

Application of Hyperspectral Imaging System to Discriminate Different Diets of Live

Rainbow Trout (*Oncorhynchus mykiss*)

Mohammadmehdi Saberioon*,

University of South Bohemia in České Budějovice, Faculty of Fisheries and Protection of Waters,
South Bohemian Research Centre of Aquaculture and Biodiversity of Hydrocenoses, Institute of
Complex Systems, Zámek 136, Nové Hrady 37 333, Czech Republic

msaberioon@frov.jcu.cz

Petr Císař,

University of South Bohemia in České Budějovice, Faculty of Fisheries and Protection of Waters,
South Bohemian Research Centre of Aquaculture and Biodiversity of Hydrocenoses, Institute of
Complex Systems, Zámek 136, Nové Hrady 37 333, Czech Republic

cisar@frov.jcu.cz

Laurent Labbé,

INRA, UE 0937 PEIMA (Pisciculture Expérimentale INRA des Monts d'Arrée), Sizun, France

Laurent.labbe@inra.fr

Pavel Souček,

University of South Bohemia in České Budějovice, Faculty of Fisheries and Protection of Waters,
South Bohemian Research Centre of Aquaculture and Biodiversity of Hydrocenoses, Institute of
Complex Systems, Zámek 136, Nové Hrady 37 333, Czech Republic

psoucek@frov.jcu.cz

Pablo Pelissier

INRA, UE 0937 PEIMA (Pisciculture Expérimentale INRA des Monts d'Arrée), Sizun, France

pablo.pelissier@inra.fr

***corresponding author**

Application of hyperspectral imaging system to discriminate different diets of live Rainbow trout (*Oncorhynchus mykiss*)

Mohammadmehdi Saberioon^{1*}, Petr Císař¹, Laurent Labbé², Pavel Souček¹, Pablo Pelissier²

¹University of South Bohemia in České Budějovice, Faculty of Fisheries and Protection of Waters, South Bohemian Research Centre of Aquaculture and Biodiversity of Hydrocenoses, Institute of Complex Systems, Zámek 136, Nové Hradky 37 333, Czech Republic

²INRA, UE 0937 PEIMA (Pisciculture Expérimentale INRA des Monts d'Arrée), Sizun, France

*Corresponding author: msaberioon@frov.jcu.cz; saberioon@gmail.com.

Abstract

The main aim of this study was to evaluate the feasibility of hyperspectral imagery for determining the influence of different diets on fish skin. Rainbow trout (*Oncorhynchus mykiss*) were fed either a commercial based diet (N= 80) or a 100 % plant-based diet (N = 80). Hyperspectral images were made using a push-broom hyperspectral imaging system in the spectral region of 394-1009 nm. All images were calibrated using dark and white reference and the average spectral data from the region of interest were extracted. Six spectral pre-treatment methods were used, including Savitzky-Golay (SG), First Derivative(FD), Second Derivative (SD), Standard Normal Variate (SNV) and Multiplicative Scatter Correction (MSC) then a support vector machine (SVM) with linear kernel was applied to establish the classification models. Additionally, the Genetic algorithm (GA) was used to select optimal wavelengths to reduce the high dimensionality from hyperspectral images in order to decrease the computational costs and simplify the classification models. Overall classification models established from full wavelengths and selected wavelengths showed the good performance (Correct Classification Rate (CCR) = 0.871, Kappa = 0.741) when coupled with SG. The overall results indicate that the integration of Vis/NIR hyperspectral imaging system and machine learning algorithms has promise for discriminating different diets based on the live fish skin.

Introduction

The processes of light interaction with tissue has fundamental importance in biology. By studying the process involve in light remission from tissues, better protocols can be developed to automatically determine physiological and pathological status of animals. Light entering the skin undergoes multiple scattering and absorption events as it propagates across the tissue(Patterson, Wilson & Wyman, 1991). Tissue absorption is a function of molecular composition. Molecules absorb photons when the photons' energy matches interval energy states, and the transition between quantum states obeys the selection rules for the species. During the absorption process, transitions between two energy levels of a molecule that are well defined at a specific wavelength which could serve as a spectral fingerprint for diagnostic purposes(Zhang et al., 2013). The reflectance signal measured from the tissue is determined by the structural and biochemical properties of the tissue; therefore, change in optical properties can be used to noninvasively probe the living microenvironment. In other words, changes in tissue status will lead to corresponding changes in the pattern of reflected light.

Tissue components that absorb light are called chromophores. Some of the most important chromophores for visible wavelength is skin colour (Lu & Fei, 2014). Skin colour, as a complex trait that is determined by genetic, cellular, and physiological factors, has important role in displaying physiological, behavioural and ecological status of aquatic organisms. Knowledge about skin colouration pattern in fish is not only representative of growth rates(Colihueque et al., 2011) but also it exhibits the welfare of fish (Pavlidis et al., 2006). The colouration of fish is caused by the overlay and stacking of several types of pigmented cells including melanophores (black cell), xanthophores (yellow cells), erythrophores (red cells) and leucophores (silvery cells) in the dermis (Kelsh, 2004). Some researchers have shown that skin colour is highly dependent on the carotenoids present in the diet (Ho, O'Shea & Pomeroy, 2013), therefore, skin colour can provide

useful information for regulating accurate feeding. Also, some patterns show when fish are stressed of if they undergo various metabolic changes(Sefc, Brown & Clotfelter, 2014). Also, consumers use colour when selecting fish in the market. For instance, Colihueque et al. (Colihueque et al., 2011) showed the importance of Blue Black (BB) phenotype in adult rainbow trout (*Oncorhynchus mycosis*) as a positive trait in some markets; fish of this type had 23% higher growth rate than other skin colour phenotypes. Farmed large yellow croacker loses its natural colour under the intensive cultured condition. This loss is one of the most important reasons leading to its low market price and low consumer acceptance(Yi et al., 2014). Thus, fish farmers use augmented diet so as to develop optimal skin colour and enhance acceptance by consumers. Farmers avoid diets with no astaxanthin supplementation, in red porgy culture to prevent dark-gray skin colour that leads to the rejection by consumers (Kalinowski et al., 2007). Skin colour is crucial for quality assessment of ornamental fishes relative to market price but also breeding of high quality fish; repeated sorting and grading is done during grow-out according to skin colour quality(Zion et al., 2008). Therefore, analysis of skin colour, interpretation of the significance of colouration pattern and colour changes of aquatic animals is an important issue.

Diet, among other factors, has strong effects on stress tolerance and health, and therefore, fish must be fed adequate diet that meet all their nutrient requirements for a good growth and resistance to stress and disease problems(Trichet, 2010). In other words, feeding fish with an inadequate nutritional diet not only affects growth and feed efficiency but also increase susceptibility to disease and induces the appearance of deficiency signs, including altered behavioral and pathological changes(Oliva-Teles, 2012). Feed is the most expensive constituent of production costs in aquaculture(Hemre et al., 2003). Currently, some considerable effort is addressing the replacement fish meal and fish oil with plant based diet (PBD) (Boucher et al., 2011; Lazzarotto et al., 2015). However, few studies have investigated the effects of different diets on fish skin. Costa et al.(Costa

et al., 2013) showed image from a digital camera could be used to discriminate fish fed with organic and commercial diet. Segade et al. (Segade et al., 2015) studied the effect of different diets on seahorse body colour and their biochemical composition.

