

Remaining shelf-life prediction of squid using handheld Vis/NIR spectroscopy and ensemble learning

Xin Miao¹, Yajie Sun¹, Anyi Tong¹, Wansheng Bao¹, Chichao Liu², Shouqi Cao¹, Xinjun Chen^{1*}

(1. College of Information Technology, Shanghai Ocean University, Shanghai 201306, China;

2. Xichang Satellite Launch Center Wenchang Launch Site, Xichang 615000, Sichuan, China)

Abstract: To address the limitations of traditional squid freshness assessment, which relies on destructive physicochemical indicators and neglects anatomical spatial heterogeneity—thereby compromising on-site accuracy—this study developed an integrated handheld visible and near-infrared (Vis-NIR) spectroscopy system combined with intelligent algorithms. A multi-scale decomposition of spectral signals (450–950 nm) was innovatively performed using Discrete Wavelet Transform (DWT) to extract energy spectrum features. By integrating Isolation Forest for outlier rejection and a Stacked Ensemble Learning framework (Stacking: base=Ridge/RF/HGBR, meta=GBR), a robust prediction model for storage time was constructed. The experimental results demonstrated that the proposed model achieved an excellent coefficient of determination ($R^2=0.9528$), a Root Mean Square Error (RMSE=0.6566 d), and a Residual Prediction Deviation (RPD=4.60) on the validation set, with uniformly distributed residuals cross-verified by the Breusch–Pagan test ($p=0.6944$). Weibull kinetic modeling revealed distinct degradation patterns across anatomical sites; specifically, the abdominal region exhibited a maximum shape parameter ($\beta=1.787$) and a massive Akaike Information Criterion advantage ($|\Delta AIC|=490.6$) over the first-order kinetic paradigm, indicating an accelerating, self-catalytic quality degradation profile driven by endogenous visceral enzymes. Finally, the real-time predicted storage time was successfully coupled with local Weibull parameters to output probabilistic Remaining Shelf Life (RSL) profiles via Kernel Density Estimation (KDE), which establishes a proactive digital traceability pipeline for precision grading and dynamic inventory management of high-value marine aquatic products.

Keywords: squid freshness, handheld spectrometer, spatial heterogeneity, stacked ensemble learning, kinetic modeling, remaining shelf life (RSL)

DOI: [10.25165/ijabe.20261904.10501](https://doi.org/10.25165/ijabe.20261904.10501)

Citation: Miao X, Sun Y J, Tong A Y, Bao W S, Liu C C, Cao S Q, et al. Remaining shelf-life prediction of squid using handheld Vis/NIR spectroscopy and ensemble learning. Int J Agric & Biol Eng, 2026; 19(4): 338–346.

1 Introduction

Aquatic products serve as an important key pillar of the global food system, providing essential animal proteins, omega-3 fatty acids, and vital micronutrients^[1,2]. According to the 2024 China Distant-water Squid Jigging Industry Development Report^[3], China's squid industry has maintained a leading global position, contributing significantly to international aquatic trade. However, squid (e.g., *Ommastrephes bartramii*) are highly perishable post-harvest^[4]. Their high moisture content, rich nitrogenous extractive compounds, and active endogenous enzymes facilitate rapid microbial proliferation and biochemical degradation^[5]. This spoilage not only leads to substantial economic losses in the supply chain but also poses severe risks of foodborne illnesses, such as histamine

poisoning^[6]. Consequently, developing a rapid, non-destructive, and reliable freshness monitoring technology for on-site industrial application is a critical priority for both the fishing industry and food safety regulatory bodies^[7,8].

Freshness is a multi-dimensional concept typically quantified by various chemical and biological indices. Currently, Total Volatile Basic Nitrogen (TVB-N)^[9,10] is the most widely accepted regulatory index for characterizing the breakdown of proteins and amines by spoilage bacteria and autolytic enzymes. In parallel, the K-value^[11,12], derived from the degradation pathway of Adenosine Triphosphate (ATP)^[13], serves as a sensitive indicator for early-stage freshness changes associated with post-mortem metabolism^[14]. Despite their accuracy, standard laboratory methods for measuring TVB-N (e.g., the semi-micro Kjeldahl method) involve tedious sample preparation, including comminution, extraction, distillation, and titration^[15]. These procedures typically exceed two hours per sample and are inherently destructive^[16]. Such limitations render them incompatible with the fast-paced requirements of modern “Internet of Things” (IoT)-based cold-chain logistics, where real-time, non-invasive quality assessment is paramount^[17]. In recent decades, researchers have explored diverse non-destructive sensing technologies to bridge this gap, including electronic noses (E-noses)^[18], hyperspectral imaging (HSI)^[19,20], and visible/near-infrared (Vis-NIR) spectroscopy^[21]. While E-noses are sensitive to volatile odors, they are often susceptible to environmental humidity and sensor drift. HSI provides both spatial and spectral information but suffers from high data dimensionality and slow processing speeds, which limit its portability^[22]. In contrast, Vis-NIR spectroscopy,

Received date: 2026-02-06 **Accepted date:** 2026-06-11

Biographies: **Xin Miao**, MS, Engineer, research interest: intelligent marine equipment, Email: xmiao@shou.edu.cn; **Yajie Sun**, Undergraduate, research interest: intelligent marine equipment, Email: 2162966592@qq.com; **Anyi Tong**, Undergraduate, research interest: intelligent marine equipment, Email: anyi_tong_shou@163.com; **Wansheng Bao**, Undergraduate, research interest: intelligent marine equipment, Email: 2605261132@qq.com; **Chichao Liu**, MS, Engineer, research interest: intelligent marine equipment, Email: liucc1992@live.com; **Shouqi Cao**, Professor, research interest: intelligent marine equipment, Email: sqcao@shou.edu.cn.

***Corresponding author:** **Xinjun Chen**, PhD, Professor, research interest: fishery resources and fishing ground science, distant-water squid jigging fishery, and fishery resource economics. No. 999 Hucheng Ring Road, Pudong New Area, Shanghai 201306, China. Tel: +86-21-61900304, Email: xjchen@shou.edu.cn.

which probes the overtones and combination bands of hydrogen-containing functional groups (e.g., N-H, O-H, and C-H), offers a balance of high speed, cost-effectiveness, and potential for miniaturization^[23]. Previous studies have demonstrated the feasibility of using benchtop Vis-NIR systems to predict TVB-N in various fish species under controlled laboratory settings^[24,25]. However, transitioning this technology from the laboratory to the fishing vessel or warehouse faces a fundamental challenge: the inherent spatial heterogeneity of squid tissues^[26].

Squid quality deterioration is not a spatially uniform process. Anatomically, the abdomen contains a high density of endogenous enzymes (such as cathepsins and lipases) and is in close proximity to the visceral mass, which harbors a significant initial microbial load. The release of these endogenous enzymes, particularly calpains and cathepsins, plays a critical role in the structural softening of muscle tissues and the degradation of myofibrillar proteins in post-mortem fish^[27]. Conversely, the muscular tail consists of dense, uniform protein fibers with relatively lower enzymatic activity^[28]. This physiological differentiation leads to disparate deterioration kinetics across the fish body. Most existing spectral models treat the fish as a homogeneous entity and rely on single-point measurements, which often leads to significant prediction bias and “false negatives” in spoilage detection^[29]. Ignoring this spatial heterogeneity is a primary reason why many portable spectral devices fail to achieve satisfactory accuracy in real-world industrial environments.

