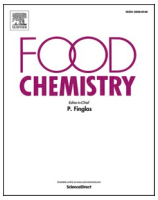
Food Chemistry 480 (2025) 143913

Contents lists available at ScienceDirect

Food Chemistry

journal homepage: www.elsevier.com/locate/foodchem



Analytical Methods

Applying deep learning algorithms for non-invasive estimation of carotenoid content in the foot muscle of Pacific abalone with different colors

Guijia Liu [a,b,1], Xiaoyong Wu [a,b,1], Yiming Wei [a,b], Tian Xu [a,b], Dongchang Li [d], Xuan Luo [a,b,c], Weiwei You [a,b,c,*], Caihuan Ke [a,b,*]

[a] State Key Laboratory of Mariculture Breeding, College of Ocean and Earth Sciences, Xiamen University, Xiamen 361102, China
[b] Fujian Key Laboratory of Genetics and Breeding of Marine Organisms, College of Ocean and Earth Sciences, Xiamen University, Xiamen 361102, China
[c] Abalone Research Center, Fujian Minruibao Marine Biotechnology Co., Ltd, Xiamen 361102, China
[d] Jinjiang Fuda Abalone Aquaculture Co., Ltd, Quanzhou 362251, China

ARTICLE INFO

Keywords:
Carotenoid
Zeaxanthin
Abalone
Non-invasive estimation
Deep learning
CIELAB
Metabolomics

ABSTRACT

Carotenoids are vital pigments influencing both the coloration and health of aquatic organisms, particularly in species such as the Pacific abalone (*Haliotis discus hannai)*. In this study, we identified the major carotenoids in abalone foot muscle using targeted metabolomics. Through differential metabolite analysis, we selected metabolites that met the following criteria: p-value <0.05, variable importance in projection (VIP) score ≥ 1, and fold change (FC) ≥ 2 or FC ≤ 0.5. The results showed that zeaxanthin had the highest content among all foot muscle colors, with the most significant p-value of 0.0079. Thus, we confirmed that zeaxanthin is the predominant carotenoid contributing to the distinct coloration of the foot muscle. We then used a deep learning model to predict carotenoid content based on color measurements in the CIELAB color space, defined by the Commission Internationale de l'Eclairage (CIE), which includes three dimensions: lightness (L^*), redness-greenness (a^*), and yellowness-blueness (b^*). Performance evaluation of 344 abalone samples showed that the Long Short-Term Memory (LSTM) model provided the best prediction results, with a root mean square error (RMSE) of 6.692 and a coefficient of determination (R^2) of 0.415. Furthermore, we developed the Color-Based Carotenoid Estimation Suite (CCES). This software features a user-friendly graphical interface, enabling users to input colorimetric data, train models, and predict carotenoid content. Compared to traditional methods, CCES offers non-destructive, rapid carotenoid estimation, improving efficiency by 450 times and reducing costs by 47 to 77 times. This method provides an efficient and scalable tool for aquaculture breeding and quality control, with applications extending beyond abalone to other aquatic and terrestrial species.

1. Introduction

Abalone is a highly valuable species in marine aquaculture, particularly in China, where it is considered a premium delicacy. According to the 2024 China Fisheries Statistical Yearbook, Fujian Province accounts for approximately 80 % of the nation's abalone production, making it a vital industry in the region. Among the various abalone species, the primary one cultivated in China is *Haliotis discus hannai*, commonly known as Pacific abalone, which is highly prized for its tender texture and superior nutritional content. This species has been the focus of most breeding programs in the region due to its economic significance and consumer demand for premium-quality abalone. Traditionally, abalone breeding programs have primarily focused on improving traits such as growth rate (Zhou et al., 2022), high-temperature tolerance (Meier et al., 2021), and food conversion efficiency (Yu et al., 2022). However, with evolving consumer preferences, breeding programs are increasingly focused on quality traits, including improved appearance and nutritional content.

Among these emerging quality traits, foot muscle color has become a critical factor influencing consumer decisions in the abalone market.

* Corresponding authors at: State Key Laboratory of Marine Environmental Science, College of Ocean and Earth Sciences, Xiamen University, Xiamen 361102, China.
E-mail addresses: wwyou@xmu.edu.cn (W. You), chke@xmu.edu.cn (C. Ke).
[1] Guijia Liu[1] and Xiaoyong Wu[1] equally contributed to this study.

https://doi.org/10.1016/j.foodchem.2025.143913
Received 22 November 2024; Received in revised form 6 March 2025; Accepted 14 March 2025
Available online 21 March 2025

The range of colors, spanning from white and green to vibrant orange, serves not only as a visual appeal but also as an indicator of nutritional value (Fig. 1). The orange hue is associated with a higher concentration of carotenoids, primarily zeaxanthin, which has been confirmed in *Haliotis gigantea* (Wei et al., 2019). However, the specific carotenoid types in Pacific abalone (*Haliotis discus hannai*) remain unreported. Given its connection to both market appeal and health benefits, carotenoid accumulation in abalone foot muscle is gaining importance in selective breeding programs aimed at improving product quality (Dall et al., 2019; Lehnert et al., 2019; Wade et al., 2017).

Despite the significance of carotenoids, traditional methods for quantifying their content—such as chemical extraction and ultra-performance liquid chromatography (UPLC-MS/MS)—are invasive and involve destroying the abalone, making them impractical for breeding programs (Sachindra et al., 2005; Xu et al., 2022; Zheng et al., 2010). This destructive process prevents the reuse of individuals for breeding, limiting the ability to improve these traits over generations. Given the growing demand for abalone with higher carotenoid content, there is a critical need for non-invasive methods that allow breeders to estimate carotenoid content without sacrificing valuable breeding stock.

In recent years, the development of deep learning and machine learning technologies has opened new possibilities for the non-invasive estimation of carotenoid content in aquaculture (Yang et al., 2021). Traditional methods such as ultra-performance liquid chromatography (UPLC), while precise, are invasive and time-consuming, making them unsuitable for large-scale applications (Dall et al., 2019; Guo et al., 2021; Vandeputte et al., 2020). In contrast, non-invasive techniques like near-infrared spectroscopy (NIR) and hyperspectral imaging (HSI) can quickly and non-destructively measure pigment content in plant and animal tissues, but their complex equipment requirements and data processing challenges limit their widespread use (Falcioni et al., 2024; Zhao et al., 2023; Sánchez et al., 2013). Chromaticity-based analysis methods have gained increasing attention in recent years for pigment content estimation due to their convenience and low cost (Bunluephan et al., 2023). Compared to more complex and expensive equipment such as NIR and HSI, colorimeters or image-based chromaticity devices are easier to operate and have lower maintenance costs. By measuring the L^*, a^*, and b^* chromaticity values of samples, these methods can quickly extract data related to pigment content (Itle & Kabelka, 2009), making them suitable for large-scale screening in aquaculture and agricultural breeding programs. Moreover, when combined with deep learning models, chromaticity data processing and analysis can achieve higher accuracy and automation, providing a cost-effective solution for non-invasive carotenoid estimation. Therefore, this study aims to develop a carotenoid estimation model based on chromaticity data and deep learning algorithms (see Fig. 1), providing a convenient, low-cost, and scalable method for estimating carotenoid content in the foot muscle of Pacific abalone.

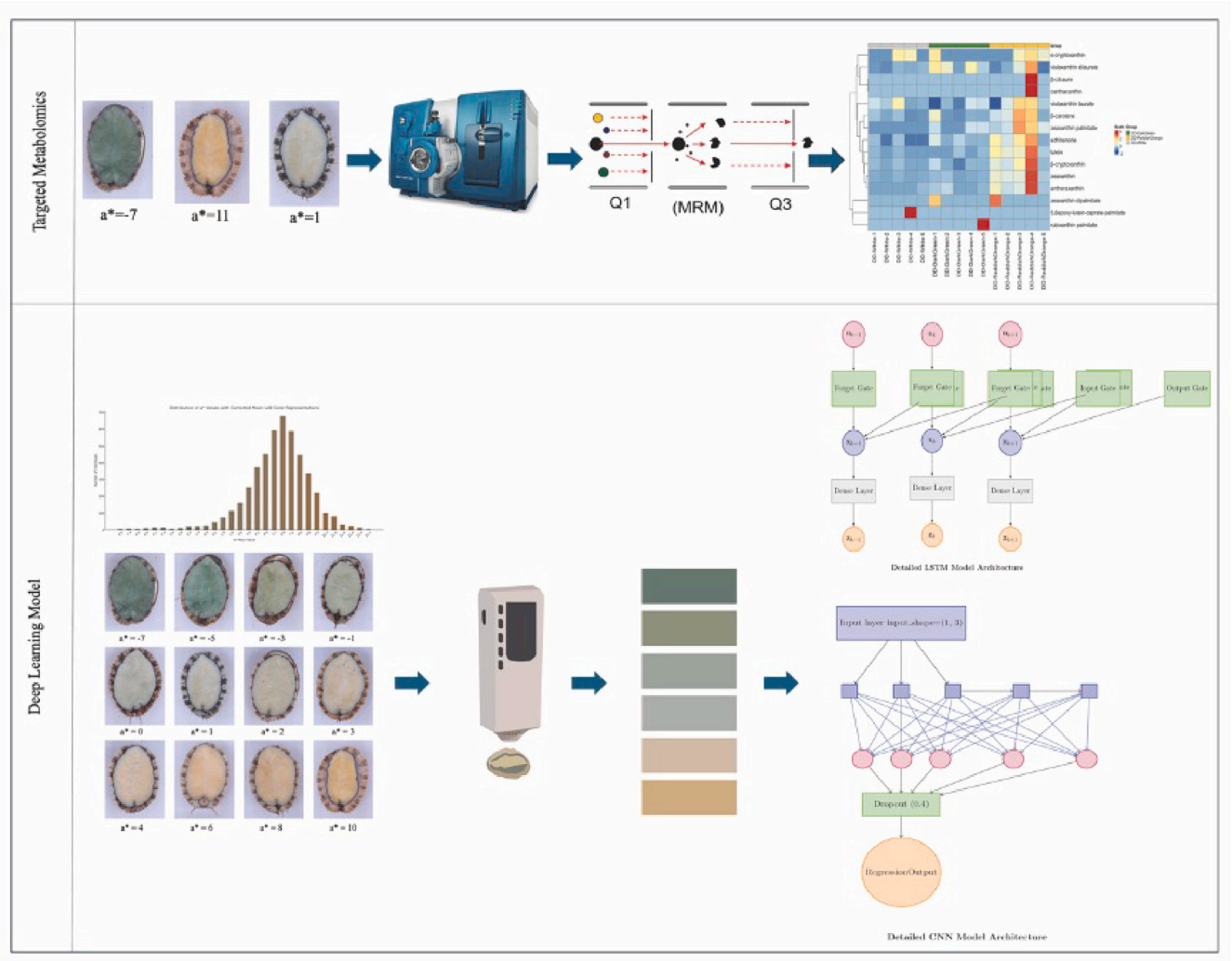


Fig. 1. The Full Process of Carotenoid Target Metabolite Identification and Non-Invasive Carotenoid Prediction Using Colorimetric Data and Machine Learning Models.

2. Materials and methods

2.1. Sample collection and preparation

The protocols for using animals in the present study were approved by the Committee on the Welfare and Ethics of Animal Experiments of Xiamen University. The population of Pacific abalone in our study came from a commercial breeding company, Fuda Abalone Aquaculture Co., Ltd., located in Jinjiang, China. A total of 344 samples of Pacific abalone, all 2 years of age, were collected for the study in April 2023. The study used 2-year-old abalones as samples, as they reach a commercially viable size at this age. This choice was made to ensure uniformity of the samples and to minimize potential confounding variables that could affect the results. The samples had an average body weight of 47.45 g, shell weight of 11.71 g, foot muscle weight of 21.93 g, shell length of 70.67 mm, and shell width of 48.32 mm.

2.1.1. Chromaticity measurement

Using a 3nh-NR110 color difference meter from Shenzhen 3nh Technology Co., Ltd., we assessed the chromaticity of the pedal sole (Porturas Olaechea et al., 1993), a specific part of the foot muscle, in the Pacific abalone samples. The CIELAB color space, also referred to as $L^*a^*b^*$, was defined by the Commission Internationale de l'Eclairage (CIE) in 1976. In this system, L^* represents lightness, with values ranging from 0 (black) to 100 (white), while a^* and b^* represent the color-opponent dimensions. The a^* value ranges from green ($-a^*$) to red ($+a^*$), and the b^* value ranges from blue ($-b^*$) to yellow ($+b^*$). This three-dimensional system allows for precise and objective measurement of color. Our investigation revealed a distribution pattern in the a^* values, with samples showing a^* values less than 0 corresponding to greener foot muscles—specifically, an a^* value of approximately −8 appeared dark green, −5 was green, and − 2 corresponded to an olive drab color—while those with a^* values greater than 0 appeared more orange, where an a^* value of 0 was white, 1 was light khaki, 3 was beige tan, 5 was yellow orange, 7 was orange, 9 corresponded to deep orange, 10 to light red orange, and 11 to reddish orange. The a^* values in this batch of Pacific abalone ranged from −8.45 to 12.86, indicating significant variation in foot muscle coloration across the sample population. This correlation between a^* values and observed colors provides a quantitative basis for categorizing foot muscle pigmentation in Pacific abalone.

2.1.2. Sample processing and storage

After recording the chromaticity data, the foot muscle of each abalone was dissected. The muscle tissues were freeze-dried for 48 h using a Labconco FreeZone 4.5 L freeze dryer. Once dried, the samples were ground into a fine powder using a Bionoon-24LD grinder from BIONOON Biotechnology Co., Ltd. in Shanghai. The resulting powder was stored at −80 °C for further analysis.

2.2. Carotenoid analysis

2.2.1. UV full-wavelength scanning

For the crude extraction of pigments, three representative abalone foot muscle samples were selected based on their a^* values: dark green ($a^* = -7$), white ($a^* = 0$), and orange ($a^* = 11$). Approximately 1 g of powdered foot muscle from each sample was extracted with 20 mL of ethyl acetate containing 0.1 % butylated hydroxytoluene (BHT) as an antioxidant. The samples were homogenized for 5 min, followed by centrifugation at 7000 rpm for 5 min to collect the supernatant. This extraction process was repeated three times to ensure complete extraction of the pigments. After extraction, 15 mL of the supernatant from each sample was transferred to a 6-well cell culture plate (Wuxi NEST Biotechnology Co., Ltd., China) and subjected to UV full-wavelength scanning using the TECAN Infinite M200 Pro microplate absorbance reader (Tecan Group Ltd., Switzerland). The scan covered wavelengths from 360 to 740 nm. Notably, the UV spectra revealed three characteristic absorbance peaks commonly associated with carotenoids. Based on these findings, targeted carotenoid detection was subsequently conducted to identify and quantify specific carotenoid compounds present in the samples.

2.2.2. Targeted carotenoid metabolomics

Following the identification of characteristic carotenoid absorbance peaks from the UV full-wavelength scan, targeted carotenoid metabolomics was conducted to specifically identify and quantify carotenoids present in the abalone foot muscle. The analysis involved the use of advanced chromatographic and mass spectrometric techniques to measure the concentrations of specific carotenoid compounds in the samples.

2.2.2.1. Chemicals and reagents. HPLC-grade solvents such as methanol, ethanol, acetonitrile, and other chemicals like BHT, acetone, and methyl tert-butyl ether (MTBE) were used for the analysis. Standard solutions were prepared at a concentration of 1 mg/mL in MTBE/MeOH and stored at −20 °C.

2.2.2.2. Sample preparation and extraction. Freeze-dried samples were ground into powder and stored at −80 °C. For extraction, 50 mg of the sample powder was mixed with n-hexane:acetone:ethanol (1:1:1, *v*/v/v). The mixture was vortexed, centrifuged, and the supernatant was collected. The residue was re-extracted using the same method, evaporated to dryness, reconstituted in dichloromethane, and filtered through a 0.22 μm filter for further LC-MS/MS analysis.

2.2.2.3. UPLC conditions. The extracts were analyzed using a UPLC-APCI-MS/MS system equipped with a YMC C30 column (2.1 × 100 mm, 3 μm). The mobile phase consisted of Solvent A (methanol:acetonitrile, 3:1, v/v, with 0.1 % formic acid and 0.1 % ammonium hydroxide) and Solvent B (methyl tert-butyl ether with 0.1 % formic acid and 0.1 % ammonium hydroxide). The gradient elution program was as follows: 0 min: 95 % A; 0–3 min: 95 % A;3–5 min: 80 % A; 5–9 min: 30 % A; 9–13 min: 30 % A;13–15 min: 95 % A. The flow rate was 0.8 mL/min, the column temperature was maintained at 28 °C, and the injection volume was 2 μL.

2.2.2.4. APCI-MS/MS conditions. The mass spectrometer was a QTRAP® 6500+ LC-MS/MS system, operating in positive ion mode. Carotenoids were detected using multiple reaction monitoring (MRM). The APCI source temperature was set at 350 °C, and parameters such as declustering potentials and collision energies were optimized for each transition. Data was acquired and quantified using Sciex software (Analyst 1.6.3 and Multiquant 3.0.3). Fig. 3d shows the integration and calibration results for carotenoids peaks across different samples, processed with MultiQuant 3.0.3 software to ensure accurate quantification. Retention time and peak shape were referenced from zeaxanthin standards to ensure accurate qualitative and quantitative analysis. The x-axis represents retention time (Time, min) and the y-axis represents ion intensity (Intensity, cps).

2.2.2.5. Data processing and differential analysis. Carotenoid contents were detected using the AB Sciex QTRAP® 6500+ LC-MS/MS platform, with the analysis performed by MetWare Biotechnology (Wuhan, China). Data acquisition and processing were conducted using Analyst 1.6.3 and MultiQuant 3.0.3 software (AB Sciex). The metabolite content data were processed using unit variance scaling (UV scaling), also known as *Z*-score standardization. This method standardizes data based on the mean (μ) and standard deviation (σ), resulting in a distribution with a mean of 0 and a standard deviation of 1, which facilitates comparison across different metabolites and samples.

The accumulation patterns of metabolites across different samples

were analyzed using Hierarchical Cluster Analysis (HCA). A heatmap displaying relative metabolite abundances was generated using the pheatmap package in R (version 4.0.2). In this heatmap, blue indicates lower accumulation and red indicates higher accumulation among the muscle color types: DD-White, DD-DarkGreen, and DD-ReddishOrange.

To identify the key metabolites responsible for the differences in foot muscle color, various statistical analysis methods were employed in this study. First, hypothesis testing was performed using the two-sided Wilcoxon test (Wilcoxon signed-rank test) (Goforth et al., 2025), and no FDR correction was applied for pairwise comparisons. The criteria for screening differential metabolites included fold change (FC) analysis (Liu et al., 2023), with thresholds set at FC $\geq$ 2 or FC $\leq$ 0.5, and the calculation of Log2FC to quantify the magnitude of the differences. When FC $\geq$ 2, Log2FC $\geq$ 1 indicates that the metabolite abundance in the experimental group is at least twice that in the control group (upregulated); when FC $\leq$ 0.5, Log2FC ≤ -1 indicates that the metabolite abundance in the experimental group is half or less than in the control group (downregulated). To further refine the selection of differential metabolites, Orthogonal Partial Least Squares-Discriminant Analysis (OPLS-DA) (Thévenot et al., 2015) was employed. The OPLS-DA model was analyzed using MetaboAnalystR 1.0.1 software (Thévenot et al., 2015), and metabolites with a variable importance in projection (VIP) score $\geq$ 1 were considered significant. Metabolites that met all these criteria were determined to be differential metabolites between the two muscle color types. Finally, Kyoto Encyclopedia of Genes and Genomes (KEGG) annotation and enrichment analysis were performed, mapping the selected metabolites to the KEGG compound database, and performing metabolite set enrichment analysis (MSEA) for pathway enrichment, with statistical significance determined by hypergeometric test *p*-values. These comprehensive analysis methods ensured that the selected differential metabolites were both reliable and biologically meaningful.

2.3. Quantitative analysis of zeaxanthin using HPLC

2.3.1. Chromatographic condition

Following the identification of zeaxanthin as the primary carotenoid in the abalone foot muscle, quantitative analysis was performed on all 344 samples using high-performance liquid chromatography (HPLC). A Thermo Scientific UltiMate 3000 UPLC system equipped with a BEH RP C18 column (2.1 $\times$ 100 mm, 1.7 μm) was utilized for the analysis. The mobile phase consisted of Solvent A—acetonitrile/methanol (7:3, *v*/v) containing 0.01 % butylated hydroxytoluene (BHT)—and Solvent B—0.1 % formic acid in water. The gradient elution program initiated with 95 % Solvent A at 0 min, then linearly increased from 95 % to 100 % Solvent A over 0 to 10 min. This composition was maintained at 100 % Solvent A from 10 to 15 min. Subsequently, the proportion of Solvent A was decreased from 100 % back to 95 % between 15 and 16 min, and finally, the system was held at 95 % Solvent A from 16 to 20 min. The flow rate was maintained at 0.2 mL/min throughout the run, with the column temperature set at 28 °C and an injection volume of 2 μL.

2.3.2. Quantification using zeaxanthin standard curve

Quantification of zeaxanthin in the samples was achieved by constructing a calibration curve using zeaxanthin standards (Sigma-Aldrich). A stock solution of 200 μg/mL was prepared by dissolving 1.0 mg of zeaxanthin in 5.0 mL of methanol, stored at −20 °C. Serial dilutions were made to obtain standard solutions at concentrations of 0.315, 0.375, 0.625, 0.75, 1.25, 1.5, 2.5, 3, 3.75, 5, 7.5, 10, 15, 20, 40, and 45 μg/mL.

The absorbance of the standards was measured at 450 nm under the same HPLC conditions. A calibration curve was plotted using the peak area versus the concentration of the standards. The zeaxanthin content in the abalone samples was calculated by comparing their peak areas to the calibration curve.

2.4. Development of non-invasive carotenoid prediction models

2.4.1. Evaluation of single-dimension models

Initially, several single-variable models were selected to individually fit each of the CIELAB color parameters (L^*, a^*, b^*) from the colorimeter to the carotenoid content. The models included linear (first-degree polynomial), quadratic (second-degree polynomial), cubic (third-degree polynomial), and quartic (fourth-degree polynomial) polynomial models, as well as models utilizing Softplus and ReLU activation functions. These models were used to separately fit the L^*, a^*, and b^* parameters. The performance of the models was evaluated using the following metrics: Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Coefficient of Determination (R^2), and Correlation. The evaluation of these single-dimension models will serve as a reference for the subsequent comparison with multi-dimension models.

2.4.2. Development of multi-dimension models

After evaluating the single-dimension models, multi-dimension models were used to fit all three $L^*a^*b^*$ parameters simultaneously. The multi-dimension models included Multiple Linear Regression, Convolutional Neural Networks (CNN), and Long Short-Term Memory Networks (LSTM). These models were applied to simultaneously consider the relationship between L^*, a^*, b^* parameters and the carotenoid content. The same performance metrics—MAE, MSE, RMSE, R^2, and Correlation—were used to evaluate the multi-dimension models. The comparison between single-dimension and multi-dimension models will provide insights into the relationship between $L^*a^*b^*$ parameters and carotenoid content.

2.4.3. Cross-validation and performance metrics

Throughout the fitting process of both single-dimension and multi-dimension models, 10-fold cross-validation was used to assess the performance of each model. Cross-validation involved dividing the dataset into multiple subsets, repeating the training and testing process several times to ensure the generalization ability and stability of the models. The evaluation of all models was based on the performance of the $L^*a^*b^*$ parameters, and their effectiveness in fitting the relationship with carotenoid content. This validation process ensures the robustness of the modeling methods for further analysis.

2.4.4. Software development and implementation

Based on the optimal prediction model, a non-invasive carotenoid prediction software was developed using Python 3.7. Deep learning frameworks such as TensorFlow or PyTorch were utilized for model training and prediction. The graphical user interface (GUI) was created with PyQt5, allowing users to input $L^*a^*b^*$ color data and receive predicted zeaxanthin content.

The software is compatible with Windows and macOS operating systems (with XQuartz required for macOS). Users can input chromaticity data through the GUI, and the software provides zeaxanthin content predictions based on the trained model. An option to retrain the model with new datasets or under different experimental conditions is also available.

3. Results

3.1. Carotenoid profiles in abalone foot muscle of different colors

3.1.1. UV full-wavelength scanning results

UV full-wavelength scanning was performed on foot muscle samples from abalone with different colors: orange, green, and white. As shown in Fig. 2, all samples exhibited characteristic carotenoid absorption peaks, with three distinct peaks located at approximately 428 nm, 450 nm, and 483 nm. These peaks are commonly associated with carotenoid compounds. The orange foot muscle sample demonstrated the highest

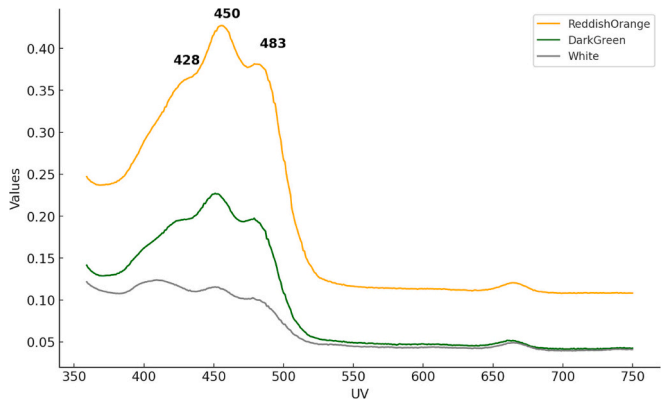


Fig. 2. UV Full-Wavelength Scan Results for Different Pedal Sole Colors of Abalone.

absorbance values, with a pronounced peak at 450 nm, followed by peaks at 428 nm and 483 nm. In comparison, the green foot muscle sample showed a lower but noticeable absorption pattern, while the white foot muscle sample displayed the weakest absorption across the same wavelength range. These results suggest that the orange-colored abalone has a higher carotenoid concentration, while the white-colored abalone has the lowest. The UV spectra confirm the presence of carotenoids across all foot muscle colors, with varying levels of concentration. This observation lays the foundation for further targeted metabolomics analysis to quantify the exact carotenoid composition and content in each sample group.

3.1.2. Targeted carotenoid metabolomics

3.1.2.1. Identification and quantification of carotenoids. A total of ten carotenoids were detected in the foot muscle of abalone, including zeaxanthin, *β*-carotene, antheraxanthin, violaxanthin dilaurate, lutein, violaxanthin laurate, *β*-cryptoxanthin, zeaxanthin palmitate, *α*-cryptoxanthin, and echinenone. