ANIMAL SCIENCE AND FOOD TECHNOLOGY
Journal homepage: https://animalscience.com.ua/en
Animal Science and Food Technology, 16(3), 70-85
Received 09.05.2025 Revised 02.09.2025 Accepted 23.09.2025

UDC 535.232:639.3
DOI: 10.31548/animal.3.2025.70

Fish spectral classification based on principal component analysis

Li Xu
Postgraduate Student, Lecturer
Bengbu University
233030, Caoshan Road, Bengbu, Anhui Province, China
https://orcid.org/0009-0000-3762-1714

Alina Menchynska*
PhD in Technical Sciences, Associate Professor
National University of Life and Environmental Sciences of Ukraine
03041, 15 Heroiv Oborony Str., Kyiv, Ukraine
https://orcid.org/0000-0001-8593-3325

Olena Ochkolyas
PhD in Technical Sciences, Associate Professor
National University of Life and Environmental Sciences of Ukraine
03041, 15 Heroiv Oborony Str., Kyiv, Ukraine
http://orcid.org/0000-0002-8483-578X

Mykola Nikolaenko
Doctor of Philosophy, Associate Professor
National University of Life and Environmental Sciences of Ukraine
03041, 15 Heroiv Oborony Str., Kyiv, Ukraine
https://orcid.org/0000-0003-2213-4985

Abstract. Spectral detection technology can be used to identify different types of fish flesh and determine the presence of ingredients in samples, which helps to control the authenticity and conformity of products. The aim of this study was to develop and compare the effectiveness of machine learning methods for the classification of different types of fish based on hyperspectral

---

Suggested Citation:
Xu, L., Menchynska, A., Ochkolyas, O., & Nikolaenko, M. (2025). Fish spectral classification based on principal component analysis. *Animal Science and Food Technology*, 16(3), 70-85. doi: 110.31548/animal.3.2025.70.

**Corresponding author*

data. In this paper, hyperspectral technology was used to obtain spectral data of 328 sampling points in the 400-1,000 nm band of four different fish, and the spectral curves were preprocessed to obtain more accurate and efficient spectral curves. The study systematically compared the preprocessing methods of Savitzky-Golay smoothing and multivariate scatter correction (MSC), and combined the principal component analysis (PCA) method to optimise the k-nearest neighbour (KNN) and support vector machine (SVM) classification models, in order to provide a more effective solution for fish classification. The principal component analysis method was used to reduce the dimensionality of the spectral data. The principal components after dimensionality reduction were compared with the full components, and the training set and test set were randomly generated. The SVM algorithm and the KNN algorithm were used in Python to predict and determine the accuracy of the model. It was shown that the average accuracy of spectral data after preprocessing and dimensionality reduction in the classification of SVM models can reach 97.25%, and the average classification accuracy of KNN models can reach 97.48%, which indicates that different types of fish flesh can be classified. Both models showed that overall, the preprocessing method that combines MSC and PCA is better than the single preprocessing method, highlighting the importance of PCA in removing redundant information and preserving key classification characteristics. When comparing the model performance under the same preprocessing conditions, the accuracy of SVM is higher than that of KNN, indicating that the SVM has greater adaptability to spectral data classification. Through the development and application of spectral detection technology, technological progress and innovation in food testing can be promoted, guiding the entire industry to a safer and more reliable future

Keywords: spectrum; Savitzky-Golay method; multiplicative scatter correction; k-nearest neighbours model; support vector machine algorithm

Introduction

As people's awareness of food safety continues to grow, food quality and authenticity have become the focal points of global concern. The food industry faces issues such as adulteration and counterfeiting. Criminals might seek profits by adding inexpensive or harmful substances to products. Thus, it is essential to develop reliable detection methods to prevent and identify such counterfeiting behaviours. However, traditional food quality detection methods are destructive, cumbersome to operate, and require trained professionals. With the rapid development and integration of image processing, spectral analysis, and food intelligence technologies, the application of hyperspectral imaging technology in food detection has attracted growing attention (Wu *et al.*, 2025). The development and advancement of spectral technologies, particularly the increasing applications of near-infrared spectroscopy, Raman spectroscopy, fluorescence spectroscopy, and other techniques, offer faster, more accurate, and non-destructive approaches for food analysis and testing. L. He *et al.* (2022) employed ultra-high performance liquid chromatography – tandem mass spectrometry (UPLC – MS/MS) to identify serum albumin target peptides in beef and horse meat. The results demonstrated that this method exhibited good linearity and excellent analytical performance within a certain concentration range. B. Li *et al.* (2024) utilised support vector machine, random forest, and LSTM algorithms to establish a classification model by fusing hyperspectral reflectance and transmittance spectral data, and combined with partial least squares regression to quantitatively analyse the concentration of beef adulterants. They found that the detection accuracy of chicken, duck, and pork adulterated in beef could be improved. The scientific work by I. Palamarchuk *et al.* (2024) used the laser-induced breakdown spectroscopy (LIBS) method for spectroscopic assessment and quantitative analysis of the trace element composition of plant additives to meat products.

Spectral detection technology can analyse numerous samples in a short time, enhancing detection and production efficiencies. Through the development and application of spectral detection technology, technological progress and innovation in food testing can be promoted, guiding the entire industry towards a safer and more reliable future (Melnychuk *et al.*, 2016). Moreover, the importance of standardised quality metrics in food production was emphasised by N. Kim (2021), who proposed a generalised quality indicator method for assessing various objects, including food products. This method integrates multiple quality metrics into a cohesive model, enabling precise and rapid evaluation of food quality, which is critical for ensuring consistency and reliability in non-destructive spectral detection applications.

Fish, being a common food item, its quality and authenticity are of great significance to consumers' health and trust. Fish spectral detection plays a crucial role in ensuring food safety and quality. By analysing the spectral characteristics of samples to identify their components and quality attributes, consumers' health and rights can be safeguarded (Wang *et al.*, 2019). T. Xu (2020) synthesised three gold nanorods (Au NRs) with different aspect ratios using the seed-growth method and, in combination with SERS technology, detected industrial dye drug residues in four types of fish flesh. Z. Chen *et al.* (2021) used these non-destructive methods to assess the freshness of pearl gentian grouper under different storage conditions. G. Tsagkatakis *et al.* (2020) proposed a new approach to fish freshness assessment using information encoded in the spectral profile obtained with a spectral imaging camera. Experimental evaluation on individuals of the Sparidae family (*Boops* sp.) confirmed that the proposed approach is a valid methodology, offering both accuracy and ease of application. Spectroscopic technology can achieve fast and non-destructive fish classification by analysing the absorption or reflection characteristics of biological tissues to electromagnetic waves (Shi *et al.*, 2022; Yuan *et al.*, 2025). However, the original spectrum is easily interfered with by factors such as instrument noise and sample surface scattering, and it is necessary to combine dimensionality reduction techniques such as principal component analysis (PCA) to improve classification accuracy.

The aim of this paper was to provide a more efficient study of fish classification based on spectral analysis. The research objectives included a systematic comparison of preprocessing methods, Savitzky-Golay (SG) smoothing and multivariate scatter correction (MSC), and a combination of PCA to optimise KNN and SVM classification models.

Materials and Methods

All experimental procedures were conducted at the Hyperspectral Imaging Laboratory, School of Food Science and Engineering, Bengbu University (Bengbu, Anhui Province, China) during March 2025. The experimental studies were conducted in compliance with the requirements of the European Convention for the Protection of Vertebrate Animals used for Experimental and other Scientific Purposes (1986), as well as the Law of Ukraine No. 3447-IV "On Protection of Animals from Cruelty" (2006).

Experimental materials. The fresh fish samples for this experiment were purchased from the Bengbu Fishery. The fish species included Snakehead, Grass carp, Bighead carp, Redfin culter, with 10 individuals of each type.

The length of each fish ranged from 25 to 40 cm. Samples were prepared from the mid-section of freshly slaughtered and cleaned fish, with 2 cm thick cross-sectional slices divided into four groups. Each container was labelled with sample identification and preparation date. Specimens were subsequently refrigerated at 3°C for 24 hours to stabilise physicochemical properties whilst preventing interference from trace microbial activity. Following this stabilisation period, samples were arranged on a motorised stage for spectral acquisition in a controlled-environment chamber maintained at 15°C to minimise wavelength-specific intensity variations caused by thermal effects. After processing to eliminate the interference of irrelevant factors, the fish samples were detected using a hyperspectral imager. Hyperspectral data of different types of fish in the wavelength band from 400 nm to 1,000 nm were collected. Following preprocessing to eliminate extraneous factors, hyperspectral imaging was performed using a GaiaField-V10 system (Jiangsu Shuangli Hepu Technology Co., Ltd) as illustrated in Figure 1. The apparatus comprised: a hyperspectral imager with internal push-broom scanning mode, a CCD camera (696×658 pixels), four 200W halogen lamps, a black background panel, a computer-controlled translation stage, and a light-tight enclosure. Samples were positioned centrally on the motorised stage within the enclosure. Operating within a 400-1,000 nm spectral range, the system acquired data at 0.7 nm/s with 328 output wavelengths. Key specifications included: spectral resolution – 2.8 nm, focal length – 25 mm, spectral sampling interval – 2.34 nm. During acquisition via SpecView software, parameters were set to: exposure time – 30 ms, gain – 1. Sample-to-camera distance was maintained at 55 cm to ensure image clarity. Radiometric calibration was performed using a 99% reflectance standard white reference prior to daily measurements to ensure a linear reflectance response.

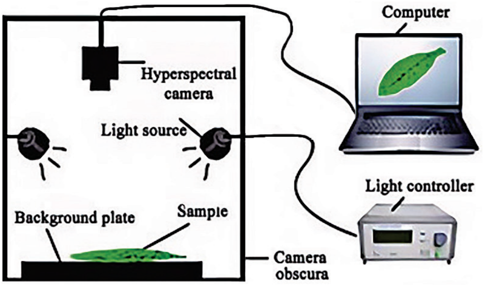

Figure 1. Hyperspectral detection device
Source: developed by the authors

Data processing methods. Hyperspectral image data of fish specimens were acquired in this study. Spectral curves of samples were extracted using ENVI (Environment for Visualising Images) software, yielding spectral information across 328 bands. Subsequent preprocessing of raw spectral data and curves was conducted within the Python 3.9.21 software environment. To mitigate noise and enhance the signal-to-noise ratio, Multiplicative Scatter Correction (MSC) and Savitzky-Golay (SG) smoothing algorithms were applied. The preprocessed data were partitioned into two subsets: one subset underwent dimensionality reduction via Principal Component Analysis (PCA) to efficiently represent features in a lower-dimensional space, while the other subset retained its original dimensionality. This process generated four distinct datasets (MSC+PCA, MSC, SG+PCA, SG), which were subsequently reimported into the Python environment. During model training, each dataset was split into training and testing sets at a 4:1 ratio, employing a cross-validation strategy. Two classification algorithms – Support Vector Machine (SVM) and K-Nearest Neighbours (KNN) – were utilised to classify spectral data from four fish categories. To meet computational and data storage requirements, the software environment was configured as follows: Development platform: PyCharm 2023.1.3 Professional Edition; Deep learning framework: PyTorch 1.12.0+cu113. Key dependencies: numpy

1.23, pandas 1.5.3, matplotlib 3.7.1, scikit-learn 1.2.2, torchvision 0.15.2, tqdm 4.65.0, tensorboard 2.12.2. Models were trained on the training sets, with final performance evaluated through confusion matrices and accuracy metrics on the test sets.

In the field of hyperspectral data preprocessing, two methods are widely used: Savitzky-Golay smoothing and multidimensional scattering correction, which are used to suppress noise and eliminate scattering interference, respectively. SG smoothing is a local weighted average algorithm based on polynomial approximation. By shifting an odd-length window on the spectral curve, a polynomial regression is performed on the data in the window, and the centre value of the approximation curve is used to replace the original value. In this study, the window size was set to 5 points during SG data processing, and the polynomial order was set to 3rd order. During preprocessing, a least-squares polynomial approximation was performed on the data points in each window, and the smoothing value of the centre point of the window was calculated. The edge of the data was processed by mirror expansion. In addition, to eliminate the scattering effect caused by the physical surface characteristics of fish samples (such as particle size and roughness), the multidimensional scattering correction (MSC) method was used to achieve spectral baseline correction.

Fish hyperspectral data usually contains hundreds of continuous bands (400-1,000 nm), which is a typical high-dimensional data feature. Principal Component Analysis (PCA) is a well-established linear dimensionality reduction method. It maps high-dimensional data to a low-dimensional space through orthogonal transformation. It effectively reduces dimensionality while retaining the main information of the data, and it has become an important preprocessing technology for fish hyperspectral data analysis. In this study, the PCA method was used to reduce the dimensionality of the data. First, the principal components were extracted by calculating the covariance matrix, and then the principal components with a cumulative variance contribution exceeding 95% were selected. The data after dimensionality reduction were then trained by KNN and SVM algorithms. This study established eight comparative experimental groups comprising the following specific combinations: SG+PCA+SVM, SG+PCA+KNN, MSC+PCA+SVM, MSC+PCA+KNN, SG+SVM, SG+KNN, MSC+SVM, and MSC+KNN. Among these, four experimental sets required initial dimensionality reduction via PCA prior to input of the data into SVM and KNN models for training. After the 8 experimental groups were preprocessed by SG and MSC, 4 of them were subjected to PCA dimensionality reduction and then substituted into the SVM and KNN models for training. The other 4 groups were not subjected to PCA dimensionality reduction and were directly substituted into the model training. Finally, the accuracy rates obtained from the training of the 8 model combinations were compared and analysed.

Results and Discussion

The original spectral curves of the four fish samples, the spectral curves after SG convolution smoothing, and the spectral curves after multivariate scattering correction (MSC) processing are shown in Figures 2, 3, and 4, respectively. Through comparative analysis, it can be observed that the Savitzky-Golay convolution smoothing algorithm effectively suppressed high-frequency random noise while fully retaining the position and intensity information of the characteristic absorption peak. This method achieved data smoothing through local polynomial fitting. Under the 5-point sliding window parameter setting, it successfully eliminated signal fluctuations caused by instrument noise and environmental interference. Compared with SG processing, the multivariate scatter correction technology demonstrated a more centralised optimisation effect in spectral preprocessing. By regressing each sample spectrum against the reference spectrum, MSC processing not only effectively eliminated the baseline drift caused by particle scattering effects but also centralised the spectral data. Experimental data indicated that both preprocessing methods can enhance the quality of spectral data.

Prior to model training, the fish hyperspectral data were subjected to both noise-reduction preprocessing and dimensionality reduction using PCA. Figures 5 and 6 show the PCA two-dimensional projection results of fish spectral data after SG and MSC preprocessing. In both figures, the horizontal axis PC1 accounted for more than 86.0% of the variance. In Figure 6, the vertical axis PC2 accounted for 6.2% of the variance, and in Figure 5, the vertical axis PC2 accounted for 10.1% of the variance. The cumulative variance contribution of PC1 and PC2 was high, reaching over 94%. PC1 was the main dimension for differentiating fish samples. Redfin culter showed the highest degree of distinction from other species. However, Bighead carp, Grass carp, and Snakehead had relatively weak distinctions because of the partial overlap of their spectral features. Further refinement was required by combining more dimensions or analysis methods.

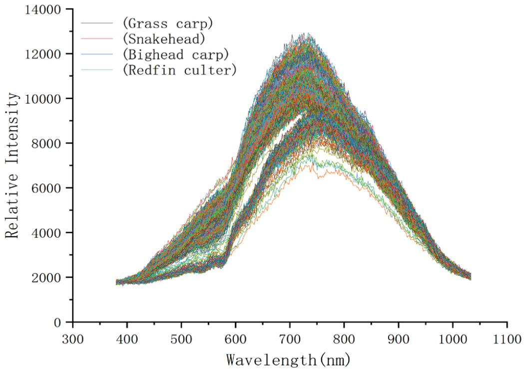

Figure 2. Raw spectral profiles of four fish species

Source: developed by the authors

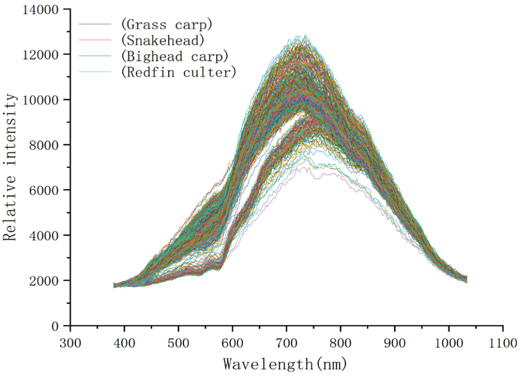

Figure 3. Spectrum after SG treatment

Source: developed by the authors

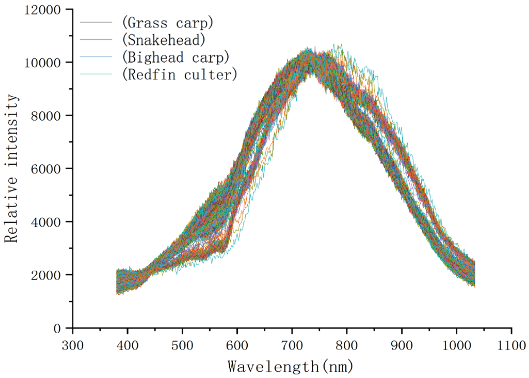

Figure 4. Spectrum after MSC treatment

Source: developed by the authors

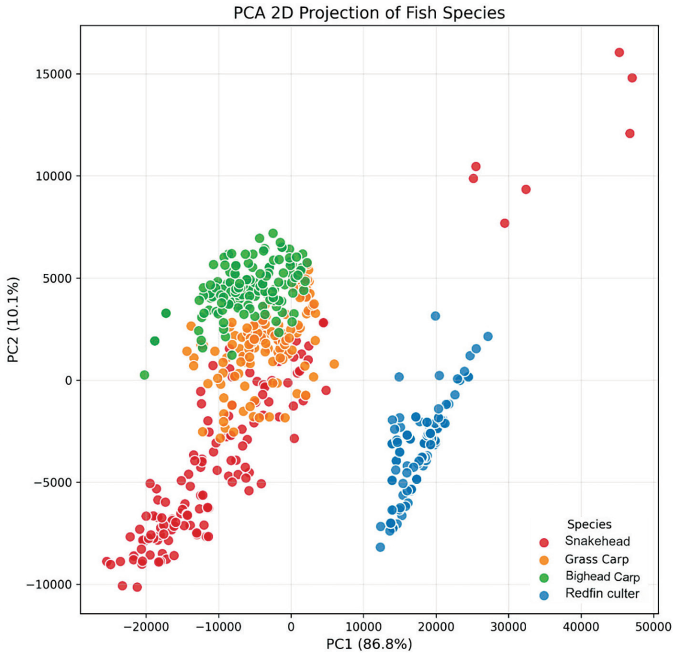

Figure 5. SG-PCA two-dimensional projection of the fish flesh spectrum
Source: developed by the authors

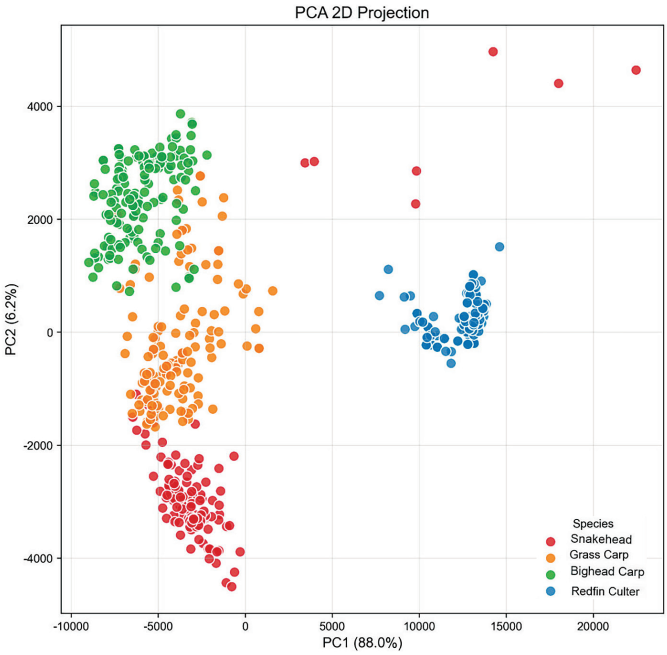

Figure 6. MSC-PCA two-dimensional projection of the fish flesh spectrum
Source: developed by the authors

Table 1 and Table 2 present the cumulative contribution variance statistics after performing PCA dimension analysis. They illustrate the contribution rates of the first 5 principal components when the fish spectral data were processed using SG-PCA and MSC-PCA.

Table 1. *Contribution rates of the top 5 principal components of SG-PCA*

Principal component number	Percentage of variance (%)	Cumulative (%)
PCA1	86.80	86.80
PCA2	10.10	96.90
PCA3	2.50	99.40
PCA4	0.30	99.70
PCA5	0.20	99.90

Source: developed by the authors

Table 2. *Contribution rates of the top 5 principal components of MSC-PCA*

Principal component number	Percentage of variance (%)	Cumulative (%)
PCA1	88.00	88.00
PCA2	6.20	94.20
PCA3	0.90	95.10
PCA4	0.40	95.20
PCA5	0.10	95.30

Source: developed by the authors

As is evident from Table 1, in the SG-PCA combination, the variance contribution rate of the first principal component, PCA1, reached as high as 86.80%. The cumulative contribution rate of the first two principal components is 96.90%, and that of the first 5 principal components was as high as 99.9%. This implied that most of the variance information could be accounted for by the first two principal components, and the first 5 principal components nearly covered all the information. In contrast, in the MSC-PCA combination, the variance contribution rate of PCA1 was 88.00%, which was higher than that of PCA1 in the SG-PCA combination. However, the cumulative contribution rate of the first two principal components was 94.20%, and the cumulative contribution rate of the first 5 principal components was 95.30%, both of which were lower than those in the SG-PCA combination. This indicated that, under the MSC-PCA combination, the explanatory power of the first 5 principal components for the data was slightly weaker, yet the explanatory power of the first principal component was stronger.

From the principal component contribution rates of the two preprocessing methods, it was observed that when the preprocessed spectral data underwent PCA analysis, more than 95% of the original information was retained while the data dimensionality was reduced to fewer than ten. This demonstrated that the combination of preprocessing and PCA effectively eliminated redundant spectral information. As a result, it not only decreased the amount of training data and shortened the training time but also improved the analysis efficiency. The training results are presented in Table 3 and Table 4.

Table 3. *SG classification model training results*

Principal component number	Accuracy	Average accuracy (with or without PCA)	Average accuracy
MSC	97.33%	96.77%	97.25%
SG	96.20%		
MSC-PCA	97.86%	97.73%	
SG-PCA	97.59%		

Source: developed by the authors

Table 4. *KNN classification model training results*

Principal component number	Accuracy	Average accuracy (with or without PCA)	Average accuracy
MSC	97.59%	96.97%	97.48%
SG	96.35%		
MSC-PCA	98.39%	97.99%	
SG-PCA	97.60%		

Source: developed by the authors

The results indicated that in the task of fish spectral classification, the preprocessing effects and classification performances of the SVM and KNN models exhibited significant differences. For the SVM model, the MSC-PCA combined preprocessing demonstrated the best performance, with an accuracy of 98.39% and an average precision of 97.99%. This was significantly better than using MSC alone (97.59%), SG alone (96.35%), or SG-PCA (97.60%), thus verifying the synergistic optimisation effect of MSC correction and PCA dimensionality reduction. When preprocessed individually, the spectral correction effect of MSC (97.59%) was better than that of SG (96.35%). The optimal preprocessing scheme for the KNN model was also MSC-PCA, with an accuracy of 97.86% and an average precision of 97.73%. This was superior to SG-PCA (97.59%), single-use of MSC (97.33%), and SG (96.20%), further validating the effectiveness of the feature extraction of this combination. Both models showed that, overall, the preprocessing method combining MSC and PCA was better than a single preprocessing method, highlighting the importance of PCA in removing redundant information and retaining key classification features. In the comparison of model performances, under the same preprocessing conditions, the accuracy of SVM was generally higher than that of KNN, indicating that SVM had a stronger adaptability to the classification of spectral data.

The results of the research are consistent with the conclusions of other scientists regarding the application of spectral analysis with data processing methods for identifying fish species and assessing the freshness of fish products. M. Rahman *et al.* (2023) used principal component analysis to differentiate the spectra of mackerel meat using fluorescence spectroscopy. Principal component analysis of FFs spectra demonstrated clear discrimination (PC1 + PC2 = 91.7%) between white and dark meat of frozen fish depending on the freshness state. Therefore, the principal component analysis method of dense matter spectra could instantly distinguish the freshness state of white and dark meat in fish products, even in the frozen state. The effectiveness of the fluorescence fingerprinting method combined with PCA for rapid and non-destructive monitoring of fish quality had also been demonstrated in the works of the aforementioned authors. In particular, M. Rahman *et al.* (2022) proposed a method using fluorescent fingerprints for non-destructive prediction of pH in frozen fish fillets. The results showed that the developed method was accurate enough to predict pH changes in horse mackerel ($R^2 = 0.71$) and spotted mackerel ($R^2 = 0.90$), therefore, it could be used as a fast and non-contact alternative to traditional pH electrodes.

K. Omwange *et al.* (2021) proposed a rapid and non-destructive method for assessing the freshness of *Tribolodon hakonensis* using multispectral imaging combined with multivariate analysis methods. The results achieved an R^2 of 0.94 and an RMSE of 2.42%. Multispectral fluorescence imaging was found to be a powerful tool for non-destructive monitoring of fish freshness during storage. The multispectral imaging method was appropriate for non-destructive detection of quality characteristics not only of fish, but also of other white meat products, such as shrimp, chicken, duck, and goose. Non-destructive methods for examining food products also included fluorescence spectroscopy and colour imaging. Based on machine learning and convolutional neural networks, these techniques could not only be used to determine the freshness and category of white meat through imaging and analysis, but could also be used to detect various harmful substances in meat products (Fan & Su, 2022; Fan *et al.*, 2024).

In the work of Y. Hu *et al.* (2023), a method for fish species authentication based on Raman spectroscopy was proposed. In this case, the PCA method was applied to analyse the data using KNN and SVM models to study the spectral features. At the initial stage of spectrum analysis, SVM models were selected to classify the sample into the group of salmon or non-salmon fish. Secondary classification models based on the PLS-DA or KNN algorithm were applied to classify the sample into one of four salmon species or one of seven non-salmon fish species. The classification model at the first level provided almost 100% accuracy, which allowed for confident identification with an accuracy of >93% at the second level for both fish groups. This fast and reliable fish authentication method was another useful tool for controlling the authenticity of fish, and therefore, for protecting consumer benefits.

In the work of J. Qin *et al.* (2020), a study was conducted to differentiate six types of fish fillets and evaluate the freshness of frozen-thawed fillets. Four types of spectra were obtained – visible and near-infrared reflectance, fluorescence, short-wave infrared reflectance, and Raman spectroscopy. The hyperspectral images of fish fillets were analysed using six models, including KNN and SVM. The highest accuracy was achieved at 100% when using the full VNIR reflectance spectra for species classification and 99.9% when using the full SWIR reflectance spectra for freshness classification. This method was proposed to quickly detect substitution and mislabelling of fish fillets (Do & Wong, 2025). Similarly, H. Li *et al.* (2022) developed rapid detection methods (RPA, RPA-LFD, and real-time RPA) targeting the serC gene of *Edwardsiella ictaluri* in yellow catfish, achieving a detection limit of 10^2 copies/μl within 30 minutes at 38°C. These methods, particularly RPA-LFD, demonstrated high sensitivity and specificity without requiring specialised equipment, offering a practical solution for on-site fish pathogen detection in aquaculture settings. R. Wang (2022) proposed a quantitative detection method for fish paste adulteration based on hyperspectral technology. By comparing and analysing the full-wavelength and simplified regression models, the optimal fish paste adulteration detection model was determined. The results indicated that both the full-wavelength model and the simplified model could effectively detect fish paste adulteration, with the prediction coefficient R^2 greater than 0.90, and the root mean square error of calibration (RMSEC) and root mean square error of prediction (RMSEP) being 1.9989% and 1.7983%, respectively.

