

# Journal Pre-proof

Integration of handheld NIR and machine learning to “Measure & Monitor” chicken meat authenticity

Hadi Parastar, Geert van Kollenburg, Yannick Weesepeel, Andre van den Doel, Lutgarde Buydens, Jeroen Jansen



PII: S0956-7135(20)30065-7

DOI: <https://doi.org/10.1016/j.foodcont.2020.107149>

Reference: JFCO 107149

To appear in: *Food Control*

Received Date: 4 December 2019

Revised Date: 27 January 2020

Accepted Date: 28 January 2020

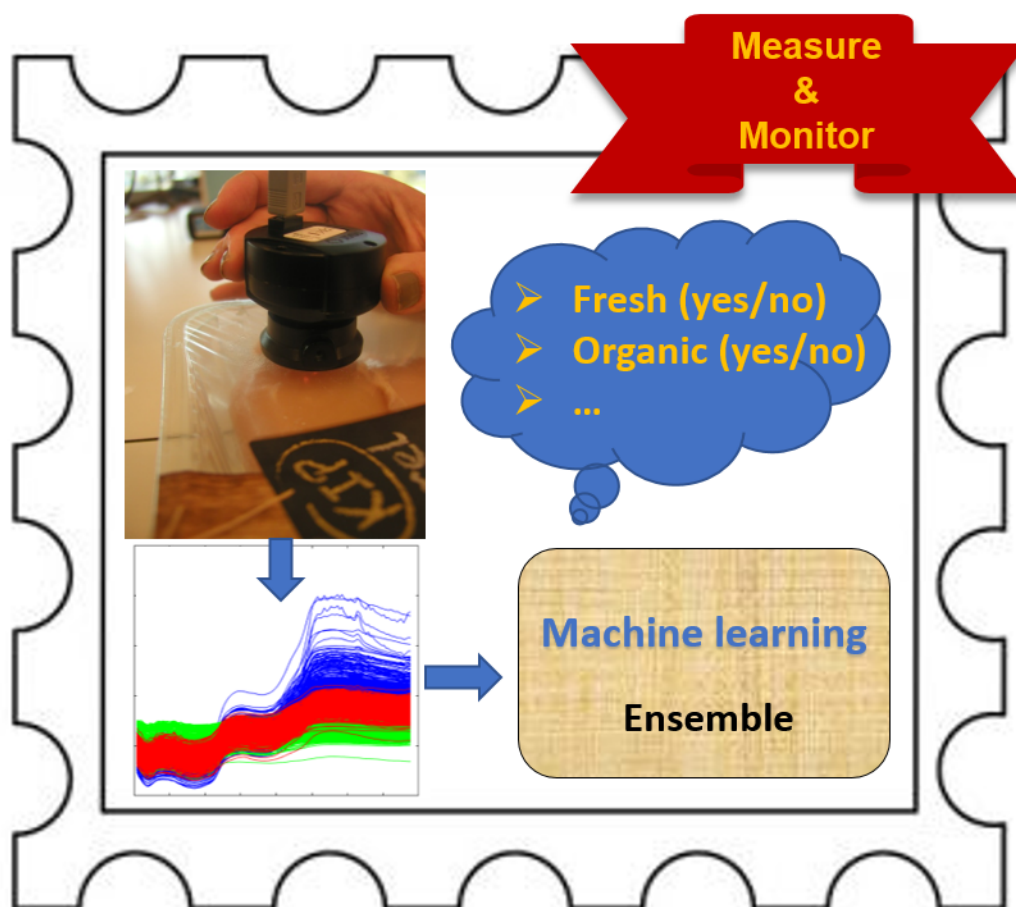
Please cite this article as: Parastar H., van Kollenburg G., Weesepeel Y., van den Doel A., Buydens L. & Jansen J., Integration of handheld NIR and machine learning to “Measure & Monitor” chicken meat authenticity, *Food Control* (2020), doi: <https://doi.org/10.1016/j.foodcont.2020.107149>.

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2020 Published by Elsevier Ltd.

**CRedit author statement**

**Hadi Parastar:** Formal Analysis, Funding acquisition, Investigation, Methodology, Software, Visualization, Writing – original draft **Geert van Kollenburg:** Conceptualization, Formal Analysis, Investigation, Methodology, Project administration, Supervision, Writing – original draft **Yannick Weesepeel:** Conceptualization, Data curation, Resources, Validation, Writing – review & editing **André van den Doel:** Data curation, Methodology, Writing – review & editing **Lutgarde Buydens:** Funding acquisition, Resources, Supervision, Validation **Jeroen Jansen:** Conceptualization, Funding acquisition, Project administration, Supervision.



**Integration of handheld NIR and machine learning to “Measure & Monitor” chicken meat authenticity.**

Hadi Parastar<sup>[1]‡</sup> & Geert van Kollenburg<sup>\*[2]‡</sup>, Yannick Weesepeel<sup>[3]</sup>, Andre van den Doel<sup>[2]</sup>, Lutgarde Buydens<sup>[2]</sup>, Jeroen Jansen<sup>[2]</sup>.

<sup>[1]</sup> *Department of Chemistry, Sharif University of Technology, Tehran, Iran*

<sup>[2]</sup> *Radboud University, Department of Analytical Chemistry/ Chemometrics, Institute for Molecules and Materials (IMM), Heyendaalseweg 135, 6525 AJ Nijmegen, The Netherlands, Tel: +31 24 3653180 e-mail: chemometrics@science.ru.nl*

<sup>[3]</sup> *Wageningen Food Safety Research (WFSR), Wageningen University & Research, Wageningen, The Netherlands*

<sup>‡</sup> These authors contributed to this work equally.

## Integration of handheld NIR and machine learning to “Measure & Monitor” chicken meat authenticity.

### Abstract

By combining portable, handheld near-infrared (NIR) spectroscopy with state-of-the-art classification algorithms, we developed a powerful method to test chicken meat authenticity. The research presented shows that it is both possible to discriminate fresh from thawed meat, based on NIR spectra, as well as to correctly classify chicken fillets according to the growth conditions of the chickens with good accuracy. In all cases, the random subspace discriminant ensemble (RSDE) method significantly outperformed other common classification methods such as partial least squares-discriminant analysis (PLS-DA), artificial neural network (ANN) and support vector machine (SVM) with classification accuracy of >95%. This study shows that handheld NIR coupled with machine learning algorithms is a useful, fast, non-destructive tool to identify the authenticity of chicken meat. By comparing and combining different protocols to measure the NIR spectra (i.e., through packaging and directly on meat), we show the possibilities for both consumers and food inspection authorities to check the authenticity and origin of packaged chicken fillet.

**Keywords:** *Handheld NIR, Chemometrics, Ensemble learning, Meat authenticity*

## 1. Introduction

The supply of sufficient healthy, safe, and authentic food to a growing world population is one of the most important challenges for the present and the future (Pischetsrieder, 2018). Detection of food adulteration such as unlabelled replacement of food components may be hindered because of the targeted focus of analytical techniques (Reid, O'Donnell, & Downey, 2006; Sentandreu & Sentandreu, 2014). From an analytical standpoint, successful detection of food adulteration faces two major challenges (Reid et al., 2006). The first challenge comprises untargeted determination of undeclared ingredients or unknown (hazardous) naturally present substances. Secondly, and more analytically challenging, are claims like animal welfare, fair trade, or eco-friendly production. While these “soft claims” are generally beyond the scope of analytical chemistry, the effects on the chemical composition of the product may still be found and quantified.

Meat authenticity (and traceability) are of particular importance in modern society (Sentandreu & Sentandreu, 2014; Vlachos, Arvanitoyannis, & Tserkezou, 2016). Recent events of meat adulteration with non-declared species such as horse meat illustrate the global need for clear and reliable checks for consumer products, but even intact fresh meat is often indistinguishable between brands or price-range. Nowadays price and lifestyle, together with religion and health concerns, determine an individual's choice for particular food products (Reid et al., 2006; Sentandreu & Sentandreu, 2014).

Detection technologies applied for food authenticity are mainly based on spectroscopic and chromatographic techniques (Gallo & Ferranti, 2016). Spectroscopic techniques have great potential for discrimination of food materials. One promising and widely used technique in this context is near infrared (NIR) spectroscopy, a rapid and non-destructive technique. NIR enables preliminary monitoring of different types of food and as an analytical technique is able to give qualitative and quantitative information about complex samples (Abasi, Minaei, Jamshidi, & Fathi, 2018; Lohumi, Lee, Lee, & Cho, 2015; Prieto, Roehe, Lavín, Batten, & Andrés, 2009).

Developments in instrumentation technology have led to the availability of portable spectroscopic devices. Modern handheld NIR instruments that have been developed for food and drug quality control are fast, lightweight and relatively inexpensive. The trade-off for using these devices is that the spectral region and resolution are limited compared to benchtop technologies (Modroño, Soldado, Martínez-Fernández, & de la Roza-Delgado, 2017; Pasquini, 2018; Zamora-Rojas, Pérez-Marín, De Pedro-Sanz, Guerrero-Ginel, & Garrido-Varo, 2012). Additionally, scattering effects and instrumental and ambient noise make robust chemometric and machine learning methods crucial to extract the relevant information from the spectra (Arvanitoyannis & Van Houwelingen-Koukaliasoglou, 2003; Curran et al., 2018).

Previously, different chemometric and machine learning approaches such as principal component analysis (PCA), partial least squares (PLS), artificial neural network (ANN), linear discriminant analysis (LDA) and support vector machine (SVM) have been used for the analysis of handheld NIR spectra in relation to food research (Acquarelli et al., 2017; Arvanitoyannis & Van Houwelingen-Koukaliasoglou, 2003; Ballabio, Consonni, & Todeschini, 2009; Brereton & Lloyd, 2010; Efenberger-Szmechtyk, Nowak, & Kregiel, 2018; Risoluti, Gregori, Schiavone, & Materazzi, 2018; Zontov, Balyklova, Titova, Rodionova, &

Pomerantsev, 2016). However, these methods perform poorly for exploration and classification of complex analytical problems like freshness and growth system of food samples. Additionally, these methods often need data preprocessing and selection of the best preprocessing strategy is challenging on its own (Rinnan, 2014; Rinnan, Berg, & Engelsen, 2009).

In the present contribution, a powerful machine learning algorithm is used based on ensemble learning (Merkwirth et al., 2004; Rokach, 2010). This method splits the data into multiple parts and combines the best models for the different parts (of the NIR spectra) to come to a majority vote classification. Random subspace discriminant ensemble (RSDE) (Ho, 1998) is proposed here as a fast and reliable method to use handheld NIR devices for food authenticity. The simplicity of the different components of our methodology will allow for “Measure & Monitor” technology to evaluate food authenticity. The goals of the presented research were (1) discrimination of fresh (*Fr*) and Thawed (*Th*) samples and (2) discrimination of growth systems based on handheld NIR spectra from three recording modes of on meat (*OM*), through the top of the package (*TP*) and through the package held bottom up (*TB*), such that the meat touched the covering foil.