Researchers have used different optical sensors for measuring and determining light interaction with fish skin. Colourimeters have been used to determine skin colour changes. These instruments, which usually provide readings in XYZ, RGB and CIE Lab colour space, allow accurate and reproducible measurements of the colour with no influence by the observed or surroundings (Clydesdale & Ahmed, 1978). For instance, Kalinowski et al. (Kalinowski et al., 2007) used a tristimulus colourimeter to characterize skin colour parameters (CIE Lab) and intensity to determine the effect of esterified astaxanthin supplement in red colouration of red porgy. Yi et al. (Yi et al., 2014) used portable Minolta Chroma meter to show the effect of astaxanthin and xanthophylls as carotenoid sources on growth and skin colour of large yellow croaker. Pavlidis et al. (Pavlidis et al., 2006) used a portable spectrocolourimeter to measure the 3D characteristics and colour appearance (CIE L, a*, b*) of wild and farmed red skin Sparidae (*Pagrus pagrus*, *Pagrus caeruleostictus*, and *Dentex gibbosus*). They introduced a new index called Entire Colour Index (ECI) for exhibiting fish skin colour pattern. Some also used visible and Near infrared (Vis/NIR) spectroscopy to document the color changes on fish skin. Solberg et al. (Solberg et al., 2003) and Folkestad et al. (Folkestad et al., 2008) demonstrate feasibility of NIR spectroscopy for determining fat in live and slaughtered Atlantic salmon (*Salmo salar*). Lin et al. (Lin et al., 2003) showed satisfactory application of Vis/NIR spectroscopy to detect bruises in pacific pink salmon (*Oncorhynchus gorbuscha*) through skin. Costa et al. (Costa et al., 2011) used Vis/NIR spectroscopy of skin to differentiate the sea bass (*Dicentrarchus labrax*) with 87% accuracy at 48hr post-mortem quality cultured in tank from sea cage. Although fish skin colour described with proximal sensors is accurate, their use has been criticized due to small area measured by the

machine, and that aspects of the overall colours are lost (Mendoza & Aguilera, 2004). Also for accurate measurement, the surface colour should be quite uniform or homogenous and that may many locations on the sample must be measured to obtain the representative colour profile which sometimes are destructive (Yam & Papadakis, 2004).

Researchers also used consumer grade cameras as a non-invasive tool for measuring skin colour parameters. Digital images from consumer-grade cameras can overcome the deficiencies of visual and instrumental techniques and offer an objective measurement of colour and other physical factors (Chen, Chao & Kim, 2002). Wallat et al. (Wallat, Lazur & Chapman, 2005) demonstrated how a compact true colour camera could be employed for objective measurement of the skin colour of live goldfish (*Carassius auratus*) and how this information was used to optimize the feeding scheme to develop the most desirable skin colour. Zatkova et al. (Zat'ková et al., 2009) utilized a digital camera to estimate changes in skin colour of wels catfish (*Silurus glanis*). They showed the feasibility of digital cameras for monitoring skin colour changes due to diet alteration. Colihueque et al. (Colihueque et al., 2011) developed a method for estimating skin colour, spottiness and darkness using consumer digital camera and digital image analysis for categorizing cultured rainbow trout (*Oncorhynchus mykiss*). Costa et al. (Costa et al., 2013) used digital camera to analyse skin colour to discriminate the effects of seabass fed organic or commercial diet. Moreover, Wade et al. (Wade et al., 2014) used three different methods to measure prawn colour (two different colourimeter and colour quantification from digital images) to quantify whether any significant variation existed between the colours of farmed tiger prawns (*P. monodon*) from different ponds or from different farms. They successfully showed capability of both methods (colourimeters and digital images) to characterize the prawns from different farms. Segade et al. (Segade et al., 2015) also showed the effect of diet on seahorse (*Hippocampus hippocampus*) body colour using images obtained from consumer-grade digital cameras. Consumer-grade cameras

provide the capability to rapidly scan both larger areas as well as smaller details but can only study colour in Visible (Vis) bands. Furthermore, all above mentioned studies have not described interaction of light in Near infrared (NIR) bands to show the chromatic changes on fish skin.

Hyperspectral imagery (HSI) is an emerging technology that integrates both spectroscopy and imaging in a single system; it has potential to capture the subtle spectral difference under different physiological and pathological conditions. HSI is enabling simultaneous acquisition of spatial and spectral information from an object. The system has the ability to image the same scene in hundreds of contiguous narrow wavebands, from the visible to the short-wave infrared region of electromagnetic spectrum (400-2500 nm). In other words, HSI has higher spatial resolution than the multispectral image which are obtained by consumer grade digital cameras. The output of the hyperspectral imaging system is three-dimensional (3D) (spatial-spectral-spatial) and is referred to as spectral cube, data cube, spectral volume or simply hypercube. Hypercube consists of a large amount of information and can be used to analyse minor and/or subtle physical and chemical features in samples (ElMasry & Nakauchi, 2016). Hyperspectral Imaging (HSI) has been used to quantify many fish and fish products features such as fat distribution (ElMasry & Sun, 2010), texture (Wu, Sun & He, 2014), pH (He, Wu & Sun, 2012), tenderness (He, Wu & Sun, 2014) freshness (Sone et al., 2012) and moisture (He, Wu & Sun, 2013) with high significant prediction ability ($r > 0.8$). Fat and water content in fillets of multiple fish species, including Atlantic halibut, catfish, cod, saithe, macherel and herring were also measured using HSI (ElMasry & Wold, 2008). Potential of HSI to estimate lactic acid bacteria and microbial spoilage of fish products also has been assessed (He, Sun & Wu, 2014; Cheng & Sun, 2015). However, the feasibility of HSI to determine the impact of diet changes in live aquatic organisms has not been studied. Our hypothesis is that with a combination of spectra and image can provide more objective, reproducible and potentially automated method for determining diet-affected fish skin biochemical composition

compared to other methods. Therefore, the main aim of this study was to evaluate the feasibility of hyperspectral imagery to evaluate the influence of two different diets on fish skin. Another objective was to develop a new, robust, accurate and non-invasive protocol using hyperspectral imagery system and machine learning algorithms in analysing subtle spectral differences of diet-affected characteristics.

Materials and Method

Fish and cultural condition

The experimental groups were produced at INRA-PEIMA (Sizun, France). The fish population, contained 250 fish which were grown in six 5.4 m³ replicated tanks using in river water (13.4 – 18.3 °C) until data acquisition. All 250 fish were tagged with passive integrated transponder (AEG-Id, ISO FDXB) for individual identification. Experiments were designed in a split-block design with three replications for each diet; therefore, 125 fish were fed a commercial based diet (3 tanks) and 125 were fed a plant-based diet (3 tanks). After three weeks, 80 fish from each treatment were selected randomly for hyperspectral image acquisition. Mean body weight of fish receiving the commercial based diet (CBD) was 228.99g and the plant based diet (PBD) was 222.46g at time of data acquisition. All methods and protocols were carried out in accordance with the French law for animal experiment with authorisation of the French veterinary service under ethical approval No. B29-777-02.

Diets and feeding controls

Diets were manufactured at the INRA NUMEA facility of Donzacq (France). The ingredient and analysis composition is given in Table 1. CBD contained fishmeal and fish oil as protein and lipid source respectively. PBD is contained mixture of wheat gluten, extruded peas, corn gluten meal, soybean meal and white lupin as protein sources; and combination of palm seed, rapeseed and linseed oil, rich in saturated, mono-unsaturated and n-3 poly-unsaturated fatty acids, respectively,

as lipid source. A mineral and a vitamin premix equally were added into both diets. Both diets fulfilled the known nutrient requirement of rainbow trout as explained by Geurden et al.(Geurden et al., 2013).

Table 1: The Ingredient of commercial based diet (CBD) and plant based diet (PBD)

| Ingredients | CBD | PBD |
|-------------------------------------------|------|------|
| Fish Oil | 11.8 | - |
| Plant oil blend* | - | 11.4 |
| Fish meal | 42.4 | - |
| Soybean Meal | 12.0 | 12.0 |
| Pea | 17.1 | 12.5 |
| Wheat | 9.6 | 4.0 |
| Lupin flour | - | 5.0 |
| Wheat gluten | - | 17.0 |
| Corn gluten | - | 17.0 |
| Fababean protein concentration | - | 10.0 |
| Dicalcium Phosphate | - | 3.0 |
| Soy lecithin powder | - | 2.0 |
| Additive (vitamin, mineral, preservative) | 4.5 | 4.5 |

*Palm seed, rapeseed and liveseed oil

Image acquisition

A push-broom, line-scanning reflectance hyperspectral imaging system was used to acquire the hyperspectral images of rainbow trout in a dark room to avoid the interference due to stray light and to get pure spectral reflectance. This system includes a high-performance CCD camera (Photonfocus 1312 CL) along with focus lens, a spectrograph (Specim V10E, Spectral Imaging Ltd., Oulu, Finland) attached to the camera acquire hyperspectral images in the wavelength range of 393-1009 nm, an illumination source (150 W halogen lamp attached to a fibre optic line light positioned at an angle of 45 degree to the moving table to reduce the shadowing effects), a moving table and a computer system equipped with an image acquisition software (SpectralScanner, DV optics).

To acquire the spectral and spatial information, each rainbow trout was placed on the sample loading device and they conveyed to the camera field of view (FoV) of camera with adjusted speed (1.6 mm/s) and exposure time (10 s) to be scanned line by line. The procedure was controlled and implemented by image acquisition software (Specim Lumo software, Spectral Imaging Ltd., Oulu, Finland). The raw hyperspectral image for each sample consisted of several congruent images representing intensities at 784 wavelength bands from 393 to 1009 nm. Due to the low signal-to-noise rate at the two ends of the spectral ranges, only the wavelength ranging from 400 to 1000 nm wavelength were used. Totally, 160 hyperspectral images were created, recorded, stored in raw format before being processed. Before measurement, each fish mildly anesthetized with Benzocaine to reduce the movement and minimize the stress. The surface of each rainbow trout also was wiped with piece of tissue paper to remove extra water on skin before data acquisition.

Hyperspectral image calibration

In order to eliminate the influences from the bright and dark response, the acquired hyperspectral images were corrected based on the methods described by Wu et al. (Wu, Sun & He, 2012). Dark reference images (~ 0 % reflectance) were obtained by closing the shutter and the white reference images (~99.9% reflectance) were acquired by collecting a spectral image from a uniform white calibration tile, afterward, reflectance hyperspectral image was calculated using Equation (1).

$$I_i = \frac{R_i - D_i}{W_i - D_i} \quad (1)$$

where I is the corrected hyperspectral image, R is the raw hyperspectral image, W is the white reference image and D is the dark reference image as well as i is the pixel index, i.e. $i = 1, 2, 3, \dots, n$ and n is the total number of pixels, which was then used as a basis for subsequent hyperspectral image extraction and spectral data analysis.

Region of Interest (ROI) corresponding to image of sample identified and the spectra of ROI were extracted and exported as a matrix (X), where the rows of this matrix represent the number of samples and the columns represent the number of variables (764 wavelengths). The reflectance spectrum from the ROI was computed by averaging the spectral value of all pixels in the ROI for each sample using the Environment for Visualizing Images software (ENVI 5.3, Harris geospatial solutions, FL, USA).

Spectral data analysis

Spectral pre-processing

Spectral pre-processing which used for correcting light scattering in reflectance measurement and data enhancement by reducing/removing physical phenomena before data is used in classifier. The most widely used spectral pre-processing methods can be commonly divided into three groups, namely smoothing, baseline removal, and scaling (Gholizadeh et al., 2015). The first category is smoothing such as Savitzky-Golay (SG) which is used for noise reduction; the second category is baseline removal such as the first derivative (FD) and second derivative (SD) which is used for correcting background signals or baseline that is far away from zero level. Multiplicative scatter correction (MSC) is also another popular transformation method used to remove the scatter effects on spectral data. Another group is range scaling, this method is applicable when the total intensity in the spectra is sample-dependent, and samples need to be scaled in such a way that intensities can be compared. Standard normal variate (SNV) is one of the popular pre-processing method which centring and scaling each individual spectrum for correcting the multiplicative interferences of light scatter (Duckworth). Six forms of spectra pre-processing were used in this study to remove the non-constituent-related effects in spectra data and to develop optimal classifier model. The six forms were SG smoothing with second order polynomial fit and 11 smoothing points, the FD and SD transformation, MSC and SNV.

Classifier

After pre-treatment, Support vector machine (SVM) as a classifier was employed to develop the classification models for discriminating two different diets. SVM uses kernel function to map input x into a high-dimensional feature space $\phi(x)$ by constructs an optimal hyperplane to separate the two kinds of data points from two classes(Duan, Keerthi & Poo, 2003) . SVM employs the lagrange multiplier to compute the partial differentiation of each feature to acquire the optimal solution. Consider a given training set of N data points, $\{x_k, y_k\}_{k=1}^N$ with input data, which is an n -dimensional data vector ($x_k \in R^N$) , and output, which is the one-dimensional vector space($y_k \in r$); SVM create the classifier as shown in Equation (2).

$$y(x) = \text{sign}[\sum_{k=1}^N \alpha_k y_k \psi(x, x_k) + b] \quad (2)$$

where α_k are positive real constants and b is a real constant. For this study, SVM with linear kernel which can be showed in Equation (3) was used.

$$\psi(x, x_k) = x_k^T x \quad (3)$$

The classifier is constructed as Equation (4).

$$y_k [w^T \phi(x_k) + b] \geq 1, k = 1, \dots, N \quad (4)$$

where $\phi(\cdot)$ is a nonlinear function which maps the input space into a higher dimensional space. Further details can be found in Vapnik(Vapnik & Vapnik, 1998). R package Caret (Kuhn, 2008)used for SVM classification model.

Selection of feature wavelengths

Some studies used the most relevant and informative wavelengths instead of full wavelengths to optimize the data analysis and reduce the cost of computation(Xu, Riccioli & Sun, 2017). Therefore, the genetic algorithm (GA) was used to identify and select the most important and informative wavelengths in addition to full range of wavelengths. GA, as one of the popular feature

selection method, is a random search method which utilizes non-local and probabilistic process based on the principal of natural genetic selection systems(Leardi, Seasholtz & Pell, 2002). In other words, GA aimed to reach the global optimum for a problem by shifting the best individuals in the population using mutation and cross-over operations. Basically, GA's process can be summarized into five steps; 1/ coding all variables 2/ initiation of population 3/ evaluation of the responses 4/ reproductions 5/ mutations(Leardi & Gonzalez, 1998). More details of GA can be found in Murthy and Chowdhury(Murthy & Chowdhury, 1996). The parameters adopted in this study were mainly related to population size of each generation (20), number of iteration (100), deletion group (5), crossover probability (0.8) and mutation probabilities (0.1). R package Caret(Kuhn, 2008) used for GA implementation. All the parameters and conditions were taken from previous studies(Li et al., 2011; Feng & Sun, 2013).

Evaluation of the classification models

Validation is an important component to test the learning status of the model. The dataset from 160 hyperspectral for rainbow trout was divided into training set (80% of total samples) used to develop the classifier models and a validation set (20% of total samples) used to assess the prediction accuracy of each model. The training set was used for fitting models, and the validation set was performed by random stratified sampling. Afterward, classifier was evaluated through the analysis of correct classification rate (CCR, %) and Cohen's Kappa coefficient in the validation set. CCR and Cohen's Kappa coefficient was calculated by the equation 5 and 6 respectively.

$$CCR = \frac{N_1}{N_0} \times 100 \% \quad (5)$$

where N_1 is number of corrected estimation of samples and N_0 is the total number of samples

$$K = \frac{\text{Pr}(a) - \text{Pr}(e)}{1 - \text{Pr}(e)} \quad (6)$$

where $\text{Pr}(a)$ is observed agreement and $\text{Pr}(e)$ is probability of random agreement.

Furthermore, sensitivity and specificity which can be obtained using Equation (7) and (8) respectively, were used to evaluate the classification model as well (Amigo, Babamoradi & Elcoroaristizabal, 2015).

$$Sensitivity = \frac{TP}{TP+FN} \quad (7)$$

$$Specificity = \frac{TN}{FP+TN} \quad (8)$$

where TP and TN are true positive and true negative respectively; and FN and FP are false negative and false positive respectively.

Additionally, the area under the Receiver Operator Characteristics (ROC) curve (AUC), known as a global measures of classifier performance, were calculated for comparing overall performance of all different classification schemes (Bradley, 1997). R package pROC (Robin et al., 2011) used in this study to create ROC curves. Figure 1 shows the schematic of methodology used in this study.

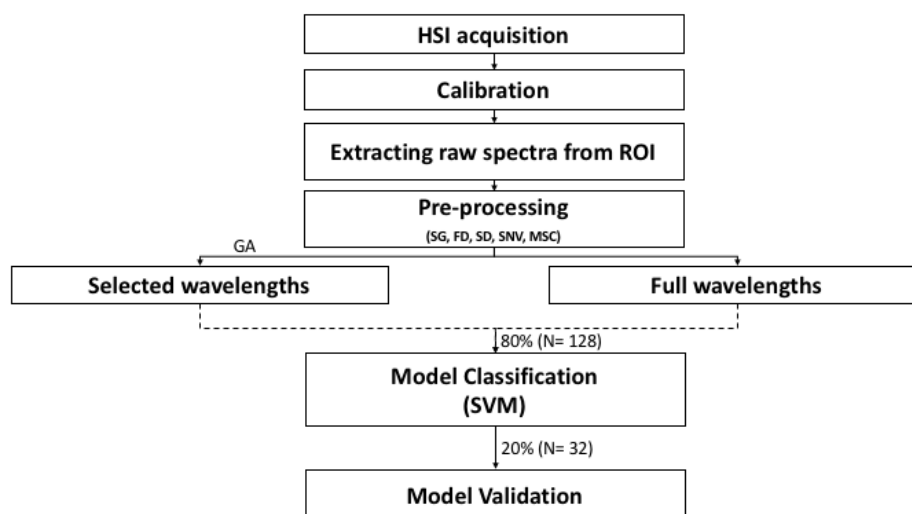


Figure 1: Schematic diagram of hyperspectral data analysis in this study

Results and Discussion

The Vis/NIR hyperspectral imaging for the two diets were similar but had differences in reflectance intensity in the range from 450- 750 nm and 900- 1000 nm (Figure 2). It can be seen that the mean reflectance values of PBD are higher than CBD in most part of visible bands (450 – 750 nm), but the mean reflectance values of CBD are higher than PBD in NIR bands (900 -1000 nm). This implied that different diets have induced significant alterations to fish skin in a way that can be detected by spectral information. The differences between spectra might be mainly associated with how the different types of lipids, influence the absorption and deposition of carotenoids. In other words, oils contain different amounts of stanol and sterols which interfere with the uptake of carotenoids (Nguyen, 1999; Kalinowski et al., 2005; Choubert, Mendes-Pinto & Morais, 2006).

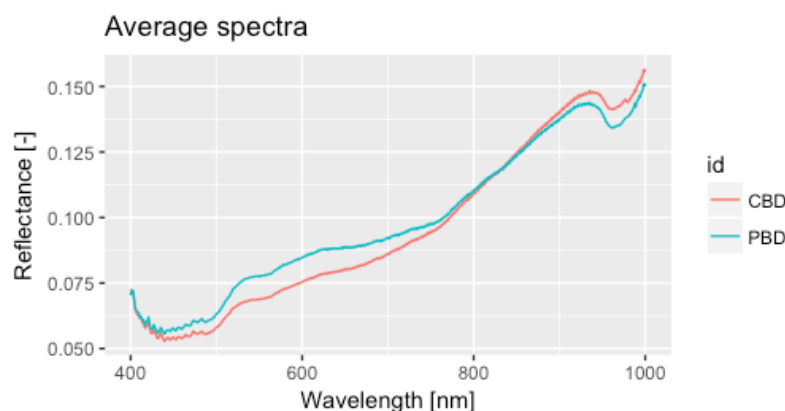


Figure 2: The mean Vis/NIR spectral reflectance of rainbow trout skin for commercial based diet (CBD) and plant based diet (PBD)

In order to establish a robust classification model and explore the influence of spectral sampling interval on the classification accuracy, different pre-processing algorithms were performed. Figure 3 shows raw and pre-processed spectra for all samples.

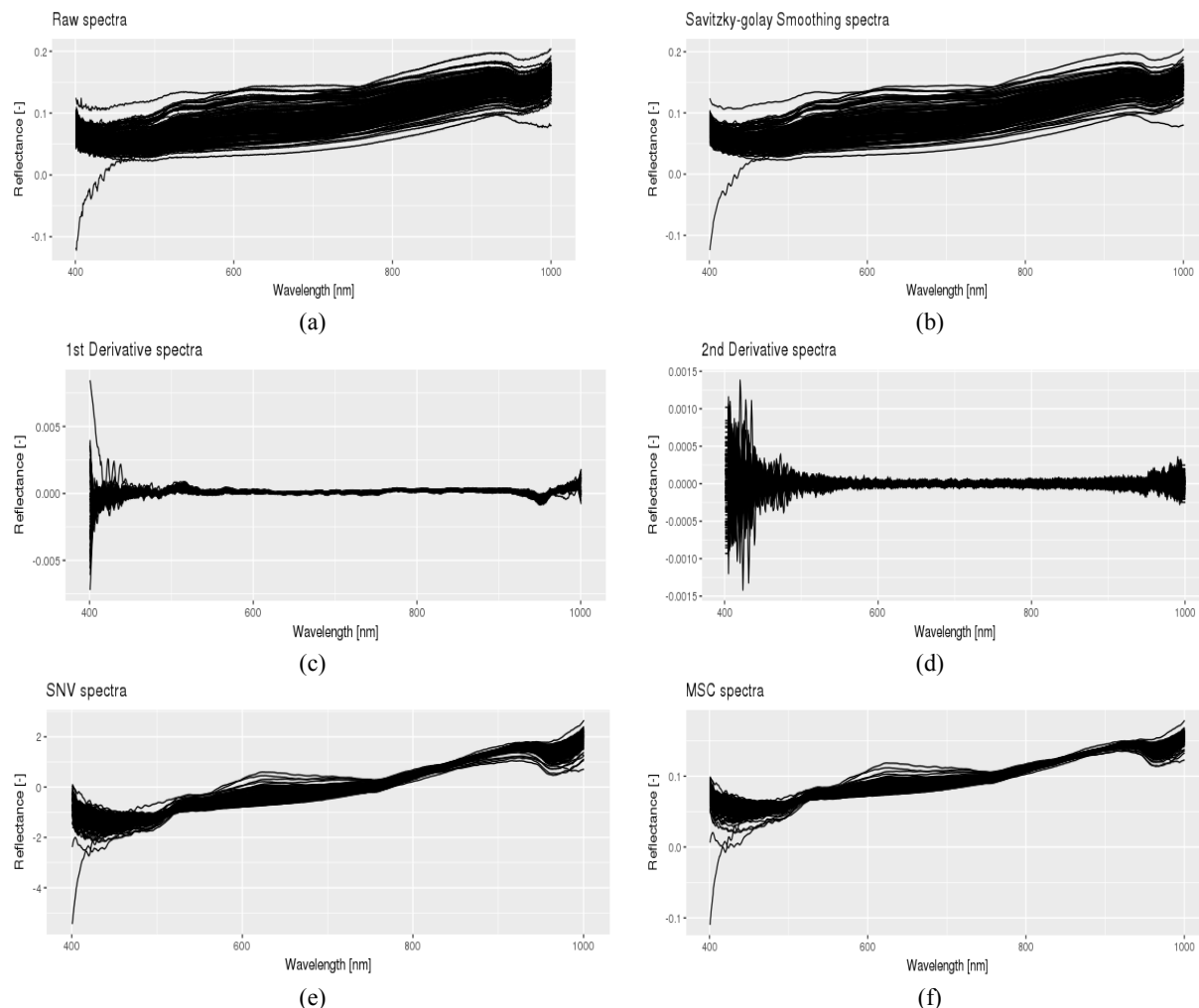


Figure 3: representative Vis/NIR Spectra of fish skin raw and different pre-processing methods; SG: Savitzky-golay (b), FD: 1st Derivative (c), SD: 2nd Derivative (d), SNV: Standard normal variate (e), and MSC: multiplicative scatter correction (f).

Classification model using full wavelengths

The performances of classification models were evaluated by the CCR, Kappa coefficient, Sensitivity and Specificity. Table 2 shows the average CCRs for testing set with six spectral pre-processing techniques based on the whole spectral range. When the raw and pre-processed spectra were used to build the classification models, CCRs ranged between 0.740 and 0.871. After application of SD, classification accuracy decreased compare to classification using spectra without pre-processing, however, application of SG, FD, SNV and MSC improved the classification compared to raw spectra. Thus, spectral pre-processing of SD was not helpful and to some extend

reduce the accuracy of classification, but other pre-processing treatments were improved the performance of classifier for full range of wavelengths. The best performance for classification with the highest CCR of 0.871 and Kappa coefficient of 0.740 achieved when smoothing solely used as pre-processing method for full wavelength.

Classification model using selected wavelengths

In this study, GA was used to select the most effective wavelengths after different pre-processing methods. The CCR and Kappa coefficient for validation set was 0.709 and 0.417 when no spectral pre-treatment used for selected wavelengths (Table 2). After using the spectral pre-processing algorithms, CCR range was between 0.709 and 0.838, thus FD, SD and MSC were not improved the classification performance, but SG and SNV improved the accuracy of classification for validation set in compare to spectra without pre-treatment. The highest CCR and Kappa coefficient for validation set was 0.838 and 0.679 respectively, achieved when selected wavelengths classified using SVM coupled with Standard normal variate.

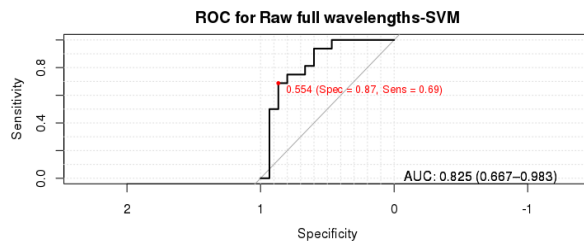
Classification models using the selected wavelengths showed decline in CCRs to correspondence full range of wavelengths except pre-processed spectra using SNV which improved the accuracy. In other words, GA is suitable for selecting the most informative and important wavelengths for differentiate the diet based on fish skin, however, it does not improve the accuracy of classification. The ROC curves also showed on Figure 4. The red dot on the figures denotes the best threshold obtained which is the closest point to the top corner where the true positive rate equals one, and the false positive rate of zero (Ariana, Guyer & Shrestha, 2006). Furthermore, AUC as an index for showing the quality of classifiers (Liu, Sun & Zeng, 2014) mentioned. AUC of one is considered as a perfect classifier, while 0.5 would be a random classifier. Based on the AUC comparison, the classification performance for full range of wavelengths was slightly better than the classification

performance for selected wavelengths. AUC based on the full wavelengths coupled with SNV as pre-treatment is the highest (0.933).

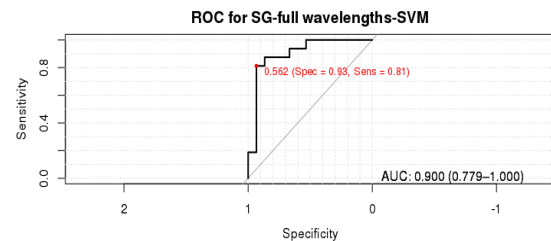
Table 2: Model performance for identification of different diet on validation set

| method | Number* | Full wavelengths | | | | Number* | Selected wavelengths | | | |
|---------|---------|------------------|--------------|--------------|--------------|---------|----------------------|--------------|--------------|--------------|
| | | CCR | Sensitivity | Specificity | Kappa | | CCR | Sensitivity | Specificity | Kappa |
| Raw-SVM | 764 | 0.741 | 0.687 | 0.800 | 0.485 | 340 | 0.709 | 0.666 | 0.750 | 0.417 |
| SG-SVM | 764 | 0.871 | 0.875 | 0.866 | 0.741 | 350 | 0.830 | 0.866 | 0.812 | 0.670 |
| FD-SVM | 764 | 0.806 | 0.812 | 0.800 | 0.612 | 242 | 0.709 | 0.666 | 0.750 | 0.417 |
| SD-SVM | 764 | 0.730 | 0.750 | 0.733 | 0.483 | 363 | 0.709 | 0.733 | 0.687 | 0.420 |
| SNV-SVM | 764 | 0.774 | 0.750 | 0.800 | 0.548 | 318 | 0.838 | 0.933 | 0.750 | 0.679 |
| MSC-SVM | 764 | 0.838 | 0.875 | 0.800 | 0.676 | 218 | 0.677 | 0.800 | 0.562 | 0.359 |

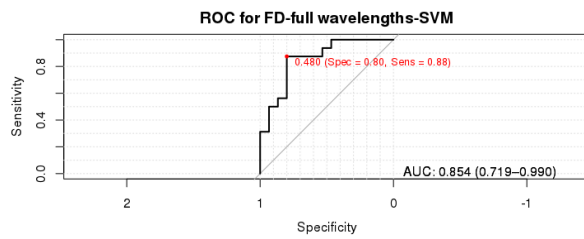
*Number of wavelengths



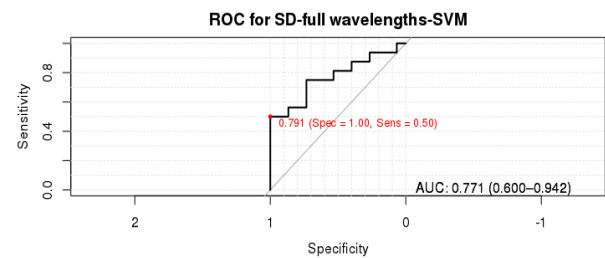
(a)



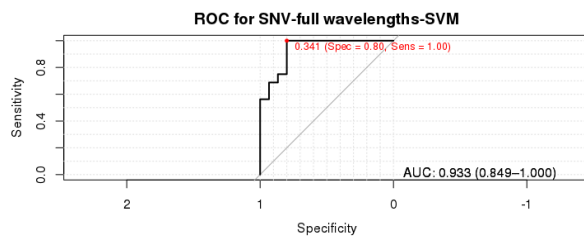
(b)



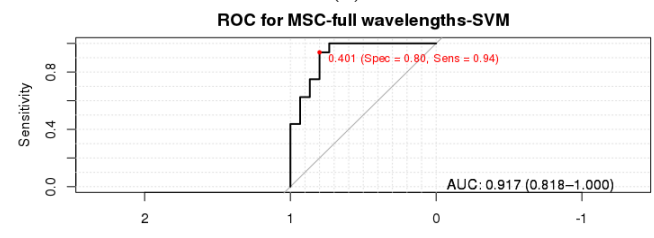
(c)



(d)



(e)



(f)

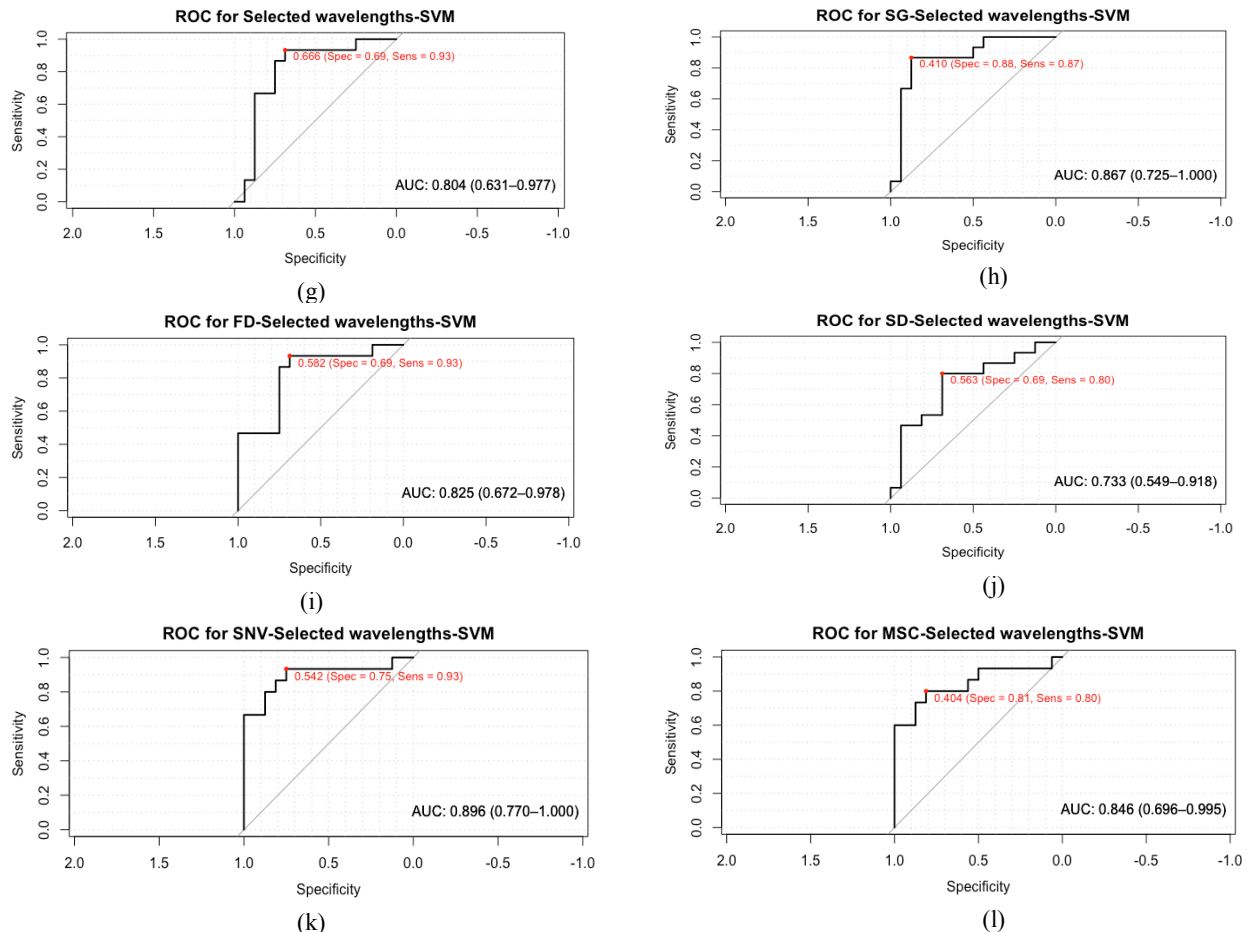


Figure 4: ROC curves obtained for the test set of all developed models

This study developed a rapid, non-invasive and automate system to discriminate a dietary effect on live rainbow trout non-invasively and rapidly. Overall, classification models established from full wavelengths and selected wavelengths showed the good performance when SG used as pre-treatment of spectra. Although selecting feature wavelength did not improve the accuracy of classification, but it did provide the same accuracy as full wavelengths. The selected optimal wavelengths would be useful to develop a multispectral imaging system, simplifying the data handle in the high dimension of HIS and more suitable for industrial on-line application. In a brief conclusion, results of this study will contribute a novel and accurate method to compare with current available approaches for categorizing the external appearance of live rainbow trout, particularly with the goal of studying impact of different diets on fish external appearance. In other

words, the skin pigmentation in farmed rainbow trout has commercial value because high visual impact on the consumers and affects marketing prospects(Kause et al., 2003), therefore, the introduced system in this study can provide more objective, accurate, fast tools for determining impact of different diets on external appearance of rainbow trout. Furthermore, it can be used in implementation of European union regulation (EC 178/2002) on traceability of food and feed, food and feed producing animals, and substances intended to be incorporated into food and feed by providing diet information based on fish skin. Further studies should be carried out to not only improve the classification accuracy using different machine learning algorithms but also increasing classification power for online evaluation of fish skin at industrial scale. Furthermore, impact of different nutrient deficiency can be investigated using HIS in future studies.

Acknowledgment

This work was funded by projects CENAKVA [CZ.1.05/2.1.00/01.0024] and CENAKVA II (the results of the project LO1205 were obtained with a financial support from the MEYS of the CR under the NPU I program); The CENAKVA Centre Development [No. CZ.1.05/2.1.00/19.0380]; and the European Union's Horizon 2020 research and innovation program under grant agreement No. 652831" (Aquaexcel2020).

References:

- Amigo J., Babamoradi H., Elcoroaristizabal S. 2015. Hyperspectral image analysis. A tutorial. *Analytica Chimica Acta* 896:34–51. DOI: 10.1016/j.aca.2015.09.030.
- Ariana D., Guyer D., Shrestha B. 2006. Integrating multispectral reflectance and fluorescence imaging for defect detection on apples. *Computers and Electronics in Agriculture* 50:148. DOI: 10.1016/j.compag.2005.10.002.
- Boucher R., Quillet E., Vandeputte M., Lecalvez J., Goardon L., Chatain B., Médale F., Dupont-Nivet M. 2011. Plant-based diet in rainbow trout (*Oncorhynchus mykiss* Walbaum): Are there genotype-diet interactions for main production traits when fish are fed marine vs. plant-based diets

- from the first meal? *Aquaculture* 321:41–48. DOI: 10.1016/j.aquaculture.2011.08.010.
- Bradley A. 1997. The use of the area under the ROC curve in the evaluation of machine learning algorithms. 30:1145–1159. DOI: 10.1016/S0031-3203(96)00142-2.
- Chen Y-R., Chao K., Kim M. 2002. Machine vision technology for agricultural applications. *Computers and Electronics in Agriculture* 36:173. DOI: 10.1016/S0168-1699(02)00100-X.
- Cheng J-H., Sun D-W. 2015. Recent Applications of Spectroscopic and Hyperspectral Imaging Techniques with Chemometric Analysis for Rapid Inspection of Microbial Spoilage in Muscle Foods. *Comprehensive Reviews in Food Science and Food Safety* 14:478. DOI: 10.1111/1541-4337.12141.
- Choubert., Mendes-Pinto., Morais. 2006. Pigmenting efficacy of astaxanthin fed to rainbow trout *Oncorhynchus mykiss*: Effect of dietary astaxanthin and lipid sources. *Aquaculture* 257:429–436. DOI: 10.1016/j.aquaculture.2006.02.055.
- Clydesdale F., Ahmed E. 1978. Colorimetry — methodology and applications* . *C R C Critical Reviews in Food Science and Nutrition* 10:243–301. DOI: 10.1080/10408397809527252.
- Colihueque N., Parraguez M., Estay F., Diaz N. 2011. Skin Color Characterization in Rainbow Trout by Use of Computer-Based Image Analysis. *North American Journal of Aquaculture* 73:249. DOI: 10.1080/15222055.2011.581578.
- Costa C., D’Andrea S., Russo R., Antonucci F., Pallottino F., Menesatti P. 2011. Application of non-invasive techniques to differentiate sea bass (*Dicentrarchus labrax*, L. 1758) quality cultured under different conditions. *Aquaculture International* 19:765. DOI: 10.1007/s10499-010-9393-9.
- Costa C., Menesatti P., Rambaldi E., Argenti L., Bianchini M. 2013. Preliminary evidences of colour differences in European sea bass reared under organic protocols. *Aquacultural Engineering* 57:82. DOI: 10.1016/j.aquaeng.2013.08.001.
- Duan K., Keerthi S., Poo A. 2003. Evaluation of simple performance measures for tuning SVM hyperparameters. *Neurocomputing* 51:41–59. DOI: 10.1016/S0925-2312(02)00601-X.
- Duckworth J. Mathematical Data Preprocessing. In: *Mathematical Data Preprocessing. Near-Infrared Spectroscopy in Agriculture*. American Society of Agronomy, Crop Science Society of America, Soil Science Society of America, American Society of Agronomy, Crop Science Society of America, Soil Science Society of America, 115.
- ElMasry G., Nakauchi S. 2016. Image analysis operations applied to hyperspectral images for non-invasive sensing of food quality – A comprehensive review. *Biosystems Engineering* 142:53. DOI: 10.1016/j.biosystemseng.2015.11.009.
- ElMasry G., Sun D-W. 2010. Principles of Hyperspectral Imaging Technology. In: *Principles of Hyperspectral Imaging Technology*. Elsevier Inc., Elsevier Inc., 3. DOI: 10.1016/B978-0-12-374753-2.10001-2.

- ElMasry G., Wold J. 2008. High-Speed Assessment of Fat and Water Content Distribution in Fish Fillets Using Online Imaging Spectroscopy. *Journal of Agricultural and Food Chemistry* 56:7672. DOI: 10.1021/jf801074s.
- 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. DOI: 10.1016/j.talanta.2013.01.057.
- Folkestad A., Wold J., Rørvik K-A., Tschudi J., Haugholt K., Kolstad K., Mørkøre T. 2008. Rapid and non-invasive measurements of fat and pigment concentrations in live and slaughtered Atlantic salmon (*Salmo salar* L.). *Aquaculture* 280:129–135. DOI: 10.1016/j.aquaculture.2008.04.037.
- Geurden., Borchert., Balasubramanian., Schrama., Dupont-Nivet., Quillet., Kaushik., Panserat., Médale. 2013. The Positive Impact of the Early-Feeding of a Plant-Based Diet on Its Future Acceptance and Utilisation in Rainbow Trout. *PLoS ONE* 8:e83162. DOI: 10.1371/journal.pone.0083162.
- Gholizadeh A., Borůvka L., Saberioon M., Kozák J., Vašát R., Němeček K. 2015. Comparing different data preprocessing methods for monitoring soil heavy metals based on soil spectral features. *Soil and Water Research* 10:218. DOI: 10.17221/113/2015-SWR.
- He H-J., Sun D-W., Wu. 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. DOI: 10.1016/j.foodres.2014.03.064.
- He H., Wu D., Sun D-W. 2012. Application of hyperspectral imaging technique for non-destructive pH prediction in salmon fillets. 17:5.
- He H-J., Wu D., Sun D-W. 2013. Non-destructive and rapid analysis of moisture distribution in farmed Atlantic salmon (*Salmo salar*) fillets using visible and near-infrared hyperspectral imaging. *Innovative Food Science & Emerging Technologies* 18:237. DOI: 10.1016/j.ifset.2013.02.009.
- He H-J., Wu D., Sun D-W. 2014. 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. DOI: 10.1016/j.jfoodeng.2013.11.015.
- Hemre., Ø K., Mangor-Jensen., Rosenlund. 2003. Digestibility of dry matter, protein, starch and lipid by cod, *Gadus morhua*: comparison of sampling methods. *Aquaculture* 225:225–232. DOI: 10.1016/s0044-8486(03)00291-6.
- Ho., O'Shea., Pomeroy. 2013. Dietary esterified astaxanthin effects on color, carotenoid concentrations, and compositions of clown anemonefish, *Amphiprion ocellaris*, skin. *Aquaculture International* 21:361–374. DOI: 10.1007/s10499-012-9558-9.
- Kalinowski C., Izquierdo M., Schuchardt D., Robaina L. 2007. Dietary supplementation time with shrimp shell meal on red porgy (*Pagrus pagrus*) skin colour and carotenoid concentration.

- 529 Aquaculture 272:451. DOI: 10.1016/j.aquaculture.2007.06.008.
- 530
- 531 Kalinowski C., Robaina L., Fernandez-Palacios H., Schuchardt D., Izquierdo MS. 2005. Effect of
- 532 different carotenoid sources and their dietary levels on red porgy (*Pagrus pagrus*) growth and skin
- 533 colour. Aquaculture 244:223–231. DOI: 10.1016/j.aquaculture.2004.11.001.
- 534
- 535 Kause A., Ritola O., Paananen T., Eskelinen U., Mantysaari E. 2003. Big and beautiful?
- 536 Quantitative genetic parameters for appearance of large rainbow trout. Journal of Fish Biology
- 537 62:610–622. DOI: 10.1046/j.0022-1112.2003.00051.x.
- 538
- 539 Kelsh R. 2004. Genetics and evolution of pigment patterns in fish. Pigment Cell Research 17:326–
- 540 336. DOI: 10.1111/j.1600-0749.2004.00174.x.
- 541
- 542 Kuhn M. 2008. Building Predictive Models in R Using the caret Package. Journal of Statistical
- 543 Software 28. DOI: 10.18637/jss.v028.i05.
- 544
- 545 Lazzarotto V., Corraze G., Leprevost A., Quillet E., Dupont-Nivet M., Médale F. 2015. Three-
- 546 Year Breeding Cycle of Rainbow Trout (*Oncorhynchus mykiss*) Fed a Plant-Based Diet, Totally
- 547 Free of Marine Resources: Consequences for Reproduction, Fatty Acid Composition and Progeny
- 548 Survival. PLOS ONE 10:e0117609. DOI: 10.1371/journal.pone.0117609.
- 549
- 550 Leardi R., Gonzalez A. 1998. Genetic algorithms applied to feature selection in PLS regression:
- 551 how and when to use them. Chemometrics and intelligent laboratory systems 41:195–207.
- 552
- 553 Leardi R., Seasholtz M., Pell R. 2002. Variable selection for multivariate calibration using a genetic
- 554 algorithm: prediction of additive concentrations in polymer films from Fourier transform-infrared
- 555 spectral data. Analytica Chimica Acta 461:189–200. DOI: 10.1016/S0003-2670(02)00272-6.
- 556
- 557 Li S., Wu H., Wan D., Systems Z-. 2011. An effective feature selection method for hyperspectral
- 558 image classification based on genetic algorithm and support vector machine. Knowledge-based
- 559 systems 24:40–48. DOI: 10.1016/j.knosys.2010.07.003.
- 560
- 561 Lin M., Cavinato A., Mayes D., Smiley S., Huang Y., Al-Holy M., Rasco B. 2003. Bruise Detection
- 562 in Pacific Pink Salmon (*Oncorhynchus gorbusha*) by Visible and Short-Wavelength Near-Infrared
- 563 (SW-NIR) Spectroscopy (600–1100 nm). Journal of Agricultural and Food Chemistry 51:6404–
- 564 6408. DOI: 10.1021/jf0346197.
- 565
- 566 Liu D., Sun D., Zeng X. 2014. Recent advances in wavelength selection techniques for
- 567 hyperspectral image processing in the food industry. Food and Bioprocess Technology 7:307–323.
- 568 DOI: 10.1007/s11947-013-1193-6.
- 569
- 570 Lu G., Fei B. 2014. Medical hyperspectral imaging: a review. Journal of Biomedical Optics
- 571 19:010901. DOI: 10.1117/1.JBO.19.1.010901.
- 572
- 573 Mendoza F., Aguilera J. 2004. Application of image analysis for classification of ripening bananas.
- 574 Journal of Food Science 69:E471–E477. DOI: 10.1111/j.1365-2621.2004.tb09932.x.
- 575

- Murthy C., Chowdhury N. 1996. In search of optimal clusters using genetic algorithms. *Pattern Recognition Letters* 17:825–832.
- Nguyen T. 1999. The cholesterol-lowering action of plant stanol esters. *Recent Advances in Nutritional Science* 129:2109–2112.
- Oliva-Teles A. 2012. Nutrition and health of aquaculture fish. *Journal of Fish Diseases* 35:83. DOI: 10.1111/j.1365-2761.2011.01333.x.
- Patterson M., Wilson B., Wyman D. 1991. The propagation of optical radiation in tissue I. Models of radiation transport and their application. *Lasers in Medical Science* 6:155–168. DOI: 10.1007/BF02032543.
- Pavlidis M., Pavlidis M., Papandroulakis N., Papandroulakis N., Divanach P., Divanach P. 2006. A method for the comparison of chromaticity parameters in fish skin: Preliminary results for coloration pattern of red skin Sparidae. *Aquaculture* 258:211. DOI: 10.1016/j.aquaculture.2006.05.028.
- Robin X., Turck N., Hainard A., Tiberti N., Lisacek F., Sanchez J., Müller M. 2011. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics* 12:1–8. DOI: 10.1186/1471-2105-12-77.
- Sefc K., Brown A., Clotfelter E. 2014. Carotenoid-based coloration in cichlid fishes. *Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology* 173:42–51. DOI: 10.1016/j.cbpa.2014.03.006.
- Segade Á., Robaina L., Ferrer O., Romero G., Domínguez M. 2015. Effects of the diet on seahorse (*Hippocampus hippocampus*) growth, body colour and biochemical composition. *Aquaculture Nutrition* 21:1. DOI: 10.1111/anu.12202.
- Solberg C., Saugen E., Swenson L., Bruun L., Isaksson T. 2003. Determination of fat in live farmed Atlantic salmon using non- invasive NIR techniques. *Journal of the Science of Food and Agriculture* 83:692–696. DOI: 10.1002/jsfa.1363.
- Sone I., Olsen R., Sivertsen A., Eilertsen G., Heia K. 2012. Classification of fresh Atlantic salmon (*Salmo salar* L.) fillets stored under different atmospheres by hyperspectral imaging. *Journal of Food Engineering* 109:482. DOI: 10.1016/j.jfoodeng.2011.11.001.
- Trichet V. 2010. Nutrition and immunity: an update. *Aquaculture Research* 41:356. DOI: 10.1111/j.1365-2109.2009.02374.x.
- Vapnik V., Vapnik V. 1998. Statistical learning theory.
- Wade N., Paulo C., Goodall J., Fischer M., Poole S., Glencross B. 2014. Quantitative methods to measure pigmentation variation in farmed Giant Tiger Prawns, *Penaeus monodon*, and the effects of different harvest methods on cooked colour. *Aquaculture* 433:513. DOI: 10.1016/j.aquaculture.2014.07.014.

- Wallat G., Lazur A., Chapman F. 2005. Carotenoids of Different Types and Concentrations in Commercial Formulated Fish Diets Affect Color and Its Development in the Skin of the Red Oranda Variety of Goldfish. *North American Journal of Aquaculture* 67:42. DOI: 10.1577/FA03-062.1.
- Wu D., Sun D., He Y. 2012. Application of long-wave near infrared hyperspectral imaging for measurement of color distribution in salmon fillet. *Innovative Food Science & Emerging Technologies* 16:361–372. DOI: 10.1016/j.ifset.2012.08.003.
- Wu D., Sun D-W., He Y. 2014. Novel non-invasive distribution measurement of texture profile analysis (TPA) in salmon fillet by using visible and near infrared hyperspectral imaging. *Food Chemistry* 145:417. DOI: 10.1016/j.foodchem.2013.08.063.
- Xu J-L., Riccioli C., Sun D-W. 2017. Comparison of hyperspectral imaging and computer vision for automatic differentiation of organically and conventionally farmed salmon. *Journal of Food Engineering* 196:170–182. DOI: 10.1016/j.jfoodeng.2016.10.021.
- Yam K., Papadakis S. 2004. A simple digital imaging method for measuring and analyzing color of food surfaces. *Journal of Food Engineering* 61:137–142. DOI: 10.1016/S0260-8774(03)00195-X.
- Yi X., Xu W., Zhou H., Zhang Y., Luo Y., Zhang W., Mai K. 2014. Effects of dietary astaxanthin and xanthophylls on the growth and skin pigmentation of large yellow croaker *Larimichthys croceus*. *Aquaculture* 433:377. DOI: 10.1016/j.aquaculture.2014.06.038.
- Zařková I., Sergejevová M., Urban J., VACHTA R., řtys D., Masojřdek J. 2009. Carotenoid-enriched microalgal biomass as feed supplement for freshwater ornamentals: albinic form of wels catfish (*Silurus glanis*). *Aquaculture Nutrition* 17:278. DOI: 10.1111/j.1365-2095.2009.00751.x.
- Zhang Y., Chen Y., Yu Y., Xue X., Tuchin V., Zhu D. 2013. Visible and near-infrared spectroscopy for distinguishing malignant tumor tissue from benign tumor and normal breast tissues in vitro. *Journal of Biomedical Optics* 18:077003–077003. DOI: 10.1117/1.JBO.18.7.077003.
- Zion B., Alchanatis V., Ostrovsky V., Barki A., Karplus I. 2008. Classification of guppies' (*Poecilia reticulata*) gender by computer vision. *Aquacultural Engineering* 38:97. DOI: 10.1016/j.aquaeng.2008.01.002.