In the paper of Y. Feng *et al.* (2023), surimi adulteration detection was used as an example for in-depth research on the performance of different spectral reconstruction algorithms. In the spectral reconstruction process, two main methods were investigated: the multivariate polynomial least squares regression (PMLR) algorithm and the hierarchical deep learning regression network (HRNet). Experimental results showed that the support vector regression model with standard normal transformation (SNV) preprocessing achieved the best results with a prediction correlation coefficient (R_p) of 0.9830 and a root mean square prediction error ($RMSE_p$) of 3.9544% in the model based on the spectrum reconstruction by the PMLR algorithm. In the model based on the reconstructed spectrum of the HRNet deep learning network, the support vector regression model, also using SNV preprocessing, achieved the best performance with a prediction correlation coefficient of 0.9987 and a root mean square prediction error of 4.0808%. Thus, the PMLR algorithm and the HRNet network reconstruction method based on RGB images not only realised efficient spectral reconstruction but also provided a reliable solution for surimi adulteration detection. Furthermore, Z. Deng *et al*. (2024) reviewed the application of deep learning models, such as convolutional neural networks (CNNs) and multilayer perceptrons (MLPs), in food authenticity, emphasising their superior performance over traditional machine learning for complex data analysis. Their work highlighted the potential of deep learning combined with spectroscopic techniques for rapid and accurate fish authenticity detection, aligning with the high accuracy achieved in the SVM-based spectral classification. In the studies by J. Zhou e*t al.* (2019) it was found that after spectral preprocessing and CARS characteristic wavelength optimisation, the correlation coefficients of the model were 0.961, 0.881, 0.955, and 0.946, respectively, and the cross-validation root mean square errors were 0.049, 1.659, 0.047, and 2.558, respectively, verifying the good predictive ability of the model and providing an effective method for rapid and non-destructive detection of freshwater fish freshness. P. Tian (2023) used spectral and hyperspectral imaging technology, combined with various chemometric methods and image processing techniques, to conduct chemometric analysis on frozen and refrigerated salmon meat and established a prediction model for qualitative analysis. It was found that the correct identification rate of the test samples could reach 97.92%. The use of CARS and SPA feature wavelength extraction could reduce the data dimension, and the correct identification rate of the CARS-CNN model could still reach 93.75%, effectively realising the identification of fresh and frozen-thawed salmon meat.

The study confirmed that combining multivariate scatter correction with principal component analysis yields the most effective preprocessing pipeline for hyperspectral fish classification, achieving the highest accuracy of 98.39% with the SVM model and demonstrating superior performance in noise reduction, feature extraction, and classification compared to SG and individual preprocessing methods.

Conclusions

Aiming at the classification problem of fish spectral data, this study proposed a dimensionality reduction method based on principal component analysis (PCA). The original spectral data were optimised by integrating two data preprocessing techniques, Savitzky-Golay (SG) and multiplicative scatter correction (MSC). This approach effectively enhanced the data quality. The high-dimensional spectral features were reduced to fewer than 10 dimensions while retaining a cumulative variance contribution rate of over 92%, which verified the effectiveness of the dimensionality reduction method in preserving spectral feature information. Based on the dimensionally reduced data, K-nearest neighbours (KNN) and support vector machine (SVM) classification models were constructed separately. The experimental results indicated that when the nearest neighbour parameter of the KNN algorithm was optimised to k = 4, it achieved an average classification accuracy of 97.25% on the test set. The KNN algorithm demonstrated stability on the dataset with a balanced category distribution, but was more sensitive to noise. Through the selection of the radial basis function (RBF) kernel and parameter tuning, the SVM algorithm achieved a classification accuracy of 97.48%. By comparing the combinations of the two preprocessing methods and dimensionality reduction strategies, it was discovered that the choice of the preprocessing method had a significant influence on the classification results. The accuracy of the MSC method for the SVM classifier was 0.53% higher than that for the KNN classifier. PCA not only decreased the computational complexity but also enhanced the model's robustness by removing redundant features.

This study still has certain limitations. For instance, there was an insufficient capability to identify the differences in spectral features among some fish species, or the generalisation ability under extreme lighting conditions remains to be verified. In the future, it would be possible to further explore the combination of a convolutional neural network (CNN), a deep learning model, with spectral features, or introduce multimodal data to improve the classification performance.

Acknowledgements

None.

Funding

None.

Conflict of Interest

None.

References

[1] Chen, Z., Wang, Q., Zhang, H., & Nie, P. (2021). Hyperspectral Imaging (HSI) technology for the non-destructive freshness assessment of pearl gentian grouper under different storage conditions. *Sensors*, 21(2), article number 583. doi: 10.3390/s21020583.

[2] Deng, Z., Wang, T., Zheng, Y., Zhang, W., & Yun, Y.-H. (2024). Deep learning in food authenticity: Recent advances and future trends. *Trends in Food Science & Technology*, 144, article number 104344. doi: 10.1016/j.tifs.2024.104344.

[3] Do, T.D., & Wong, L.L. (2025). Investigation of seafood mislabeling in Asia: A review of the literature and mitigation strategies. *Journal of Agriculture and Food Research*, 22, article number 101702. doi: 10.1016/j.jafr.2025.101702.

[4] European Convention for the Protection of Vertebrate Animals used for Experimental or other Scientific Purposes. (1986, March). Retrieved from https://rm.coe.int/168007a67b.

[5] Fan, K.-J., & Su, W.-H. (2022). Applications of fluorescence spectroscopy, RGB- and multispectral imaging for quality determinations of white meat: A review. *Biosensors*, 12(2), article number 76. doi: 10.3390/bios12020076.

[6] Fan, Y., Dong, R., Luo, Y., Tan, Y., Hong, H., Ji, Z., & Shi, C. (2024). Deep learning models with optimised fluorescence spectroscopy to advance freshness of rainbow trout predicting under nonisothermal storage conditions. *Food Chemistry*, 454, article number 139774. doi: 10.1016/j.foodchem.2024.139774.

[7] Feng, Y., Wan, S., Pan, S., Luan, P., Kong, L., & Zhu, M. (2023). Quantitative detection of surimi adulteration based on spectral reconstruction of RGB images. *Transactions of the Chinese Society of Agricultural Engineering*, 39(20), 275-282. doi: 10.11975/j.issn.1002-6819.202307013.

[8] He, L., Ma, J., Li, Q., Wang, L., Fan, S., & Zhang, Y. (2022). Determination of ingredients in livestock and poultry meat based on liquid chromatography-tandem mass spectrometry. *Journal of Future Foods*, 2(1), 53-60. doi: 10.1016/j.jfutfo.2022.03.017.

[9] Hu, Y., Huang, S.Y., & Lu, X. (2023). A green analytical method for fish species authentication based on Raman spectroscopy. *Green Analytical Chemistry*, 4, article number 100052. doi: 10.1016/j.greeac.2023.100052.

[10] Kim, N. (2021). Generalised indicator of the quality of qualimetry objects of various nature. *Ukrainian Black Sea Region Agrarian Science*, 25(1), 94-101. doi: 10.31521/2313-092X/2021-1(109)-12.

[11] Law of Ukraine No. 3447-IV "On the Protection of Animals from Cruelty". (2006, February). Retrieved from https://zakon.rada.gov.ua/laws/show/3447-15#Text.

[12] Li, B., Lu, Y., Liu Y., & Wan, X. (2024). Detection of adulterated beef mince based on hyperspectral reflectance and transmission fusion technology. *Transactions of the Chinese Society of Agricultural Engineering*, 40(16), 251-260. doi: 10.11975/j.issn.1002-6819.202402097.

[13] Li, H., Zhang, L., Yu, Y., Ai, T., Zhang, Y., & Su, J. (2022). Rapid detection of *Edwardsiella ictaluri* in yellow catfish (*Pelteobagrus fulvidraco*) by real-time RPA and RPA-LFD. *Aquaculture*, 552, article number 737976. doi: 10.1016/j.aquaculture.2022.737976.

[14] Melnychuk, D.O., Melnychuk, S.D., Voitsitskyi, V.M., Hryshchenko, V.A., Kalachniuk, L.H., Khyzhniak, S.V., & Tsvilikhovskyi, V.I. (2016). *Analytical research methods. Spectroscopic analysis methods: Theoretical foundations and techniques*. Kyiv: Komprynt.

[15] Omwange, K.A., Saito, Y., Zichen, H., Khaliduzzaman, A., Kuramoto, M., Ogawa, Y., Kondo, N., & Suzuki, T. (2021). Evaluating Japanese dace (*Tribolodon hakonensis*) fish freshness during storage using multispectral images from visible and UV excited fluorescence. *LWT*, 151, article number 112207. doi: 10.1016/j.lwt.2021.112207.

[16] Palamarchuk, I., Yuanxia, F., Zhuravel, D., Petrychenko, I., Blishch, R., Holovatyuk, A., Domin, O., & Kostiuk, T. (2024). Spectroscopic assessment and quantitative analysis of the trace element composition of vegetable additives to meat products. *Potravinarstvo Slovak Journal of Food Sciences*, 18, 480-496. doi: 10.5219/1965.

[17] Qin, J., Vasefi, F., Hellberg, R.S., Akhbardeh, A., Isaacs, R.B., Yilmaz, A.G., Hwang, C., Baek, I., Schmidt, W.F., & Kim, M.S. (2020). Detection of fish fillet substitution and mislabeling using multimode hyperspectral imaging techniques. *Food Control*, 114, article number 107234. doi: 10.1016/j.foodcont.2020.107234.

[18] Rahman, M., Shibata, M., Nakazawa, N., Rithu, N.A., Nakauchi, S., Hagiwara, T., Osako, K., & Okazaki, E. (2022). Nondestructive approach for the prediction of pH in frozen fish meat using fluorescence fingerprints in tandem with chemometrics. *Fishes*, 7, article number 364. doi: 10.3390/fishes7060364.

[19] Rahman, M., Shibata, M., Nakazawa, N., Rithu, N.A., Okazaki, E., & Nakauchi, S. (2023). Potential of fluorescence fingerprints for fish meat authentication: Differences in freshness evaluation in white and dark meat at frozen state. *Journal of Food Science*, 88(12), 5339-5354. doi: 10.1111/1750-3841.16825.

[20] Shi, X., Zhang, J., Shi, C., Tan, Y., Hong, H., & Luo, Y. (2022). Nondestructive prediction of freshness for bighead carp (*Hypophthalmichthys nobilis*) head by Excitation-Emission Matrix (EEM) analysis based on fish eye fluid: Comparison of BPNNs and RBFNNs. *Food Chemistry*, 382, article number 132341. doi: 10.1016/j.foodchem.2022.132341.

[21] Tian, P. (2023). *Research on salmon quality detection based on VIS-NIR spectroscopy and hyperspectral imaging technology*. Beijing: Tongfang Knowledge Network (Beijing) Technology. doi: 10.27147/d.cnki.ghdju.2023.000624.

[22] Tsagkatakis, G., Nikolidakis, S., Petra, E., Kapantagakis, A., Grigorakis, K., Katselis, G., Vlahos, N., & Tsakalides, P. (2020). Fish freshness estimation though analysis of multispectral images with convolutional neural networks. *Electronic Imaging*, 32(12), 171-175. doi: 10.2352/issn.2470-1173.2020.12.fais-171.

[23] Wang, H., *et al*. (2019). Recent progress in hyperspectral imaging for nondestructive evaluation of fish quality. *Food Science*, 40(05), 329-338. doi: 10.7506/spkx1002-6630-20180129-392.

[24] Wang, R. (2022). *Study on detection of surimi starch adulteration based on spectral reconstruction*. Beijing: Tongfang Knowledge Network (Beijing) Technology. doi: 10.27158/d.cnki.ghznu.2022.001742.

[25] Wu, L., Ma, L., Zhang, Y., Tian, Y., Zhu, Y., & Zhang, Y. (2025). Detection of coated cabbage seeds color based on hyperspectral imaging technology. *Journal of Analysis and Testing*, 44(3), 454-463. doi: 10.12452/j.fxcsxb.240711212.

[26] Xu, T. (2020). *Detection of drug residues in fish based on surface enhanced Raman spectroscopy*. Beijing: Tongfang Knowledge Network (Beijing) Technology. doi: 10.27314/d.cnki.gsscu.2020.000414.

[27] Yuan, Y., Ji, Z., Fan, Y., Xu, Q., Shi, C., Lyu, J., & Ertbjerg, P. (2025). Deep learning-assisted fluorescence spectroscopy for food quality and safety analysis. *Trends in Food Science & Technology*, 156, article number 104821. doi: 10.1016/j.tifs.2024.104821.

[28] Zhou, J., Wu, X., & Chen, Z. (2019). Rapid prediction of freshness of *Megalobrama amblycephala* using near infrared spectroscopy. *Journal of Huazhong Agricultural University*, 38(4), 120-126. doi: 10.13300/j.cnki.hnlkxb.2019.04.016.

Спектральна класифікація риб на основі основного компонентного аналізу

Лі Сюй
Аспірант, викладач
Університет Бенгбу
233030, дорога Каошань, м. Бенгбу, провінція Аньхой, Китай
https://orcid.org/0009-0000-3762-1714

Аліна Менчинська
Кандидат технічних наук, доцент
Національний університет біоресурсів і природокористування України
03041, вул. Героїв Оборони, 15, м. Київ, Україна
https://orcid.org/0000-0001-8593-3325

Олена Очколяс
Кандидат технічних наук, доцент
Національний університет біоресурсів і природокористування України
03041, вул. Героїв Оборони, 15, м. Київ, Україна
http://orcid.org/0000-0002-8483-578X

Микола Ніколаєнко
Доктор філософії, доцент
Національний університет біоресурсів і природокористування України
03041, вул. Героїв Оборони, 15, м. Київ, Україна
https://orcid.org/0000-0003-2213-4985

Анотація. Спектральна технологія виявлення може бути використана для розпізнавання різних видів м'яса риби та визначення наявності у зразках інгредієнтів, що допомагає контролювати автентичність та відповідність продукції. Метою дослідження була розробка та порівняння ефективності методів машинного навчання для класифікації різних видів риби на основі гіперспектральних даних. У роботі гіперспектральну технологію використано для отримання спектральних даних 328 точок вибірки у смузі 400-1000 нм чотирьох різних риб, а спектральні криві були попередньо оброблені для отримання більш точних та ефективних спектральних кривих. У дослідженні систематично порівняно методи попереднього оброблення Савицького-Голея згладжування та багатовимірної корекції розсіювання (MSC) та поєднано метод аналізу головних компонентів (PCA) для оптимізації моделей класифікації k-найближчих сусідів (KNN) і машини опорного вектору (SVM) з метою забезпечити більш ефективне рішення для класифікації риб. Метод аналізу головних компонентів використано для зменшення розмірності спектральних даних. Основні компоненти після зменшення розмірності порівняно з повними компонентами, тренувальний набір та тестовий набір згенеровано випадковим чином. Алгоритм SVM та алгоритм KNN використано у Python для прогнозування та визначення точності моделі. Показано, що середня точність спектральних даних після попереднього оброблення та зменшення розмірності у класифікації моделей SVM може досягати 97,25 %, а середня точність класифікації моделей KNN – 97,48 %. Це свідчить про те, що різні види м'яса риби можуть бути класифіковані. Обидві моделі показали, що в цілому метод попереднього оброблення, який поєднує MSC і PCA, є кращим, ніж один метод попереднього оброблення, підкреслюючи важливість PCA для видалення зайвої інформації та збереження ключових характеристик класифікації. Під час порівняння продуктивності моделі за однакових умов попереднього оброблення встановлено вищу точність SVM, ніж KNN, що вказує на більшу адаптивність SVM до класифікації спектральних даних. Завдяки розробленню та застосуванню спектральної технології виявлення можна сприяти технологічному прогресу та інноваціям у тестуванні харчових продуктів, направляючи всю галузь до безпечнішого та надійнішого майбутнього

Ключові слова: спектр; метод Савицького-Голея; мультиплікативна корекція розсіювання; модель k-найближчих сусідів; алгоритм машини опорного вектору