of this study was to investigate the potential of multispectral imaging supported by multivariate data analysis for the detection of minced beef fraudulently substituted with pork and vice versa. Multispectral images in 18 different wavelengths of 220 meat samples in total from four independent experiments (55 samples per experiment) were acquired for this work. The appropriate amount of beef and pork-minced meat was mixed in order to achieve nine different proportions of adulteration and two categories of pure pork and beef. After an image processing step, data from the first three experiments were used for partial least squares-discriminant analysis (PLS-DA) and linear discriminant analysis (LDA) so as to discriminate among all adulteration classes, as well as among adulterated, pure beef and pure pork samples. Results showed very good discrimination between pure and adulterated samples, for PLS-DA and LDA, yielding 98.48% overall correct classification. Additionally, 98.48% and 96.97% of the samples were classified within a ±10% category of adulteration for LDA and PLS-DA respectively. Lastly, the models were further validated using the data of the fourth experiment for independent testing, where all pure and adulterated samples were classified correctly in the case of PLS-DA, while LDA was proved to be less accurate.

1. Introduction

Nowadays European consumers are increasingly demanding information and reassurance not only on the origin but also on the content of their food. Protecting consumer rights and preventing fraudulent or deceptive practices such as food adulteration are important and challenging issues facing the European food industry, as manufacturers are required to provide and confirm the authenticity and point of origin of food products and their components. Furthermore, adulterants can be revealed with great difficulty in the context of methods commonly applied in laboratories. Several standard analytical techniques, such as immunological and enzymatic techniques, DNA and protein based assays and triacylglycerol analysis have been applied to authenticate food commodities (Ballin, 2010; Soares, Amaral, Mafra, & Oliveira, 2010). However, while these methods are usually capable of detecting low levels of adulteration (Ballin, 2010) they are expensive, invasive, sophisticated, laborious, and technically demanding (Ding & Xu, 1999).

Indeed meat adulteration is a growing challenge for EU meat manufacturers since most adulterants are unknown and unpredictable (e.g., horse meat). For this reason attention should also be paid to the safety and authenticity of meat and meat products, as they can be attractive targets for adulteration in many ways, including substitution or partial substitution of high commercial value meat with cheaper, such as pork or offal or by adding proteins from several origins (Kamruzzaman, Sun, ElMasry, & Allen, 2013; Tian, Wang, & Cui, 2013). With minced meat being the basic ingredient for burgers, adulteration of beef minced meat is a current problem, involving economic, quality, safety and socio-religious issues (Alamprese, Casale, Sinelli, Lanteri, & Casiraghi, 2013). Thus, the meat industry urgently needs methods that will screen non-targeted food samples for contaminants in order to provide proof of origin and prevent deliberate or accidental undeclared admixture to food samples, in a rapid and cost efficient way.

Hyperspectral and multispectral imaging spectroscopy have been used as rapid techniques to monitor quality attributes of food products (Wu & Sun, 2013). The former has been used for the rapid detection of total viable counts in pork (Barbin, Sun, & Su, 2013; Huang, Zhao, Chen, & Zhang, 2013) and of the water-holding capacity of fresh beef (ElMasry, Sun, & Allen, 2011) and pork (Prevolnik, Čandek-Potokar, & Škorjanc, 2010). Meanwhile, multispectral image analysis has high potency for the evaluation of food quality systems during handling, processing and storage (Løkke et al., 2013), and it has been previously used for the conversion of meat color in L*, a*, b* values (Sharifzadeh, Clemmensen, Borggaard, Støier, & Ersbøll, 2014) and for quality

* Corresponding author.Tel.: +30 210 5294938.
E-mail address: gjn@aua.gr (G.-J.E. Nychas).
[1] These authors contributed equally to this study.

http://dx.doi.org/10.1016/j.foodres.2014.10.032
0963-9969/

assessment of beef (Dissing et al., 2013; Panagou, Papadopoulou, Carstensen, & Nychas, 2014). Despite the fact that hyperspectral imaging has been used for the detection of minced lamb adulteration (Kamruzzaman et al., 2013) and gelatine adulteration in prawn (Wu, Shi, He, Yu, & Bao, 2013), to the best of our knowledge the use of multispectral image analysis for meat adulteration, especially in the case of minced beef with pork, has never been previously explored.

Surface chemistry, such as multispectral image spectroscopy, is introduced in the present study as a new approach in tandem with advanced statistical approaches, for the discrimination of raw minced beef meat, which has been fraudulently substituted or combined with raw minced pork. Thus, the objective of this study was to (a) evaluate the potential use of multispectral imaging to discriminate pork from beef, (b) identify if possible, the lowest percentage of minced pork adulteration in minced beef that can be safely detected and (c) establish a rapid and non-invasive technique that can potentially give results in a few minutes.

2. Materials and methods

2.1. Sample preparation

Different levels of adulteration of minced beef and pork were prepared as follows; fresh beef and pork fillets *Longissimus* muscle of normal pH (5.6–5.8) were purchased from central butcher shops in Athens and transported under refrigeration to the laboratory within 30 min. The fillets were cut into smaller pieces and grinded separately one at a time, using a domestic meat-mincing machine. The machine parts coming in contact with the meat were initially disinfected by washing with detergent and hot water, and rinsing with pure ethanol. To achieve different levels of adulteration, ranging from 10 to 90% with a 10% increment, the appropriate amount of each type of meat was used and mixed in conditions that simulate industrial processing. From each level of adulteration, five different portions of ca. 75–80 g were placed in Petri dishes, and snapshots were taken using VideometerLab vision system (Videometer A/S, Hørsholm, Denmark). For every level of adulteration (nine categories of mixed meat and two categories of pure pork and beef), each Petri dish was considered as a replicate in the experiment (5×11 samples in total per experiment).

All experimental procedure took place aseptically and was repeated four times. One hundred and sixty five (165) samples from three independent experiments (i.e., 55 samples per batch) were used to develop the model, and 55 samples from the fourth experiment were employed for the purpose of external validation. It should be noted that 220 samples from different batches were analyzed in total. From this point on, meat samples from the previously mentioned independent experiments will be referred to as samples from batches 1, 2, 3 and 4.

2.2. Image acquisition and analysis

Images from every sample were captured using VideometerLab, a system which acquires multispectral images in 18—non-uniformly distributed—different wavelengths ranging from 405 to 970 nm. Analytically, the wavelengths are 405, 430, 450, 470, 505, 565, 590, 630, 645, 660, 850, 870, 890, 910, 920, 940, 950 and 970 nm. The system has been developed by the Technical University of Denmark and commercialized by "Videometer A/S" (Carstensen & Hansen, 2003; http://www.videometer.com). A detailed description of the instrument has been reported elsewhere (Panagou et al., 2014). The advantage of this instrument is that it not only uses the information of visible and short-NIR spectral regions, but moreover uses the spatial information of each pixel.

The system was first calibrated radiometrically and geometrically using well-defined standard targets, followed by a light setup based on the type of object to be recorded (Folm-Hansen, 1999) called "autolight". In autolight, it is always the brightest sections in the image that dictate the final result. Petri dishes (75–80 g meat portions) were placed inside an Ulbricht sphere in which the camera is top-mounted. For every random dish in each level of adulteration, a different autolight procedure was employed.

The resulting image includes redundant information, such as the Petri dish and its surrounding background, as well as the fat and connective tissue of the meat. For this reason an image-processing step is needed that will result in an image mask where only meat tissue is included. This step, which includes transformation and segmentation procedures, was implemented using the respective routines of the VideometerLab software (version 2.12.39) that controls the operation of the instrument. Canonical discriminant analysis (CDA) was employed as a two-step supervised transformation building method to divide the images into regions of interest (Daugaard, Adler-Nissen, & Carstensen, 2010). Following this transformation, the separation was distinct, and a simple threshold was enough to separate meat from non-meat pixels. The result of this processing is a segmented image for each meat sample with the isolated part of the meat tissue as the main region of interest (ROI) to be used for the extraction of spectral data that were further employed in statistical analysis. The procedure is graphically presented in Fig. 1.

2.3. Data analysis

For each image, the mean reflectance spectrum was calculated by averaging the intensity of pixels within the ROI at each wavelength. Furthermore, the standard deviation of the pixels' intensity per wavelength was extracted. The resulting data consisted of 18 mean values and 18 standard deviations of the reflectance, as it was recorded by the camera for the pixels that were included in each image's ROI, and were further analyzed with various classification methods.

Two methods, partial least squares discriminant analysis (PLS-DA) (Barker & Rayens, 2003; De Jong, 1993) and linear discriminant analysis (LDA) (Fisher, 1936), were performed in order to discriminate among all adulteration classes (11 in total), as well as among adulterated, pure beef and pure pork samples.

As both methods are supervised, the data were partitioned in two sets: the training set used for model calibration and the test set used for validation. A 60–40% stratified partition was applied on the first three batches, meaning 60% of the dataset was chosen in a random way for calibration (99 samples out of 165) as long as all classes and batches were included and equally represented. The fourth batch was also reserved for independent model validation.

Model performance was measured in terms of recall (sensitivity) and precision, as well as overall correct classification (OCC) (Sokolova & Lapalme, 2009). Especially in the case of PLS-DA, the optimum number of PLS components was estimated using stratified three-fold cross-validation.

Lastly, hierarchical cluster analysis—HCA (Everitt, Landau, Leese, & Stahl, 2011) was performed per batch as an unsupervised technique to explore the relationship between variables and adulteration classes, using Euclidean Distance and Ward's minimum variance agglomeration method. Then, principal component analysis—PCA (Jolliffe, 2002) was performed per batch, as well as with all three batches so as to visualize whether there were significant differences among samples from different batches, as well as among different classes.

The partitioning algorithms of the dataset and the LDA algorithm were implemented in MATLAB, 2012a (The MathWorks, Inc., Natick, Massachusetts, United States), while HCA, PCA and PLS-DA were implemented in R v.3.0.2 (RStudio, n.d.), using the "plsgenomics" package (Boulesteix, 2004; Boulesteix & Strimmer, 2007; De Jong, 1993). Lastly, a heatmap was created using the MetaboAnalyst 2.0 software (Xia, Mandal, Sinelnikov, Broadhurst, & Wishart, 2012).

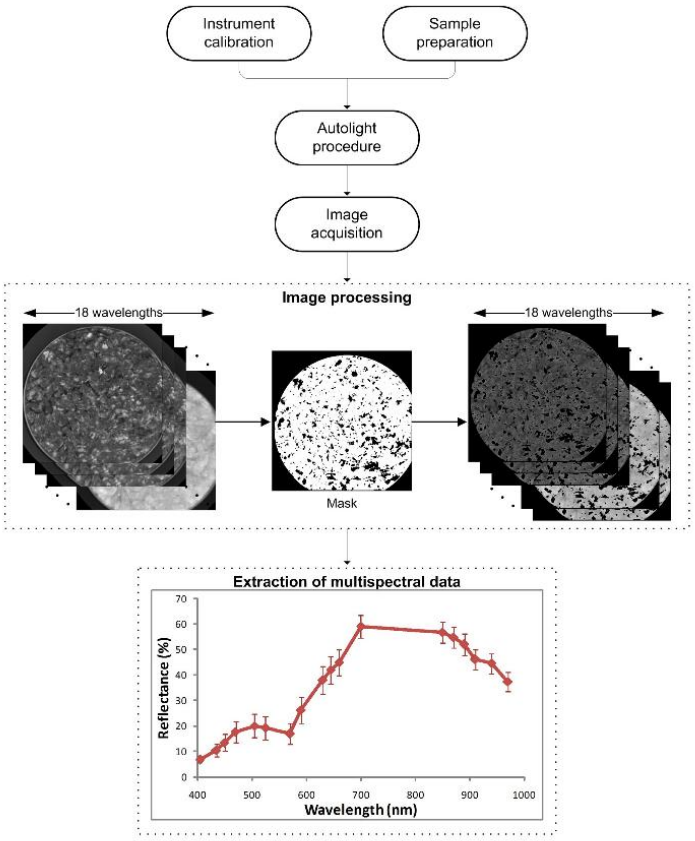


Fig. 1. Process of multispectral image acquisition and extraction of data.

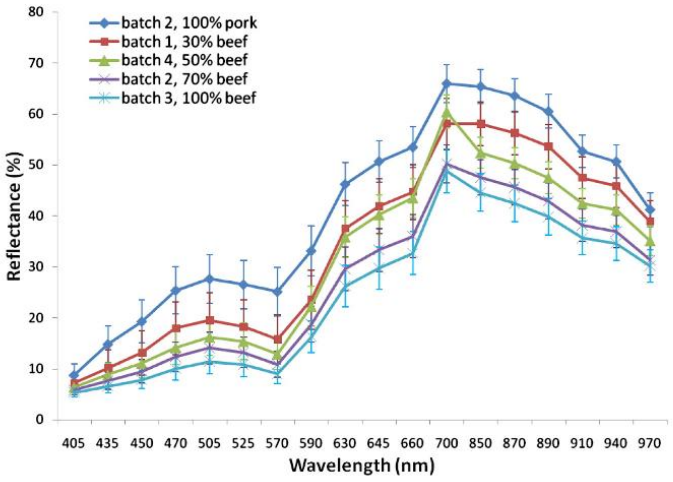


Fig. 2. Selected spectra of the examined samples corresponding to different ratios of adulteration.

3. Results

3.1. Multispectral image data

Selected spectra of minced beef in various adulteration levels are presented in Fig. 2. It is characteristic that the reflectance of the sample increased in most wavelengths of the spectrum with increasing proportions of pork meat in the mixture, providing strong evidence for the effectiveness of multispectral imaging in discriminating meat adulteration.

Initial analysis with HCA and PCA showed that the use of different batches is critical, as we must take into account not only the variability within a batch (different samples of the same class), but also the variability among batches. This was evident either when examining the results of HCA per batch (see S-Fig. 1 and 3 in supplement file) or the heatmap for all three batches (Fig. 3). When all three batches were compared, samples belonging in the same category showed great differences among batches. Nevertheless, a potential for good discrimination could be concluded after PCA analysis. Different principal components (PCs) contributed differently in terms of the variability explained (results not shown), but the first two PC scores are presented in Fig. 4. In all cases, pure pork and pure beef were found on the far left and right of the plot respectively, and the discrimination between classes was more evident. Only adjacent categories sometimes overlap. On the other hand, classes of mixed samples seem to be represented in a different way for each batch. Furthermore, when all three batches were included (Fig. 4d), a definite trend to the right of the plot was seen as the percentage of beef in the mix increases, but the discrimination among classes was less evident.

Based on the above observations, a 60–40% stratified partition was applied in the first three batches of meat corresponding to three samples per batch per category for calibration and two for validation (i.e., 99 and 66 samples in total, respectively). The samples of the fourth batch (55 in total) were reserved for independent/external model validation.

3.2. Validation

The analysis of the data acquired for each class of adulteration is shown in Table 1, where the results for per-class recall and precision

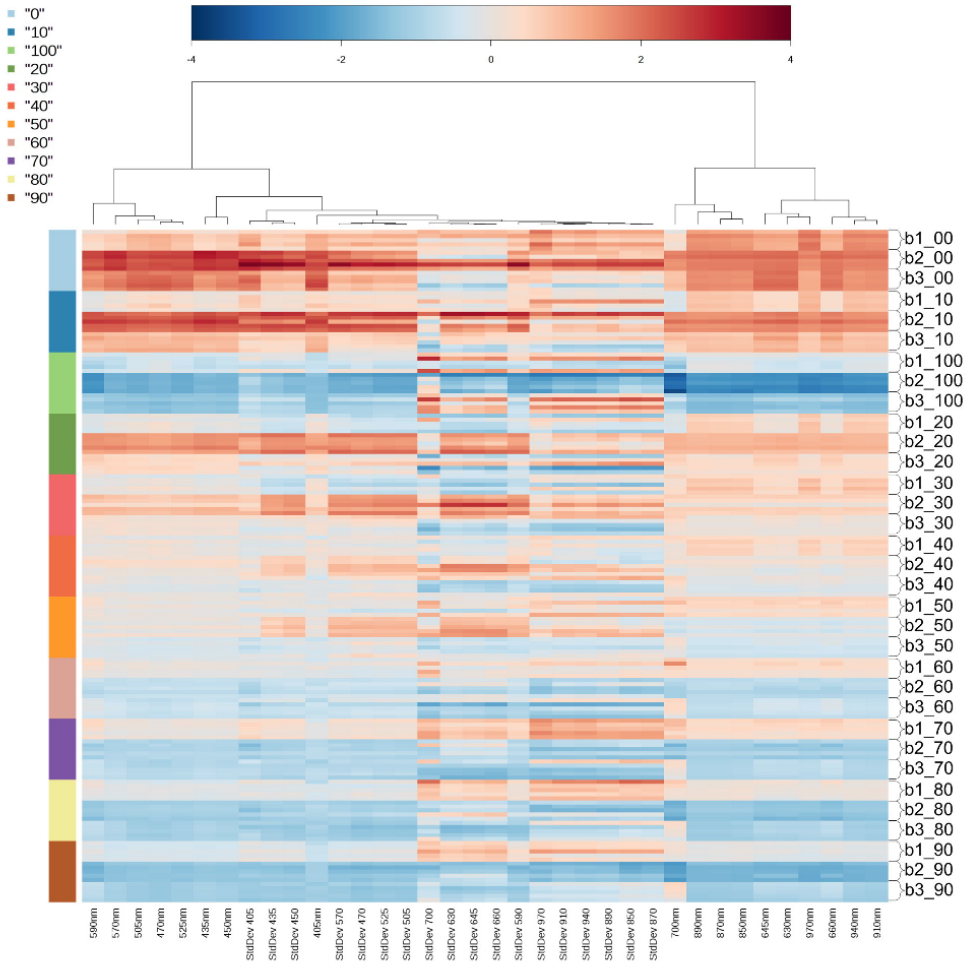


Fig. 3. Heatmap of all samples from batches 1–3; different colors on the left side correspond to different ratios of adulteration. Samples named with "00" correspond to pure pork, all other categories correspond to the percentage of beef in the mix, and consequently "100" refers to pure beef samples, whereas "b1", "b2", "b3" correspond to the number of batch.

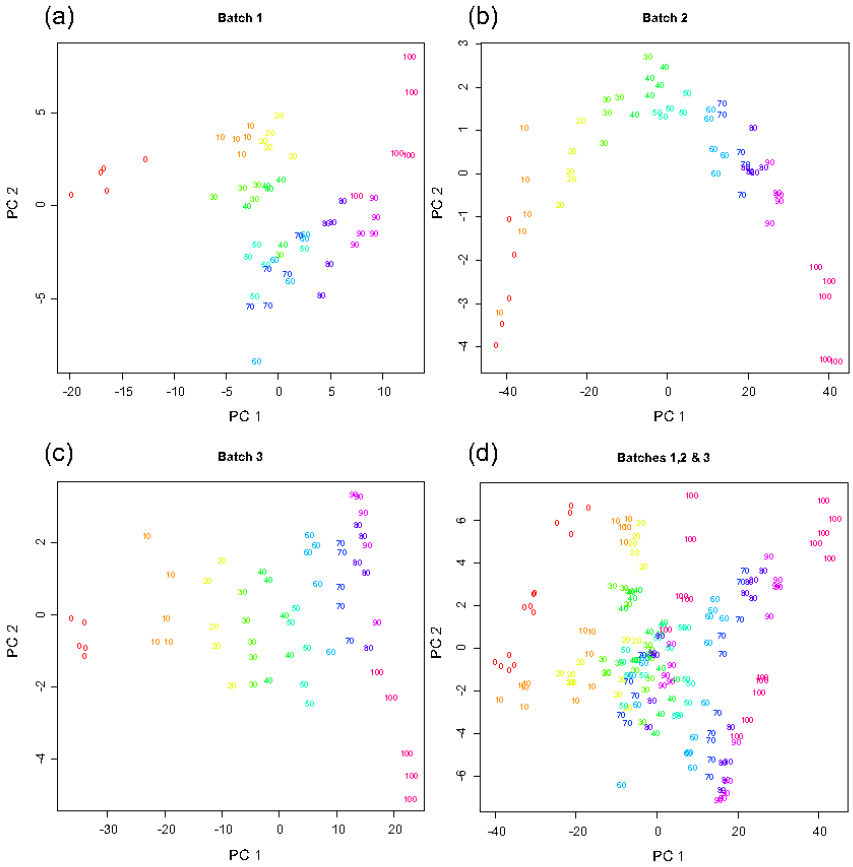


Fig. 4. Principal component analysis scores for batches 1(a), 2(b), 3(c) and all batches (d). Samples named with "0" correspond to pure pork, all other categories correspond to the percentage of beef in the mix and consequently "100" refers to pure beef samples.

are presented. The overall correct classification, mean per-class recall and precision were 83.33%, 83.33% and 84.46% respectively; the classification error for 98.48% of the samples was at most 10%, for LDA. The classification of each sample of the validation set is presented along with the per-class precision and recall in S-Table 1 in the supplement file. Very good results were also acquired when classification among pure pork, adulterated and pure beef was tested (Table 2), and where the overall correct classification, mean per-class recall and precision

Table 1
LDA vs. PLS-DA for both validation set and external validation batch with 11 classes ranging from pure pork (0%) to beef (100%).

Sample	Validation set						External validation set					
	LDA			PLS-DA			LDA			PLS-DA		
	Recall	Precision	±10% error	Recall	Precision	±10% error	Recall	Precision	±10% error	Recall	Precision	±10% error
Is 0%	83.33%	100.00%	100.00%	66.67%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%
Is 10%	100.00%	85.71%	100.00%	100.00%	75.00%	100.00%	20.00%	16.67%	100.00%	0.00%	0.00%	20.00%
Is 20%	83.33%	100.00%	100.00%	83.33%	100.00%	100.00%	20.00%	7.14%	20.00%	0.00%	0.00%	20.00%
Is 30%	83.33%	71.43%	83.33%	50.00%	75.00%	100.00%	0.00%	0.00%	100.00%	0.00%	0.00%	40.00%
Is 40%	66.67%	80.00%	100.00%	66.67%	50.00%	100.00%	20.00%	7.69%	80.00%	20.00%	14.29%	20.00%
Is 50%	83.33%	71.43%	100.00%	100.00%	75.00%	100.00%	0.00%	0.00%	0.00%	0.00%	0.00%	60.00%
Is 60%	66.67%	80.00%	100.00%	33.33%	50.00%	83.33%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Is 70%	66.67%	57.14%	100.00%	83.33%	50.00%	100.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Is 80%	83.33%	83.33%	100.00%	16.67%	100.00%	83.33%	0.00%	0.00%	0.00%	0.00%	0.00%	20.00%
Is 90%	100.00%	100.00%	100.00%	83.33%	71.43%	100.00%	0.00%	0.00%	0.00%	100.00%	33.33%	100.00%
Is 100%	100.00%	100.00%	100.00%	100.00%	85.71%	100.00%	0.00%	0.00%	0.00%	100.00%	55.56%	100.00%
Mean per-class	83.33%	84.46%	98.48%	71.21%	75.65%	96.97%	14.55%	11.95%	36.36%	29.09%	18.47%	43.64%

Table 2
LDA and PLS-DA (12 PLS components) for both validation set and external validation batch with 3 classes (pork—adulterated—beef).

	LDA					PLS-DA			
	Validation set								
	Classified as					Classified as			
	Pork	Adulterated	Beef	Recall		Pork	Adulterated	Beef	Recall
Is pork	5	1	0	83.33%	Is pork	5	1	0	83.33%
Is adulterated	0	54	0	100.00%	Is adulterated	0	54	0	100.00%
Is beef	0	0	6	100.00%	Is beef	0	0	6	100.00%
Precision	100.00%	98.18%	100.00%		Precision	100.00%	98.18%	100.00%	
External validation batch									
Is pork	4	0	1	80.00%	Is pork	5	0	0	100.00%
Is adulterated	0	35	10	77.78%	Is adulterated	0	45	0	100.00%
Is beef	0	0	5	100.00%	Is beef	0	0	5	100.00%
Precision	100.00%	100.00%	31.25%		Precision	100.00%	100.00%	100.00%	

was over 94% (mean recall: 94.44%, precision: 99.39% and overall correct classification: 98.48%). In fact, only one out of 66 samples was misclassified.

Application of PLS-DA for the three-class case, yielded the same results (98.48% correct classification) using 12 PLS components after cross-validation. PLS-DA for all categories gave similar results to the LDA. The calibration of the model was done with cross-validation, as described previously, based on overall correct classification criterion, using 20 PLS components (Table 1, and S-Table 2 in the supplement file). It would be interesting to note that, although the overall correct classification dropped considerably, this method classified 96.97% of the samples within the ±10% error of prediction.

In the case of external validation, PLS-DA performed well, classifying all samples correctly in the three-class problem of pure vs. adulterated samples (Table 2). In the case of 11 categories, all pure samples were classified correctly, but the prediction of adulteration levels in the samples was less accurate (Table 1). However, only 4 out of 45 adulterated samples were classified as pure (see S-Table 4 in the supplement file).

The LDA model was less successful in predicting pure beef samples (Table 1), as well as the adulteration level for the case of 11 categories (S-Table 3 in the supplement file), whereas in the case of pure vs. adulterated samples, the results were better yielding an overall correct classification of 80% and mean per-class recall and precision of 85.93% and 77.08%, respectively (Table 2).

4. Discussion

This work investigated whether an emerging, non-destructive, fast and low cost technique (in the long term) based on imaging technology combined with multivariate data analysis can be a promising tool for the detection of minced meat adulteration. Since various standard analytical methods are now available to identify meat's adulteration at a very low level (Ballin, 2010; Ballin, Vogensen, & Karlsson, 2009), the proposed method showed great potential. In comparison with these reported methods (Ballin et al., 2009) which are time-consuming, expensive, use harmful reagents, need expert laboratory staff and are strongly dependent on rigorously following a standardized protocol to obtain accuracy, the multispectral imaging is non-destructive, requires only basic training in a user-friendly software, a few minutes for image acquisition and processing and no cost at all—excluding initial instrument and software purchase.

Compared to other similar published studies on rapid techniques, in most cases, the main objective is the differentiation among different types of meat, e.g. beef vs. kangaroo (Ding & Xu, 1999), pork vs. beef vs. lamb (Kamruzzaman, Barbin, ElMasry, Sun, & Allen, 2012), chicken vs. turkey (Ellis, Broadhurst, Clarke, & Goodacre, 2005), and beef vs. horsemeat (Boyacı et al., 2014; Ebrahim, Sowoidnich, & Kronfeldt, 2013). Few studies has been published on the adulteration of poultry with pork (Soares et al., 2010), pork in minced mutton (Tian et al., 2013), pork in beef meatball (Rohman, Sismindari, Erwanto, & Che Man, 2011), beef offal in fresh and frozen beefburger (Zhao, Downey, & O'Donnell, 2014), pork meat in raw beef burger (Giaretta, Di Giuseppe, Lippert, Parente, & Di Maro, 2013), minced lamb (Kamruzzaman et al., 2013), gelatin in prawn (Wu et al., 2013), different types of caviar in Caspian caviar (Mohamadi Monavar et al., 2013). Some have been reported in the case of classification between adulterated and pure samples with different percentages of adulteration (Alamprese et al., 2013; Kamruzzaman et al., 2013; Tian et al., 2013), and the case of beef adulterated with pork (Morsy & Sun, 2013; Rohman et al., 2011). In this study, pure beef, pure pork and nine levels of adulteration were employed in order not only to discriminate but also to quantify the minimum possible level of adulteration detected.

As in most of the above cases, HCA, PCA, LDA and PLS-DA were the predominant methods used for data analysis. Results showed that multispectral imaging has the potential to identify adulterated beef samples with pork and vice versa in a rapid, non-invasive way. Furthermore, the variability between meat batches was taken into account—an important issue that is not always presented in the available literature—by using three different batches for model training and testing, and a fourth external batch for validation. Results showed that a 10% adulteration with pork in beef and vice versa could be successfully identified and could thus be considered as a detection limit of the applied method, which can be related to the results by Morsy and Sun (Morsy & Sun, 2013) using NIR spectroscopy, although no external validation was performed in this work, and the results by Alamprese et al.(2013), in the case of adulteration with turkey.

It should be noted that very few of the abovementioned studies use external batch validation; results demonstrated—especially for LDA—the necessity of such an approach in order to exclude cases of overoptimistic results. On the other hand the quantification of the level of adulteration was proved to be a more difficult task. A large number of adulteration classes were used (11 classes in total), whereas in other studies discriminant analysis was performed with fewer categories. For example, Alamprese et al. (2013) used 5 classes for cross-validation, grouping very low adulteration with pure samples. However, in this study the applied method was found to provide additional information on the detection limit of 10% and as such can be considered as an advantage even if the quantification per se is of great importance.

In conclusion, multispectral imaging was used for the first time as a rapid method for food authentication and detection of adulteration of raw meat, illustrating a clear separation of pure vs. adulterated samples. PLS-DA performed better compared to LDA in the case of external batch validation. Moreover, the quantification of the percentage of adulteration was proved to be more challenging. The applied method managed to detect relatively small percentages of adulteration (10% w/w) of pure beef with pork and vice versa. It is therefore evident that multispectral

imaging could be used as a rapid, non-invasive method for the detection of adulteration.

Acknowledgements

This work has been supported by the project "Intelligent multi-sensor system for meat analysis –iMeatSense 550" co-financed by the European Union (European Social Fund – ESF) and Greek national funds through the Operational Program "Education and Lifelong Learning" of the National Strategic Reference Framework (NSRF)—Research Funding Program: ARISTEIA-I.

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.foodres.2014.10.032.

References

Alamprese, C., Casale, M., Sinelli, N., Lanteri, S., & Casiraghi, E. (2013). Detection of minced beef adulteration with turkey meat by UV–vis, NIR and MIR spectroscopy. *LWT - Food Science and Technology, 53*(1), 225–232, http://dx.doi.org/10.1016/j.lwt.2013.01.027.

Ballin, N.Z. (2010). Authentication of meat and meat products. *Meat Science*, *86*(3), 577–587, http://dx.doi.org/10.1016/j.meatsci.2010.06.001.

Ballin, N.Z., Vogensen, F.K., & Karlsson, A.H. (2009). Species determination - Can we detect and quantify meat adulteration? *Meat Science*, *83*(2), 165–174, http://dx.doi.org/10.1016/j.meatsci.2009.06.003.

Barbin, D.F., Sun, D. -W., & Su, C. (2013). NIR hyperspectral imaging as non-destructive evaluation tool for the recognition of fresh and frozen–thawed porcine longissimus dorsi muscles. *Innovative Food Science & Emerging Technologies*, *18*, 226–236, http://dx.doi.org/10.1016/j.ifset.2012.12.011.

Barker, M., & Rayens, W. (2003). Partial least squares for discrimination. *Journal of Chemometrics*, *17*(3), 166–173, http://dx.doi.org/10.1002/cem.785.

Boulesteix, A. (2004). PLS dimension reduction for classification with microarray data. *Statistical Applications in Genetics and Molecular Biology*, *3*(14), http://dx.doi.org/10.2202/1544-6115.1075 Article33.

Boulesteix, A. -L., & Strimmer, K. (2007). Partial least squares: A versatile tool for the analysis of high-dimensional genomic data. *Briefings in Bioinformatics*, *8*(1), 32–44, http://dx.doi.org/10.1093/bib/bbl016.

Boyacı, İ. H., Temiz, H.T., Uysal, R.S., Velioğlu, H.M., Yadegari, R.J., & Rishkan, M.M. (2014). A novel method for discrimination of beef and horsemeat using Raman spectroscopy. *Food Chemistry*, *148*, 37–41, http://dx.doi.org/10.1016/j.foodchem.2013.10.006.

Carstensen, J.M., & Hansen, J.F. (2003). *An apparatus and a method of recording an image of an object, patent family, EP1051660.*

Daugaard, S.B., Adler-Nissen, J., & Carstensen, J.M. (2010). New vision technology for multidimensional quality monitoring of continuous frying of meat. *Food Control*, *21*(5), 626–632, http://dx.doi.org/10.1016/j.foodcont.2009.09.007.

De Jong, S. (1993). SIMPLS: An alternative approach to partial least squares regression. *Chemometrics and Intelligent Laboratory Systems*, *18*(3), 251–263, http://dx.doi.org/10.1016/0169-7439(93)85002-X.

Ding, H.B., & Xu, R.J. (1999). Differentiation of beef and kangaroo meat by visible/near-infrared reflectance spectroscopy. *Journal of Food Science*, *64*(5), 814–817, http://dx.doi.org/10.1111/j.1365-2621.1999.tb15918.x.

Dissing, B.S., Papadopoulou, O.S., Tassou, C., Ersbøll, B.K., Carstensen, J.M., Panagou, E.Z., et al. (2013). Using multispectral imaging for spoilage detection of pork meat. *Food and Bioprocess Technology*, *6*(9), 2268–2279, http://dx.doi.org/10.1007/s11947-012-0886-6.

Ebrahim, H. Al, Sowoidnich, K., & Kronfeldt, H. -D. (2013). Raman spectroscopic differentiation of beef and horse meat using a 671 nm microsystem diode laser. *Applied Physics B*, *113*(2), 159–163, http://dx.doi.org/10.1007/s00340-013-5677-x.

Ellis, D.I., Broadhurst, D., Clarke, S.J., & Goodacre, R. (2005). Rapid identification of closely related muscle foods by vibrational spectroscopy and machine learning. *The Analyst*, *130*(12), 1648–1654, http://dx.doi.org/10.1039/b511484e.

ElMasry, G., Sun, D. -W., & Allen, P. (2011). Non-destructive determination of water-holding capacity in fresh beef by using NIR hyperspectral imaging. *Food Research International*, *44*(9), 2624–2633, http://dx.doi.org/10.1016/j.foodres.2011.05.001.

Everitt, B., Landau, S., Leese, M., & Stahl, D. (2011). *Cluster analysis* (5th ed.). Wiley.

Fisher, R.A. (1936). The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, *7*(2), 179–188, http://dx.doi.org/10.1111/j.1469-1809.1936.tb02137.x.

Folm-Hansen, J. (1999). *On chromatic and geometrical calibration.* Ph.D. thesis). Technical University of Denmark (DTU), Department of Mathematical Modeling.

Giaretta, N., Di Giuseppe, A.M.A., Lippert, M., Parente, A., & Di Maro, A. (2013). Myoglobin as marker in meat adulteration: A UPLC method for determining the presence of pork meat in raw beef burger. *Food Chemistry*, *141*(3), 1814–1820, http://dx.doi.org/10.1016/j.foodchem.2013.04.124.

Huang, L., Zhao, J., Chen, Q., & Zhang, Y. (2013). Rapid detection of total viable count (TVC) in pork meat by hyperspectral imaging. *Food Research International*, *54*(1), 821–828, http://dx.doi.org/10.1016/j.foodres.2013.08.011.

Jolliffe, I.T. (2002). *Principal component analysis* (2nd ed.). New York: Springer, http://dx.doi.org/10.1007/b98835.

Kamruzzaman, M., Barbin, D., ElMasry, G., Sun, D. -W., & Allen, P. (2012). Potential of hyperspectral imaging and pattern recognition for categorization and authentication of red meat. *Innovative Food Science & Emerging Technologies*, *16*, 316–325, http://dx.doi.org/10.1016/j.ifset.2012.07.007.

Kamruzzaman, M., Sun, D. -W., ElMasry, G., & Allen, P. (2013). Fast detection and visualization of minced lamb meat adulteration using NIR hyperspectral imaging and multivariate image analysis. *Talanta*, *103*, 130–136, http://dx.doi.org/10.1016/j.talanta.2012.10.020.

Løkke, M.M., Seefeldt, H.F., Skov, T., & Edelenbos, M. (2013). Color and textural quality of packaged wild rocket measured by multispectral imaging. *Postharvest Biology and Technology*, *75*, 86–95, http://dx.doi.org/10.1016/j.postharvbio.2012.06.018.

MATLAB 2012a, The MathWorks, Inc., Natick, Massachusetts, United States. (n.d.). The MathWorks, Inc., Natick, Massachusetts, United States.

Mohamadi Monavar, H., Afseth, N.K., Lozano, J., Alimardani, R., Ornid, M., & Wold, J.P. (2013). Determining quality of caviar from Caspian Sea based on Raman spectroscopy and using artificial neural networks. *Talanta*, *111*, 98–104, http://dx.doi.org/10.1016/j.talanta.2013.02.046.

Morsy, N., & Sun, D. -W. (2013). Robust linear and non-linear models of NIR spectroscopy for detection and quantification of adulterants in fresh and frozen-thawed minced beef. *Meat Science*, *93*(2), 292–302, http://dx.doi.org/10.1016/j.meatsci.2012.09.005.

Panagou, E.Z., Papadopoulou, O., Carstensen, J.M., & Nychas, G. -J.E. (2014). Potential of multispectral imaging technology for rapid and non-destructive determination of the microbiological quality of beef filets during aerobic storage. *International Journal of Food Microbiology*, *174*, 1–11, http://dx.doi.org/10.1016/j.ijfoodmicro.2013.12.026.

Prevolnik, M., Čandek-Potokar, M., & Škorjanc, D. (2010). Predicting pork water-holding capacity with NIR spectroscopy in relation to different reference methods. *Journal of Food Engineering*, *98*(3), 347–352, http://dx.doi.org/10.1016/j.jfoodeng.2009.11.022.

Rohman, A., Sismindari, Erwanto, Y., & Che Man, Y.B. (2011). Analysis of pork adulteration in beef meatball using Fourier transform infrared (FTIR) spectroscopy. *Meat Science*, *88*(1), 91–95, http://dx.doi.org/10.1016/j.meatsci.2010.12.007.

RStudio, v. 0.97.551, RStudio, Inc., Boston, Massachusetts, United States. (n.d.).

Sharifzadeh, S., Clemmensen, L.H., Borggaard, C., Støier, S., & Ersbøll, B.K. (2014). Supervised feature selection for linear and non-linear regression of L*a*b* color from multispectral images of meat. *Engineering Applications of Artificial Intelligence*, *27*, 211–227, http://dx.doi.org/10.1016/j.engappai.2013.09.004.

Soares, S., Amaral, J.S., Mafra, I., & Oliveira, M.B.P.P. (2010). Quantitative detection of poultry meat adulteration with pork by a duplex PCR assay. *Meat Science*, *85*(3), 531–536, http://dx.doi.org/10.1016/j.meatsci.2010.03.001.

Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information Processing & Management*, *45*(4), 427–437, http://dx.doi.org/10.1016/j.ipm.2009.03.002.

Tian, X., Wang, J., & Cui, S. (2013). Analysis of pork adulteration in minced mutton using electronic nose of metal oxide sensors. *Journal of Food Engineering*, *119*(4), 744–749, http://dx.doi.org/10.1016/j.jfoodeng.2013.07.004.

Videometer A/S (official website). (n.d.). Retrieved August 29, 2014, from http://www.videometer.com.

Wu, D., Shi, H., He, Y., Yu, X., & Bao, Y. (2013). Potential of hyperspectral imaging and multivariate analysis for rapid and non-invasive detection of gelatin adulteration in prawn. *Journal of Food Engineering*, *119*(3), 680–686, http://dx.doi.org/10.1016/j.jfoodeng.2013.06.039.

Wu, D., & Sun, D. -W. (2013). Advanced applications of hyperspectral imaging technology for food quality and safety analysis and assessment: A review — Part II: Applications. *Innovative Food Science & Emerging Technologies*, *19*, 15–28, http://dx.doi.org/10.1016/j.ifset.2013.04.016.

Xia, J., Mandal, R., Sinelnikov, I.V., Broadhurst, D., & Wishart, D.S. (2012). MetaboAnalyst 2.0: A comprehensive server for metabolomic data analysis. *Nucleic Acids Research*, *40*(Web Server issue), W127–W133, http://dx.doi.org/10.1093/nar/gks374.

Zhao, M., Downey, G., & O'Donnell, C.P. (2014). Detection of adulteration in fresh and frozen beefburger products by beef offal using mid-infrared ATR spectroscopy and multivariate data analysis. *Meat Science*, *96*(2 Pt A), 1003–1011, http://dx.doi.org/10.1016/j.meatsci.2013.10.015.