## 2. Materials and methods.

### 2.1. Sampling and data collection

Fresh chicken breast fillet samples were kindly provided Albert Heijn B.V. (The Netherlands) and Musgraves Group Ltd. (Ireland) in their standard supermarket packages in June 2015. The animal welfare classification system differs between the countries of origin.

Albert Heijn B.V. has provided a set of 70 fresh chicken fillet samples from different production systems and batches, divided over a time span of 3 weeks. Animal welfare was expressed on the packaging by “no star” representing the lowest level of welfare and three stars representing the highest level of welfare.

Conventional chicken (CONV) (18 samples)

Free-range (1 star, 1\*) (17 samples)

Specialty (2 stars, 2\*) (17 samples)

Organic (ORG) (3 stars, 3\*) (18 samples)

In the same period, Musgraves Group has provided a set of 83 fresh chicken fillet samples from different production systems and batches, divided over a time span of 2 weeks:

Standard chicken (STD) (18 samples)

Free range (FR) (15 samples)

Corn fed chicken (CF) (15 samples)

Marinated chicken (MAR) (35 samples)

Samples (153 total) were shipped in ‘fresh packs’, guaranteeing a temperature between 4 and 7°C for 96h. Samples arrived within that time-span. Marinated chicken fillets acted as controls, since these were expected to be highly identifiable.

Thawed samples (133 in total) were obtained by freezing at  $-18^{\circ}\text{C}$  for 48 h and thawing for 24 h at  $+4^{\circ}\text{C}$ . Twenty fresh samples were used for  $\beta$ -hydroxyacyl-CoA-dehydrogenase (HADH) reference measurements (13% of the total sample set) to assess their storage history, i.e. whether the samples had been chilled or frozen (Boerrigter-Eenling, Alewijn, Weesepeel, & van Ruth, 2017). For the Dutch set, three samples of each class were used for HADH, whilst for the Irish set two samples were subjected to HADH per category. No deviations were found in the freshness of the samples. Samples which were subjected to HADH measurements were not subjected to NIR measurements for the thawed category. No reference methods were available for confirmation of the growing system of the chicken fillet samples. Providers have confirmed that the indicated growing system is correct. Note that growth conditions may be similar across countries (e.g., CONV and STD), but that different labels have been attached in order to classify between-country variation.

NIR data was acquired using a MicroNIR Pro NIR (Viavi Solutions, Milpitas, CA, USA), powered by the MicroNIR Pro software (version 2.2, Viavi Solutions) in diffuse reflectance mode in wavelength range of approximately 908-1676 nm with an evenly distributed spectral resolution, resulting in 125 variables/measurement. A 99% white diffuse reflectance standard was used for calibration followed by a dark measurement. This calibration was repeated in 10 minute cycles. The 153 chicken fillet samples were subjected to non-destructive NIR measurements by applying the NIR with standard collar in three different ways: on meat (*OM*), through package (*TP*) and through packaging bottom up (*TB*). First, *TP* measurements were acquired by placing the package on a flat surface and applying the NIR on the transparent top foil without pressure above the fillet sample. In most cases an air pocket was between the foil and the sample. Secondly, the *TB* measurements were performed by flipping the package bottom up, letting the fillet sample lean on the top transparent foil, followed by NIR measurements through this transparent foil. Finally, the transparent top foil was removed and NIR measurements were taken directly on the fillet sample without applying considerable pressure. Prior to freezing, the fillet package was covered with a new layer of identical transparent top foil. Five replicates were taken per *OM/TP/TB*, with a total of 4590 raw NIR measurements. Scheme A1 of the appendix illustrates how the samples were collected.

## 2.2. Data handling and preprocessing

Spectral data was labelled to ensure replicate measurements and measurements from different modes of the same fillet could be connected. Training and test sets were created using the Duplex method (Daszykowski, Walczak, & Massart, 2002; Puzyn, Mostrag-Szlichtyng, Gajewicz, Skrzyński, & Worth, 2011; Snee, 1977) on the entire data set in order to ensure a representative test set including boundary cases (Reitermanova, 2010; Westad & Marini, 2015). All classes were represented in the test set. Importantly, all measurements (i.e., spectra) of a sample were assigned to either training set (70%) or test set (30%) in order to ensure that the test set did not include data from the same sample that the model was trained on.

When looking for the optimum pre-processing strategy a design of experiments was used (Gerretzen et al., 2015). The predictive classification model was built (PLS-DA, CP-ANN, SVM and RSDE) and validated using cross validation (CV). In every CV, spectra belonging to the same chicken sample were removed from the train set (leave-chicken-out).



After training, tuning and evaluation of the model, the test set was used for final performance estimation. The data analysis pipeline of the presented work is shown in Scheme 1.

*Scheme 1. Data analysis pipeline for each presented study.*

[Scheme 1 about here]

### 2.3. Random subspace discriminant ensemble.

Classification of the NIR spectra was done using Random Subspace Discriminant Ensemble. This method divides the spectra into a number of random subspaces (30 random subspaces as standard in this case), selected from the spectral domain (e.g., a random subset of 60 wavelengths is the default in Matlab). Discriminant analysis (DA) was used to classify the spectra in each subspace (Ho, 1998; Tan, Li, & Qin, 2008). Each subspace may result in different classification probabilities. These probabilities are combined by taking their average across all subspaces to come to a single classification model of the full spectra. Figure 1 shows the general architecture of RSDE algorithm. The potential of RSDE in high dimensional data comes from the fact that each model requires only a limited number of variables.

*Figure 1. Principle of the RSDE framework (sequential subspaces used for illustrative purposes).*

[Insert Figure 1 about here]

### 2.4. Software

Chemometric data analysis was performed in MATLAB environment R2016a, with the exception of the leave-class-out validation (Section 3.5), which was done in R2019B (Mathworks, MA, USA). The PLS-Toolbox v7.8 (Eigenvector, WA, USA) was used for PLS-DA modelling, the pre-processing toolbox (Gerretzen et al., 2015) was used to choose the best preprocessing strategy (based on an experimental design), the CP-ANN toolbox (Milano Chemometrics and QSAR Research Group) was used for optimization of the Kohonen network and supervised classification and the Classification Learner toolbox of MATLAB was used for SVM and RSDE modelling

## 3. Classification of NIR spectra

### 3.1. Fresh vs. Thawed

The RSDE algorithm was first used to discriminate *Fr* and *Th* samples for each of the three different spectral recording modes. For the preprocessing of the data, an experimental design was used to find the best strategy with minimal classification error (Gerretzen et al., 2015). Classification performance was evaluated using accuracy (*Acc*), precision (*Pre*), sensitivity (*Sen*), specificity (*Spe*) and error rate (*ER*) (Ballabio & Consonni, 2013).

Figure 2 shows the NIR spectra of *Fr* and *Th* samples in three different spectral recording modes of OM/TP/TB. Coloring the spectra by recording mode shows that there are clear differences in absorbance related to how the spectra were obtained. The differences are similar for *Fr* and *Th* samples. Because *Fr* and *Th* samples have similar spectra, the first challenge in this study was to discriminate *Fr* and *Th* samples based on NIR spectra.

Figure 2. NIR spectra from different modes of fresh (left) and thawed (right) chicken fillets

[Insert Figure 2 about here]

The RSDE performed well in discriminating individual spectra of *Fr* and *Th* samples; *Acc* values were 90.2% for training set, 87.6% for cross validation (CV) and 85.2% for test set of *OM* samples. For *TP* samples, the values were 96.4%, 95.4% and 92.0% for train, CV and test sets, respectively. The *Acc* values for *TB* samples were respectively 95.4%, 93.3% and 91.0% for train, CV and test sets. Details on the classification power can be found in Table A1 in the appendix.

### 3.2 Classification in growth conditions

The ability of the RSDE method to classify individual NIR spectra was promising. The next objective was to evaluate whether the RSDE method could also be used to discriminate between the growth systems of the chickens. The RSDE algorithm was used for classification of seven growth conditions of 1\*/2\*/ORG/CONV/STD/FR/CF as well as MAR samples in *Fr* and *Th* conditions in *OM*, *TP* and *TB* modes (details in section 2.1). As an example, Figure 3 depicts the discrimination performance of RSDE for classification of chickens in different growth conditions in terms of *Acc* in *OM*, *TP* and *TB* modes. As can be seen, the values of *Acc* for training, validation and test sets are between 80-90 % for *OM* (Figure 3a), *TP* (Figure 3b) and *TB* (Figure 3c). The values of *Acc* are low because of the complexity of the samples and similarity in the NIR spectra of samples in different conditions.

To compare the results of RSDE model with common classification methods in chemometrics, NIR data of chickens in different growth conditions were classified by partial least squares-discriminant analysis (PLS-DA) (Ballabio & Consonni, 2013; Gromski et al., 2015), counter propagation-artificial neural network (CP-ANN) (Ballabio et al., 2009; Ballabio & Vasighi, 2012) and support vector machine with quadratic kernel function (Q-SVM) (Brereton & Lloyd, 2010). Model performance of the RSDE was better than that of the other methods. In Figure 3, the classification results for PLS-DA, CP-ANN and Q-SVM for training, validation and test sets in terms of *Acc* are shown in comparison with RSDE. Due to the type of subspace selection, the RSDE is only slightly affected by noise and is less prone to overfitting (shown by similar *Acc* values for train, validation and test sets).

Figure 3. Classification *Acc* in *OM* (a), *TP* (b) and *TB* (c) modes.

[Insert Figure 3 about here]

For PLS-DA, the best preprocessing strategy was chosen according to experimental design approach (lowest classification error) (Gerretzen et al., 2015). In this regard, mean-centering and pareto scaling were the best pre-processing strategies. Other attempts such as outlier detection using Q-residuals/Hotelling's  $T^2$  (Ballabio & Consonni, 2013) and variable selection using variable importance in projection (VIP) with "greater than one" rule (Andersen & Bro, 2010) were performed to improve PLS-DA classification. These methods slightly improved the models but not to acceptable levels (see Table A2 for more details).

For CP-ANN, firstly, the genetic algorithm (GA) (Ballabio, Vasighi, Consonni, & Kompany-Zareh, 2011) was used to optimize the network topology including neurons and

number of epochs, resulting in 40 neurons and 150 epochs. As shown in Figure 3, the performance of CP-ANN is not good in the CV and test sets.

