Current Research in Food Science 9 (2024) 100789

Contents lists available at ScienceDirect

Current Research in Food Science

journal homepage: www.sciencedirect.com/journal/current-research-in-food-science

Review Article

What the fish? Tracing the geographical origin of fish using NIR spectroscopy

Nidhi Dalal *, Raffaela Ofano, Luigi Ruggiero, Antonio Giandonato Caporale, Paola Adamo

Department of Agricultural Sciences, University of Naples 'Federico II', Italy

ARTICLE INFO

Keywords:
Fish authentication
Food traceability
NIR based modelling
NIR data analysis
Seafood fraud
Food fraud

ABSTRACT

Food authentication is a growing concern with rising complexities of the food supply network, with fish being an easy target of food fraud. In this regard, NIR spectroscopy has been used as an efficient tool for food authentication. This article reviews the latest research advances on NIR based fish authentication. The process from sampling/sample preparation to data analysis has been covered. Special attention was given to NIR spectra pre-processing and its unsupervised and supervised analysis. Sampling is an important aspect of traceability study and samples chosen ought to be a true representative of the population. NIR spectra acquired is often laden with overlapping bands, scattering and highly multicollinear. It needs adequate pre-processing to remove all undesirable features. The pre-processing technique can make or break a model and thus need a trial-and-error approach to find the best fit. As for spectral analysis and modelling, multicollinear nature of NIR spectra demands unsupervised analysis (PCA) to compact the features before application of supervised multivariate techniques such as LDA, PLS-DA, QDA etc. Machine learning approach of modelling has shown promising result in food authentication modelling and negates the need for unsupervised analysis before modelling.

1. Introduction

The global fish production is forecasted to rise to 200 MT (Megatons) by 2029 at a pace of 1.3% per annum, and aquaculture will surpass wild caught fish production by 2024 (OCED/FAO, 2020). The fish production goes hand in hand, with the fish consumption expected to rise 16.3% worldwide and 5.3% in Europe. Fish protein will also make up 21.4 kg in LWE (live weight equivalent) by 2029 as opposed to 20.4 kg in 2017–2019. The global fish consumption is increasing at 0.5% per annum with increase in global fish trade. The future population trends also predict a major role of the fisheries and aquaculture to feed the increasing world population (El Sheikha & Xu, 2017). Globally, fisheries employ 33 million people directly (in fishing) and indirectly (transporting, processing, preparing, selling (FAO, 2022)

Despite the flourishing trade, the seafood industry faces fraud. The fish supply chain isn't linear as most of the general public assumes, but a complex network of intermediaries and middlemen, and also opaque despite dealing with high-value products, which makes fraud rather easy. The fact that fishing is done in national and international waters, away from any direct governmental oversight helps in hiding fraud. Food fraud is the practice of misleading consumers or customers about a product for financial gain (Robson et al., 2021). Fish fraud is done intentionally at any point in its supply chain and for financial gains and/or to hide illegal harvesting of fish from restricted waters (Reilly, 2018). The most common fish frauds are origin mislabelling and species substitution. Origin mislabelling involves misrepresentation of the fish's origin for financial gain. For example fish grown in aquaculture conditions can be sold as wild-caught to fetch a higher price. This might not feel like a serious crime, but fishing location and weather a fish was reared in an aquaculture facility of was wild-caught influences its nutritional composition (Fuentes et al., 2010). Feeding and breeding conditions are known to affect the nutritional quality of the fish, with wild fish and well-fed farmed fish known for their high content of essential polyunsaturated omega-3 and omega-6 fatty acids (Grigorakis, 2007; Tomić et al., 2017). Additionally, studies have reported noticeable differences in the taste and sensory qualities of fish caught in open seas (wild caught) compared to those farmed in controlled/semi-controlled systems with pre-determined feed (Grigorakis, 2007; Tomić et al., 2017). Another form of food fraud in the fish industry is species substitution, where high-value species are substituted with cheaper alternatives for economic gains (Grigorakis, 2007; Tomić et al., 2017)

Reports by FAO, EU and various other agencies have pointed out the

* Corresponding author.
E-mail address: nidhi.dalal@unina.it (N. Dalal).

https://doi.org/10.1016/j.crfs.2024.100789
Received 16 March 2024; Received in revised form 11 June 2024; Accepted 14 June 2024
Available online 17 June 2024

fish supply chain is at higher risk of fraud as compared to other food categories with multiple cases of seafood fraud involving products worth millions of dollars (Lawrence et al., 2024). Seafood fraud accounts for 12.4% of all food fraud in the last 10 years, second only to dairy ingredients (Everstine et al., 2024). INTERPOL and EC have placed fish as the third and second highest risk category food susceptible to food fraud (Reilly, 2018). The Rapid Alert System for Food and Feed (RASFF) under Alert and Cooperation Network in EU reported 107 instances of food fraud in fish and its products in 2020 (EC, 2021).

While fish fraud, and food fraud on a whole, may look like a victimless crime, it has been proven otherwise. The victims are thousands, if not millions of customers, the small food sellers/businesses, and can also can cause health problems in severe cases (Elliot, 2014). Some estimates suggest that food fraud affects 1% of the global food industry costing stakeholders $10–15 billions every year, while recent estimates put this value closer to $40 billions (FDA, 2023). Even though the economic ramifications of food fraud are explicitly larger than the global heroin trade, it doesn't receive the same seriousness from the government, police and even the research funding bodies as other crimes and research subjects (Elliot, 2014; Kwasi Bannor et al., 2023).

Multitude of biological and chemical methods have been developed and successfully used for authentication and traceability of seafood. Molecular strategies such as DNA barcoding (including mitochondrial-DNA), PCR-based screening and targeted proteomics such as MALDI-TOF focus on evaluation of organic compounds for accurate seafood authentication (Asensio Gil, 2007; Carrera et al., 2013; Mazzeo et al., 2008; Mazzeo and Siciliano, 2016). Inorganic fractions in fish are also used for traceability with geochemical profile (in terms of major, trace and ultra-trace elements) and stable isotope ratios (Dempson and Power, 2004; Molkentin et al., 2007; Varrà et al., 2021b). Although these methods are accurate and precise, they are time-consuming and require extensive destructive sample preparation and are expensive as well. Near Infrared Spectroscopy (NIR) offers a promising solution to these challenges. NIR spectroscopy is quick and requires minimal sample preparation. It has gained popularity in traceability studies as it provides rapid, non-destructive and cost-efficient analysis. The focus of this paper is to provide a comprehensive and critical review of studies that apply NIR spectroscopy combined with chemometrics for tracing the geographical origin of fish. Rather than issuing a summary of existing work on fish traceability, this review aims to critically examine each aspect involved in geographical provenance of fish. Beginning from sample collection and its processing to final model building, all the steps are examined for their pros and cons.

2. Laws and regulations

Combatting seafood fraud is a complex task and a single law can't address the challenges it poses in different areas of the world. The first step in combatting fish fraud, or any kind of food fraud, as food crime in order to realise the seriousness of the situation (Elliot, 2014). The contemporary global seafood industry is subject to a multitude of regulations, guidelines, and standards aimed at ensuring sustainable production and safe distribution of seafood products. In addition, the consumers now are better informed and want to shop/buy local and authentic seafood. This is driven by the perception that local caught seafood is fresh, of better quality and has environmental benefits. Consumers are increasingly willing to pay a premium for products with certified origins, such as Protected Designation of Origin (PDO) or Protected Geographical Indication (PGI) labels developed by EU.

When it comes to the authenticity of seafood products in EU, three key factors come into play, which are often associated with fraud and are defined under EU law (EC Reg. 2065/2001): production method (wild-caught or farmed, from sea or freshwater), geographical origin, and biological species. The European Union has registered 22 PDO and 47 PGI denominations for "Fresh fish, molluscs, and crustaceans and products derived therefrom" (as of June 2024, according to the DOOR database) (Fig. 1). 52 of these PDOs and PGIs are registered for EU member states, while the rest are from non-EU member states of China, Ireland, Norway, United Kingdom and Vietnam. While 75–80% of all PDO and PGI designations in Europe and Italy are assigned to plant-based products, it is important to note that the majority of economic trade involves products of animal origin (EC, 2021b). This report also pointed out that the combined trade of PGI + PDO products has seen a growth of 90% between 2005 and 2017, whereas 'fresh fish, molluscs, and crustaceans' witnessed a growth rate of 150%.

European regulations (Commission Regulation (EC) No 2065/2001) pertaining to fishery and aquaculture products require the indication of whether a product was "caught" or "farmed," as well as the catch area or the reference to the Member State or third country where the final development stage of farmed products takes place. Additionally, as an animal product, fish is subject to EU Regulation 178/2002, which mandates traceability of food and food-producing animals throughout all stages of production, processing, and distribution. The most recent regulation, EU No. 1379/2013, specifically regulates the labelling of fishery products, requiring the inclusion of the commercial designation, proper scientific name of the species, production method (caught or farmed), fishing gear used (such as hook, trap, or trawl), and the catch or production FAO area.

Similar regulations have also been released by Food and Agriculture Organisation (FAO) regarding the harmonisation of food fraud related laws and detection methodologies globally. The FAO has conducted a comprehensive review and analysis of current traceability practices in order to identify gaps and similarities among different tracking systems, with the goal of developing best practices and guidelines for traceability (Andre, 2018). The fourteenth session of the Committee on Fisheries (COFI: FT/XIV/2014/7), under the consultancy of the FAO, stressed the importance of analysing different traceability systems and formulating best practice guidelines for traceability. The objective is to establish harmonized traceability regulations among regions and countries to facilitate smooth and efficient seafood trade while maintaining the integrity and transparency of the global seafood supply chain.

Outside of EU, individual countries have also developed guidelines for seafood authenticity and traceability. The Seafood Import Monitoring Program (SIMP, 2018) by the United States outlines rigorous requirements for traceability importing seafood in US (He, 2018). Fisheries Management Act, 1991 by Australia, Act on Sustainable Fisheries (2018) in Japan and China's Food Safety Law Amendment (2015) also aim to regulate fish traceability backed with strict documentation and labelling of seafood. These frameworks come from developed economies with clear definitions and possible ability for enforcement. Developing economies, on the other hand, are still developing such frameworks for better integration in into international seafood trade. Private certification programs have also been instrumental in establishing seafood traceability. They guide consumer choice (Yang et al., 2024) and bring about transparency to the opaque seafood network (Robson et al., 2021). The Marine Stewardship Council (MSC) label for wild-caught fish and Aquaculture Stewardship Council (ASC) for farmed seafood are the most widely recognised certification for seafood (Davis and Boyd, 2021; Jones et al., 2023). MSC has issued 529 total certification with certification process being rigorous, time consuming and thorough.

3. Near-Infrared spectroscopy: Working principle

Mid and Near-Infrared spectroscopy are molecular/vibrational spectroscopy techniques used to evaluate and study the interactions of electromagnetic waves with a sample. (Cozzolino, 2015; D. Liu et al., 2013). Chemical bonds presents in food exhibit vibrational movements, such as bending, stretching, rocking, wagging or scissoring, which enables them to absorb infrared (IR) spectra. As these actions occur solely at specific energy levels, chemical bonds absorb IR radiation of defined energies at distinct wavenumbers or wavelengths. Infrared

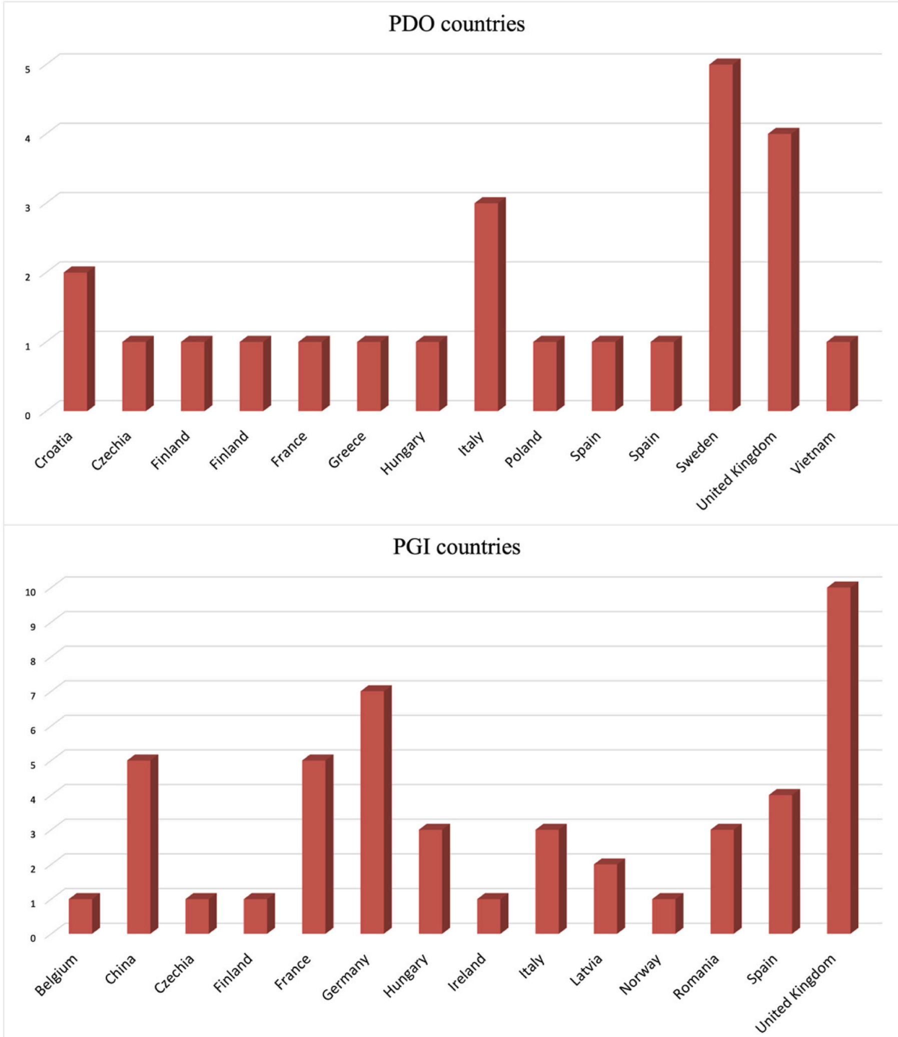

Fig. 1. List of countries with PDO (Protected Designation of Origin)and PGI (Protected Geographical Indication) fish and fish products.

spectroscopy, thus, is the measure of this absorption, resulting in a spectrum with peaks representing the chemical bonds present in the given sample (Cozzolino and Murray, 2012; Weeranantanaphan et al., 2011).

The NIR region of the spectrum is located between 700 and 2500 nm. NIR spectroscopy has been established as an efficient analytical technique in the food industry when combined with chemometric tools (e.g., data mining and data processing), as the weaker bands in the NIR region are less intense than those in the MIR (between 10 and 100 times) (Cozzolino, 2015; Cozzolino and Murray, 2012; Karoui et al., 2006). This characteristic of the NIR energy facilitates that samples can be analyzed directly without the need for pre-processing (e.g., drying, homogenisation, grinding) (Weeranantanaphan et al., 2011). Overall, the NIR spectra are the result of overtone and combination bands of fundamental vibrations of C–H, O–H, and N–H bonds (D. Liu et al., 2013; Weeranantanaphan et al., 2011).

4. Brief history of NIR spectroscopy in food and FISH

Spectroscopy has become an integral part of evaluating food quality in every part of the food supply chain in the past 55 years. Every possible kind of food matrix can be successfully assessed for its quality and chemical composition using spectroscopic-based evaluation. Initial years of spectroscopic applications in food industry saw its use for adulterant detection in various food matrices (Beć et al., 2022). With the development of high-quality PGI and PDO products around the world, quality authentication naturally turned to geographical authentication to keep pace with contemporary food regulations and food industry requirements. Osborne et al. (1993) reported the first application of NIR to authenticate basmati rice from other long-grain rice samples.

The NIR spectra of fish are a function of their composition. The composition of fish is affected by seasonal variability (Albo-Puigserver et al., 2020; Sağlam & Sağlam, 2013; Yalcin and Hulay, 2008), wild v/s farmed (Trocino et al., 2012a) and level of fish maturity (Albo-Puigserver et al., 2020) and/or fish size. Any work on fish traceability must take all these factors into account to have a holistic view of fish traceability. The first instance of NIR for fish quality determination was reported in 1999 by Pink et al. (1999). They studied the application of FT-IR to observe the quality changes in fresh and frozen Atlantic red hake (*Urophycis chuss*). Similar studies were performed to differentiate between fresh and frozen horse mackerel (*Trachurus japonicus*) using NIR spectroscopy (Uddin and Okazaki, 2004), and whiting (*Merlangius merlangus*) using MIR spectroscopy (Karoui et al., 2007) among many. These studies only aimed to determine the fish quality paying no mind to its origin and hence are not in the scope of this review. The methodology of a fish authentication study has been compiled from these studies and depicted in Fig. 2.

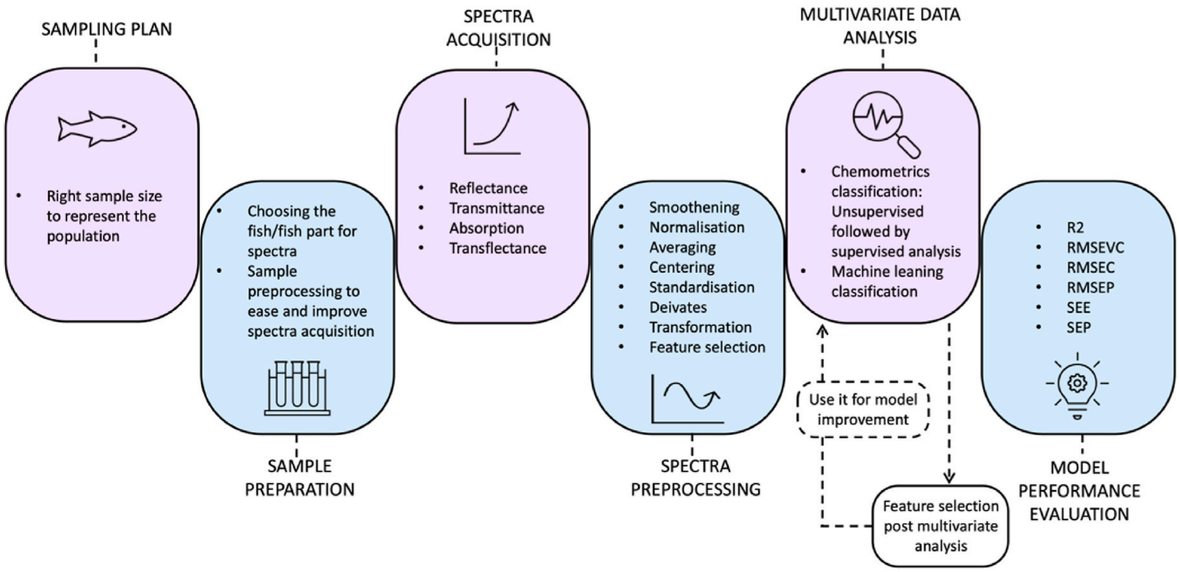

Fig. 2. Procedure for a fish authentication study.

5. Sampling and sample preparation

The most important step of any study about food (and fish) traceability lies in choosing the sampling area and number of samples. An ideal sample should represent the natural variance (of the analyte of interest) present in the target population (Donarski et al., 2019). A representative sample may be defined as "an aliquot of a material taken from a consignment and selected in such a way that it possesses all the essential characteristics of the bulk" (Murray and Cowe, 2004). In the case of fish, the natural variation can be a function of fish age or size, the season of fishing, geographical location, fishing type (wild vs farmed), freshness, pre-processing methods, the salinity of water and, water temperature, etc. (Albrecht-Ruiz and Salas-Maldonado, 2015; Gencbay and Turhan, 2016; Kaya and Turan, 2010; Romotowska et al., 2016; Xiccato et al., 2004b; Zlatanos and Laskaridis, 2007). In addition, sampling can be affected by the budget and timeframe of the project, its objectives and logistics for sample collection and access to the authentic sample. A fish traceability fraud can either replace a fish of wild origin with farmed fish or substitute a high-value species with a similar-looking low-value species (Olsen and Borit, 2018). Most studies of fish traceability have worked on differentiating between fish from different rearing systems at one or multiple locations (Costa et al., 2011; Ghidini et al., 2019; Majolini et al., 2009; Trocino et al., 2012b; Xiccato et al., 2004b) or between wild and artificially reared fish (Ghidini et al., 2019; Ottavian et al., 2012). Few studies have also dealt with species substitution (Alamprese and Casiraghi, 2015; Lv et al., 2017; O'Brien et al., 2013) Only two studies differentiated between wild fish from different geographical locations (Currò et al., 2021; Y. Liu et al., 2015b). Animals growing in the wild have a non-specific but largely varied diet whereas reared animals are fed with specific feeds to meet the market requirement. The diet of animals has been proven to affect the chemical composition of their muscles and other edible tissues (Al Hafedh, 1999; Izquierdo et al., 2003; Øverland et al., 2009; Webster et al., 1999). These changes are reflected in the NIR spectra of the fish tissue and thus aid in differentiating easily between wild and artificially reared fish. Even differences in diets of artificially reared fish can be quantified with the help of fish meal composition (Izquierdo et al., 2003; Øverland et al., 2009; Webster et al., 1999). Another important factor to consider is the sampling time. Fish composition has been known to changes with seasonal variations due to changes in food availability (Albrecht-Ruiz and Salas-Maldonado, 2015; Zlatanos and Laskaridis, 2007). In order to obtain all variability in the population, the sampling need to be performed across the seasons throughout the year.

The next step of the study is sample preparation for spectra acquisition. NIR spectra are a result of energy absorbed by the bonds present in chemical constituents of the food. Fish, like most meat-based foods, have 75–85% water (fresh weight basis). And water, with an –OH group, is a strong absorber of NIR radiation and thus NIR spectra of fresh fish are strongly dominated by water, thereby obscuring the detection of minor elements/constituents. Water has a large absorption coefficient and is absorbed in the regions between 1400–1440 nm and 1900–1950 nm (Andueza et al., 2019; Büning-Pfaue, 2003) and these regions have been used to quantify water content in food (Dalle Zotte et al., 2014). The presence of water in a food sample can also lead to better penetration of light in high-moisture foods while simultaneously causing attenuated refraction at particular wavelengths (Murray and Cowe, 2004). Apart from free water, it may also be present in association with ions, polymers and organic molecules in food using hydrogen bonds. Thus, water absorption bands can also be influenced by the association of water with solutes present in the food (Büning-Pfaue, 2003). In an intact fish muscle, most water is present as free water. The ratio of free water to bound water decreases when the fish is homogenised since fish proteins tend to form gels (Stone and Stanley, 1992; Yi et al., 2020). Thus spectra of fresh fish, homogenised fish and freeze-dried meat samples will differ with respect to the peaks presented by water molecules (Andueza et al., 2019; Dalle Zotte et al., 2014; Giaretta et al., 2019). Homogenisation of sample leads to better presentation of samples to the NIR wavelengths and have yielded better results (high calibration coefficients low standard errors) in predicting properties of meat products such as pork sausages, pork muscles, oysters, lamb, beef, chicken and European Sea Bass (Cozzolino et al., 2003; Cozzolino and Murray, 2002; Madigan et al., 2012; Ortiz-Somovilla et al., 2007; Xiccato et al., 2004b). Apart from homogenisation, freeze-drying the meat sample also improves the quality of NIR spectra. Better spectra quality has led to better quality predictions in beef, trout, beef meat patty, and European sea bass (Andueza et al., 2019; Dalle Zotte et al., 2014; Giaretta et al., 2019; Trocino et al., 2012b; Xiccato et al., 2004a). Most studies on fish traceability used fresh fish samples to acquire the spectra (Alamprese and Casiraghi, 2015; Currò et al., 2021; Ghidini et al., 2019; Y. Liu et al., 2015; Lv et al., 2017; O'Brien et al., 2013; Ottavian et al., 2012) and did not comment on the use of homogenised or freeze-dried samples. Xiccato et al. (2004) however, did compare the efficiency of whole, homogenised and freeze-dried muscles in assessing the rearing system and chemical composition of European sea bass. They concluded that three samples were performed in the order freeze-dried > homogenised > whole muscle to identify the rearing system and predict chemical composition.

6. Data analysis

Data analysis of NIR spectra and it's modelling is long process and is depicted in Fig. 3.

6.1. Spectral pre-processing

Raw NIR spectra obtained after scanning the sample contains information in form of numerous variables (thousands) relating to the chemical constituents of the sample. However, this information is contained in many overlapping bands, where an individual analyte also might absorb at several wavelengths, resulting in multi-collinearity (Goedhart, 1990). Some of these absorption bands are also weak and can lead to misrepresentation of data, while others can be influenced by the non-linearities introduced by light scattering (Rinnan et al., 2009; Yu, 2021). The spectra can also contain undesirable baseline shifts and non-linear scatter effects, caused by light scattering and differences in effective path length (for solid samples) (Rinnan et al., 2009). This necessitates the need to spectral pre-processing before moving forward with exploratory analysis.

Spectral (Duckworth, 2003; Rinnan et al., 2009) pre-processing improves the subsequent exploratory analysis and the classification model by removing or reducing un-modellable variability (Bouveresse, 1997; Rinnan et al., 2009). This in turn enhances spectral features of interest in the spectra. The choice of the pre-processing technique however must not be too severe. On one hand, the spectral feature of interest can be enhanced using pre-processing, but choosing a high degree of pre-processing or choosing the wrong type of pre-treatment can also lead to the loss of key information (Rinnan et al., 2009). For example, smoothening can remove noise when used at appropriate degrees, but its overuse can lead to loss of spectral features. Similarly, derivatives can decrease SNR by enhancing noise while enhancing spectral differences (Bouveresse, 1997). The choice of accurate pre-processing is hard to determine before model building since there is no standard procedure to decide on the pre-processing. Lv et al. (2017) checked the effects of pre-processing methods on the prediction ability of four multivariate models to differentiate between different species of carp. The pre-processing methods were evaluated based on accuracy, specificity, sensitivity and precision in initial fitting and cross-validation. They concluded that pre-processing depended on the kind of meat being used for spectra acquisition since the pre-processing they used for fish authentication gave different results when used for beef. The pre-processing performance depended on the multivariate methodology used for modelling. Similarly, Alamprese & Casiraghi (2015) also evaluated the efficiency of pre-treatments to differentiate between species of red mullet and Atlantic mullet, and flounder and plaice. The choice of pre-treatment of spectra from the same instrument depended on the species. While SNV pre-treated FT-NIR (optical fibre) spectra classified mullets with 100% accuracy, the same spectral data could only give 84.97% correct classification with SNV and first derivate pre-treated spectra. The pre-treatments also changed by integrating spheres and optical fibre probes.

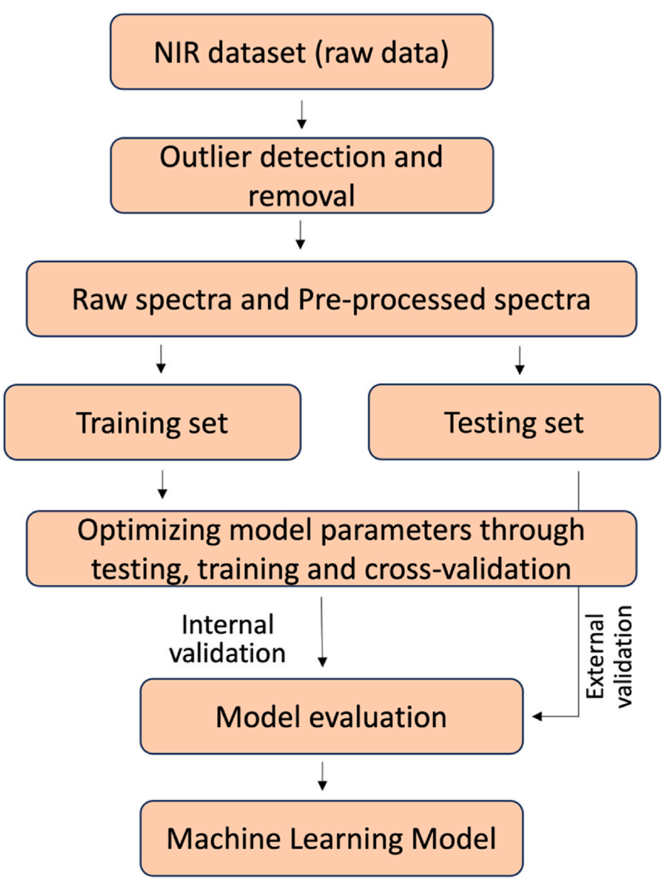

Fig. 3. Process of data analysis for NIR spectra in an authentication study.

While some of the studies on fish traceability chose to pre-process the spectral data (Alamprese and Casiraghi, 2015; Currò et al., 2021; Ghidini et al., 2019; Lv et al., 2017; Majolini et al., 2009; O'Brien et al., 2013; Ottavian et al., 2012; Trocino et al., 2012b; Xiccato et al., 2004a), others (Costa et al., 2011; Y. Liu et al., 2015) chose not to use any pre-processing technique. While the pre-processing technique is rather essential for developing multivariate models with NIR spectra, Lv et al. (2017) demonstrated that models developed with raw spectra performed considerably similar to the ones developed with pre-processes spectra in the terms of validation. However, they had lower performance in cross-validation and even lower in prediction. Most studies used a combination of pre-treatments, with SNV and first or second-order derivatives of SG being the most common choice, for best model performance (Alamprese and Casiraghi, 2015; Currò et al., 2021; Ghidini et al., 2019; Ottavian et al., 2012). Studies by Majolini et al. (2009) Trocino et al. (2012), and Xiccato et al. (2004) used second-order SG derivate calculated over 20 points of smoothening.

The large size of NIR spectra can hinder model building by including irrelevant data that doesn't provide relevant information. Feature or wavelength selection can be used to reduce the bulk of data in addition to improving the model operation efficiency and model interpretability. When the uncorrelated and non-linear variables of NIR spectra are removed with feature selection, the prediction ability and robustness of the final model is improved (Yun, 2022). It has also been proven to improve the prediction accuracy in fish (Currò et al., 2021; Lv et al., 2017; O'Brien et al., 2013; Varrà et al., 2021), quinoa flour (Wang et al., 2022), asparagus (Richter et al., 2019), garlic (Han et al., 2024), tea (Jin et al., 2022), mushrooms (Chen et al., 2022) and many other food products. It involves the identification and selection of relevant spectral regions or features that carry the most discriminatory information about the geographical origin of the food. The selection can be made pre or post multivariate analysis on preprocessed spectra. While the goal of feature selection before multivariate analysis is reduce dataset dimensionality and remove redundant variable, feature selection after multivariate analysis is performed with the help of latent variable/components to select most influential variables from spectra to refine the final model. Feature selection is performed by the means of specially designed algorithms/methods in data science programming languages such as R, Python, MatLab, Julia, SQL, Java, Scala, etc. (Yun et al., 2019). The choice of feature selection algorithm depends on the type of model and the spectra obtained. For example, if the NIR based model is needed for moisture content prediction, the wavelengths corresponding to the O–H bonds would be logically targeted as variables of interest (Yun et al., 2019). The same algorithm might not work for authenticating different food products with equal robustness. For example, Competitive adaptive re-weighted sampling (CARS) in MatLab for feature selection worked efficiently for the authenticating the geographical provenance of mushrooms (Chen et al., 2022), while the same algorithm couldn't classify tea based on its geographical origin (Jin et al., 2022).

Lv et al. (2017) also applied CARS to choose optimal wavelengths by evaluating weight values and large absolute regression coefficients, ensuring interpretability to the desired chemical property. CARS, however, did not produce optimum result and the final model did not exhibit sufficient robustness as compared to the one with complete spectra. Currò et al. (2021) on the other hands demonstrated the application of Boruta package of random forest feature selection procedure based on the Boruta algorithm in R. It uses a random forest algorithm to identify important features by comparing them with randomly generated shadow features. Features that outperform shadow features significantly are considered important, aiding in the selection of relevant variables for predictive modeling tasks (Kursa and Rudnicki, 2010). Only one instance of post-multivariate feature selection in fish authentication was observed by building Ghidini et al. (2019). They used VIP (Variable Influence on Projection) to find the most influential absorption bands of the spectra to explain the chemical composition of fish and but did not use it for further model improvement. VIP feature selection was applied by (Ottavian et al., 2012) for classification of European Sea bass but the model obtained after VIP feature selection was inferior to the one with complete spectra. While there are only a handful of instances of feature selection for fish traceability, it has been used extensively for authentication of other food products. Feature selection while reducing the bulk of data can also improve the model accuracy as demonstrated by Han et al. (2024) for determining the geographic origin of garlic. They applied feature selection after initial to modelling with machine learning algorithms to pick the features with most relevant input for the model. The feature selection in this case was done by Genetic algorithm (GA), which is a Python based stochastic method for function optimization inspired by natural genetics and biological evolution (Altarabichi et al., 2023). The same method of feature selection was also used by Jin et al. (2022) to authenticate high-quality tea from China. The compared GA with continuous projection algorithm (SPA), CARS methodologies for feature selection. The GA feature selection outperformed CARS and SPA even when GA selected less wavelengths as compared for CARS for modelling. Richter et al. (2019), on the other hand, used a much simpler approach for feature selection for geographical provenance of asparagus. Instead of using pre-existing feature selection algorithms, they simply binned 10 adjacent features to reduce the data bulk 10 times which subsequently improved the model performance. This goes on to show the need for an appropriate feature selection algorithm for the dataset, food product and the need of the model at hand.

6.2. Data modelling approach

The next step after spectral preprocessing is the multivariate statistical analysis. A reliable model for traceability and authentication needs sound statistical backing. Spectroscopic data obtained from NIR spectra contain strongly correlated hundreds, if not thousands of variables. This is essentially a mathematical and computational challenge for any chemometrician. Thus, the data needs to be compressed, with the least amount of information being lost and accuracy maintained, to a lower variable number such that it is easily explorable and visualized. Principal Component Analysis is the most widely used unsupervised (class membership not known) data reduction method and has been proven to work well with the multicollinear nature of NIR spectra. It transforms the originally large data matrix into a simpler representation that uses a reduced number of compressed variables called principal components (PCs) (Berrueta et al., 2007). PCA shows the hidden trends in the data with each PC being the linear combination of the original variable and variable importance being defined by the loadings of the respective PC. Most studies on fish authentication (Table 1) used PCA as the unsupervised chemometric methodology (Alamprese and Casiraghi, 2015; Currò et al., 2021; Ghidini et al., 2019; Y. Liu et al., 2015; Lv et al., 2017; Majolini et al., 2009; O'Brien et al., 2013; Ottavian et al., 2012; Trocino et al., 2012b; Xiccato et al., 2004a), except Costa et al. (2011). In the end, PCA provides the knowledge of class membership of the analyzed spectra which is crucial to the development of a predictive model. Apart from PCA, PLS (Partial least square) can also be used to obtain a small variable set in the form of orthogonal factors from a larger dataset to reduce the bulk of data for classification. Though unlike PCA, PLS is a supervised form of data reduction applied based on linear regression with a much lower risk of chance correlation.

The challenge of fish authentication and traceability is that of classification, where samples need to be classified into more than one group. Thus, predictive models for fish authentication have all used classification modelling approaches instead of calibrative ones. PCs obtained by PCA are thus used to generate predictive models with supervised methods (class membership known) to classify existing and possibly unknown samples in future. Samples groups are differentiated based on spectral (quality) attributes or pre-defined classes based on the information gathered by unsupervised methods to enable the classification of the new unknown samples (Berrueta et al., 2007; Héberger et al., 2003). Linear discriminant analysis (LDA), quadratic discriminant analysis (QDA), partial least squares regression-discriminant analysis (PLS-DA), soft independent modelling of class analogies (SIMCA), Orthogonal PLS-DA (OPLS-DA), SVM and KNN are the most commonly used classification techniques for NIR spectra of meat products. LDA determines the linear discriminant to maximize the between-class variance ratio and minimize the within-class variance ratio. The discriminant function thus obtained is a linear combination of independent or original variables that can discriminate between the categories of dependent variables with the highest accuracy. Classes determined by LDA follow a multivariate normal distribution and variance and co-variance matrices of the datasets must be of equal size for LDA to work and calculate a discriminatory function (Berrueta et al., 2007; Héberger et al., 2003). Apart from a linear function, a quadratic function can also be used for discrimination (QDA) which is less subject to the constraints of object distribution, unlike LDA and establishes parabolic boundaries. Both LDA and QDA, however, require that the number of samples is higher than the number of variables and also affected negatively by collinearity among variables (Bevilacqua et al., 2013). This constraint of having more variables than samples can either be overcome by feature reduction using PCA, PLS and/or variable selection. In SIMCA, however, objects are classified by measuring how far they are from class models. SIMCA can assign objects to multiple classes, but in traceability studies, objects should belong to only one class. SIMCA uses a fixed number of principal components (PCs) with the most variation, which may not align with the direction that separates the classes. PLS-DA combines PLS regression with LDA for classification and provides class-separating information used for discrimination. Both, SIMCA and PLS-DA are, however, sensitive to outliers and PLS-DA is also prone to overfitting which can result in the misclassification of samples (we assume that the outliers are caused by mis-acquisition of spectra and not the sample variability). PLS-DA works better in instances of low within-class variability, as class separation is maximised (Bylesjö et al., 2006). Another method used for classification in traceability studies is OPLS-DA which combines the strengths of SIMCA and PLS-DA. It's a continuation of PLS regression featuring the use of the integrated OSC filter initially developed for spectral pre-processing and works better with collinear and noisy data like NIR (Bylesjö et al., 2006).

Xiccato et al., (2004) used NIR based SIMCA model to differentiate between European Sea bass (*Dicentrarchus labrax*) from four different growing systems and were only able to obtain 80% correct classification

Table 1
Studies dealing fish Traceability using NIR spectroscopy.

Aim of the study	Geogrpahical Location	Equipment required	Fish species	Sample type	Spectral pre-processing	Un-supervised analysis	Supervised analysis	Reference
Differentiate between fish rearing systems	Central Tyrrhenian Sea and North Adriatic Sea	NIR spectrometer (1100–2500 nm)	European sea bass (Dicentrarchus labrax)	fresh Fish fillets and freeze dried fillet	Transformed in second derivative, calculated over a 20-point smoothing.	PCA	SIMCA	Xiccato et al., 2004
Differentiate between fish rearing systems	Northern and Southern Adriatic Sea, Tyrrhenian Sea	NIR spectrometer (1100–2500 nm)	European sea bass (Dicentrarchus labrax)	Fish fillets	Transformed in second derivative, calculated over a 20-point smoothing.	PCA	None	Majolini et al. (2009)
Differentiate between fish rearing systems	Southern Adriatic Sea	Vis-NIR spectrometer (400–970 nm)	European sea bass (Dicentrarchus labrax)	Whole fish, 48 h and 96 h post mortem	none	None	PLS	Costa et al. (2011)
Differentiate between fish rearing systems	Northern Adriatic Sea	NIR spectrometer (1100–2500 nm)	European sea bass (Dicentrarchus labrax)	Fresh minced meat (left fillet) and freeze-dried minced meat (right fillet)	Transformed in second derivative, calculated over a 20-point smoothing.	PCA	SIMCA	Trocino et al., 2012
Descriminate between wild and farmed fish	Not available	NIR System 5000, 1100–2500 nm	European sea bass (Dicentrarchus labrax)	fish	SNV and first and second order derivates of SG	PCA	PLS-DA and Wavelet-based WPTER (wavelet packet transform for efficient pattern recognition)	Ottavian et al. (2012)
Distinguish superior from lower quality fish species	Not available	MicroNIR 1700 spectrometer (887–1667 nm)	Red mullet and mullet, winter cod and cod and samlet and salmon trout	skin and meat of fillet	Wavelength selection and EMSC	PCA	SIMCA	O'Brien et al. (2013)
Distinguish superior from lower quality fish species	Not available	FT-IR spectrometer (700–10000 nm) fitted with integrating sphere (800–2667 nm) and optical fibre (910–2273 nm)	Red mullet (*Mulls surmuletus*) v/s Atlantic mullet (*Pesudupeneus prayensis*) and Plaic (Pleuronectes platessa) v/s Flounder (Platichthya flesus flesus)	Homogenised fish fillets	Individual and/or combination with MSC, SNV and SG algoritm.	PCA	LDA and SIMCA	Alamprese and Casiraghi (2015)
Determine fish origin	Guangdong, Hainan, Guangxi and Fujian province in South China Sea (exact longitudes and latitudes mentioned)	NIRFlex N-500 (1000–2500 nm)	tilapia	Homogenised fish flllets	none	PCA	SIMCA	Liu et al. (2015)
Differentiate between species from same family	Hubei Province, China	NIR spectrometer (1100–1799 nm)	Black carp (*Mylopharyngodon piceus*), grass carp (*Ctenopharyngodon idellus*), silver carp (*Hypophthalmichthys molitrix*), bighead carp (*Aristichthys nobilis*), common carp (*Cyprinus carpio*), crucian (*Carassius auratus*), and bream (*Parabramis pekinensis*)	homogenised dorsal muscle	MSC and SNV individually; feature selection with Competitive Adaptive Re-weighted Sampling (CARS); spectral transformation with Fast Fourier Transformation (FFT)	PCA	PCA-LDA, PLS-LDA, CARS-LDA, FFT-LDA	Lv et al. (2017)
Differentiate between fish rearing systems, wild v/s farmed fish and determine fish provenance	Western, Central or Eastern Mediterranean Sea	NIR spectrometer (1100–2500 nm)	European sea bass (*Dicentrarchus labrax*)	Homogenised fish fillets	SNV, second derivate and 15-points SG smoothening	PCA	OPLS-DA	Ghidini et al. (2019)

(*continued on next page*)

Table 1 (*continued*)

Aim of the study	Geogrpahical Location	Equipment required	Fish species	Sample type	Spectral pre-processing	Un-supervised analysis	Supervised analysis	Reference
Determine fish origin	Adriatic Sea, north-eastern and eastern central Atlantic Oceans, and eastern Indian and western central Pacific Oceans	Potable NIR spectrometer (902–1680 nm)	Cuttlfish (*Sepia officinalis and Sepiella inermis*)	fresh fillets and patties	SNV, detrending, smoothing, and second derivate followed by feature selection	PCA	SVM and KNN	Currò et al. (2021)

between them after extensive derivate based spectral preprocessing. SIMCA led to more samples not classified rather than being classified to the wrong group. The same fish was again the topic of authenticity based on three rearing systems instead of four with vis-NIR spectra (Costa et al., 2011). They however, obtained better classification of 87% with PLS based modelling without any spectral preprocessing. The study was repeated again Trocino et al. (2012) for differentiating between organic and semi-intensive fish rearing systems. The SIMCA results differed with kind of sample used for spectra acquisition. While >70% correct classification was seen for freeze-dried fillets, fresh minced fillets had a much lower classification percentage. Again, more sample were left unclassified than wrongly classified by SIMCA. Wild and farmed European Sea bass was differentiated using NIR. Pre-processed spectra were modelled with PLS-DA with and without VIP based feature selection. The model with complete spectra exhibited better performance with only 3 misclassified samples. They also proposed classification based on Wavelet-based WPTER (wavelet packet transform for efficient pattern recognition). WPTER is algorithm package that decomposes the data or signals into smaller parts to get detailed information which is then used for internal feature selection, dimensionality reduction and final classification (Ma et al., 2023). WPTER based classification also had only 1 extra misclassified sample and was almost on par with PLS-DA in terms of model performance. SIMCA was again applied by O'Brien et al. (2013) to distinguish red mullet from mullet, winter cod from cod and salmon trout from samlet. The former ones are high quality (higher price) fish while the latter ones are of lower quality (lower price). They successfully discriminated red mullet from mullet and winter cod from cod. Salmon trout, however, was not completely discriminated from trout with SIMCA modelling. SIMCA was also used to differentiate between tilapia from four provinces from China (Guangdong, Hainan, Guangxi and Fujian province in South China Sea). The spectra were not pre-processed and data was reduced with PCA followed by SIMCA. They observed 85%, 82%, 75% and 83% fish were correctly classified from each of the 4 locations respectively.

While only one modelling approach was evaluated in the previous studies, Alamprese and Casiraghi (2015) and Lv et al. (2017) different modelling approached for fish authentication (Table 1). (Alamprese and Casiraghi, 2015) Red mullet (*Mulls surmuletus*) v/s Atlantic mullet (*Pesudupeneus prayensis*) since they both have similar appearance but Red mullet fetches a higher price. Spectra was acquired in three different wavelength: FT-IR spectrometer (700–10000 nm) fitted with integrating sphere (800–2667 nm) and optical fibre (910–2273 nm). The spectra was pre-processed and then classified with LDA and SIMCA. They observed 100% correct classification with LDA in all three wavelengths ranges. SIMCA on the other hand had 99.50%, 100% and 100% correct classification with integrating sphere, optical fibre and FT-IR spectra. They used the same methodology for differentiating between Plaice (*Pleuronectes platessa*) v/s Flounder (*Platichthya flesus*). Consecutively, SIMCA performed at par with LDA and exhibited 84.95%, 84.97% and 100% correct classification with integrating sphere, optical fibre and FT-IR spectra. LDA demonstrated 92%, 88% and 100% correct classification with integrating sphere, optical fibre and FT-IR spectra. Lv et al. (2017) demonstrated the difference in the prediction ability of models built with raw and various kinds of pre-processed spectra and for carp species [Black carp (*Mylopharyngodon piceus*), grass carp (*Ctenopharyngodon idellus*), silver carp (*Hypophthalmichthys molitrix*), bighead carp (*Aristichthys nobilis*), common carp (*Cyprinus carpio*), crucian (*Carassius auratus*), and bream (*Parabramis pekinensis*)]. Fast fourier transformation (FFT) was applied to the spectra along with MSC and SNV pre-processing and CARS feature selection. PCA-LDA model with MSC pre-treated spectra had the highest classification of 100%. Model based on spectral transformation (FFT-LDA) did not achieve 100% classification. The PCA-LDA model worked better than PLS-LDA with higher accuracy, specificity, sensitivity, and precision across training, testing and cross-validation sets. They also isolated wavelengths with CARS with the highest discrimination power from the PCA-LDA model and then used them for the modelling with LDA (CARS-LDA). This new model proved 100% accurate and successfully proved that NIR modelling doesn't need a large dataset.

The predictive ability of OPLS-DA was exhibited by Ghidini et al. (2019) for European sea bass. They attempted to differentiate European Sea bass based on production method (wild or farmed), rearing system (extensive, semi-intensive or intensive), and provenance. The SNV and derivate pre-treated spectra was used for PCA based data reduction. PCA scores rather than the whole spectra were used to build the OPLS-DA model and 88.89% correct classification was observed for provenance. Better classification of 100% and 94.44% was reported for production method and farming system respectively.

Apart from traditional multivariate analysis, Machine learning (ML) based classification has been gaining traction. ML based classification has been used extensively for food quality estimation using NIR. The use of NIR couples with ML for food traceability has been less prominent, and in fact we could only find one study with the application on fish traceability by Currò et al. (2021) on cuttlefish. They used two-step ML classification to classify cuttlefish from five distinct FAO fishing area. Random forest was used for feature selection and the selected wavelengths were used for classification. Support vector machine (SVM) and K-nearest neighbor (KNN) were compared for the accuracy in differentiating fish from different regions. ML techniques, however, were not compared with traditional multivariate classification algorithms.

Common to both unsupervised and supervised methods is the importance of careful model validation to avoid over-fitting and thus over-optimistic classification results (Bro et al., 2008), which currently represents one of the drawbacks in authenticity studies. An principal solution to combat this challenge is a collection of samples across the spectrum of season throughout the year. Fish composition changes with seasonal availability of food (Albo-Puigserver et al., 2020; Albrecht-Ruiz and Salas-Maldonado, 2015) and these changes ought to be encompassed in a traceability model to improve the robustness.

7. Conclusions

The work done on fish traceability has come a long way with the advanced chemometric methods that are being used now. Several food classes have seen the use of techniques such as Machine Learning for their authentication. However, fish, so far, haven't seen its application for their authentication. The biggest challenge with conducting studies on fish authentication is the collection of authentic samples that represent the variance present in the population to generate feasible and stable models. Another challenge we observed was the lack of studies that worked on geographical traceability, and not authentication (freshness).

In conclusion, the growing seafood trade presents challenges related to fraud and mislabelling. However, NIR spectroscopy coupled with chemometrics offers a promising solution for tracing the geographical origin of fish. This paper provides a detailed analysis of the different aspects involved in fish traceability studies, aiming to construct robust and reliable traceability models. By addressing these challenges, the seafood industry can enhance transparency, ensure product authenticity, and meet the demands of consumers for accurate information about the origin and quality of the fish they consume.

CRediT authorship contribution statement

Nidhi Dalal: Conceptualization, Writing – original draft, Writing – review & editing, Visualization, Literature search. **Raffaela Ofano:** Writing – original draft, Visualization, Literature search. **Luigi Ruggiero:** Conceptualization, Writing – original draft. **Antonio Giandonato Caporale:** Supervision, Writing – review & editing. **Paola Adamo:** Conceptualization, Supervision, Writing – review & editing.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Nidhi Dalal reports financial support was provided by Partnership for Research and Innovation in the Mediterranean Area (PRIMA). If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

References

Al Hafedh, Y.S., 1999. Effects of dietary protein on growth and body composition of Nile tilapia, Oreochromis niloticus L. Aquacult. Res. 30, 385–393.

Alamprese, C., Casiraghi, E., 2015. Application of FT-NIR and FT-IR spectroscopy to fish fillet authentication. LWT 63 (1), 720–725. https://doi.org/10.1016/j.lwt.2015.03.021.

Albo-Puigserver, M., Sánchez, S., Coll, M., Bernal, M., Sáez-Liante, R., Navarro, J., Palomera, I., 2020. Year-round energy dynamics of sardine and anchovy in the north-western Mediterranean Sea. Mar. Environ. Res. 159 https://doi.org/10.1016/j.marenvres.2020.105021.

Albrecht-Ruiz, M., Salas-Maldonado, A., 2015. Chemical composition of light and dark muscle of Peruvian anchovy (Engraulis ringens) and its seasonal variation. J. Aquat. Food Prod. Technol. 24 (2), 191–196. https://doi.org/10.1080/10498850.2012.762705.

Altarabichi, M.G., Nowaczyk, S., Pashami, S., Mashhadi, P.S., 2023. Fast Genetic Algorithm for feature selection — a qualitative approximation approach. Expert Syst. Appl. 211, 118528 https://doi.org/10.1016/j.eswa.2022.118528.

Andre, V., 2018. Good practice guidelines (GPG) on national seafood traceability systems. https://openknowledge.fao.org/server/api/core/bitstreams/1cece1f0-a44e-4d90-b3c9-42608277e89f/content.

Andueza, D., Listrat, A., Durand, D., Normand, J., Mourot, B.P., Gruffat, D., 2019. Prediction of beef meat fatty acid composition by visible-near-infrared spectroscopy was improved by preliminary freeze-drying. Meat Sci. 158 https://doi.org/10.1016/j.meatsci.2019.107910.

Asensio Gil, L., 2007. PCR-based methods for fish and fishery products authentication. Trends Food Sci. Technol. 18 (11), 558–566. https://doi.org/10.1016/j.tifs.2007.04.016.

Beć, K.B., Grabska, J., Huck, C.W., 2022. Miniaturized NIR spectroscopy in food analysis and quality control: Promises, challenges, and Perspectives. Foods 11 (10). https://doi.org/10.3390/foods11101465.

Berrueta, L.A., Alonso-Salces, R.M., Héberger, K., 2007. Supervised pattern recognition in food analysis. J. Chromatogr. A 1158 (Issues 1–2), 196–214. https://doi.org/10.1016/j.chroma.2007.05.024.

Bevilacqua, M., Bucci, R., Magrì, A.D., Magrì, A.L., Nescatelli, R., Marini, F., 2013. Classification and class-modelling. Data Handling Sci. Technol. 28, 171–233. https://doi.org/10.1016/B978-0-444-59528-7.00005-3. Elsevier Ltd.

Bouveresse, E., 1997. *Maintenance and transfer of multivariate calibration models based on near-infrared spectoscopy* [Vrije Universiteit, Brussles]. https://researchportal.vub.be/en/studentTheses/maintenance-and-transfer-of-multivariate-calibration-models-based.

Bro, R., Kjeldahl, K., Smilde, A.K., Kiers, H.A.L., 2008. Cross-validation of component models: a critical look at current methods. Anal. Bioanal. Chem. 390 (5), 1241–1251. https://doi.org/10.1007/s00216-007-1790-1.

Büning-Pfaue, H., 2003. Analysis of water in food by near infrared spectroscopy. Food Chem. 82 (1), 107–115. https://doi.org/10.1016/S0308-8146(02)00583-6.

Bylesjö, M., Rantalainen, M., Cloarec, O., Nicholson, J.K., Holmes, E., Trygg, J., 2006. OPLS discriminant analysis: Combining the strengths of PLS-DA and SIMCA classification. J. Chemometr. 20 (8–10), 341–351. https://doi.org/10.1002/cem.1006.

Carrera, M., Cañas, B., Gallardo, J.M., 2013. Fish authentication. In: Proteomics in Foods. Springer US, pp. 205–222. https://doi.org/10.1007/978-1-4614-5626-1_12.

Chen, X., Liu, H., Li, J., Wang, Y., 2022. A geographical traceability method for Lanmaoa asiatica mushrooms from 20 township-level geographical origins by near infrared spectroscopy and ResNet image analysis techniques. Ecol. Inf. 71, 101808 https://doi.org/10.1016/J.ECOINF.2022.101808.

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. Aquacult. Int. 19 (4), 765–778. https://doi.org/10.1007/s10499-010-9393-9.

Cozzolino, D., 2015. ScienceDirect Foodomics and infrared spectroscopy : from compounds to functionality. Curr. Opin. Food Sci. 4, 39–43. https://doi.org/10.1016/j.cofs.2015.05.003.

Cozzolino, D., Murray, I., 2002. Effect of sample presentation and animal muscle species on the analysis of meat by near infrared reflectance spectroscopy. J. Near Infrared Spectrosc. 10 (1), 37–44. https://doi.org/10.1255/jnirs.319.

Cozzolino, D., Murray, I., 2012. A review on the application of infrared technologies to determine and monitor composition and other quality characteristics in raw fish, fish products, and seafood. Appl. Spectrosc. Rev. 47 (3), 207–218. https://doi.org/10.1080/05704928.2011.639106.

Cozzolino, D., Barlocco, N., Vadell, A., Ballesteros, F., Gallieta, G., 2003. The Use of Visible and Near-Infrared Reflectance Spectroscopy to Predict Colour on Both Intact and Homogenised Pork Muscle, vol 36.

Currò, S., Balzan, S., Serva, L., Boffo, L., Ferlito, J.C., Novelli, E., Fasolato, L., 2021. Fast and green method to Control frauds of geographical origin in traded cuttlefish using a Portable infrared reflective instrument. https://doi.org/10.3390/foods10081678.

Dalle Zotte, A., Ottavian, M., Concollato, A., Serva, L., Martelli, R., Parisi, G., 2014. Authentication of raw and cooked freeze-dried rainbow trout (Oncorhynchus mykiss) by means of near infrared spectroscopy and data fusion. Food Res. Int. 60, 180–188. https://doi.org/10.1016/j.foodres.2013.10.033.

Davis, R.P., Boyd, C.E., 2021. A comparison of the technical efficiency of Aquaculture Stewardship Council certified shrimp farms to non-certified farms. Current Research in Environmental Sustainability 3, 100069. https://doi.org/10.1016/J.CRSUST.2021.100069.

Dempson, J.B., Power, M., 2004. Use of stable isotopes to distinguish farmed from wild Atlantic salmon, Salmo salar. Ecol. Freshw. Fish 13 (3), 174–184. https://doi.org/10.1111/j.1600-0633.2004.00057.x.

Donarski, J., Camin, F., Fauhl-Hassek, C., Posey, R., Sudnik, M., 2019. Sampling guidelines for building and curating food authenticity databases. Trends Food Sci. Technol. 90, 187–193. https://doi.org/10.1016/J.TIFS.2019.02.019.

Duckworth, J., 2003. Mathematical data preprocessing. In: Roberts, C.A., Workman Jr., J., Reeves III, J.B. (Eds.), Near-Infrared Spectroscopy in Agriculture44, pp. 113–132. https://doi.org/10.2134/agronmonogr44.c6. John Wiley & Sons.

El Sheikha, A.F., Xu, J., Jp, 2017. Traceability as a key of seafood safety: Reassessment and possible applications. Reviews in Fisheries Science and Aquaculture 25 (2), 158–170. https://doi.org/10.1080/23308249.2016.1254158. Taylor and Francis Inc.

Elliot, C., 2014. Elliott review into the integrity and assurance of food supply networks-final report. https://www.gov.uk/government/publications/elliott-review-into-the-integrity-and-assurance-of-food-supply-networks-final-report.

European Commission, 2021. The Rapid Alert System for Food and Feed (RASFF) - annual report 2020. https://doi.org/10.2875/259374.

European Commission, Directorate-General for Agriculture and Rural Development, 2021b. Study on economic value of EU quality schemes, geographical indications (GIs) and traditional specialities guaranteed (TSGs) : final Report. https://data.europa.eu/doi/10.2762/396490.

Everstine, K.D., Chin, H.B., Lopes, F.A., Moore, J.C., 2024. Database of food fraud Records: summary of data from 1980 to 2022. J. Food Protect. 87 (3), 100227 https://doi.org/10.1016/J.JFP.2024.100227.

FAO, 2022. The state of world fisheries and aquaculture 2022. Towards Blue transformation. In: The State of World Fisheries and Aquaculture 2022. FAO. https://doi.org/10.4060/cc0461en.

Fuentes, A., Fernández-Segovia, I., Barat, J.M., Serra, J.A., 2010. Physicochemical characterization of some smoked and marinated fish products. J. Food Process. Preserv. 34 (1), 83–103. https://doi.org/10.1111/j.1745-4549.2008.00350.x.

FDA, 2023. Economically motivated adulteration (food fraud). Food and Drug Administration (FDA). https://www.fda.gov/news-events/press-announcements/federal-judge-enters-consent-decree-against-washington-state.

Gencbay, G., Turhan, S., 2016. Proximate composition and nutritional profile of the Black sea anchovy (Engraulis encrasicholus) whole fish, fillets, and by-products.

J. Aquat. Food Prod. Technol. 25 (6), 864–874. https://doi.org/10.1080/10498850.2014.945199.
Ghidini, S., Varrà, M.O., Dall'Asta, C., Badiani, A., Ianieri, A., Zanardi, E., 2019. Rapid authentication of European sea bass (Dicentrarchus labrax L.) according to production method, farming system, and geographical origin by near infrared spectroscopy coupled with chemometrics. Food Chem. 280, 321–327. https://doi.org/10.1016/j.foodchem.2018.12.075.
Giaretta, E., Mordenti, A., Palmonari, A., Brogna, N., Canestrari, G., Belloni, P., Cavallini, D., Mammi, L., Cabbri, R., Formigoni, A., 2019. NIRs calibration models for chemical composition and fatty acid families of raw and freeze-dried beef: a comparison. J. Food Compos. Anal. 83, 103257 https://doi.org/10.1016/J.JFCA.2019.103257.
Goedhart, P.W., 1990. Comparison of multivariate calibration methods for prediction of feeding value by near infrared reflectance spectroscopy. Neth. J. Agric. Sci. 38, 449–460.
Grigorakis, K., 2007. Compositional and organoleptic quality of farmed and wild gilthead sea bream (Sparus aurata) and sea bass (Dicentrarchus labrax) and factors affecting it: a review. Aquaculture 272 (1–4), 55–75. https://doi.org/10.1016/J.AQUACULTURE.2007.04.062.
Han, H., Sha, R., Dai, J., Wang, Z., Mao, J., Cai, M., 2024. Garlic origin traceability and identification based on fusion of multi-Source Heterogeneous spectral information. Foods 13 (7), 1016. https://doi.org/10.3390/foods13071016.
He, J., 2018. From country-of-origin labelling (COOL) to seafood import monitoring program (SIMP): how far can seafood traceability rules go? Mar. Pol. 96, 163–174. https://doi.org/10.1016/J.MARPOL.2018.08.003.
Héberger, K., Csomós, E., Simon-Sarkadi, L., 2003. Principal component and linear discriminant Analyses of free Amino acids and Biogenic Amines in Hungarian Wines. J. Agric. Food Chem. 51 (27), 8055–8060. https://doi.org/10.1021/jf034851c.
Izquierdo, M.S., Obach, A., Arantzamendi, L., Montero, D., Robaina, L., Rosenlund, & G., 2003. Dietary Lipid Sources for Seabream and Seabass: Growth Performance, Tissue Composition and Flesh Quality.
Jin, G., Xu, Y., Cui, C., Zhu, Y., Zong, J., Cai, H., Ning, J., Wei, C., Hou, R., 2022. Rapid identification of the geographic origin of Taiping Houkui green tea using near-infrared spectroscopy combined with a variable selection method. J. Sci. Food Agric. 102 (13), 6123–6130. https://doi.org/10.1002/jsfa.11964.
Jones, S.T., Allison, E.H., Kroetz, K., Ota, Y., Jardine, S.L., 2023. Enrollment, retention, and inclusivity of Marine Stewardship Council (MSC) eco-labelling certifications. Mar. Pol. 155, 105734 https://doi.org/10.1016/J.MARPOL.2023.105734.
Karoui, R., Thomas, E., Dufour, E., 2006. Utilisation of a rapid technique based on front-face fluorescence spectroscopy for differentiating between fresh and frozen-thawed fish fillets. Food Res. Int. 39 (3), 349–355. https://doi.org/10.1016/j.foodres.2005.08.007.
Karoui, R., Lefur, B., Grondin, C., Thomas, E., Demeulemester, C., De Baerdemaeker, J., Guillard, A.S., 2007. Mid-infrared spectroscopy as a new tool for the evaluation of fish freshness. Int. J. Food Sci. Technol. 42 (1), 57–64. https://doi.org/10.1111/j.1365-2621.2006.01208.x.
Kaya, Y., Turan, H., 2010. Comparison of Protein, lipid and fatty acids composition of anchovy (Engraulis encrasicolus L. 1758) during the commercial catching season. J. Muscle Foods 21, 474–483.
Kursa, M.B., Rudnicki, W.R., 2010. Feature selection with the Boruta package. JSS Journal of Statistical Software 36. http://www.jstatsoft.org/.
Kwasi Bannor, R., Arthur, K.K., Oppong, D., Oppong-Kyeremeh, H., 2023. A comprehensive systematic review and bibliometric analysis of food fraud from a global perspective. Journal of Agriculture and Food Research 14. https://doi.org/10.1016/j.jafr.2023.100686.
Lawrence, S., van Ruth, S., Elliott, C., Huisman, W., 2024. Characteristics and situational aspects of seafood fraud: a comparative crime script analysis. Crime Law Soc. Change. https://doi.org/10.1007/s10611-024-10149-7.
Liu, D., Zeng, X.A., Sun, D.W., 2013. NIR spectroscopy and imaging techniques for evaluation of fish quality - a review. Appl. Spectrosc. Rev. 48 (8), 609–628. https://doi.org/10.1080/05704928.2013.775579.
Liu, Y., Ma, D. hong, Wang, X. chang, Liu, L. ping, Fan, Y. xia, Cao, J. xuan, 2015. Prediction of chemical composition and geographical origin traceability of Chinese export tilapia fillets products by near infrared reflectance spectroscopy. LWT - Food Sci. Technol. (Lebensmittel-Wissenschaft -Technol.) 60 (2), 1214–1218. https://doi.org/10.1016/J.LWT.2014.09.009.
Lv, H., Xu, W., You, J., Xiong, S., 2017. Classification of freshwater fish species by linear discriminant analysis based on near infrared reflectance spectroscopy. J. Near Infrared Spectrosc. 25 (1), 54–62. https://doi.org/10.1177/0967033516678801.
Ma, R., Zhao, H., Wang, K., Zhang, R., Hua, Y., Jiang, B., Guo, X., Ruan, Z., Huang, L., 2023. A Novel wavelet packet transform-Fuzzy pattern recognition-based method for Leakage fault Diagnosis of Sail Slewing Hydraulic system. Machines 11 (2), 286. https://doi.org/10.3390/machines11020286.
Madigan, T., Kiermeier, A., de Barros Lopes, M., Cozzolino, D., 2012. The effect of homogenisation and Storage on the near-infrared spectra of Half Shell Pacific oysters (Crassostrea gigas). Food Anal. Methods 5 (5), 995–1002. https://doi.org/10.1007/s12161-011-9329-7.
Majolini, D., Trocino, A., Xiccato, G., Santulli, A., 2009. Near infrared reflectance spectroscopy (NIRS) characterization of European sea bass (Dicentrarchus labrax) from different rearing systems. Ital. J. Anim. Sci. 8 (Suppl. 2), 860–862. https://doi.org/10.4081/ijas.2009.s2.860.
Mazzeo, M.F., Siciliano, R.A., 2016. Proteomics for the authentication of fish species. J. Proteonomics 147, 119–124. https://doi.org/10.1016/j.jprot.2016.03.007.
Mazzeo, M.F., Giulio, B. De, Guerriero, G., Ciarcia, G., Malorni, A., Russo, G.L., Siciliano, R.A., 2008. Fish authentication by MALDI-TOF mass Spectrometry. J. Agric. Food Chem. 56 (23), 11071–11076. https://doi.org/10.1021/jf8021783.
Molkentin, J., Meisel, H., Lehmann, I., Rehbein, H., 2007. Identification of organically farmed Atlantic salmon by analysis of stable isotopes and fatty acids. European Food Research and Technology 224 (5), 535–543. https://doi.org/10.1007/s00217-006-0314-0.
Murray, I., Cowe, I., 2004. Sample preparation. In: Roberts, C.A., Workman Jr., J., Reeves III, J.B. (Eds.), Near-Infrared Spectroscopy in Agriculture, vol 44. John Wiley & Sons, Ltd., pp. 75–112. https://doi.org/10.2134/agronmonogr44.c5
OECD/FAO, 2020. OECD-FAO Agricultural Outlook 2020-2029. FAO, Rome/OECD Publishing, Paris. https://doi.org/10.1787/1112c23b-en.
Olsen, P., Borit, M., 2018. The components of a food traceability system. Trends Food Sci. Technol. 77, 143–149. https://doi.org/10.1016/j.tifs.2018.05.004. Elsevier Ltd.
Ortiz-Somovilla, V., España-España, F., Gaitán-Jurado, A.J., Pérez-Aparicio, J., de Pedro-Sanz, E.J., 2007. Proximate analysis of homogenized and minced mass of pork sausages by NIRS. Food Chem. 101 (3), 1031–1040. https://doi.org/10.1016/j.foodchem.2006.02.058.
Osborne, B.G., Mertens, B., Thompson, M., Fearn, T., 1993. The authentication of basmati rice using near infrared spectroscopy. J. Near Infrared Spectrosc. 1 (2), 77–83. https://doi.org/10.1255/jnirs.8.
Ottavian, M., Facco, P., Fasolato, L., Novelli, E., Mirisola, M., Perini, M., Barolo, M., 2012. Use of near-infrared spectroscopy for fast fraud detection in seafood: application to the authentication of wild European sea bass (Dicentrarchus labrax). J. Agric. Food Chem. 60 (2), 639–648. https://doi.org/10.1021/jf203385e.
Øverland, M., Sørensen, M., Storebakken, T., Penn, M., Krogdahl, Å., Skrede, A., 2009. Pea protein concentrate substituting fish meal or soybean meal in diets for Atlantic salmon (Salmo salar)—Effect on growth performance, nutrient digestibility, carcass composition, gut health, and physical feed quality. Aquaculture 288 (3–4), 305–311. https://doi.org/10.1016/J.AQUACULTURE.2008.12.012.
O'Brien, N., Hulse, C.A., Pfeifer, F., Siesler, H.W., 2013. Near infrared spectroscopic authentication of seafood. J. Near Infrared Spectrosc. 21 (4), 299–305. https://doi.org/10.1255/jnirs.1063.
Pink, J., Naczk, M., Pink, D., 1999. Evaluation of the quality of frozen minced red hake: use of Fourier transform near-infrared spectroscopy. J. Agric. Food Chem. 47 (10), 4280–4284. https://doi.org/10.1021/jf990170z.
Reilly, A., 2018. Overview of Food Fraud in the Fisheries Sector. https://openknowledge.fao.org/items/556a13f7-fc8d-4e3d-aeda-72b34c7e49e6.
Richter, B., Rurik, M., Gurk, S., Kohlbacher, O., Fischer, M., 2019. Food monitoring: screening of the geographical origin of white asparagus using FT-NIR and machine learning. Food Control 104, 318–325. https://doi.org/10.1016/j.foodcont.2019.04.032.
Rinnan, Å., Berg, F. van den, Engelsen, S.B., 2009. Review of the most common pre-processing techniques for near-infrared spectra. TrAC, Trends Anal. Chem. 28 (10), 1201–1222. https://doi.org/10.1016/j.trac.2009.07.007.
Robson, K., Dean, M., Haughey, S., Elliott, C., 2021. A comprehensive review of food fraud terminologies and food fraud mitigation guides. Food Control 120, 107516. https://doi.org/10.1016/j.foodcont.2020.107516.
Romotowska, P.E., Karlsdóttir, M.G., Gudjónsdóttir, M., Kristinsson, H.G., Arason, S., 2016. Seasonal and geographical variation in chemical composition and lipid stability of Atlantic mackerel (Scomber scombrus) caught in Icelandic waters. J. Food Compos. Anal. 49, 9–18. https://doi.org/10.1016/j.jfca.2016.03.005.
Sağlam, N.E., Sağlam, C., 2013. Age, growth and mortality of anchovy Engraulis encrasicolus in the south-eastern region of the Black Sea during the 2010-2011 fishing season. J. Mar. Biol. Assoc. U. K. 93 (8), 2247–2255. https://doi.org/10.1017/S0025315413000611.
Stone, A.P., Stanley, D.W., 1992. Mechanisms of fish muscle gelation. Food Res. Int. 25 (5), 381–388. https://doi.org/10.1016/0963-9969(92)90113-J.
Tomić, M., Lucević, Z., Tomljanović, T., Matulić, D., 2017. Wild-caught versus farmed fish – consumer perception. Ribarstvo, Croatian Journal of Fisheries 75 (2), 41–50. https://doi.org/10.1515/cjf-2017-0007.
Trocino, A., Xiccato, G., Majolini, D., Tazzoli, M., Bertotto, D., Pascoli, F., Palazzi, R., 2012a. Assessing the quality of organic and conventionally-farmed European sea bass (Dicentrarchus labrax). Food Chem. 131 (2), 427–433. https://doi.org/10.1016/j.foodchem.2011.08.082.
Trocino, A., Xiccato, G., Majolini, D., Tazzoli, M., Bertotto, D., Pascoli, F., Palazzi, R., 2012b. Assessing the quality of organic and conventionally-farmed European sea bass (Dicentrarchus labrax). Food Chem. 131 (2), 427–433. https://doi.org/10.1016/j.foodchem.2011.08.082.
Uddin, M., Okazaki, E., 2004. Classification of fresh and frozen-thawed fish by near-infrared spectroscopy. J. Food Sci. 69 (8) https://doi.org/10.1111/j.1750-3841.2004.tb18015.x.
Varrà, M.O., Ghidini, S., Ianieri, A., Zanardi, E., 2021a. Near infrared spectral fingerprinting: a tool against origin-related fraud in the sector of processed anchovies. Food Control 123. https://doi.org/10.1016/j.foodcont.2020.107778.
Varrà, M.O., Husáková, L., Patočka, J., Ghidini, S., Zanardi, E., 2021b. Multi-element signature of cuttlefish and its potential for the discrimination of different geographical provenances and traceability. Food Chem. 356, 129687 https://doi.org/10.1016/j.foodchem.2021.129687.
Wang, Z., Wu, Q., Kamruzzaman, M., 2022. Portable NIR spectroscopy and PLS based variable selection for adulteration detection in quinoa flour. Food Control 138. https://doi.org/10.1016/j.foodcont.2022.108970.
Webster, C.D., T I U ~, L.G., Morgan, A.M., Gannam, A., 1999. Effect of partial and total Replacement of fish meal on growth and body composition of Sunshine bass Morone chrysops x M. saxatilis fed practical diets. J. World Aquacult. Soc. 30 (Issue 4).
Weeranantanaphan, J., Downey, G., Allen, P., Sun, D.W., 2011. A review of near infrared spectroscopy in muscle food analysis: 2005-2010. J. Near Infrared Spectrosc. 19 (2), 61–104. https://doi.org/10.1255/jnirs.924.

Xiccato, G., Trocino, A., Tulli, F., Tibaldi, E., 2004a. Prediction of chemical composition and origin identification of european sea bass (Dicentrarchus labrax L.) by near infrared reflectance spectroscopy (NIRS). Food Chem. 86 (2), 275–281. https://doi.org/10.1016/j.foodchem.2003.09.026.

Xiccato, G., Trocino, A., Tulli, F., Tibaldi, E., 2004b. Prediction of chemical composition and origin identification of european sea bass (Dicentrarchus labrax L.) by near infrared reflectance spectroscopy (NIRS). Food Chem. 86 (2), 275–281. https://doi.org/10.1016/j.foodchem.2003.09.026.

Yalcin, K., Hulay, T., 2008. Comparison of protein, lipid and fatty acids composition of anchovy. J. Muscle Foods 21, 474–483, 2010.

Yang, H., He, S., Feng, Q., Xia, S., Zhou, Q., Wu, Z., Zhang, Y., 2024. Navigating the depths of seafood authentication: technologies, regulations, and future prospects. Measurement: Food 14, 100165. https://doi.org/10.1016/j.meafoo.2024.100165.

Yi, S., Ji, Y., Guo, Z., Zhu, J., Xu, Y., Li, X., Li, J., 2020. Gel properties and flavor characteristics of blended anchovy (Engraulis japonicus) mince and silver carp (Hypophthalmichthys molitrix) surimi. https://doi.org/10.1039/c9ra10847e.

Yu, W., 2021. Applications of near infrared spectroscopy for fish and fish products quality: a review. IOP Conf. Ser. Earth Environ. Sci. 657 (1) https://doi.org/10.1088/1755-1315/657/1/012115.

Yun, Y.-H., 2022. Wavelength selection methods. In: Chemometric Methods in Analytical Spectroscopy Technology. Springer Nature, Singapore, pp. 169–207. https://doi.org/10.1007/978-981-19-1625-0_5.

Yun, Y.H., Li, H.D., Deng, B.C., Cao, D.S., 2019. An overview of variable selection methods in multivariate analysis of near-infrared spectra. TrAC, Trends Anal. Chem. 113, 102–115. https://doi.org/10.1016/j.trac.2019.01.018. Elsevier B.V.

Zlatanos, S., Laskaridis, K., 2007. Seasonal variation in the fatty acid composition of three Mediterranean fish - sardine (Sardina pilchardus), anchovy (Engraulis encrasicholus) and picarel (Spicara smaris). Food Chem. 103 (3), 725–728. https://doi.org/10.1016/j.foodchem.2006.09.013.