Supplementary material for Section 2.3

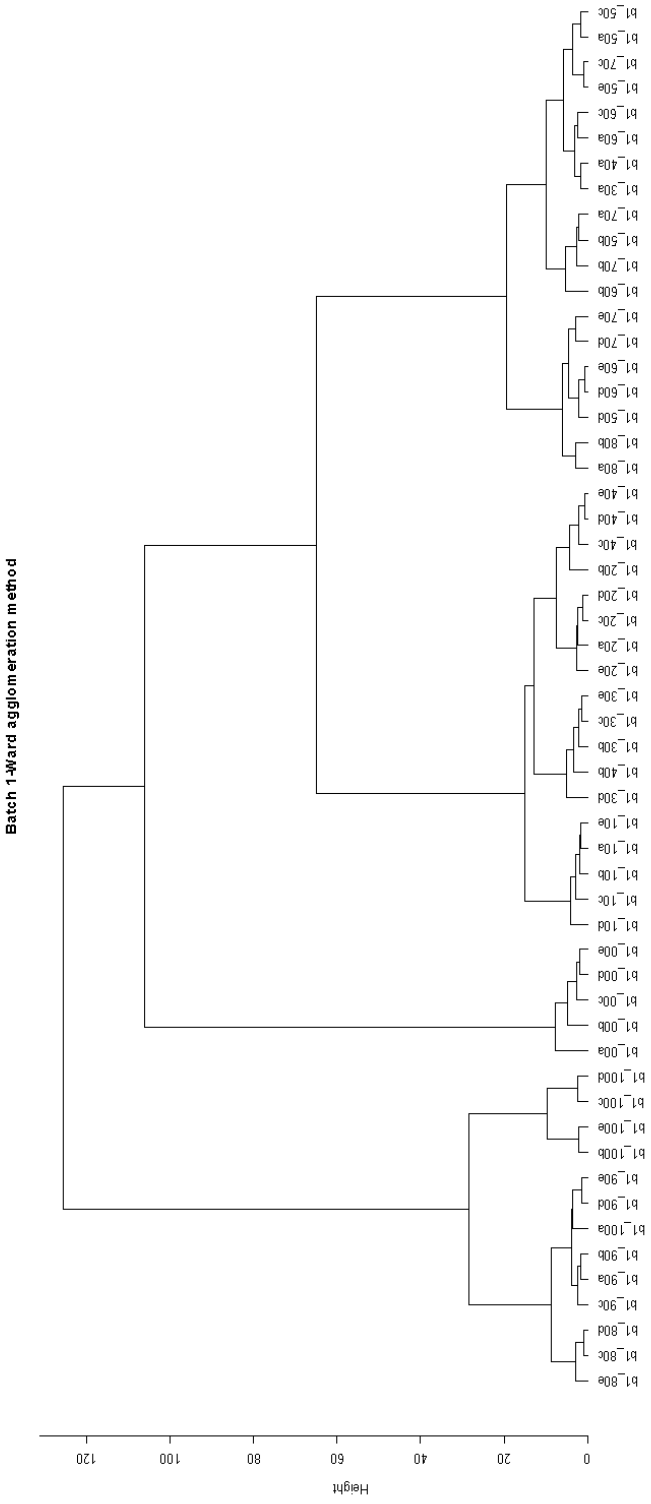


Fig.1(a)

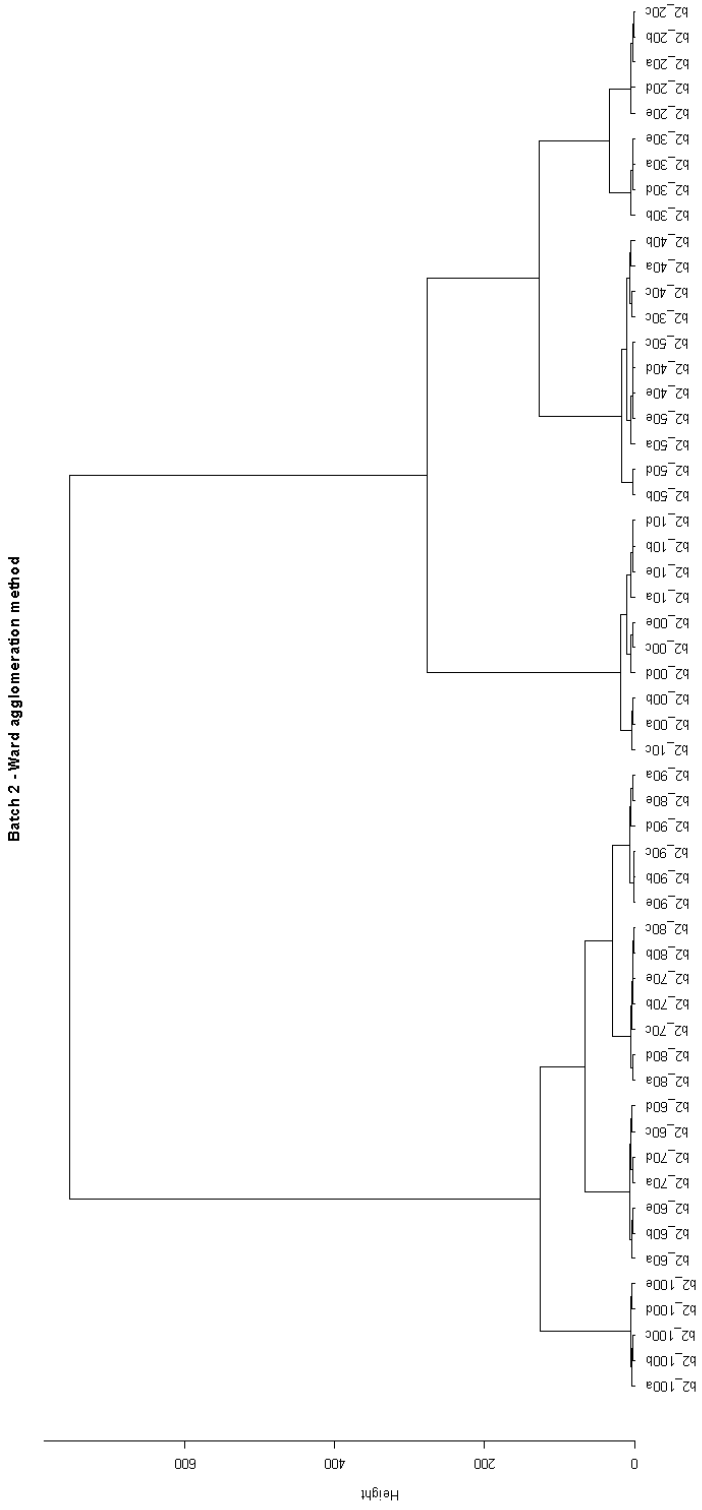

Batch 2 - Ward agglomeration method
Height
0
200
400
600
b2_20c
b2_20b
b2_20a
b2_20d
b2_20e
b2_30e
b2_30a
b2_30d
b2_30b
b2_40b
b2_40a
b2_40c
b2_30c
b2_50c
b2_40d
b2_40e
b2_50e
b2_50a
b2_50d
b2_50b
b2_10d
b2_10b
b2_10e
b2_10a
b2_00e
b2_00c
b2_00d
b2_00b
b2_00a
b2_10c
b2_90a
b2_80e
b2_90d
b2_90c
b2_90b
b2_90e
b2_80c
b2_80b
b2_70e
b2_70b
b2_70c
b2_80d
b2_80a
b2_60d
b2_60c
b2_70d
b2_70a
b2_60e
b2_60b
b2_60a
b2_100e
b2_100d
b2_100c
b2_100b
b2_100a

Fig.1(b)

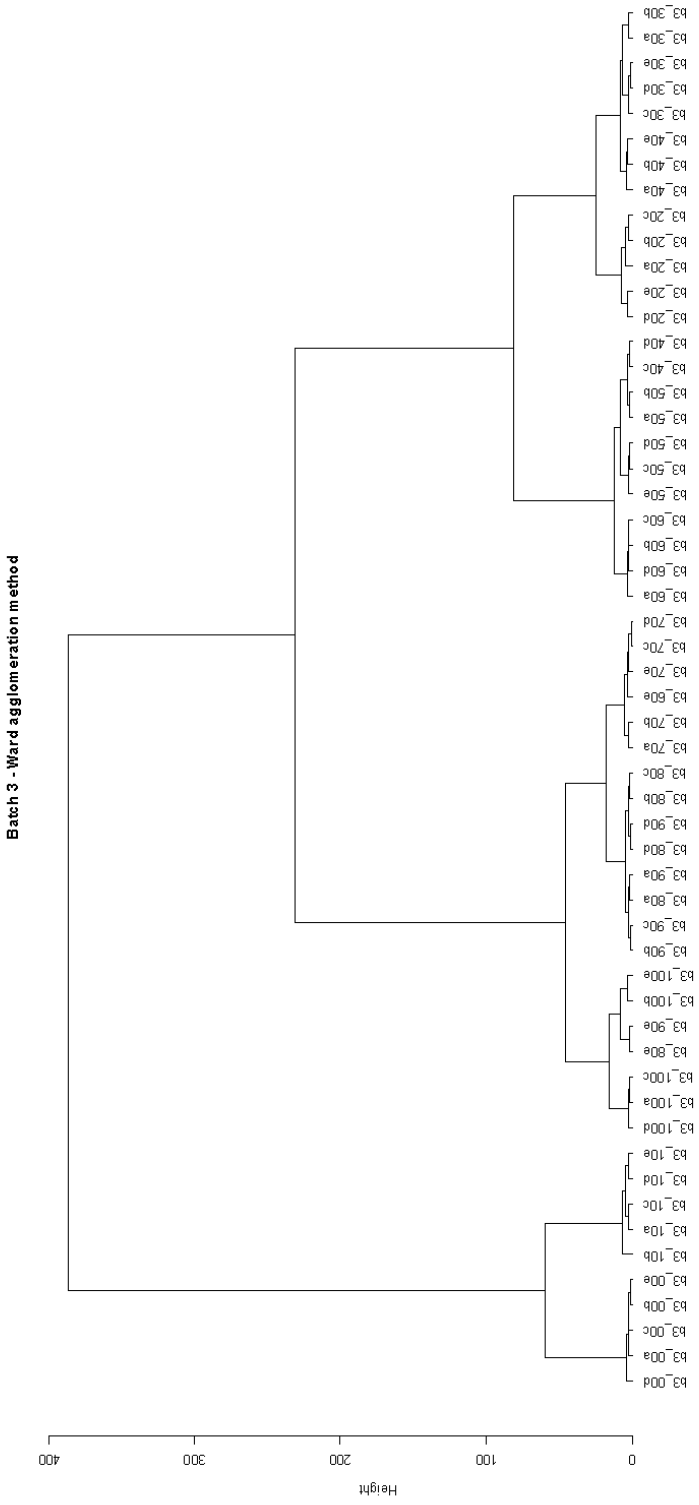


Fig.1(c)

Fig.1 Hierarchical Cluster Analysis for Batches 1(a), 2(b), 3(c) using Ward's minimum variance agglomeration method. Samples named with "00" correspond to pure pork, all other categories correspond to the percentage of beef in the mix and consequently "100" refers to pure beef samples. E.g. b3_70c is an adulterated sample consisting of 70% beef and 30% pork.

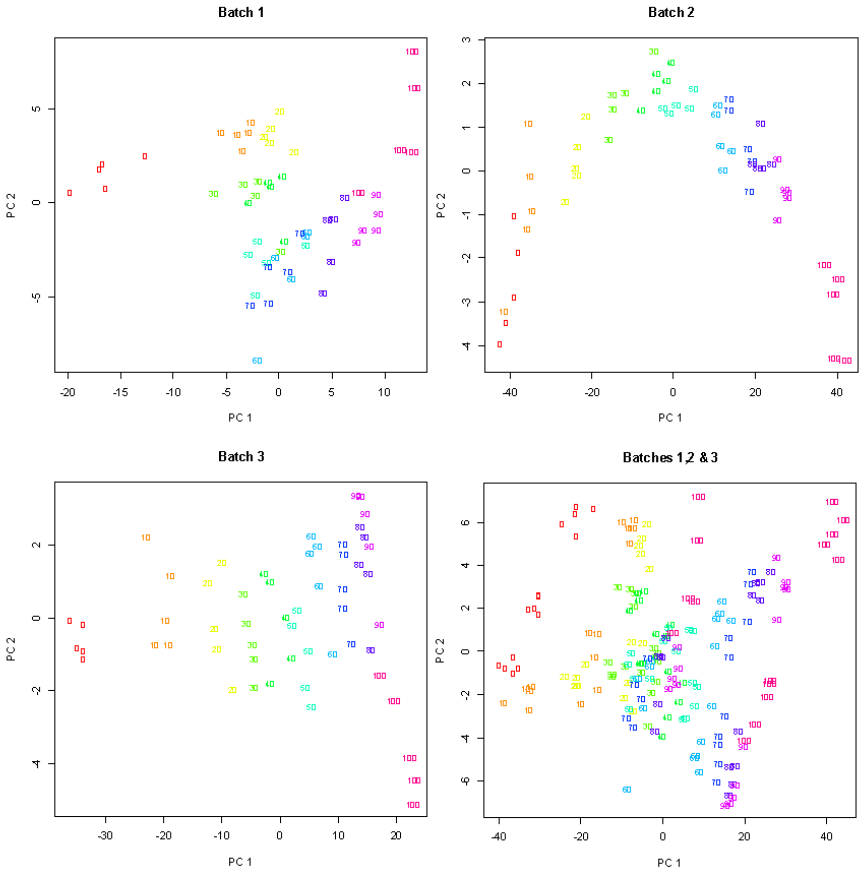


Fig.2.Principal Component Analysis scores for Batches 1(a), 2(b), 3(c) and all batches (d). Samples named with "0" correspond to pure pork, all other categories correspond to the percentage of beef in the mix and consequently "100" refers to pure beef samples.

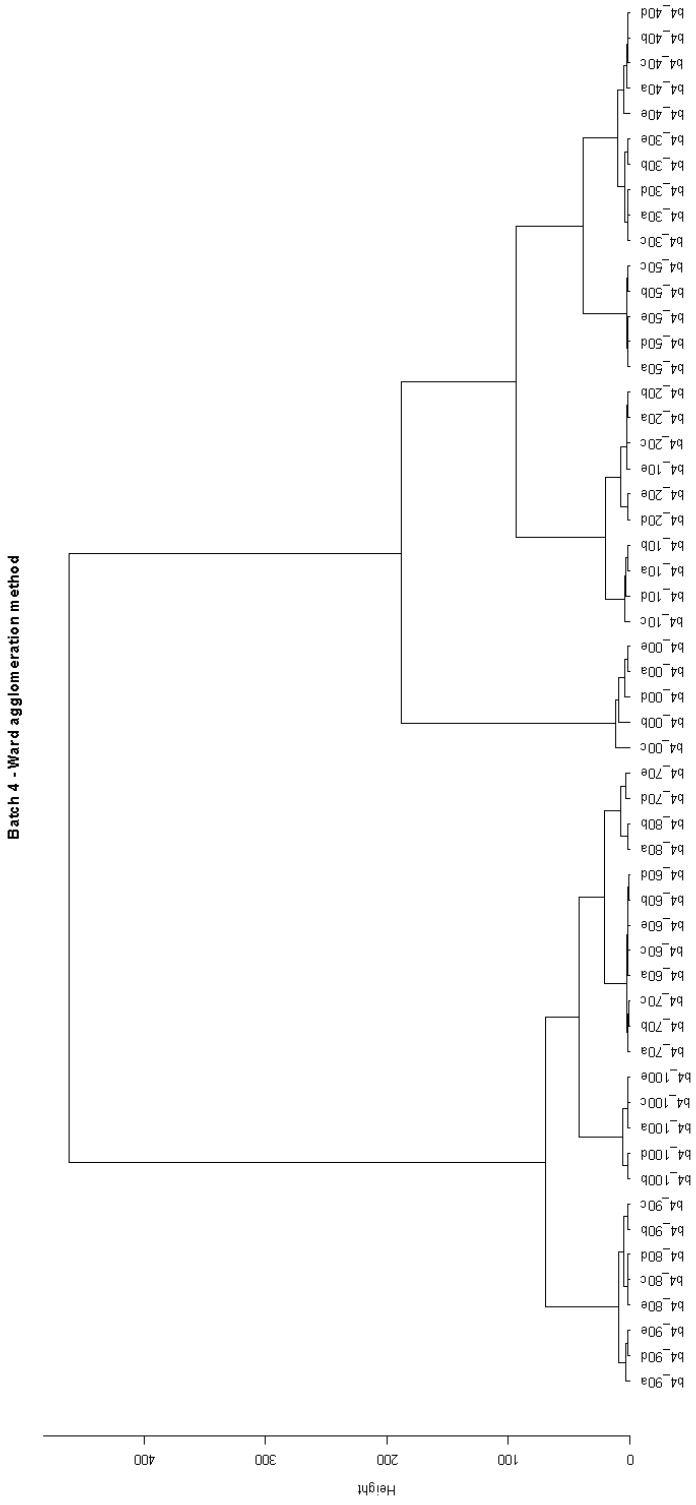


Fig.3 Hierarchical Cluster Analysis for external validation batch using Ward's minimum variance agglomeration method. Samples named with "00" correspond to pure pork, all other categories correspond to the percentage of beef in the mix and consequently "100" refers to pure beef samples

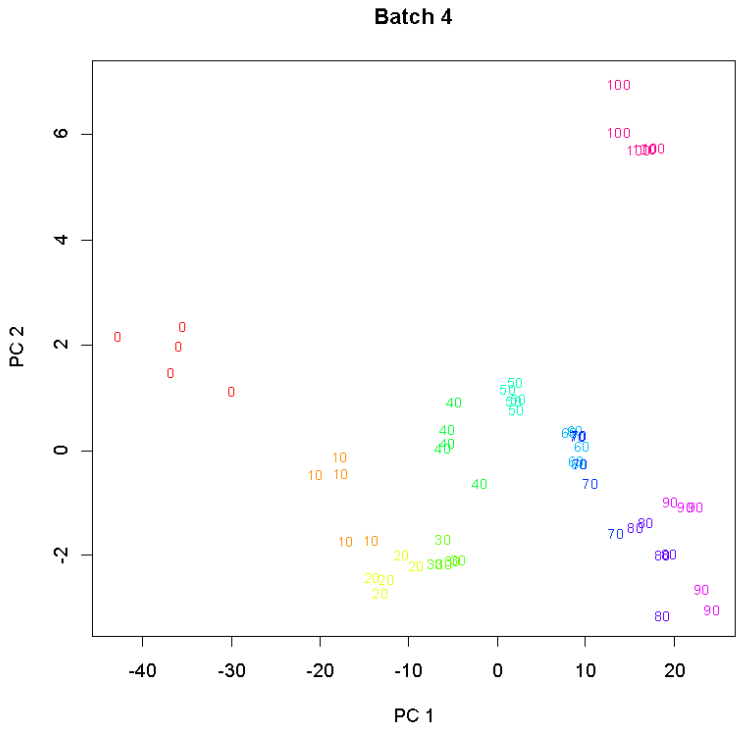


Fig.4.Principal Component Analysis scores for external validation batch. Samples named with "0" correspond to pure pork, all other categories correspond to the percentage of beef in the mix and consequently "100" refers to pure beef samples

Table 1. LDA for validation set with 11 classes of adulteration, 98.48% of the samples are classified within a 10% error.

	Classified as													
sample	0% Beef	10% beef	20% beef	30% beef	40% beef	50% beef	60% beef	70% beef	80% beef	90% beef	100% beef	Recall	±10% error
Is 0%	5	1	0	0	0	0	0	0	0	0	0	83.33%	100.00%
Is 10%	0	6	0	0	0	0	0	0	0	0	0	100.00%	100.00%
Is 20%	0	0	5	1	0	0	0	0	0	0	0	83.33%	100.00%
Is 30%	0	0	0	5	0	1	0	0	0	0	0	83.33%	83.33%
Is 40%	0	0	0	1	4	1	0	0	0	0	0	66.67%	100.00%
Is 50%	0	0	0	0	1	5	0	0	0	0	0	83.33%	100.00%
Is 60%	0	0	0	0	0	0	4	2	0	0	0	66.67%	100.00%
Is 70%	0	0	0	0	0	0	1	4	1	0	0	66.67%	100.00%
Is 80%	0	0	0	0	0	0	0	1	5	0	0	83.33%	100.00%
Is 90%	0	0	0	0	0	0	0	0	0	6	0	100.00%	100.00%
Is 100%	0	0	0	0	0	0	0	0	0	0	6	100.00%	100.00%
precision	100.00%	85.71%	100.00%	71.43%	80.00%	71.43%	80.00%	57.14%	83.33%	100.00%	100.00%		

Table 2. PLS-DA (20 PLS components) for validation set with 11 classes of adulteration.

	Classified as												
sample	0% Beef	10% beef	20% beef	30% beef	40% beef	50% beef	60% beef	70% beef	80% beef	90% beef	100% beef	Recall	±10% error
Is 0%	4	2	0	0	0	0	0	0	0	0	0	66.67%	100.00%
Is 10%	0	6	0	0	0	0	0	0	0	0	0	100.00%	100.00%
Is 20%	0	0	5	1	0	0	0	0	0	0	0	83.33%	100.00%
Is 30%	0	0	0	3	3	0	0	0	0	0	0	50.00%	100.00%
Is 40%	0	0	0	0	4	2	0	0	0	0	0	66.67%	100.00%
Is 50%	0	0	0	0	0	6	0	0	0	0	0	100.00%	100.00%
Is 60%	0	0	0	0	1	0	2	3	0	0	0	33.33%	83.33%
Is 70%	0	0	0	0	0	0	1	5	0	0	0	83.33%	100.00%
Is 80%	0	0	0	0	0	0	1	2	1	2	0	16.67%	83.33%
Is 90%	0	0	0	0	0	0	0	0	0	5	1	83.33%	100.00%
Is 100%	0	0	0	0	0	0	0	0	0	0	6	100.00%	100.00%
precision	100.00%	75.00%	100.00%	75.00%	50.00%	75.00%	50.00%	50.00%	100.00%	71.43%	85.71%		

Table 3. LDA for external validation batch with 11 classes of adulteration.

	Classified as												
sample	0% Beef	10% beef	20% beef	30% beef	40% beef	50% beef	60% beef	70% beef	80% beef	90% beef	100% beef	Recall	±10% error
Is 0%	5	0	0	0	0	0	0	0	0	0	0	100.00%	100.00%
Is 10%	0	1	4	0	0	0	0	0	0	0	0	20.00%	100.00%
Is 20%	0	0	1	0	4	0	0	0	0	0	0	20.00%	20.00%
Is 30%	0	0	5	0	0	0	0	0	0	0	0	0.00%	100.00%
Is 40%	0	0	1	3	1	0	0	0	0	0	0	20.00%	80.00%
Is 50%	0	0	0	5	0	0	0	0	0	0	0	0.00%	0.00%
Is 60%	0	0	3	1	1	0	0	0	0	0	0	0.00%	0.00%
Is 70%	0	0	0	2	3	0	0	0	0	0	0	0.00%	0.00%
Is 80%	0	0	0	0	4	1	0	0	0	0	0	0.00%	0.00%
Is 90%	0	0	0	0	0	5	0	0	0	0	0	0.00%	0.00%
Is 100%	0	5	0	0	0	0	0	0	0	0	0	0.00%	0.00%
precision	100.00%	16.67%	7.14%	0.00%	7.69%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%		

Table 4. PLS-DA (20 PLS components) for external validation batch with 11 classes of adulteration. All pure samples are classified correctly.

	Classified as												
sample	0% Beef	10% beef	20% beef	30% beef	40% beef	50% beef	60% beef	70% beef	80% beef	90% beef	100% beef	Recall	±10% error
Is 0%	5	0	0	0	0	0	0	0	0	0	0	100.00%	100.00%
Is 10%	0	0	1	4	0	0	0	0	0	0	0	0.00%	20.00%
Is 20%	0	0	0	1	4	0	0	0	0	0	0	0.00%	20.00%
Is 30%	0	0	0	0	2	0	3	0	0	0	0	0.00%	40.00%
Is 40%	0	0	0	0	1	0	2	2	0	0	0	20.00%	20.00%
Is 50%	0	0	0	1	0	0	3	1	0	0	0	0.00%	60.00%
Is 60%	0	0	0	0	0	0	0	0	0	5	0	0.00%	0.00%
Is 70%	0	0	0	0	0	0	0	0	0	4	1	0.00%	0.00%
Is 80%	0	0	0	0	0	0	1	0	0	1	3	0.00%	20.00%
Is 90%	0	0	0	0	0	0	0	0	0	5	0	100.00%	100.00%
Is 100%	0	0	0	0	0	0	0	0	0	0	5	100.00%	100.00%
precision	100.00%	0.00%	0.00%	0.00%	14.29%	0.00%	0.00%	0.00%	0.00%	33.33%	55.56%		

Table 5. Loading matrix for the first two Principal Components (PCs).

	X405	X435	X450	X470	X505	X525	X570	X590	X630	X645	X660	X700	X850	X870	X890	X910	X940	X970
batch 1	-0.0468	-0.0588	-0.0903	-0.1657	-0.1782	-0.1578	-0.1391	-0.128	-0.2922	-0.307	-0.2843	-0.1273	-0.3124	-0.3331	-0.3429	-0.308	-0.3037	-0.2675
	0.0358	-0.0202	0.0013	0.1388	0.0653	0.0199	-0.0091	-0.2892	-0.1618	-0.1626	-0.2569	-0.7888	0.108	0.1599	0.1746	0.148	0.1546	0.1227
batch 2	-0.0384	-0.0991	-0.1384	-0.1833	-0.2083	-0.2038	-0.1992	-0.2538	-0.3005	-0.3081	-0.3131	-0.3097	-0.2748	-0.2737	-0.2684	-0.2295	-0.2193	-0.1686
	-0.0839	-0.1998	-0.2365	-0.2738	-0.2287	-0.2334	-0.3343	-0.0678	-0.0556	-0.0496	0.0204	0.5118	0.0797	0.0403	0.0175	0.0593	0.0805	0.1542
batch 3	-0.0482	-0.0882	-0.1349	-0.2207	-0.2441	-0.2238	-0.1948	-0.2026	-0.327	-0.335	-0.3107	-0.1109	-0.2899	-0.3014	-0.3044	-0.2469	-0.2266	-0.1417
	-0.0357	-0.0209	-0.0601	-0.1845	-0.1673	-0.1297	-0.0802	0.1257	0.0882	0.142	0.2551	0.8325	-0.0978	-0.1198	-0.1216	-0.0646	-0.0442	0.0204
batch 4	-0.0411	-0.0903	-0.1264	-0.1718	-0.2095	-0.1997	-0.1725	-0.2506	-0.3658	-0.3742	-0.3657	-0.2545	-0.252	-0.2556	-0.2554	-0.207	-0.1902	-0.1178
	-0.0192	-0.1206	-0.1847	-0.2522	-0.1931	-0.2053	-0.2792	-0.1938	0.4077	0.4537	0.3996	-0.1324	-0.2018	-0.1795	-0.1511	-0.0857	-0.0815	-0.0197
batches 1,2 & 3	-0.0397	-0.0893	-0.1292	-0.1878	-0.2105	-0.1998	-0.1858	-0.2245	-0.3085	-0.3181	-0.3134	-0.2453	-0.2914	-0.2949	-0.2927	-0.247	-0.2338	-0.1736
	0.0298	-0.0209	0.0011	0.1294	0.081	0.0327	-0.026	-0.253	-0.0575	-0.0574	-0.1646	-0.8016	0.1926	0.2404	0.252	0.1863	0.1708	0.0877
	StdDev. 405	StdDev. 435	StdDev .450	StdDev. 470	StdDev. 505	StdDev. 525	StdDev. 570	StdDev. 590	StdDev. 630	StdDev. 645	StdDev. 660	StdDev. 700	StdDev. 850	StdDev. 870	StdDev. 890	StdDev. 910	StdDev. 940	StdDev. 970
batch 1	-0.0095	-0.0118	-0.0164	-0.0257	-0.0299	-0.0297	-0.0306	-0.0199	0.0017	0.0065	0.0095	0.0158	-0.0032	-0.0041	-0.0047	-0.0043	-0.0065	-0.0116
	-0.0007	-0.0156	-0.0169	-0.005	-0.014	-0.0136	-0.0055	-0.0295	-0.0467	-0.048	-0.049	-0.0532	-0.0452	-0.0456	-0.0465	-0.0388	-0.0369	-0.0242
batch 2	-0.0124	-0.0293	-0.0368	-0.0409	-0.0449	-0.0457	-0.0475	-0.0448	-0.0285	-0.024	-0.0214	-0.0113	-0.0171	-0.0178	-0.0197	-0.0165	-0.0165	-0.0152
	-0.0323	0.0055	0.0511	0.1086	0.1524	0.1445	0.1111	0.1828	0.2453	0.2281	0.21	0.0475	0.0477	0.0589	0.0724	0.0757	0.0723	0.0577
batch 3	-0.011	-0.0199	-0.0253	-0.0323	-0.034	-0.0338	-0.0327	-0.0257	-0.0098	-0.0047	-0.0014	0.0089	0.0015	0.0019	0.0009	0.0009	0.0005	-0.0028
	-0.0207	-0.0262	-0.0371	-0.0628	-0.0591	-0.0539	-0.0424	-0.0572	-0.092	-0.0945	-0.0977	-0.0847	-0.0422	-0.0407	-0.0424	-0.0401	-0.0403	-0.0433
batch 4	-0.012	-0.0261	-0.032	-0.0338	-0.0372	-0.0381	-0.0377	-0.0363	-0.0114	-0.0061	-0.0042	-0.0016	-0.0065	-0.0067	-0.0068	-0.0061	-0.0067	-0.0084
	0.004	-0.0203	-0.0329	-0.0426	-0.0299	-0.0314	-0.0435	0.0032	0.0732	0.0753	0.0765	0.0622	0.034	0.0304	0.0326	0.0257	0.0264	0.0257
batches 1,2 & 3	-0.0116	-0.0246	-0.031	-0.0361	-0.0392	-0.0397	-0.0406	-0.0382	-0.0215	-0.0171	-0.0147	-0.0068	-0.0124	-0.0128	-0.0142	-0.0122	-0.0126	-0.0134
	0.0035	-0.0069	-0.0089	-0.0028	-0.0104	-0.0086	-0.0038	0.0034	-0.0125	-0.0138	-0.0099	-0.0041	-0.0261	-0.0258	-0.0244	-0.0189	-0.0165	-0.006

Food Control 73 (2017) 57–63

Contents lists available at ScienceDirect

Food Control

journal homepage: www.elsevier.com/locate/foodcont

Multispectral imaging (MSI): A promising method for the detection of minced beef adulteration with horsemeat

Athina I. Ropodi, Efstathios Z. Panagou, George-John E. Nychas*

Laboratory of Microbiology and Biotechnology of Foods, Department of Food Science and Human Nutrition, Faculty of Foods, Biotechnology and Development, Agricultural University of Athens (AUA), IeraOdos 75, Athens, 11855, Greece

ARTICLE INFO

Article history:
Received 20 February 2016
Received in revised form
26 May 2016
Accepted 28 May 2016
Available online 29 May 2016

Keywords:
Minced beef
Horsemeat
Adulteration
Multispectral imaging
Data analysis

ABSTRACT

In recent years, detection of fraudulent and deceptive practices has become a major priority in the food industry and inspection authorities. The aim of this study was to investigate the potential of multispectral imaging coupled with data analysis methods for the detection of minced beef adulteration with horsemeat, as well as to explore model performance during storage in refrigerated conditions. For this reason, multispectral images of 110 samples from three different batches of minced beef and horsemeat in 18 wavelengths were acquired. Images were taken again after samples were stored at 4 °C for 6, 24 and 48 h. Classification models (partial least squares discriminant analysis, random forest, support vector machines) based on the first two batches were developed while the third batch was set aside for external/independent validation. Results showed that freshly-ground and stored samples were clearly distinguishable, whereas classification model performance for detection of adulterated samples was significantly affected by changes in meat color during storage. Using a two-step SVM model however, all pure and freshly-groundsamples were classified correctly and the overall correct classification was equal to 95.31% for independent batch validation.

1. Introduction

In recent years, detection of fraudulent and deceptive practices has become a major priority for food monitoring agencies as well as the food industry worldwide, since such practices can lead to consumer loss of confidence. Meat and meat products are important food commodities that have been targets for adulteration in the past and, whether deliberately or accidentally, undeclared admixtures and previously unknown and unpredictable adulterants have been observed. While some of these cases include substitution or partial substitution of high commercial value food commodities with cheaper ones, such as beef adulteration with pork or offal or by adding proteins from several origins(Kamruzzaman, Makino, & Oshita, 2015; Ropodi, Pavlidis, Mohareb, Panagou, & Nychas, 2015; Tian, Wang, & Cui, 2013; Zhao, Downey, & O'Donnell, 2014), non-compliance to label has not only economic, but also quality, safety and socio-religious consequences(Alamprese, Casale, Sinelli, Lanteri, & Casiraghi, 2013).

Standard analytical techniques (e.g. immunological and enzymatic techniques, DNA and protein based assays, triacylglycerol analysis) have been used in the past for the authentication of food products and have been very effective in detecting low levels of adulteration (Ballin, 2010; Soares, Amaral, Mafra, & Oliveira, 2010). However, these methods are expensive, invasive, sophisticated, laborious, and technically demanding and thus cannot be used in large-scale in-, on- or at-line applications (Ding & Xu, 1999; Ropodi, Panagou, & Nychas, 2016).

Hyperspectral and multispectral imaging (HSI and MSI) have been proposed as non-invasive rapid methods for monitoring quality, safety and authenticity of foods and in particular meat and meat products (Ropodi et al., 2016; Wu & Sun, 2013). Specifically, minced meat adulteration has been explored in previously published articles using imaging or spectroscopic techniques. Ropodi et al. (2015) investigated the case of minced beef adulteration with pork using MSI, whereas HSI was used for the detection of minced lamb adulterated with pork (Kamruzzaman, Sun, ElMasry, & Allen, 2013). In terms of minced beef adulteration with horsemeat, Raman spectroscopy has been applied recently with promising results(Boyacı et al., 2014; Zając, Hanuza, & Dymińska, 2014). To our knowledge, MSI has not been used previously in the case of minced beef adulteration with horsemeat. Furthermore, no comparison has been performed so far between freshly-ground meat

* Corresponding author.
E-mail address: gjn@aua.gr (G.-J.E. Nychas).

http://dx.doi.org/10.1016/j.foodcont.2016.05.048
0956-7135/

and meat stored in refrigerated conditions where changes in meat color naturally occur.

Thus, the objective of this study was to (a) evaluate the potential of multispectral imaging in tandem with data analysis techniques to identify and/or quantify horsemeat in minced beef, and (b) explore model performance under refrigerated storage of both type meat samples.

2. Materials and methods

2.1. Experimental design

The experimental procedure consisted of three distinct stages. In the first stage, different levels of adulteration with a 20% step were prepared based on the procedure described by Ropodi et al. (2015). Briefly, fresh beef and horsemeat *Longissimus* muscle fillets were purchased, cut into smaller pieces and ground separately using a domestic meat-mincing machine. The appropriate portions of each meat were mixed in order to achieve fourlevels of adulteration, 20–80%, 40–60%, 60-40% and 80-20% (w/w), as well as pure beef and horsemeat. From each level of adulteration, five different portions of ca.75–80 g ($5 \times 6 = 30$ samples in total) were placed in Petri dishes and snapshots were taken using VideometerLab vision system(Videometer A/S, Hørsholm,Denmark). After preliminary analysis of the data, it was decided to create more samples per category and focus on the pure samples, and the levels of 60–40% and 80–20%(w/w) for beef and horsemeat, respectively, where beef is adulterated with horsemeat. A 90–10% level for the second stage was added in order to explore smaller levels of adulteration. Additionally, the opposite case of adulteration horsemeat with beef would create similar models as shown by Ropodi et al. (2015). Furthermore, eight samples per level ($8 \times 5 = 40$ samples) were prepared and multispectral images were acquired at the time, as well as after the samples were stored in high-precision incubators at 4 °C for 6, 24 and 48 h. A graphical representation of the experimental design is shown in Fig. 1. Lastly, in the third stage, the previous procedure was repeated for validation purposes. From now on, meat batches from each experimental stage will be referred to as batch 1, 2 or 3 (b1, b2 or b3).

In total, 110 samples were prepared and 350 images were acquired (i.e., 30 images from batch 1, $40 \times 4 = 160$ images from batch 2, and $40 \times 4 = 160$ images from batch 3).

2.2. Image acquisition and segmentation

Multispectral images were captured in 18 non-uniformly distributed different wavelengths ranging from 405 to 970 nm. The VideometerLab instrument used was commercialized by "Videometer A/S" (Carstensen & Hansen, 2003) and a more detailed description can be found elsewhere (Panagou, Papadopoulou, Carstensen, & Nychas, 2014; Ropodi et al., 2015). Instrument calibration is a two stage procedure, where a light setup called "autolight" based on the type of object to be recorded and a geometrical and radiometrical calibration using well-defined standard targets are performed (Folm-Hansen, 1999).

While the resulting images provide spectral as well as spatial information, they also include redundant information (e.g.sample background, Petri dish). Image segmentation is an image-processing step applied in order to removeimage background and the Petridishfrom the actual sample, as well as separate adipose from lean tissue. Using the respective routines of the VideometerLab software (version 2.12.39) which controls the operation of the instrument, canonical discriminant analysis (CDA) was employed as a two-step supervised transformation building method to divide the images into regions of interest and using a simple threshold to separate between pixels of lean tissue and other pixels.

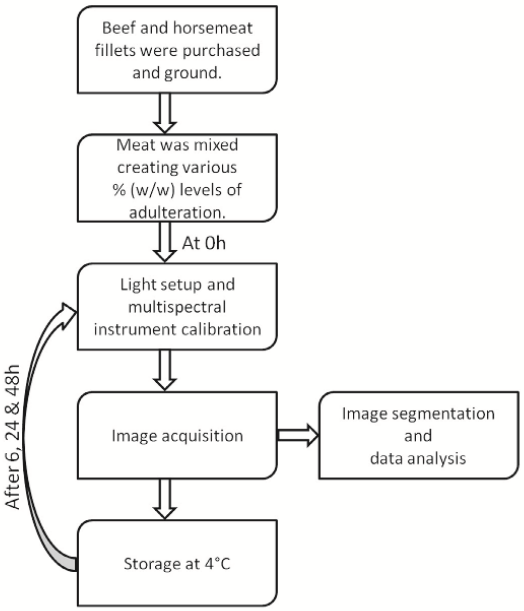


Fig. 1. Graphical representation of experimental design.

2.3. Data analysis

Following image segmentation, the average reflectance values and their standard deviation per wavelength based on pixel intensity values were extracted, resulting in 36 variables (18 mean values and 18 standard deviations) and various data analysis techniques were employed.

Specifically, the unsupervised method of principal component analysis (PCA) (Jolliffe, 2002) was used in order to visualize and interpret data compared to previous works as well as among different experimental stages. Furthermore, various supervised classification techniques were employed in order to discriminate among different levels of adulteration and other classes, such as freshly-ground vs. stored minced meat samples and pure vs. adulterated samples. These techniques are presented below.

(i) Partial least squares discriminant analysis (PLS-DA) (Barker & Rayens, 2003; de Jong, 1993) was employed for various class combinations. The optimum number of PLS components was estimated based on the overall correct classification (OCC) using cross-validation (CV) results of 100 random partitions (80% for training, 20% for testing). Other CV combinations were also performed (results not shown).

(ii) Random forest (RF) (Breiman, 2001) is a supervised learning algorithm which uses an ensemble of classification trees. Ensemble methodologies involve generating multiple classifiers and aggregating their results (e.g. bagging) (Breiman, 2001; Liu, Wang, Wang, & Li, 2013). In RF various parameters were explored and models were chosen based on Out-Of-Bag (OOB) classification error, as a subset of the training

instances is left out in order to obtain an unbiased estimate of the error.

(iii) Support Vector Machines (SVMs) (Cortes & Vapnik, 1995) map the original data points from the input space into a higher dimensional feature space using a kernel function, in order to construct a maximal separating hyper-plane. Various kernel functions were employed and training as well as model selection was performed based on the OOC criterion for 3-fold CV coupled with a grid search for the optimal hyper-parameters and the OOC of a small stratified subset of the training samples. For the two-step model, a SVM model (SVM-1) with b2 training data (including classes 0, 60, 80, 90, and 100% beef) and a radial basis function (RBF) kernel was calibrated using grid search coupled with 3-fold cross validation for calculating the optimal parameters (Capacity = 3, gamma = 0.028). Next, a new SVM model (SVM-2) was developed using as input all freshly-ground data (common levels of adulteration and pure samples of b1 & 2) and a linear kernel (Capacity = 1).

In all supervised methods, while b1 and/or b2 samples were used for model development, b3 samples were reserved for independent model validation, as proposed by Ropodi et al. (2015). This was done so that the models could be tested with an unknown and unbiased dataset in order to exclude overoptimistic results (Ropodi et al., 2015). In addition, b1 and b2 were used for model calibration to take into account the variability within (replicate samples of the same adulteration level) and among meat batches. Furthermore, only the common levels of adulteration were used during model development.

Different classification problems were explored including classification among different percentages of beef (4 classes), pure beef vs. pure horsemeat vs. adulterated samples, pure meat (both beef and horse) vs. adulterated samples and freshly-ground vs. stored minced meat. Model performance was measured mainly in terms of OCC, as well as recall (sensitivity) and precision (Sokolova & Lapalme, 2009).

PCA was performed using MetaboAnalyst 3.0 software (Xia, Sinelnikov, Han, & Wishart, 2015) and PLS-DA was implemented in R v.3.0.2 (http://www.r-project.org/, The R Foundation for Statistical Computing, Vienna, Austria) and Rstudio v.0.97.551 interface (RStudio, Inc., Boston, Massachusetts, USA), using the "plsgenomics" package (Boulesteix & Strimmer, 2007; Boulesteix, 2004; de Jong, 1993). RF was performed in MATLAB 2012a software (The MathWorks, Inc., Natick, Massachusetts, USA) and lastly SVMs were employed using Statistica v.8.0 software (Statsoft Inc., Tulsa, OK, USA).

3. Results and discussion

3.1. Multispectral image data

In Fig. 2, the mean reflectance values in 18 wavelengths is presented for a pure horsemeat(a) and a pure beef (b) sample before and after storage for 6, 24 and 48 h. It is evident that, although stored samples are more difficult to distinguish among themselves, freshly-ground sample are easily differentiated. Additionally, PCA scores for batch 1 are presented for principal components (PCs) 1and 3 are found in Fig. 3. Indeed, PC1 vs. PC3 scores displayed a distinct separation of pure vs. adulterated samples. Pure beef and horsemeat samples were located on the top right of the plot and although they seemed close, discrimination was more evident in the PC2 vs. PC5 plot (Supplementary, Fig. 1). Furthermore, various levels of adulteration were easily distinguishable with only adjacent categories sometimes overlapping. These results are in good agreement with PCA analysis performed on Raman spectra of fat samples with a 25% step among different levels of adulteration, where different categories were evident in the PC1 vs. PC2 plot (Boyacı et al., 2014). They are also consistent with the results in the case of adulteration with pork where the same experimental design (with a 10% step) was implemented (Ropodi et al., 2015). However, it should be noted that, while in the latter two cases the differences were apparent using the first two components and therefore the PCs with the highest variance explained (%), this was not the case in b1 data. In fact, PCs 1 vs. 3 and PCs 2 vs. 3 (Supplementary, Fig. 2) displayed higher discriminatory power than PC1 vs. PC2 (Supplementary, Fig. 3), especially in the case of distinguishing between pure and adulterated samples, denoting a significant difference between the datasets.

Different levels of adulteration were also evident when PCA was applied to b2 freshly-ground image data with the 90-10% (w/w) to seem more difficult to separate from pure beef samples (Supplementary, Fig. 4). Furthermore, freshly-ground and especially pure samples were clearly noticeable indicating the major changes occurring in meat color during storage (Fig. 4). Lastly, as storage time elapsed, less image data from different storage times

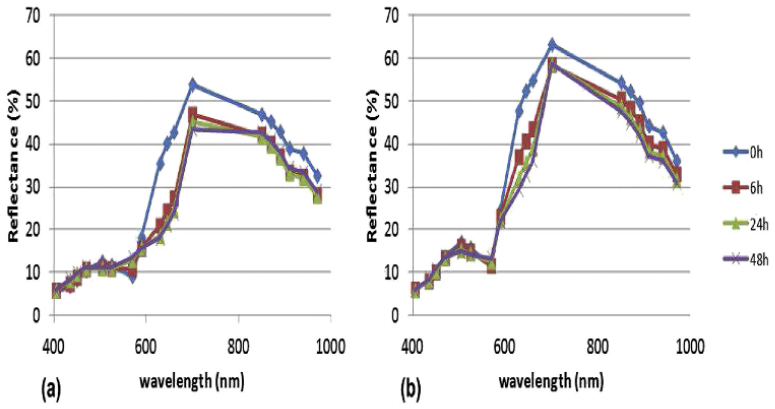


Fig. 2. Mean reflectance (%) values of (a) a beef and (b) a horsemeat sample for 18 distinct wavelengths ranging from 405 to 970 nm. Samples are freshly-ground (0 h) and stored in refrigerated conditions for 6, 24 & 48 h minced.

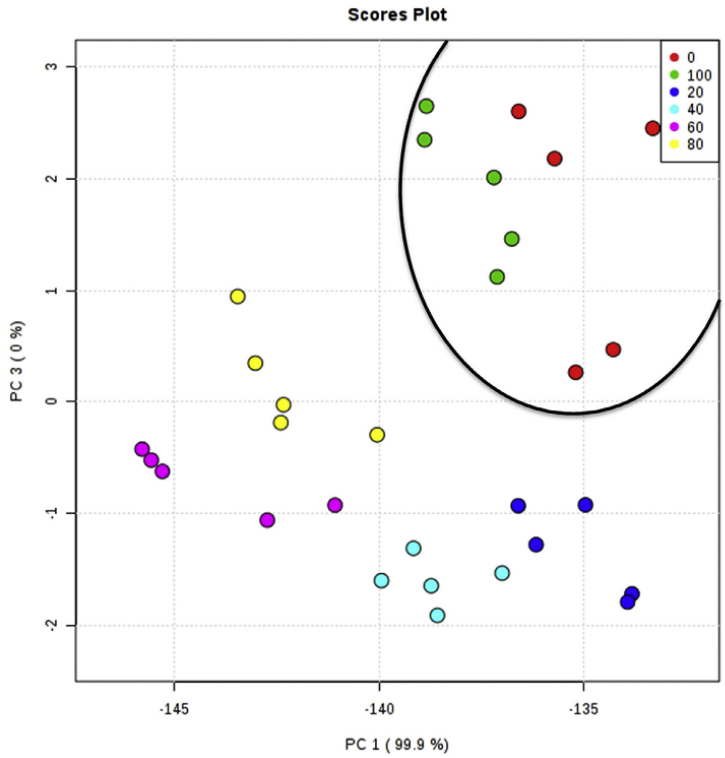


Fig. 3. Principal component analysis scores (PC1 vs. PC3)for batch 1. Samples indicated with "0" correspond to pure horse, all other categories correspond to thepercentage of beef in the mix and consequently "100" refers to pure beef samples.

and percentages of adulteration could be distinguishable.

3.2. Model development and validation

In Table 1, OCC classification results for PLS-DA and RF are presented. In the case of PLS-DA for the different classification problems explored including classification among different percentages of beef (4 classes), pure beef vs. pure horsemeat vs. adulterated samples, pure meat (both beef and horse)vs. adulterated samples and freshly-ground vs. stored minced meat. Results showed that the majority of the models significantly underperformed in the stage of independent validation, although they exhibited very good performance during calibration. PLS-DA yielded a 66.41% OCC in the case of freshly-ground vs. stored meat samples, however only one freshly-ground sample was misclassified as stored. Furthermore, when the proportion of observations to be included in the training set at each cross-validation iteration was set to 0.2, results improved and mean per-class Recall (sensitivity) and Precision were 89.06 and 81.00%, respectively (Supplementary Table 1). While this may imply that another methodology should be used, lack of generalization in the former modelsmay also be attributed to a case of overfitting due to inadequate representative training samples. Indeed, b1 and b2 datasets include two batches of freshly ground samples, but only one batch of stored samples in refrigerated conditions. This may have led to a lack of sufficient data for adulteration detection. However, the improved results of the latter PLS-DA model and PCA scores for b2 (Fig. 4) suggest that data were sufficient for the "freshly-ground vs. stored" classification problem and possibly "freshly-ground and pure vs. other samples.

SVM models were calibrated in order to solve this problemand yielded very good results. The SVM model (SVM-1) with b2 training data and a radial basis function (RBF) kernel classified all training samples correctly (OCC = 100%) in the subset retained for internal model validation. Additionally, independent model validation yielded an OCC equal to 99.38%, as well as 99.58% and 98.78% mean per-class recall and precision. In fact, only one testing sample, a minced horsemeat sample stored for 6 h, was misclassified(Supplementary, Table 2). The above classification results led to considering a new two-step approach in order to identify pure as well as freshly-ground samples, namely (a) classifying samples as freshly-ground or stored, and (b) classifying the former as pure or adulterated.

In Fig. 5, a graphical presentation of the classification results is presented. Using SVM-1, only one "stored" sample was misclassified. Applying the SVM-2 model, all pure samples were classified as pure, but 5 adulterated samples were misclassified yielding an OCC of 84.38%. However, the misclassified samples were all adulterated with 20% horsemeat, therefore the largest (w/w) percentage of horsemeat where samples were misclassified was 20% after independent validation. The misclassified sample of SVM-1 was classified as pure and as a consequence the final results yielded an OOC equal to 95.31%. The performance of the two-step model is presented in Table 2.

Similar studies on meat authenticity have been published

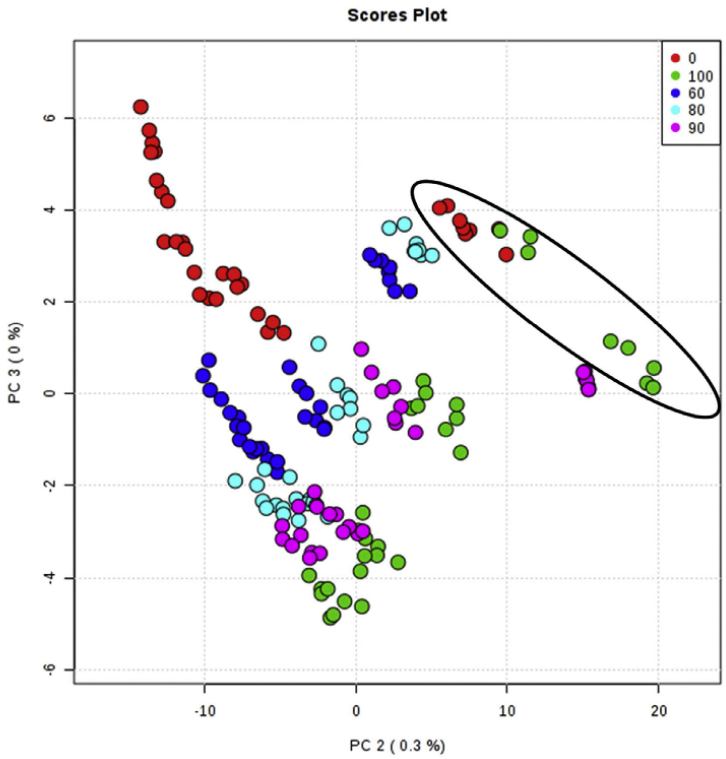


Fig. 4. Principal component analysis scores (PC2 vs. PC3) for batch 2. Samples indicated with "0" correspond to pure horse, all other categories correspond to thepercentage of beef in the mix and consequently "100" refers to pure beef samples. Samples within the ellipse are freshly-ground pure beef and horsemeat.

Table 1
Overall correct classification results for training and validation sets for partial least squares discriminant analysis and random forest.

	Overall correct classification (%)			
	PLS-DA		RF	
	Training set (b1&b2)	Validation set (b3)	OOB prediction	Validation set (b3)
Freshly ground vs. stored (0, 60, 80, 100)	100.00	66.41	96.62	75.00
Pure vs. adulterated (0, 60, 80, 100)	97.30	52.34	95.27	50.00
Pure horse vs. pure beef vs. adulterated (0, 60, 80, 100)	100.00	50.00	95.27	25.00
4 classes 0, 60, 80&100	98.65	37.50	92.57	25.00

previously, but they usually focus on classification among different types of meat such as chicken vs. turkey (Ellis, Broadhurst, Clarke, & Goodacre, 2005), beef vs. kangaroo (Ding & Xu, 1999) and beef vs.horsemeat (Boyacı et al., 2014; Ebrahim, Sowoidnich, & Kronfeldt, 2013) using FT-IR, visible/NIR and Raman spectroscopy. In addition, hyperspectral imaging has been employed for the differentiation of various types of meat such as pork vs. beefvs. lamb (Kamruzzaman, Barbin, ElMasry, Sun, & Allen, 2012). MSI has only been used in the detection of adulteration of minced beef with pork (Ropodi et al., 2015). In fact, little work has been undertaken in terms of multiple levels of meat adulteration with horsemeat (Boyacı et al., 2014), but without extended model development and validation using different batches of meat. On the other hand, published studies have presented cases of adulteration of poultry with pork (Soares et al., 2010), pork in beef meatball and in minced mutton(Rohman, Sismindari, Erwanto, & Che Man, 2011; Tian et al., 2013), beef offal in fresh and frozen beef burger(Zhao et al., 2014), pork meat in raw beef burger and minced lamb (Giaretta, Di Giuseppe, Lippert, Parente, & Di Maro, 2013; Kamruzzaman et al., 2013). Although some of these studies displayed good results, they cannot be compared to this study due to the different type of meat and/or adulterant. While the maximum error was 20%, lower detection limits were found by Morsy and Sun (Morsy & Sun, 2013) using NIR spectroscopy, but with no external validation. Alamprese et al. (2013) used independent samples for validation with very good results in the case of beef adulteration with turkey, but pure samples were grouped together with low levels of adulteration for classification. Lastly, none of the above studies has taken into consideration changes occurring during refrigerated storage.

In conclusion, this study proved that MSI is capable of

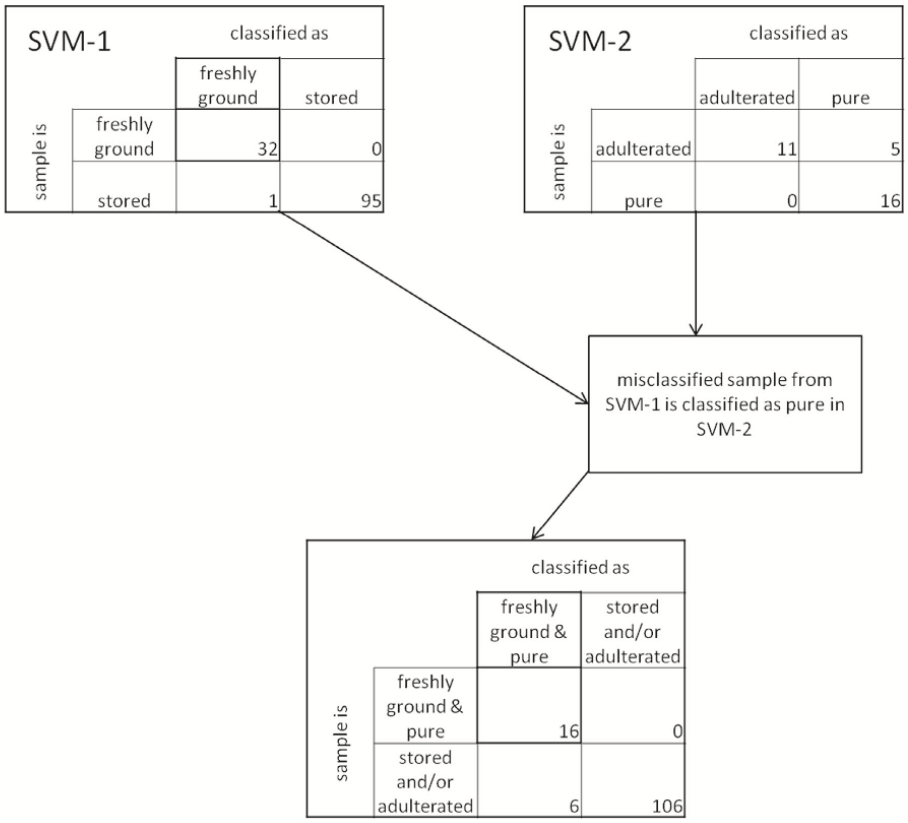


Fig. 5. SVM two-step classification scheme and classification results for the identification of freshly-ground and non-adulterated samples.

Table 2
Final confusion matrix for 2-step SVM model and classification performance indices for validation batch (b3) and samples of common levels of adulteration (100-0, 80-20, 60-40, 0-100% w/w beef and horsemeat respectively).

		To			
		Freshly-ground & pure	Stored and/or adulterated	Recall	Mean recall
From	Freshly-ground & pure	16	0	100.00%	97.32%
	Stored and/or adulterated	6	106	94.64%	
	Precision	72.73%	100.00%		
	Mean precision	86.36%			
				OOC–	95.31%

identifying pure and freshly-groundsamples with a maximum error of 20%. The classification models proved less successful compared to the case of adulteration with pork. However, it also emphasized the necessity of independent sample validation and the significant changes occurring in color during storage. Although these changes affected the model performance greatly, pureand freshly-ground samples were clearly distinguishable, proving that MSI could be used for large scale quality control applications in the future.

Acknowledgements

This work has been supported by the project "Intelligent multisensory system for meat analysis - iMeatSense_550" co-financed bythe European Union (European Social Fund-ESF) and Greek national funds through the Operational Program "Education and Lifelong Learning" of the National Strategic Reference Framework (NSRF)Research Funding Program: ARISTEIA-I.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.foodcont.2016.05.048.

References

Alamprese, C., Casale, M., Sinelli, N., Lanteri, S., & Casiraghi, E. (2013). Detection of

minced beef adulteration with turkey meat by UV–vis, NIR and MIR spectroscopy. *LWT – Food Science and Technology, 53*(1), 225–232. http://dx.doi.org/10.1016/j.lwt.2013.01.027.
Ballin, N. Z. (2010). Authentication of meat and meat products. *Meat Science, 86*(3), 577–587. http://dx.doi.org/10.1016/j.meatsci.2010.06.001.
Barker, M., & Rayens, W. (2003). Partial least squares for discrimination. *Journal of Chemometrics, 17*(3), 166–173. http://dx.doi.org/10.1002/cem.785.
Boulesteix, A.-L. (2004). PLS dimension reduction for classification with microarray data. *Statistical Applications in Genetics and Molecular Biology, 3*(14). http://dx.doi.org/10.2202/1544-6115.1075. Article33.
Boulesteix, A.-L., & Strimmer, K. (2007). Partial least squares: a versatile tool for the analysis of high-dimensional genomic data. *Briefings in Bioinformatics, 8*(1), 32–44. http://dx.doi.org/10.1093/bib/bbl016.
Boyacı, İ. H., Temiz, H. T., Uysal, R. S., Velioğlu, H. M., Yadegari, R. J., & Rishkan, M. M. (2014). A novel method for discrimination of beef and horsemeat using Raman spectroscopy. *Food Chemistry, 148*, 37–41. http://dx.doi.org/10.1016/j.foodchem.2013.10.006.
Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5–32. http://dx.doi.org/10.1023/A:1010933404324.
Carstensen, J. M., & Hansen, J. F.. (2003). An apparatus and a method of recording an image of an object, Patent family EP1051660.
Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning, 20*(3), 273–297. http://dx.doi.org/10.1007/BF00994018.
Ding, H. B., & Xu, R. J. (1999). Differentiation of beef and kangaroo meat by visible/near-infrared reflectance spectroscopy. *Journal of Food Science, 64*(5), 814–817. http://dx.doi.org/10.1111/j.1365-2621.1999.tb15918.x.
Ebrahim, H. Al, Sowoidnich, K., & Kronfeldt, H.-D. (2013). Raman spectroscopic differentiation of beef and horse meat using a 671 nm microsystem diode laser. *Applied Physics B, 113*(2), 159–163. http://dx.doi.org/10.1007/s00340-013-5677-x.
Ellis, D. I., Broadhurst, D., Clarke, S. J., & Goodacre, R. (2005). Rapid identification of closely related muscle foods by vibrational spectroscopy and machine learning. *The Analyst, 130*(12), 1648–1654. http://dx.doi.org/10.1039/b511484e.
Folm-Hansen, J. (1999). *On chromatic and geometrical calibration (Ph.D. thesis)*. Technical University of Denmark (DTU), Department of Mathematical Modeling.
Giaretta, N., Di Giuseppe, A. M. A., Lippert, M., Parente, A., & Di Maro, A. (2013). Myoglobin as marker in meat adulteration: a UPLC method for determining the presence of pork meat in raw beef burger. *Food Chemistry, 141*(3), 1814–1820. http://dx.doi.org/10.1016/j.foodchem.2013.04.124.
Jolliffe, I. T. (2002). *Principal component analysis* (2nd ed.). New York: Springer. http://dx.doi.org/10.1007/b98835.
de Jong, S. (1993). SIMPLS: an alternative approach to partial least squares regression. *Chemometrics and Intelligent Laboratory Systems, 18*(3), 251–263. http://dx.doi.org/10.1016/0169-7439(93)85002-X.
Kamruzzaman, M., Barbin, D., ElMasry, G., Sun, D.-W., & Allen, P. (2012). Potential of hyperspectral imaging and pattern recognition for categorization and authentication of red meat. *Innovative Food Science & Emerging Technologies, 16*, 316–325. http://dx.doi.org/10.1016/j.ifset.2012.07.007.
Kamruzzaman, M., Makino, Y., & Oshita, S. (2015). Non-invasive analytical technology for the detection of contamination, adulteration, and authenticity of meat, poultry, and fish: a review. *Analytica Chimica Acta, 853*, 19–29. http://dx.doi.org/10.1016/j.aca.2014.08.043.
Kamruzzaman, M., Sun, D.-W., ElMasry, G., & Allen, P. (2013). Fast detection and visualization of minced lamb meat adulteration using NIR hyperspectral imaging and multivariate image analysis. *Talanta, 103*, 130–136. http://dx.doi.org/10.1016/j.talanta.2012.10.020.
Liu, M., Wang, M., Wang, J., & Li, D. (2013). Comparison of random forest, support vector machine and back propagation neural network for electronic tongue data classification: application to the recognition of orange beverage and Chinese vinegar. *Sensors and Actuators B: Chemical, 177*, 970–980. http://dx.doi.org/10.1016/j.snb.2012.11.071.
Morsy, N., & Sun, D.-W. (2013). Robust linear and non-linear models of NIR spectroscopy for detection and quantification of adulterants in fresh and frozen-thawed minced beef. *Meat Science, 93*(2), 292–302. http://dx.doi.org/10.1016/j.meatsci.2012.09.005.
Panagou, E. Z., Papadopoulou, O., Carstensen, J. M., & Nychas, G.-J. E. (2014). Potential of multispectral imaging technology for rapid and non-destructive determination of the microbiological quality of beef filets during aerobic storage. *International Journal of Food Microbiology, 174*, 1–11. http://dx.doi.org/10.1016/j.ijfoodmicro.2013.12.026.
Rohman, A., Sismindari, Erwanto, Y., & Che Man, Y. B. (2011). Analysis of pork adulteration in beef meatball using Fourier transform infrared (FTIR) spectroscopy. *Meat Science, 88*(1), 91–95. http://dx.doi.org/10.1016/j.meatsci.2010.12.007.
Ropodi, A. I., Panagou, E. Z., & Nychas, G.-J. E. (2016). Data mining derived from food analyses using non-invasive/non-destructive analytical techniques; determination of food authenticity, quality & safety in tandem with computer science disciplines. *Trends in Food Science & Technology, 50*, 11–25. http://dx.doi.org/10.1016/j.tifs.2016.01.011.
Ropodi, A. I., Pavlidis, D. E., Mohareb, F., Panagou, E. Z., & Nychas, G.-J. E. (2015). Multispectral Image Analysis approach to detect adulteration of beef and pork in raw meats. *Food Research International, 67*, 12–18. http://dx.doi.org/10.1016/j.foodres.2014.10.032.
Soares, S., Amaral, J. S., Mafra, I., & Oliveira, M. B. P. P. (2010). Quantitative detection of poultry meat adulteration with pork by a duplex PCR assay. *Meat Science, 85*(3), 531–536. http://dx.doi.org/10.1016/j.meatsci.2010.03.001.
Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information Processing & Management, 45*(4), 427–437. http://dx.doi.org/10.1016/j.ipm.2009.03.002.
Tian, X., Wang, J., & Cui, S. (2013). Analysis of pork adulteration in minced mutton using electronic nose of metal oxide sensors. *Journal of Food Engineering, 119*(4), 744–749. http://dx.doi.org/10.1016/j.jfoodeng.2013.07.004.
Videometer A/S (official website). (n.d.). Retrieved August 29, 2014, from http://www.videometer.com.
Wu, D., & Sun, D.-W. (2013). Advanced applications of hyperspectral imaging technology for food quality and safety analysis and assessment: a review – part II: applications. *Innovative Food Science & Emerging Technologies, 19*, 15–28. http://dx.doi.org/10.1016/j.ifset.2013.04.016.
Xia, J., Sinelnikov, I. V., Han, B., & Wishart, D. S. (2015). MetaboAnalyst 3.0—making metabolomics more meaningful. *Nucleic Acids Research, 43*(W1), W251–W257. http://dx.doi.org/10.1093/nar/gkv380.
Zając, A., Hanuza, J., & Dymińska, L. (2014). Raman spectroscopy in determination of horse meat content in the mixture with other meats. *Food Chemistry, 156*, 333–338. http://dx.doi.org/10.1016/j.foodchem.2014.02.002.
Zhao, M., Downey, G., & O'Donnell, C. P. (2014). Detection of adulteration in fresh and frozen beefburger products by beef offal using mid-infrared ATR spectroscopy and multivariate data analysis. *Meat Science, 96*(2 Pt A), 1003–1011. http://dx.doi.org/10.1016/j.meatsci.2013.10.015.

Supplementary material for Section 2.4

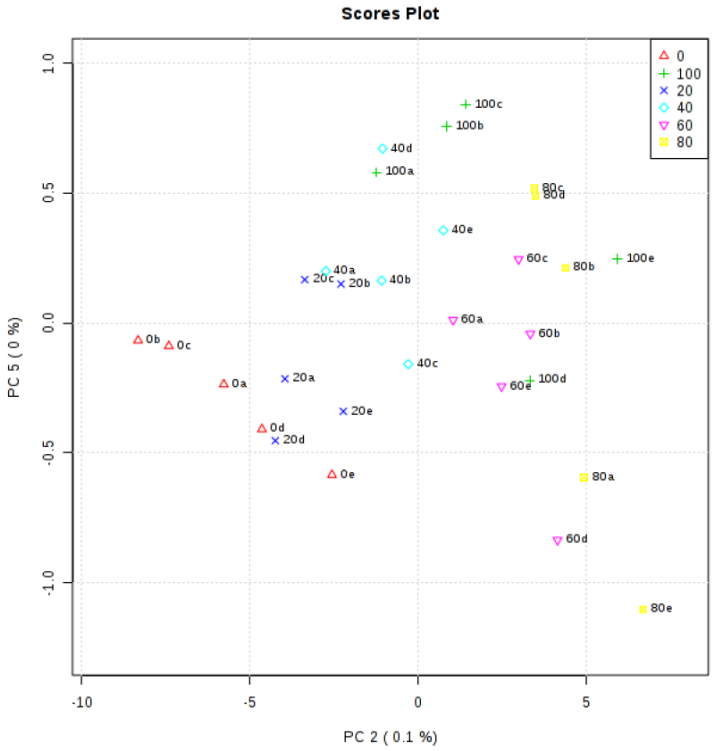


Fig. 2 PC2 vs. PC5 for batch 1

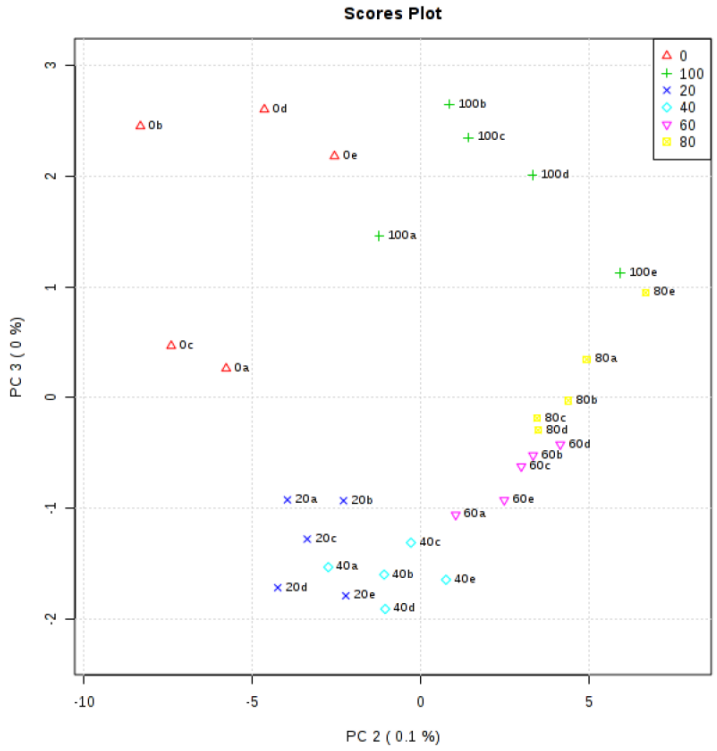


Fig. 2 PC2 vs. PC3 for batch 1

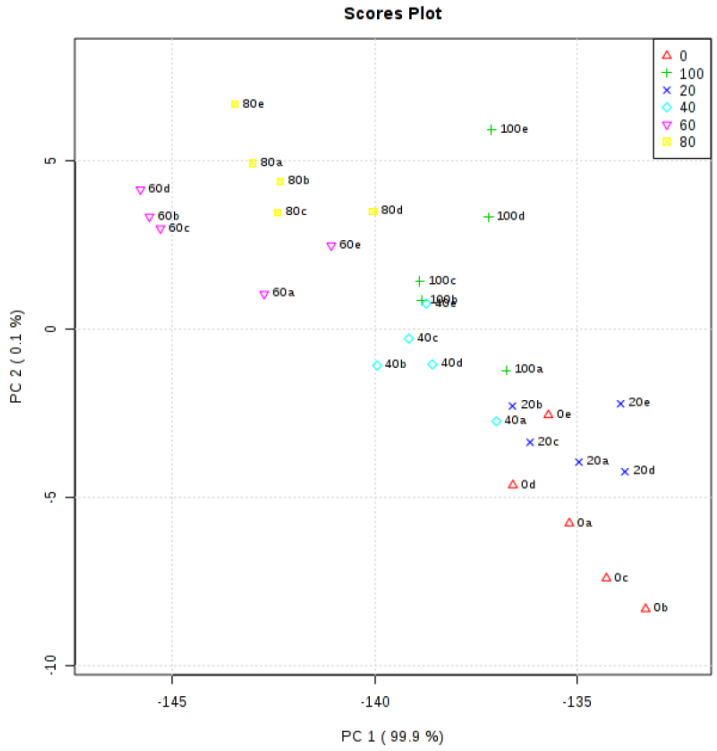


Fig. 3 PC1 vs. PC2 for batch 1

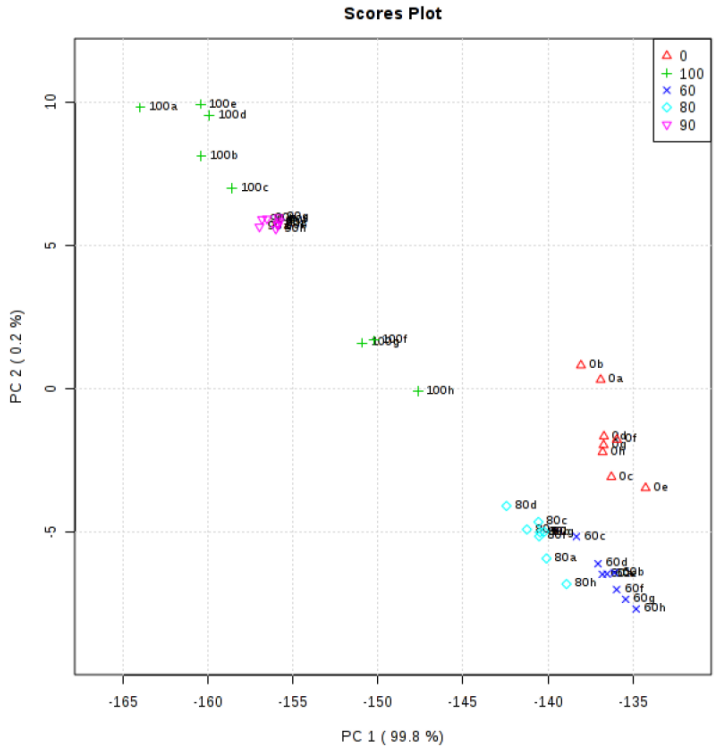


Fig. 4 PC1 vs. PC2 for batch 2 at 0h.

Table 1. PLS-DA classification results for freshly-ground vs. stored samples (classes 0, 60, 80, 100)

b1&b2		To			
		freshly-ground	**stored**	Recall	Mean Recall
From	**freshly-ground**	52	0	100.00%	100.00%
	stored	0	96	100.00%	
	Precision	100.00%	100.00%		
	Mean precision	100.00%			
				OOC=	**100.00%**
b3		**To**			
		freshly-ground	**stored**	Recall	Mean Recall
From	**freshly-ground**	31	1	96.88%	89.06%
	stored	18	78	81.25%	
	Precision	63.27%	98.73%		
	Mean precision	81.00%			
				OOC=	**85.16%**

Table 2. SVM classification results for freshly-ground vs. stored samples (classes 0, 60, 80, 90, 100)

b3 (validation)		To			
		freshly-ground	**stored**	Recall	Mean Recall
From	**freshly-ground**	40	0	100.00%	99.58%
	stored	1	119	99.17%	
	Precision	97.56%	100.00%		
	Mean precision	98.78%			
				OOC=	**99.38%**

Supplementary material for Chapter 3

Table S1. Microbiological analysis (log CFU/gr) of the meat batches (Total Viable Counts - TVC, *Pseudomonas spp.* - CFC, *Brochothrix thermosphacta* - STAA, Lactic acid bacteria – MRS, Enterobacteriaceae – VRBGA).

	PCA	CFC	STAA	MRS	VRBGA	pH
Batch 1	7.85	5.79	7.10	6.22	2.63	5.65
Batch 2	6.84	6.83	6.79	5.34	4.08	5.47
Batch 3	7.14	5.75	5.85	5.08	4.58	5.68
Batch 4	8.06	7.24	6.63	4.93	2.98	5.64
Batch 5	6.09	5.98	5.54	4.15	2.00	5.62
Batch 6	5.85	4.51	4.20	3.20	2.81	5.66
Batch 7	4.85	4.79	3.20	3.77	1.00	5.47

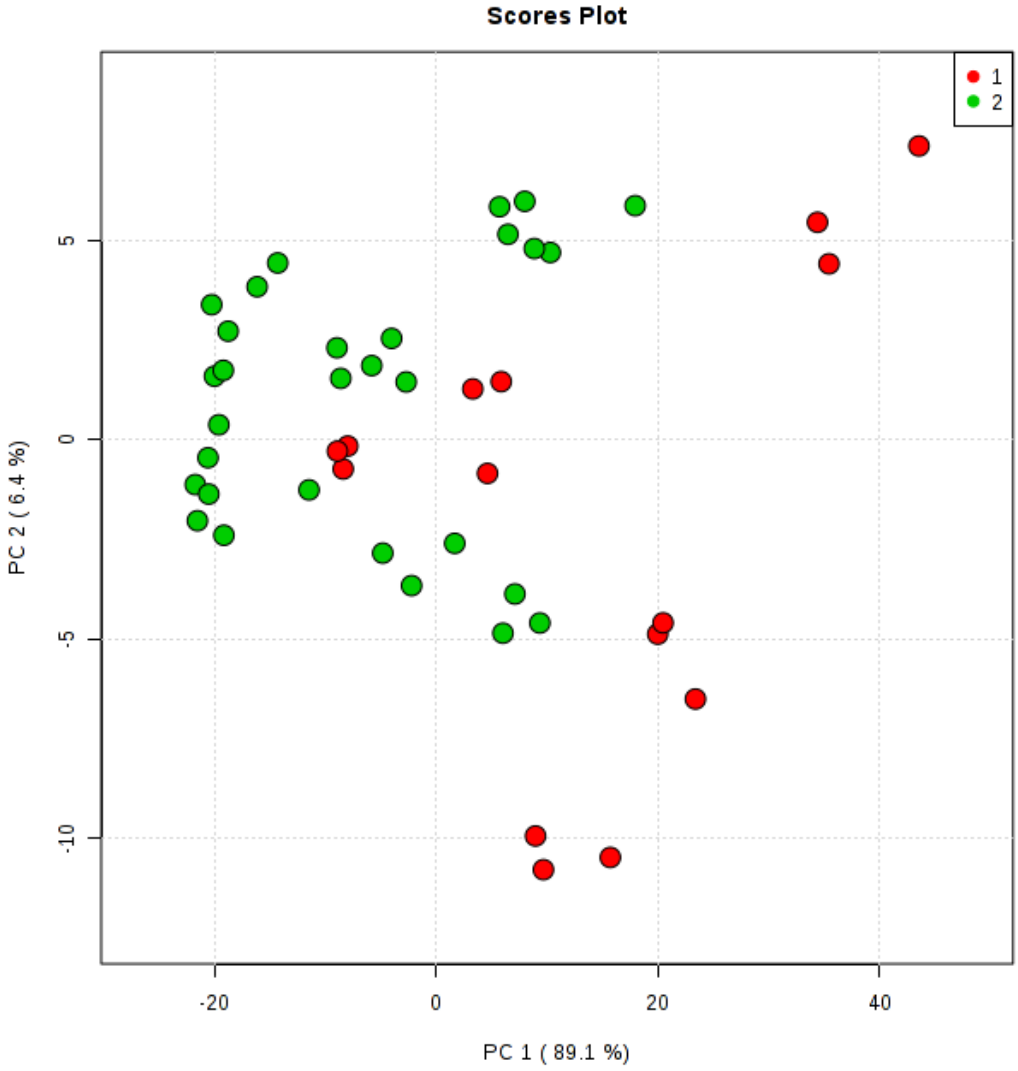


Figure S1. Principal component analysis scores for multispectral imaging training data. Red (class 1) and green (class 2) dots correspond to fresh and frozen-then-thawed samples.

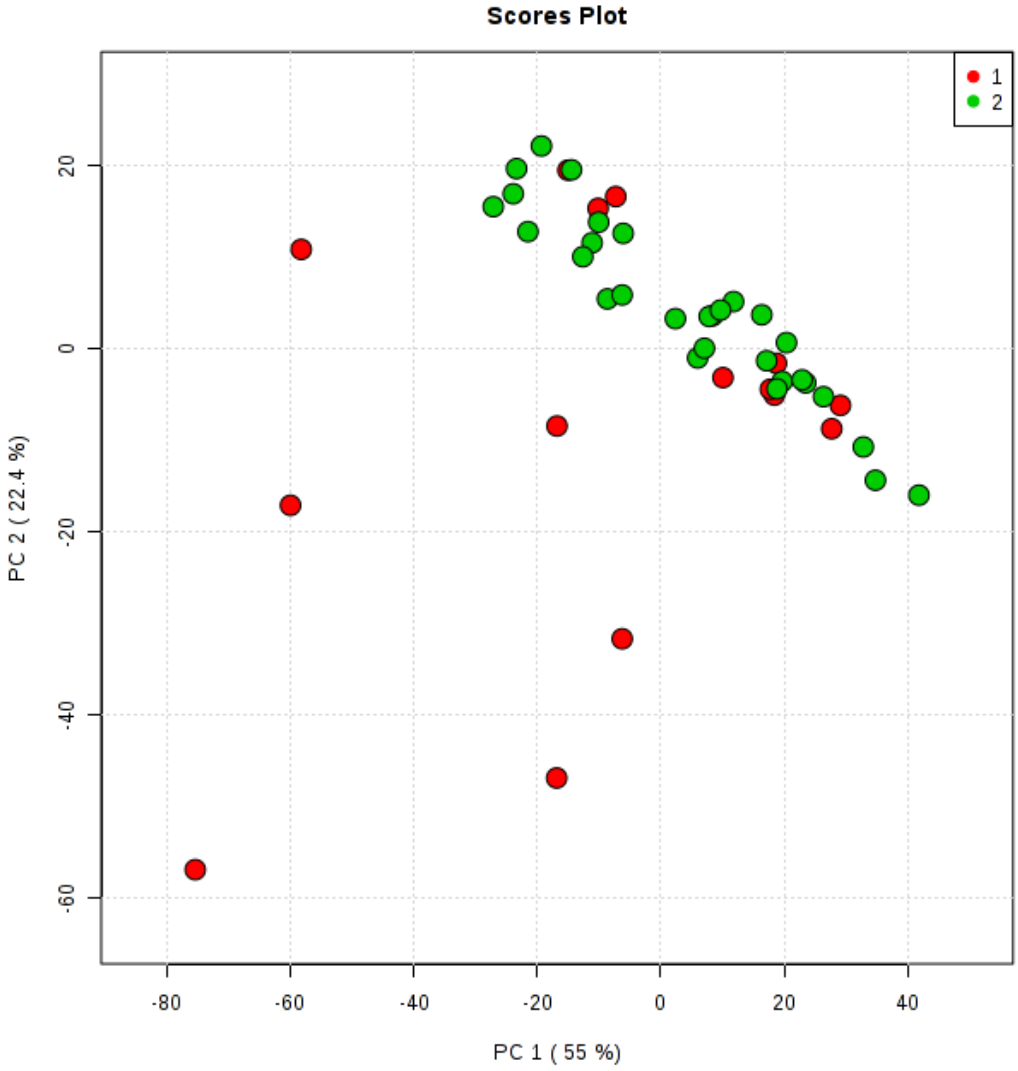


Figure S2. Principal component analysis scores for FTIR training data. Red (class 1) and green (class 2) dots correspond to fresh and frozen-then-thawed samples.

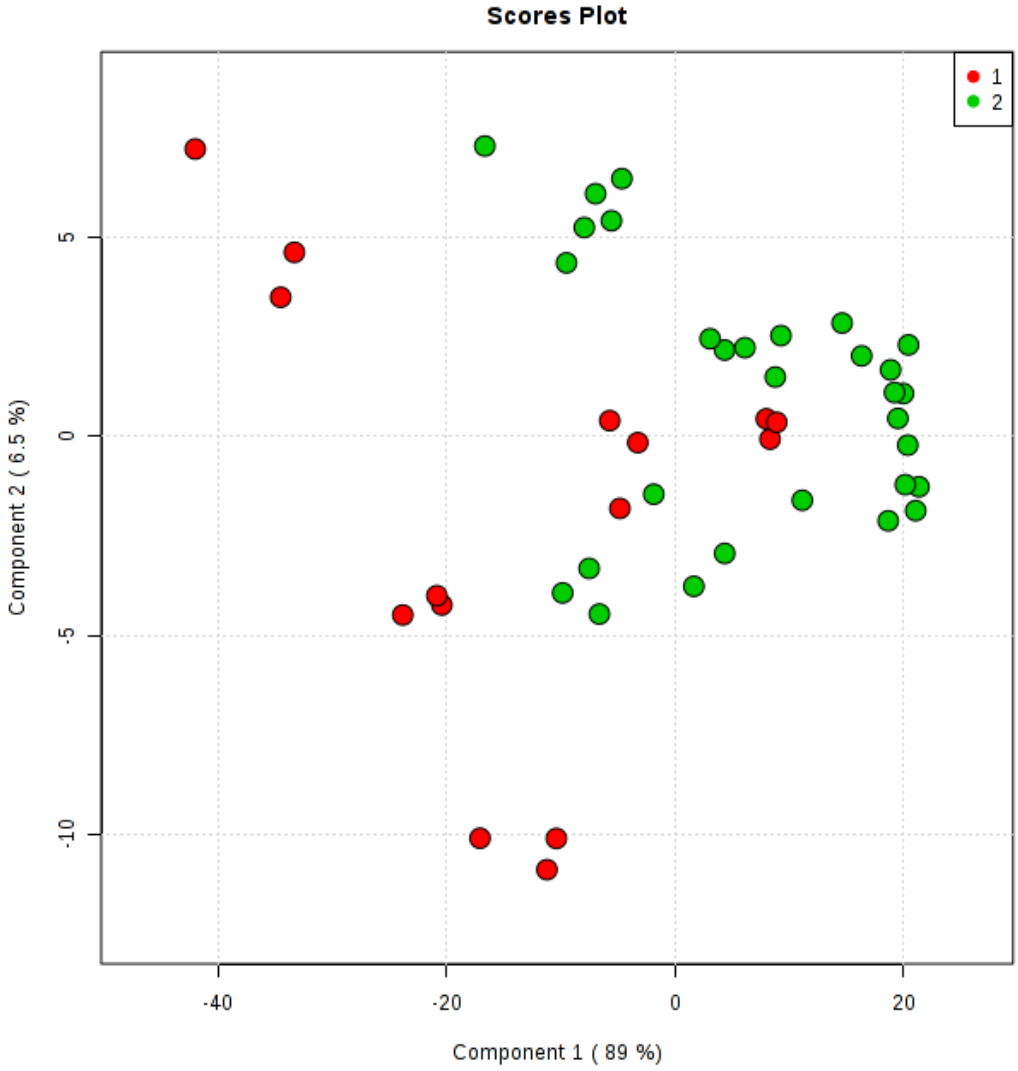


Figure S3. Partial least-squares discriminant analysis scores for multispectral imaging training data (components 1 vs. 2). Red (class 1) and green (class 2) dots correspond to fresh and frozen-then-thawed samples.

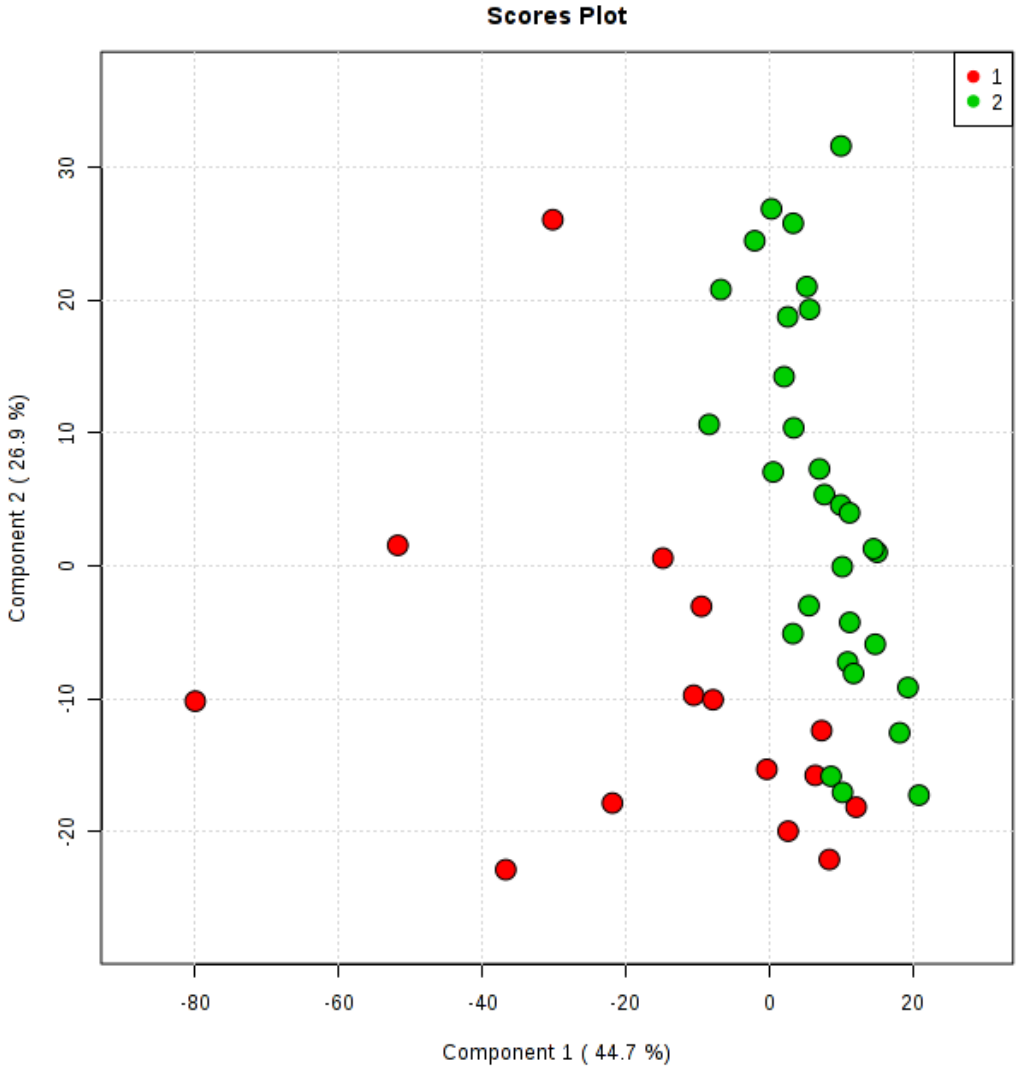


Figure S4. Partial least-squares discriminant analysis scores for FTIR training data (components 1 vs. 2). Red (class 1) and green (class 2) dots correspond to fresh and frozen-then-thawed samples.

Accepted Manuscript

An automated ranking platform for machine learning regression models for meat spoilage prediction using multi-spectral imaging and metabolic profiling

Lucia Estelles-Lopez, Athina Ropodi, Dimitris Pavlidis, Jenny Fotopoulou, Christina Gkousari, Audrey Peyrodie, Efstathios Panagou, George-John Nychas, Fady Mohareb

PII: S0963-9969(17)30205-3
DOI: doi: 10.1016/j.foodres.2017.05.013
Reference: FRIN 6704

To appear in: *Food Research International*

Received date: 6 March 2017
Revised date: 16 May 2017
Accepted date: 18 May 2017

Please cite this article as: Lucia Estelles-Lopez, Athina Ropodi, Dimitris Pavlidis, Jenny Fotopoulou, Christina Gkousari, Audrey Peyrodie, Efstathios Panagou, George-John Nychas, Fady Mohareb , An automated ranking platform for machine learning regression models for meat spoilage prediction using multi-spectral imaging and metabolic profiling, *Food Research International* (2017), doi: 10.1016/j.foodres.2017.05.013

An automated ranking platform for machine learning regression models for meat spoilage prediction using multi-spectral imaging and metabolic profiling

Lucia Estelles-Lopez[a], Athina Ropodi[b], Dimitris Pavlidis[b], Jenny Fotopoulou[b], Christina Gkousari[b] Audrey Peyrodie[a], Efstathios Panagou[b], George-John Nychas[b], Fady Mohareb[a,*]

[a] Bioinformatics Group, Department of Agrifood, School of Water, Energy and Environment Cranfield University, College Road, Cranfield, Bedfordshire, MK43 0AL, UK

[b]Laboratory of Microbiology and Biotechnology of Foods, Department of Food Science and Technology, Agricultural University of Athens, Iera Odos 75, Athens, Greece, GR-11855

Bioinformatics Group, Department of Agrifood, School of Water, Energy and Environment, Cranfield University, College Road, Cranfield, Bedfordshire, MK43 0AL, UK, Tel: +44 (0) 777 588 4779. Email address: f.mohareb@cranfield.ac.uk

Keywords: Machine Learning – Pattern Recognition – Meat spoilage – Metabolic profiling – Data science – Food Quality.

Abbreviations:

SVM-R: Support Vector machines regression
RF-R: Random forests regressions
kNN: k-nearest neighbour
PCA: Principal Component Analysis
OLS-R Ordinary least squares regression
SL-R: Stepwise Linear regression
ML: Machine Learning
RMSE: Root mean square of error
PC-R Principal Component regression

Abstract

Over the past decade, analytical approaches based on vibrational spectroscopy, hyperspectral / multispectral imagining and biomimetic sensors started gaining popularity as rapid and efficient methods for assessing food quality, safety and authentication; as a sensible alternative to the expensive and time-consuming conventional microbiological techniques. Due to the multi-dimensional nature of the data generated from such analyses, the output needs to be coupled with a suitable statistical approach or machine-learning algorithms before the results can be interpreted. Choosing the optimum pattern recognition or machine learning approach for a given analytical platform is often challenging and involves a comparative analysis between various algorithms in order to achieve the best possible prediction accuracy.

In this work, "MeatReg", a web-based application is presented, able to automate the procedure of identifying the best machine learning method for comparing data from several analytical techniques, to predict the counts of microorganisms responsible of meat spoilage regardless of the packaging system applied. In particularly up to 7 regression methods were applied and these are ordinary least squares regression, stepwise linear regression, partial least square regression, principal component regression, support vector regression, random forest and *k*-nearest neighbours.

MeatReg" was tested with minced beef samples stored under aerobic and modified atmosphere packaging and analysed with electronic nose, HPLC, FT-IR, GC-MS and Multispectral imaging instrument. Population of total viable count, lactic acid bacteria, pseudomonads, Enterobacteriaceae and *B. thermosphacta*, were predicted. As a result, recommendations of which analytical platforms are suitable to predict each type of bacteria and which machine learning methods to use in each case were obtained. The developed system is accessible *via* the link: http://elvis.misc.cranfield.ac.uk/SORF/.

1. Introduction

Consumers demand food products, which should not be only perfectly safe for human consumption, but visually attractive as well. Additionally, to be more likely to meet the customers' expectations, foodstuff should include fewer additives, be minimally processed and high in quality. (Van Wezemael, Verbeke, de Barcellos, Scholderer, & Perez-Cueto, 2010). In general, food products are considered spoiled if it is unacceptable by the customer, though not necessarily unsafe (Koutsoumanis, 2009). This is also the case with fresh meat, where spoilage is quite a subjective judgement among consumers, often based in the presence of gross discoloration, strong off-odours and development of slime due to intrinsic or extrinsic factors (G. J. Nychas, P. N. Skandamis, C. C. Tassou, & K. P. Koutsoumanis, 2008).

Meat spoilage is a very complex phenomenon, which involves significant changes and activities of very different microbial groups depending on the storage conditions e.g. packaging, temperature (Doulgeraki, Ercolini, Villani, & Nychas, 2012). However, according to EU authorities (Commission, 2005), the quality of fresh meat is evaluated only by viable counts of bacteria able to grow on very generic medium (Total viable count) or on counts of the Enterobacteriaceae family. On the other hand, it is well established that pseudomonads are the major cause of spoilage in aerobic conditions as once they have used all glucose and lactate available, they start to metabolise the nitrogen sources, producing slime and off-odours (Mohareb et al., 2015). Under modified atmosphere packaging, other microorganisms like *B. thermosphacta*, Enterobacteriaceae and lactic acid bacteria are dominating, whilst under vacuum packaging the dominant species are *Pseudomonas* spp., *B. thermosphacta* and *S. putrefaciens* (G-J.E; Nychas, Marshall, & Sofos, 2007). Counting colonies is certainly time-consuming, costly and provide retrospective information (G. J. E. Nychas, P. N. Skandamis, C. C. Tassou, & K. P. Koutsoumanis, 2008). Moreover, both the analysis of limited samples and/or their low counts, can significantly underestimate the microbial contribution to meat quality. Recently, rapid, non-invasive methods relying on processing large datasets using computational analysis are gaining popularity (George-John E. Nychas, Panagou, & Mohareb, 2016). Although such instruments represent an efficient alternative to conventional microbiological analysis, the experimental output is far more complex and usually needs processing before the results can be interpreted. In the food sector, a plethora of machine learning approaches has been followed by different authors in order to predict spoilage in meat samples using metabolomics data (Comprehensive examples are highlighted in (Ropodi, Panagou, & Nychas, 2016). internet technologies and more specifically open sources platforms will enhance food safety management system (George-John E. Nychas et al., 2016) while will allow supply chains to use virtualizations dynamically in operational management processes. This will improve support for food companies dealing with perishable products, unpredictable supply variations and stringent food

safety and sustainability requirements.

While research involving metabolomics data in tandem with machine learning techniques is extensive, guidelines to choose the machine learning method that provides the best results for a specific type of data are still needed. Furthermore, the actual procedure of optimising and validating the spoilage prediction models is computationally extensive, and often requires the availability of suitable resources and statistical knowledge.

Therefore, the aim of this work is (i) to develop spoilage prediction models from data derived from different analytical instruments, and (ii) to implement an accuracy ranking system through a platform (MeatReg), which assesses the suitability of machine learning methods to specific type of metabolic data provided by a certain analytical process. For this study, metabolomics data from minced beef samples stored under aerobic and modified atmosphere packaging were collected using 5 different analytical and imaging instruments: electronic nose (eNose), High Performance Liquid Chromatography (HPLC), Fourier Transformed Infrared Spectroscopy (FT-IR), Gas Chromatography coupled to Mass Spectrometry (GC-MS) and multi-spectral imaging (MSI).

In particular in this study, seven machine learning methods; namely Ordinary Least Squares regression (OLS-R), Stepwise Linear regression (SL-R), Principal Component regression (PC-R) , Partial Least Squares Regression (PLS-R), Support Vector Regression (SVM-R), Random Forests Regression (RF-R) and k-Nearest Neighbours' Regression (kNN-R) were used to predict bacterial counts for Pseudomonads, Lactobacilli, *B. thermosphacta* and Enterobacteriaceae, as well as for the bacterial total viable count. This way, the most suitable analytical platforms to predict bacterial counts for each type of bacteria present in meat stored under aerobic or modified atmosphere conditions were identified and machine-learning methods were ranked for each scenario according to their performance. Additionally, "MeatReg" was made available online as a web-based application in order to provide a flexible and setup-free mean to automate the whole analysis process since internet technologies and more specifically open sources platforms will enhance food safety management system (George-John E. Nychas et al., 2016) while will allow supply chains to use virtualizations dynamically in operational management processes. This will improve support for food companies dealing with perishable products, unpredictable supply variations and stringent food safety and sustainability requirements.

2. Material and methods

2.1. Sample preparation & microbiological analyses

Fresh minced meat was obtained from a central butcher shop in Athens and transported under refrigeration to the laboratory within 30 min. Portions of approximately 75-80 g were placed on styrofoam trays, were stored in air or under modified air packaging (MAP) conditions (20% CO_2/ 80% O_2) at 4 and 10°C. For aerobic storage, samples were covered with plastic food membrane for domestic use and for MAP storage, samples were packed into plastic pouches of gas permeability at 20°C and 50% relative humidity of *ca*. 25 and 90 cm^3/m^2 per day/10^5 Pa for CO_2 and O_2 respectively, using a HenkoVac 1900 Machine. At appropriate time intervals (approximately every 24 and 12 hours for the case of 4°C and 10°C respectively), multispectral images of duplicate samples were captured and samples were analyzed microbiologically until spoilage was pronounced and sub-samples were stored (-20°C) for FTIR, HPLC, GC-MS and e-nose measurements. Additionally, three more samples at 0 hours (control samples) were analyzed. In total, 105 samples (three control samples and 11-14 duplicate samplings per packaging condition per storage temperature) were analyzed. Generally speaking, meat spoilage and sample discoloration was not evident to the naked eye, except between extreme storage times (See Supplementary Figure S4). Twenty-five gram-portions from each meat samples were weighted aseptically in 400ml sterile stomacher bags (Seward Medical, London, United Kingdom), containing 225ml of sterile quarter Ringer's solution (LabM Limited, Lancashire, United Kingdom) and were homogenized for 60 sec (Lab Blender 400, Seward Medical). Appropriate serial dilutions were prepared with the same Ringer's solution and duplicate 0.1 or 1 mL samples of the appropriate dilutionswere spread or mixed on the following media: plate count agar (PCA, Biolife 4021452, Milano, Italy) for total viable counts (TVC), incubated at 30 °C for 48–72 h; Pseudomonas agar base (PAB, Biolife 401961,Milano, Italy) for Pseudomonas spp., incubated at 25 °C for 48–72 h;streptomycin thallous acetate-actidione agar (STAA, Biolife 402079,Milano, Italy) for *B. thermosphacta*, incubated at 25 °C for 72 h; and de Man–Rogosa–Sharpe medium (MRS, Biolife, 4017282, Milano, Italy) with pH adjusted to 5.7 with 10 N HCl, for lactic acid bacteria overlaid with the same medium and incubated at 30 °C for 48–72 h. All plates were examined visually for typical colony types and morphological characteristics that were associated with each growth medium. Moreover, the selectivity of each medium was routinely checked by Gram staining and microscopic examination of smears prepared from randomly selected colonies obtained from the media.

2.2 Fourier Transform Infrared (FTIR) Spectroscopy

FTIR spectra were acquired using an FT/IR 6200 JASCO spectrometer (Jasco Corp., Tokyo, Japan). Small portions (~3 g) were placed on the surface of a ZnSe 45° HATR (Horizontal Attenuated Total Reflectance) crystal (PIKE Technologies, Madison, Wisconsin, United States) and, using the Spectra Manager software version 2 (Jasco Corp.), spectra were collected from 4000 to 400 cm^{-1} (100 scans, resolution of 4 cm^{-1}) within a period of 2 min. Prior to the measurements, reference spectra were acquired using the crystal with no added meat. After each measurement,

the crystal's surface was cleaned using firstly detergent and distilled water and secondly with analytical grade acetone, and dried with lint-free tissue.

2.3 Image Acquisitions and Analysis

Images from every sample were captured using Multispectral imaging (MSI), a system which acquires multispectral images in 18—non-uniformly distributed wavelengths ranging from 405 to 970nm in visible and short-NIR spectral region. The system's detailed description commercialized by "Videometer A/S (VM)" (http://www.videometer.com) has been reported elsewhere (Panagou, Papadopoulou, Carstensen, & Nychas, 2014). The system was calibrated radiometrically and geometrically using well-defined standard targets, after a light setup based on the type of object to be recorded called "autolight". Petri dishes (75–80 g meat portions) were placed inside an Ulbricht sphere in which the camera is top-mounted. An image-processing step is needed that will result in an image mask where only meat tissue is included. This step was implemented using VideometerLab software (version 2.12.39). The resulting image naturally includes irrelevant information such as the Petri dish and its surrounding background, as well as the fat and connective tissue within the meat. For this reason, an image pre-processing step was performed; which includes transformation and segmentation procedures using the default Videometer lab software which controls the operation of the instrument. This procedure is described in details elsewhere (Ropodi, Pavlidis, Mohareb, Panagou, & Nychas, 2015). Briefly, Canonical discriminant analysis (CDA) was applied to separate the images according to the regions of interest (Daugaard, Adler-Nissen, & Carstensen, 2010). Following this transformation, the separation was distinct, and a simple threshold was enough to separate meat from non-meat pixels. The resulting segmented image for each sample only contains the isolated part of the meat tissue as the main region of interest (ROI) was used for the extraction of spectral data used for the statistical modelling. For this step, the mean reflectance spectrum was calculated for each image by averaging the intensity of pixels within the ROI at each wavelength. Furthermore, the standard deviation of the pixels' intensity per wavelength was extracted. The resulting data consisted of 18 mean values and 18 standard deviations of the reflectance, as it was recorded by the camera for the pixels that were included in each image's ROI, and were further analyzed with various mathematical modelling methods.

2.4 Electronic nose

The volatile compounds chemical fingerprints of beef fillet samples were analysed using a gas sensor array system (LibraNose, Technobiochip, Napoli, Italy). The platform uses an array of 8 quartz crystal microbalance (QMB) non-selective sensors coated with different poly-pyrrole derivatives, synthesized at Technobiochip. Further details on the LibraNose instrumentation and mode of action can be found elsewhere (Baietto, Wilson, Bassi, & Ferrini, 2010). For each measurement, a beef fillet sample of 5 g was introduced inside a 100 ml volume glass jar and left at

room temperature (20°C ± 2°C) for 15 min to enhance desorption of volatile compounds from the meat into the headspace. The headspace was then pumped over the sensors of the electronic nose and the generated signal was continuously and in real time recorded and stored to a laptop computer.

2.5 High Performance Liquid Chromatography (HPLC)

The analysis was performed as described by (Skandamis & Nychas, 2001) equipped with a Model PU-980 Intelligent pump, a Model LG-980-02 ternary gradient unit pump and a MD-910 multi wavelength detector. The injection valve was connected with a 20 μL loop, whilst 50 μL of the sample was injected each time. The sample was eluted isocratically with a solution of 0.009 N H_2SO_4 (using HPLC grade solvent and ultra-pure water) through an Amminex HPX-87H column (300 × 7.8 mm, Bio-Rad Laboratories, Richmond, CA, USA) at a rate of 0.7 mL/min and oven temperature set at 65 °C. The software used for the collection and the processing of the spectra was the Jasco Chrompass Chromatography Data system v1.7.403.1. Spectral data were collected from 200 to 600 nm, however chromatogram integration was performed at 210 nm and the purity of the peaks was examined through the software using all spectral ranges. Solutions of oxalic, citric, malic, lactic, acetic, formic, tartaric, succinic and propionic acids (HPLC grade) were used as reference substances, analysed using the same programme and their spectra were compared with the samples for the identification of the peaks.

2.6 Gas Chromatography coupled to Mass Spectrometry (GC-MS)

GC-MS analyses was performed as described by (A. A. Argyri, Mallouchos, Panagou, & Nychas, 2015). Briefly, Agilent 7890A gas chromatograph coupled to Agilent 5973C mass spectrometer was deployed using helium as a carrier gas at a constant flow rate of 1 mL/min. The injection port was equipped with a liner (0.75 mm i.d.) suitable for SPME analysis. It was operated in splitless mode for 1 min at 250 °C. Separation of compounds was performed on an HP-5MS column (30 m, 0.25 mm i.d., 0.25 μm film thickness, Agilent). Oven temperature was maintained at 40 °C for 5 min, programmed at 4 °C/min to 150 °C and then it was raised to 250 °C with a rate of 30 °C/min and held for 5 min. The interface temperature was set at 280 °C. The mass spectrometer was operated in electron impact mode with the electron energy set at 70 eV and a scan range of 29–350 m/z (scan rate: 4.37 scans/s, gain factor: 1, resulting EM voltage: 1188 V). The temperature of MS source and quadrupole was set at 230 and 150 °C, respectively. Identification of the compounds was conducted as described elsewhere (A. A. Argyri et al., 2015).

2.7 Mathematical modelling

Three sets of minced beef samples stored under different packaging conditions; aerobic packaging

(AIR), modified atmosphere packaging (MAP) and mixed packaging storage (AIR + MAP) were analysed using different analytical and imaging platforms as shown is Figure 1, where output datasets were used to implement and test the performance implemented with "MeatReg". For each of the sets, data from 5 analytical platforms used in metabolomics in order to predict bacterial counts grown in 5 different mediums (a total of 50 combinations; 2 packaging systems x 5 platform x 5 species counts) was analysed. The analysis was performed using the open-source software environment R as it contains several resources and support to perform machine learning. For each dataset corresponding an analytical platform, a Principal Component Analysis (PCA) plot was generated for exploratory purposes. The models were generated using seven different methods for machine learning as previously outlined. In order to generate each of the models, the dataset was randomly split over 20 iterations into a training dataset, which contained 70% of the samples, and a testing dataset composed by the remaining samples. The training dataset was then used to build the model while the testing dataset was used to calculate the performance of such model.

Ordinary least squares regression (Wilkinson & Rogers, 1973), OLS-R, and stepwise linear regression (Friedman, Hastie, & Tibshirani, 2010), SL-R, are linear regression methods, with the difference that linear stepwise linear regression performs feature selection. These were performed using the "lm" function present in the package "stats" in R base. Principal component regression (Kendall, 1957), PC-R, and partial least squares regression (Höskuldsson, 1988), PLS-R, perform the regression models based on the reduction the number of variables to a few components that explain the variability within the samples. These methods were implemented using the "pcr" and "pls" functions with 3 components from the "pls" package (Mevik & Wehrens, 2007).
Random forest (Breiman, 2001) is a method based on ensemble learning that combines the predictions of several decision trees in order to provide a more accurate prediction than the individual trees on their own. It was implemented with 200 trees using the "randomForest" function from the "randomForest" package (Liaw & Wiener, 2002). For Support Vector Machines Regression (Boser, Guyon, & Vapnik, 1992), SVM-R, an optimal hyperplane that minimizes the distance to all the data points is built. As it is a parametric method, 3 parameters, cost of constraints violation, epsilon and gamma, were adjusted in order to optimize the model. To find the best values for these parameters grid search was carried out. The implementation was performed using the "svm" function from the "e1071" package (Dimitriadou, Hornik, Leisch, Meyer, & Weingessel, 2011) and the kernel selected was the radial basis kernel (RBF). The values for grid search were 0, 0.03, 0.06, 0.09, 0.20, 0.30, 0.60 and 0.90 for gamma; 0, 0.02, 0.04, 0.06, 0.10, 0.20, 0.30 and 0.60 for epsilon; and 1, 4, 7, 10, 13, 16, 19, 30, 50, 70, 90 and 110 for the cost. Finally, *k*-nearest neighbours (Silverman & Jones, 1989), kNN, predicts the value of an unknown sample based on the values of the *k* samples closer in distance. It was implemented using the function "knn.reg" from the "FNN" package (Beygelzimer, Kakadet, Langford, Arya, & Mount, 2013). The best *k* was selected using

grid search from k=4 to 10.

For the methods that required grid search the dataset was split into training and testing. The training dataset was used to perform a model for each one of the possibilities of the grid search. The performance of each one of the models was calculated using the testing dataset and the parameters, which provided the model with the lowest RMSE, were selected.

2.7.1 Models validation

The performance of the models was assessed in terms of how close the predicted values were to the observed. Quantification was performed by calculating the root-mean-square error (RMSE), the accuracy, the maximum difference between predicted and observed values (Δ_{max}), the bias factor (B_f) and the accuracy (A_f) factor (Baranyi, Pin, & Ross, 1999). The root mean-square error (Equation 1) quantifies the difference between predicted and observed value. If the difference is small, the RMSE is a positive number close to 0.

$$RMSE = \sqrt{\frac{\sum(predicted - observed)^2}{n}}$$

Equation 1. Root mean-square error formula.

The accuracy factor (Equation 2) is an indicator of the difference between observed and predicted values. The bias factor (Equation 3) indicates if there is systematic bias, providing information about how much the model over-predicts or under-predicts. Both factors are close to 1 if the model provides good predictions.

$$A_f = 10^{\left(\frac{\sum log\left(\left|\frac{predicted}{observed}\right|\right)}{n}\right)}$$

Equation 2. Accuracy factor formula.

$$B_f = 10^{\left(\frac{\sum log\left(\frac{predicted}{observed}\right)}{n}\right)}$$

Equation 3. Bias factor formula.

The accuracy is the percentage of samples correctly predicted (Equation 4). A sample is considered correctly predicted if the difference between predicted and observed values is less than 1 log (TVC) (Mohareb, Papadopoulou, Panagou, Nychas, & Bessant, 2016) as it is the equivalent to 1 dilution.

$$Accuracy = \frac{samples\ correctly\ predicted}{overall\ number\ of\ samples} \times 100$$

Equation 4. Accuracy formula.

The maximum difference between predicted and observed values was calculated as well. Even if it is not a statistical parameter, it is informative as it shows the worst prediction. If the highest difference between predicted and observed is below or close to 1, the model is providing good predictions but if it is higher it means that the worst prediction deviates too much from the observed values.

Validation is a crucial part of the analysis. To assess the performance of the machine learning algorithms, Monte Carlo cross validation (Qing-Song Xu & Liang, 2001) was used. For K=20, the samples from each dataset were randomly distributed into training and testing datasets in 20 different splits. Then, the performance is calculated as an average of the performance of the 20 models. Performance plots for each scenario and machine learning method applied were generated to observe how the calculations of accuracy vary depending on the value of N.

2.7.2 Ranking and visualization

Machine learning methods were ranked according to the RMSE for each scenario. The first machine learning method in the ranking were used to generate a summary plot in which the intersection of analytical platform and the medium used to grow microorganisms indicates the performance of the best machine learning technique (Figure 2). Light green tones represent that at least one of the machine learning methods applied provided a good prediction of the bacterial count whilst red tones illustrate that no machine learning methods provided a good performance.

3. Results

3.1 Microbial Association

The microbial association of minced meat comprises of Total Viable Counts (initial counts; 5.65 log cfu/g), Pseudomonads (4.13 log cfu/g), *Brochothrix thermosphacta* (4.24 log cfu/g), Lactic acid bacteria (4.33 log cfu/g), Enterobacteriaceae (1.92 log cfu/g) and yeasts (4.80 log cfu/g). Tables 1 and 2 summarize kinetic parameters of the member of microbial association following data fitting using the Baranyi and Roberts' model (Baranyi & Roberts, 1994). Lag phase was observed only at 0°C, while there was an increase in the maximum specific growth rate (μ_{max}) of various groups correlated to the increase of the storage temperature regardless the packaging type. The influence of the packaging type was also evident. For example, TVC μ_{max} followed the increase order MAP <

AIR, regardless the storage temperature.

3.2 Performance of Machine-learning models in data acquired from different metabolomics fingerprint profiles

Ranking of the tested machine learning models for AIR and MAP, and AIR+MAP is provided in the supplemented Table S1, S2 and S3 respectively; while the best performed models are indicated in Figures 2. It should be noted that no data pre-treatment was performed prior to analysis. Taking RMSE as a measure of model's performance validation, the data analysis revealed that these values ranged from 0.370 to 1.321 across different best machine learning methods to predict each of the species counts for AIR, MAP, conditions (Figures 2a & 2b). Furthermore, for mixed storage conditions (AIR + MAP) showed comparable results to storage-specific models; with RMSE values ranged from 0.388 to 1.343 (Figure 2c).

Under AIR packaging storage (Figure 2a), the best overall prediction outcomes were obtained from data derived with HPLC and MSI. HPLC achieved 100% prediction accuracy for Lactobacilli (medium MRS, RMSE = 0.466) when combined with RF-R, 93.9% and 92.6% when combined with kNN-R for total viable count (medium PCA, RMSE = 0.508) and *B. thermosphacta* (medium STAA, RMSE = 0.564) respectively. eNose achieved 86.6% for Enterobacteriaceae (medium VRBG, RMSE = 0.724). Pseudomonads (medium CFC), were best predicted using MSI combined with PLS-R (RMSE = 0.853). The ranking of the entire machine learning methods suite applied to beef stored under aerobic packaging conditions for each combination is available in the Supplementary Materials – Table 1S.

For MAP storage (Figure 2b), the RMSE ranged from 0.370 to 0.866 across different machine learning methods to predict each of the species counts. RF-R provided the best predictions in most cases. For electronic nose, the RMSE ranged from 0.431 for Lactobacilli (medium MRS) to 0.654 for Pseudomonads. For FT-IR, the RMSE ranged from 0.574 for Lactobacilli to 0.866 for Enterobacteriaceae. For GC-MS data, the RMSE ranged from 0.426 for Lactobacilli to 0.621 for *B. Thermosphacta* and the accuracy is above 89% for all bacterial counts. On the other hand, *k*NN-R and RF-R provided the best results for HPLC and MSI. For HPLC data, the RMSE ranged from 0.378 and 100% accuracy using RF-R for Lactobacilli to 0.668 and 91.5% accuracy using kNN-R for Enterobacteriaceae. For MSI measurements, the RMSE ranged from 0.506 and 94.1% accuracy for Lactobacilli, to 0.822 and 73.0% accuracy for *B. thermosphacta* (medium STAA), both using RF-R. The best predictions for Enterobacteriaceae, were obtained using GC-MS, with an RMSE of 0.480 and an accuracy of 93.9%. For Lactobacilli (medium MRS) the highest RMSE was obtained using HPLC in tandem with RF-R. For the total viable counts, GC-Ms and eNose in tandem with SV-R and

RF-R provided predictions with an RMSE of 0.561 and 0.564 respectively. The best predictions for pseudomonads, were obtained using GC-MS with *k*NN-R, with an RMSE of 0.471 and 96.0% accuracy. Finally, the best predictions for *B. thermosphacta*, was obtained using eNose combined with RF-R; achieving RMSE of 0.566 and an accuracy of 92.8%. The ranking of the entire machine learning methods applied to beef stored under modified atmosphere packaging conditions for each combination is available in the Supplementary Materials – Table S2.

In the case of mixed packaging system, the input dataset consisted of a random mixture of AIR and MAP within a single dataset and the packaging type was not included as a model input variable, in order to evaluate the performance of the developed pipeline in achieving a good prediction accuracy regardless of the packaging system applied. For this experiment, GCMS coupled with RF-R achieved the best prediction accuracy throughout all growth media, followed by HPLC coupled with KNN-R, while data derived from FT-IR had the worst prediction performance (Figure 2c). The best predictions for Enterobactriaceae (VRBG), *B, thermosphacta* (STAA), Lactobacilli (MRS), Pseudomonads (CFC) and total viable counts (PCA), were all achieved using random forest at an RMSE at 0.558, 0.568, 0.368, 0.66, and 0.471 respectively. The ranking of the entire machine learning methods applied to beef stored under mixed packaging conditions for each combination is available in the Supplementary Materials – Table S3.

3.3 Automation of the ML ranking system *via* an interactive web-based platform

In order to maximise benefit to the food research community from the developed protocol, the entire process of models development, optimisation and validation pipeline has been deployed within an tailored R library package, "MeatReg" deployed within a Web application. The purpose of this package is to automate the procedure of creating prediction models based on data from analytical platforms and microbiological counts in CSV format. Figure 3 shows the general workflow of the analysis. The program loops through machine learning methods, datasets from analytical platforms and microbial data selected (highlighted in red) in order to calculate the performance of the models for all possible combinations. For each dataset from an analytical platform, a Principle Component analysis (PCA) plot is provided (highlighted in blue). For each combination of machine learning method, data from the analytical platform and microbial counts are combined and divided into training and testing datasets, models are created and statistical parameters are calculated K times and averaged (highlighted in green).

The "MeatReg" package is available to use *via* the sorfML web application platform

(http://elvis.cranfield.ac.uk/SORF/), allowing users to perform regression analysis online (Figure 4). For a specific dataset already uploaded into the sorfML database, users select "Regression analysis" and they are able to select the analytical platforms, bacterial growth mediums and machine learning methods to include in their analysis. Additionally, other parameters can be modified, such as select a pre-treatment, convert bacterial counts to logarithmic scale if they were not already converted, modify the number of trees for random forest or modify the parameters for grid search for *k*NN-R and SVM-R. When the user clicks "Submit" it passes all the selected options to a Python script as command line arguments. A Python script generates and executes an R script to perform the regression analysis using the "MeatReg" package, gathers the information into a LaTeX file and it is presented to the user as a PDF report containing the summary plot, the ranking for machine learning methods, performance plots, PCA plots and the R code to run the analysis.

4. Discussion

Data mining derived from food analyses using rapid analytical techniques based on limited or non-destructive / non-invasive sensors is a growing area in the food sector (George-John E. Nychas et al., 2016). Indeed, the implementation of these rapid techniques coupled with data analysis methods has given promising results in several food products (Ropodi et al., 2016). To our knowledge, "MeatReg" is the first attempt to compare machine-learning methods to determine the suitability of several analytical instruments in predicting microbiological quality, including spoilage of foods and more specifically meat, once a suitable dataset is uploaded to the platform. With "MeatReg", it is possible to visualize how data derived from different analytical instruments are suitable to predict not only the general microbiological quality of meat but also determine the contribution of individual bacterial species in the food spoilage process, using a combination of seven regression methods. All the machine-learning methods applied are ranked according to their performance, in order to provide information about how each of them performed individually, and in relation to one another. This multi 'dimensional' analyses could be used to evaluate (i) spoilage, regardless the storage conditions e.g. temperature, packaging etc, and (ii) the feasibility of using in food sector specific non-destructive/non-invasive instruments/sensors.

For example, Pseudomonads and Enterobacteriaceae counts are well known to be related with meat spoilage under aerobic packaging conditions (Nychas et al. 2008). In this case the most suitable sensor is MSI (Figure 2) and it is quite convenient due to its portability, i.e. it can be placed on line within the food industry production line, and it is able to provide very fast and accurate results. On the other hand, although HPLC and GC-MS usually provide accurate prediction, they are not portable and cannot be deployed within the production chain, however instead can be used by competent authorities in their analytical laboratories and to conclude regarding the

microbiological quality as requested by EFSA's legislation (EFSA 2005). Electronic nose and FT-IR are also convenient methods and although do provide good prediction >86% for TVC, the lack of accuracy in the case of pseudomonads in meat stored under aerobic packaging, can be considered as a drawback if these organisms is requested to be used as quality index. In general, this pipeline helps to decide which analytical technique to be used by whom (e.g. meat industry, food authority) and where (on in at line) to predict all types of bacteria of interest according to the food product being tested, and which machine learning algorithms will provide accurate predictions of their bacterial counts without having to invest in numerous and usually expensive platforms.

Models developed using "MeatReg" are validated using the Monte Carlo cross validation as opposed to the more widely applied methods such as leave-one-out cross-validation (LOOCV). Though LOOCV is a very popular validation method (Argyri et al., 2010; Mohareb et al., 2016; Wang, Wang, Liu, & Liu, 2012), it has been shown that using Monte Carlo validation can avoid an unnecessary large model and decrease the risk of overfitting (Qing-Song Xu, Liang, & Du, 2004). Other common strategy is to randomly split the original dataset into training and testing datasets (Anthoula A. Argyri et al., 2013; Mataragas, Skandamis, Nychas, & Drosinos, 2007; Panagou, Mohareb, Argyri, Bessant, & Nychas, 2011; Rajamäki et al., 2006), but the calculation of the statistical parameters highly depends on which samples fell into each dataset. However, with Monte-Carlo validation the original datasets are split into training and testing datasets K number of times (as specified by the user). Training datasets are used to build the statistical model and testing datasets are used to test it. K models are built and tested and the statistical parameters for all of them are averaged. In contrast to splitting the dataset into training and testing datasets only once, this method is more reliable to assess the performance as it does not depend that much in which samples randomly fall into the training and testing subsets (Mohareb et al., 2016). Moreover, a particular effort has been placed in providing a comprehensive and easy to understand report. A heatmap that ranks all included instruments/sensors, algorithms and species counts using a red-green colour-ramp; where red represents unsuitability and green represents suitability (Figure 2). In this way, it became possible to compare and rank several possible scenarios within a single plot. Detailed information about each scenario is also provided within the generated report. For each combination of analytical platform and bacteria, all the tested machine-learning methods are ranked and statistics are provided for all of them. Performance plots are shown as well. Principal Component Analysis (PCA) plots are provided hereafter for all the data from analytical platform. PCA is a very popular unsupervised method for exploratory purposes and it is worth checking the plots for the experimental data in order to detect if it behaves as expected or to find possible outliers as well as giving insights about the potential for a given dataset to be used to optimise an accurate model; through the identification of correlation between samples . Finally, the R code applied is shown in the report as well in order to keep track of the selected parameters.

As far as the models performance for each of the applied analytical instruments is concerned, MSI, HPLC and GC-MS provide good predictions for all types of counts. The predictions regarding Enterobacteriaceae and *B. thermosphacta* are limited for FT-IR and electronic nose respectively while both failed to predict pseudomonads counts under aerobic packaging. Although most of the research has been performed using PLS-R or support vector machines, our results show that other machine learning methods such as kNN-R and RF-R could provide higher prediction accuracy. Regarding the performance of FTIR, it should be mentioned that from this instrument a huge amount of data are derived in comparison with e.g. HPLC (18 peaks) enose (12-20 peaks) MSI (18 wavelength) GC/MS (130 peaks). This should be taken into account if data mining should be applied before the implementation of 'MeatReg'.

Supplementary Tables S1&S2 show that HPLC data is very suitable to predict bacterial counts as the regression methods provide good results in most cases. MSI and GC-MS data are very suitable as well in tandem with kNN-R, SVM-R, RF-R, PLS-R or PC-R. However, GC-MS is not likely to be introduced in the industry due to its high cost, high maintenance and size, compared to more portable MSI platforms such as Videometer.

Aside from the difficulties mentioned before regarding the predictions of certain bacterial counts, electronic nose data is suitable to predict bacterial counts using most methods for modified atmosphere packaging but the appropriate methods for aerobic packaging. For the total bacterial counts under aerobic packaging support vector regression is recommended. For lactic acid bacteria *(*MRS*)* all the methods provided accurate predictions while for Enterobacteria (VRBG) SVM-R, linear regression and stepwise linear regression provided the best results. For FT-IR data, aside from pseudomonads and Enterobacteria under aerobic conditions, RF-R provided the best predictions in all cases. Additionally, the results rank PLS-R, *k*NN-R, PC-R and SVM-R next, which work generally fine for all methods with some exceptions.

5. Conclusions

An interactive online platform has been developed to determine the most suitable regression machine learning method that provides the best prediction accuracy of different types of microorganisms involved in beef spoilage using data from five different analytical methods. A thorough approach in selecting the most appropriate strategy for validation of the developed models has been implemented, by providing several methods for calculating the difference between the predicted and the measured values, such as RMSE, the accuracy, the maximum difference between predicted and observed values (Δ_{max}), the bias factor (B_f) and the accuracy (A_f) factor. The

models development, optimisation, and validation is taking place on the server-side via a user-friendly interface, therefore abolishing the need for installing specialised statistical software at the client-side (i.e. the user computer). The analysis output is provided through a comprehensive PDF report which include all optimisation and validation results for each chosen analytical platform and ML method, as well as all the parameters used during the analysis. As a future extension of the developed platform, we aim to provide a classification ML suite to complement the regression suite presented hereby, to predict spoilage-related discrete values, such as freshness profiles and sensory. Additionally, we aim to extend the compatibility of the data repository to support additional platform commonly applied in food research such as Raman and MALDI-ToF Mass Spectroscopy. Data collected from either FT-IR, image analysis, HPLC or GC-MS combined with appropriate machine learning strategies (for example partial least squares regression, artificial neural networks) could become an interesting tool to monitor food spoilage/freshness through the measurement of biochemical changes occurring in food substrate, without the history of sample to be known (e.g. temperature of storage, the initial contamination, pH).

A primary goal of data visualization is to communicate information clearly and efficiently to users via the information graphics selected, such as tables and charts. Effective visualization helps users in analysing and reasoning about data and evidence. It makes complex data more accessible, understandable and usable. Data visualization is both an art and a science.

Authors Contribution

Lopez developed the "MeatReg" library and the machine learning models described in this work. Ropodi, Pavlidis, Fotopolou, and Gkousari performed the laboratory work, microbiological and instrumental analysis. Peyrodie contributed in developing the sorfML platform and integrated the MeatReg library with the online platform. Panagou managed the data analysis. Nychas designed the experiment and led the wet-lab work. Mohareb managed the implementation the sorfML platform the machine learning work, and led the article preparation together with Nychas. All authors contributed equally to article preparation and approved the final article.

Declaration of interest

There are no conflicts of interest among authors. All authors have met the criteria for authorship, and have read and approved the final version of the manuscript. This work is not under consideration elsewhere.

References

Argyri, A., A., Panagou, Z., E., Tarantilis, A., P., . . . E., G.-J. (2010). Rapid qualitative and quantitative detection of beef fillets spoilage based on Fourier transform infrared spectroscopy data and artificial neural networks. *145*(1), 9.

Argyri, A. A., Jarvis, R. M., Wedge, D., Xu, Y., Panagou, E. Z., Goodacre, R., & Nychas, G.-J. E. (2013). A comparison of Raman and FT-IR spectroscopy for the prediction of meat spoilage. *Food Control, 29*(2), 461-470. doi: http://dx.doi.org/10.1016/j.foodcont.2012.05.040

Argyri, A. A., Mallouchos, A., Panagou, E. Z., & Nychas, G. J. (2015). The dynamics of the HS/SPME-GC/MS as a tool to assess the spoilage of minced beef stored under different packaging and temperature conditions. *Int J Food Microbiol, 193*, 51-58. doi: 10.1016/j.ijfoodmicro.2014.09.020

Baietto, M., Wilson, A. D., Bassi, D., & Ferrini, F. (2010). Evaluation of three electronic noses for detecting incipient wood decay. *Sensors (Basel), 10*(2), 1062-1092. doi: 10.3390/s100201062

Baranyi, J., Pin, C., & Ross, T. (1999). Validating and comparing predictive models. *Int J Food Microbiol, 48*(3), 159-166.

Baranyi, J., & Roberts, T. A. (1994). A dynamic approach to predicting bacterial growth in food. *Int J Food Microbiol, 23*(3-4), 277-294.

Beygelzimer, A., Kakadet, S., Langford, J., Arya, S., & Mount, D. (2013). FNN: Fast Nearest Neighbor Search Algorithms and Applications. (Version 1.1) [R package]. https://cran.r-project.org/package=FNN.

Boser, B. E., Guyon, I. M., & Vapnik, V. N. (1992). A Training Algorithm for Optimal Margin Classifiers %U http://doi.acm.org/10.1145/130385.130401. In D. Haussler (Ed.), *Proceedings of the 5th Annual Workshop on Computational Learning Theory (COLT'92)* (pp. 144--152). Pittsburgh, PA, USA: ACM Press.

Breiman, L. (2001). Random Forests. *Mach. Learn., 45*(1), 5-32. doi: 10.1023/a:1010933404324

Commission regulation (EC) No 2073/2005 on microbiological criteria for foodstuffs, OJ L C.F.R. (2005).

Daugaard, S. B., Adler-Nissen, J., & Carstensen, J. M. (2010). New vision technology for multidimensional quality monitoring of continuous frying of meat. *Food Control, 21*(5), 626-632. doi: https://doi.org/10.1016/j.foodcont.2009.09.007

Dimitriadou, E., Hornik, K., Leisch, F., Meyer, D., & Weingessel, A. (2011). *e1071: Misc Functions of the Department of Statistics (e1071), TU Wien*.

Doulgeraki, A. I., Ercolini, D., Villani, F., & Nychas, G. J. (2012). Spoilage microbiota associated to the storage of raw meat in different conditions. *Int J Food Microbiol, 157*(2), 130-141. doi: 10.1016/j.ijfoodmicro.2012.05.020

Friedman, J., Hastie, T., & Tibshirani, R. (2010). Regularization Paths for Generalized Linear Models via Coordinate Descent. *J Stat Softw, 33*(1), 1-22.

Höskuldsson, A. (1988). PLS regression methods. *Journal of Chemometrics, 2*(3), 211-228. doi: 10.1002/cem.1180020306

Kendall, M. G. (1957). *A Course in Multivariate Analysis*: Griffin.

Koutsoumanis, K. (2009). Modeling food spoilage in microbial risk assessment. *J Food Prot, 72*(2), 425-427.

Liaw, A., & Wiener, M. (2002). Classification and Regression by randomForest. *R News, 2*(3), 18-22. doi: citeulike-article-id:1121494

Mataragas, M., Skandamis, P., Nychas, G. J., & Drosinos, E. H. (2007). Modeling and predicting

spoilage of cooked, cured meat products by multivariate analysis. *Meat Sci, 77*(3), 348-356. doi: 10.1016/j.meatsci.2007.03.023

Mevik, B.-H., & Wehrens, R. (2007). The pls Package: Principal Component and Partial Least Squares Regression in R. *Journal of Statistical Software, 18*(2), 1-24.

Mohareb, F., Iriondo, M., Doulgeraki, A. I., Van Hoek, A., Aarts, H., Cauchi, M., & Nychas, G.-J. E. (2015). Identification of meat spoilage gene biomarkers in Pseudomonas putida using gene profiling. *Food Control, 57*, 152-160. doi: http://dx.doi.org/10.1016/j.foodcont.2015.04.007

Mohareb, F., Papadopoulou, O., Panagou, E., Nychas, G.-J., & Bessant, C. (2016). Ensemble-based support vector machine classifiers as an efficient tool for quality assessment of beef fillets from electronic nose data. [10.1039/C6AY00147E]. *Analytical Methods, 8*(18), 3711-3721. doi: 10.1039/C6AY00147E

Nychas, G.-J. E., Marshall, D., & Sofos, J. (2007). *Chapter 6; Meat Poultry and Seafood*. Eds Doyle: Beuchat and Montville ASM Press.

Nychas, G.-J. E., Panagou, E. Z., & Mohareb, F. (2016). Novel approaches for food safety management and communication. *Current Opinion in Food Science, 12*, 13-20. doi: http://dx.doi.org/10.1016/j.cofs.2016.06.005

Nychas, G. J., Skandamis, P. N., Tassou, C. C., & Koutsoumanis, K. P. (2008). Meat spoilage during distribution. *Meat Sci, 78*(1-2), 77-89. doi: 10.1016/j.meatsci.2007.06.020

Nychas, G. J. E., Skandamis, P. N., Tassou, C. C., & Koutsoumanis, K. P. (2008). Meat spoilage during distribution. *Meat Sci, 78*(1-2), 77-89. doi: Doi 10.1016/J.Meatsei.2007.06.020

Panagou, E. Z., Mohareb, F. R., Argyri, A. A., Bessant, C. M., & Nychas, G. J. (2011). A comparison of artificial neural networks and partial least squares modelling for the rapid detection of the microbial spoilage of beef fillets based on Fourier transform infrared spectral fingerprints. *Food Microbiol, 28*(4), 782-790. doi: 10.1016/j.fm.2010.05.014

Panagou, E. Z., Papadopoulou, O., Carstensen, J. M., & Nychas, G. J. (2014). Potential of multispectral imaging technology for rapid and non-destructive determination of the microbiological quality of beef filets during aerobic storage. *Int J Food Microbiol, 174*, 1-11. doi: 10.1016/j.ijfoodmicro.2013.12.026

Rajamäki, T., Alakomi, H.-L., Ritvanen, T., Skyttä, E., Smolander, M., & Ahvenainen, R. (2006). Application of an electronic nose for quality assessment of modified atmosphere packaged poultry meat. *Food Control, 17*(1), 5-13. doi: http://dx.doi.org/10.1016/j.foodcont.2004.08.002

Ropodi, A. I., Panagou, E. Z., & Nychas, G. J. E. (2016). Data mining derived from food analyses using non-invasive/non-destructive analytical techniques; determination of food authenticity, quality & safety in tandem with computer science disciplines. *Trends in Food Science & Technology, 50*, 11-25. doi: http://dx.doi.org/10.1016/j.tifs.2016.01.011

Ropodi, A. I., Pavlidis, D. E., Mohareb, F., Panagou, E. Z., & Nychas, G. J. E. (2015). Multispectral image analysis approach to detect adulteration of beef and pork in raw meats. *Food Research International, 67*, 12-18. doi: https://doi.org/10.1016/j.foodres.2014.10.032

Silverman, B. W., & Jones, M. C. (1989). E. Fix and J.L. Hodges(1951): an important contribution to nonparametric discriminant analysis and density estimation. *International Statistical Review, 57*(3), 233-247.

Skandamis, P. N., & Nychas, G. J. (2001). Effect of oregano essential oil on microbiological and physico-chemical attributes of minced meat stored in air and modified atmospheres. *J Appl Microbiol, 91*(6), 1011-1022.

Van Wezemael, L., Verbeke, W., de Barcellos, M. D., Scholderer, J., & Perez-Cueto, F. (2010). Consumer perceptions of beef healthiness: results from a qualitative study in four European

countries. *BMC Public Health, 10*, 342. doi: 10.1186/1471-2458-10-342

Wang, D., Wang, X., Liu, T., & Liu, Y. (2012). Prediction of total viable counts on chilled pork using an electronic nose combined with support vector machine. *Meat Science, 90*(2), 373-377. doi: http://dx.doi.org/10.1016/j.meatsci.2011.07.025

Wilkinson, G. N., & Rogers, C. E. (1973). Symbolic descriptions of factorial models for analysis of variance. *Applied Statistics, 22*, 392-399.

Xu, Q.-S., & Liang, Y.-Z. (2001). Monte Carlo cross validation. *Chemometrics and Intelligent Laboratory Systems, 56*(1), 1-11. doi: http://dx.doi.org/10.1016/S0169-7439(00)00122-2

Xu, Q.-S., Liang, Y.-Z., & Du, Y.-P. (2004). Monte Carlo cross-validation for selecting a model and estimating the prediction error in multivariate calibration. *Chemometrics, 18*, 112-120.

Table 1.: Kinetic parameters (following data fitting, calculated with the model of Baranyi and Robert 1994) of the members of microbial association of minced meat [Total Viable Counts; initial counts 5.48 ± 0.2 log cfu, pseudomonads spp.] stored under aerobic conditions at 4 and 10°C, as function of the storage temperature and the packaging type.

	Microbial group	μ_{max} [a]	Lag phase (h)	y0 (log cfu/g)	yEnd (log cfu/g)	Standard error of fitting	Adjusted R^2, statistics of the fitting
4°C	Total Viable count	0.057	35.78	5.60	9.38	0.180	0.988
	Pseudomonads	0.080	18.93	4.12	9.32	0.200	0.990
	B. thermosphacta	0.075	21.03	4.24	7.82	0.160	0.986
	Lactic acid bacteria	0.036	64.73	4.30	6.35	0.205	0.907
	Enterobacteriaceae	0.041	45.78	1.75	nd	0.203	0.967
	Yeasts-molds	0.025	17.58	4.70	6.67	0.159	0.943
10°C	Total Viable count	0.126	nd	5.50	10.90	0.218	0.964
	Pseudomonads	0.124	nd	5.72	10.78	0.300	0.931
	B. thermosphacta	0.143	nd	4.36	7.73	0.237	0.958
	Lactic acid bacteria	0.126	nd	3.40	7.15	0.200	0.957
	Enterobacteriaceae	0.053	nd	0.85	nd	0.463	0.658
	Yeasts-molds	0.103	nd	3.82	5.7	0.306	0.806

[a] Maximum specific growth rate h_1.

Table 2.: Kinetic parameters (following data fitting, calculated with the model of Baranyi and Robert 1994) of the members of microbial association of minced meat [Total Viable Counts; initial counts 5.48 ± 0.2 log cfu, pseudomonads] stored under modified atmosphere conditions at 4 and 10°C, as function of the storage temperature and the packaging type

	Microbial group	μ_{max} [a]	Lag phase (h)	y0 (log cfu/g)	yEnd (log cfu/g)	Standard error of fitting	Adjusted R^2, statistics of the fitting
	Microbial group	μ_{max}	lag	y0	yEnd	Se(fit)	R^2_stat
4°C	Total Viable count	0.040	74.23	5.86	7.81	0.166	0.963
	Pseudomonads	0.050	60.75	4.09	7.35	0.287	0.917
	B. thermosphacta	0.070	36.35	4.51	7.78	0.227	0.965
	Lactic acid bacteria	0.021	62.29	4.49	6.05	0.105	0.972
	Enterobacteriaceae	0.052	112.35	2.01	3.29	0.144	0.944
	Yeasts-molds	0.024	25.60	4.83	6.07	0.155	0.871
10°C	Total Viable count	0.116	nd	5.41	7.99	0.223	0.935
	Pseudomonads	0.075	nd	5.40	7.10	0.420	0.892
	B. thermosphacta	0.100	nd	4.20	7.06	0.247	0.949
	Lactic acid bacteria	0.094	nd	3.29	6.35	0.177	0.979
	Enterobacteriaceae	0.019	nd	0.25	nd	0.326	0.419
	Yeasts-molds	0.042	nd	3.87	nd	0.654	0.796

List of Figures

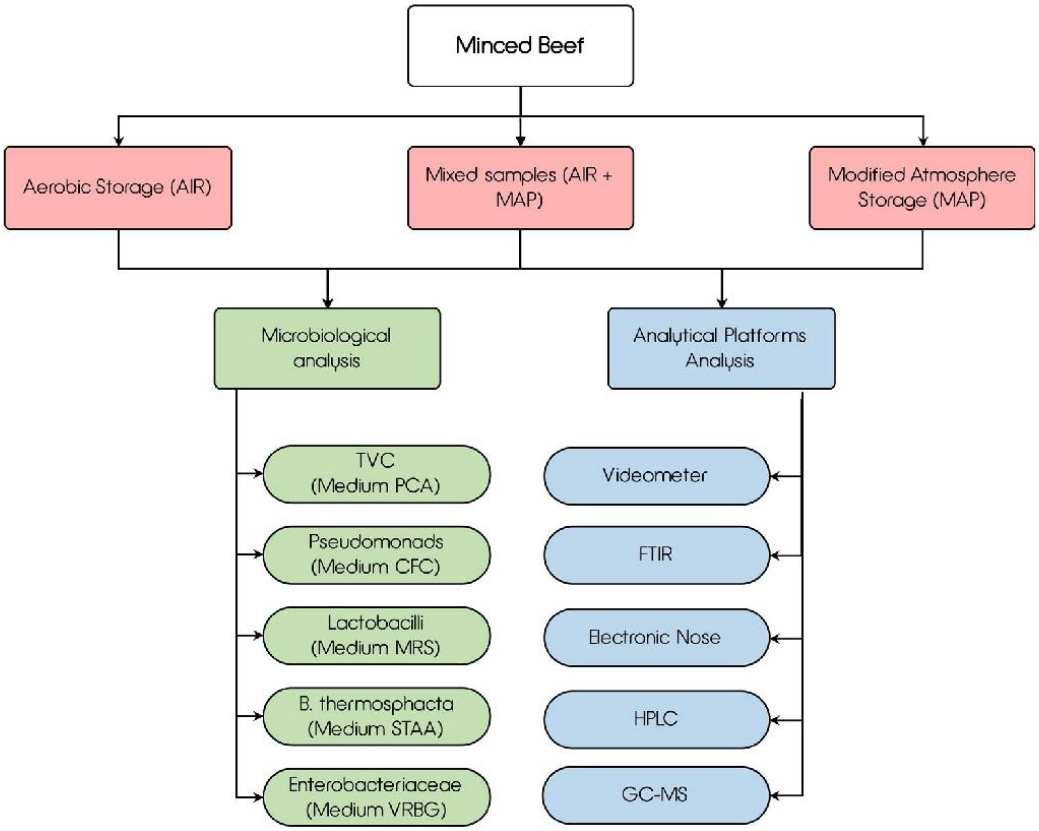


Figure 1

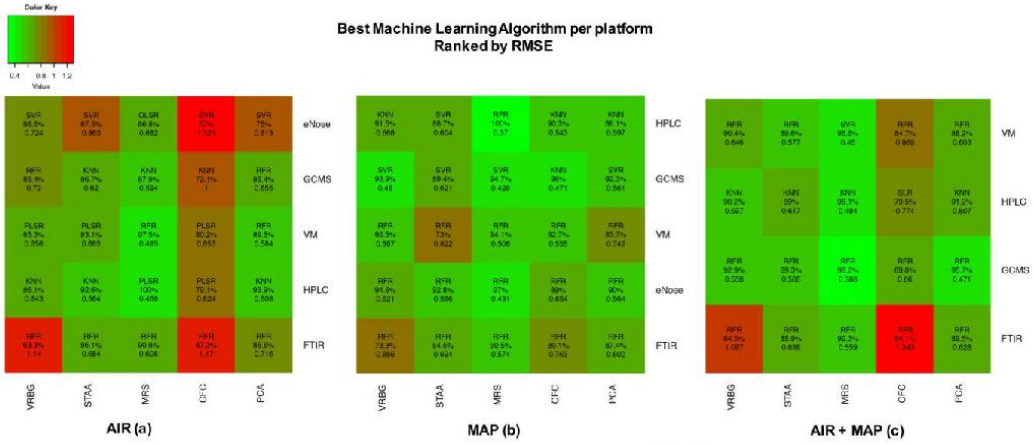


Figure 2

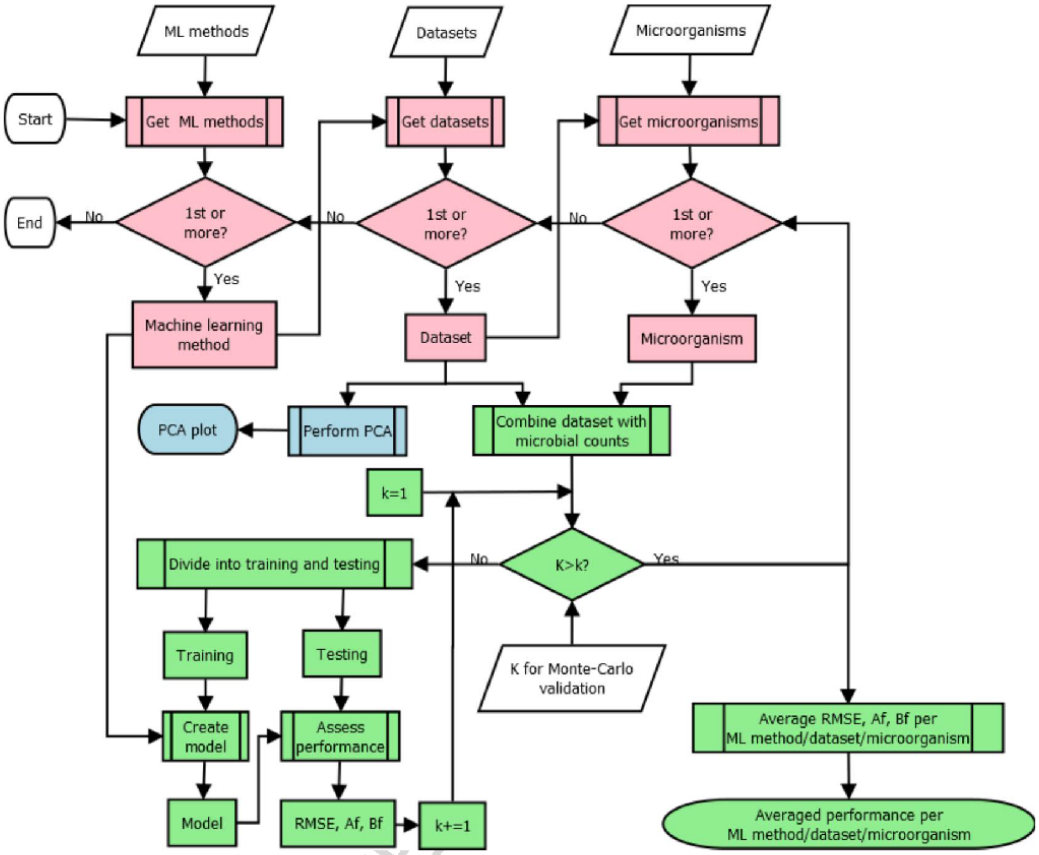


Figure 3

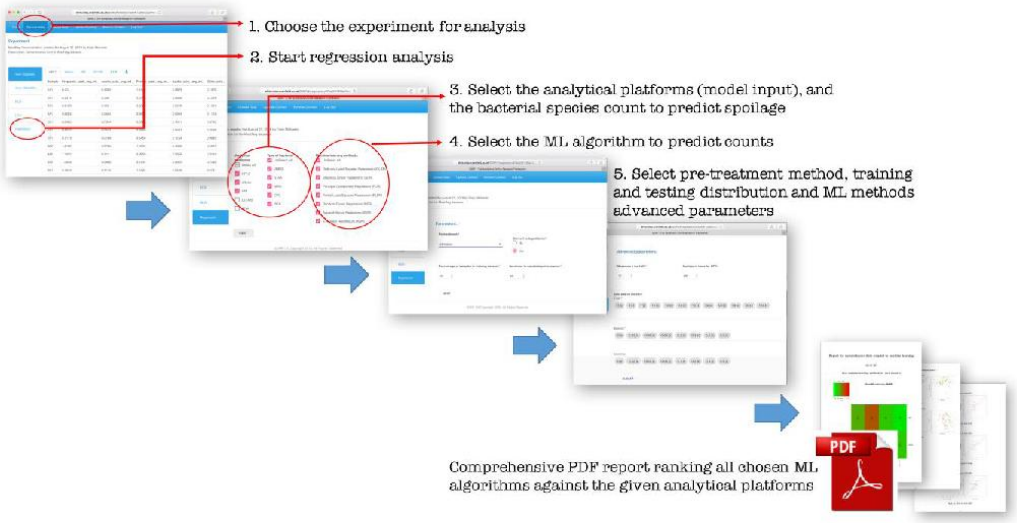


Figure 4

Supplementary material for Section 4.1 -S1

Report for metabolomics data coupled to machine learning

February 25, 2017

Best machine learning methods for each scenario

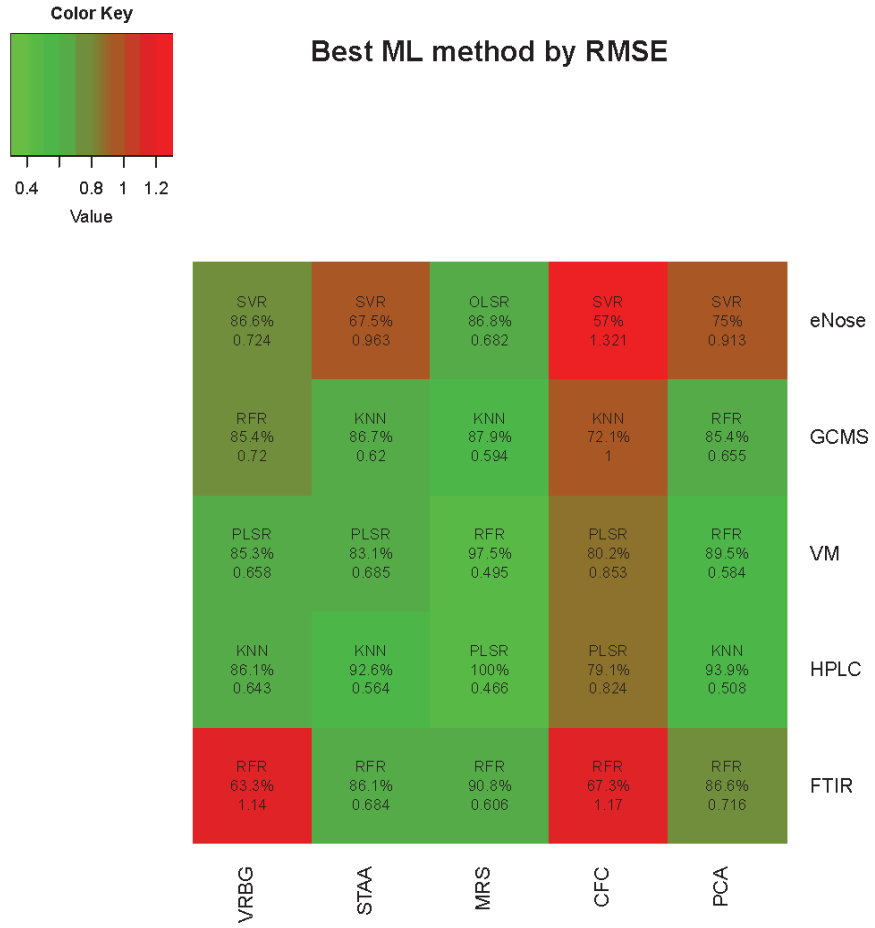


Figure 1: Best machine learning method by RMSE

1

Table 1: Ranking of machine methods by RMSE in first place and accuracy in second place.

	rank #1	rank #2	rank #3	rank #4	rank #5	rank #6	rank #7
eNose VREG	SVR (c=110, γ=0.06, ε=0.02) RMSE: 0.724 Acc: 86.6% Δ_{max}: 1.97 A_f: 1.13 B_f: 1	OLSR RMSE: 0.781 Acc: 78.7% Δ_{max}: 1.84 A_f: 1.15 B_f: 1.01	SLR RMSE: 0.876 Acc: 81.4% Δ_{max}: 2.26 A_f: NaN B_f: NaN	PLSR RMSE: 0.899 Acc: 70.9% Δ_{max}: 1.79 A_f: 1.18 B_f: 1.01	RFR RMSE: 0.971 Acc: 68.3% Δ_{max}: 2.14 A_f: 1.19 B_f: 1.01	PCR RMSE: 1.013 Acc: 67.4% Δ_{max}: 2.23 A_f: 1.2 B_f: 1.01	KNN (k=8) RMSE: 1.088 Acc: 63.7% Δ_{max}: 2.45 A_f: 1.2 B_f: 0.98
eNose STAA	SVR (c=70, γ=0.3, ε=0.04) RMSE: 0.963 Acc: 67.5% Δ_{max}: 2.14 A_f: 1.16 B_f: 1	RFR RMSE: 0.98 Acc: 66.3% Δ_{max}: 2.03 A_f: 1.16 B_f: 1.01	OLSR RMSE: 1.001 Acc: 66.8% Δ_{max}: 2.15 A_f: 1.16 B_f: 1	KNN (k=9) RMSE: 1.057 Acc: 58.5% Δ_{max}: 2.21 A_f: 1.18 B_f: 1.02	PLSR RMSE: 1.1 Acc: 57.2% Δ_{max}: 2.02 A_f: 1.18 B_f: 1.02	SLR RMSE: 1.106 Acc: 66.6% Δ_{max}: 2.68 A_f: NaN B_f: NaN	PCR RMSE: 1.14 Acc: 50.1% Δ_{max}: 2.13 A_f: 1.2 B_f: 1.02
eNose MRS	OLSR RMSE: 0.682 Acc: 86.8% Δ_{max}: 1.68 A_f: 1.1 B_f: 1	SLR RMSE: 0.722 Acc: 89.4% Δ_{max}: 1.96 A_f: NaN B_f: NaN	PLSR RMSE: 0.73 Acc: 83.4% Δ_{max}: 1.42 A_f: 1.12 B_f: 1.01	SVR (c=70, γ=0, ε=0.04) RMSE: 0.755 Acc: 83.4% Δ_{max}: 1.91 A_f: 1.11 B_f: 0.99	RFR RMSE: 0.756 Acc: 79.7% Δ_{max}: 1.56 A_f: 1.12 B_f: 1.01	KNN (k=9) RMSE: 0.785 Acc: 79.7% Δ_{max}: 1.57 A_f: 1.12 B_f: 1	PCR RMSE: 0.796 Acc: 75.5% Δ_{max}: 1.49 A_f: 1.13 B_f: 1
eNose CFC	SVR (c=110, γ=0.3, ε=0.02) RMSE: 1.321 Acc: 57% Δ_{max}: 3.16 A_f: 1.2 B_f: 1.01	OLSR RMSE: 1.373 Acc: 50.8% Δ_{max}: 3.22 A_f: 1.22 B_f: 1.01	RFR RMSE: 1.474 Acc: 35.2% Δ_{max}: 2.88 A_f: 1.25 B_f: 1.03	PLSR RMSE: 1.543 Acc: 38.1% Δ_{max}: 2.85 A_f: 1.26 B_f: 1.04	KNN (k=8) RMSE: 1.587 Acc: 39.8% Δ_{max}: 3.21 A_f: 1.26 B_f: 1	SLR RMSE: 1.595 Acc: 50.9% Δ_{max}: 4.23 A_f: NaN B_f: NaN	PCR RMSE: 1.649 Acc: 30.8% Δ_{max}: 3.11 A_f: 1.28 B_f: 1.04
eNose PCA	SVR (c=90, γ=0.09, ε=0.04) RMSE: 0.913 Acc: 75% Δ_{max}: 2.27 A_f: 1.11 B_f: 1	OLSR RMSE: 0.972 Acc: 67.9% Δ_{max}: 2.11 A_f: 1.12 B_f: 1	RFR RMSE: 0.976 Acc: 62.3% Δ_{max}: 1.81 A_f: 1.12 B_f: 1	PLSR RMSE: 1.025 Acc: 58.5% Δ_{max}: 1.9 A_f: 1.13 B_f: 1.01	KNN (k=8) RMSE: 1.063 Acc: 51.1% Δ_{max}: 2.03 A_f: 1.13 B_f: 1	SLR RMSE: 1.07 Acc: 69.4% Δ_{max}: 2.63 A_f: NaN B_f: NaN	PCR RMSE: 1.108 Acc: 50.1% Δ_{max}: 2.01 A_f: 1.14 B_f: 1.01
GCMS VREG	RFR RMSE: 0.72 Acc: 85.4% Δ_{max}: 1.43 A_f: 1.14 B_f: 1.05	PLSR RMSE: 0.807 Acc: 82% Δ_{max}: 1.52 A_f: 1.17 B_f: 1.02	PCR RMSE: 0.849 Acc: 79.1% Δ_{max}: 1.63 A_f: 1.17 B_f: 1	SVR (c=4, γ=0, ε=0.02) RMSE: 0.863 Acc: 79.3% Δ_{max}: 1.77 A_f: 1.16 B_f: 1	KNN (k=7) RMSE: 0.984 Acc: 78.5% Δ_{max}: 1.81 A_f: 1.18 B_f: 1.02	OLSR RMSE: 1.014 Acc: 70.8% Δ_{max}: 2.11 A_f: 1.21 B_f: 1.02	SLR RMSE: 1.221 Acc: 64% Δ_{max}: 2.45 A_f: 1.28 B_f: 1.01
GCMS STAA	KNN (k=5) RMSE: 0.62 Acc: 86.7% Δ_{max}: 1.18 A_f: 1.1 B_f: 1.02	RFR RMSE: 0.766 Acc: 74.7% Δ_{max}: 1.49 A_f: 1.12 B_f: 1.03	SVR (c=30, γ=0.9, ε=0.04) RMSE: 1.073 Acc: 70.2% Δ_{max}: 2.07 A_f: 1.17 B_f: 1.03	PCR RMSE: 1.143 Acc: 60.1% Δ_{max}: 2.16 A_f: 1.21 B_f: 1	PLSR RMSE: 1.153 Acc: 60.9% Δ_{max}: 2.12 A_f: NaN B_f: NaN	OLSR RMSE: 1.446 Acc: 52.9% Δ_{max}: 2.88 A_f: 1.27 B_f: 1	SLR RMSE: 1.535 Acc: 53.6% Δ_{max}: 2.97 A_f: NaN B_f: NaN
GCMS MRS	KNN (k=7) RMSE: 0.594 Acc: 87.9% Δ_{max}: 1.04 A_f: 1.09 B_f: 1	RFR RMSE: 0.61 Acc: 89.6% Δ_{max}: 1.17 A_f: 1.09 B_f: 1.01	PLSR RMSE: 0.674 Acc: 86.8% Δ_{max}: 1.25 A_f: 1.11 B_f: 1.01	SVR (c=1, γ=0, ε=0.1) RMSE: 0.691 Acc: 83.9% Δ_{max}: 1.3 A_f: 1.1 B_f: 0.99	PCR RMSE: 0.698 Acc: 84.6% Δ_{max}: 1.3 A_f: 1.12 B_f: 0.99	OLSR RMSE: 0.733 Acc: 82.5% Δ_{max}: 1.45 A_f: 1.12 B_f: 1.02	SLR RMSE: 0.943 Acc: 71% Δ_{max}: 1.81 A_f: 1.16 B_f: 0.99
GCMS CFC	KNN (k=6) RMSE: 1 Acc: 72.1% Δ_{max}: 2.02 A_f: 1.14 B_f: 1.03	RFR RMSE: 1.059 Acc: 74.4% Δ_{max}: 2.23 A_f: 1.16 B_f: 1.05	SVR (c=10, γ=0.6, ε=0.02) RMSE: 1.23 Acc: 61.3% Δ_{max}: 2.41 A_f: 1.18 B_f: 1.03	PLSR RMSE: 1.35 Acc: 64.6% Δ_{max}: 2.57 A_f: NaN B_f: NaN	PCR RMSE: 1.467 Acc: 54.6% Δ_{max}: 3.03 A_f: NaN B_f: NaN	OLSR RMSE: 1.737 Acc: 52.1% Δ_{max}: 3.58 A_f: NaN B_f: NaN	SLR RMSE: 2.088 Acc: 43.6% Δ_{max}: 4.16 A_f: NaN B_f: NaN
GCMS PCA	RFR RMSE: 0.655 Acc: 85.4% Δ_{max}: 1.2 A_f: 1.08 B_f: 1.01	KNN (k=6) RMSE: 0.793 Acc: 72.8% Δ_{max}: 1.39 A_f: 1.1 B_f: 1.01	PCR RMSE: 0.932 Acc: 69.4% Δ_{max}: 1.71 A_f: 1.12 B_f: 0.99	SVR (c=7, γ=0.9, ε=0.1) RMSE: 0.967 Acc: 67.3% Δ_{max}: 1.83 A_f: 1.11 B_f: 1.03	PLSR RMSE: 0.984 Acc: 70.4% Δ_{max}: 1.77 A_f: 1.12 B_f: 1.01	OLSR RMSE: 1.112 Acc: 63.1% Δ_{max}: 2.16 A_f: 1.14 B_f: 1.02	SLR RMSE: 1.307 Acc: 61.3% Δ_{max}: 2.57 A_f: NaN B_f: NaN
VM VREG	PLSR RMSE: 0.658 Acc: 85.3% Δ_{max}: 1.36 A_f: 1.12 B_f: 1.01	PCR RMSE: 0.676 Acc: 84.4% Δ_{max}: 1.37 A_f: 1.13 B_f: 1.02	RFR RMSE: 0.684 Acc: 83.1% Δ_{max}: 1.26 A_f: 1.14 B_f: 1	KNN (k=6) RMSE: 0.86 Acc: 75.4% Δ_{max}: 1.74 A_f: 1.17 B_f: 0.99	SVR (c=1, γ=0.06, ε=0) RMSE: 1.062 Acc: 69.7% Δ_{max}: 2.4 A_f: 1.19 B_f: 0.93	OLSR RMSE: 24.228 Acc: 23.8% Δ_{max}: 70.61 A_f: NaN B_f: NaN	
VM STAA	PLSR RMSE: 0.685 Acc: 83.1% Δ_{max}: 1.28 A_f: 1.11 B_f: 0.99	RFR RMSE: 0.739 Acc: 83.2% Δ_{max}: 1.28 A_f: 1.13 B_f: 1.01	PCR RMSE: 0.799 Acc: 78.5% Δ_{max}: 1.56 A_f: 1.14 B_f: 1.03	KNN (k=4) RMSE: 0.957 Acc: 62.3% Δ_{max}: 1.69 A_f: 1.17 B_f: 1.04	SVR (c=1, γ=0.9, ε=0.02) RMSE: 1.107 Acc: 46.2% Δ_{max}: 1.91 A_f: 1.2 B_f: 1.02	OLSR RMSE: 152.079 Acc: 11.6% Δ_{max}: 478.03 A_f: NaN B_f: NaN	
VM MRS	RFR RMSE: 0.495 Acc: 97.5% Δ_{max}: 0.91 A_f: 1.08 B_f: 1.01	PLSR RMSE: 0.503 Acc: 96.6% Δ_{max}: 0.96 A_f: 1.08 B_f: 1	PCR RMSE: 0.563 Acc: 93.2% Δ_{max}: 1.06 A_f: 1.09 B_f: 1.02	KNN (k=11) RMSE: 0.682 Acc: 82.5% Δ_{max}: 1.28 A_f: 1.11 B_f: 1	SVR (c=1, γ=0.2, ε=0) RMSE: 0.789 Acc: 74.9% Δ_{max}: 1.49 A_f: 1.12 B_f: 0.98	OLSR RMSE: 48.788 Acc: 18.9% Δ_{max}: 148.98 A_f: NaN B_f: NaN	

VM OFO	PLSR RMSE: 0.853 Acc: 80.2% Δ_{max}: 1.87 A_f: 1.13 B_f: 1.01	RFR RMSE: 0.904 Acc: 68.9% Δ_{max}: 1.76 A_f: 1.14 B_f: 1.02	PCR RMSE: 0.968 Acc: 70.6% Δ_{max}: 2 A_f: 1.16 B_f: 1.03	KNN (k=5) RMSE: 1.158 Acc: 59% Δ_{max}: 2.3 A_f: 1.19 B_f: 1.03	SVR (c=1, γ=0, ε=0) RMSE: 1.726 Acc: 42.3% Δ_{max}: 3.4 A_f: 1.27 B_f: 0.95	OLSR RMSE: 152.542 Acc: 11.3% Δ_{max}: 477 A_f: NaN B_f: NaN	
VM PCA	RFR RMSE: 0.884 Acc: 89.5% Δ_{max}: 1.14 A_f: 1.07 B_f: 1	PLSR RMSE: 0.641 Acc: 84.8% Δ_{max}: 1.31 A_f: 1.07 B_f: 0.99	PCR RMSE: 0.788 Acc: 77% Δ_{max}: 1.54 A_f: 1.09 B_f: 1.01	KNN (k=6) RMSE: 0.989 Acc: 65.2% Δ_{max}: 1.83 A_f: 1.12 B_f: 1.01	SVR (c=1, γ=0, ε=0) RMSE: 1.109 Acc: 41.1% Δ_{max}: 1.79 A_f: 1.15 B_f: 1.03	OLSR RMSE: 134.106 Acc: 13% Δ_{max}: 422.27 A_f: NaN B_f: NaN	
HPLC VREG	KNN (k=3) RMSE: 0.643 Acc: 86.1% Δ_{max}: 1.57 A_f: 1.1 B_f: 1.02	PLSR RMSE: 0.652 Acc: 87.8% Δ_{max}: 1.44 A_f: 1.11 B_f: 1.01	PCR RMSE: 0.682 Acc: 85.3% Δ_{max}: 1.44 A_f: 1.12 B_f: 1	SVR (c=4, γ=0, ε=0.02) RMSE: 0.79 Acc: 78.1% Δ_{max}: 1.83 A_f: 1.12 B_f: 1.01	RFR RMSE: 0.86 Acc: 72.3% Δ_{max}: 1.92 A_f: 1.13 B_f: 1.02	OLSR RMSE: 7.219 Acc: 28.8% Δ_{max}: 17.3 A_f: NaN B_f: NaN	
HPLC STAA	KNN (k=3) RMSE: 0.564 Acc: 92.6% Δ_{max}: 1.52 A_f: 1.07 B_f: 1.02	PLSR RMSE: 0.579 Acc: 88.8% Δ_{max}: 1.24 A_f: 1.08 B_f: 1.01	PCR RMSE: 0.597 Acc: 87.3% Δ_{max}: 1.27 A_f: 1.09 B_f: 1.01	SVR (c=4, γ=0.2, ε=0.02) RMSE: 0.731 Acc: 82.7% Δ_{max}: 1.68 A_f: 1.1 B_f: 1.02	RFR RMSE: 0.748 Acc: 81.8% Δ_{max}: 1.69 A_f: 1.1 B_f: 1.02	OLSR RMSE: 6.3 Acc: 21% Δ_{max}: 15.04 A_f: NaN B_f: NaN	
HPLC MRS	PLSR RMSE: 0.466 Acc: 100% Δ_{max}: 0.85 A_f: 1.07 B_f: 1.01	PCR RMSE: 0.469 Acc: 99.8% Δ_{max}: 0.87 A_f: 1.07 B_f: 1	KNN (k=3) RMSE: 0.492 Acc: 96.2% Δ_{max}: 1.07 A_f: 1.07 B_f: 1.01	SVR (c=4, γ=0.06, ε=0.02) RMSE: 0.579 Acc: 89.1% Δ_{max}: 1.26 A_f: 1.08 B_f: 1.01	RFR RMSE: 0.633 Acc: 87.5% Δ_{max}: 1.27 A_f: 1.09 B_f: 1.01	OLSR RMSE: 8.775 Acc: 33.2% Δ_{max}: 21.14 A_f: NaN B_f: NaN	
HPLC OFO	PLSR RMSE: 0.824 Acc: 79.1% Δ_{max}: 1.67 A_f: 1.12 B_f: 1.01	PCR RMSE: 0.838 Acc: 76.9% Δ_{max}: 1.69 A_f: 1.12 B_f: 1	KNN (k=3) RMSE: 0.957 Acc: 72.5% Δ_{max}: 2.43 A_f: 1.11 B_f: 1.04	SVR (c=1, γ=0.3, ε=0.02) RMSE: 1.062 Acc: 63.6% Δ_{max}: 2.2 A_f: 1.15 B_f: 1.03	RFR RMSE: 1.095 Acc: 63.2% Δ_{max}: 2.32 A_f: 1.15 B_f: 1.02	OLSR RMSE: 8.334 Acc: 22% Δ_{max}: 20.29 A_f: NaN B_f: NaN	
HPLC PCA	KNN (k=3) RMSE: 0.508 Acc: 93.9% Δ_{max}: 1.53 A_f: 1.05 B_f: 1.01	PLSR RMSE: 0.561 Acc: 92.6% Δ_{max}: 1.29 A_f: 1.06 B_f: 1	PCR RMSE: 0.598 Acc: 90.9% Δ_{max}: 1.29 A_f: 1.07 B_f: 1	SVR (c=4, γ=0, ε=0.02) RMSE: 0.668 Acc: 87.4% Δ_{max}: 1.65 A_f: 1.07 B_f: 1.02	RFR RMSE: 0.771 Acc: 79.3% Δ_{max}: 1.82 A_f: 1.08 B_f: 1.01	OLSR RMSE: 10.792 Acc: 29.9% Δ_{max}: 25.56 A_f: NaN B_f: NaN	
FTIR VREG	RFR RMSE: 1.14 Acc: 63.3% Δ_{max}: 2.76 A_f: 1.18 B_f: 1.03	PLSR RMSE: 1.408 Acc: 47.1% Δ_{max}: 3.12 A_f: 1.24 B_f: 1.02	KNN (k=5) RMSE: 1.559 Acc: 40.5% Δ_{max}: 3.35 A_f: 1.27 B_f: 1.11	PCR RMSE: 1.583 Acc: 37.1% Δ_{max}: 3.31 A_f: 1.28 B_f: 1.04	SVR (c=4, γ=0, ε=0.02) RMSE: 1.704 Acc: 30.8% Δ_{max}: 2.9 A_f: 1.31 B_f: 1.05	OLSR RMSE: 16.708 Acc: 10.6% Δ_{max}: 39.33 A_f: NaN B_f: NaN	
FTIR STAA	RFR RMSE: 0.684 Acc: 86.1% Δ_{max}: 1.95 A_f: 1.08 B_f: 1.01	PLSR RMSE: 0.816 Acc: 77.5% Δ_{max}: 1.94 A_f: 1.12 B_f: 1	KNN (k=5) RMSE: 0.994 Acc: 71.3% Δ_{max}: 2.47 A_f: 1.13 B_f: 1.04	PCR RMSE: 1.031 Acc: 65.9% Δ_{max}: 2.24 A_f: 1.15 B_f: 1.02	SVR (c=4, γ=0, ε=0.02) RMSE: 1.082 Acc: 70.7% Δ_{max}: 2.3 A_f: 1.17 B_f: 1.02	OLSR RMSE: 9.561 Acc: 12.3% Δ_{max}: 22.41 A_f: NaN B_f: NaN	
FTIR MRS	RFR RMSE: 0.606 Acc: 90.8% Δ_{max}: 1.55 A_f: 1.08 B_f: 1	PLSR RMSE: 0.73 Acc: 84.1% Δ_{max}: 1.73 A_f: 1.1 B_f: 1	KNN (k=5) RMSE: 0.839 Acc: 74.9% Δ_{max}: 1.85 A_f: 1.12 B_f: 1.03	PCR RMSE: 0.855 Acc: 71.3% Δ_{max}: 1.71 A_f: 1.13 B_f: 1.01	SVR (c=1, γ=0, ε=0.02) RMSE: 0.911 Acc: 73% Δ_{max}: 1.7 A_f: 1.14 B_f: 1.03	OLSR RMSE: 10.18 Acc: 12.6% Δ_{max}: 24.02 A_f: NaN B_f: NaN	
FTIR OFO	RFR RMSE: 1.17 Acc: 67.3% Δ_{max}: 3.19 A_f: 1.14 B_f: 1.02	PLSR RMSE: 1.585 Acc: 44.7% Δ_{max}: 3.75 A_f: 1.22 B_f: 1.02	KNN (k=5) RMSE: 1.829 Acc: 51.1% Δ_{max}: 4.37 A_f: 1.23 B_f: 1.09	PCR RMSE: 1.86 Acc: 28.2% Δ_{max}: 3.54 A_f: 1.27 B_f: 1.04	SVR (c=1, γ=0, ε=0.02) RMSE: 1.972 Acc: 18.6% Δ_{max}: 3.71 A_f: 1.3 B_f: 1.06	OLSR RMSE: 21.965 Acc: 9.2% Δ_{max}: 52.41 A_f: NaN B_f: NaN	
FTIR PCA	RFR RMSE: 0.716 Acc: 86.6% Δ_{max}: 2.05 A_f: 1.07 B_f: 1.01	PLSR RMSE: 0.9 Acc: 74.8% Δ_{max}: 2.25 A_f: 1.1 B_f: 1	KNN (k=5) RMSE: 1.079 Acc: 69.5% Δ_{max}: 2.7 A_f: 1.11 B_f: 1.04	PCR RMSE: 1.147 Acc: 56.6% Δ_{max}: 2.31 A_f: 1.13 B_f: 1.01	SVR (c=1, γ=0, ε=0.02) RMSE: 1.233 Acc: 60.3% Δ_{max}: 2.56 A_f: 1.14 B_f: 1.03	OLSR RMSE: 10.508 Acc: 15% Δ_{max}: 25.51 A_f: NaN B_f: NaN	

Supplementary material for Section 4.1 –S2

Report for metabolomics data coupled to machine learning

February 25, 2017

Best machine learning methods for each scenario

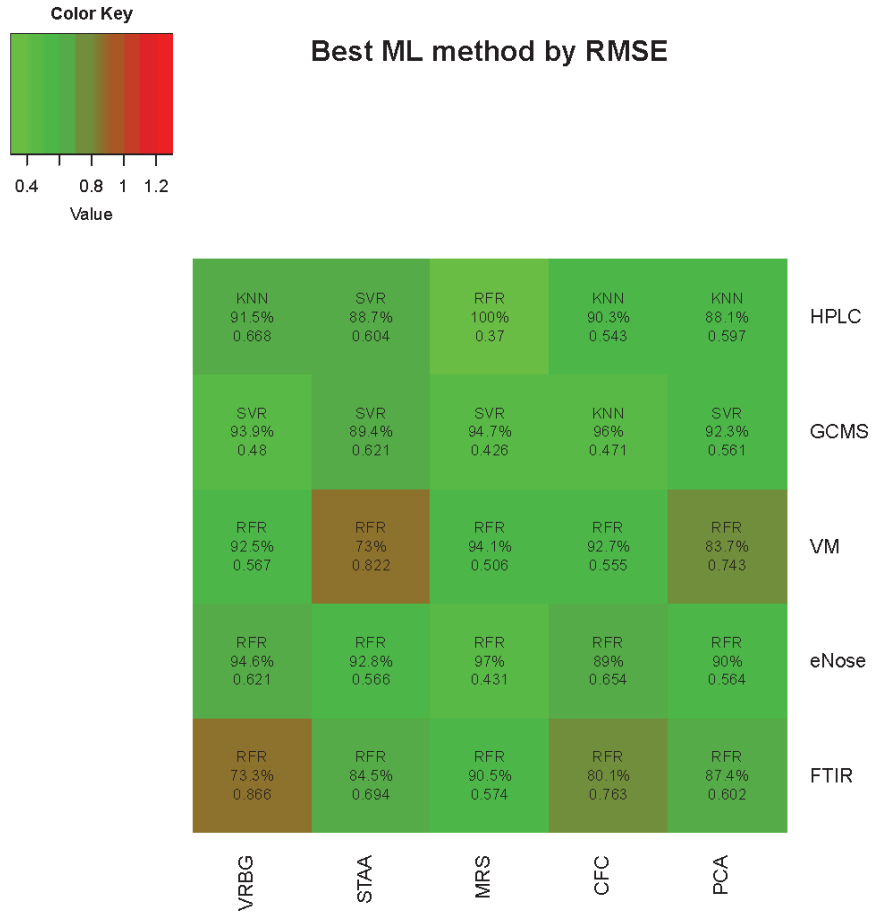


Figure 1: Best machine learning method by RMSE

1

Table 1: Ranking of machine methods by RMSE in first place and accuracy in second place.

	rank #1	rank #2	rank #3	rank #4	rank #5	rank #6	rank #7
HPLC VRBG	KNN (k=5) RMSE: 0.668 Acc: 91.5% Δ_{max}: 1.26 A_f: 1.15 B_f: 1	RFR RMSE: 0.681 Acc: 88.9% Δ_{max}: 1.21 A_f: 1.15 B_f: 1	PLSR RMSE: 0.697 Acc: 83.6% Δ_{max}: 1.35 A_f: 1.15 B_f: 0.98	SLR RMSE: 0.717 Acc: 81.1% Δ_{max}: 1.41 A_f: 1.14 B_f: 0.99	SVR (c=10, γ=0, ε=0.06) RMSE: 0.763 Acc: 82.3% Δ_{max}: 1.49 A_f: 1.16 B_f: 1	OLSR RMSE: 0.783 Acc: 76.6% Δ_{max}: 1.47 A_f: 1.17 B_f: 0.98	PCR RMSE: 0.8 Acc: 78.7% Δ_{max}: 1.53 A_f: 1.18 B_f: 1
HPLC STAA	SVR (c=7, γ=0.09, ε=0.04) RMSE: 0.604 Acc: 88.7% Δ_{max}: 1.23 A_f: 1.09 B_f: 0.98	KNN (k=5) RMSE: 0.672 Acc: 83.8% Δ_{max}: 1.32 A_f: 1.11 B_f: 1	SLR RMSE: 0.704 Acc: 84.1% Δ_{max}: 1.47 A_f: 1.12 B_f: 0.97	RFR RMSE: 0.705 Acc: 84.2% Δ_{max}: 1.27 A_f: 1.12 B_f: 1.01	PCR RMSE: 0.708 Acc: 83.3% Δ_{max}: 1.39 A_f: 1.12 B_f: 1	PLSR RMSE: 0.724 Acc: 80.8% Δ_{max}: 1.39 A_f: 1.12 B_f: 0.99	OLSR RMSE: 0.78 Acc: 82.6% Δ_{max}: 1.45 A_f: 1.13 B_f: 0.98
HPLC MRS	RFR RMSE: 0.37 Acc: 100% Δ_{max}: 0.76 A_f: 1.06 B_f: 1	KNN (k=4) RMSE: 0.386 Acc: 99.5% Δ_{max}: 0.73 A_f: 1.06 B_f: 1	SVR (c=10, γ=0.6, ε=0.04) RMSE: 0.402 Acc: 96.8% Δ_{max}: 0.92 A_f: 1.06 B_f: 1.01	PLSR RMSE: 0.416 Acc: 99.5% Δ_{max}: 0.8 A_f: 1.07 B_f: 0.99	SLR RMSE: 0.42 Acc: 96.7% Δ_{max}: 0.88 A_f: 1.06 B_f: 0.99	PCR RMSE: 0.449 Acc: 99.5% Δ_{max}: 0.8 A_f: 1.07 B_f: 1	OLSR RMSE: 0.45 Acc: 94.8% Δ_{max}: 0.98 A_f: 1.07 B_f: 0.99
HPLC CFC	KNN (k=5) RMSE: 0.543 Acc: 90.3% Δ_{max}: 1.09 A_f: 1.1 B_f: 0.99	RFR RMSE: 0.56 Acc: 92% Δ_{max}: 1.14 A_f: 1.1 B_f: 1.02	PLSR RMSE: 0.607 Acc: 88.5% Δ_{max}: 1.27 A_f: 1.11 B_f: 0.99	PCR RMSE: 0.645 Acc: 87% Δ_{max}: 1.35 A_f: 1.12 B_f: 0.99	SLR RMSE: 0.663 Acc: 86.4% Δ_{max}: 1.31 A_f: 1.12 B_f: 1.01	OLSR RMSE: 0.759 Acc: 79.5% Δ_{max}: 1.51 A_f: 1.15 B_f: 0.98	SV (c=110, γ=0 RMSE Acc: Δ_{ma} A_f: B_f:
HPLC PCA	KNN (k=5) RMSE: 0.597 Acc: 88.1% Δ_{max}: 1.18 A_f: 1.07 B_f: 1.01	PCR RMSE: 0.648 Acc: 86.6% Δ_{max}: 1.23 A_f: 1.08 B_f: 1	SLR RMSE: 0.657 Acc: 85% Δ_{max}: 1.3 A_f: 1.08 B_f: 0.99	RFR RMSE: 0.683 Acc: 84.4% Δ_{max}: 1.41 A_f: 1.08 B_f: 1.01	SVR (c=19, γ=0.3, ε=0.02) RMSE: 0.714 Acc: 83.3% Δ_{max}: 1.45 A_f: 1.08 B_f: 1	PLSR RMSE: 0.739 Acc: 78.7% Δ_{max}: 1.46 A_f: 1.09 B_f: 0.99	OLSR RMSE: 0.75 Acc: 79.3% Δ_{max}: 1.42 A_f: 1.1 B_f: 1
GCMS VRBG	SVR (c=7, γ=0.06, ε=0.02) RMSE: 0.48 Acc: 93.9% Δ_{max}: 1.11 A_f: 1.08 B_f: 1	KNN (k=3) RMSE: 0.502 Acc: 93.2% Δ_{max}: 1.27 A_f: 1.08 B_f: 1.02	RFR RMSE: 0.623 Acc: 90.1% Δ_{max}: 1.49 A_f: 1.11 B_f: 1.01	PLSR RMSE: 0.707 Acc: 85.8% Δ_{max}: 1.61 A_f: 1.13 B_f: 1.02	PCR RMSE: 0.751 Acc: 83% Δ_{max}: 1.64 A_f: 1.14 B_f: 1.02	OLSR RMSE: 49.356 Acc: 23.6% Δ_{max}: 106.19 A_f: NaN B_f: NaN	
GCMS STAA	SVR (c=4, γ=0, ε=0.04) RMSE: 0.621 Acc: 89.4% Δ_{max}: 1.37 A_f: 1.09 B_f: 1.01	RFR RMSE: 0.665 Acc: 85.1% Δ_{max}: 1.46 A_f: 1.1 B_f: 1.02	KNN (k=3) RMSE: 0.678 Acc: 88.4% Δ_{max}: 1.72 A_f: 1.09 B_f: 1.04	PLSR RMSE: 0.775 Acc: 77.7% Δ_{max}: 1.56 A_f: 1.12 B_f: 1.02	PCR RMSE: 0.788 Acc: 74.9% Δ_{max}: 1.48 A_f: 1.12 B_f: 1.01	OLSR RMSE: 67.589 Acc: 16.7% Δ_{max}: 139.34 A_f: NaN B_f: NaN	
GCMS MRS	SVR (c=4, γ=0, ε=0.02) RMSE: 0.426 Acc: 94.7% Δ_{max}: 1.11 A_f: 1.05 B_f: 1	KNN (k=3) RMSE: 0.465 Acc: 95.9% Δ_{max}: 1.2 A_f: 1.06 B_f: 1.02	RFR RMSE: 0.469 Acc: 95.4% Δ_{max}: 1.17 A_f: 1.06 B_f: 1.01	PLSR RMSE: 0.57 Acc: 92.1% Δ_{max}: 1.51 A_f: 1.07 B_f: 1.01	PCR RMSE: 0.604 Acc: 89.9% Δ_{max}: 1.47 A_f: 1.08 B_f: 1.01	OLSR RMSE: 62.271 Acc: 23.4% Δ_{max}: 126.72 A_f: NaN B_f: NaN	
GCMS CFC	KNN (k=3) RMSE: 0.471 Acc: 96% Δ_{max}: 1.01 A_f: 1.08 B_f: 1	SVR (c=4, γ=0, ε=0.02) RMSE: 0.49 Acc: 94.1% Δ_{max}: 1.19 A_f: 1.08 B_f: 1.02	RFR RMSE: 0.511 Acc: 93.9% Δ_{max}: 1.09 A_f: 1.08 B_f: 1.01	PLSR RMSE: 0.589 Acc: 90.1% Δ_{max}: 1.35 A_f: 1.1 B_f: 1.02	PCR RMSE: 0.613 Acc: 90.1% Δ_{max}: 1.41 A_f: 1.1 B_f: 1.01	OLSR RMSE: 50.188 Acc: 27.6% Δ_{max}: 108.51 A_f: NaN B_f: NaN	
GCMS PCA	SVR (c=19, γ=0.3, ε=0.02) RMSE: 0.561 Acc: 92.3% Δ_{max}: 1.33 A_f: 1.06 B_f: 1	KNN (k=5) RMSE: 0.645 Acc: 87% Δ_{max}: 1.8 A_f: 1.06 B_f: 1.01	RFR RMSE: 0.658 Acc: 85.3% Δ_{max}: 1.66 A_f: 1.06 B_f: 1.01	PLSR RMSE: 0.733 Acc: 81.4% Δ_{max}: 1.71 A_f: 1.08 B_f: 1.01	PCR RMSE: 0.752 Acc: 79.5% Δ_{max}: 1.68 A_f: 1.08 B_f: 1.01	OLSR RMSE: 52.428 Acc: 20% Δ_{max}: 110.62 A_f: NaN B_f: NaN	
VM VRBG	RFR RMSE: 0.567 Acc: 92.5% Δ_{max}: 1.23 A_f: 1.12 B_f: 1.02	SVR (c=50, γ=0.2, ε=0.02) RMSE: 0.597 Acc: 89.1% Δ_{max}: 1.63 A_f: 1.12 B_f: 0.99	PLSR RMSE: 0.605 Acc: 92% Δ_{max}: 1.52 A_f: 1.13 B_f: 1.02	KNN (k=3) RMSE: 0.611 Acc: 89.1% Δ_{max}: 1.36 A_f: 1.13 B_f: 0.99	SLR RMSE: 0.655 Acc: 89.4% Δ_{max}: 1.65 A_f: 1.13 B_f: 1.01	PCR RMSE: 0.67 Acc: 88.3% Δ_{max}: 1.8 A_f: 1.15 B_f: 1.01	OLSR RMSE: 0.67 Acc: 88.8% Δ_{max}: 1.89 A_f: 1.13 B_f: 1.01
VM STAA	RFR RMSE: 0.822 Acc: 73% Δ_{max}: 1.72 A_f: 1.13 B_f: 1.01	PLSR RMSE: 0.864 Acc: 75.5% Δ_{max}: 2.05 A_f: 1.14 B_f: 1.02	KNN (k=3) RMSE: 0.867 Acc: 74.2% Δ_{max}: 2.05 A_f: 1.13 B_f: 1	SLR RMSE: 0.913 Acc: 75.5% Δ_{max}: 2.68 A_f: 1.13 B_f: 1.01	OLSR RMSE: 0.924 Acc: 75.6% Δ_{max}: 2.79 A_f: 1.13 B_f: 1.01	PCR RMSE: 0.965 Acc: 62.8% Δ_{max}: 1.89 A_f: 1.17 B_f: 1.02	SVF (c=90, γ=0. RMSE: Acc: 6 Δ_{max}: A_f: 1 B_f: 0
VM MRS	RFR RMSE: 0.506 Acc: 94.1% Δ_{max}: 1.37 A_f: 1.07 B_f: 1.01	KNN (k=3) RMSE: 0.576 Acc: 92.7% Δ_{max}: 1.74 A_f: 1.08 B_f: 0.99	PLSR RMSE: 0.576 Acc: 91.2% Δ_{max}: 1.36 A_f: 1.08 B_f: 1.01	OLSR RMSE: 0.592 Acc: 90.2% Δ_{max}: 1.53 A_f: 1.08 B_f: 1	SLR RMSE: 0.602 Acc: 88.6% Δ_{max}: 1.48 A_f: 1.09 B_f: 1	PCR RMSE: 0.606 Acc: 90.2% Δ_{max}: 1.8 A_f: 1.09 B_f: 1	SVI (c=70, γ=0. RMSE: Acc: 9 Δ_{max}: A_f: 1 B_f:

VM OFO	RFR RMSE: 0.555 Acc: 92.7% Δ_{max}: 1.3 A_f: 1.1 B_f: 1	KNN (k=7) RMSE: 0.594 Acc: 91.2% Δ_{max}: 1.21 A_f: 1.11 B_f: 1	PLSR RMSE: 0.597 Acc: 91% Δ_{max}: 1.32 A_f: 1.11 B_f: 1.01	SVR (c=4, γ=0.9, ε=0.6) RMSE: 0.607 Acc: 92% Δ_{max}: 1.31 A_f: 1.12 B_f: 1	POR RMSE: 0.627 Acc: 90.9% Δ_{max}: 1.24 A_f: 1.12 B_f: 1	OLSR RMSE: 0.698 Acc: 88% Δ_{max}: 2.06 A_f: 1.12 B_f: 1.01	SLR RMSE: 0.69 Acc: 85.9% Δ_{max}: 1.8 A_f: 1.13 B_f: 1.01
VM POA	RFR RMSE: 0.743 Acc: 83.7% Δ_{max}: 1.76 A_f: 1.08 B_f: 1.01	SVR (c=50, γ=0, ε=0.06) RMSE: 0.765 Acc: 79.8% Δ_{max}: 1.91 A_f: 1.08 B_f: 1	KNN (k=8) RMSE: 0.783 Acc: 80.3% Δ_{max}: 1.81 A_f: 1.09 B_f: 0.99	PLSR RMSE: 0.804 Acc: 79.1% Δ_{max}: 1.96 A_f: 1.09 B_f: 1.01	POR RMSE: 0.84 Acc: 78.7% Δ_{max}: 1.95 A_f: 1.1 B_f: 1.01	SLR RMSE: 0.84 Acc: 76.9% Δ_{max}: 2.34 A_f: 1.09 B_f: 1.01	OLSR RMSE: 0.86 Acc: 77.3% Δ_{max}: 2.57 A_f: 1.09 B_f: 1.01
eNose VREG	RFR RMSE: 0.621 Acc: 94.6% Δ_{max}: 1.02 A_f: 1.14 B_f: 1.01	SVR (c=4, γ=0.3, ε=0.02) RMSE: 0.674 Acc: 88.7% Δ_{max}: 1.11 A_f: 1.15 B_f: 1.01	KNN (k=9) RMSE: 0.689 Acc: 90.2% Δ_{max}: 1.08 A_f: 1.16 B_f: 0.99	PLSR RMSE: 0.691 Acc: 84.4% Δ_{max}: 1.31 A_f: 1.15 B_f: 1	POR RMSE: 0.711 Acc: 87.5% Δ_{max}: 1.15 A_f: 1.16 B_f: 1.01	OLSR RMSE: 4602.54 Acc: 21.4% Δ_{max}: 15644.91 A_f: NaN B_f: NaN	
eNose STAA	RFR RMSE: 0.566 Acc: 92.8% Δ_{max}: 1.03 A_f: 1.1 B_f: 1.01	PLSR RMSE: 0.629 Acc: 89% Δ_{max}: 1.28 A_f: 1.1 B_f: 1	POR RMSE: 0.823 Acc: 71.6% Δ_{max}: 1.44 A_f: 1.14 B_f: 1.01	SVR (c=4, γ=0, ε=0.02) RMSE: 0.964 Acc: 60% Δ_{max}: 1.57 A_f: 1.17 B_f: 1.01	KNN (k=3) RMSE: 1.004 Acc: 62.9% Δ_{max}: 1.8 A_f: 1.17 B_f: 0.99	OLSR RMSE: 3708.561 Acc: 12.6% Δ_{max}: 12542.54 A_f: NaN B_f: NaN	
eNose MRS	RFR RMSE: 0.431 Acc: 97% Δ_{max}: 0.89 A_f: 1.07 B_f: 1.01	PLSR RMSE: 0.506 Acc: 94% Δ_{max}: 1.16 A_f: 1.07 B_f: 1	POR RMSE: 0.577 Acc: 94.5% Δ_{max}: 1.2 A_f: 1.09 B_f: 1	SVR (c=4, γ=0, ε=0.02) RMSE: 0.646 Acc: 86.3% Δ_{max}: 1.33 A_f: 1.09 B_f: 1	KNN (k=3) RMSE: 0.649 Acc: 88.1% Δ_{max}: 1.26 A_f: 1.1 B_f: 1.02	OLSR RMSE: 2784.458 Acc: 28.7% Δ_{max}: 9333.96 A_f: NaN B_f: NaN	
eNose OFO	RFR RMSE: 0.654 Acc: 89% Δ_{max}: 1.13 A_f: 1.13 B_f: 1.02	SVR (c=4, γ=0, ε=0.02) RMSE: 0.666 Acc: 88.1% Δ_{max}: 1.1 A_f: 1.14 B_f: 1.02	POR RMSE: 0.667 Acc: 88.5% Δ_{max}: 1.11 A_f: 1.14 B_f: 1.01	PLSR RMSE: 0.679 Acc: 85.2% Δ_{max}: 1.32 A_f: 1.13 B_f: 1.01	KNN (k=3) RMSE: 0.759 Acc: 76.8% Δ_{max}: 1.2 A_f: 1.16 B_f: 0.94	OLSR RMSE: 6750.417 Acc: 15% Δ_{max}: 22879.25 A_f: NaN B_f: NaN	
eNose POA	RFR RMSE: 0.564 Acc: 90% Δ_{max}: 1.13 A_f: 1.06 B_f: 0.99	PLSR RMSE: 0.654 Acc: 87.8% Δ_{max}: 1.29 A_f: 1.08 B_f: 1	POR RMSE: 0.837 Acc: 78.1% Δ_{max}: 1.66 A_f: 1.1 B_f: 1	KNN (k=3) RMSE: 0.908 Acc: 70.8% Δ_{max}: 1.8 A_f: 1.11 B_f: 0.98	SVR (c=4, γ=0, ε=0.02) RMSE: 0.913 Acc: 68.8% Δ_{max}: 1.68 A_f: 1.12 B_f: 1	OLSR RMSE: 4384.675 Acc: 13.2% Δ_{max}: 14710.48 A_f: NaN B_f: NaN	
FTIR VREG	RFR RMSE: 0.866 Acc: 73.3% Δ_{max}: 1.85 A_f: 1.16 B_f: 1.04	KNN (k=9) RMSE: 1.11 Acc: 57.5% Δ_{max}: 2.11 A_f: 1.21 B_f: 1.06	POR RMSE: 1.111 Acc: 57% Δ_{max}: 2.16 A_f: 1.21 B_f: 1.02	PLSR RMSE: 1.129 Acc: 69.6% Δ_{max}: 2.71 A_f: 1.21 B_f: 1.02	SVR (c=1, γ=0.3, ε=0.02) RMSE: 1.147 Acc: 55.4% Δ_{max}: 2.19 A_f: 1.22 B_f: 1.02	OLSR RMSE: 29.131 Acc: 10.6% Δ_{max}: 67.02 A_f: NaN B_f: NaN	
FTIR STAA	RFR RMSE: 0.694 Acc: 84.5% Δ_{max}: 1.54 A_f: 1.1 B_f: 1	PLSR RMSE: 0.863 Acc: 75.2% Δ_{max}: 2.06 A_f: 1.13 B_f: 1.01	KNN (k=9) RMSE: 0.866 Acc: 79.6% Δ_{max}: 2.13 A_f: 1.13 B_f: 1.03	POR RMSE: 0.892 Acc: 77.3% Δ_{max}: 2.11 A_f: 1.14 B_f: 1.01	SVR (c=1, γ=0.6, ε=0.02) RMSE: 0.935 Acc: 67% Δ_{max}: 1.86 A_f: 1.16 B_f: 0.99	OLSR RMSE: 23.418 Acc: 13.7% Δ_{max}: 53.24 A_f: NaN B_f: NaN	
FTIR MRS	RFR RMSE: 0.574 Acc: 90.5% Δ_{max}: 1.27 A_f: 1.08 B_f: 1	PLSR RMSE: 0.686 Acc: 85.5% Δ_{max}: 1.56 A_f: 1.1 B_f: 1	POR RMSE: 0.716 Acc: 83.2% Δ_{max}: 1.49 A_f: 1.11 B_f: 1	KNN (k=9) RMSE: 0.725 Acc: 83% Δ_{max}: 1.63 A_f: 1.11 B_f: 1.01	SVR (c=1, γ=0.9, ε=0) RMSE: 0.736 Acc: 88% Δ_{max}: 1.68 A_f: 1.11 B_f: 1	OLSR RMSE: 16.072 Acc: 14.8% Δ_{max}: 38.29 A_f: NaN B_f: NaN	
FTIR OFO	RFR RMSE: 0.763 Acc: 80.1% Δ_{max}: 1.67 A_f: 1.13 B_f: 1.01	KNN (k=9) RMSE: 0.931 Acc: 67% Δ_{max}: 1.91 A_f: 1.16 B_f: 1.03	POR RMSE: 0.958 Acc: 62.6% Δ_{max}: 1.97 A_f: 1.16 B_f: 1.01	SVR (c=1, γ=0, ε=0.02) RMSE: 0.98 Acc: 60.3% Δ_{max}: 1.92 A_f: 1.16 B_f: 1.01	PLSR RMSE: 1.042 Acc: 64.4% Δ_{max}: 2.45 A_f: 1.18 B_f: 1.01	OLSR RMSE: 19.746 Acc: 11.2% Δ_{max}: 46.07 A_f: NaN B_f: NaN	
FTIR POA	RFR RMSE: 0.602 Acc: 87.4% Δ_{max}: 1.4 A_f: 1.06 B_f: 1	PLSR RMSE: 0.826 Acc: 78.5% Δ_{max}: 2.03 A_f: 1.09 B_f: 1	KNN (k=9) RMSE: 0.831 Acc: 83.7% Δ_{max}: 2.05 A_f: 1.09 B_f: 1.01	POR RMSE: 0.846 Acc: 82.6% Δ_{max}: 1.99 A_f: 1.1 B_f: 1	SVR (c=4, γ=0.09, ε=0.02) RMSE: 0.88 Acc: 84% Δ_{max}: 1.94 A_f: 1.1 B_f: 1	OLSR RMSE: 12.799 Acc: 14.4% Δ_{max}: 29.79 A_f: NaN B_f: NaN	

Supplementary material for Section 4.1 –S3

Report for metabolomics data coupled to machine learning

February 25, 2017

Best machine learning methods for each scenario

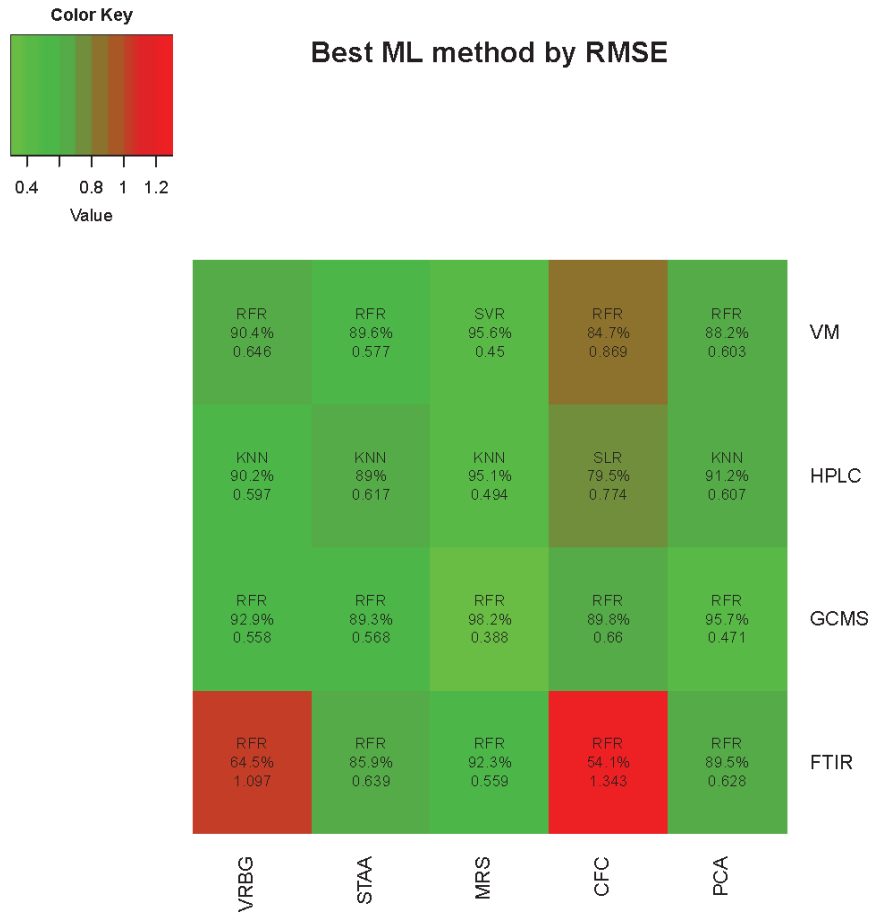


Figure 1: Best machine learning method by RMSE

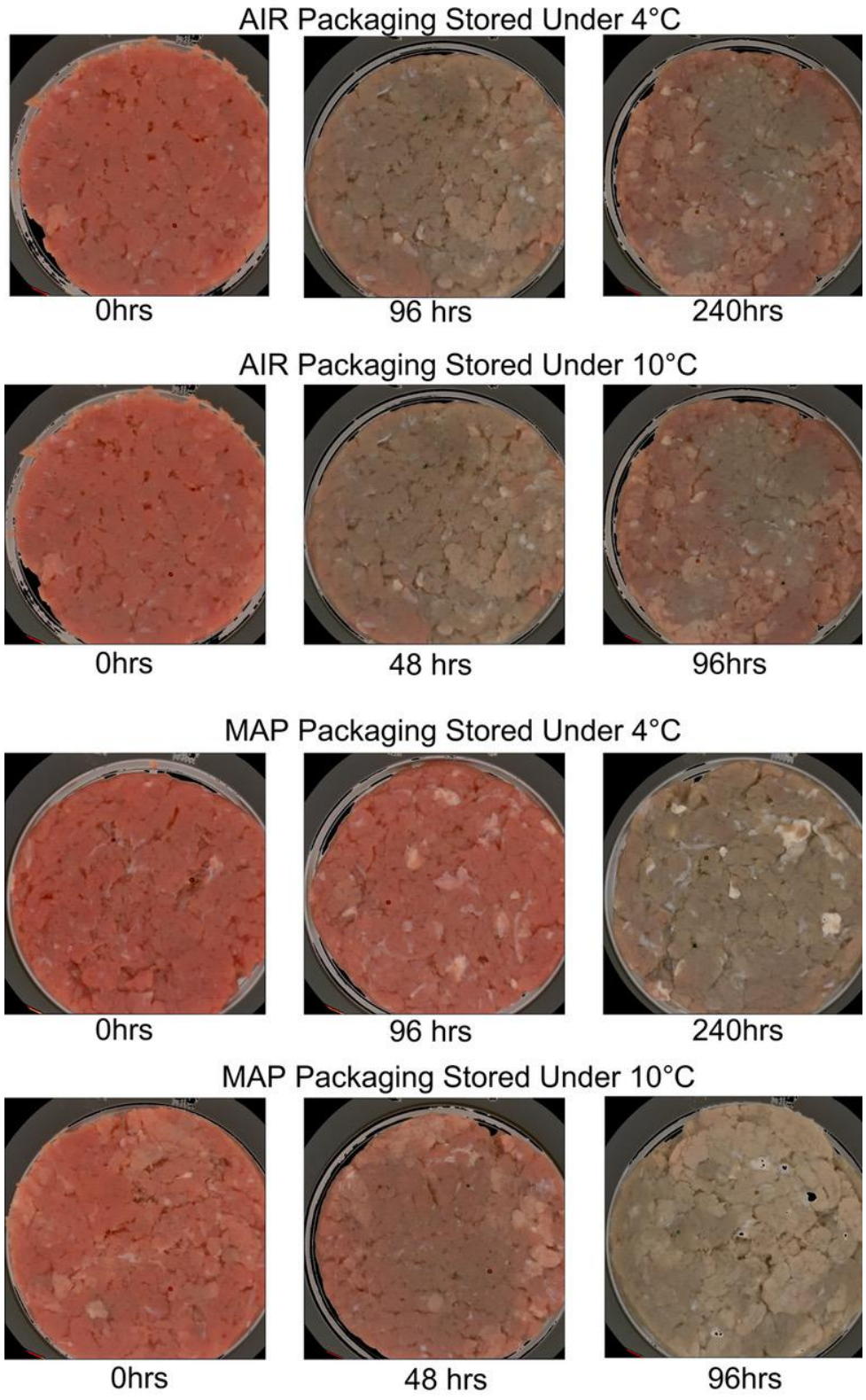
1

Table 1: Ranking of machine methods by RMSE in first place and accuracy in second place.

	rank #1	rank #2	rank #3	rank #4	rank #5	rank #6	rank #7
VM VREG	RFR RMSE: 0.646 Acc: 90.4% Δ_{max}: 1.72 A_f: 1.12 B_f: 1.02	SVR (c=4, γ=0.6, ε=0.1) RMSE: 0.725 Acc: 87.3% Δ_{max}: 1.83 A_f: 1.14 B_f: 1.01	OLSR RMSE: 0.773 Acc: 82.8% Δ_{max}: 1.68 A_f: 1.17 B_f: 1.01	KNN (k=11) RMSE: 0.779 Acc: 86.2% Δ_{max}: 2 A_f: 1.15 B_f: 0.99	PLSR RMSE: 0.843 Acc: 81.5% Δ_{max}: 2.05 A_f: NaN B_f: NaN	SLR RMSE: 0.845 Acc: 79.2% Δ_{max}: 1.9 A_f: 1.19 B_f: 1	POR RMSE: 0.90 Acc: 79.2% Δ_{max}: 2.2 A_f: NaN B_f: NaN
VM STAA	RFR RMSE: 0.577 Acc: 89.6% Δ_{max}: 1.47 A_f: 1.08 B_f: 1	KNN (k=5) RMSE: 0.627 Acc: 85% Δ_{max}: 1.38 A_f: 1.09 B_f: 0.99	SVR (c=7, γ=0, ε=0.2) RMSE: 0.701 Acc: 77.5% Δ_{max}: 1.48 A_f: 1.1 B_f: 1.01	OLSR RMSE: 1.018 Acc: 74.1% Δ_{max}: 2.62 A_f: NaN B_f: NaN	POR RMSE: 1.04 Acc: 74.4% Δ_{max}: 3 A_f: NaN B_f: NaN	SLR RMSE: 1.087 Acc: 73.9% Δ_{max}: 2.86 A_f: NaN B_f: NaN	PLSR RMSE: 1.11 Acc: 66.1% Δ_{max}: 2.8 A_f: NaN B_f: NaN
VM MRS	SVR (c=4, γ=0.09, ε=0.1) RMSE: 0.45 Acc: 95.6% Δ_{max}: 1.21 A_f: 1.06 B_f: 1	RFR RMSE: 0.466 Acc: 95.2% Δ_{max}: 1.27 A_f: 1.06 B_f: 0.99	KNN (k=11) RMSE: 0.474 Acc: 96.9% Δ_{max}: 1.02 A_f: 1.08 B_f: 1.01	OLSR RMSE: 0.567 Acc: 92.9% Δ_{max}: 1.41 A_f: 1.09 B_f: 1	PLSR RMSE: 0.644 Acc: 91.4% Δ_{max}: 1.62 A_f: 1.1 B_f: 0.99	SLR RMSE: 0.649 Acc: 89.9% Δ_{max}: 1.58 A_f: 1.1 B_f: 0.99	POR RMSE: 0.68 Acc: 90.1% Δ_{max}: 1.8 A_f: 1.11 B_f: 0.99
VM CFC	RFR RMSE: 0.869 Acc: 84.7% Δ_{max}: 2.67 A_f: 1.11 B_f: 1	KNN (k=13) RMSE: 0.967 Acc: 81.2% Δ_{max}: 2.83 A_f: 1.15 B_f: 1.01	SVR (c=19, γ=0.03, ε=0.06) RMSE: 1.093 Acc: 72.1% Δ_{max}: 2.83 A_f: 1.15 B_f: 1	OLSR RMSE: 1.239 Acc: 62.9% Δ_{max}: 3.1 A_f: NaN B_f: NaN	POR RMSE: 1.286 Acc: 71.8% Δ_{max}: 3.69 A_f: NaN B_f: NaN	PLSR RMSE: 1.311 Acc: 66.8% Δ_{max}: 3.74 A_f: NaN B_f: NaN	SLR RMSE: 1.32 Acc: 60.8% Δ_{max}: 3.48 A_f: NaN B_f: NaN
VM POA	RFR RMSE: 0.603 Acc: 88.2% Δ_{max}: 1.46 A_f: 1.06 B_f: 1	KNN (k=6) RMSE: 0.627 Acc: 88.3% Δ_{max}: 1.4 A_f: 1.07 B_f: 1.01	SVR (c=19, γ=0.2, ε=0.02) RMSE: 0.74 Acc: 82.5% Δ_{max}: 1.8 A_f: 1.08 B_f: 1.01	OLSR RMSE: 0.896 Acc: 76.9% Δ_{max}: 2.16 A_f: 1.11 B_f: 1	POR RMSE: 0.933 Acc: 73.8% Δ_{max}: 2.42 A_f: 1.12 B_f: 0.99	PLSR RMSE: 0.965 Acc: 73.7% Δ_{max}: 2.43 A_f: 1.12 B_f: 0.98	SLR RMSE: 0.99 Acc: 72.6% Δ_{max}: 2.4 A_f: NaN B_f: NaN
HPLC VREG	KNN (k=4) RMSE: 0.597 Acc: 90.2% Δ_{max}: 1.58 A_f: 1.1 B_f: 1.02	SLR RMSE: 0.706 Acc: 84.4% Δ_{max}: 1.71 A_f: 1.12 B_f: 1.01	SVR (c=50, γ=0, ε=0.02) RMSE: 0.722 Acc: 88.7% Δ_{max}: 1.87 A_f: 1.11 B_f: 1.01	RFR RMSE: 0.741 Acc: 88.4% Δ_{max}: 1.92 A_f: 1.12 B_f: 1.02	OLSR RMSE: 0.741 Acc: 81.3% Δ_{max}: 1.83 A_f: 1.12 B_f: 1	PLSR RMSE: 0.838 Acc: 76.4% Δ_{max}: 1.95 A_f: 1.16 B_f: 1.01	POR RMSE: 0.92 Acc: 71.4% Δ_{max}: 2.1 A_f: 1.17 B_f: 1.02
HPLC STAA	KNN (k=4) RMSE: 0.617 Acc: 89% Δ_{max}: 1.68 A_f: 1.08 B_f: 1.02	SVR (c=7, γ=0.2, ε=0.04) RMSE: 0.674 Acc: 86% Δ_{max}: 1.78 A_f: 1.09 B_f: 1.01	RFR RMSE: 0.683 Acc: 85% Δ_{max}: 1.7 A_f: 1.09 B_f: 1.01	SLR RMSE: 0.765 Acc: 81.4% Δ_{max}: 1.93 A_f: 1.11 B_f: 1.01	OLSR RMSE: 0.777 Acc: 80.9% Δ_{max}: 2.06 A_f: 1.11 B_f: 1.01	PLSR RMSE: 0.824 Acc: 77.7% Δ_{max}: 1.84 A_f: 1.12 B_f: 1.01	POR RMSE: 0.87 Acc: 77.7% Δ_{max}: 1.9 A_f: 1.13 B_f: 1.01
HPLC MRS	KNN (k=4) RMSE: 0.494 Acc: 95.1% Δ_{max}: 1.34 A_f: 1.07 B_f: 1.01	RFR RMSE: 0.545 Acc: 92.7% Δ_{max}: 1.43 A_f: 1.07 B_f: 1	SVR (c=4, γ=0.03, ε=0.02) RMSE: 0.551 Acc: 92.8% Δ_{max}: 1.5 A_f: 1.07 B_f: 1	SLR RMSE: 0.569 Acc: 91.6% Δ_{max}: 1.52 A_f: 1.08 B_f: 1	PLSR RMSE: 0.597 Acc: 88.3% Δ_{max}: 1.6 A_f: 1.08 B_f: 1	OLSR RMSE: 0.601 Acc: 89.8% Δ_{max}: 1.62 A_f: 1.08 B_f: 1	POR RMSE: 0.61 Acc: 87.9% Δ_{max}: 1.47 A_f: 1.09 B_f: 1.01
HPLC CFC	SLR RMSE: 0.774 Acc: 79.5% Δ_{max}: 1.92 A_f: 1.11 B_f: 1.01	OLSR RMSE: 0.781 Acc: 80.6% Δ_{max}: 1.93 A_f: 1.11 B_f: 1	KNN (k=4) RMSE: 0.785 Acc: 82.9% Δ_{max}: 2.33 A_f: 1.1 B_f: 1.03	SVR (c=50, γ=0.03, ε=0.02) RMSE: 0.799 Acc: 80.9% Δ_{max}: 2.19 A_f: 1.1 B_f: 1.02	RFR RMSE: 0.917 Acc: 72.1% Δ_{max}: 2.2 A_f: 1.12 B_f: 1.02	PLSR RMSE: 1.091 Acc: 65.6% Δ_{max}: 2.56 A_f: 1.16 B_f: 1.01	POR RMSE: 1.2 Acc: 62.7% Δ_{max}: 2.76 A_f: 1.17 B_f: 1.02
HPLC POA	KNN (k=4) RMSE: 0.607 Acc: 91.2% Δ_{max}: 1.93 A_f: 1.06 B_f: 1.02	RFR RMSE: 0.705 Acc: 84.7% Δ_{max}: 1.82 A_f: 1.07 B_f: 1.01	SVR (c=1, γ=0, ε=0.04) RMSE: 0.719 Acc: 85% Δ_{max}: 1.91 A_f: 1.07 B_f: 1.02	SLR RMSE: 0.769 Acc: 82.1% Δ_{max}: 1.97 A_f: 1.08 B_f: 1	OLSR RMSE: 0.787 Acc: 82% Δ_{max}: 2.09 A_f: 1.08 B_f: 1	PLSR RMSE: 0.853 Acc: 75.3% Δ_{max}: 2.02 A_f: 1.09 B_f: 1.01	POR RMSE: 0.87 Acc: 75.9% Δ_{max}: 2.08 A_f: 1.1 B_f: 1.01
GCMS VREG	RFR RMSE: 0.558 Acc: 92.9% Δ_{max}: 1.42 A_f: 1.1 B_f: 1.01	PLSR RMSE: 0.654 Acc: 88.7% Δ_{max}: 1.57 A_f: 1.13 B_f: 1	SVR (c=4, γ=0, ε=0.02) RMSE: 0.731 Acc: 87.2% Δ_{max}: 1.89 A_f: 1.13 B_f: 1.03	POR RMSE: 0.782 Acc: 77.8% Δ_{max}: 1.7 A_f: 1.16 B_f: 1.02	KNN (k=17) RMSE: 0.81 Acc: 83.1% Δ_{max}: 2.04 A_f: 1.17 B_f: 1	OLSR RMSE: 9965.109 Acc: 29.6% Δ_{max}: 44045.23 A_f: NaN B_f: NaN	
GCMS STAA	RFR RMSE: 0.568 Acc: 89.3% Δ_{max}: 1.89 A_f: 1.08 B_f: 1	PLSR RMSE: 0.647 Acc: 87.9% Δ_{max}: 1.47 A_f: 1.11 B_f: 1.01	KNN (k=8) RMSE: 0.84 Acc: 76.8% Δ_{max}: 1.97 A_f: 1.13 B_f: 1	SVR (c=1, γ=0.3, ε=0.02) RMSE: 0.849 Acc: 73% Δ_{max}: 1.82 A_f: 1.14 B_f: 1.01	POR RMSE: 0.896 Acc: 68.6% Δ_{max}: 1.78 A_f: 1.16 B_f: 1.02	OLSR RMSE: 13428.148 Acc: 24% Δ_{max}: 59413.1 A_f: NaN B_f: NaN	
GCMS MRS	RFR RMSE: 0.388 Acc: 98.2% Δ_{max}: 1 A_f: 1.06 B_f: 1	PLSR RMSE: 0.504 Acc: 96% Δ_{max}: 1.34 A_f: 1.07 B_f: 1	SVR (c=4, γ=0, ε=0.02) RMSE: 0.543 Acc: 93.3% Δ_{max}: 1.24 A_f: 1.07 B_f: 1	KNN (k=8) RMSE: 0.579 Acc: 92.1% Δ_{max}: 1.4 A_f: 1.09 B_f: 1	POR RMSE: 0.617 Acc: 90.5% Δ_{max}: 1.31 A_f: 1.1 B_f: 1.01	OLSR RMSE: 4807.732 Acc: 35.7% Δ_{max}: 21273.07 A_f: NaN B_f: NaN	

GCMS OFC	RFR RMSE: 0.66 Acc: 89.8% Δ_{max}: 1.88 A_f: 1.1 B_f: 1.02	PLSR RMSE: 0.82 Acc: 81.1% Δ_{max}: 2.06 A_f: 1.14 B_f: 1.01	KNN (k=11) RMSE: 1.033 Acc: 74.8% Δ_{max}: 2.7 A_f: 1.17 B_f: 1.01	PCR RMSE: 1.048 Acc: 67.7% Δ_{max}: 2.33 A_f: 1.19 B_f: 1.02	SVR (c=4, γ=0, ε=0.02) RMSE: 1.06 Acc: 74.1% Δ_{max}: 3.01 A_f: 1.14 B_f: 1.01	OLSR RMSE: 9274.027 Acc: 28.2% Δ_{max}: 41033.4 A_f: NaN B_f: NaN	
GCMS PCA	RFR RMSE: 0.471 Acc: 95.7% Δ_{max}: 1.06 A_f: 1.05 B_f: 1	PLSR RMSE: 0.637 Acc: 87% Δ_{max}: 1.4 A_f: 1.08 B_f: 1	SVR (c=4, γ=0.03, ε=0.02) RMSE: 0.738 Acc: 82.2% Δ_{max}: 1.7 A_f: 1.08 B_f: 0.99	KNN (k=8) RMSE: 0.773 Acc: 79.2% Δ_{max}: 1.71 A_f: 1.09 B_f: 1	PCR RMSE: 0.904 Acc: 70.7% Δ_{max}: 1.91 A_f: 1.11 B_f: 1.02	OLSR RMSE: 5340.837 Acc: 27.6% Δ_{max}: 23629.92 A_f: NaN B_f: NaN	
FTIR VREG	RFR RMSE: 1.097 Acc: 64.5% Δ_{max}: 2.85 A_f: 1.18 B_f: 1.02	PLSR RMSE: 1.383 Acc: 50.6% Δ_{max}: 3.4 A_f: 1.24 B_f: 1.02	KNN (k=13) RMSE: 1.452 Acc: 43.9% Δ_{max}: 3.12 A_f: 1.26 B_f: 1.08	PCR RMSE: 1.477 Acc: 45.2% Δ_{max}: 3.42 A_f: 1.27 B_f: 1.05	SVR (c=1, γ=0, ε=0.02) RMSE: 1.496 Acc: 41.8% Δ_{max}: 3.37 A_f: 1.26 B_f: 1	OLSR RMSE: 20.463 Acc: 11.1% Δ_{max}: 52.08 A_f: NaN B_f: NaN	
FTIR STAA	RFR RMSE: 0.639 Acc: 85.9% Δ_{max}: 1.76 A_f: 1.09 B_f: 1	PLSR RMSE: 0.792 Acc: 81.8% Δ_{max}: 2.12 A_f: 1.11 B_f: 1.01	KNN (k=10) RMSE: 0.89 Acc: 78.6% Δ_{max}: 2.45 A_f: 1.13 B_f: 1.03	SVR (c=4, γ=0.6, ε=0.02) RMSE: 0.987 Acc: 62.8% Δ_{max}: 2.02 A_f: 1.16 B_f: 0.99	PCR RMSE: 1.032 Acc: 68.9% Δ_{max}: 2.48 A_f: 1.16 B_f: 1.03	OLSR RMSE: 9.608 Acc: 13.7% Δ_{max}: 26.18 A_f: NaN B_f: NaN	
FTIR MRS	RFR RMSE: 0.559 Acc: 92.3% Δ_{max}: 1.37 A_f: 1.08 B_f: 1	PLSR RMSE: 0.682 Acc: 86.7% Δ_{max}: 1.8 A_f: 1.1 B_f: 1.01	KNN (k=11) RMSE: 0.755 Acc: 78.2% Δ_{max}: 1.54 A_f: 1.11 B_f: 1.02	SVR (c=4, γ=0.6, ε=0.02) RMSE: 0.796 Acc: 71.9% Δ_{max}: 1.59 A_f: 1.12 B_f: 1	PCR RMSE: 0.817 Acc: 73.4% Δ_{max}: 1.68 A_f: 1.12 B_f: 1.01	OLSR RMSE: 10.322 Acc: 14.7% Δ_{max}: 26.01 A_f: NaN B_f: NaN	
FTIR OFC	RFR RMSE: 1.343 Acc: 54.1% Δ_{max}: 3.65 A_f: 1.19 B_f: 1.01	PLSR RMSE: 1.809 Acc: 36.8% Δ_{max}: 4.28 A_f: 1.27 B_f: 1.03	SVR (c=1, γ=0.03, ε=0.02) RMSE: 1.837 Acc: 40.3% Δ_{max}: 3.35 A_f: 1.28 B_f: 1	KNN (k=12) RMSE: 1.838 Acc: 28.8% Δ_{max}: 3.6 A_f: 1.29 B_f: 1.08	PCR RMSE: 1.869 Acc: 29.3% Δ_{max}: 3.59 A_f: 1.3 B_f: 1.05	OLSR RMSE: 20.401 Acc: 8.7% Δ_{max}: 55.37 A_f: NaN B_f: NaN	
FTIR PCA	RFR RMSE: 0.628 Acc: 89.5% Δ_{max}: 1.82 A_f: 1.06 B_f: 1	PLSR RMSE: 0.836 Acc: 76% Δ_{max}: 2.09 A_f: 1.09 B_f: 1	KNN (k=10) RMSE: 0.989 Acc: 69.4% Δ_{max}: 2.36 A_f: 1.11 B_f: 1.02	SVR (c=4, γ=0.6, ε=0.02) RMSE: 1.027 Acc: 60.8% Δ_{max}: 2.1 A_f: 1.12 B_f: 1	PCR RMSE: 1.086 Acc: 60.2% Δ_{max}: 2.32 A_f: 1.12 B_f: 1.01	OLSR RMSE: 7.857 Acc: 16.5% Δ_{max}: 21.67 A_f: NaN B_f: NaN	

Supplementary material for Section 4.1 –S4



APPENDIX II

Other publications

Trends in Food Science & Technology 50 (2016) 11–25

Contents lists available at ScienceDirect

Trends in Food Science & Technology

journal homepage: http://www.journals.elsevier.com/trends-in-food-science-and-technology

Review

Data mining derived from food analyses using non-invasive/non-destructive analytical techniques; determination of food authenticity, quality & safety in tandem with computer science disciplines

A.I. Ropodi, E.Z. Panagou, G.-J.E. Nychas*

Laboratory of Microbiology and Biotechnology of Foods, Department of Food Science and Human Nutrition, Faculty of Foods, Biotechnology and Development, Agricultural University of Athens (AUA), Iera Odos 75, Athens 11855, Greece

ARTICLE INFO

Article history:
Received 19 October 2015
Received in revised form
14 January 2016
Accepted 18 January 2016
Available online 23 January 2016

Keywords:
Spectroscopy
Hyper/multispectral imaging
e-nose
Chemometrics
Machine learning

ABSTRACT

Background: Food quality, safety and authenticity are important issues for consumers, governments, as well as the food industry. In the last decade, several researchers have attempted to go beyond traditional microbiological, DNA-based and other methods using rapid techniques. This broad term involves a variety of sensors such as hyperspectral and multispectral imaging, vibrational spectroscopy, as well as biomimetic receptors.

Scope and approach: The resulting data acquired from the above-mentioned sensors require the application of various case-specific data analysis methods for the purpose of simple understanding and visualization of the acquired high-dimensional dataset, but also for classification and prediction purposes.

Key findings and conclusions: It is evident that rapid techniques coupled with data analysis methods have given promising results in several food products with various sensors. Additionally there are several applications, new sensors and new algorithms that remain to be explored and validated in the future.

1. Introduction

At the dawn of the 21st century in an environment of tremendous technological progress and evolution of consumer life-styles, but also of economic problems, the European food industry is called to operate under seemingly contradictory market demands. While in general highly advanced technologies are rapidly being accepted and absorbed, the position of consumers with regard to their expectations for food products is often ambivalent. They seek food products of enhanced sensory quality, increased functional and nutritional properties combined with a traditional, wholesome image, as well as guaranteed safety but yet inexpensive foods with less processing, fewer additives and "technological" interventions. At the same time they expect extended shelf life and convenience in preparation and use.

To remedy this, the food industry and other stakeholders (e.g. Authorities from USA, EU and elsewhere, retailers) need to support development and application of effective quality and safety assurance systems based on controlling, monitoring, and recording the critical parameters throughout the food chain. This cannot be implemented with conventional microbiology (e.g. colony counting methods) or molecular based techniques that are considered more reliable and accurate (Velusamy, Arshak, Korostynska, Oliwa, & Adley, 2010). The fact that both are time-consuming, destructive and require highly trained personnel limits their potential to be used on-, in- or at-line (Nychas, Skandamis, Tassou, & Koutsoumanis, 2008; Papadopoulou, Panagou, Tassou, & Nychas, 2011). Furthermore, in the case of molecular tools, results may be misleading, as these techniques are focused so far on pathogenic rather than specific groups of the microbial association which contributes to spoilage and depends on storage and packaging conditions (Doulgeraki, Ercolini, Villani, & Nychas, 2012). This molecular approach is also costly, as high-tech instruments are required. In addition, due to the complexity of molecular techniques, the number of verified samples/measurements in many cases is severely limited. Therefore, efforts have been made to replace both conventional and molecular microbiological analyses with detection of biochemical changes occurring in food that could be used to assess food spoilage or safety.

Recently, some promising analytical approaches are being

* Corresponding author.
E-mail address: gjn@aua.gr (G.-J.E. Nychas).

http://dx.doi.org/10.1016/j.tifs.2016.01.011
0924-2244/

forwarded for the rapid and quantitative monitoring of food quality/safety. These approaches are based on (i) arrays of biomimetic sensors, (ii) vibrational spectroscopy (Fourier transform infrared -FT-IR, Raman) and (iii) surface chemistry (hyper/multispectral imaging) (Argyri, Panagou, & Nychas, 2014; Kamruzzaman, Maniko, & Oshita, 2014; Loutfi, Coradeschi, Mani, Shankar, & Rayappan, 2015; Nunes, 2014; Qin, Chao, Kim, Lu, & Burks, 2013; Sun, 2009; Teena, Manickavasagan, Mothershaw, El Hadi, & Jayas, 2013; Velusamy et al., 2010; Xiong, Sun, Zeng, & Xie, 2014). The usage of these sensors is based mainly on the principle that by-products of either metabolic activity of microorganisms, and/or the product's history itself and its origin display different biochemical profiles resulting in a characteristic sensor fingerprint that could be used for quality/safety evaluation (Ellis & Goodacre, 2001; Nychas et al., 2008). Besides, these techniques have several advantages compared to traditional methods since they are direct and non-invasive, require a minimum amount of sample or can be used on line. However, the enormous amount of data generated by this technology is very difficult to interpret. Indeed, although these rapid techniques are either non-invasive or require a minimum amount of sample, the acquired data are more complex and their analysis demands a multi-disciplinary approach. Depending on the type of sensor and data complexity, analysis involves disciplines such as computer vision/image processing, signal processing, statistical analysis/chemometrics, machine learning, computational intelligence techniques, etc. (Goodacre, 2003; Haralick & Shapiro, 1991; Hastie, Tibshirani, & Friedman, 2009; Liu, Wang, Wang, & Li, 2013; Loutfi et al., 2015; Marini, 2009; Nunes, 2014; Oliveri & Downey, 2012).

There are currently some reviews available on rapid methods applied in food commodities; however, they usually focus on a specific instrument, type of food, and application or data analysis methodology (Cheng & Sun, 2014; Cozzolino, Roumeliotis, & Eglinton, 2014; Domingo, Tirelli, Nunes, Guerreiro, & Pinto, 2014; Duchesne, Liu, & MacGregor, 2012; Kamruzzaman, Makino, & Oshita, 2015; Loutfi et al., 2015; Marini, 2009; Sankaran, Khot, & Panigrahi, 2012; Teena et al., 2013; Woodcock, O'Donnell, & Downey, 2008; Zhang et al., 2014). This article will focus on (1) data mining and data analysis derived from non-destructive, non-invasive instruments reported in the literature in recent years, regardless of sensor, food type and application, and (2) the aspects related to the implementation of these techniques to the food industry and other stakeholders, such as food managers who are not familiar with these approaches.

2. Instrumental techniques and resulting data

Data mining and data analysis from sensors, based on spectroscopy, hyperspectral and/or multispectral imaging, has been included in this review and the basic principles of each type of sensor and resulting data will be described briefly in this section.

2.1. Vibrational spectroscopy

Vibrational Spectroscopy (***VS***) is a collective term used to describe two analytical techniques infrared and Raman spectroscopy. Infrared (IR) and Raman spectroscopy are non-destructive, non-invasive tools that provide information about the molecular composition, structure and interactions within a sample. These techniques measure vibrational energy levels, which are associated with the chemical bonds in the sample. The sample spectrum is characteristic, like a fingerprint, and vibrational spectroscopy is used for identification, characterization, structure elucidation, reaction monitoring, quality control, and quality assurance. Infrared and Raman spectroscopy provide complementary information about molecular structure. It is important to stress that IR and Raman spectroscopy have found a wide variety of applications in food quality analysis. In the case of IR, several studies have been applied in the evaluation of food spoilage, including animal origin foods such as meat, milk and cheese, as well as plant origin foods like wheat, fruit spirits and beer (Argyri et al., 2014; Damez & Clerjon, 2013). On the other hand, the studies reported for evaluating food spoilage through the use of Raman spectroscopy are rather limited, including food products such as meat and milk.

The resulting data for both types of spectroscopy are two dimensional, i.e. spectra consisting of a wavenumber and the value of the measured parameter. Fourier Transformation (FT-IR spectroscopy) may be used before exploiting the spectra. In Fig. 1(a) and (b), examples of FT-IR and Raman spectra of minced beef samples are presented, where for a specific wavenumber a single value of absorbance and intensity is acquired. Several methods for pre-processing before data analysis, and their combinations, have been proposed, as they are affected by noise and sometimes display baseline and scatter effects (Engel et al., 2013; Jarvis & Goodacre, 2005). While pre-processing methods tend to be case-specific, some of these methods include first or second derivatives, Wavelet Transform (WT), detrending, Multiplicative Scatter Correction (MSC) or its extended form (EMSC), and Standard Normal Variate (SNV) transformation (Argyri et al., 2014; Engel et al., 2013; Jarvis & Goodacre, 2005). In the available literature, usually one or more pre-processing methods are applied and -depending on the application-different methods provide different results (Biancolillo, Bucci, Magrì, Magrì, & Marini, 2014; Coppa et al., 2014). Lastly, a genetic algorithm was employed in order to decide the appropriate method or combination of methods for FT-IR spectra pretreatment (Jarvis & Goodacre, 2005).

2.2. Multispectral (MSI) and hyperspectral imaging (HSI)

Hyperspectral and Multispectral imaging (***HSI*** **and** ***MSI***) is a technique with which both spectral and spatial information from chemical targets can be obtained. This chemical imaging technique combines *Vibrational spectroscopy*, and *Computer vision*. The former technology is an optical technology that depends on the interaction between incident light and molecules in matter. As spectrometers analyze only a small portion of the food sample (therefore, the spectra, strictly speaking, are sometimes not representative of the whole sample), technologies that take into account the whole sample or a large part of it may provide more representative and detailed measurements. In order to obtain spatial information, another technology, namely, *Computer Vision* is available. This discipline imitates the principle of human vision, using three bands (red, green and blue) to acquire the characteristics of objects. Working in the visible range, the features obtained by computer vision include shape, color, size, and texture. However, only occasionally is this method reported to be sufficient for detecting chemical and biological parameters. Both spectroscopy and computer vision techniques have found a wide range of applications in the food industry. However, both techniques have their own disadvantages. The merits in spectroscopy and computer vision are both combined in ***hyperspectral imaging***, which can also be used to generate *chemical maps* to show distributions of parameters of interest. However, the rich information in hyperspectral imaging also results in difficulties in data processing, which makes it hard for industrial online applications. To overcome this problem, a simplified version called ***multispectral imaging*** (MSI) is available. The difference between the two lies only in the number of bands involved. For HSI, there are normally more than 100 bands, while for MSI, it is usually less than 20. The success of MSI deeply relies on the efficiency of HSI for providing the important wavelengths. In

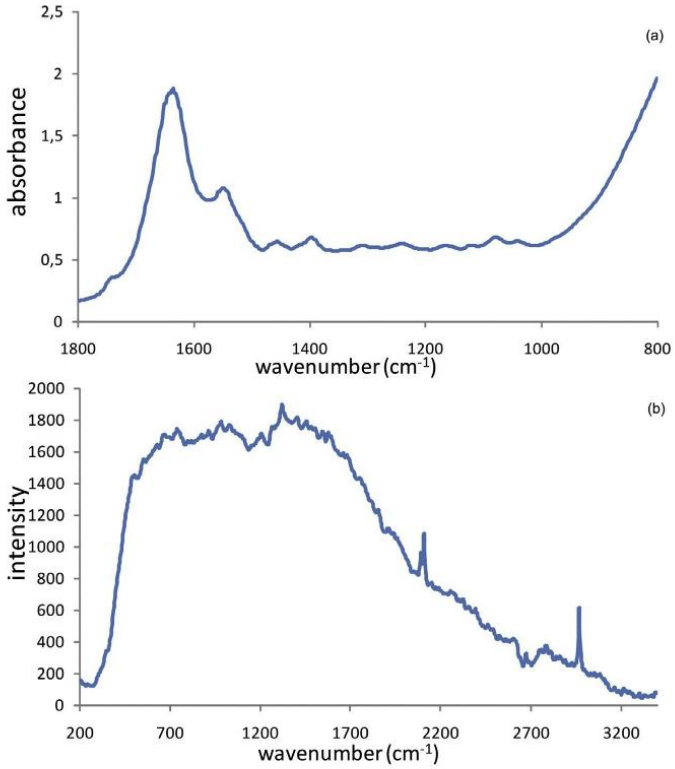


Fig. 1. Typical FT-IR and Raman spectra of minced beef samples.

hyperspectral imaging, several choices can be taken to carry out the detection work. These options involve near-infrared HSI, fluorescence HSI, as well as Raman HSI, which provide great flexibility in finding solutions for all sorts of detection problems. In other words, the versatility of hyperspectral imaging has built up its wide applications in food inspection. Several review papers have been published regarding the applicability of hyperspectral imaging in different aspects of food quality and safety (Gowen, Feng, Gaston, & Valdramidis, 2015; Liu, Zeng, & Sun, 2015).

The data captured is a three-dimensional (3D) image cube, providing both spatial and spectral information, as shown in Fig. 2. "Spatial" information implies pixel intensity when choosing a specific wavelength and "spectral" information is each pixel's intensity at different bands. It is evident that this type of data is extremely informative, but also very complex. Furthermore, not only are they susceptible to noise and other artifacts, but in the case of HSI the spectral bands are highly collinear and have redundant information. Therefore, an image-processing step on the spectral and spatial level is necessary for analysis. Pre-processing involves removal of noise, blur and geometric distortions, radiometrics, spectral axis calibrations, etc. Noise in the spectra may be decreased using median filtering and Savitzky Golay (SG) smoothing, which can also be applied in the spatial domain. Also, techniques mentioned in section 2.1 may be used to overcome unwanted spectral variation due to the natural morphology of food samples and/or non-uniform lighting (Gowen et al., 2015; Kamruzzaman,

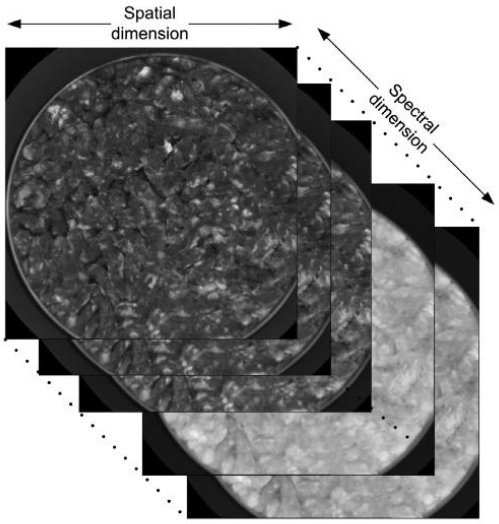


Fig. 2. Example of multispectral image hypercube of minced meat presentation of spatial vs. spectral dimensions.

Barbin, ElMasry, Sun, & Allen, 2012). Additionally, an image segmentation step may be necessary in order to remove unwanted areas of an image, e.g. to remove image background and select a region of interest (ROI) (Teena et al., 2013). From then on, basic features are extracted for further analysis, using simple statistical measures, e.g. mean pixel intensity per wavelength (Panagou, Papadopoulou, Carstensen, & Nychas, 2014; Ropodi, Pavlidis, Mohareb, Panagou, & Nychas, 2015), or more complex data, such as textural characteristics (Duchesne et al., 2012; Haralick, Shanmugam, & Dinstein, 1973).

2.3. Biomimetic sensors

While trained sensory panels are commonly employed in quality evaluation, this approach has some drawbacks, including discrepancy that might occur due to human fatigue or stress. Furthermore, sensory panels can be expensive and cannot be used for large-scale monitoring. A solution can be found in biomimetic systems, which are nature inspired arrays of sensors designed to mimic the olfactory and gustatory systems of humans, called electronic nose (e-nose) and e-tongue, respectively (Ghasemi-Varnamkhasti, Mohtasebi, & Siadat, 2010).

Recognition of odor in humans occurs when there is a chemical interaction between volatile odor compounds and receptors in the nasal cavity and the signals are then transferred to the brain through neurons. In accordance to this principle, the chemical sensors of an e-nose resemble the primary neurons with different sensitivity to different odors. Chemical interaction between odor compounds and the sensors, give rise to electrical signals recorded by the instrument. Based on the same idea, tastes are sensed from different, discrete regions on the tongue including specific receptors. E-tongue is an analytical tool consisting of an array of non-specific, low selective chemical sensors with partial specificity (cross-sensitivity) to different components in a solution (Ghasemi-Varnamkhasti et al., 2010; Loutfi et al., 2015).

Unlike traditional analytical methods, biomimetic sensor arrays do not obtain information on the nature of the compounds under consideration, but only present a digital fingerprint of the food material, meaning the signals from the individual sensors involved in the system show a pattern which is thereafter, analyzed by data analysis tools. Nowadays, e-noses are more often used in food science compared to e-tongues. A variety of different sensor types have been developed, to which three types of materials are commonly used: metal oxides, conducting polymers composites and intrinsically conducting polymers. However, these sensors also present a number of shortcomings that have to be taken into account during data analysis, including the stability of sensor behavior meaning the reproducibility of a sensor response, sensor drift, sensitivity and selectivity. Regardless of these drawbacks, various applications have been presented in milk, fish, meat, tea and coffee as well as fruit concerning freshness, shelf-life, ripeness, quality grading etc. (Baietto & Wilson, 2015; Baldwin, Bai, Plotto, & Dea, 2011; Loutfi et al., 2015; Wilson & Baietto, 2009; Wilson, 2013).

3. Chemometrics, machine learning & evolutionary computational methods

Acquiring data from one or more of the previously introduced sensors results in multivariate datasets, i.e. a large number of variables (x-data) connected to an observed value or category (y-data). Therefore, the resulting datasets display high complexity and are obviously difficult to inspect and visualize. Chemometrics, machine learning and evolutionary computational methods contribute in visualization, dimensionality reduction and presentation of each variable's contribution to the final result. Chemometrics is the science of extracting information from chemical systems by data-driven means and it was developed as consequence of the computer application in chemistry while an actual definition of chemometrics is "the chemical discipline that uses mathematical and statistical methods, to (i) design or select optimal measurement procedures and experiments, and (ii) provide maximum chemical information by analyzing chemical or signal data generated by modern analytical instrumentation" (Otto, 2007).

Since the use of computers, either personal or those linked with analytical instruments, by scientists was inevitable, it provided a new dimension for the acquisition, processing and interpretation of (chemical) data derived from the continuously developed instrumentation. In fact, this data flood is another yet reason, that analytical chemists in particular develop applications of chemometric methods. As a consequence, the necessity emerges for a deeper understanding of those methods. It should be noted that the education in mathematics and statistics in most academic disciplines is usually unsatisfactory mainly due to lack of practice. Therefore, one of the initial aims of chemometrics was to make complicated mathematical methods practicable. Meanwhile, the commercialized statistical and numerical software simplifies this process, so that all important chemometric methods can be taught in appropriate computer demonstrations (Brereton, 2014). In general, chemometrics is a highly interfacial discipline, using methods frequently employed in core data-analytic disciplines, such as multivariate statistics, applied mathematics, and computer science, to address problems not only in chemistry, but also to biochemistry, medicine, biology, chemical engineering and in this case-food science.

Machine learning is the scientific discipline that concentrates on the development and application of algorithms that are designed to "learn" from data. Specifically, in machine learning example inputs are used to construct a model in order to make predictions and/or decisions (Kohavi & Provost, 1998). As such, it is a subfield of computer science, employing a range of computing tasks that can go far beyond explicit programming and overlaps with computational intelligence. Furthermore, it is connected with fields such as statistics and mathematics, which provide a major part of the theory behind the constructed models. It is also a great tool for exploratory data analysis and has been the basis of various applications/subfields, e.g. computer vision and optical character recognition (OCR).

Evolutionary computation is another subfield of computational intelligence in which the main focus of the applied algorithms is the development of an optimization process though an iterative procedure based on the idea of evolution and natural genetics. In detail, its algorithms deal with the iterative growth of a population inspired by the mechanisms of evolution and a careful selection process. The procedure is repeated numerous times based on stopping criteria. Evolutionary computation, while computationally expensive, provides highly optimized processes and has many applications in diverse fields, including food science. Some of these algorithms include genetic algorithms (GAs), evolution strategies, evolutionary programming, and genetic programming (GP) (Argyri et al., 2013).

All chemometric methods can be included in two major categories of learning techniques, supervised and unsupervised. Reinforcement learning is another new category where the model interacts in the context of the dynamic environment and improves its performance based on goal without explicitly "knowing" if the goal is achieved (Sutton & Barto, 1998). However, this approach has not been used in the food industry yet and thus will not be discussed. Unsupervised learning is based on cluster analysis (Everitt, Landau, Leese, & Stahl, 2011) and focuses on finding how similar is one sample to another. On the other hand, supervised learning

methods try to model and/or map the input variables (x-data) based on the output (y-data) (Goodacre, 2003). Whereas unsupervised techniques focus mostly in grouping samples/distinguish objects and populations unrelated to the target output, supervised techniques -both qualitative and quantitative-tend to work well, as they take into account the actual knowledge of the target output. The downside to the latter techniques is that they tend to be more complex, as one or more parameters have to be adjusted based on metrics such as the Root Mean Squared Error (RMSE). This procedure is called calibration or training and may be simple or highly complicated depending on the method. The resulting model can be used to predict the category or the output value of a new sample.

3.1. Unsupervised methods

Among the unsupervised methods, Hierarchical Cluster Analysis (HCA) and Principal Component Analysis (PCA) have been extensively used in food applications (Argyri et al., 2014; Everitt et al., 2011; Jolliffe, 2002). In HCA, the distance (Euclidean or other) between objects (samples) is calculated and is further processed with an agglomerative distance algorithm in order to construct a dendrogram that combines the samples according to their similarities to each other (Everitt et al., 2011). On the other hand, PCA is a widely used method that enables a significant reduction of dimensionality while at the same time choosing uncorrelated variables (called Principal Components PCs) explaining most of the variance in the dataset. The main advantage of this method is that it can "translate" a multivariate dataset into a dataset of a few Principal Components and therefore help visualize the samples, e.g. PCA scores plot groups samples based on their similarities or dissimilarities usually using a 2D or 3D projection of the samples. Because of this ability, PCA is usually the first analysis performed when presented with this type of data. Furthermore, it can be used in outlier detection and highlighting important features to be used for further analysis, a process known as variable selection (Argyri et al., 2014; Goodacre, 2003). In Fig. 3, PCA is performed on the MSI data from samples of minced beef adulterated with pork (Ropodi et al., 2015), where samples named with "0" correspond to pure pork, all other categories correspond to the percentage (w/w) of beef in the mix and consequently "100" refers to pure beef samples. It is evident that PCA manages to visualize a multivariate dataset (a total of 36 variables in this case) in a 2-D score plot, showing a good discrimination among samples with different levels of adulteration along the axis of the first PC (PC1).

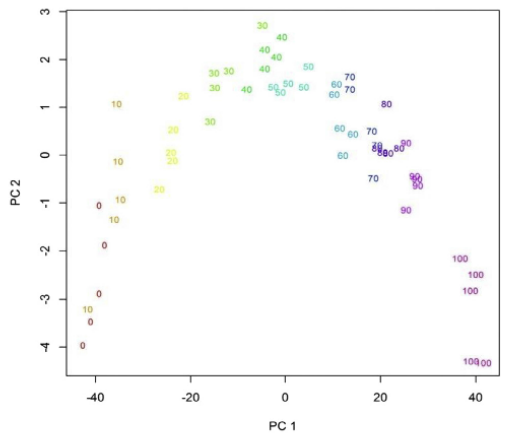


Fig. 3. PCA scores for minced beef samples adulterated with pork based on MSI data. Samples named with "0" correspond to pure pork, all other categories correspond to the percentage of beef in the mix and consequently "100" refers to pure beef samples.

3.2. Supervised methods

Depending on the nature of the problem there are discriminant and regression algorithms for qualitative and quantitative approaches, respectively. Discriminant Function Analysis (DFA) or Canonical Variate Analysis (CVA) is a cluster analysis based method that is used to classify individuals into two or more predetermined groups. It can be considered as a posterior procedure of multivariate analysis of variance (Berrueta, Alonso-Salces, & Héberger, 2007). Linear regression methods are extensively used in food applications, as they are relatively easy to apply. These methods include Multiple Linear Regression (MLR), Principal Component Regression (PCR) and Partial Least Squares Regression (PLSR) (Goodacre, 2003). MLR is the simplest form of linear regression, based on least squares. However, it does not do well with colinearity. PLSR (de Jong, 1993; Wold, Sjöström, & Eriksson, 2001), like PCR, is a linear method in the sense that the new components are linear combinations of the original variables (Boulesteix & Strimmer, 2007). While PCR uses PCA to extract the new components (principal components-PCs), PLSR projects both observed and predicted values in a feature space and a linear regression model is established. PLS can be combined with various variable selection algorithms and several of its variations have been modified to accommodate interval selection methodologies, e.g. interval partial least squares (iPLS), windows PLS and iterative PLS (Xiaobo, Jiewen, Povey, Holmes, & Hanpin, 2010). Furthermore, coupled with Linear Discriminant Analysis (LDA), it can be extended to PLS-DA for classification purposes (Barker & Rayens, 2003).

3.3. Machine learning/evolutionary computation

While the abovementioned methods have received great attention, lately methods based on more complicated algorithms involving machine learning and computational intelligence have been introduced. These terms apply to heterogeneous algorithms, such as neural networks, fuzzy systems, rough set, evolutionary computation, swarm intelligence, probabilistic reasoning, multiagent systems, etc. Very common computational intelligence methodologies found in research articles and reviews are Artificial Neural Networks (ANNs), Support Vector Machines (SVMs) and evolutionary-based algorithms, including Genetic Algorithms (GAs) and Genetic Programming (GP). The latter algorithms are used for optimization purposes.

ANNs are inspired from the biological paradigms of the human brain and the function of neurons. Depending on the type of network they can be used for both classification and regression purposes. The building block of every neural network is a neuron that has a number of inputs. Each of these inputs is multiplied by a connection weight and in the simplest case, these products are simply added together, fed through a transfer function to generate an output. The transfer functions that are commonly supported are sigmoid, sine, etc. As there are various ways that these neurons can be clustered together, understanding of the diverse network topologies, learning rules, transfer functions, summation functions is very important (Sumathi & Surekha, 2010). The most commonly applied ANNs are Multilayer Perceptrons (MLPs) and Radial Basis Function (RBF).

SVMs (Cortes & Vapnik, 1995) map the original data points from the input space into a higher dimensional feature space using a kernel function, in order to construct a maximal separating hyper-

plane (SP). While there are various choices for the kernel functions, namely linear, polynomial, sigmoid, and radial basis function (RBF), RBF is among the most common kernel functions. There are also SVM regression (SVM-R) models for value estimation purposes (Balabin & Lomakina, 2011). For a given regression problem, the goal of SVM is to find the optimal hyper-plane from which the distance to all the data points is minimum. The performance of SVMs depends on several factors including the kernel function type and its corresponding hyperparameter(s). The optimization of these parameters can be complex and computationally intensive.

Lastly, the evolutionary-based algorithms are iterative processes resembling growth or development in a population. The optimal solution is then obtained using parallel processing among the population. GAs use Darwin's "survival of the fittest" strategy and reproduction operators to select the optimal set of features or parameters. However, GAs do not really solve any problem. A given GA can be used for feature selection and/or the optimal adjustment of parameters, but these features or parameters are evaluated by a predetermined function, e.g. using a PLS method (Luke, 2003).

3.4. Model calibration, parameter estimation and validation

Supervised -both qualitative and quantitative-techniques take the actual response (y-data) into consideration. At the same time, however, these techniques are more complex as a number of parameters -depending on the method-have to be estimated by the data analyst. Specifically, in the case of PCR and PLSR the optimum number of components has to be decided and used for regression. A simple criterion could be the percentage of variance explained especially in the case of principal components. The decision is usually based on simple criteria, such as minimizing the Mean Squared Error (MSE) or Root Mean Squared Error (RMSE) which are based on the residual values (Residual = observed − estimated value). Regression is performed for a range of different number of components and the decision is made based on the chosen criterion. Similarly, in the case of PLS-DA and other qualitative methods, criteria such as correct classification, sensitivity, specificity and F-score can be applied (Sokolova & Lapalme, 2009).

While the same criteria can be used for model validation, the appropriate validation strategy may vary depending on the actual dataset, its size, distribution of samples or possible stratification based on background information. Specifically, a model based on a small number or samples is very sensitive to random results (Nunes, Alvarenga, de Souza Sant'Ana, de Sousa Santos, & Granato, 2015). According to Westad and Marini (2015), while it is encouraged to further validate models using a test set, occasionally this is not always possible, since the number of training samples must be enough for the development of a robust model and the test samples should be representative of the obtained results. As a rule of thumb, **cross-validation** (CV) is used if the number of samples is smaller than 40. They also emphasize the importance of CV as a method of understanding the inherent structure of the dataset/observed data and help in the assessment of dimensionality of a model (rank) and therefore CV is mostly applied for calibration. In Fig. 4(a), an example of variance explained depending on the number of components is presented. While using a large number or all the components would seem logical in order to create a model based on the observed data, this is not the case. Except for the obvious dimensionality issues leading to computationally demanding calculations, it could lead to overfitting. In simple terms, **overfitting** occurs when a model is overly-optimistic on the data used for calibration and lacks ability to generalize on other data. Indeed, as more components are used, the percentage of variance explained increases also but by a very small margin (Fig. 4a); however, the MSE of CV does not necessarily decrease as illustrated in Fig. 4b. A good choice for the number of components would be the number where the minimum MSECV is achieved. Several approaches for cross-validation have been applied in the past, the most common being Leave-One-Out (LOOCV) and k-fold cross-validation. The first method leaves one sample out of the calibration process, which is used for validation. Then, this process is performed for all samples in the dataset. In the second method, the calibration data are separated in k equal subsets. Then each subset is used iteratively for testing the model based on the rest of the data. In fact, LOOCV is a common method as all samples are used in an exhaustive way providing repeatability of the results compared with other random methods of partitioning of the training dataset. Furthermore, it can also be applied in smaller datasets where a k-fold partitioning or a random partitioning could result in folds where all categories/types of samples are not represented in an equal manner. Indeed, an important parameter for the selection of CV strategy is whether the data can be categorized into groups based on background information, i.e. stratification across replicates, treatment, etc. In the case of LOOCV, using one sample for testing and all other samples for training each time may lead to over-optimistic results. In the interest of equal category representation stratified partitioning or -given no stratification Leave-Multiple-Out CV is preferred (Westad & Marini, 2015). In short, cross-validation may contribute in avoiding overfitting by not using the same dataset to both model building and prediction error estimation. If possible a representative percentage of the original dataset is left out for **testing/validation,** whereas for further validation purposes, independent data can be used as **external validation**.

Training/test split is an equally important matter for a validation scheme. While a random partition can be applied, it can lead to high variance of model estimates when not repeated a number of times. Other methods that span the sample space as uniformly as possible are also applied, e.g. the Kennard–Stone (KS) algorithm (Kamruzzaman et al., 2012). Other methods include the Duplex algorithm, selection based on D-optimal criterion or even use of clustering techniques like k-means and Kohonen mapping. Several of these schemes were compared using NIR spectra of oat flour samples by Westad and Marini (2015). On one hand, KS and D-optimal approaches tended to capture as much as possible of the sample diversity in the training set. On the other hand, Kohonen and mostly Duplex maintained the same diversity among both sets.

The same principles may apply in the case of machine learning methodologies, but the estimation of the hyperparameters can be even more complex relying significantly on the experience of the analyst and also a trial-and-error process. In the case of SVMs, for example, depending on the kernel function chosen, different- and more than one-hyperparameters have to be calculated. In most cases, a grid-search coupled with a CV methodology is used for model training. Configuring ANNs is also a very complex process as it depends on the type of ANN, its architecture, initialization, transfer functions in both hidden and output layers, the learning algorithm employed, the learning rate, etc. While Back-Propagation Neural networks (BPNN), also called MLPs, are simpler and more commonly used, they have been accused of poor generalization ability and difficulty to control the training process (whether the model converges and the speed of convergence). On the other hand, ANNs have been used successfully in several applications including foodstuff analyses (Marini, 2009; Pan, Zhang, Zhu, Mao, & Tu, 2014; Panagou, Sahgal, Magan, & Nychas, 2008). GAs have also been applied for ANN configuration in the past (Argyri et al., 2013). Lastly, evolutionary programming requires defining several other parameters and/or criteria such as the number of generations, population size, probability of mutation/crossover, etc (Ellis, Broadhurst, Clarke, & Goodacre, 2005).

We can safely conclude that a successful model is highly

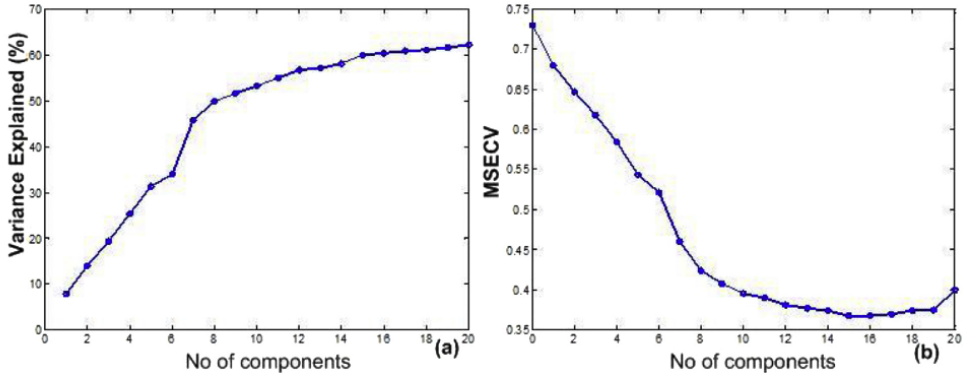


Fig. 4. Examples of percentage of variance explained in response data for up to 20 PLS components and the corresponding MSE values (log CFU/g)^2 after 10-fold cross-validation (unpublished multispectral minced beef data).

dependent on the type, size and quality of the initial data. It is self-evident that the dataset has to provide a representative number of samples for all cases/parameters investigated in order to take into account the variability within the same category and/or conditions. A good **model performance** during validation and/or CV ensures the model's ability to generalize. In Table 1, the most common performance criteria and their definitions are presented for regression and classification models, respectively, as described in various articles (Argyri, Panagou, Tarantilis, Polysiou, & Nychas, 2010; Ross, 1996; Sokolova & Lapalme, 2009). Usually, the performance of a calibration model is evaluated in terms of root mean squared error (RMSE) of calibration (RMSEC), cross-validation (RMSECV) and prediction (RMSEP). In addition, several other criteria such as the bias (B_f) and accuracy (A_f) factors may be employed. In the case of classification, the percentage of correctly classified samples, recall/sensitivity, precision and specificity are usually employed.

In Table 2, algorithms and calibration methods are presented for various sensor applications. LOOCV is the more common method used for cross validation, as it can be applied in small datasets and uses all samples in an exhaustive way. Panagou et al. (2014) used the Predicted Residuals Sum of Squares (PRESS), the model was validated with completely independent samples from two separate experiments/meat batches and at different temperatures with good results. Further on, da Costa Filho (2014) also used PRESS as a PLSR calibration criterion, where a component was included only if it improved PRESS by at least 2%, for the determination of *trans* fatty acids (TFAs) in edible oils. The models exhibited coefficient of correlation >0.982. RMSEP was also utilized for GA-MLR algorithm for the discrimination among chicken and turkey samples (Ellis et al., 2005). Other researchers (Alamprese, Casale, Sinelli, Lanteri, & Casiraghi, 2013; Ropodi et al., 2015) applied 5-fold and 3-fold cross-validation in PLS-DA, respectively, while both models were further validated with independent meat batches. This is important as the variability among samples of the same meat batch and between batches is included in model development and validation. In the case of SVM models (classification or regression), a grid-search methodology coupled with CV was employed (Argyri et al., 2013; He, Wu, & Sun, 2014; Qiu, Wang, & Gao, 2014; Wu, Shi, He, Yu, & Bao, 2013). Moreover, Papadopoulou, Panagou, Mohareb, and Nychas (2013) created three random partitions for SVM calibration and validation of 177 samples of beef fillets with the validation set consisting of 59 measurements and SVM calibration was performed with grid search and 10-fold CV., whereas Sharifzadeh, Clemmensen, Borggaard, Støier, and Ersbøll (2014) applied 25 random partitions for CV. PLSR and SVR calibration was performed based on RMSE of LOOCV by Argyri et al. (2013). However, a small validation set was employed for GP and GA-ANN calibrations, where the best model was based on the R^2 and MSE criterion, respectively. Lastly, after LOOCV, Zhao, Downey, and O'Donnell (2014) selected models based on a number of criteria including the position of the first local minimum in the leverage and residual X-variance plots, the correct classification rate and the number of loadings or components involved. The PLS-DA model performance was calculated based on Efficiency (geometrical mean of sensitivity and specificity).

Table 1
Formulas of popular performance measures for regression and binary classification (Argyri et al., 2010; Ross, 1996; Sokolova & Lapalme, 2009).

Performance measure	Formula
Regression[a]	
Bias factor B_f	$10^{\frac{\sum \log(y_{pred}/y_{obs})}{n}}$
Accuracy factor A_f	$10^{\frac{\sum \lvert \log(y_{pred}/y_{obs}) \rvert}{n}}$
Mean relative percentage residual MRPR (%)	$\frac{1}{n}\sum \frac{100(y_{obs}-y_{pred})}{y_{obs}}$
Mean absolute percentage residual MAPR (%)	$\frac{1}{n}\sum \frac{100\lvert(y_{obs}-y_{pred})\rvert}{y_{obs}}$
Root mean squared error RMSE	$\sqrt{\frac{\sum (y_{obs}-y_{pred})^2}{n}}$
Standard error of prediction SEP (%)	$\frac{100}{y_{obs,mean}}\sqrt{\frac{\sum (y_{obs}-y_{pred})^2}{n}}$
Classification[b]	
Accuracy	$\frac{t_p+t_n}{t_p+f_n+f_p+t_n}$
Precision	$\frac{t_p}{t_p+f_p}$
Recall (sensitivity)	$\frac{t_p}{t_p+f_n}$
Specificity	$\frac{t_n}{f_p+t_n}$

[a] n: number of samples, y_{pred}: predicted value, y_{obs}: observed value, $y_{obs,mean}$: mean observed value.
[b] t_p, t_n: positive and negative samples classified correctly, f_p : negative samples classified as positive, f_n: positive samples classified as negative.

4. Applications of instrumental techniques in the food industry

The multivariate data analysis techniques presented in the previous sections are powerful tools regardless of the instrument

Table 2
List of representative calibration and testing techniques in various sensor-based food applications.

Reference	Data analysis method	Number of samples	Model calibration	Validation	External validation	Variable selection	Preprocessing
(Panagou et al., 2014)	HCA, PLSR, PLS-DA	106	86 PLSR,PLS-DA: PRESS for LOOCV	No	20 (two independent experiments/ different temperatures)	modified Jack- knife estimation procedure	SNV
(He, Sun, et al., 2014)	LS-SVM (RBF kernel), MLR	118	79 grid search using RMSE, LOOCV	39	No	competitive adaptive reweighted sampling (CARS) algorithm	
(Gowen et al., 2008)	PCA (application to hypercube data)	172	100	No	72		MSC, MaxN, MedN, MeanN
(Sharifzadeh et al., 2014)	PCA vs. linear, non-linear and kernel-based regression methods (ANNs & SVM)	52	52 25 random partitions (38 14)	No	No		
(Ropodi et al., 2015)	HCA, PCA, LDA, PLS-DA	220	99 PLS-DA: % correct classification for 3-fold CV	66	55		
(Wu et al., 2013)	LS-SVM	160	100 grid search using RMSE, LOOCV	60	No	Successive projections algorithm (SPA) Uninformative variable elimination (UVE)	
(Argyri et al., 2013)	PLSR,GA-GP, GA-ANN, SVR (various kernel functions)	52 × several replicates (98 FTIR & 260 Raman Spectra)	50/98 FTIR spectra 130/260 Raman spectra PLS: RMSE - LOOCV SVR: grid search RMSE - LOOCV GP, GA-GP, GA-ANN: using validation subset, MSE/R^2	48/98 FTIR spectra 130/260 Raman spectra	No	GA	SNV
(Coppa et al., 2014)	PLS and modified PLS (MPLS)	250 (×6 spectra fresh & frozen)	200 critical values (Student's T) for removing outliers & two elimination passes, full CV (6 CV groups)	No	50	PCA (highest loadings)	No correction, SNV, detrend (D), standard normal variate and detrend (SNVD), and multiplicative scatter correction (MSC).
(da Costa Filho, 2014)	PLSR	171 (28 soybean oil, 33 sunflower oil, 38 rapeseed oil, 25 palm oil, 27 peanut oil, 11 edible oil from factories & 9 samples of oil/fat extracted from finished products	~90 PRESS improvement > 2%	50% per case	No	samples of edible oils collected from various regions and oil/fat extracted from finished product were exclusively applied to validate the calibration model.	Second derivative, smoothing
(Ellis et al., 2005)	PC-DFA, GA-D-MLR (discriminant multiple linear regression)	180 (chicken breast, chicken leg, turkey breast, turkey leg)	120 PC-DFA: 20 PCs (98.7% of explained variance) used in DFA with a priori knowledge of the muscle and species type	60	No	GA for most discriminatory wavelengths	
(Alamprese et al., 2013)	PCA, LDA, PLSR	242 22 per batch	176 PLSR: RMSE 5-fold CV	No	66 (3 independent batches)	algorithm SELECT	SNV, SNV & 1st derivative
(Zhao et al., 2014)	PLS-DA, SIMCA	164 (82 fresh & 82 frozen)	82 (41 per case) Efficiency (geometric mean of sensitivity & specificity)	82 (41 per case)	No	Martens Uncertainty Test	No, MSC, SNV, 1st & 2nd derivative
(Papadopoulou et al., 2013)	PC-DFA, SVM, SVR (RBF kernel))	177 (three random partitions for training/validation)	118 grid search with 10-fold CV	59	No	PCA prior to DFA	
(Panagou et al., 2008)	PCA, HCA, DFA, MLP-NN	78	78 LOOCV	No	No		Normalized based on min & max responses
(Pan et al., 2014)	PCA, MLP-NN	240	180 45/180for testing	60	No	Multiple comparison test (MCT) of one way ANOVA	
(Wilson et al., 2013)	PCA, QF, ANN	13 fillet samples (batches) with multiple measurements	10 batches	No	3 batches (70 samples)		Raw & normalized intensity values

Table 2 (*continued*)

Reference	Data analysis method	Number of samples	Model calibration	Validation	External validation	Variable selection	Preprocessing
(Zhang et al., 2012)	PCA, LDA, PCR, PLSR	90	60 PCR, PLSR: based on SEC & R indices	30	No		
(Zakaria et al., 2012)	PCA, LDA, LDA-ANN	480 samples	LDA:480 (LOOCV) LDA-ANN:120		LDA-ANN:360	Wilks' lamda test for sensor selection	Baseline manipulation & normalization, low level fusion
(Qiu et al., 2014)	LDA(stepwise), PLSR, SVM, RF	100 samples	75 samples LDA: stepwise using Wilks' lamda test, LOOCV, OCC PLSR: based on R2 & RMSE SVM, SVR: grid search, 5-fold CV	25	No	One-way ANOVA	Autoscaling

type or the application. There are several examples of different applications where the same or similar methods are used. Similarly, more than one data analysis method may be applied for the same purpose or dataset for comparison purposes.

All applications can be divided into two main groups: (a) food quality and safety assessment and (b) authenticity claims.

4.1. Quality and safety assessment

The first category involves a large number of characteristics, such as microbial quality, freshness and food spoilage, sensory characteristics, and other quality parameters (e.g. color, fatty acid composition, firmness, moisture content, pH, etc), as well as contamination, detection of mould, pathogens, toxins and specific harmful compounds. In Table 3, applications of various multivariate methods are presented grouped by sensor type.

4.1.1. Microbial/sensory quality

Vibrational spectroscopy, imaging, and e-nose have been used to predict microbial quality. Huang, Zhao, Chen, and Zhang (2014) attempted to measure total volatile basic nitrogen (TVB-N) content in pork meat integrating near infrared spectroscopy (NIRS), computer vision (CV), and electronic nose (e-nose) techniques and yielded a coefficient of determination equal to 0.9527 in the prediction set. A variety of methodologies was employed by Argyri et al. (2013) for the comparison of FT-IR and Raman spectroscopic techniques based on microbial and sensory data from minced beef samples. Analysis was performed using machine learning and evolutionary computing methods, including PLSR, GP, GA, ANNs and support vector machines regression (SVR) including linear, polynomial, radial basis and sigmoid kernel functions. Results showed that FT-IR models performed only slightly better. Furthermore, PLS and SVM models performed better in predicting microbial counts, but the GA-GP model was superior in predicting sensory scores using the FT-IR data, whereas the GA-ANN model performed better in predicting the sensory scores using the Raman data. Panagou et al. (2014) explored the potential of multispectral imaging in 18 wavelengths in the visible and shortwave near infrared area (405–970 nm) in assessing the microbial quality of beef fillets stored at different isothermal conditions. The authors employed HCA, PLSR for prediction and PLS-DA for discrimination among 3 microbial classes. Results were promising and were further validated with independent test samples stored at two new temperatures with satisfactory results. Additionally, hyperspectral imaging has also been used to evaluate surface spoilage of farmed salmon flesh by lactic acid bacteria (LAB) during cold storage applying the LS-SVM and Competitive adaptive reweighted sampling (CARS) algorithm in order to reduce spectral redundancy and identify the most informative wavelengths (MIWs) related with LAB prediction (He, Sun, & Wu, 2014). The derived LS-SVM model using 239 wavelengths yielded a regression coefficient of prediction R_p equal to 0.929 and RMSEP 0.515, very similar to the CARS-LS-SVM model using only 8 wavelengths (R_p = 0.925, RMSEP = 0.531). Microbial and sensory quality were also assessed using a portable e-nose and SVMs for regression and classification purposes (Papadopoulou et al., 2013). The resulting SVR model exhibited a mean correlation coefficient between observed and predicted counts in the test datasets equal to 0.863 for TVC, whereas in the case of the sensory categories fresh, semi-fresh, and spoiled, the average sensitivity values were 85.7, 87.3, and 88.9%, respectively. Lastly, e-nose has been applied for the discrimination of table olives' quality based on sensory score, using ANNs and specifically MLPs, where the classification accuracy for acceptable, marginal and unacceptable samples was ca. 90%, 78% and 52%, respectively (Panagou et al., 2008).

4.1.2. External features, bruise detection and other quality parameters

Imaging techniques have also been used in order to recognize and/or discriminate based on external features (i.e., shape, size, and color) in fruits and vegetables (Zhang et al., 2014) and for bruise detection in mushrooms (Gowen et al., 2008) by two classification methodologies based on PCA: (a) applying PCA to entire hypercube and classify based on the 2nd PC, and (b) multiplying hypercube by loading function and classify resultant virtual image. Results showed that compared with applying PCA to the entire hypercube, using virtual prediction image resulted in better classification in all cases except for the classification of one category of bruised samples (damaged by shaking at 400 rpm for 20 min), where the first and second method resulted in 100% and 90% correct classification, respectively. Furthermore, multispectral imaging has been suggested as a method for monitoring meat color instead of using RGB measurements based on L*, a*, b* (Sharifzadeh et al., 2014). The authors compared PCA with a wide range of linear, non-linear, kernel-based regression and sparse regression methods coupled with variable selection methodologies, and linear ridge regression combined with the proposed Elastic net-based feature selection strategy provided the best results. Other applications include assessment of meat tenderness and freshness, moisture and firmness of fruits, etc. (Dai, Sun, Xiong, Cheng, & Zeng, 2014) where the majority of these applications involve the use of PLSR, but also other methods mentioned above. PLSR and PCR were also used in order to predict quality indices of peaches, such as firmness, acidity and sugar content based on e-nose measurements (Zhang, Wang,

Table 3
List of representative rapid methods application and corresponding data handling methodologies.

Sensor type	Reference	Food type	Purpose	Data analysis method
Imaging	(Panagou et al., 2014)	Beef fillets	Spoilage	HCA, PLSR, PLS-DA
Imaging	(He, Sun, et al., 2014)	Salmon	Spoilage (LAB)	LS-SVM
Imaging	(Gowen et al., 2008)	Mushrooms	Bruise detection	PCA (application to hypercube data)
Imaging	(Sharifzadeh et al., 2014)	Meat	Monitoring meat color	PCA vs. linear, non-linear and kernel-based regression methods (ANNs & SVM)
Imaging	(Fu et al., 2014)	Milk powders	Adulteration	Spectral similarity measures (spectral angle measure-SAM, spectral correlation measure-SCM, Euclidian distance measure-EDM).
Imaging	(Ropodi et al., 2015)	Pork, beef	Detection of adulteration	HCA, PCA, LDA, PLS-DA
Imaging	(Wu et al., 2013)	Prawn	Detection of adulteration	UVE-SPA-LS-SVM
Spectroscopy	(Argyri et al., 2013)	Minced beef	Spoilage/sensory	PLSR,GA-GP, GA-ANN, SVR (various kernel function)
Spectroscopy	(Coppa et al., 2014)	Milk	Fatty acid composition	PLS and modified PLS (MPLS)
Spectroscopy	(da Costa Filho, 2014)	Edible oils	Trans-fatty acid determination	PLSR
Spectroscopy	(Ellis et al., 2005)	Chicken, pork, turkey, lamb and beef	Authentication of species and the distinct muscle groups within these species	PC-DFA, GA-MLR
Spectroscopy	(Alamprese et al., 2013)	Beef, turkey	Detection of adulteration	PCA, LDA, PLSR
Spectroscopy	(Zhao et al., 2014)	Beefburger	Detection of adulteration	PLS-DA, SIMCA, low & mid level fusion strategies based on PLS
e-nose	(Papadopoulou et al., 2013)	Beef fillets	Spoilage/sensory	PCA-DFA SVM & SVR (RBF kernel)
e-nose	(Panagou et al., 2008)	Table olives	Sensory	PCA, HCA, DFA, MLP-NN
e-nose	(Concina et al., 2009)	Tomatoes	Detection of microbial contamination	PCA
e-nose	(Pan et al., 2014)	Strawberry	Detection of fungal disease	PCA, ANN (MLP)
e-nose	(Wilson et al., 2013)	Catfish fillets	Sensory/off-flavor detection	PCA, ANN, QF
e-nose	(Zhang et al., 2012)	Peach	Firmness, sugar content, acidity	PCA, LDA, PCR, PLSR
e-nose, acoustic sensor	(Zakaria et al., 2012)	Mangoes	Ripeness/maturity	PCA, LDA, LDA-ANN
e-nose, e-tongue	(Qiu et al., 2014)	Strawberry juice	Discrimination among processing approaches	LDA, PLSR, SVM, RF
e-tongue	(Liu et al., 2013)	Orange beverage, Chinese vinegar	Authentication/discrimination among brands	PCA, BP-ANN, SVM, RF
Imaging, e-nose, spectroscopy	(Huang et al., 2014)	Pork	Freshness (TVB-N content)	PCA, BP-ANN

Ye, & Chang, 2012). Models were carefully chosen so as to achieve high performance in terms of SEC, SEP and correlation coefficient. In fact, PLS models displayed correlation coefficients ranging from 0.83 to 0.86. Additionally, a low-level fusion of e-nose and acoustic sensor data was also performed by Zakaria et al. (2012) in order to improve classification among different levels of ripeness and maturity of mangoes. While PCA and LDA were unable to differentiate among all categories, results improved in the case of LDA after the fusion process, where 99.8% of grouped cases were correctly classified after LOOCV. However, very promising testing results were achieved using three-element input vectors (from LDA output) for the development of an unsupervised Competitive Learning ANN as the prediction accuracy was improved from 66.7% to 84.4%. E-nose sensors have also been applied for the discrimination between good-flavor vs. off-flavor catfish meat samples (Wilson, Oberle, & Oberle, 2013) applying PCA for exploratory analysis and a statistical algorithm called Quality Factor (QF) analysis that determines statistical distances between profiles of classes measured using Euclidean distance. In addition, a profile measurement library was developed based on ANNs for class prediction and 91.4% of all catfish samples were classified correctly. As fatty-acid composition is an important quality parameter, chemometric methods (PCA and PLS variants) have been applied to spectroscopic data in order to predict fatty-acid composition in milk and barley and trans-fatty acid in edible oils (Coppa et al., 2014; Cozzolino, 2014; da Costa Filho, 2014). In the case of fresh and thawed milk (Coppa et al., 2014) regressions were calculated with both partial least square (PLS) and modified partial least square (MPLS) either after a preprocessing step (SNV, detrending-D, SNV and D, MSC) or with no correction. PCA loadings were used for feature selection and the calibration was performed using Student'sT for outlier removal and two elimination passes with 6 CV groups. Most models showed high performance indicating that Near and Medium IR spectroscopy could be used for routine milk FA composition recording.

4.1.3. Contamination, detection of pathogens/parasites and harmful compounds

A lot of work has also been done on the contribution of e-nose sensors on food quality assessment (Loutfi et al., 2015; Wilson,

2013). Lately, e-nose instruments were able to detect artificially spoiled canned peeled tomatoes with bacteria including *E.coli*, yeasts and fungi using the k-nearest neighbors (kNN) method with k = 1. PCA was also used for exploratory analysis and 5-fold stratified CV was applied for kNN calibration (Concina et al., 2009). In addition, MLP networks have been applied for detection and classification of pathogenic fungal diseases in strawberries (Pan et al., 2014). Strawberries were inoculated with three common pathogenic fungi (*Botrytis* sp., *Penicillium* sp. and *Rhizopus* sp.) and stored for 10 days along with non-inoculated control samples. Decay was evident with PCA and accuracy of the fungal infection type for the four groups reached 96.6% using an ANN model (MLP). HSI has several other applications in food safety, e.g. detection of parasitic nematodes (cod worms, seal worms, whale worms) in fish fillets (Kamruzzaman et al., 2015). In fact, various methods have been presented for the detection of melamine in milk and milk products based on IR spectroscopy (Domingo et al., 2014), as well as HSI (Fu et al., 2014). Fu et al. (2014) compared milk powder and melamine samples with samples of milk-melamine mixtures (melamine concentrations ranging from 0.02% to 1%) using three different spectral similarity measures, (a) spectral correlation measure (SCM), (b) spectral angle measure (SAM) and (c) Euclidean distance measure (EDM). All measures proved to have similar performance and in most cases melamine particles were detected.

4.2. Authenticity

The second category's applications cater to authentication of origin, traceability, compliance with label description and adulteration.

Chemometrics have been widely used for authenticity claims (Oliveri & Downey, 2012). There are several articles involving chemometrics for authenticity and adulteration of edible oils and fats (Nunes, 2014). Additionally, Ellis et al. (2005) applied Principal Component-DFA between muscle foods using infrared spectroscopy (Raman & FT-IR) for the discrimination among closely related poultry species, chicken and turkey and distinct muscle groups. A GA-MLR methodology was also utilized and results showed very good discrimination. Alamprese et al. (2013) and Ropodi et al. (2015) investigated detection methods for minced beef adulteration with turkey meat by UV visible (UV vis), NIR and MIR spectroscopy and with pork utilizing a MSI instrument, respectively. Results showed a somewhat better performance for PLS techniques compared to Linear Discriminant Analysis (LDA). In the first case, PLS regression was employed for the prediction of the percentage of turkey meat adulteration in minced beef. Better results were obtained with NIR and MIR spectroscopy than the UV vis results. In the case of MSI and three categories (beef, pork or adulterated), only one sample among the validation and external validation samples was misclassified using PLS-DA. However, PLS-DA results were not as accurate, in the classification in various categories of adulteration (11 classes in total) of external validation samples. In addition, Wu et al. (2013) studied the adulteration of prawn with gelatin-like chemicals. The authors' combination of uninformative variable elimination (UVE) and successive projections algorithm (SPA) followed by LS-SVM reduced the number of wavelengths drastically (from 462 to 13), improved and led to a coefficient of determination of prediction equal to 0.965. Raman spectroscopy coupled with PCA was employed for rapid determination of beef adulteration with 0, 25, 50,75, 100% w/w horsemeat presenting a good discrimination among adulteration levels (Boyacı et al., 2014). However, an extraction procedure was necessary in order to obtain pure animal fat from the samples. Offal-adulteration of fresh and frozen beef burger products has been investigated using MIR spectroscopy and chemometric data analysis (Zhao et al., 2014). Both discriminant (PLS-DA) and class-modeling (SIMCA) methods were used and the former achieved 100% correct classification accuracies for fresh and frozen-then-thawed material. Biancolillo et al. (2014) used five different instrumental techniques -thermogravimetry, mid-infrared, near-infrared, ultra-violet and visible spectroscopy-which were then combined on a PLS component (extracted feature) level in order to discriminate between different types of Italian craft beef with PLS-DA and SIMCA. The data fusion strategy provided 100% correct classification, which was higher than each instrument separately. Articles have explored the possible discrimination among different processing approaches using a combination of sensors, i.e. in strawberry juice using e-nose and e-tongue instruments coupled with LDA, PLSR as well as random forest (RF), and SVM for classification and regression for quality parameters (Qiu et al., 2014). Results showed that in most cases e-tongue results were better compared to e-nose and CV results improved for LDA when both sensors were used. However, the accuracy rate in RF and SVM was 100% using only e-tongue, while SVM performance dropped to 88% when both instruments were combined from 100% for e-tongue only. Lastly, Liu et al. (2013) uses an e-tongue instrument coupled with random forest, support vector machine and back propagation neural network algorithms (BP-ANN) for type and brand recognition of orange beverage and Chinese vinegar on 4 diverse datasets For each data set, the performance parameters of RF are superior to those of SVM and BPNN. In Table 3, some of the publications related with various sensors and their corresponding food applications are presented.

5. Existing data mining and analysis software: advantages and limitations in their implementation in the food science and food industry

As mentioned above, the increased use of computers has lead to the advancement of data acquisition as well as data analysis.

Nowadays, all commercial instruments (spectrometers, imaging devices, e-nose sensors) are directly connected with computers for two basic reasons: instrument control and data acquisition. Specifically, all instruments are dependent on their official software package for their activation, calibration, and measurement acquisition. This software's features are basic in all packages and can be used for data inspection and basic data pretreatment methods (e.g., Spectra Manager software version 2, Jasco Corp., Tokyo, Japan), while some other packages provide extended capabilities, such as image processing/segmentation (e.g., VideometerLab software version 2.12.39, Videometer A/S, Hørsholm, Denmark). Regardless of the extent of these capabilities, in most cases data can be exported in generic file formats that are compatible with the majority of data analysis softwares, e.g. Excel files or CSV (comma separated values) files in the case of spectroscopy, as well as multiple bitmap image files (BMP) in the case of HSI/MSI. That way they can subsequently be imported in available on-line tools e.g. Metaboanalyst 3.0, (http://www.metaboanalyst.ca/) or in standalone chemometric softwares such as Unscrambler (CAMO AS, Trondheim, Norway), Statistica (Statsoft, OK, USA) or XLSTAT (Addinsoft, Paris, France).

These packages are easy-to-use and users may apply many of the chemometric methods mentioned previously. While this type of software has provided users with the ability to perform fast complex calculations and automated visualization of the results, it has also increased the number of users who apply data analysis methods without an in depth understanding of the theoretical background as well as the reason for their application (Granato, de Araújo Calado, & Jarvis, 2014). In order to clearly interpret the results, there is an increasing demand for scientists who combine knowledge in food science with an understanding in mathematics,

statistics and computational intelligence. These users who may want a more hands-on approach on, e.g. how the parameters are estimated and have programming experience, have other options, the most common being two programming environments, MATLAB (The MathWorks, Inc., Natick, Massachusetts, United States) and R (https://www.r-project.org/, The R Foundation for Statistical Computing, Vienna, Austria) often combined with Rstudio (RStudio, Inc., Boston, Massachusetts, United States) an interface for R. Both MATLAB and R provide a variety of data analysis methods, ready-to-use toolboxes specific to each user's needs that go beyond food science applications and are more user-friendly. Both have similar capabilities, are widely used and often users contribute free code. However the former software is commercial and the latter is open source reducing the cost for scientific research. On the other hand, commercial software guarantees user support and minimizes software problems. MATLAB and the official toolboxes (http://www.mathworks.com) combined with in-house scripts have been used extensively in research articles (Fu et al., 2014; Gowen et al., 2008; He, Sun, et al., 2014; Huang et al., 2014; Kamruzzaman et al., 2012; Zakaria et al., 2012). Several commercial and open-source toolboxes have also been utilized for data analysis, such as *libSVM, libPLS, LS-SVM, PLS Toolbox* etc. Similarly, libraries for R are now available in CRAN package repository, a network of ftp and web servers around the world, and elsewhere. Various packages, such as *randomForest* and *plsgenomics* have been used (Liu et al., 2013; Qiu et al., 2014; Ropodi et al., 2015). In Table 4, a list of MATLAB toolboxes and R packages supporting chemometric and machine learning applications is presented.

To sum up, as there are advantages and disadvantages in all types of software; users should carefully balance their data, abilities, theoretical background as well as resources and funding in order to choose the one designed for their needs. For more information on software, there are available reviews in the literature (Nunes et al., 2015).

6. Discussion

Over the past decades, Food Authorities e.g. European Food Safety Authority (EFSA) of European Union (EU), FDA and USDA in USA, Food Standards Australia New Zealand have identified assurance of food safety and quality as one of their ultimate goals (European Commission, 2000; http://www.usda.gov/wps/portal/usda/usdahome?navid=food-safety; http://www.foodstandards.gov.au/Pages/default.aspx). It is well established that the potential of causing huge economic losses to the food industry are due to microbial spoilage and other biological or chemical hazards (Anonymous, 1993a,b, 2000; 2010; EFSA, 2006, 2008; Evans, 2005; Nychas et al., 2008). On the other hand, monitoring and predicting quality and safety of highly perishable foods, such as muscle foods, still relies heavily on regulatory inspection (European Commission, 2005; http://www.fsis.usda.gov/wps/portal/fsis/topics/inspection). This process is costly, time-consuming, and unable to provide in-line or at-line inspection. To overcome these limitations, conventional microbiological analyses should be integrated with advances in signal analysis. Both microbiological and spectroscopic profiling data can be collected at various time points along the food chain, reflecting the 'current' status of the tested product. The spectroscopic profiling data will, more specifically, include compiled vibrational spectroscopy data and surface chemistry measurements using rapid, non-invasive methods and appropriate sensor devices. All the pertinent collected data, corresponding to different knowledge levels, can be utilized with the previous mentioned methodologies aiming, ultimately, at the assembly of a complete holistic picture for food quality and safety, and contributing to a long-term vision for new technologies. Various methods based on chemometrics, machine learning and/or computational intelligence have been applied in research for food products. However, no specific method can be safely characterized as more effective, as all applications are case-specific. Depending on the sensor, methods

Table 4
List of various MATLAB and R libraries available for chemometric and machine learning applications.

	Library	Methods	Source
MATLAB	Statistics and Machine Learning Toolbox	HCA, k-means, ANOVA, MLR, LDA, kNN, SVM, RF and other methods	http://www.mathworks.com/
	Neural Network Toolbox	ANNs	http://www.mathworks.com/
	PLS Toolbox	MLR, PLS, PCR and preprocessing methods	http://www.eigenvector.com
	LibPLS	PLS-R & -DA, LDA and various methods for preprocesing, variable selection and outlier detection	http://www.libpls.net/
	iToolbox	PLS variants with intervals (iPLS, BiPLS, FiPLS, SiPLS,mwPLS)	http://www.models.life.ku.dk/iToolbox
	PLS-Genetic Algorithm Toolbox	GA-PLS	http://www.models.life.ku.dk/GAPLS
	libSVM	SVM	http://www.csie.ntu.edu.tw/~cjlin/libsvm/
	LS-SVMlab	LS-SVM	http://www.esat.kuleuven.be/sista/lssvmlab/
R	The R Stats Package	HCA, PCA, k-means and other Statistics' functions	http://www.r-project.org
	Schemometrics	PCA, PLSR, other regression methods (lasso, ridge), tools for CV, clustering, etc.	http://cran.r-project.org/web/packages/
	Chemometrics With R	PCA, GA for variable selection	http://cran.r-project.org/web/packages/
	pls	PLSR, PCR	http://cran.r-project.org/web/packages/
	plsgenomics	PLSR & DA, ridge PLS	http://cran.r-project.org/web/packages/
	gpls	generalized PLS	http://bioconductor.org/packages/
	cluster	CA methods	http://cran.r-project.org/web/packages/
	neuralnet	ANNs	http://cran.r-project.org/web/packages/
	e1071	SVM (libSVM) and other clustering methods	http://cran.r-project.org/web/packages/
	random Forest	RF	http://cran.r-project.org/web/packages/

tend to be dissimilar as the acquired data are quite different. While vibrational spectrometry can be quick and highly efficient, it lacks the ability to describe the spatial distribution of the measured attribute. In addition, as HSI provides both spectral and spatial information, analysis of the resulting data can be quite challenging and might require a wavelength selection step. MSI on the other hand, provides limited spectral information, but it is not as expensive as HSI. Furthermore, techniques such as imaging are non-invasive whereas others may require a small amount of sample.

As far as data handling techniques are concerned, depending on the dataset one or more methodologies may be applied (Liu et al., 2013; Panagou, Mohareb, Argyri, Bessant, & Nychas, 2011; Ropodi et al., 2015) or more than one instrumental techniques may be used (Argyri et al., 2013). However, very little has been done in terms of combining dissimilar instrumental techniques (Alamprese et al., 2013; Biancolillo et al., 2014; Huang et al., 2014). In this direction, combining different types of sensor data can be quite difficult as it involves acquiring data from multiple sensors, discovering the advantages and disadvantages of each sensor in a specific application, as well as understanding when there is a case of redundant or complementary application. Usually, data of the same type, e.g. spectral data, have been mostly used for this "fusion" process. For different types of sensors different possibilities will have to be explored as to the stage of the fusion, whether it will be in the data, decision level or intermediate/feature level (Dasarathy, 1997). The different instrument brands and experimental setups only add to the complexity of comparing and combining results.

In conclusion, it is evident that rapid techniques coupled with data analysis methods have given promising results in several food products with various sensors. However, there are several applications, new sensors and new algorithms that remain to be explored and validated in the future. In any case, the benefits of introducing such technologies/approaches in the food industry are:

(a) Interdisciplinarity, as these approaches will bring together scientists from different research backgrounds (food microbiologists, veterinarians, molecular biologists, bioinformaticians, software developers, IT experts, etc.) allowing, therefore, for a strong synergy among these disparate fields for the benefit of science;
(b) The massive amount of high-throughput, analytical and imaging meta-data collected with these instruments will provide the cornerstone for a holistic view of the decaying process of the various food products across diverse storage conditions (temperature and packaging). The online availability of this knowledge will ensure the continuous benefit and growing of this repository with more data becoming available beyond the course of scientific research;
(c) The implementation of this multidisciplinary approach will provide the food industry with a new generation of quality and safety monitoring tools, oriented around fast, efficient and non-invasive tools that could be easily implemented on-site using portable devices. The online connectivity of the developed system will offer an unprecedented degree of accuracy compared to conventional inspection methods and will provide more accurate and automated means of product monitoring that will substantially increase consumer confidence for the food product, and would, in the long run, improve the competitiveness of these industries;
(d) Contribution to the reduction of food waste, through a more efficient control of the processes;
(e) Reassuring trust between consumers and the food industry through the introduction of a novel system for quality and safety monitoring that can be applied throughout the production chain;
(f) Contribution to Food Regulation, in the context of fulfilling (EC) No 2073/2005 on microbiological criteria for foodstuffs.

Acknowledgment

This work has been supported by the project "Intelligent multi-sensor system for meat analysis – iMeatSense_550" co-financed by the European Union (European Social Fund-ESF) and Greek national funds through the Operational Program "Education and Lifelong Learning" of the National Strategic Reference Framework (NSRF) Research Funding Program: ARISTEIA-I.

References

Alamprese, C., Casale, M., Sinelli, N., Lanteri, S., & Casiraghi, E. (2013). Detection of minced beef adulteration with turkey meat by UV–vis, NIR and MIR spectroscopy. *LWT – Food Science and Technology, 53*(1), 225–232. http://dx.doi.org/10.1016/j.lwt.2013.01.027.

Anonymous. (1993a). Council directive 93/43/EEC on the hygiene of foodstuffs. *Official Journal of the European Communities*.

Anonymous. (1993b). Mandatory safe handling statement on labeling of raw meat and poultry products. *Federal Register (Food Safety and Inspection Service), 58*(156), 9. CFR Parts 317 and 381.

Anonymous. (2000). *WHO surveillance programme for control of foodborne infections and intoxications in Europe*, 7th Report.

Anonymous. (2010). *Food contaminants – dioxins and PCBs slide*. Retrieved September 8, 2015, from: http://ec.europa.eu/food/food/chemicalsafety/contaminants/dioxins_en.htm.

Argyri, A. A., Jarvis, R. M., Wedge, D., Xu, Y., Panagou, E. Z., Goodacre, R., et al. (2013). A comparison of Raman and FT-IR spectroscopy for the prediction of meat spoilage. *Food Control, 29*(2), 461–470. http://dx.doi.org/10.1016/j.foodcont.2012.05.040.

Argyri, A. A., Panagou, E. Z., & Nychas, G.-J. (2014). Monitoring microbial spoilage of foods by vibrational spectroscopy (FT-IR & raman). In *Novel food preservation and microbial assessment techniques* (pp. 386–434). CRC Press. http://dx.doi.org/10.1201/b16758-17.

Argyri, A. A., Panagou, E. Z., Tarantilis, P. A., Polysiou, M., & Nychas, G.-J. E. (2010). Rapid qualitative and quantitative detection of beef fillets spoilage based on Fourier transform infrared spectroscopy data and artificial neural networks. *Sensors and Actuators B: Chemical, 145*(1), 146–154. http://dx.doi.org/10.1016/j.snb.2009.11.052.

Baietto, M., & Wilson, A. (2015). Electronic-nose applications for fruit identification, ripeness and quality grading. *Sensors, 15*. http://dx.doi.org/10.3390/s150100899.

Balabin, R. M., & Lomakina, E. I. (2011). Support vector machine regression (LS-SVM) – an alternative to artificial neural networks (ANNs) for the analysis of quantum chemistry data? *Physical Chemistry Chemical Physics: PCCP, 13*(24), 11710–11718. http://dx.doi.org/10.1039/c1cp00051a.

Baldwin, E. A., Bai, J., Plotto, A., & Dea, S. (2011). Electronic noses and tongues: applications for the food and pharmaceutical industries. *Sensors, 11*(12), 4744–4766. http://dx.doi.org/10.3390/s110504744.

Barker, M., & Rayens, W. (2003). Partial least squares for discrimination. *Journal of Chemometrics, 17*(3), 166–173. http://dx.doi.org/10.1002/cem.785.

Berrueta, L. A., Alonso-Salces, R. M., & Héberger, K. (2007). Supervised pattern recognition in food analysis. *Journal of Chromatography A, 1158*(1–2), 196–214. http://dx.doi.org/10.1016/j.chroma.2007.05.024.

Biancolillo, A., Bucci, R., Magrì, A. L., Magrì, A. D., & Marini, F. (2014). Data-fusion for multiplatform characterization of an Italian craft beer aimed at its authentication. *Analytica Chimica Acta, 820*, 23–31. http://dx.doi.org/10.1016/j.aca.2014.02.024.

Boulesteix, A.-L., & Strimmer, K. (2007). Partial least squares: a versatile tool for the analysis of high-dimensional genomic data. *Briefings in Bioinformatics, 8*(1), 32–44. http://dx.doi.org/10.1093/bib/bbl016.

Boyaci, İ. H., Temiz, H. T., Uysal, R. S., Velioğlu, H. M., Yadegari, R. J., & Rishkan, M. M. (2014). A novel method for discrimination of beef and horsemeat using Raman spectroscopy. *Food Chemistry, 148*, 37–41. http://dx.doi.org/10.1016/j.foodchem.2013.10.006.

Brereton, R. G. (2014). A short history of chemometrics: a personal view. *Journal of Chemometrics, 28*(10), 749–760. http://dx.doi.org/10.1002/cem.2633.

Cheng, J.-H., & Sun, D.-W. (2014). Hyperspectral imaging as an effective tool for quality analysis and control of fish and other seafoods: current research and potential applications. *Trends in Food Science & Technology, 37*(2), 78–91. http://dx.doi.org/10.1016/j.tifs.2014.03.006.

Concina, I., Falasconi, M., Gobbi, E., Bianchi, F., Musci, M., Mattarozzi, M., … Sberveglieri, G. (2009). Early detection of microbial contamination in processed tomatoes by electronic nose. *Food Control, 20*(10), 873–880. http://dx.doi.org/10.1016/j.foodcont.2008.11.006.

Coppa, M., Revello-Chion, A., Giaccone, D., Ferlay, A., Tabacco, E., & Borreani, G. (2014). Comparison of near and medium infrared spectroscopy to predict fatty

acid composition on fresh and thawed milk. *Food Chemistry, 150*, 49–57. http://dx.doi.org/10.1016/j.foodchem.2013.10.087.
Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning, 20*(3), 273–297. http://dx.doi.org/10.1007/BF00994018.
da Costa Filho, P. A. (2014). Developing a rapid and sensitive method for determination of trans-fatty acids in edible oils using middle-infrared spectroscopy. *Food Chemistry, 158*, 1–7. http://dx.doi.org/10.1016/j.foodchem.2014.02.084.
Cozzolino, D. (2014). An overview of the use of infrared spectroscopy and chemometrics in authenticity and traceability of cereals. *Food Research International, 60*, 262–265. http://dx.doi.org/10.1016/j.foodres.2013.08.034.
Cozzolino, D., Roumeliotis, S., & Eglinton, J. (2014). Evaluation of the use of attenuated total reflectance mid infrared spectroscopy to determine fatty acids in intact seeds of barley (*Hordeum vulgare*). *LWT – Food Science and Technology, 56*(2), 478–483. http://dx.doi.org/10.1016/j.lwt.2013.11.019.
Dai, Q., Sun, D.-W., Xiong, Z., Cheng, J.-H., & Zeng, X.-A. (2014). Recent advances in data mining techniques and their applications in hyperspectral image processing for the food Industry. *Comprehensive Reviews in Food Science and Food Safety, 13*(5), 891–905. http://dx.doi.org/10.1111/1541-4337.12088.
Damez, J.-L., & Clerjon, S. (2013). Quantifying and predicting meat and meat products quality attributes using electromagnetic waves: an overview. *Meat Science, 95*(4), 879–896. http://dx.doi.org/10.1016/j.meatsci.2013.04.037.
Dasarathy, B. V. (1997). Sensor fusion potential exploitation-innovative architectures and illustrative applications. *Proceedings of the IEEE, 85*(1), 24–38. http://dx.doi.org/10.1109/5.554206.
Domingo, E., Tirelli, A. A., Nunes, C. A., Guerreiro, M. C., & Pinto, S. M. (2014). Melamine detection in milk using vibrational spectroscopy and chemometrics analysis: a review. *Food Research International, 60*, 131–139. http://dx.doi.org/10.1016/j.foodres.2013.11.006.
Doulgeraki, A. I., Ercolini, D., Villani, F., & Nychas, G.-J. E. (2012). Spoilage microbiota associated to the storage of raw meat in different conditions. *International Journal of Food Microbiology, 157*(2), 130–141. http://dx.doi.org/10.1016/j.ijfoodmicro.2012.05.020.
Duchesne, C., Liu, J. J., & MacGregor, J. F. (2012). Multivariate image analysis in the process industries: a review. *Chemometrics and Intelligent Laboratory Systems, 117*, 116–128. http://dx.doi.org/10.1016/j.chemolab.2012.04.003.
EFSA (European Food Safety Authority). (2006). The community summary report on trends and sources of zoonoses, zoonotic agents, antimicrobial resistance and foodborne outbreaks in the European Union. *The EFSA Journal, 290*. http://dx.doi.org/10.2903/j.efsa.2006.94r.
EFSA (European Food Safety Authority). (2008). Statement of EFSA on the risks for public health due to the presence of dioxins in pork from Ireland. *The EFSA Journal, 911*, 1–15.
Ellis, D. I., Broadhurst, D., Clarke, S. J., & Goodacre, R. (2005). Rapid identification of closely related muscle foods by vibrational spectroscopy and machine learning. *The Analyst, 130*(12), 1648–1654. http://dx.doi.org/10.1039/b511484e.
Ellis, D. I., & Goodacre, R. (2001). Rapid and quantitative detection of the microbial spoilage of muscle foods: current status and future trends. *Trends in Food Science & Technology, 12*(11), 414–424. http://dx.doi.org/10.1016/S0924-2244(02)00019-5.
Engel, J., Gerretzen, J., Szymańska, E., Jansen, J. J., Downey, G., Blanchet, L., et al. (2013). Breaking with trends in pre-processing? *TrAC Trends in Analytical Chemistry, 50*, 96–106. http://dx.doi.org/10.1016/j.trac.2013.04.015.
European Commission. (2000). *White paper on food safety* (12 January 2000 COM (1999) 719 final) (p. 52).
European Commission. (2005). Commision regulation (EC) No 2073/2005 on microbiological criteria for foodstuffs. *Official Journal of the European Union, OJ L, 338*, 1–26.
Evans, D. (2005). Future trends in the European beef industry: a global view. In *8th Annual Langford food Industry Conference, Bristol -UK* (pp. 45–48).
Everitt, B., Landau, S., Leese, M., & Stahl, D. (2011). *Cluster analysis* (5th ed.). Wiley.
Fu, X., Kim, M. S., Chao, K., Qin, J., Lim, J., Lee, H. …Ying, Y. (2014). Detection of melamine in milk powders based on NIR hyperspectral imaging and spectral similarity analyses. *Journal of Food Engineering, 124*, 97–104. http://dx.doi.org/10.1016/j.jfoodeng.2013.09.023.
Ghasemi-Varnamkhasti, M., Mohtasebi, S. S., & Siadat, M. (2010). Biomimetic-based odor and taste sensing systems to food quality and safety characterization: an overview on basic principles and recent achievements. *Journal of Food Engineering, 100*(3), 377–387. http://dx.doi.org/10.1016/j.jfoodeng.2010.04.032.
Goodacre, R. (2003). Explanatory analysis of spectroscopic data using machine learning of simple, interpretable rules. *Vibrational Spectroscopy, 32*(1), 33–45. http://dx.doi.org/10.1016/S0924-2031(03)00045-6.
Gowen, A. A., Feng, Y., Gaston, E., & Valdramidis, V. (2015). Recent applications of hyperspectral imaging in microbiology. *Talanta, 137*, 43–54. http://dx.doi.org/10.1016/j.talanta.2015.01.012.
Gowen, A. A., O'Donnell, C. P., Taghizadeh, M., Cullen, P. J., Frias, J. M., & Downey, G. (2008). Hyperspectral imaging combined with principal component analysis for bruise damage detection on white mushrooms (*Agaricus bisporus*). *Journal of Chemometrics, 22*(3–4), 259–267. http://dx.doi.org/10.1002/cem.1127.
Granato, D., de Araújo Calado, V. M., & Jarvis, B. (2014). Observations on the use of statistical methods in Food Science and Technology. *Food Research International, 55*, 137–149. http://dx.doi.org/10.1016/j.foodres.2013.10.024.
Haralick, R. M., Shanmugam, K., & Dinstein, I. (1973). Textural features for image classification. *IEEE Transactions on Systems, Man, and Cybernetics, 3*(6), 610–621. http://dx.doi.org/10.1109/TSMC.1973.4309314.
Haralick, R. M., & Shapiro, L. G. (1991). *Computer and robot vision* (1st ed.). Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc.
Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning* (2nd ed.). New York: Springer.
He, H.-J., Sun, D.-W., & Wu, D. (2014a). Rapid and real-time prediction of lactic acid bacteria (LAB) in farmed salmon flesh using near-infrared (NIR) hyperspectral imaging combined with chemometric analysis. *Food Research International, 62*, 476–483. http://dx.doi.org/10.1016/j.foodres.2014.03.064.
He, H.-J., Wu, D., & Sun, D.-W. (2014b). Potential of hyperspectral imaging combined with chemometric analysis for assessing and visualising tenderness distribution in raw farmed salmon fillets. *Journal of Food Engineering, 126*, 156–164. http://dx.doi.org/10.1016/j.jfoodeng.2013.11.015.
Huang, L., Zhao, J., Chen, Q., & Zhang, Y. (2014). Nondestructive measurement of total volatile basic nitrogen (TVB-N) in pork meat by integrating near infrared spectroscopy, computer vision and electronic nose techniques. *Food Chemistry, 145*, 228–236. http://dx.doi.org/10.1016/j.foodchem.2013.06.073.
Jarvis, R. M., & Goodacre, R. (2005). Genetic algorithm optimization for pre-processing and variable selection of spectroscopic data. *Bioinformatics (Oxford, England), 21*(7), 860–868. http://dx.doi.org/10.1093/bioinformatics/bti102.
Jolliffe, I. T. (2002). *Principal component analysis* (2nd ed.). New York: Springer. http://dx.doi.org/10.1007/b98835.
de Jong, S. (1993). SIMPLS: an alternative approach to partial least squares regression. *Chemometrics and Intelligent Laboratory Systems, 18*(3), 251–263. http://dx.doi.org/10.1016/0169-7439(93)85002-X.
Kamruzzaman, M., Barbin, D., ElMasry, G., Sun, D.-W., & Allen, P. (2012). Potential of hyperspectral imaging and pattern recognition for categorization and authentication of red meat. *Innovative Food Science & Emerging Technologies, 16*, 316–325. http://dx.doi.org/10.1016/j.ifset.2012.07.007.
Kamruzzaman, M., Makino, Y., & Oshita, S. (2015). Non-invasive analytical technology for the detection of contamination, adulteration, and authenticity of meat, poultry, and fish: a review. *Analytica Chimica Acta, 853*, 19–29. http://dx.doi.org/10.1016/j.aca.2014.08.043.
Kamruzzaman, M., Maniko, Y., & Oshita, S. (2014). Non-invasive analytical technology for the detection of contamination, adulteration, and authenticity of meat, poultry, and fish: a review. *Analytica Chimica Acta*. http://dx.doi.org/10.1016/j.aca.2014.08.043.
Kohavi, R., & Provost, F. (1998). Glossary of terms. *Machine Learning, 30*(2–3), 271–274. Retrieved from: http://dl.acm.org/citation.cfm?id=288808.288815.
Liu, M., Wang, M., Wang, J., & Li, D. (2013). Comparison of random forest, support vector machine and back propagation neural network for electronic tongue data classification: application to the recognition of orange beverage and Chinese vinegar. *Sensors and Actuators B: Chemical, 177*, 970–980. http://dx.doi.org/10.1016/j.snb.2012.11.071.
Liu, D., Zeng, X.-A., & Sun, D.-W. (2015). Recent developments and applications of hyperspectral imaging for quality evaluation of agricultural products: a review. *Critical Reviews in Food Science and Nutrition, 55*(12), 1744–1757. http://dx.doi.org/10.1080/10408398.2013.777020.
Loutfi, A., Coradeschi, S., Mani, G. K., Shankar, P., & Rayappan, J. B. B. (2015). Electronic noses for food quality: a review. *Journal of Food Engineering, 144*, 103–111. http://dx.doi.org/10.1016/j.jfoodeng.2014.07.019.
Luke, B. T. (2003). Nature-inspired methods in chemometrics: genetic algorithms and artificial neural networks. *Data Handling in Science and Technology, 23*. http://dx.doi.org/10.1016/S0922-3487(03)23001-X. Elsevier.
Marini, F. (2009). Artificial neural networks in foodstuff analyses: trends and perspectives A review. *Analytica Chimica Acta, 635*(2), 121–131. http://dx.doi.org/10.1016/j.aca.2009.01.009.
Nunes, C. A. (2014). Vibrational spectroscopy and chemometrics to assess authenticity, adulteration and intrinsic quality parameters of edible oils and fats. *Food Research International, 60*, 255–261. http://dx.doi.org/10.1016/j.foodres.2013.08.041.
Nunes, C. A., Alvarenga, V. O., de Souza Sant'Ana, A., de Sousa Santos, J., & Granato, D. (2015). The Use of statistical software in food science and technology: advantages, limitations and misuses. *Food Research International*. http://dx.doi.org/10.1016/j.foodres.2015.06.011.
Nychas, G.-J. E., Skandamis, P. N., Tassou, C. C., & Koutsoumanis, K. P. (2008). Meat spoilage during distribution. *Meat Science, 78*, 77–89. http://dx.doi.org/10.1016/j.meatsci.2007.06.020.
Oliveri, P., & Downey, G. (2012). Multivariate class modeling for the verification of food-authenticity claims. *TrAC Trends in Analytical Chemistry, 35*, 74–86. http://dx.doi.org/10.1016/j.trac.2012.02.005.
Otto, M. (2007). *Chemometrics* (2nd ed.). Weinheim: WILEY-VCH Verlag GmbH & Co. KGaA.
Panagou, E. Z., Mohareb, F. R., Argyri, A. A., Bessant, C. M., & Nychas, G. J. E. (2011). A comparison of artificial neural networks and partial least squares modelling for the rapid detection of the microbial spoilage of beef fillets based on Fourier transform infrared spectral fingerprints. *Food Microbiology, 28*(4), 782–790. http://dx.doi.org/10.1016/j.fm.2010.05.014.
Panagou, E. Z., Papadopoulou, O., Carstensen, J. M., & Nychas, G.-J. E. (2014). Potential of multispectral imaging technology for rapid and non-destructive determination of the microbiological quality of beef filets during aerobic storage. *International Journal of Food Microbiology, 174*, 1–11. http://dx.doi.org/10.1016/j.ijfoodmicro.2013.12.026.
Panagou, E. Z., Sahgal, N., Magan, N., & Nychas, G.-J. E. (2008). Table olives volatile fingerprints: potential of an electronic nose for quality discrimination. *Sensors and Actuators B: Chemical, 134*(2), 902–907. http://dx.doi.org/10.1016/j.snb.2008.06.038.

Pan, L., Zhang, W., Zhu, N., Mao, S., & Tu, K. (2014). Early detection and classification of pathogenic fungal disease in post-harvest strawberry fruit by electronic nose and gas chromatography–mass spectrometry. *Food Research International, 62*, 162–168. http://dx.doi.org/10.1016/j.foodres.2014.02.020.

Papadopoulou, O. S., Panagou, E. Z., Mohareb, F. R., & Nychas, G.-J. E. (2013). Sensory and microbiological quality assessment of beef fillets using a portable electronic nose in tandem with support vector machine analysis. *Food Research International, 50*(1), 241–249. http://dx.doi.org/10.1016/j.foodres.2012.10.020.

Papadopoulou, O., Panagou, E. Z., Tassou, C. C., & Nychas, G.-J. E. (2011). Contribution of Fourier transform infrared (FTIR) spectroscopy data on the quantitative determination of minced pork meat spoilage. *Food Research International, 44*(10), 3264–3271. http://dx.doi.org/10.1016/j.foodres.2011.09.012.

Qin, J., Chao, K., Kim, M. S., Lu, R., & Burks, T. F. (2013). Hyperspectral and multispectral imaging for evaluating food safety and quality. *Journal of Food Engineering, 118*(2), 157–171. http://dx.doi.org/10.1016/j.jfoodeng.2013.04.001.

Qiu, S., Wang, J., & Gao, L. (2014). Discrimination and characterization of strawberry juice based on electronic nose and tongue: comparison of different juice processing approaches by LDA, PLSR, RF, and SVM. *Journal of Agricultural and Food Chemistry, 62*(27), 6426–6434. http://dx.doi.org/10.1021/jf501468b.

Ropodi, A. I., Pavlidis, D. E., Mohareb, F., Panagou, E. Z., & Nychas, G.-J. E. (2015). Multispectral image analysis approach to detect adulteration of beef and pork in raw meats. *Food Research International, 67*, 12–18. http://dx.doi.org/10.1016/j.foodres.2014.10.032.

Ross, T. (1996). Indices for performance evaluation of predictive models in food microbiology. *Journal of Applied Bacteriology, 81*(5), 501–508. http://dx.doi.org/10.1111/j.1365-2672.1996.tb03539.x.

Sankaran, S., Khot, L. R., & Panigrahi, S. (2012). Biology and applications of olfactory sensing system: a review. *Sensors and Actuators B: Chemical, 171–172*, 1–17. http://dx.doi.org/10.1016/j.snb.2012.03.029.

Sharifzadeh, S., Clemmensen, L. H., Borggaard, C., Støier, S., & Ersbøll, B. K. (2014). Supervised feature selection for linear and non-linear regression of L*a*b* color from multispectral images of meat. *Engineering Applications of Artificial Intelligence, 27*, 211–227. http://dx.doi.org/10.1016/j.engappai.2013.09.004.

Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information Processing & Management, 45*(4), 427–437. http://dx.doi.org/10.1016/j.ipm.2009.03.002.

Sumathi, S., & Sureldha, P. (2010). *Computational intelligence paradigms. Computational intelligence paradigms*. CRC Press. http://dx.doi.org/10.1201/9781439809037.

Sun, D.-W. (2009). *Infrared spectroscopy for food quality analysis and control. Infrared spectroscopy for food quality analysis and control*. Elsevier. http://dx.doi.org/10.1016/B978-0-12-374136-3.00021-3.

Sutton, R. S., & Barto, A. G. (1998). *Introduction to reinforcement learning* (1st ed.). Cambridge, MA, USA: MIT Press.

Teena, M., Manickavasagan, a., Mothershaw, a., El Hadi, S., & Jayas, D. S. (2013). Potential of machine vision techniques for detecting fecal and microbial contamination of food products: a review. *Food and Bioprocess Technology, 6*(7), 1621–1634. http://dx.doi.org/10.1007/s11947-013-1079-7.

Velusamy, V., Arshak, K., Korostynska, O., Oliwa, K., & Adley, C. (2010). An overview of foodborne pathogen detection: in the perspective of biosensors. *Biotechnology Advances, 28*(2), 232–254. http://dx.doi.org/10.1016/j.biotechadv.2009.12.004.

Westad, F., & Marini, F. (2015). Validation of chemometric models – a tutorial. *Analytica Chimica Acta, 893*, 14–24. http://dx.doi.org/10.1016/j.aca.2015.06.056.

Wilson, A. (2013). Diverse applications of electronic-nose technologies in agriculture and forestry. *Sensors, 13*(2), 2295–2348. http://dx.doi.org/10.3390/s130202295.

Wilson, A. D., & Baietto, M. (2009). Applications and advances in electronic-nose technologies. *Sensors, 9*(7), 5099–5148. http://dx.doi.org/10.3390/s90705099.

Wilson, A., Oberle, C., & Oberle, D. (2013). Detection of off-flavor in catfish using a conducting polymer electronic-nose technology. *Sensors, 13*(12), 15968–15984. http://dx.doi.org/10.3390/s131215968.

Wold, S., Sjöström, M., & Eriksson, L. (2001). PLS-regression: a basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems, 58*(2), 109–130. http://dx.doi.org/10.1016/S0169-7439(01)00155-1.

Woodcock, T., O'Donnell, C., & Downey, G. (2008). Review: better quality food and beverages: the role of near infrared spectroscopy. *Journal of Near Infrared Spectroscopy, 16*(1), 1–29. http://dx.doi.org/10.1255/jnirs.758.

Wu, D., Shi, H., He, Y., Yu, X., & Bao, Y. (2013). Potential of hyperspectral imaging and multivariate analysis for rapid and non-invasive detection of gelatin adulteration in prawn. *Journal of Food Engineering, 119*(3), 680–686. http://dx.doi.org/10.1016/j.jfoodeng.2013.06.039.

Xiaobo, Z., Jiewen, Z., Povey, M. J. W., Holmes, M., & Hanpin, M. (2010). Variables selection methods in near-infrared spectroscopy. *Analytica Chimica Acta, 667*(1–2), 14–32. http://dx.doi.org/10.1016/j.aca.2010.03.048.

Xiong, Z., Sun, D.-W., Zeng, X.-A., & Xie, A. (2014). Recent developments of hyperspectral imaging systems and their applications in detecting quality attributes of red meats: a review. *Journal of Food Engineering, 132*, 1–13. http://dx.doi.org/10.1016/j.jfoodeng.2014.02.004.

Zakaria, A., Md Shakaff, A. Y., Masnan, M. J., Saad, F. S. A., Adom, A. H., Ahmad, M. N., … Kamarudin, L. M. (2012). Improved maturity and ripeness classifications of Magnifera Indica cv. harumanis mangoes through sensor fusion of an electronic nose and acoustic sensor. *Sensors (Switzerland), 12*(5), 6023–6048. http://dx.doi.org/10.3390/s120506023.

Zhang, B., Huang, W., Li, J., Zhao, C., Fan, S., Wu, J., et al. (2014). Principles, developments and applications of computer vision for external quality inspection of fruits and vegetables: a review. *Food Research International, 62*, 326–343. http://dx.doi.org/10.1016/j.foodres.2014.03.012.

Zhang, H., Wang, J., Ye, S., & Chang, M. (2012). Application of electronic nose and statistical analysis to predict quality indices of peach. *Food and Bioprocess Technology, 5*, 65–72. http://dx.doi.org/10.1007/s11947-009-0295-7.

Zhao, M., Downey, G., & O'Donnell, C. P. (2014). Detection of adulteration in fresh and frozen beefburger products by beef offal using mid-infrared ATR spectroscopy and multivariate data analysis. *Meat Science, 96*(2 Pt A), 1003–1011. http://dx.doi.org/10.1016/j.meatsci.2013.10.015.

APPENDIX III

Conference contributions

Oral and poster presentations:

1. Ropodi, A.I., Panagou, E.Z. and Nychas G.-J.E. (2016).Efficacy of Multi-spectral Imaging and FTIR Spectroscopy as Methods for Detection of Frozen-Then-Thawed Minced Beef. *IAFP - European Symposium on Food Safety, Athens-Greece, 11-13 May 2016, Poster Presentation.*
2. Gkoussari, C.N., Fotopoulou, E.I., Ropodi, A.I., Pavlidis, D.E., Panagou, E.Z. and Nychas G.-J.E. (2016). Assessment of Minced Beef Microbiological Quality Based on Multiple Sensor Data and Validation with Independent Datasets. *IAFP - European Symposium on Food Safety, Athens-Greece, 11-13 May 2016, Poster Presentation.*
3. Ropodi, A.I., Panagou, E.Z. and Nychas G.-J.E. (2015). Multispectral Imaging (MSI); a Promising Method for the Detection of Minced Beef Adulteration with Horsemeat. *29th EFFoST International Conference, Athens-Greece, 10-12 November 2015, Oral Presentation.*
4. Ropodi A.I., Panagou E.Z. and Nychas G.-J.E. (2015). Assessment of minced beef spoilage using Fourier Transform InfraRed (FTIR) spectroscopy, ensemble learning and Artificial Neural Networks (ANNs). *9th International Conference on Predictive Modelling in Food (ICPMF9), Rio de Janeiro, Brazil, 8-12 September 2015, Poster Presentation.*
5. Papadopoulou, O., Ropodi, A.I., Panagou, E.Z. and Nychas G.-J.E. (2015). Monitoring Spoilage of Sterile Pork Meat Fillets Inoculated with Specific Spoilage Microorganisms (*Lactobacillus sakei, Leuconostocmesenteroides*) Packaged under Modified Atmospheres in Tandem with GC/MS Analysis and Chemometrics. *IAFP - European Symposium on Food Safety, Cardiff-Wales, 20-22th April, Poster Presentation.*
6. Ropodi, A.I., Pavlidis, D.E., Loukas, D., Tsakanikas, P., Panagou, E.Z. and Nychas, G.-J.E. (2015). A Dual-sensor Method for the Assessment of Minced Beef Microbial Quality. *IFPAC-2015 Food Quality, Safety & Analysis PAT Applications, 29th Conference & Exhibition, Arlington, VA, (Washington, DC)-U.S.A., Jan. 25-28, Oral Presentation.*

7. Ropodi, A.I., Pavlidis, D.E., Loukas, D., Panagou, E.Z. and Nychas, G.-J.E. (2015). Quantification of Minced Beef Adulteration with pork using FTIR Spectroscopy and Partial Least Squares Regression. *IFPAC-2015 Food Quality, Safety & Analysis PAT Applications, 29th Conference & Exhibition, Arlington, VA, (Washington, DC)-U.S.A., Jan. 25-28, Poster Presentation.*

8. Ropodi, A.I., Pavlidis, D.E., Loukas, D., Panagou, E.Z. and Nychas, G.-J.E. (2015). Detection of Minced Meat Adulteration using Multispectral Imaging and Fourier Transform Infrared (FTIR) Spectroscopy. *IFPAC-2015 Food Quality, Safety & Analysis PAT Applications, 29th Conference & Exhibition, Arlington, VA, (Washington, DC)-U.S.A., Jan. 25-28, Poster Presentation.*

9. Loukas, D., Ropodi, A., &Nychas, G.-J. (2014). Regression modeling for spectral data sets : A multi-objective genetic approach. In *3rd International Symposium & 25th National Conference on Operational Research,Volos-Greece, 26-27 June, Oral Presentation, Book of Proceedings* (pp. 140–147).

10. Pavlidis, D.E., Ropodi, A.I., Loukas, D., Panagou, E.Z. and Nychas, G.-J.E. (2014). Multispectral Imaging vs. Fourier Transform InfraRed (FTIR) Spectroscopy for Monitoring Meat Spoilage. *European Symposium on Food Safety, Budapest - Hungary, 7-9th May, Oral Presentation.*

11. Ropodi, A.I., Tsakanikas, P., Loukas, D., Panagou, E.Z. and Nychas, G.-J.E. (2014). Multispectral image analysis for the assessment of pork minced meat quality and mapping of microbial contamination. *EUROPT(R)ODE XII Conference on Optical and Chemical Sensors, Athens - Greece, 13-16th April, Poster Presentation.*

12. Pavlidis, D.E., Ropodi, A.I., Panagou, E.Z. and Nychas, G.-J.E. (2014). Multispectral Image Analysis: a promising tool for the detection of minced meat adulteration. *EUROPT(R)ODE XII Conference on Optical and Chemical Sensors, Athens - Greece, 13-16th April, Poster Presentation.*

13. Ropodi, A.I., Loukas, D., Carstensen, J.-M., Panagou, E.Z., &Nychas, G.-J. (2014). Multispectral Image Analysis; a tool to assess minced meat quality & authenticity in the context of application of PAT in the meat sector. *ASSET2014*

Food Integrity & Traceability Conference, Belfast - UK, 8-10th April, Poster Presentation.

14. Ropodi, A.I., Panagou, E.Z. and Nychas, G.-J.E. (2013). Assessment of microbiological quality and authenticity of minced meat using multispectral image analysis. *8th International Conference on Predictive Modelling in Food, Paris - France, 16-20th September, Oral Presentation.*
15. Ropodi, A.I., Panagou, E.Z. and Nychas, G.-J.E. (2013). Evaluation of meat spoilage using FT-IR and multispectral imaging analysis. *Microbial Spoilers in Food 2013, Quimper - France, 1st-3rd July, Oral Presentation.*