In SVM, the quadratic kernel gave the best accuracy (among linear, quadratic, cubic and radial basis function) (Brereton & Lloyd, 2010). The performance of Q-SVM was better than PLS-DA and CP-ANN in terms of *Acc* (see Table A3 for more details of Q-SVM performance), but were still far from ideal (accuracy values were below 77.7%). In summary, going from linear PLS-DA to non-linear CP-ANN and Q-SVM improved classification performance, but results were deemed insufficient.

The RSDE outperformed other classification methods for discrimination of growth conditions. To obtain a more detailed view of the classification power of this method, the classification performance of RSDE in terms of *Acc*, *Sen*, *Prec* for *Fr* samples in *OM* mode is presented here; *Acc* value for eight classes was 79.4 %, the *Sen* values ranged from 55.8-95.4 % and *Prec* values ranged from 63.6-90.5% for the test set (467 spectra). Though the *Acc* value of RSDE (79.4%) was significantly higher than that of the closest *Acc* of Q-SVM (79.4% vs. 71.1%,  $z = 2.9389$ ,  $p = .00164$ ) the classification performance strongly showed room for improvement. Table A3 shows more details of the classification performance of RSDE in terms of *Acc*, *Sen*, *Prec* for *Fr* samples in *OM* mode.

One of the surprising aspects of RSDE algorithm is its insensitivity to preprocessing. In other words, conventional chemometric spectral preprocessing does not affect the performance of this algorithm and therefore, raw data can be used as input for this algorithm (Figure A1 and Table A4) (Ho, 1998; Tan et al., 2008; Zheng, Hu, Tong, & Du, 2014). Additionally, the detector of the MicroNIR is especially sensitive in the region of approximately 1425 – 1575 nm. In the raw spectra in Figure 2 it could be observed that absorbance units of 3-3.5 were recorded. The detector operated at its limits in this region and noise is visible with some large spikes. Still, there were no issues in classifying the samples, including in the external model validation, indicating the power of RSDE in NIR spectroscopy.

### 3.3 Combining modes

In the previous analyses, we classified single NIR measurements. Of course, it is also possible to take multiple NIR scans of a sample through multiple sample handling protocols (i.e., *OM*, *TP* and *TB*) and to combine the spectra (Borràs et al., 2015). This is cost-insensitive as multiple measurements are easy to obtain with handheld technologies. By simply concatenating the measurements of *OM*, *TP* and *TB* NIR spectra, we can boost the performance of RSDE. In this manner, the spectral dimension is increased and the RSDE has more flexibility to select random subspaces and as a result classification may be improved.

Two different options for data combination were tested; (i) different measurement modes (i.e., *TP/TB* for consumers and *OM/TP/TB* for food administration) and (ii) multiple spectra from same mode (i.e., *OM/OM/OM*, *TP/TP/TP* and *TB/TB/TB*). To combine different measurement modes, we randomly selected one of replicate spectra from each mode *TP* and *TB* (and *OM*) of a sample to simulate ‘uncontrolled consumer measurements’. Table 1 gives detailed classification *Acc* values of the RSDE when classifying the individual or combined NIR spectra. The results in Table 1 confirmed that the data combination resulted in improvement of the classification over individual modes.

For the second combination method (multiple measurements of the same mode) there was no significant improvement of the classification performance compared to the individual measurements. Single measurements closer to the meat lead to better performance of the RSDE, as can be seen from the increasing values going from column 3 to column 5. Apparently, a single measurement is already highly representative of the sample, and combining data will improve the classification performance only if new aspects of the sample are added to the data (e.g., measurements in *OM*, *TP* and *TB*).

Table 1. Accuracy of RSDE (in %) for individual and combined spectra.

[Insert Table 1 about here]

For a fair comparison with other classification methods, the *TP/TB* and *OM/TP/TB* combined data sets were also analysed with Q-SVM, CP-ANN and PLS-DA. In summary, the RSDE again strongly outperformed the other methods. The classification accuracy of 99.4% for the *OM/TP/TB* test set was significantly higher than the 82% of the Q-SVM ( $z = 5.3586$ ,  $p < .00001$ ), the 65.4% of the CP-ANN and the 63.5% of the PLS-DA. Also the classification accuracy of 96.9% for the *TP/TB* test set was significantly higher than the 81.3% of the Q-SVM ( $z = 4.4773$ ,  $p < .00001$ ), the 63.6% of the CP-ANN and the 62.4% of the PLS-DA. More details on the differences between models for the classification of *TP/TB* and *OM/TP/TB* test set (160 spectra) are provided in Table A5.

Efforts to validate the developed RSDE models were made by using two shuffling methods (y-randomization and the permutation test) (Rücker, Rücker, & Meringer, 2007). After permutation, classification accuracy of the RSDE deteriorated. As an example, in the classification of growth conditions of *Fr* samples in *OM* mode, CV classification accuracy was reduced from 99.0% (non-shuffled) to an average of 17.4% for permuted data. For the combined data, it is noteworthy that the RSDE is so powerful that it can get >80% accuracy in a training set. However, cross validation and test set reveal that no structure was present in the data, as the accuracies drop to values which are no better than random assignment (See Table A6 for more details).

### 3.4. Leave-class-out analysis

In the previous analyses, the RSDE model was trained on data from all growth conditions. But what if spectra from a not before seen growth condition were classified by the RSDE? To evaluate this, a final study was done based on a leave-class-out (LCO) methodology. Several RSDE models were trained (and cross-validated) using  $8 - 1 = 7$  classes, while the left out class was completely used as a test set. This method was used 8 times, such that every class was left out, and classified, once.

It is important to note that RSDE classifications are done according to the highest classification probability, regardless of the absolute value of that probability. Because there is no 'correct classification' in the LCO situation, cut-offs were imposed on the classification probability, before a classification would be accepted. This can protect a researcher from classifying highly deviating spectra. The results of the LCO analysis, with varying cut-offs are shown in Table A6. The classification accuracy of each 7-class models was over >98.5 % (see column 2). One of the a priori expectation was to find that CONV and STD (having similar growth conditions) would be classified as the other in most situations. By increasing the minimum classification probability, we expected to see this pattern more clearly.

Even when no cut-off was used for the classification probability, the results defied expectations. The CONV spectra were classified as either 1\* (63.3%) or ORG (36.7%), while the majority (54.4%) of STD spectra were classified as FR. Furthermore, the majority (66.7%) of the 1\* and the majority (83.3%) of ORG spectra were classified as CONV. Somewhat in line with expectations was that CF spectra were mainly classified as FR (74.7%) and FR spectra as CF (62.7%).

Though the total number of spectra that could be classified decreased when we increased the minimum classification probability the results became more distinct (See sub-tables of Table A7). For example, with a cut-off of .90, an unexpected pattern became apparent. Namely, the classifications reveal that Dutch fillets (1\*, 2\*, ORG and CONV) are always classified as other Dutch fillets, and Irish (FR, STD, CF and MAR) fillets are always classified as other Irish fillets. We expect that this pattern is related to the difference in lifetime of the chickens (Irish chickens on average live longer than Dutch chickens). Interestingly, the control fillets of the MAR condition work well in the sense that they are mostly classified as STD, and much less as the more premium fillets FR and CF.

#### 4. Conclusion

A RSDE was used as a fast and reliable machine learning algorithm for authentication of the growth condition of chicken fillets and their freshness within a thoroughly validated chemometric workflow with several specific practical implementations. The RSDE considerably outperformed other common classification models such as PLS-DA, CP-ANN and SVM. Also, combining spectra improved the classification performance of this method even further. We demonstrated that the use of a relatively inexpensive portable device was able to provide very fast (~3 s per measurement) results in the application of NIR spectroscopy in food authenticity. Considering the measurement time of approximately 8s, ~3.0s per NIR measurement and a few seconds to flip the package, a complete analysis (measure and monitor) would require approximately 20s including data analysis. The combination of handheld NIR with RSDE algorithm may offer a very interesting and reliable tool for monitoring meat authenticity (and quality) directly in the field.

The RSDE algorithm was so powerful that it could not only clearly discriminate between NIR spectra based on the growth conditions of the chickens, the Leave-Class-Out validation provided the authors with new insights about the differences between country of origin and the differences in meat. Our analyses do indicate that some adjustments to the existing implementation are needed before the methodology can be applied in a real-world setting. Imposing minimal classification probabilities can protect from classifying meat of known origin (i.e., chicken) into just any class. However, it is not advised to use this method for classification of meats from unknown origin (i.e., other animals). Therefore, future work could implement a pre-screening based on, for example, the Mahalanobis distance of a new spectrum to the spectra of the known classes. After these adaptation, the combined approach presented in this work is very fast and if applied throughout the supply chain, it could improve the quality of meat that reaches consumers' tables in everyday life.

**Acknowledgements**

The authors would like to thank Albert Heijn B.V. (The Netherlands) and Musgrave Group Ltd. (Ireland) for providing chicken fillet samples. This work was made possible through financial support from Sharif University of Technology (SUT) and the Netherlands Organisation for Scientific Research (NWO) through the PTA-COAST3 consortium “Outfitting the Factory of the Future with Online Analysis (OFF/On)”



## References

- Abasi, S., Minaei, S., Jamshidi, B., & Fathi, D. (2018). Dedicated non-destructive devices for food quality measurement: A review. *Trends in Food Science and Technology*, 78, 197-205. doi:10.1016/j.tifs.2018.05.009
- Acquarelli, J., van Laarhoven, T., Gerretzen, J., Tran, T. N., Buydens, L. M. C., & Marchiori, E. (2017). Convolutional neural networks for vibrational spectroscopic data analysis. *Analytica Chimica Acta*, 954, 22-31. doi:10.1016/j.aca.2016.12.010
- Andersen, C. M., & Bro, R. (2010). Variable selection in regression-a tutorial. *Journal of Chemometrics*, 24(11-12), 728-737.
- Arvanitoyannis, I. S., & Van Houwelingen-Koukaliaroglou, M. (2003). Implementation of Chemometrics for Quality Control and Authentication of Meat and Meat Products. *Critical Reviews in Food Science and Nutrition*, 43(2), 173-218. doi:10.1080/10408690390826482
- Ballabio, D., & Consonni, V. (2013). Classification tools in chemistry. Part 1: Linear models. PLS-DA. *Analytical Methods*, 5(16), 3790-3798. doi:10.1039/c3ay40582f
- Ballabio, D., Consonni, V., & Todeschini, R. (2009). The Kohonen and CP-ANN toolbox: A collection of MATLAB modules for Self Organizing Maps and Counterpropagation Artificial Neural Networks. *Chemometrics and Intelligent Laboratory Systems*, 98(2), 115-122. doi:10.1016/j.chemolab.2009.05.007
- Ballabio, D., & Vasighi, M. (2012). A MATLAB toolbox for Self Organizing Maps and supervised neural network learning strategies. *Chemometrics and Intelligent Laboratory Systems*, 118, 24-32. doi:10.1016/j.chemolab.2012.07.005
- Ballabio, D., Vasighi, M., Consonni, V., & Kompany-Zareh, M. (2011). Genetic Algorithms for architecture optimisation of Counter-Propagation Artificial Neural Networks. *Chemometrics and Intelligent Laboratory Systems*, 105(1), 56-64. doi:10.1016/j.chemolab.2010.10.010
- Boerrigter-Eenling, R., Alewijn, M., Weesepeel, Y., & van Ruth, S. (2017). New approaches towards discrimination of fresh/chilled and frozen/thawed chicken breasts by HADH activity determination: Customized slope fitting and chemometrics. *Meat science*, 126, 43-49.
- Borràs, E., Ferré, J., Boqué, R., Mestres, M., Aceña, L., & Busto, O. (2015). Data fusion methodologies for food and beverage authentication and quality assessment - A review. *Analytica Chimica Acta*, 891, 1-14. doi:10.1016/j.aca.2015.04.042
- Brereton, R. G., & Lloyd, G. R. (2010). Support Vector Machines for classification and regression. *Analyst*, 135(2), 230-267. doi:10.1039/b918972f
- Curran, K., Underhill, M., Grau-Bové, J., Fearn, T., Gibson, L. T., & Strlič, M. (2018). Classifying Degraded Modern Polymeric Museum Artefacts by Their Smell. *Angewandte Chemie - International Edition*, 57(25), 7336-7340. doi:10.1002/anie.201712278
- Daszykowski, M., Walczak, B., & Massart, D. (2002). Representative subset selection. *Analytica chimica acta*, 468(1), 91-103.
- Efenberger-Szmechtyk, M., Nowak, A., & Kregiel, D. (2018). Implementation of chemometrics in quality evaluation of food and beverages. *Critical Reviews in Food Science and Nutrition*, 58(10), 1747-1766. doi:10.1080/10408398.2016.1276883
- Gallo, M., & Ferranti, P. (2016). The evolution of analytical chemistry methods in foodomics. *Journal of Chromatography A*, 1428, 3-15. doi:10.1016/j.chroma.2015.09.007
- Gerretzen, J., Szymańska, E., Jansen, J. J., Bart, J., Van Manen, H. J., Van Den Heuvel, E. R., & Buydens, L. M. C. (2015). Simple and Effective Way for Data Preprocessing Selection Based on Design of Experiments. *Analytical Chemistry*, 87(24), 12096-12103. doi:10.1021/acs.analchem.5b02832
- Gromski, P. S., Muhamadali, H., Ellis, D. I., Xu, Y., Correa, E., Turner, M. L., & Goodacre, R. (2015). A tutorial review: Metabolomics and partial least squares-discriminant analysis - a marriage of convenience or a shotgun wedding. *Analytica Chimica Acta*, 879, 10-23. doi:10.1016/j.aca.2015.02.012



- Ho, T. K. (1998). The random subspace method for constructing decision forests. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(8), 832-844. doi:10.1109/34.709601
- Lohumi, S., Lee, S., Lee, H., & Cho, B. K. (2015). A review of vibrational spectroscopic techniques for the detection of food authenticity and adulteration. *Trends in Food Science and Technology*, 46(1), 85-98. doi:10.1016/j.tifs.2015.08.003
- Merkwirth, C., Mauser, H., Schulz-Gasen, T., Roche, O., Stahl, M., & Lengauer, T. (2004). Ensemble methods for classification in cheminformatics. *Journal of Chemical Information and Computer Sciences*, 44(6), 1971-1978. doi:10.1021/ci049850e
- Modroño, S., Soldado, A., Martínez-Fernández, A., & de la Roza-Delgado, B. (2017). Handheld NIRS sensors for routine compound feed quality control: Real time analysis and field monitoring. *Talanta*, 162, 597-603. doi:10.1016/j.talanta.2016.10.075
- Pasquini, C. (2018). Near infrared spectroscopy: A mature analytical technique with new perspectives – A review. *Analytica Chimica Acta*, 1026, 8-36. doi:10.1016/j.aca.2018.04.004
- Pischetsrieder, M. (2018). Global Food-Related Challenges: What Chemistry Has Achieved and What Remains to Be Done. *Angewandte Chemie - International Edition*, 57(36), 11476-11477. doi:10.1002/anie.201803504
- Prieto, N., Roehe, R., Lavín, P., Batten, G., & Andrés, S. (2009). Application of near infrared reflectance spectroscopy to predict meat and meat products quality: A review. *Meat Sci*, 83(2), 175-186. doi:10.1016/j.meatsci.2009.04.016
- Puzyn, T., Mostrag-Szlichtyng, A., Gajewicz, A., Skrzyński, M., & Worth, A. P. (2011). Investigating the influence of data splitting on the predictive ability of QSAR/QSPR models. *Structural Chemistry*, 22(4), 795-804. doi:10.1007/s11224-011-9757-4
- Reid, L. M., O'Donnell, C. P., & Downey, G. (2006). Recent technological advances for the determination of food authenticity. *Trends in Food Science and Technology*, 17(7), 344-353. doi:10.1016/j.tifs.2006.01.006
- Reitermanova, Z. (2010). *Data splitting*. Paper presented at the WDS.
- Rinnan, Å. (2014). Pre-processing in vibrational spectroscopy-when, why and how. *Analytical Methods*, 6(18), 7124-7129. doi:10.1039/c3ay42270d
- Rinnan, Å., Berg, F. v. d., & Engelsen, S. B. (2009). Review of the most common pre-processing techniques for near-infrared spectra. *TrAC - Trends in Analytical Chemistry*, 28(10), 1201-1222. doi:10.1016/j.trac.2009.07.007
- Risoluti, R., Gregori, A., Schiavone, S., & Materazzi, S. (2018). "click and Screen" Technology for the Detection of Explosives on Human Hands by a Portable MicroNIR-Chemometrics Platform. *Analytical Chemistry*, 90(7), 4288-4292. doi:10.1021/acs.analchem.7b03661
- Rokach, L. (2010). Ensemble-based classifiers. *Artificial Intelligence Review*, 33(1-2), 1-39. doi:10.1007/s10462-009-9124-7
- Rücker, C., Rücker, G., & Meringer, M. (2007). Y-randomization and its variants in QSPR/QSAR. *Journal of Chemical Information and Modeling*, 47(6), 2345-2357. doi:10.1021/ci700157b
- Sentandreu, M. Á., & Sentandreu, E. (2014). Authenticity of meat products: Tools against fraud. *Food Research International*, 60, 19-29. doi:10.1016/j.foodres.2014.03.030
- Snee, R. D. (1977). Validation of regression models: methods and examples. *Technometrics*, 19(4), 415-428.
- Tan, C., Li, M., & Qin, X. (2008). Random subspace regression ensemble for near-infrared spectroscopic calibration of tobacco samples. *Analytical Sciences*, 24(5), 647-653. doi:10.2116/analsci.24.647
- Vlachos, A., Arvanitoyannis, I. S., & Tserkezou, P. (2016). An Updated Review of Meat Authenticity Methods and Applications. *Critical Reviews in Food Science and Nutrition*, 56(7), 1061-1096. doi:10.1080/10408398.2012.691573
- Westad, F., & Marini, F. (2015). Validation of chemometric models—a tutorial. *Analytica chimica acta*, 893, 14-24.

- Zamora-Rojas, E., Pérez-Marín, D., De Pedro-Sanz, E., Guerrero-Ginel, J. E., & Garrido-Varo, A. (2012). Handheld NIRS analysis for routine meat quality control: Database transfer from at-line instruments. *Chemometrics and Intelligent Laboratory Systems*, 114, 30-35. doi:10.1016/j.chemolab.2012.02.001
- Zheng, K., Hu, H., Tong, P., & Du, Y. (2014). Ensemble Regression Coefficient Analysis for Application to Near-Infrared Spectroscopy. *Analytical Letters*, 47(13), 2238-2254. doi:10.1080/00032719.2014.900776
- Zontov, Y. V., Balyklova, K. S., Titova, A. V., Rodionova, O. Y., & Pomerantsev, A. L. (2016). Chemometric aided NIR portable instrument for rapid assessment of medicine quality. *Journal of Pharmaceutical and Biomedical Analysis*, 131, 87-93. doi:10.1016/j.jpba.2016.08.008

## **APPENDIX**

### **Scheme A1. Details of the data collection**

[Insert Scheme 1A here]

Journal Pre-proof

**Table A1. Classification results (in %) for freshness classification of chicken fillets based on *Fr* and *Th*. Sensitivity (*Sen*) and precision (*Pre*) are reported. Precision (*Pre*) indicates how confident one can be about the given classification**

[Insert Table A1 here]

**Table A2. PLS-DA classification results (in %) after outlier detection and variable selection. Sensitivity (*Sen*) and specificity (*Spe*) are reported.**

[Insert Table A2 here]

**Table A3. Classification performance (in %) of RSDE and Q-SVM models for classification of eight growth conditions. Precision (*pre*) indicates how confident one can be about the given classification. The test set contained 467 spectra.**

[Insert Table A3 here]

**Table A4. RSDE classification performance (in %) for raw and preprocessed Fr data in OM mode for classification of growth conditions.**

[Insert Table A4 here]

**Table A5. Comparing classification performances for the 160 combined *Fr* test set data into growth conditions.**

[Insert Table A5 here]

Journal Pre-proof



**Table A6. Permutation test (y-randomization) for evaluation of RSDE performance on combined data**

[Insert Table A6 here]

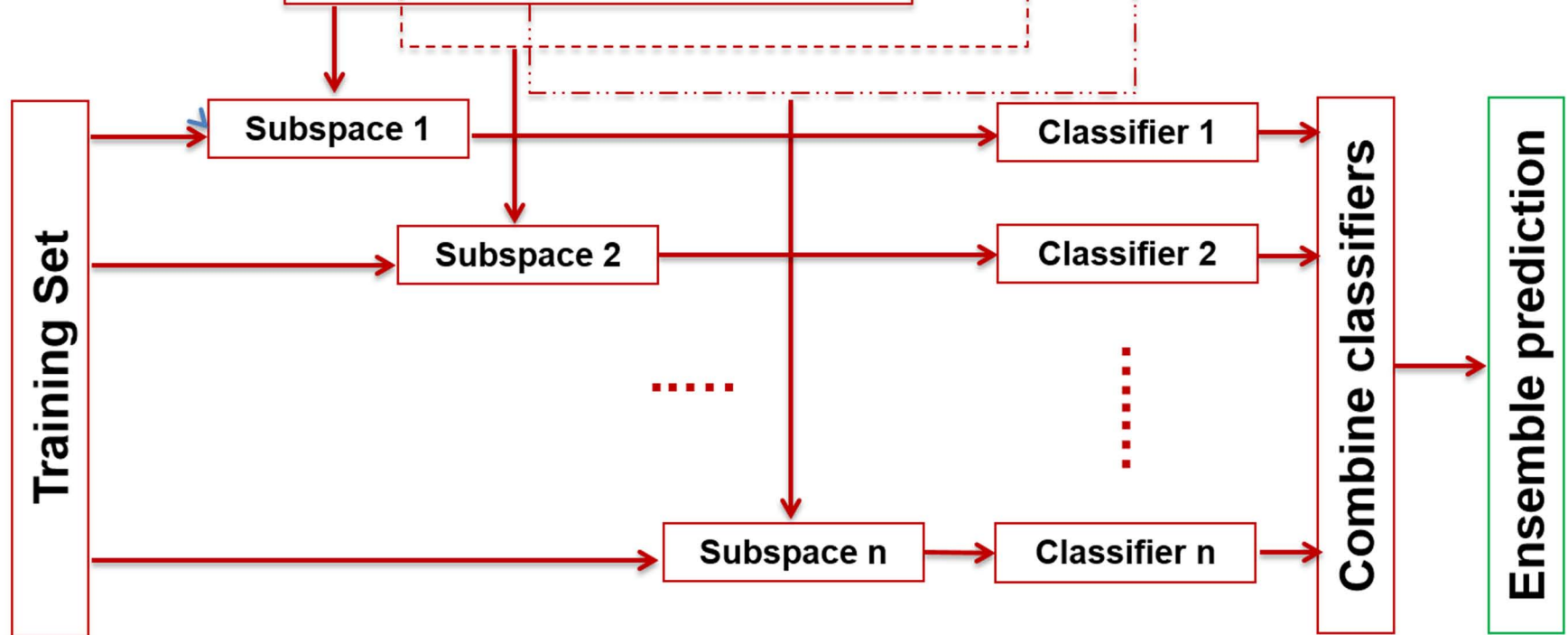
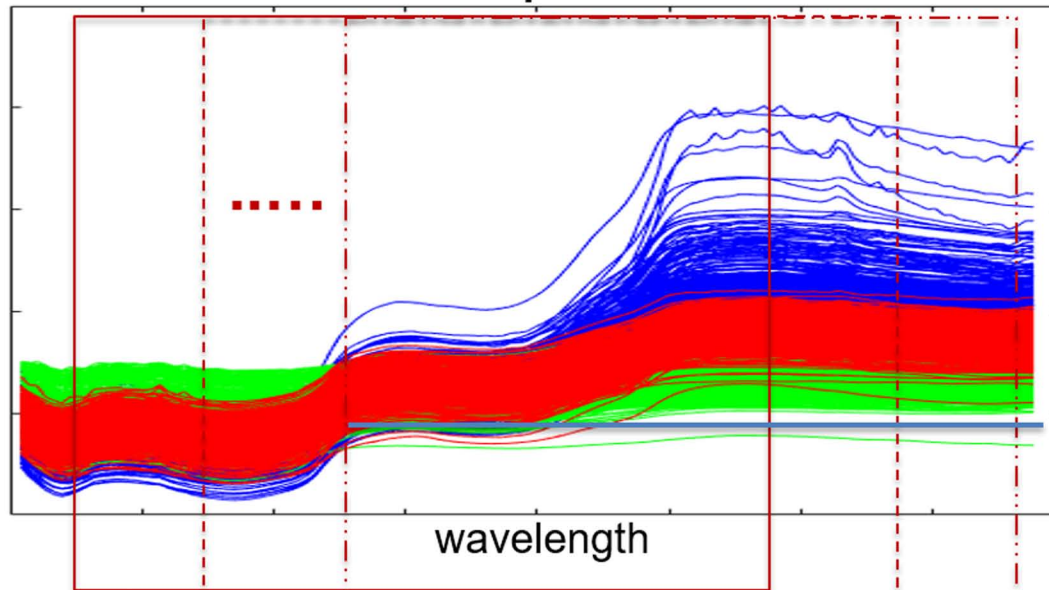
**Table A7. RSDE Classification performance for *Fr* combined data (*OM/TP/TB*) in leave-class-out validation. *Acc* values (column 2) are based on 10 Fold Cross validation of the remaining 7-class data. Percentages sum to 100 (to 1 decimal place) over rows.**

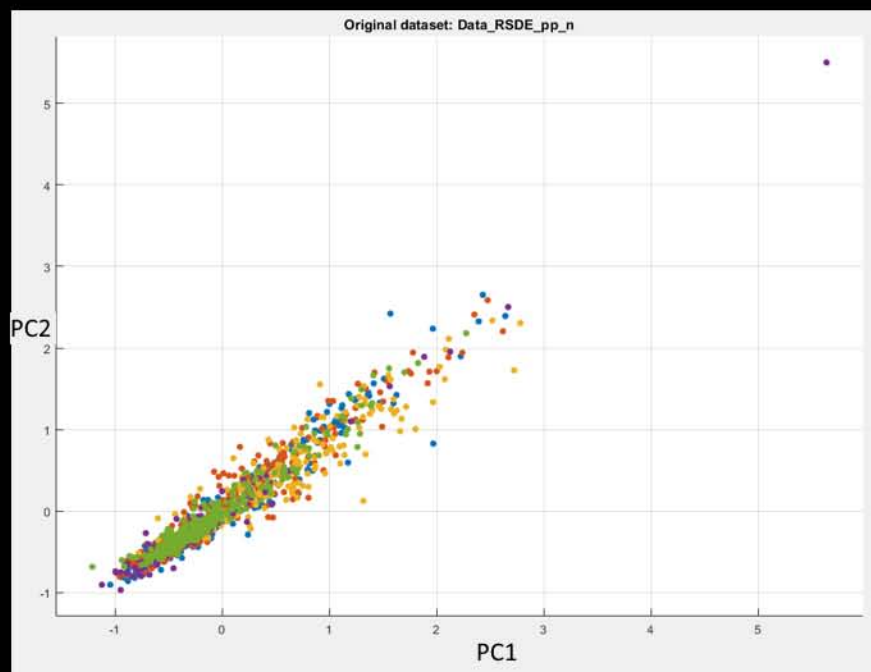
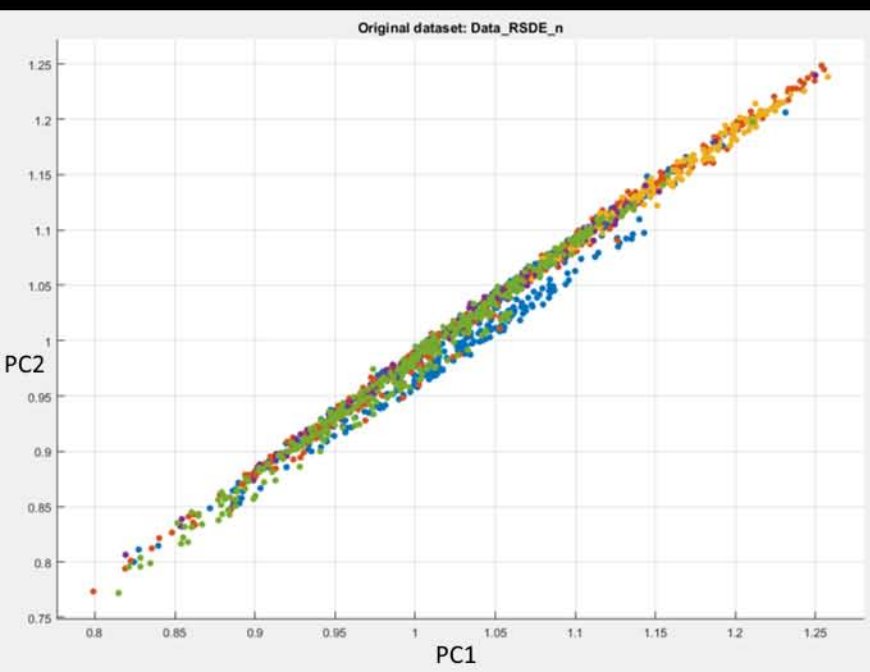
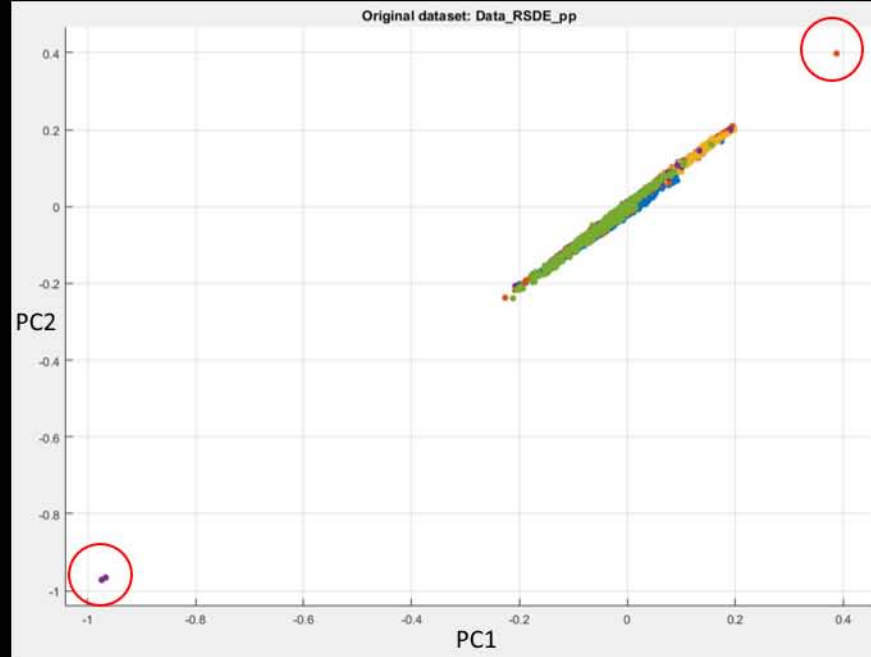
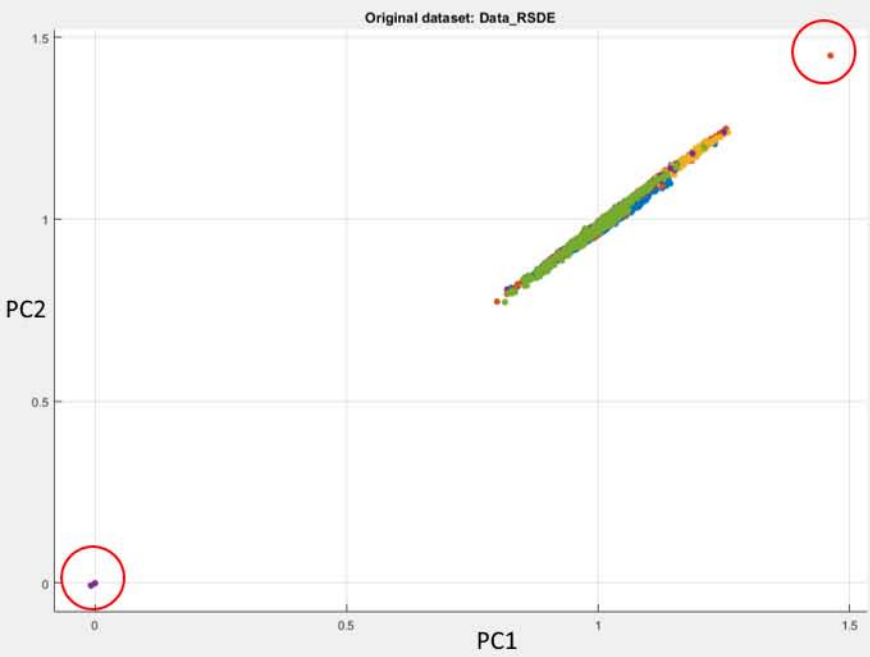
[Insert Table A7 here]

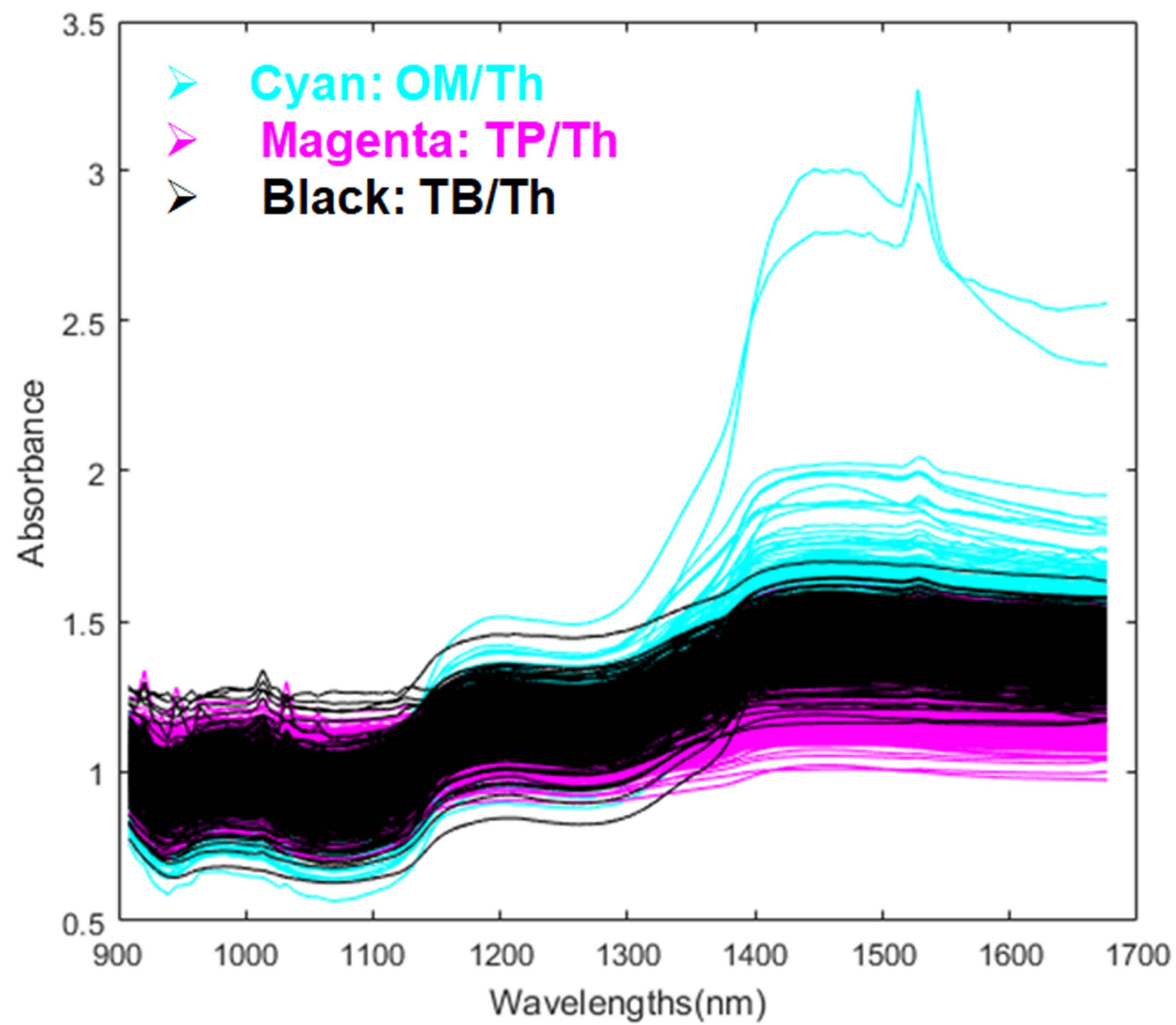
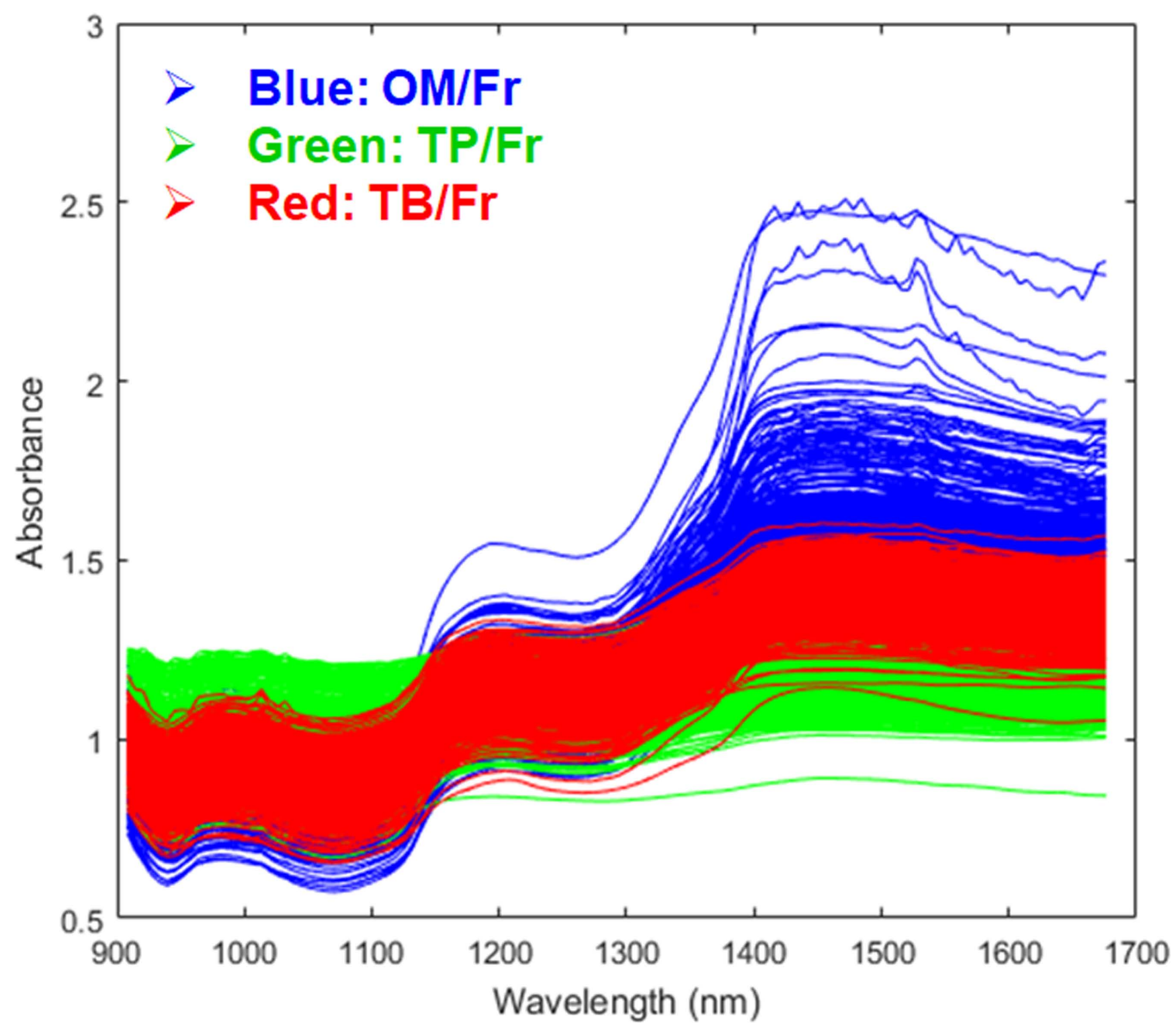
**Figure A1. Effect of preprocessing on data scattering in PCA space. Raw data (upper left), preprocessed data by mean centering (MC) and standard normal variate (SNV) (upper right), original data without outliers (bottom left) and preprocessed data after outlier removal using MC and SNV (bottom right). Red circles show extreme values which were removed in bottom figures according to Q-residuals and Hotelling's  $T^2$  tests.**

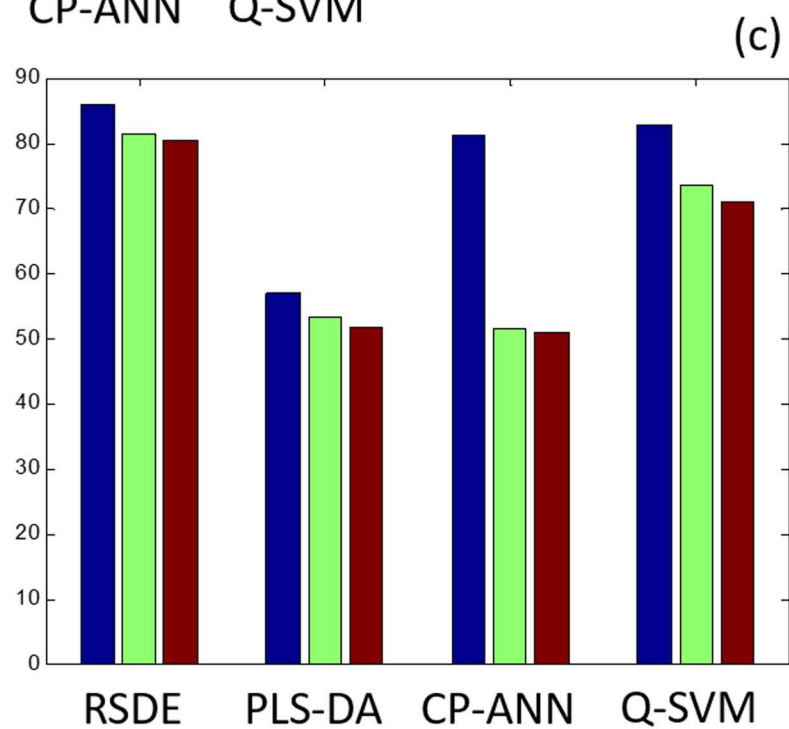
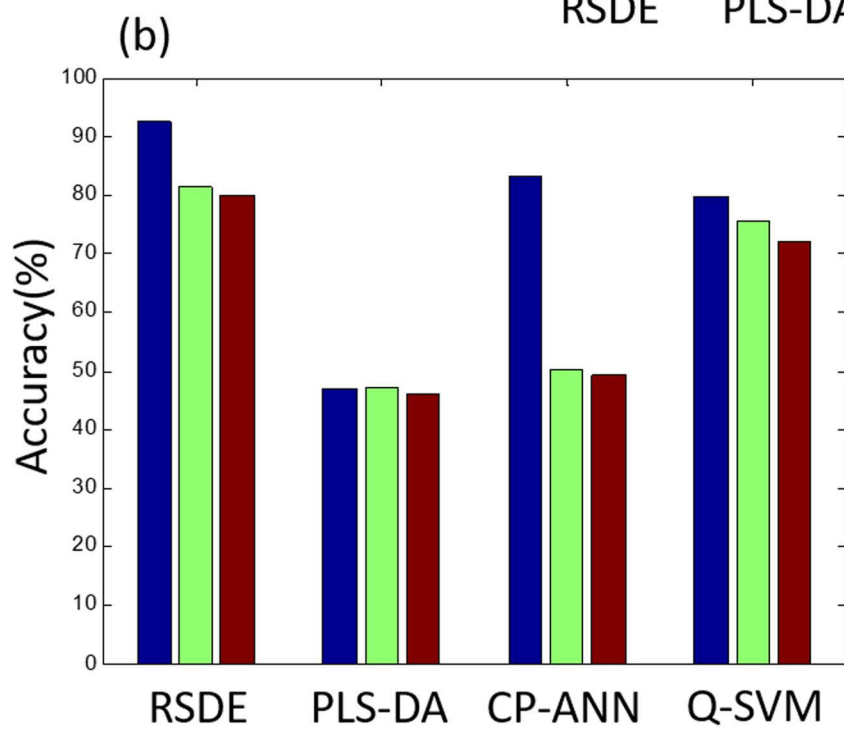
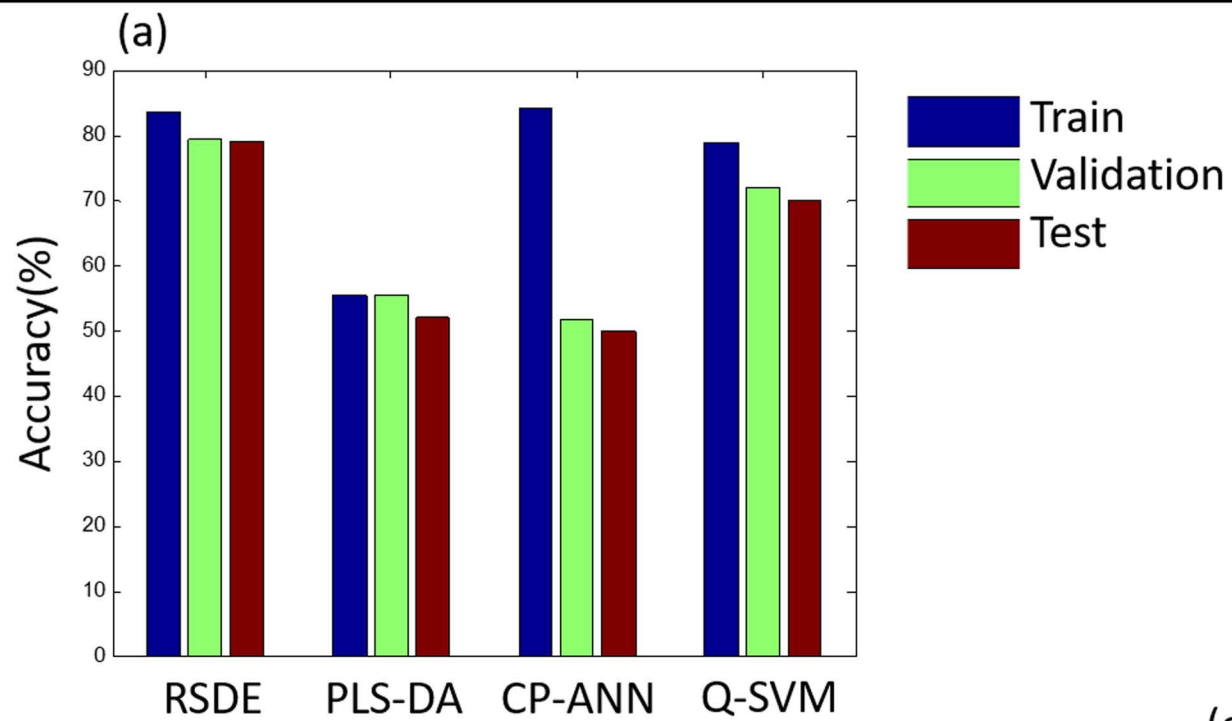
[Insert Figure A1 here]

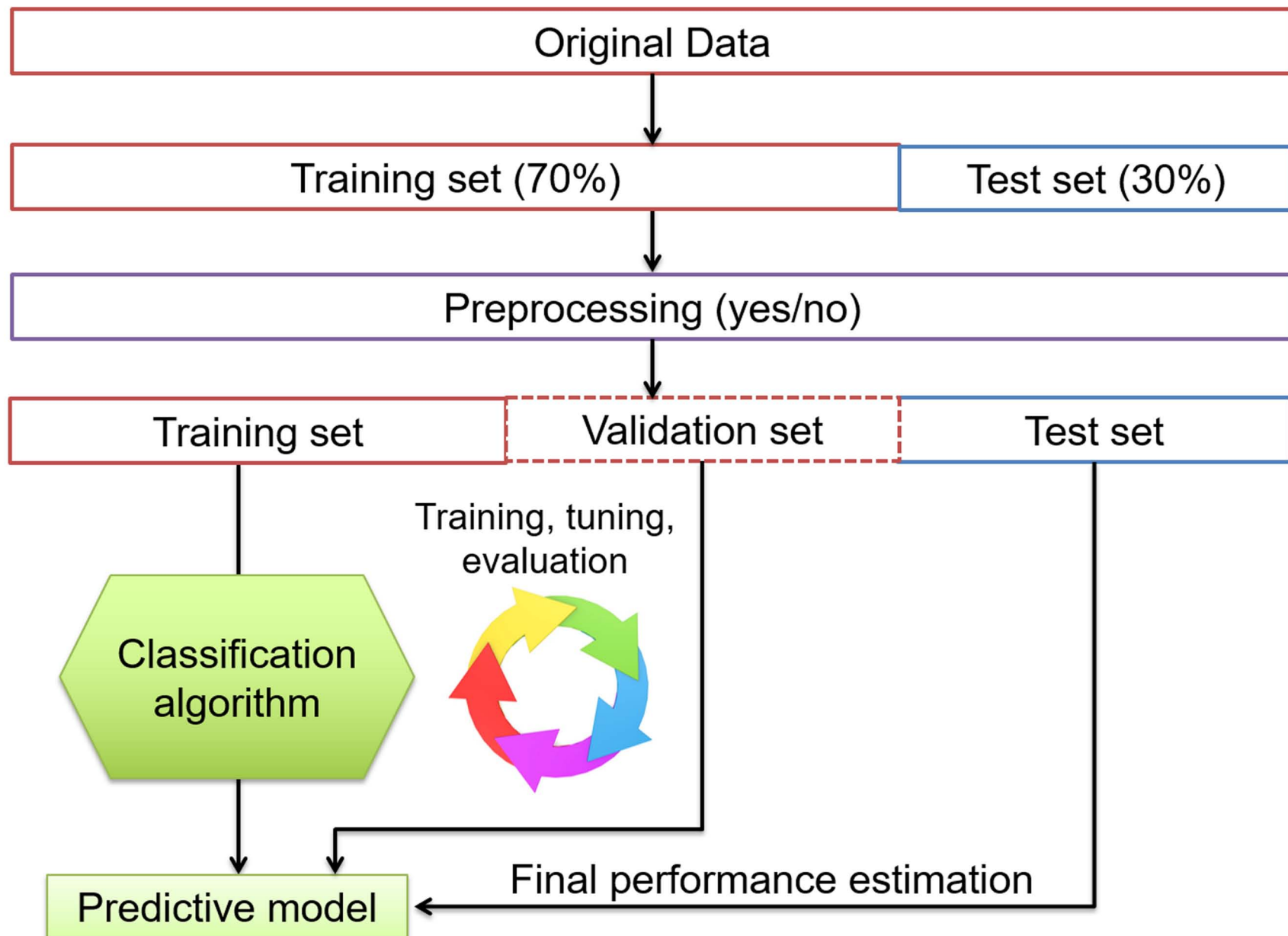
NIR spectra



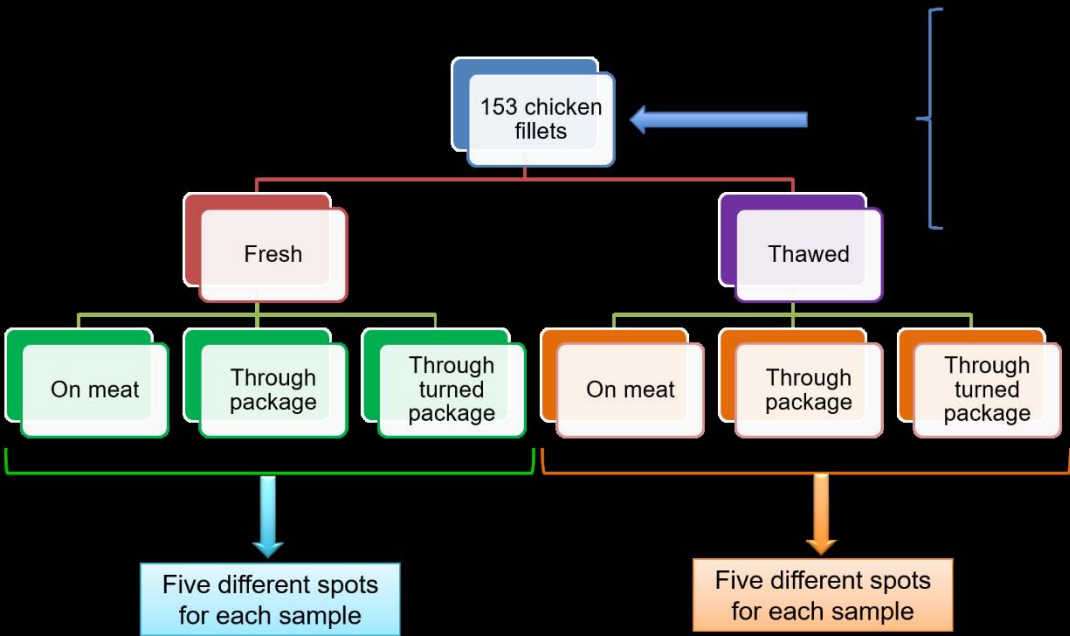












	Data	<i>TP</i>	<i>TB</i>	<i>OM</i>	<i>TP/TB</i>	<i>OM/TP/TB</i>
Fresh	Train	94.4	95.4	96.2	100.0	100.0
	<i>CV</i>	86.2	88.7	92.0	95.3	98.4
	<i>Test</i>	85.1	86.7	91.4	96.9	99.4
Thawed	<i>Train</i>	90.3	91.5	94.5	100.0	100.0
	<i>CV</i>	82.6	82.2	84.2	94.4	95.8
	<i>Test</i>	80.9	81.1	83.0	91.7	94.7

		<b>OM</b>			<b>TP</b>			<b>TB</b>		
		Train	CV	Test	Train	CV	Test	Train	CV	Test
	<i>Acc</i>	90.2	87.6	85.2	96.4	95.4	92.0	95.4	93.3	91.0
<b>Fr</b>	<i>Sen</i>	90.1	89.1	88.4	97.0	95.1	94.3	97.2	94.3	91.1
	<i>Pre</i>	90.3	89.4	87.1	98.3	96.3	94.0	96.3	93.1	90.0
<b>Th</b>	<i>Sen</i>	89.4	86.3	84.0	98.1	96.0	93.2	95.1	92.0	92.2
	<i>Pre</i>	90.0	87.2	85.2	96.2	94.2	95.1	96.0	93.2	91.3

		<b>OM</b>		<b>TP</b>		<b>TB</b>	
		Trai	CV	Train	CV	Train	CV
Explained Variance		97.1	97.1	100.0	100.0	99.8	99.8
Optimal number of LVs		6	6	5	5	6	6
Preprocessing used?		yes	yes	no	no	yes	yes
1*	<i>Sen</i>	84.4	84.1	49.4	49.1	82.4	82.3
	<i>Spe</i>	64.2	64.0	90.2	90.0	82.0	81.3
2*	<i>Sen</i>	81.8	81.6	52.5	51.0	80.9	80.2
	<i>Spe</i>	61.6	61.5	89.0	88.8	81.1	80.5
FR	<i>Sen</i>	88.9	88.1	92.0	91.0	82.9	81.7
	<i>Spe</i>	69.3	67.9	75.6	75.7	81.5	80.3
CONV	<i>Sen</i>	78.2	75.6	92.5	91.2	86.3	83.1
	<i>Spe</i>	57.1	59.4	31.3	31.1	52.6	53.1
STD	<i>Sen</i>	78.5	78.0	88.7	88.1	82.1	80.3
	<i>Spe</i>	64.6	65.0	56.7	55.9	50.6	52.1
ORG	<i>Sen</i>	81.5	83.0	94.4	93.0	84.3	80.6
	<i>Spe</i>	67.4	67.1	73.5	73.2	83.2	83.2
CF	<i>Sen</i>	83.2	78.1	89.8	89.2	79.2	77.7
	<i>Spe</i>	60.5	60.1	51.3	50.1	57.2	57.3
MAR	<i>Sen</i>	86.1	86.0	71.1	71.0	83.2	81.0
	<i>Spe</i>	92.2	92.2	88.3	88.2	93.3	93.2

OM samples		RSDE			Q-SVM		
		Train	CV	Test	Train	CV	Test
1*	<i>Acc</i>	85.1	78.5	79.4	77.7	68.7	71.1
	<i>Prec</i>	83.0	76.0	74.8	73.0	66.0	62.2
	<i>Sen</i>	85.0	80.0	79.5	90.0	84.0	80.3
2*	<i>Prec</i>	85.0	78.0	80.0	72.0	61.0	68.8
	<i>Sen</i>	84.0	77.0	76.2	69.0	55.0	62.9
FR	<i>Prec</i>	83.1	76.8	73.2	80.7	72.0	66.4
	<i>Sen</i>	79.2	76.0	69.1	73.8	64.0	63.2
CONV	<i>Prec</i>	80.0	70.0	82.8	88.0	68.0	42.9
	<i>Sen</i>	70.0	59.0	55.8	30.0	14.0	14.0
STD	<i>Prec</i>	90.2	89.0	88.5	89.0	84.2	85.2
	<i>Sen</i>	93.2	90.1	93.0	91.2	90.2	91.0
ORG	<i>Prec</i>	81.0	72.0	63.6	82.0	66.0	65.5
	<i>Sen</i>	77.0	66.0	67.7	72.0	59.0	58.1
CF	<i>Prec</i>	82.4	78.0	80.3	88.2	71.2	53.3
	<i>Sen</i>	77.1	62.2	59.8	35.1	24.3	22.5
MAR	<i>Prec</i>	93.0	90.0	90.5	88.0	82.0	88.4
	<i>Sen</i>	97.0	93.0	95.4	91.0	89.0	93.8

The Q-SVM had the closest overall test set *Acc* to RSDE. Still the RSDE significantly outperformed Q-SVM ( $z = 2.9389$ ,  $p = .00164$ ).

Data	Acc		
	Train	CV	Test
Raw data	90.4	82.4	81.9
Preprocessed data*	91.4	81.2	81.1
Preprocessed data without outliers**	91.7	80.8	80.9
Preprocessed data after outlier removal***	90.6	82.7	81.9

\*Mean centering and standard normal variate (SNV) as preprocessing

\*\*Q-residuals and Hotelling's  $T^2$  were used for outlier detection

\*\*\*Mean-centering and SNV on the data without outliers

Data	Method	Acc		
		Train	CV	Test
OM/TP/TB	RSDE	100.0	98.4	99.4
	SVM	92.7	81.5	82.0
	CP-ANN	90.2	68.3	65.4
	PLS-DA	78.8	66.5	63.5
TP/TB	RSDE	100.0	95.3	96.9
	SVM	88.8	82.2	81.3
	CP-ANN	87.2	66.2	63.6
	PLS-DA	76.3	64.3	62.4

Accuracy of RSDE for the test set was significantly higher than the accuracies of other methods, with all p-values < .00001 based on one-sided z-tests.

<b>RSDE model</b>	<b>Training</b>	<b>5-fold CV</b>	<b>Holdout CV(25%)</b>
	<i>Acc %</i>		
Data	100.0	99.0	100.0
Shuffled y-1	82.0	14.9	15.0
Shuffled y-2	81.0	19.8	19.0
Shuffled y-3	82.2	15.1	6.8
Shuffled y-4	80.7	19.5	17.0
Shuffled y-5	83.1	17.6	13.6



Assignment of spectra, no cut-off										
Left out	Acc	#Spectra	1*	2*	FR	ORG	STD	CONV	CF	MAR
1*	99.3	30	---		10.0	6.7		66.7	10.0	6.7
2*	99.6	85	38.8	---	3.5	31.8		24.7		1.2
FR	99.8	75			---		37.3		62.7	
ORG	98.8	30	16.7			---		83.3		
STD	99.2	90		1.1	54.4		---		28.9	15.6
CONV	99.3	30	63.3			36.7		---		
CF	99.4	75			74.7		16.0		---	9.3
MAR	99.0	175	3.4	2.3	26.9		41.7		25.7	---
Assignment of spectra (cut-off p > .50)										
Left out	Acc	#Spectra	1*	2*	FR	ORG	STD	CONV	CF	MAR
1*	99.3	19	---				5.3	78.9	5.3	10.5
2*	99.6	73	45.2	---		34.2		19.2		1.4
FR	99.8	75			---		37.3		62.7	
ORG	98.8	30	16.7			---		83.3		
STD	99.2	77		1.3	57.1		---		28.6	13.0
CONV	99.3	26	65.4			34.6		---		
CF	99.4	63			79.4		14.3		---	6.3
MAR	99.0	138	1.4		26.1		46.4		26.1	---
Assignment of spectra, (cut-off p > .75)										
		#Spectra	1*	2*	FR	ORG	STD	CONV	CF	MAR
1*	99.3	9	---					100.0		
2*	99.6	33	75.8	---		21.2		3.0		
FR	99.8	47			---		19.1		80.9	
ORG	98.8	19	5.3			---		94.7		
STD	99.2	35			74.3		---		11.4	14.3
CONV	99.3	19	73.7			26.3		---		
CF	99.4	33			90.9		6.1		---	3.0
MAR	99.0	71			19.7		59.2		21.1	---
Assignment of spectra, (cut-off p > .90)										
		#Spectra	1*	2*	FR	ORG	STD	CONV	CF	MAR
1*	99.3	3	---					100.0		
2*	99.6	24	91.7	---		8.3				
FR	99.8	26			---		15.4		84.6	
ORG	98.8	8	12.5			---		87.5		
STD	99.2	8			62.5		---		12.5	25.0
CONV	99.3	14	71.4			28.6		---		
CF	99.4	17			100.0				---	
MAR	99.0	31			16.1		67.7		16.1	---

## **Integration of handheld NIR and machine learning for the development of a “Measure & Monitor” technology for food authenticity**

### **Highlights**

- Handheld spectroscopy becomes even more powerful through Ensemble Learning.
- Authenticity of packaged chicken fillet could be monitored non-invasively.
- Single scans can provide more than 95% classification accuracy
- This work can be used towards consumer empowerment and forensic research.

**Declaration of interests**

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐ The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: