Article

Rapid Identification of Rainbow Trout Adulteration in Atlantic Salmon by Raman Spectroscopy Combined with Machine Learning

Zeling Chen [1], Ting Wu [2], Cheng Xiang [1], Xiaoyan Xu [1,*] and Xingguo Tian [1,3,*]

1   College of Food, South China Agricultural University, Guangzhou 510642, China
2   School of Information Science and Technology, Zhongkai University of Agriculture and Engineering, Guangzhou 510225, China
3   New Rural Development Research Institute, South China Agricultural University, Guangzhou 510225, China
*   Correspondence: yanzixu_2001@scau.edu.cn (X.X.); xingguot@scau.edu.cn (X.T.);
    Tel.: +86-020-8528-0266 (X.X.); +86-020-8528-3433 (X.T.)

check for updates

Abstract: This study intends to evaluate the utilization potential of the combined Raman spectroscopy and machine learning approach to quickly identify the rainbow trout adulteration in Atlantic salmon. The adulterated samples contained various concentrations (0–100% *w/w* at 10% intervals) of rainbow trout mixed into Atlantic salmon. Spectral preprocessing methods, such as first derivative, second derivative, multiple scattering correction (MSC), and standard normal variate, were employed. Unsupervised algorithms, such as recursive feature elimination, genetic algorithm (GA), and simulated annealing, and supervised K-means clustering (KM) algorithm were used for selecting important spectral bands to reduce the spectral complexity and improve the model stability. Finally, the performances of various machine learning models, including linear regression, nonlinear regression, regression tree, and rule-based models, were verified and compared. The results denoted that the developed GA–KM–Cubist machine learning model achieved satisfactory results based on MSC preprocessing. The determination coefficient (R^2) and root mean square error of prediction sets (RMSEP) in the test sets were 0.87 and 10.93, respectively. These results indicate that Raman spectroscopy can be used as an effective Atlantic salmon adulteration identification method; further, the developed model can be used for quantitatively analyzing the rainbow trout adulteration in Atlantic salmon.

Keywords: Atlantic salmon; adulteration; Raman spectroscopy; machine learning

1. Introduction

Atlantic salmon (Salmo salar) has attracted consumer interest because of its unique taste and rich nutritional value. Even though there is a huge demand for Atlantic salmon in the Chinese market, imported Atlantic salmon is often in short supply. Under these circumstances, rainbow trout (Oncorhynchus mykiss) is often used to imitate or adulterate Atlantic salmon meat or products. Rainbow trout is considerably less expensive than Atlantic salmon but looks similar, making it difficult for consumers to distinguish between the two. Adulterated Atlantic salmon meat not only infringes the legitimate rights and interests of consumers but also causes serious food safety problems, leading to widespread concern among consumers, producers, retailers, and food regulatory agencies.

Therefore, an accurate and expedient method is required for identifying Atlantic salmon adulteration. Traditional meat identification methods have mainly used enzyme-linked immunosorbent assays, deoxyribonucleic acid (DNA) [1–6], proteome [7–9], and triacylglycerol-based analytical techniques [10]. Although these methods have proved to be accurate, they exhibit long processing

times and complicated technologies, making them unsuitable for rapid on-site detection under market supervision environments.

Among the emerging detection technologies, infrared spectroscopy, hyperspectral imaging, and Raman spectroscopy have been proved to be valuable for characterizing the chemical structure of the meat molecules [11,12]. Further, these techniques have been regularly employed to assess the quality [13–15], safety [16–18], and classification [19,20] of the meat products [21–23]. Among them, Raman spectroscopy has been considered to be a promising meat detection tool because it is a non-invasive, fast, and convenient technique that requires little to no sample pretreatment [20,24–28]. Recently, especially after the "horse meat storm" incident [29], Raman spectroscopy has gradually been used for identifying meat adulteration [29–39]. For example, Boyaci et al., used Raman spectroscopy to scan beef and horse fat samples and used the principle component analysis method for performing data processing and modeling [29]. The results denoted that this model could distinguish among different horse meat contents (25%, 50%, and 75%) in beef samples. Zhou Yaling et al. [32] used Raman spectroscopy combined with chemometrics to establish a discriminant model based on support vector regression. The results showed that the model was able to accurately identify the adulterated beef stuffing with different proportions of chicken meat.

In previous studies, Raman spectroscopy has often been used to identify sausages [38], fish [23], beef [24], and other common meat. In these examples, the full potential of Raman spectroscopy could not be evaluated because the minimum detection limit of the employed model was often large. In addition, Raman spectroscopy has mostly been used for qualitative discrimination, which does not lend itself to predicting and quantifing the adulteration level in meat. Furthermore, to the best of our knowledge, Raman spectroscopy has not been used to identify Atlantic salmon adulteration. Therefore, we were interested in evaluating the ability of Raman spectroscopy to detect low adulteration levels in Atlantic salmon meat. This study discloses our efforts in developing a convenient, high-sensitivity, low-cost Atlantic salmon meat adulteration detection technology based on Raman spectroscopy combined with a stable, fast, and accurate machine learning model.

The objectives of this study are as follows: (1) to evaluate the effectiveness of Raman spectroscopy in identifying the adulteration of Atlantic salmon meat and to develop a portable Raman spectroscopy and machine learning method to quickly identify rainbow trout adulteration in Atlantic salmon; (2) to compare the four pretreatment methods [first derivation (FD), second derivation (SD), multiple scattering correction (MSC), and standard normal variate (SNV)] for obtaining the best pretreatment method; (3) to combine the supervised (RFE, GA, and SA) and unsupervised (KM) dimensionality reduction and variable selection methods to find the best characteristic wavelengths; and (4) to test the performance of mainstream machine learning models, such as linear regression, nonlinear regression, regression trees, and rule-based methods, in case of Atlantic salmon adulteration.