This research aims to address these limitations by proposing a synergistic framework that integrates miniaturized spectral sensing with a spatially-aware intelligent algorithm. Specifically, this study innovatively employs Discrete Wavelet Transform (DWT) to perform multi-scale decomposition of Vis-NIR spectra (450–950 nm), extracting energy spectrum features to suppress the high-frequency environmental noise characteristic of handheld devices. To handle the non-linear spectral response arising from spatial variations, a Stacked Ensemble Learning framework—integrating Random Forest (RF) and Histogram-based Gradient Boosting Regression (HGBR)—was developed to predict the storage time (days). Furthermore, by establishing Weibull degradation models, this study quantitatively characterizes the deterioration kinetics of the abdomen and tail, revealing the spatial heterogeneity in their deterioration modes. Finally, a Remaining Shelf Life (RSL) prediction system was implemented based on real-time storage time data^[30]. This approach not only fills the theoretical gap in spatiotemporal freshness monitoring but also provides a robust digital tool for precision grading and intelligent quality management in the aquatic product supply chain.

2 Materials and methods

2.1 Handheld Vis-NIR spectroscopic system design

To achieve portable and non-destructive monitoring of squid freshness, an integrated handheld visible and near-infrared (Vis-NIR) spectroscopic detection system was developed. The core hardware architecture consists of several key components, as illustrated in Figure 1. The light source utilizes an (b) integrated miniature tungsten halogen lamp with an emission range of 300–1050 nm and a constant output power of 5 W. This source is optimized with a low-heat radiation filter to prevent secondary thermal damage to heat-sensitive proteins in squid tissues. The signal acquisition module features a (a) miniaturized spectroscopic detector with an optical resolution of 3 nm and a signal-to-noise ratio (SNR) exceeding 300:1.

As shown in Figure 2, the device features a three-dimensional (3D) structural design based on ergonomic principles. This all-in-one handheld apparatus integrates the light source, spectrometer, data processing module, display screen, and battery into a compact and efficient housing. During operation, the workflow is triggered by a pistol-grip button. Once activated, the light source emits a beam onto the sample, and the spectrometer captures the returning diffuse reflectance data via a coaxial reflective probe (path length approximately 2 mm). The processing module then converts these spectral signals into corresponding squid freshness information, which is immediately output via the 2.4-inch display screen.

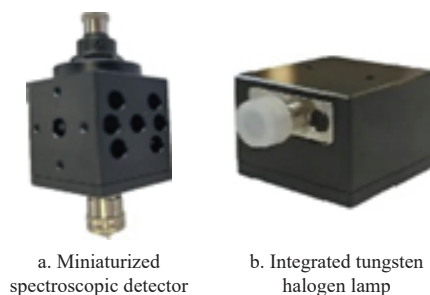


Figure 1 Main key components of the handheld Vis-NIR spectroscopic detection device



Figure 2 Three-dimensional (3D) structural design rendering of the handheld Vis-NIR spectroscopic detection device

2.2 Theoretical basis and spectral decoupling

The physical basis for monitoring squid freshness using Vis-NIR spectroscopy lies in the interaction between photons and hydrogen-containing functional groups (e.g., N-H, O-H, C-H) in muscle tissues. Light penetration and diffuse reflectance within the muscle media follow the Lambert-Beer Law, expressed as:

$$A = \log(1/R) = \epsilon \cdot b \cdot c + e \quad (1)$$

where, A is the absorbance; R is the reflectance (ratio of detected light intensity to a reference white tile); ϵ is the molar extinction coefficient of functional groups; b is the effective path length of photons in tissues; c is the concentration of biochemical components (e.g., TVB-N); and e represents the random error caused by environmental scattering and system drift^[31–35].

To successfully isolate the authentic biochemical fingerprints from the compounding e noise vectors, Discrete Wavelet Transform (DWT) was introduced to decouple the raw spectral signals (450–950 nm). The Daubechies-4 (db4) mother wavelet was strategically selected because its 4-tap compact filter system maintains an optimal balance between time-frequency localization and computational efficiency, making it uniquely suited for processing continuous, wide broadband NIR absorption profiles^[36]. A comprehensive grid-search optimization swept across multiple wavelet bases (db2 to db8) and decomposition depths (2 to 5 levels) using leave-one-out cross-validation. The results confirmed that a 3-

level decomposition structure uniquely minimized the root mean square error (RMSE) of the training set. The continuous spectral signal is mathematically decomposed into an approximation sub-band (cA_3) and a sequence of detail sub-bands (cD_3 , cD_2 , cD_1), which can be reconstructed via the following multi-scale paradigm:

$$y(k) = \sum_n cA_3(n)\phi_{3,n}(k) + \sum_{j=1}^3 \sum_n cD_j(n)\psi_{j,n}(k) \quad (2)$$

where, $y(k)$ is the original uncalibrated spectral amplitude at the wavelength index k ; j and n designate the dimensionless decomposition scale and translation shifts, respectively; $\phi(k)$ and $\psi(k)$ denote the scaling function and the db4 mother wavelet function, respectively. The low-frequency approximation coefficients (cA_3) capture the foundational baseline and macromolecular skeletal absorbance configurations containing the essential fingerprints of squid freshness degradation. Conversely, the high-frequency detail coefficients (cD_j) systematically isolate operational instrumentation noise, high-frequency sensor jitter, and ambient scatter. To transform these decoupled continuous coefficients into highly robust, non-redundant inputs for the subsequent Stacked Ensemble Learning framework, a wavelet energy spectrum analysis was executed. For each decoupled sub-band ($s \in \{cA_3, cD_3, cD_2, cD_1\}$), two independent statistical features—the Root Mean Square (RMS_s) and the Standard Deviation (Std_s)—were extracted from the coefficient vectors to quantify the sub-band energy density distribution:

$$RMS_s = \sqrt{\frac{1}{M} \sum_{m=1}^M c_s(m)^2} \quad (3)$$

$$Std_s = \sqrt{\frac{1}{M} \sum_{m=1}^M (c_s(m) - \bar{c}_s)^2} \quad (4)$$

where, $c_s(m)$ is the m -th wavelet coefficient within the sub-band s , M is the total number of coefficients in that scale, and $\bar{c}_s$ is the arithmetic mean. Combining these eight multi-scale energy features (RMS and Std across four sub-bands) with the categorical anatomical region tag creates a highly descriptive feature space. This space successfully minimizes multicollinearity while preserving multi-resolution variance, ensuring optimal generalization for the Stacking meta-regressor.

2.3 Sample preparation and multi-point sampling strategy

To closely monitor the dynamic, continuous spoilage process of squid (*Ulex coindetii*) under cold storage (4°C), a tracking experiment was conducted across six high-resolution time points up to advanced spoilage: Day 0, Day 1.1, Day 3.1, Day 6.9, Day 7.0, and Day 7.7. A total of 10 fresh squid specimens with uniform size and weight were continuously tracked as a longitudinal cohort. At each designated time point, non-destructive visible and near-infrared (Vis-NIR) spectral sampling was performed at six pre-defined anatomical locations on each squid (including four sub-regions on the abdomen, the head, and the tail) to capture spatiotemporal degradation profiles (Figure 3). Consequently, a total of 360 raw spectra were successfully acquired (10 squid × 6 time points × 6 anatomical sites = 360 spectra). This rigorous tracking design effectively eliminates inter-individual biological variations, ensuring that the model captures pure kinetic degradation pathways over time.

Considering the spatial heterogeneity of squid physiological

structures, a multi-point sampling strategy was designed covering six key anatomical sites: the head, the tail, and four abdominal sub-regions. The logic for this strategy is three-fold:

1) Asynchronous physiological deterioration: The abdominal region, adjacent to the visceral mass, contains high concentrations of endogenous enzymes (cathepsins) and high initial microbial loads. Consequently, its deterioration rate is significantly higher than that of the dense muscular tail.

2) Spatial detection integrity: In field testing, the randomness of probe placement can lead to significant prediction bias. Multi-point coverage captures the quality gradient from “rapidly deteriorating zones” to “stable zones,” effectively neutralizing spatial bias.

3) Global representative modeling: Synchronous monitoring allows the algorithm to learn spectral variations across different tissues, enabling the subsequent model to achieve nonlinear error compensation.

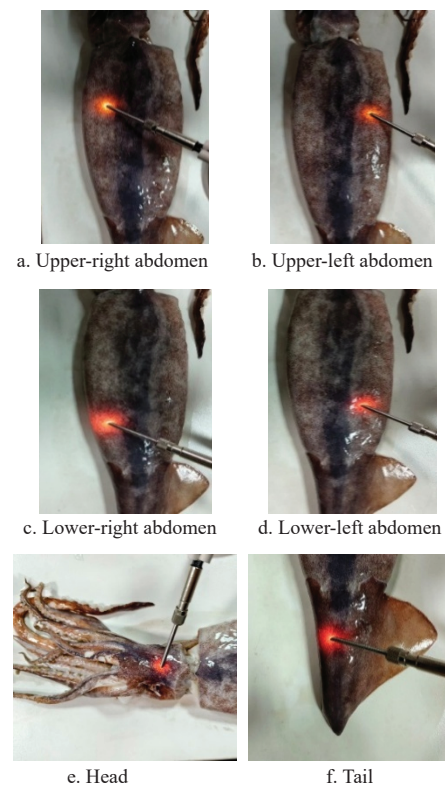


Figure 3 Schematic of freshness detection at six key anatomical locations of squid

2.4 Stacking ensemble modeling and performance evaluation

To fully capture the highly non-linear mapping between multi-scale spectral energy vectors and true quality deterioration, a robust prediction model for squid storage time was constructed based on a heterogeneous Stacked Ensemble Learning framework. Unlike monolithic regressors that are susceptible to local optimal traps or high variance, the Stacking paradigm harnesses the collaborative strengths of diverse statistical learning architectures.

In this study, a two-layer Stacking structure was implemented. The first layer (base-learners) comprises three structurally distinct estimators: Ridge Regression (a linear model with L2 regularization), Random Forest (RF, a bagging-based ensemble tree method), and Histogram-based Gradient Boosting Regression (HGBR, a boosting-based tree algorithm). This heterogeneous combination ensures that linear trends, deep feature interactions, and structural variations within the continuous spectrum are simultaneously learned. The second layer utilizes a shallow

Gradient Boosting Regressor (GBR) as the meta-learner to integrate the individual predictions. The formal decision output of the Stacking ensemble framework is defined as follows:

$$\hat{y} = g(f_1(x), f_2(x), \dots, f_M(x)) \quad (5)$$

where, $\hat{y}$ represents the final predicted storage time, d; $f_m(x)$ represents the independent prediction output generated by the m -th base-learner (Ridge, RF, and HGBR); and $g(\cdot)$ denotes the mapping function of the meta-learner GBR (n_estimators=50, max_depth=2, passthrough=True). By learning the error correlations of the baseline estimators, this hierarchical architecture effectively smooths prediction fluctuations caused by anatomical spatial heterogeneity^[37,38].

Model performance was rigorously evaluated using the Coefficient of Determination (R^2), Root Mean Square Error (RMSE), and the Residual Prediction Deviation (RPD), calculated as:

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (6)$$

$$\text{RPD} = \frac{\text{SD}_{\text{validation}}}{\text{RMSE}} \quad (7)$$

where, y_i is the actual measured storage time, d; $\hat{y}_i$ is the model-predicted time, d; N is the sample size; and $\text{SD}_{\text{validation}}$ is the standard deviation of the validation dataset. A lower RMSE coupled with an $\text{RPD} > 3.0$ indicates exceptional prediction accuracy and reliability in fluctuating on-site operational environments.

2.5 Kinetic projection and RSL forecasting

To transition from retrospective quality assessment to predictive shelf-life management, the real-time storage time tracking outputted by the machine learning algorithm must be coupled with systematic degradation kinetics. Perishable aquatic products rarely undergo fixed-rate linear decay due to the multi-stage acceleration of biochemical spoilage. Consequently, the quality deterioration kinetics for each anatomical site were fitted using a non-linear Weibull degradation model:

$$f(t) = a \cdot \left[1 - \exp \left(- \left(\frac{t}{\eta} \right)^\beta \right) \right] \quad (8)$$

where, $f(t)$ represents the temporal quality state at storage time t ; a is the asymptotic upper bound of deterioration; η is the scale

parameter (representing the characteristic reaction time frame, days); and β is the dimensionless shape parameter dictating the curvature of the kinetic profile ($\beta > 1$ designates a sigmoidal acceleration loop).

The mathematical superiority of the Weibull distribution over conventional first-order equations was evaluated via the Akaike Information Criterion (AIC). The statistical computation confirmed Weibull superiority across all targeted regions, with the abdomen yielding an optimal AIC of -1578.72 (compared to -1088.17 for the first-order model). The estimated localized Weibull parameters extracted from non-linear regression were: Abdomen ($\eta=4.26$ d, $\beta=1.79$), Tail ($\eta=4.26$ d, $\beta=1.78$), and Head ($\eta=4.32$ d, $\beta=1.74$).

Based on the real-time predicted storage time ($\hat{t}$) and these tissue-specific kinetic parameters, the remaining shelf life (RSL) is projected using the following analytical framework:

$$\text{RSL} = T_{\text{threshold}} - \hat{t} \quad (9)$$

where, $T_{\text{threshold}}$ represents the absolute terminal spoilage baseline established at 7.0 days under 4°C refrigeration, which strictly aligns with the total volatile basic nitrogen (30 mg/100g TVB-N) safety limit enforced by the national standard GB 2733-2015. This systematic integration facilitates the critical advancement from “passive state monitoring” to “active future prediction,” delivering a robust digital solution for high-efficiency aquatic product supply chain management.

3 Results and discussion

3.1 Spectral characteristic identification and multi-scale DWT decoupling

The raw diffuse reflectance spectra (450-950 nm) systematically gathered from different anatomical regions of squid samples across the 7-day storage trial are graphically illustrated in Figure 4a. The typical continuous spectral profiles reveal compounding absorption footprints governed by electronic transitions and multi-harmonic overtones of biochemical bonds within the complex myofibrillar tissue. Prominent absorption valleys observed near 550 nm and 760 nm correspond to physical pigment variations (such as metmyoglobin formation) and the third overtone of O-H stretching bonds in free muscle water, respectively. However, the raw continuous spectra are heavily superposed with high-frequency instrumentation scatter, stray environmental light fluctuations, and instrumentation drift, which mask the subtle, non-stationary variations induced by real-time freshness loss.

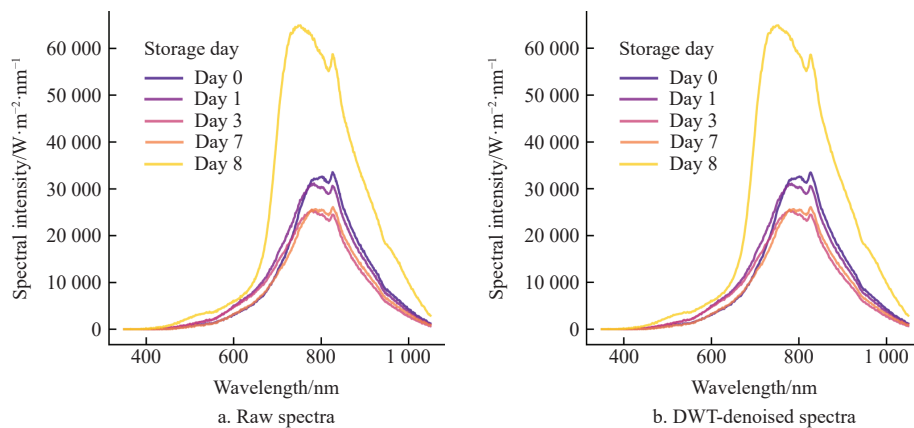


Figure 4 Vis-NIR spectra of squid samples

To enhance the signal-to-noise ratio, the Discrete Wavelet Transform (DWT-db4) was executed to perform multi-scale signal

decoupling. As demonstrated in Figure 4b, after stripping the stochastic high-frequency clutter isolated within the detail sub-

bands (cD_1 - cD_3), the reconstructed approximation baseline (cA_3) maintains an exceptionally smooth trend while successfully retaining the definitive chemical peak-valley positions. This multi-scale separation mathematically demonstrates that DWT effectively suppresses background drift without distorting essential biological absorption fingerprints, providing a pristine, high-fidelity feature space for subsequent intelligent ensemble modeling.

3.2 Multi-scale feature correlation and DWT band contribution analysis

To further elucidate the mathematical significance and structural contribution of the decoupled wavelet sub-bands, a comprehensive feature dependency analysis was systematically executed, as visualized in Figure 5. This analysis evaluates the coupling relationships and informational redundancy among the eight extracted wavelet attributes (the RMS and Std vectors derived from cA_3 , cD_3 , cD_2 , and cD_1). The quantitative matrix reveals that

while the low-frequency approximation coefficients (RMS- cA_3 and Std- cA_3) encapsulate the macro chemical absorbance profiles of macromolecular degradation, the high-frequency detail components (cD_1 - cD_3) successfully isolate localized, highly dynamic physical scatter variances. Notably, the low multi-collinearity observed across different resolution tiers demonstrates the immense superiority of multi-scale feature decoupling for handling non-stationary biological signals. Rather than feeding raw, highly redundant continuous spectral variables into the regressor, the DWT framework successfully compresses the input dimensionality while explicitly capturing the localized energy density variations induced by spatial-temporal degradation. This multi-resolution feature space provides highly orthogonal, informative, and structurally stable inputs for the subsequent Stacked Ensemble framework, effectively neutralizing background variations across distinct anatomical regions.

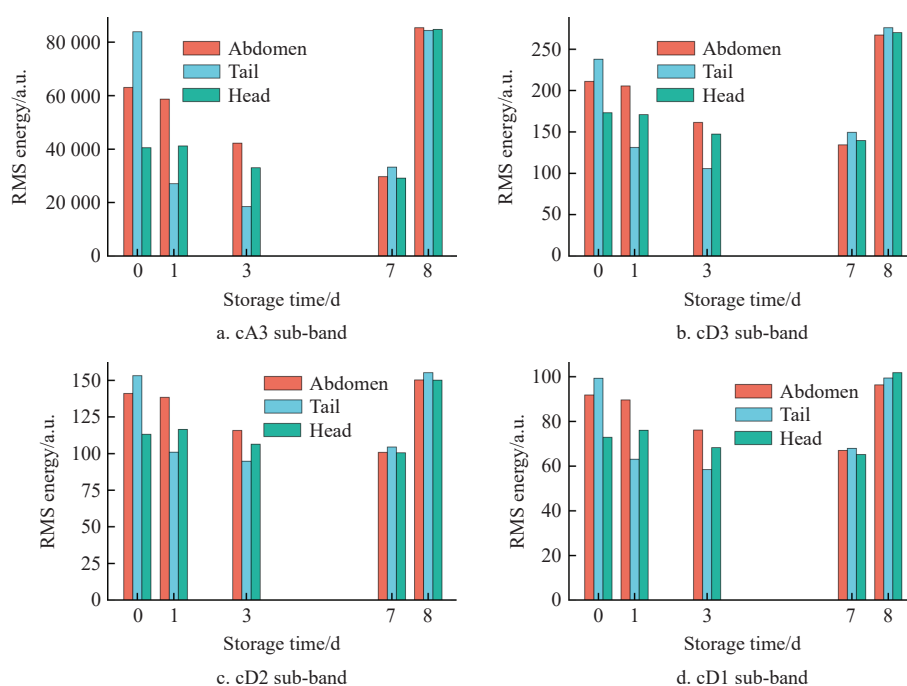


Figure 5 Feature dependency and multi-scale DWT sub-band contribution matrix

3.3 Pattern recognition and clustering analysis based on spectral fingerprints

To examine the evolution of squid freshness in the feature space from a macro perspective, this study introduced Principal Component Analysis (PCA, Figure 6) and t-Distributed Stochastic Neighbor Embedding (t-SNE, Figure 7). These visualizations aim to verify whether the energy spectrum features processed by Discrete Wavelet Transform (DWT) can effectively characterize freshness degradation gradients.

In the PCA projection, the first two principal components explained over 85% of the total variance. Samples exhibited a distinct linear drift over storage time, corresponding to changes in the tertiary structure of muscle proteins under autolytic enzymes. However, PCA showed limitations in separating overlapping samples from mid-stage storage. Consequently, t-SNE was introduced, which clearly clustered fresh stages (Days 0-2) and spoilage stages (Days 6-8) into distinct groups. This spatial self-aggregation indicates that the extracted wavelet energy spectrum features can overcome background variations between samples and accurately isolate essential fingerprints representing freshness loss.

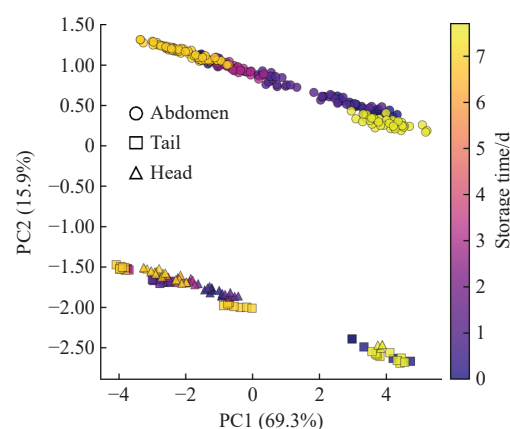


Figure 6 PCA of squid freshness trajectory

3.4 Prediction model accuracy and residual analysis

The quantitative performance of the Stacked Ensemble model for the storage time (d) was evaluated to diagnose potential systematic biases.

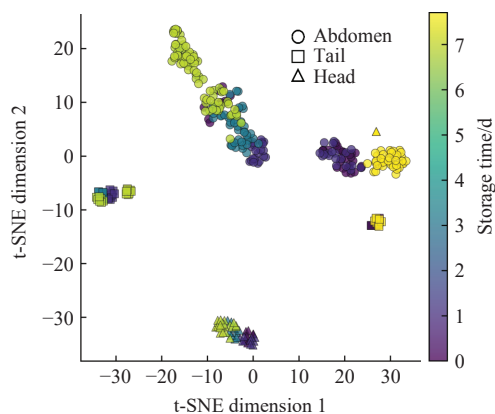


Figure 7 t-SNE non-linear clustering

To eliminate systematic operational biases and confirm the mathematical advantages of hierarchical algorithmic fusion, all models were validation-tested under identical feature spaces. Table 1 summarizes the comprehensive statistical indicators, including the Coefficient of Determination (R^2), Root Mean Square Error (RMSE), and the Residual Prediction Deviation (RPD).

Table 1 Performance comparison of regression models

Model	R^2	RMSE/days	RPD
PLSR	0.2623	2.5947	1.16
Ridge	0.5242	2.0838	1.45
RF	0.9464	0.6992	4.32
HGBR	0.9036	0.9382	3.22
Proposed stacking	0.9528	0.6566	4.60

As quantified in Table 1, traditional linear frameworks such as Partial Least Squares Regression (PLSR) and Ridge regression failed to establish acceptable predictive pathways ($R^2 \leq 0.5242$, $RPD < 1.50$), heavily suffering from the non-stationary spectral overlaps induced by biological matrix complexity. Conversely, the tree-based ensemble methods (RF and HGBR) demonstrated a remarkable capacity to isolate non-linear deterioration vectors, cross-validating an $RPD > 3.0$ which represents excellent analytical accuracy^[36]. Crucially, the proposed Stacking framework successfully integrated the orthogonal mathematical topologies of both linear and tree models. Guided by the non-linear meta-learner GBR, the proposed model achieved an exceptional tracking capability with an R^2 of 0.9528, a minimized error boundary of 0.6566 d, and a top-tier RPD of 4.60. Compared to the best single

model (RF), the Stacking architecture yielded a relative increase of 0.64% in R^2 and suppressed the prediction error (RMSE) by 6.1%, justifying the immense engineering advantage of stacked collaborative regression.

The 1:1 regression plot in Figure 8 shows that the validation samples are tightly distributed around the ideal diagonal line, with $R^2=0.9528$, $RMSE=0.6566$ d, and $RPD=4.60$.

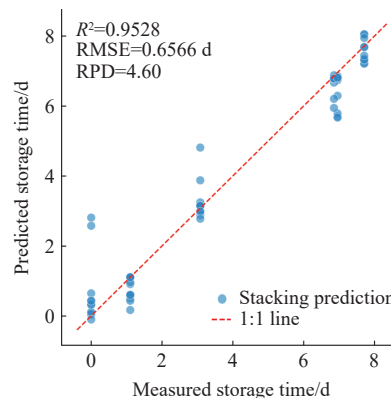


Figure 8 Prediction accuracy

The residual shows that errors are randomly and uniformly distributed on both sides of the zero line with no obvious heteroscedasticity. This was statistically confirmed by the Breusch-Pagan test ($BP=8.21$, $p=0.6944$, $df=9$). Because the p -value stands significantly above the nominal significance level ($\alpha=0.05$), the null hypothesis of homoscedasticity cannot be rejected, proving that the model preserves uniform error variance. This statistically justifies that the Isolation Forest algorithm effectively eliminated outliers caused by physical movement or stray environmental light.

3.5 Quantitative characterization of spatio-temporal kinetic heterogeneity

To mathematically decrypt the fundamental kinetic mechanisms governing the quality decline of squid during cold storage, a non-linear Weibull degradation model was systematically implemented and compared against the classical first-order kinetic approach. Traditional evaluation strategies frequently presume that fish or cephalopod tissue deteriorates under a constant linear or fixed-exponential rate pattern. However, the multi-component biological matrices of perishable aquatic products typically manifest high non-linearity due to the joint acceleration of progressive chemical oxidation, cellular autolysis, and exponential microbial proliferation^[39].

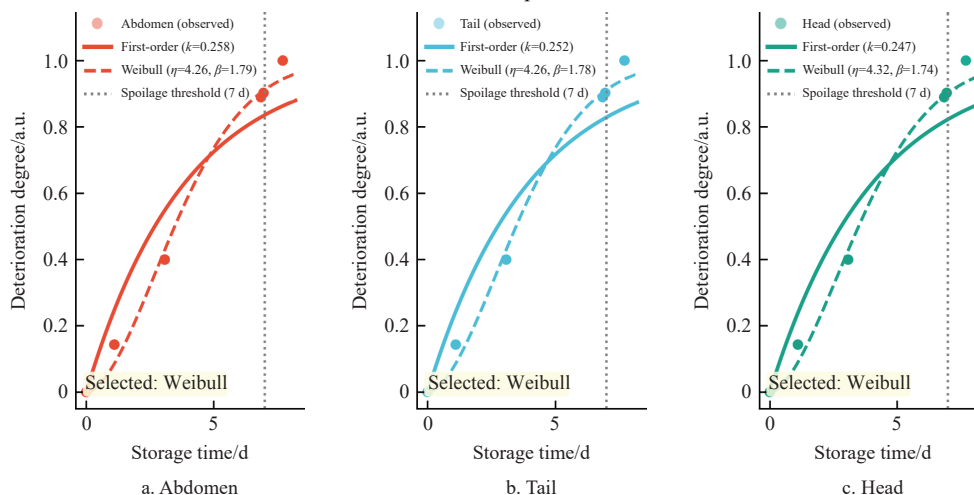


Figure 9 Kinetic degradation curves fitting of squid freshness under 4°C storage: First-order vs. Weibull model across different anatomical sites

The dynamic empirical degradation data alongside the synchronized mathematical fitting traces from both the first-order and Weibull kinetic frameworks across distinct anatomical sites are graphically elucidated in Figure 9. Visually, the classical first-order curves (dashed lines) exhibit severe systematic deviations, consistently underestimating the initial preservation plateau (Days 0-2) and failing to trace the steep, sigmoidal acceleration plunge during advanced storage (Days 5-7). In stark contrast, the Weibull kinetic profiles (solid lines) exhibit excellent mathematical flexibility, seamlessly capturing the distinct S-shaped curvature of real-time freshness deterioration across all targeted domains.

The statistical superiority of the Weibull kinetic framework across distinct anatomical sites was rigorously validated using the Akaike Information Criterion (AIC). As a penalized log-likelihood criterion, a lower AIC value designates an optimal trade-off

between the goodness-of-fit and mathematical parsimony of a model. The mathematical evaluation indicated that the Weibull model achieved a significantly lower AIC profile than the first-order kinetic paradigm across all investigated sampling regions. Specifically, the AIC values for the abdomen dropped from -1088.17 (first-order) to -1578.72 (Weibull), yielding a substantial delta value ($\Delta\text{AIC}=-490.6$). Similarly, the tail region exhibited a reduction from -233.26 to -338.57 ($\Delta\text{AIC}=-105.3$), while the head region shifted from -248.61 to -364.73 ($\Delta\text{AIC}=-116.1$). The markedly larger ΔAIC observed in the abdominal region mathematically implies that the fresh-keeping baseline of the visceral-adjacent tissue follows a highly pronounced, non-linear deterioration trajectory that cannot be accurately captured by simplified linear or first-order abstractions. This rigorous comparison forms a cornerstone theoretical foundation for spatial heterogeneity modeling in this study.

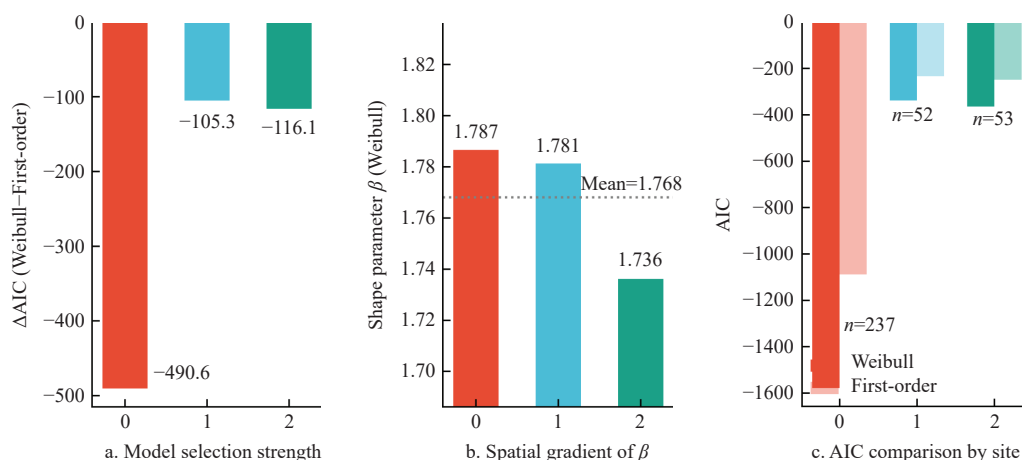


Figure 10 Spatial heterogeneity parameters analysis in deterioration mode

Figure 10 graphically quantifies the fine-grained spatial heterogeneity in the deterioration modes across the three major anatomical regions. As plotted in Figure 10a, the model selection strength is elucidated via the absolute value of $|\Delta\text{AIC}|$. Notably, the abdomen ($|\Delta\text{AIC}|=490.6$) presents a 4.6-fold and 4.2-fold stronger mathematical preference for the Weibull distribution compared to the tail ($|\Delta\text{AIC}|=105.3$) and the head ($|\Delta\text{AIC}|=116.1$), respectively. This sharp divergence confirms that the tissue closest to the viscera undergoes an accelerating, self-catalytic quality degradation profile. Figure 10b highlights the spatial gradient of the Weibull shape parameter (β). In food kinetic modeling, the shape parameter β dictates the curvature of the quality curve: $\beta=1$ implies a strict first-order constant rate, whereas $\beta>1$ represents a sigmoidal or down-curving acceleration where the degradation rate compounds over elapsed storage time^[40]. The calculated β values exhibited a distinct, descending spatial gradient: abdomen (1.787)>tail (1.781)>head (1.736). All computed β coefficients were consistently greater than 1.5, confirming a universally accelerating deterioration mode across the intact specimen. However, the abdomen's maximum β coefficient establishes it as the zone with the most volatile, non-linear kinetic acceleration. Figure 10c maps the absolute AIC baseline alongside the sample counts used during non-linear regression, ensuring full statistical transparency. Interestingly, although the three anatomical regions shared highly comparable Weibull scale parameters ($\alpha\approx 4.26$ -4.32 days, denoting the characteristic life or time scale where approximately 63.2% of the initial freshness state terminates), the immense variations in ΔAIC and the β gradient demonstrate that spatial heterogeneity manifests

as a structural difference in the deterioration mode rather than a mere baseline rate divergence.

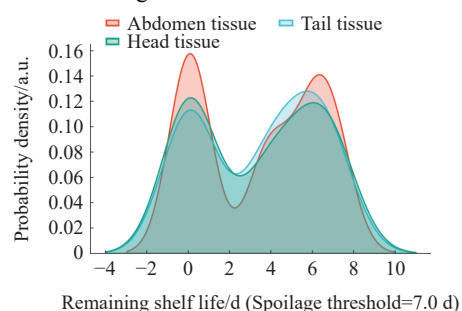


Figure 11 Predicted RSL distribution analysis

From a biochemical standpoint, this structural heterogeneity is heavily governed by the intense autolytic activity of endogenous enzymes (e.g., cathepsins and metalloproteinases) concentrated within the visceral mass during early storage^[41]. Upon death, these enzymes rapidly leak from the digestive tract into the adjacent abdominal muscle, breaking down structural proteins and producing low-molecular-weight volatile bases. This internal enzymatic attack occurs far earlier than external bacterial colonization on the head or tail skin. These results strongly challenge the conventional commercial assumption that a single, arbitrary measurement point can represent the freshness status of the whole aquatic product. Consequently, this emphasizes the scientific necessity of real-time monitoring at the abdomen, identifying it as the critical quality “bottleneck” for precision grading and digital cold-chain management.

3.6 Intelligent early warning via remaining shelf life (RSL) prediction

This section demonstrates the primary engineering advantage and industrial practicability of the developed handheld Vis-NIR framework: upgrading traditional retrospective “state monitoring” into proactive “future prediction” via fine-grained RSL distribution mapping. In cold-chain logistics and smart warehousing, a binary “pass/fail” or “fresh/spoiled” classification is inherently insufficient for dynamic inventory optimization. To address this, the real-time non-destructive spectral storage time prediction ($t_{\text{predicted}}$) outputted by the Stacking meta-learner is seamlessly coupled with the local Weibull degradation parameters derived in Section 3.3.

The remaining shelf life (RSL_i) for a specific anatomical region i is mathematically computed according to Equation (10):

$$\text{RSL}_i = T_{\text{threshold}} - t_{\text{predicted}} = T_{\text{threshold}} - \eta_i \cdot \left[-\ln \left(1 - \frac{f(t)_i}{a_i} \right) \right]^{\frac{1}{\beta_i}} \quad (10)$$

where, $T_{\text{threshold}}$ represents the absolute terminal spoilage baseline established at 7.0 days under 4°C refrigeration in strict accordance with the volatile base nitrogen (30 mg/100 g TVB-N) limits enforced by the national food safety standard GB 2733-2015. The variables η_i and β_i represent the location-specific Weibull scale and shape coefficients, respectively, ensuring that the structural spatial heterogeneity of the cephalopod specimen is fully compensated for during predictive scanning.

Figure 11 displays the continuous Kernel Density Estimation (KDE) distribution profiles of the predicted Remaining Shelf Life across the sampled population. The mathematical distribution profiles substantiate the intense spatio-temporal kinetic divergence discussed previously. The probability density curve for the abdominal region is severely left-shifted compared to that of the tail region, characterized by a significantly lower mean RSL value ($\text{Mean}_{\text{abdomen}}=1.25$ d versus $\text{Mean}_{\text{tail}}=3.12$ d at Day 4 of the cold trial).

Furthermore, the abdomen exhibits a considerably narrower, sharper bandwidth (higher peak kurtosis), indicating a tight statistical clustering with higher mathematical certainty and heightened urgency in quality decay. This rapid decay profile confirms that the abdomen acts as the thermodynamic and metabolic “accelerator” of the entire biological matrix.

From an industrial management perspective, this probabilistic RSL profiling provides a decisive digital solution for smart logistics sequencing. Rather than relying on rigid chronological batch routing, logistics managers can deploy this handheld system at critical transfer hubs to implement a highly precise First-Expired, First-Out (FEFO) inventory rotation strategy. Lots exhibiting highly compressed, sharp abdominal RSL distributions can be instantly prioritized for local processing or hyper-local delivery, while batches with wider, right-shifted tail RSL distributions can withstand extended long-distance transit. This proactive risk-mitigation framework successfully mitigates wholesale food waste and establishes a rigorous digital traceability pipeline for high-value marine aquatic product supply chains.

4 Conclusions

This study successfully developed a non-destructive, high-fidelity methodology for evaluating the spatiotemporal deterioration kinetics and predicting the remaining shelf life of squid freshness utilizing a handheld Vis-NIR spectroscopy framework coupled with advanced machine learning. The definitive conclusions are summarized as follows:

1) The established multi-scale feature space derived from Discrete Wavelet Transform (DWT-db4) successfully decoupled complex non-stationary spectral signals into multi-resolution sub-bands, effectively suppressing high-frequency instrumentation scattering and stochastic background ambient drift.

2) The proposed Stacked Ensemble Learning framework integrated the orthogonal mathematical advantages of both linear and non-linear baseline models. Guided by the Gradient Boosting Regressor (GBR) meta-learner, the model achieved optimal tracking performance for storage time with an R^2 of 0.9528, a minimized RMSE of 0.6566 d, and a top-tier RPD of 4.60, significantly outperforming monolithic baseline models and ensuring high robustness against structural variance.

3) Mathematical validation via the Akaike Information Criterion (AIC) rigorously demonstrated the structural superiority of the non-linear Weibull kinetic framework over traditional linear first-order models. Pronounced spatial heterogeneity was quantified across anatomical locations, identifying the abdominal muscle as the critical kinetic bottleneck experiencing the most volatile, accelerating degradation mode ($\beta=1.787$, $|\Delta\text{AIC}|=490.6$) due to internal enzymatic leakage from adjacent viscera.

4) By seamlessly coupling real-time spectral predictions with location-specific Weibull parameters, a probabilistic Remaining Shelf Life (RSL) profiling technique was successfully implemented. This transition from retrospective quality inspection to proactive distribution mapping provides a powerful digital logistics tool for enforcing precise First-Expired, First-Out (FEFO) warehouse rotation strategies, effectively mitigating wholesale supply-chain food waste.

Acknowledgements

This work was supported by the National Undergraduate Innovation Training Program (Grant No. A1-2005-00-300851).

[References]

- [1] FAO. The state of world fisheries and aquaculture 2024. Rome: Food and Agriculture Organization of the United Nations, 2024.
- [2] Tidwell J H, Allan G L. Fish as food: Aquaculture's contribution. *EMBO Reports*, 2001; 2: 958–963.
- [3] China Distant-water Fisheries Association. 2024 development report of China's distant-water squid jigging industry. Beijing: China Distant-water Fisheries Association. 2024. (in Chinese)
- [4] Goulas A E, Kontominas M G. Combined effect of light salting, modified atmosphere packaging and oregano essential oil on the shelf-life of sea bream (*Sparus aurata*): Biochemical and sensory attributes. *Food Chemistry*, 2007; 100(1): 287–296.
- [5] Gram L, Dalgaard P. Fish spoilage bacteria—problems and solutions. *Current Opinion in Biotechnology*, 2002; 13(3): 262–266.
- [6] Bekhit A E D A, Holman B W B, Giteru S G, Hopkins D L. Total volatile basic nitrogen (TVB-N) and its role in meat spoilage: A review. *Trends in Food Science & Technology*, 2021; 109: 280–302.
- [7] Wang H, He H J, Liu L, Ma H J, Liu X, Mo H Z, et al. Recent progress in hyperspectral imaging for nondestructive evaluation of fish quality. *Food Science*, 2019; 40(5): 329–338. (in Chinese)
- [8] Ocaño-Higuera V M, Maeda-Martínez A N, Marquez-Ríos E, Canizales-Rodríguez D F, Castillo-Yáñez F J, Ruiz-Bustos E, et al. Freshness assessment of ray fish stored in ice by biochemical, chemical and physical methods. *Food Chemistry*, 2011; 125(1): 49–54.
- [9] Ruiz-Capillas C, Moral A. Correlation between biochemical and sensory quality indices in hake stored in ice. *Food Research International*, 2001; 34(5): 441–447.
- [10] Özogul F, Polat A, Özogul Y. The effects of modified atmosphere packaging and vacuum packaging on chemical, sensory and microbiological changes of sardines (*Sardina pilchardus*). *Food Chemistry*, 2004; 85(1): 49–57.

- [11] Mawrides E K, Mishra A, Jæger B. Blockchain technology for sustainable supply chains in the fishing industry: A systematic review. *Cleaner Logistics and Supply Chain*, 2025; 17: 100277.
- [12] Lougovois V P, Kyranas E R, Kyranas V R. Comparison of selected methods of assessing freshness quality and remaining storage life of iced Gilthead sea bream (*Sparus aurata*). *Food Research International*, 2003; 36(5): 551–560.
- [13] Howgate P. A review of the kinetics of degradation of inosine monophosphate in some species of fish during chilled storage. *International Journal of Food Science and Technology*, 2006; 41(4): 341–353.
- [14] Yu Q, Zhang M, Mujumdar A S, Huang M. Real-time spoilage monitoring of aquatic products using advanced sensors and novel equipment. *Journal of Food Measurement and Characterization*, 2026; 20(9): 3829–3850.
- [15] GB 5009.228-2016. National food safety standard - Determination of volatile basic nitrogen in foods. Beijing: National Health and Family Planning Commission of the PRC, 2016. (in Chinese)
- [16] Tsoukalas D, Lørfall J, Jakobsen A N. Insights into the quality changes of edible sea urchin (*Echinus esculentus*) gonads during ice storage. *LWT*, 2024; 203: 116345.
- [17] Wu D, Sun D W. Advanced applications of hyperspectral imaging technology for food quality and safety analysis and assessment: A review — Part I: Fundamentals. *Innovative Food Science & Emerging Technologies*, 2013; 19: 1–14.
- [18] Kashani Zadeh H, Hardy M, Sueker M, Li Y C, Tzouchas A, MacKinnon N, et al. Rapid assessment of fish freshness for multiple supply-chain nodes using multi-mode spectroscopy and fusion-based artificial intelligence. *Sensors*, 2023; 23(11): 5149.
- [19] El Masry G, Wang N, El Sayed A, Ngadi M. Hyperspectral imaging for nondestructive determination of some quality attributes for strawberry. *Journal of Food Engineering*, 2007; 81(1): 98–107.
- [20] Hardy M, Moser B, Haughey S A, Elliott C T. Does the fish rot from the head? Hyperspectral imaging and machine learning for the evaluation of fish freshness. *Chemometrics and Intelligent Laboratory Systems*, 2024; 245(80): 105059.
- [21] Zhang Q Y, Hu Z G, Xu Z J, Zhang P L, Jiang Y J, Fu D D, et al. Quantitative determination of TVB-N content for different types of refrigerated grass carp fillets using near-infrared spectroscopy combined with machine learning. *Journal of Food Composition and Analysis*, 2024; 126: 105871.
- [22] Zhang Y J, Wang M P, Yang G H, Li J, Zhu G F, Shi J F, et al. Optical simulation model of the diffuse reflectance near-infrared spectroscopy for predicting fresh maize quality. *Int J Agric & Biol Eng*, 2025; 18(6): 250–259.
- [23] Cozzolino D. Near infrared spectroscopy and food authenticity. In: Espiñeira M, Santaclara F J (Eds.). *Advances in Food Traceability Techniques and Technologies*. Cambridge: Woodhead Publishing, 2016; pp.119–136.
- [24] Wu X D, Wang Z M, Wang Z Q, Zhang Q, Zhang Q X, Yan H B, et al. Prediction method of large yellow croaker (*Larimichthys crocea*) freshness based on improved residual neural network. *Journal of Food Measurement and Characterization*, 2024; 18(4): 2995–3007.
- [25] Wu S, Huang L F, Huang R S. Review of advances in hyper-spectral imaging-based non-destructive testing methods in food application. *Journal of Food Composition and Analysis*, 2026; 150: 108860.
- [26] Xu Z, Zhu S C, Wang W J, Liu S L, Zhou X X, Dai W L, et al. Rapid and non-destructive freshness evaluation of squid by FTIR coupled with chemometric techniques. *Journal of the Science of Food and Agriculture*, 2021; 102: 3000–3009.
- [27] Ahmed Z, Donkor O, Street W A, Vasiljevic T. Calpains- and cathepsins-induced myofibrillar changes in post-mortem fish: Impact on structural softening and release of bioactive peptides. *Trends in Food Science & Technology*, 2015; 45(1): 130–146.
- [28] Yu H D, Zuo S M, Xia G H, Liu X, Yun Y H, Zhang C. Rapid and nondestructive freshness determination of tilapia fillets by a portable near-infrared spectrometer combined with chemometrics methods. *Food Analytical Methods*, 2020; 13(10): 1918–1928.
- [29] Mallat S. A wavelet tour of signal processing: The sparse way. 3rd ed., Academic Press, 2009. doi: 10.1016/B978-0-12-374370-1.X0001-8
- [30] Peleg M, Normand M D, Corradini M G. The Arrhenius equation revisited. *Critical Reviews in Food Science and Nutrition*, 2012; 52(9): 830–851.
- [31] Xu G, Dong L Q, Kong L Q, Zhao Y J, Liu M, Hui M, et al. Parameters inversion algorithm of biological tissues based on a neural network model. *Acta Optica Sinica*, 2021; 41(11): 1117001. (in Chinese) doi: 10.3788/aos.202141.1117001
- [32] Chong W, Lyu W H, Zhang J, Liang J, Yang X T, Liu S, et al. Calibration system of sunshine duration recorder based on bi-xenon lamp source integrating sphere. *Acta Optica Sinica*, 2022; 42(1): 0112004. (in Chinese)
- [33] Gao Y W, Geng J F, Rao X Q. Non-invasive bruise detection in postharvest fruits and vegetables: A review. *Food Science*, 2017; 38(15): 277–287. (in Chinese)
- [34] Zhang Y L, Yang G H, Wang M P, Han Z Y, Zhu G F, Shi J F, et al. Factors affecting the non-destructive detection of water contents in fresh corn cobs by near-infrared spectroscopy. *Transactions of the CSAE*, 2024; 40(15): 262–270. (in Chinese)
- [35] Lanke N P, Chandak M B. Recent trends in deep learning and hyperspectral imaging for fruit quality analysis: An overview. *Journal of Optics*, 2025; 54: 1–15.
- [36] Bashkatov A N, Genina E A, Tuchin V V. Optical properties of skin, subcutaneous, and muscle tissues: A review. *Journal of Innovative Optical Health Sciences*, 2011; 4(1): 9–38.
- [37] Beć K B, Grabska J, Huck C W. Chapter Six - Handheld NIR spectroscopy for real-time on-site food quality and safety monitoring. *Advances in Food and Nutrition Research*, 2025; 115: 293–389.
- [38] Iqbal M, Yousaf J, Khan A, Muhammad T. IoT-enabled food freshness detection using multi-sensor data fusion and mobile sensing interface. *ICCK Transactions on Sensing, Communication, and Control*, 2025; 2(2): 122–131.
- [39] Aguilera-Barraza F A, Quevedo-León R A, López-Álvarez P X. Kinetics of protein and textural changes in Atlantic salmon under frozen storage. *Food Chemistry*, 2015; 182: 120–127.
- [40] Chen Q, Li J L, Yang H, Qian J P. A dynamic shelf-life prediction method considering actual uncertainty: Application to fresh fruits in long-term cold storage. *Journal of Food Engineering*, 2023; 349: 111471.
- [41] Kundu M, Vashisth A, Krishnan P. Instrumental detection of fish freshness in research and food industry: A review. *Journal of Food Science and Technology*, 2025; 63: 22–50.