2. Results

2.1. Raman Spectral Analyses

As the Raman spectra of lower than 500 cm^{-1} and higher than 2000 cm^{-1} obtained in various regions exhibited a large amount of spectral noise and no obvious absorption peaks, bands with rich spectral information (500–2000 cm^{-1}) were selected for conducting spectral analysis. The Raman spectra of Atlantic salmon and rainbow trout fat within these regions are depicted in Figure 1.
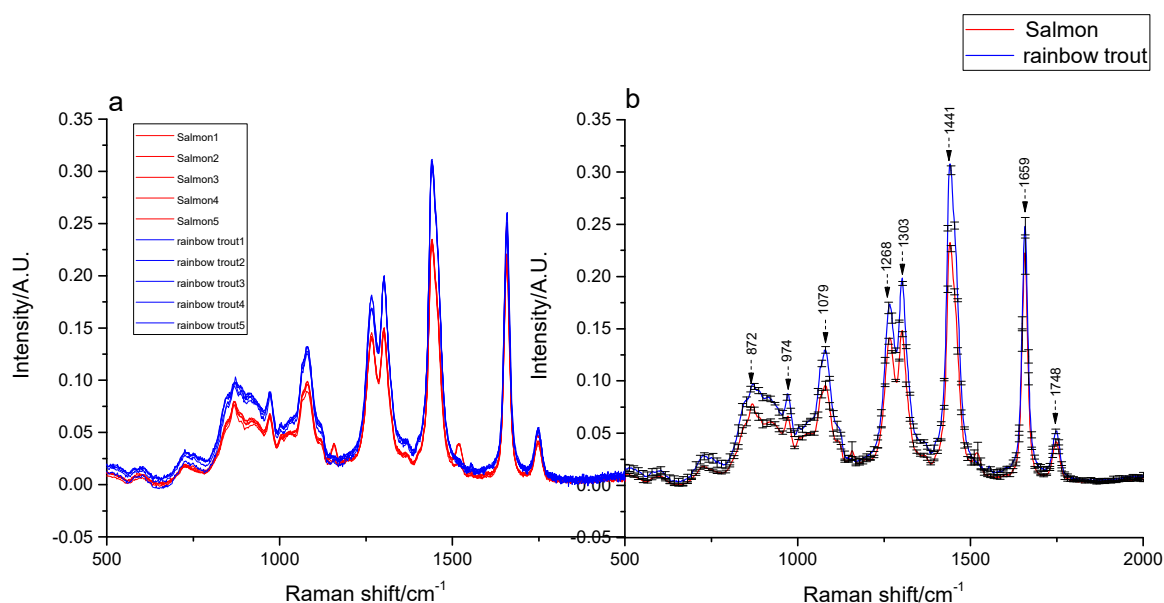
Figure 1. The Raman spectra of fat in Atlantic salmon and rainbow trout. (**a**) The five spectra of salmon and rainbow trout after baseline correction and multiple scattering correction (MSC). (**b**) The mean and standard deviation spectra of salmon and rainbow trout.

To eliminate the effects of baseline drift and scattering distortion on the spectra and to compare the spectral differences among the two fish, preprocessing methods, including baseline correction and MSC, were conducted, as depicted in Figure 1a. Some significant differences were observed between the intensities of the spectral absorption peaks of the two fish. The mean and standard deviation spectra of salmon and rainbow trout have been depicted in Figure 1b. Eight overlapping peaks with different intensities (with the exception of 1748 cm^{-1}) were identified in two spectra; the peaks associated with the rainbow trout were characterized as the stronger of the two. These peak intensity differences formed the basis for distinguishing between rainbow trout and Atlantic salmon. The component functional groups corresponding to the Raman peaks of the two fish fats were analyzed and have been provided in Table 1.

Table 1. Raman spectral distribution of the Atlantic salmon fat.

Band/cm^{-1}	Vibration Mode	Functional Groups	Intensity
1748	ν(C=O)	Ester (RC=OOR)	Weak
1659	ν(C=C)	Unsaturated band (cis RHC=CHR)	Strong
1441	$\delta\gamma$(C–H)	Methylene (CH$_2$)	Strong
1303	$\delta\tau$(C–H)	Methylene (CH$_2$)	Medium
1268	δIP(=C–H)	Non-conjugated cis (RHC=CHR)	Medium
1079	ν(C–C)	–(CH$_2$)$_n$–	Medium
974	δ(=C–H)	Trans RHC=CHR	Medium
872	ν(C–C)	–(CH$_2$)$_n$–	Medium

As presented in Table 1, the peak at 1748 cm^{-1} is weak in strength and is attributed to the C=O ester stretching mode (C=O). The peak at 1659 cm^{-1} corresponds to a Z-alkene, ν(C=C), in the fatty acid chain, whereas the strong peak at 1441 cm^{-1} corresponds to the C–H bending stretching modes. The peak at 1303 cm^{-1} is attributed to the CH$_2$ twisting modes (C–H), the peak at 1268 cm^{-1} is due to the Z conformation stretching modes (=CH) from the unsaturated fatty acids, the peaks at 1079 cm^{-1} and 872 cm^{-1} are due to gauche C–C stretching vibrations (C–C), and the peak at 974 cm^{-1} is caused by the bending vibration of trans (=CH). These peaks exhibited medium strength. The eight

characteristic peaks have also been reported as common features in the Raman spectra of edible oils below 2000 cm^{-1} [40–42], and the peak positions and intensities of different fatty acids have been observed to be slightly different [43].

The Raman spectra in case of Atlantic salmon with different proportions of rainbow trout adulteration are depicted in Figure 2.
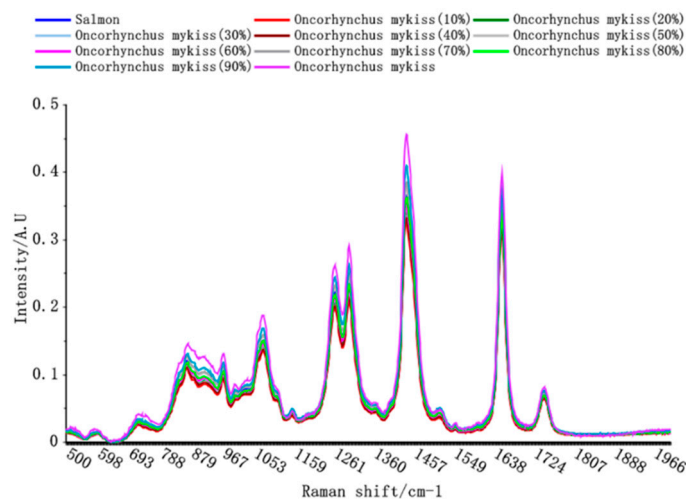

Figure 2. The Raman spectra observed in case of different proportions of rainbow trout adulteration in Atlantic salmon.

The Raman spectra of the adulterated Atlantic salmon fat were very similar, with characteristic absorption peaks being observed at 1748, 1659, 1441, 1303, 1268, 1079, 974, and 872 cm^{-1}. The Raman peak intensity increased with increasing amounts of rainbow trout meat, enabling the Raman spectra of the different Atlantic salmon meat samples to be distinguished. These differences in absorption peaks provided the basis for further model development.

2.2. Preprocessing Analysis

The spectra of the Atlantic salmon samples obtained using different pretreatment methods, such as baseline correction, MSC, SNV, FD, and SD, are depicted in Figure 3a–f.
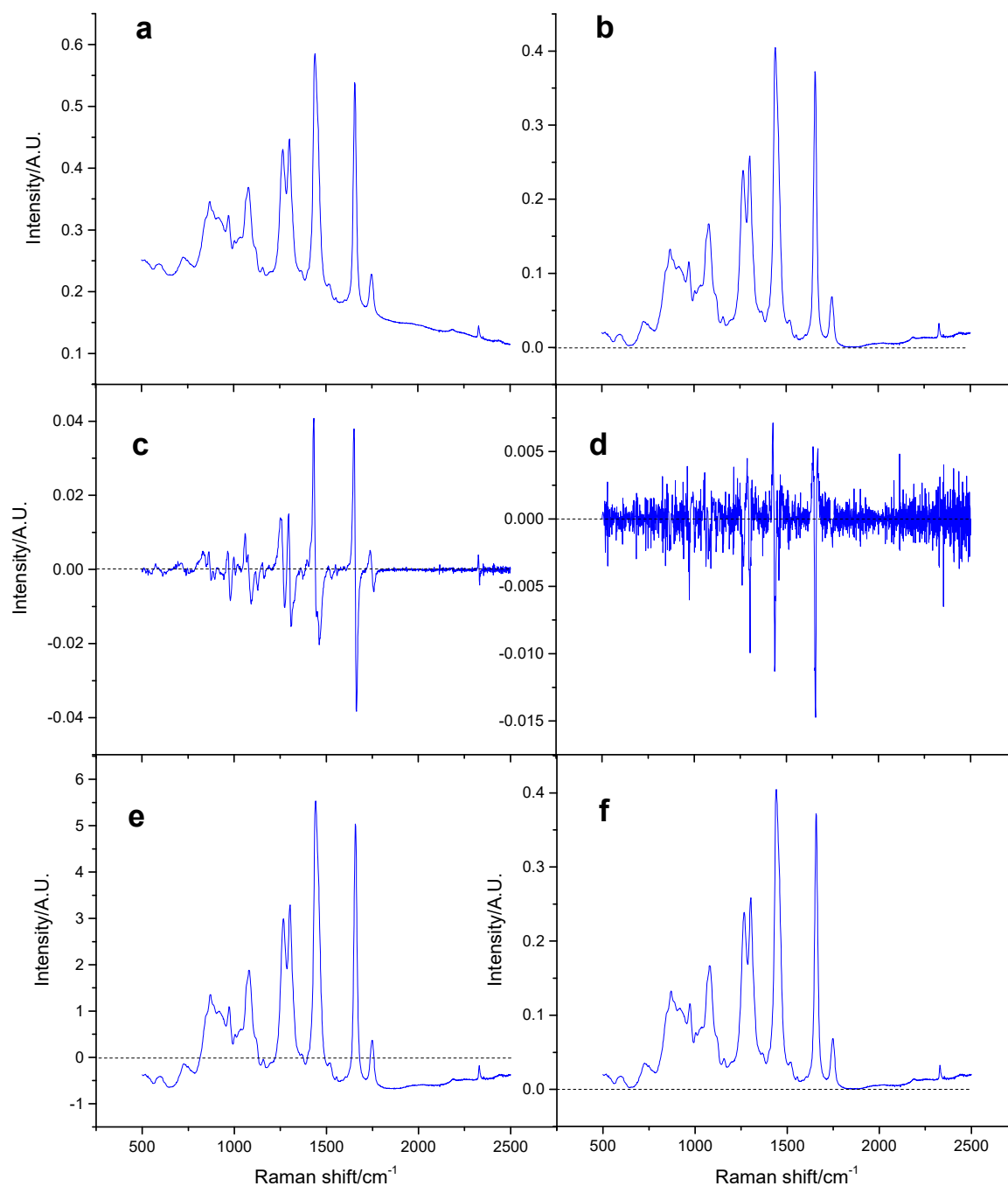
Figure 3. The Raman spectra of samples obtained using different pretreatments: (**a**) the original spectrum; (**b**) the spectrum after baseline fitting; (**c**) the spectrum after applying the first derivative; (**d**) the spectrum after applying the second derivative; (**e**) the spectrum after applying standard normal variate (SNV); (**f**) the spectrum after applying MSC.

To compare the effects of different pretreatment methods, a partial least squares regression (PLSR) model was used to evaluate the results of the pretreatment methods. The experiment conducted without a pretreatment step was used as a reference. The results are presented in Table 2.

Table 2. The partial least squares regression (PLSR) modeling results for four preprocessing methods.

Pretreatment Methods	Ncomp	Calibration Sets		Test Sets		
		RMSE (%)	R^2	RMSEP (%)	R^2P	MAE
NONE	10	14.79	0.79	17.27	0.70	13.19
FD	10	21.38	0.58	23.10	0.48	18.18
SD	10	29.26	0.19	30.15	0.12	25.11
SNV	10	13.66	0.82	13.28	0.81	10.49
MSC	9	13.68	0.82	13.32	0.81	10.57

MAE (Mean Square Error) denotes the average absolute error.

The model performance was the highest when the principal component number of PLSR modeling was 10. While observing the test sets, RMSEP and R^2 could reach values of 17.27 and 0.70 in case of the usage of raw spectral modeling, respectively. In contrast, the RMSEP values obtained using FD and SD were 23.10 and 30.15 and R^2 values were 0.48 and 0.12, respectively. These results demonstrated that FD and SD modeling were less effective than raw spectral modeling. The FD and SD methods amplified the noise in the spectra, which can explain the poor performance of these models. Further studies using MSC and SNV revealed that both the methods could achieve better results when compared with the original spectra, i.e., the resulting RMSEP was smaller and R^2 was larger. The two methods also eliminated the scattering effect that negatively influenced the spectral data.

By comparing the two methods in cases in which they performed similarly, the number of principal components required for SNV modeling was observed to be larger than that required for MSC. Therefore, the MSC modeling performance could be considered to be slightly better than SNV. Further, while comparing different machine learning modeling methods in this study, MSC was the only method employed for spectral preprocessing.

2.3. Important Spectral Band Selection

In this study, different supervised methods (RFE, GA, and SA) and unsupervised methods (KM) were combined to reduce the dimension of spectral wavelengths, and optimal bands were selected (Table 3).

Table 3. Three feature wavelength selection methods based on the PLSR modeling results.

Dimension Reduction Methods	Number of Wavelengths	Calibration Sets		Test Sets		
		RMSE (%)	R^2	RMSEP (%)	R^2P	MAE
NONE	882	13.68	0.82	13.32	0.81	10.57
RFE–KM	75	14.47	0.79	14.93	0.77	12.24
GA–KM	431	14.36	0.80	13.34	0.81	10.69
SA–KM	322	14.55	0.79	13.84	0.80	11.11

MAE denotes the average absolute error.

The results presented in Table 3 denote that the performances of the three methods were relatively similar and not distinct from those of the full-spectra model. However, the number of required spectral bands was considerably reduced in comparison with that in the full-spectra method, improving both the efficiency and stability of the model. Among them, GA–KM was considered to be the best method; the required wavelengths of the model were considerably reduced from 882 to 431, and this method exhibited improved prediction performance (R^2 = 0.81, RMSEP = 13.34%).

2.4. Results of the Cubist Model

After applying the MSC pretreatment method and selecting the optimal feature bands by GA–KM, the Cubist model was established to identify adulterated Atlantic salmon samples containing different

proportions of rainbow trout. To optimize the Cubist model, different sizes and numbers of committees and instances in the model were examined. The cross-validation curve of the Cubist model is presented in Figure 4.
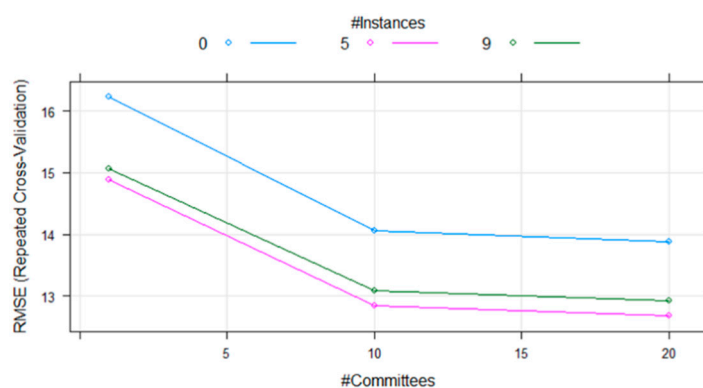

Figure 4. The cross-validated RMSE (Root Mean Square Error) curve for different commit sizes and instance numbers.

Regardless of the number of instances, the error significantly decreased as the commits gradually increased to 10. While increasing the number of commits from 10 to 20, the error only slightly decreased. When the Cubist model used 20 commits and 5 instances, the error was the smallest and the modeling effect was the largest. Furthermore, when the number of instances was too low or too high, the performance of the model would decrease. The adulteration ratio of Atlantic salmon was predicted based on the aforementioned parameters, and the results are presented in Figure 5.
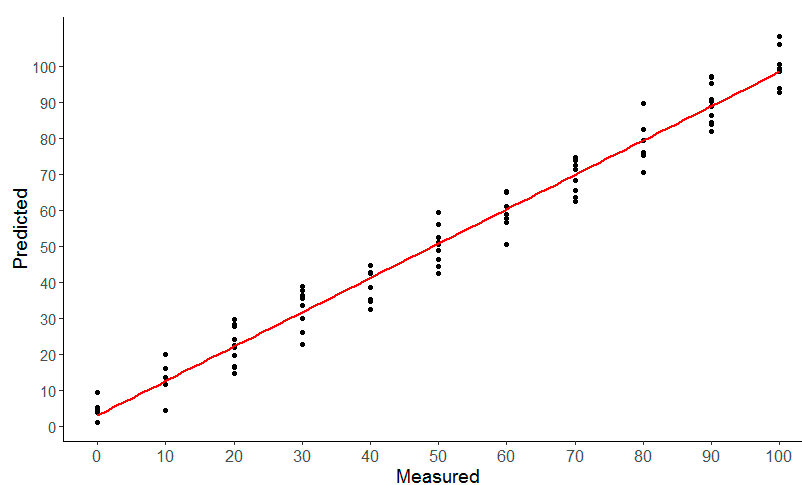

Figure 5. The predicted and true values of Atlantic salmon meat adulteration ratios based on the Cubist model in test sets.

The RMSE in the calibration sets was 12.67, and R^2 was 0.84; these values were 10.93 and 0.87, respectively, in the test sets. The experimental data exhibited a high degree of agreement with the predicted data, and the modeling performance was good, suggesting that this technique could be used to quickly identify Atlantic salmon adulteration.

3. Discussion

To denote the advantages of the Cubist algorithm with respect to model prediction, 13 types of machine learning methods and the PLSR method were used to model the selected spectral bands. The results are summarized in Table 4. For the test sets, the Cubist model was observed to have

the smallest RMSEP (10.98) followed by PLSR (13.34). This result indicated that the modeling performance of the linear regression model (Cubist and PLS) may be better than those of other models. One explanation for this result is that adulteration using the rainbow trout followed a linear relation in the Raman spectra; as the adulteration ratio increased, the peak strength of the Raman spectra increased. The RMSEP of the Cubist method was much smaller than that of the remaining models, which could have been due to the Cubist method being a rule-based model. The model tree leaf node was a linear regression model, and the regression equation modeling on the node was more flexible and more accurate when compared with the other regression models [44]. Furthermore, complex linear regression models did not yield better performances. The RMSEP of Glmboost, Enet, Ridge, and Rqlasso were not as suitable as that of the PLS model. The modeling performances obtained using nonlinear models, such as random forests and neural networks, were even worse than those of the complex linear models. This indicated that the more complex linear or nonlinear models were not globally optimal when the number of samples was not sufficiently large and that they were prone to over-fitting, leading to a decrease in the accuracy of the model. In summary, the modeling performance of the linear model was generally better than those of the other models, and there was a linear relation between the rainbow trout adulteration and peak intensity of the Raman spectra. The Cubist model exhibited the best modeling performance and was combined with Raman spectroscopy to develop a new technique for identifying Atlantic salmon adulteration.

Table 4. Performance comparison of different machine learning models.

Models	RMSE (%)		R^2		MAE	
	Calibration Sets	Test Sets	Calibration Sets	Test Sets	Calibration Sets	Test Sets
PLS	14.36	13.34	0.80	0.81	11.18	10.69
Ridge	17.09	14.84	0.74	0.78	13.39	11.81
Enet	15.23	14.38	0.77	0.78	12.11	11.60
Rqlasso	15.72	14.92	0.76	0.77	12.46	11.94
Earth	16.30	16.84	0.74	0.71	12.93	13.14
Kknn	16.44	16.02	0.75	0.74	12.79	12.38
ParRF	15.91	14.87	0.77	0.79	12.92	11.91
Qrf	15.66	14.81	0.76	0.77	11.99	10.99
Rf	15.92	14.99	0.77	0.78	12.95	11.98
Ctree	21.74	22.71	0.55	0.48	17.09	16.95
Cubist	12.67	10.93	0.84	0.87	9.78	8.37
Glmboost	15.20	14.38	0.77	0.78	12.17	11.57
XgbTree	29.67	29.22	0.33	0.30	22.70	22.86
Msaene	15.33	14.39	0.77	0.78	12.37	11.69

MAE was the average absolute error.

4. Materials and Methods

4.1. Sample Preparation

Different amounts of rainbow trout were added to Atlantic salmon to create adulterated samples. To expand the sample diversity and improve the credibility of the experimental results, Atlantic salmon was obtained from different batches, at different times, and from different regions (Denmark, Scotland, Chile, and Norway). To ensure the authenticity of the samples, import-certified Atlantic salmon stores were selected. Danish Atlantic salmon meat was purchased from Hippo Fresh Food in Guangzhou, China; Chilean Atlantic salmon meat was purchased from Jingdong Supermarket in Guangzhou, China; Scottish Atlantic salmon meat was purchased from Haidi Wang Fresh Seafood in Shanghai, China; and Norwegian Atlantic salmon meat was purchased from the Yuesheng official store in Shenzhen, China. Rainbow trout was also purchased from different regions and stores. Qinghai

rainbow trout was purchased from the Tmall Longyangxia store in Gonghe, China; Gansu rainbow trout was purchased from the Tmall Shangzhi store in Lanzhou, China; Shandong rainbow trout was purchased from the Laoshan ecological farm in the Taobao store in Qingdao, China; and Liaoning rainbow trout was purchased from the Tmall supermarket in Benxi, China.

The Atlantic salmon meat obtained from four regions (Denmark, Chile, Scotland, and Norway) and the rainbow trout obtained from Qinghai, Gansu, Shandong, and Liaoning were all crushed using a grinder in KRUPS, Germany. Further, the Atlantic salmon and rainbow trout were mixed according to the following weight percentages: 0%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, and 100% (*w/w*). Each of the 176 mixed samples weighed 50 g, and 2–3 parallel samples were prepared for each gradient, affording 516 prepared samples. Each sample was subsequently homogenized using a meat grinder and centrifuged at a speed of 10,000 rpm and a centrifugation time of 5 min. After centrifugation, the upper layer of an oily material was pipetted into a chemical reaction plate for conducting the Raman spectroscopy measurements.

4.2. Raman Spectral Data Measurements

A portable Raman spectrometer (FoodDefend RM, Thermo Fisher, Waltham, MA, USA) was used to collect the Raman spectra. The Raman system was equipped with a laser excited at 785 nm. When scanning, the laser power source was set to 250 mW, and the spectral range was 250–2500 cm^{-1}. The spectral resolution was 7 cm^{-1}, the exposure time was 5 ms, the scanning delay was 60 s, and the operating temperature was set to 30 °C to achieve the optimal Raman peaks. To eliminate noise and ensure data repeatability, each sample was scanned thrice; the average values were used as the sample spectra.

4.3. Spectral Pretreatment

Before modeling, spectral preprocessing was required to reduce noise or to eliminate random and systematic changes in the data [26]. The four preprocessing methods (FD, SD, MSC, and SNV) had different effects on the spectral data. For example, FD was used to remove the baselines, SD was used to remove the baselines and linear trends [45], and MSC and SNV were typically used to eliminate the unwanted scattering effects [46]. In this study, PLSR was used to model the same spectral data, and the effects of four pretreatment methods were evaluated for conducting the Atlantic salmon contamination analyses.

4.4. Analytical Methods

4.4.1. Spectral Band Selection Methods

Selection of important spectral bands was critical to reducing the high dimensionality of the spectral data and increasing the processing speed [16]. Some researchers used unsupervised methods, such as clustering algorithms, for conducting feature selection [47,48]. As the clustering method has been shown to result in the selection of irrelevant features [47–49], irrelevant features were deleted before feature clustering. Some supervisory feature selection methods, such as RFE [50], GA [51], and SA [52], have been shown to be effective approaches for removing unrelated features. Furthermore, we attempted to combine the supervised and unsupervised methods for performing dimensionality reduction and variable selection. Firstly, the RFE, GA, and SA algorithms were used to remove the uncorrelated wavelengths and select the relevant characteristic bands. Then, KM [53–55] was used to optimize the feature wavelengths, and PLSR was used to monitor the modeling error of these selections. The optimal feature wavelengths could be obtained based on these data.

4.4.2. Modeling Methods

Certain machine learning models have proven to be effective for identifying food adulteration [56–58]. In this study, the applicability of several machine learning models for predicting

Atlantic salmon adulteration were evaluated. The R language was used for modeling, and a total of 14 mainstream machine learning algorithms, including linear regression models, nonlinear regression models, tree-based, and rule-based models, were used for training and testing. Linear regression models included PLSR, the boosted generalized linear model (Glmboost) [59], Elasticnet regression (Enet) [60], ridge regression (Ridge) [61], quantile regression with LASSO penalty (Rqlasso) [62], multi-step adaptive MCP-net (Msaene) [63], quantile random forest (Qrf) [64], parallel random forest (parRF) [65], random forest (Rf) [66], k-nearest neighbors (Kknn) [67], and multivariate adaptive regression spline (Earth) [68]. The tree-based models include conditional inference tree (ctree) [69] and extreme gradient boosting (xgbTree) [70]. The Cubist model (Cubist) [71] was the only rule-based model considered in this study. The selected wavelength sets and adulteration levels were used as the input and output variables, respectively, for the model, and the input and output data and other conditions were observed to be consistent while evaluating and comparing the model performance. A random sampling method was selected to divide the data sets into two subsets: training data (75%) and test data (25%). When modeling, 10-fold cross-validation was used, and five training times were repeated and averaged for the final results. The aforementioned process was implemented using the R language Caret package.

4.5. Model Evaluation

The determination coefficient (R^2), root mean square error of calibration sets (RMSEC), and root mean square error of test sets (RMSEP) were used to evaluate the performance of the regression model. The definitions were as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^{N} (\hat{y}_i - y_i)^2}{\sum_{i=1}^{N} (\hat{y}_i - \overline{y}_i)^2} \tag{1}$$

$$\text{RMSEC or RMSEP} = \sqrt{\frac{\sum_{i=1}^{N} (\hat{y}_i - y_i)^2}{N}} \tag{2}$$

where $\hat{y}_i$ is the predicted adulteration level of the ith sample, y_i is the true adulterated level of the ith sample, $\overline{y}_i$ is the average of y_i, and N is the number of samples.

4.6. Software

All the Raman spectral data pretreatments were performed using TheUnscrambler X14.1 software (CAMO, Oslo, Norway). All the calculations were performed using the R program (version 3.5.1). The Kknn package (version 1.3.1) was used for variable clustering, and the Caret package (version 6.0-82) was used for performing feature wavelength selections and machine learning modeling.

5. Conclusions

In this study, we evaluated the ability of a combined Raman spectroscopy and machine learning approach to rapidly detect the adulteration of Atlantic salmon using rainbow trout. A linear relation can be observed between the adulteration ratio of Atlantic salmon and the Raman spectra intensity. In this experiment, MSC was shown to be a better pretreatment method when compared with FD, SD, and SNV. GA was used to delete the irrelevant wavelengths, and KM was used to optimize the spectral bands. The Cubist method achieved the highest performance while modeling the spectra. Thus, the machine learning model developed in this study based on the MSC–GA–KM–Cubist method is an effective tool for quickly identifying the adulteration of Atlantic salmon meat.

Author Contributions: Conceptualization, X.T.; methodology, X.X.; software, X.T.; validation, T.W., and C.X.; formal analysis, C.X.; data curation, Z.C.; writing—original draft preparation, T.W.; writing—review and editing, X.X.; project administration, Z.C.; funding acquisition, X.T.

References

1. Ali, M.E.; Hashim, U.; Mustafa, S.; Man, Y.C.; Dhahi, T.S.; Kashif, M.; Uddin, M.K.; Hamid, S.A. Analysis of pork adulteration in commercial meatballs targeting porcine-specific mitochondrial cytochrome b gene by TaqMan probe real-time polymerase chain reaction. *Meat Sci.* **2012**, *91*, 454–459. [CrossRef] [PubMed]

2. Haider, N.; Nabulsi, I.; Al-Safadi, B. Identification of meat species by PCR-RFLP of the mitochondrial COI gene. *Meat Sci.* **2012**, *90*, 490–493. [CrossRef] [PubMed]

3. Mane, B.G.; Mendiratta, S.K.; Tiwari, A.K. Beef specific polymerase chain reaction assay for authentication of meat and meat products. *Food Control* **2012**, *28*, 246–249. [CrossRef]

4. Sakaridis, I.; Ganopoulos, I.; Argiriou, A.; Tsaftaris, A. A fast and accurate method for controlling the correct labeling of products containing buffalo meat using High Resolution Melting (HRM) analysis. *Meat Sci.* **2013**, *94*, 84–88. [CrossRef] [PubMed]

5. Soares, S.; Amaral, J.S.; Oliveira, M.B.P.; Mafra, I. A SYBR Green real-time PCR assay to detect and quantify pork meat in processed poultry meat products. *Meat Sci.* **2013**, *94*, 115–120. [CrossRef]

6. Zhang, C. Semi-nested multiplex PCR enhanced method sensitivity of species detection in further-processed meats. *Food Control* **2013**, *31*, 326–330. [CrossRef]

7. Mamani-Linares, L.W.; Gallo, C.; Alomar, D. Identification of cattle, llama and horse meat by near infrared reflectance or transflectance spectroscopy. *Meat Sci.* **2012**, *90*, 378–385. [CrossRef]

8. Kim, G.; Seo, J.; Yum, H.; Jeong, J.; Yang, H. Protein markers for discrimination of meat species in raw beef, pork and poultry and their mixtures. *Food Chem.* **2017**, *217*, 163–170. [CrossRef]

9. Alamprese, C.; Casale, M.; Sinelli, N.; Lanteri, S.; Casiraghi, E. Detection of minced beef adulteration with turkey meat by UV–vis, NIR and MIR spectroscopy. *LWT-Food Sci. Technol.* **2013**, *53*, 225–232. [CrossRef]

10. Ropodi, A.I.; Pavlidis, D.E.; Mohareb, F.; Panagou, E.Z.; Nychas, G. Multispectral image analysis approach to detect adulteration of beef and pork in raw meats. *Food Res. Int.* **2015**, *67*, 12–18. [CrossRef]

11. Marquardt, B.J.; Wold, J.P. Raman analysis of fish: A potential method for rapid quality screening. *LWT-Food Sci. Technol.* **2004**, *37*, 1–8. [CrossRef]

12. Wong, H.; Choi, S.; Phillips, D.L.; Ma, C. Raman spectroscopic study of deamidated food proteins. *Food Chem.* **2009**, *113*, 363–370. [CrossRef]

13. Prieto, N.; Lopez-Campos, O.; Aalhus, J.L.; Dugan, M.; Juarez, M.; Uttaro, B. Use of near infrared spectroscopy for estimating meat chemical composition, quality traits and fatty acid content from cattle fed sunflower or flaxseed. *Meat Sci.* **2014**, *98*, 279–288. [CrossRef] [PubMed]

14. Ripoll, G.; Lobón, S.; Joy, M. Use of visible and near infrared reflectance spectra to predict lipid peroxidation of light lamb meat and discriminate dam's feeding systems. *Meat Sci.* **2018**, *143*, 24–29. [CrossRef] [PubMed]

15. Moran, L.; Andres, S.; Allen, P.; Moloney, A.P. Visible and near infrared spectroscopy as an authentication tool: Preliminary investigation of the prediction of the ageing time of beef steaks. *Meat Sci.* **2018**, *142*, 52–58. [CrossRef] [PubMed]

16. Kamruzzaman, M.; Makino, Y.; Oshita, S. Rapid and non-destructive detection of chicken adulteration in minced beef using visible near-infrared hyperspectral imaging and machine learning. *J. Food Eng.* **2016**, *170*, 8–15. [CrossRef]

17. Alamprese, C.; Amigo, J.M.; Casiraghi, E.; Engelsen, S.B. Identification and quantification of turkey meat adulteration in fresh, frozen-thawed and cooked minced beef by FT-NIR spectroscopy and chemometrics. *Meat Sci.* **2016**, *121*, 175–181. [CrossRef]

18. Kamruzzaman, M.; ElMasry, G.; Sun, D.; Allen, P. Application of NIR hyperspectral imaging for discrimination of lamb muscles. *J. Food Eng.* **2011**, *104*, 332–340. [CrossRef]

19. Wang, W.; Peng, Y.; Sun, H.; Zheng, X.; Wei, W. Spectral detection techniques for non-destructively monitoring the quality, safety, and classification of fresh red meat. *Food Anal. Method* **2018**, *11*, 2707–2730. [CrossRef]

20. Dixit, Y.; Casado Gavalda, M.P.; Cama Moncunill, R.; Cama Moncunill, X.; Markiewicz Keszycka, M.; Cullen, P.J.; Sullivan, C. Developments and challenges in online NIR spectroscopy for meat processing. *Compr Rev. Food Sci. Food Saf.* **2017**, *16*, 1172–1187. [CrossRef]

21. Boyacı, I.H.; Temiz, H.T.; Uysal, R.S.; Velioğlu, H.M.; Yadegari, R.J.; Rishkan, M.M. A novel method for discrimination of beef and horsemeat using Raman spectroscopy. *Food Chem.* **2014**, *148*, 37–41. [CrossRef]

22. Zając, A.; Hanuza, J.; Dymińska, L. Raman spectroscopy in determination of horse meat content in the mixture with other meats. *Food Chem.* **2014**, *156*, 333–338. [CrossRef] [PubMed]

23. Velioğlu, H.M.; Temiz, H.T.; Boyaci, I.H. Differentiation of fresh and frozen-thawed fish samples using Raman spectroscopy coupled with chemometric analysis. *Food Chem.* **2015**, *172*, 283–290. [CrossRef] [PubMed]

24. Zhao, M.; Downey, G.; O Donnell, C.P. Dispersive Raman spectroscopy and multivariate data analysis to detect offal adulteration of thawed beefburgers. *J. Agric. Food Chem.* **2015**, *63*, 1433–1441. [CrossRef] [PubMed]

25. Yaling, Z. Rapid discrimination of minced beef adulterated with chicken using Raman spectroscopy. *Meat Res.* **2018**, *32*, 26–29.

26. Devos, O.; Downey, G.; Duponchel, L. Simultaneous data pre-processing and SVM classification model selection based on a parallel genetic algorithm applied to spectroscopic data of olive oils. *Food Chem.* **2014**, *148*, 124–130. [CrossRef] [PubMed]

27. Jović, O. Durbin-Watson partial least-squares regression applied to MIR data on adulteration with edible oils of different origins. *Food Chem.* **2016**, *213*, 791–798. [CrossRef]

28. Li, Y.; Fang, T.; Zhu, S.; Huang, F.; Chen, Z.; Wang, Y. Detection of olive oil adulteration with waste cooking oil via Raman spectroscopy combined with iPLS and SiPLS. *Spectrochim. Acta Part A* **2018**, *189*, 37–43. [CrossRef]

29. Fowler, S.M.; Schmidt, H.; van de Ven, R.; Hopkins, D.L. Preliminary investigation of the use of Raman spectroscopy to predict meat and eating quality traits of beef loins. *Meat Sci.* **2018**, *138*, 53–58. [CrossRef]

30. Fowler, S.M.; Schmidt, H.; Scheier, R.; Hopkins, D.L. Raman spectroscopy for predicting meat quality traits. In *Advanced Technologies for Meat Processing*, 2nd ed.; CRC Press: Boca Raton, FL, USA, 2017; pp. 83–112.

31. Nian, Y.; Zhao, M.; O'Donnell, C.P.; Downey, G.; Kerry, J.P.; Allen, P. Assessment of physico-chemical traits related to eating quality of young dairy bull beef at different ageing times using Raman spectroscopy and chemometrics. *Food Res. Int.* **2017**, *99*, 778–789. [CrossRef]

32. Bauer, A.; Scheier, R.; Eberle, T.; Schmidt, H. Assessment of tenderness of aged bovine gluteus medius muscles using Raman spectroscopy. *Meat Sci.* **2016**, *115*, 27–33. [CrossRef] [PubMed]

33. Zhao, M.; Nian, Y.; Allen, P.; Downey, G.; Kerry, J.P.; O'Donnell, C.P. Application of Raman spectroscopy and chemometric techniques to assess sensory characteristics of young dairy bull beef. *Food Res. Int.* **2018**, *107*, 27–40. [CrossRef] [PubMed]

34. Zhao, M.; Nian, Y.; Allen, P.; Downey, G.; Kerry, J.P.; O Donnell, C.P. Performances of full cross-validation partial least squares regression models developed using Raman spectral data for the prediction of bull beef sensory attributes. *Data Brief* **2018**, *19*, 1355–1360. [CrossRef] [PubMed]

35. Yaseen, T.; Sun, D.; Cheng, J. Raman imaging for food quality and safety evaluation: Fundamentals and applications. *Trends Food Sci. Technol.* **2017**, *62*, 177–189. [CrossRef]

36. Rohman, A.; Che Man, Y.B.; Hashim, P.; Ismail, A. FTIR spectroscopy combined with chemometrics for analysis of lard adulteration in some vegetable oils. *Cyta-J. Food* **2011**, *9*, 96–101. [CrossRef]

37. Jaswir, I.; Mirghani, M.E.S.; Hassan, T.H.; Said, M.Z.M. Determination of lard in mixture of body fats of mutton and cow by Fourier transform infrared spectroscopy. *J. Oleo Sci.* **2003**, *52*, 633–638. [CrossRef]

38. Tomasevic, I.; Nedeljković, A.; Stanišić, N.; Pudja, P. Authenticity assessment of cooked emulsified sausages using Raman spectroscopy and chemometrics. *J. Meat Prod. Meat Process.* **2016**, *3*, 70–73.

39. De Biasio, M.; Stampfer, P.; Leitner, R.; Huck, C.W.; Wiedemair, V.; Balthasar, D. Micro-Raman spectroscopy for meat type detection. In *Next-Generation Spectroscopic Technologies VIII*; International Society for Optics and Photonics: Bellingham, WA, USA, 2015; p. 94821J.

40. Beattie, J.R.; Bell, S.E.; Moss, B.W. A critical evaluation of Raman spectroscopy for the analysis of lipids: Fatty acid methyl esters. *Lipids* **2004**, *39*, 407–419. [CrossRef]

41. Mahesar, S.A.; Sherazi, S.; Khaskheli, A.R.; Kandhro, A.A. Analytical approaches for the assessment of free fatty acids in oils and fats. *Anal. Method* **2014**, *6*, 4956–4963. [CrossRef]

42. Czamara, K.; Majzner, K.; Pacia, M.Z.; Kochan, K.; Kaczor, A.; Baranska, M. Raman spectroscopy of lipids: A review. *J. Raman Spectrosc.* **2015**, *46*, 4–20. [CrossRef]

43. Muik, B.; Lendl, B.; Molina-Díaz, A.; Ayora-Cañada, M.J. Direct monitoring of lipid oxidation in edible oils by Fourier transform Raman spectroscopy. *Chem. Phys. Lipids* **2005**, *134*, 173–182. [CrossRef] [PubMed]

44. Quinlan, J.R. Learning with continuous classes. In Proceedings of the 5th Australian Joint Conference on Artificial Intelligence, Hobart, Tasmania, 16–18 November 1992; pp. 343–348.

45. Rinnan, Å.; Van Den Berg, F.; Engelsen, S.B. Review of the most common pre-processing techniques for near-infrared spectra. *TrAC Trends Anal. Chem.* **2009**, *28*, 1201–1222. [CrossRef]

46. Vidal, M.; Amigo, J.M. Pre-processing of hyperspectral images. Essential steps before image analysis. *Chemom. Intell. Lab. Syst.* **2012**, *117*, 138–148. [CrossRef]

47. Burger, J.; Gowen, A. Data handling in hyperspectral image analysis. *Chemom. Intell. Lab. Syst.* **2011**, *108*, 13–22. [CrossRef]

48. Agrawal, R.; Gehrke, J.; Gunopulos, D.; Raghavan, P. *Automatic Subspace Clustering of High Dimensional Data for Data Mining Applications*; ACM: New York, NY, USA, 1998; Volume 27.

49. Mirkin, B. Concept learning and feature selection based on square-error clustering. *Mach. Learn.* **1999**, *35*, 25–39. [CrossRef]

50. Guyon, I.; Weston, J.; Barnhill, S.; Vapnik, V. Gene selection for cancer classification using support vector machines. *Mach. Learn.* **2002**, *46*, 389–422. [CrossRef]

51. Goldberg, D.E.; Holland, J.H. Genetic algorithms and machine learning. *Mach. Learn.* **1988**, *3*, 95–99. [CrossRef]

52. Bohachevsky, I.O.; Johnson, M.E.; Stein, M.L. Generalized simulated annealing for function optimization. *Technometrics* **1986**, *28*, 209–217. [CrossRef]

53. Ienco, D.; Meo, R. Exploration and reduction of the feature space by hierarchical clustering. In Proceedings of the 2008 SIAM International Conference on Data Mining, Atlanta, GA, USA, 24–26 April 2008; pp. 577–587.

54. Witten, D.M.; Tibshirani, R. A framework for feature selection in clustering. *J. Am. Stat. Assoc.* **2010**, *105*, 713–726. [CrossRef]

55. Liu, H.; Wu, X.; Zhang, S. Feature selection using hierarchical feature clustering. In Proceedings of the 20th ACM International Conference on Information and Knowledge Management, Glasgow, UK, 24–28 October 2011; pp. 979–984.

56. Rady, A.; Adedeji, A. Assessing different processed meats for adulterants using visible-near-infrared spectroscopy. *Meat Sci.* **2018**, *136*, 59–67. [CrossRef]

57. Tripathy, S.; Reddy, M.S.; Vanjari, S.R.K.; Jana, S.; Singh, S.G. A step towards miniaturized milk adulteration detection system: Smartphone-based accurate pH sensing using electrospun halochromic nanofibers. *Food Anal. Method* **2019**, *12*, 612–624. [CrossRef]

58. Large, J.; Kemsley, E.K.; Wellner, N.; Goodall, I.; Bagnall, A. Detecting Forged Alcohol Non-invasively Through Vibrational Spectroscopy and Machine Learning. In Proceedings of the Pacific-Asia Conference on Knowledge Discovery and Data Mining 2018, Melbourne, Australia, 3–6 June 2018; Springer: Cham, Switzerland, 2018; pp. 298–309.

59. Hofner, B.; Mayr, A.; Robinzonov, N.; Schmid, M. Model-based boosting in R: A hands-on tutorial using the R package mboost. *Comput. Stat.* **2014**, *29*, 3–35. [CrossRef]

60. Simon, N.; Friedman, J.; Hastie, T.; Tibshirani, R. Regularization paths for Cox's proportional hazards model via coordinate descent. *J. Stat. Soft* **2011**, *39*, 1–13. [CrossRef] [PubMed]

61. Chadeau-Hyam, M.; Hoggart, C.J.; O'Reilly, P.F.; Whittaker, J.C.; De Iorio, M.; Balding, D.J. Simulation of realistic sequence-level data in populations and ascertained samples. *BMC Bioinform.* **2008**, *9*, 364. [CrossRef] [PubMed]

62. Koenker, R.; Mizera, I. Convex optimization in R. *J. Stat. Soft* **2014**, *60*, 1–23. [CrossRef]

63. Xiao, N.; Qing-Song, X. Multi-step adaptive elastic-net: Reducing false positives in high-dimensional variable selection. *J. Stat. Comput. Simul.* **2015**, *85*, 3755–3765. [CrossRef]

64. Meinshausen, N. Quantile regression forests. *J. Mach. Learn. Res.* **2006**, *7*, 983–999.

65. Scholkopf, B.; Smola, A.; Williamson, R.C.; Bartlett, P. New support vector algorithms. *Neural Comput.* **2000**, *12*, 1207–1245. [CrossRef] [PubMed]

66. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]

67. Hechenbichler, K.; Schliep, K.P. *Weighted k-Nearest-Neighbor Techniques and Ordinal Classification*; Discussion Paper; Ludwig-Maximilians University Munich: Munich, Germany, 2004; pp. 86–399.

68. Leathwick, J.R.; Rowe, D.; Richardson, J.; Elith, J.; Hastie, T. Using multivariate adaptive regression splines to predict the distributions of New Zealand's freshwater diadromous fish. *Freshw. Biol.* **2005**, *50*, 2034–2052. [CrossRef]

69. Strobl, C.; Malley, J.; Tutz, G. An introduction to recursive partitioning: Rationale, application, and characteristics of classification and regression trees, bagging, and random forests. *Psychol. Methods* **2009**, *14*, 323–348. [CrossRef] [PubMed]

70. Lundberg, S.M.; Lee, S.I. A unified approach to interpreting model predictions. In Proceedings of the Advances in Neural Information Processing Systems 2017, Long Beach, CA, USA, 4–9 December 2017; pp. 4765–4774.

71. Wang, Y.; Witten, I. Inducing model trees for continuous classes. In Proceedings of the Ninth European Conference on Machine Learning, Prague, Czech Republic, 23–25 April 1997; pp. 128–137.

Sample Availability: Samples of the compounds are available from the authors.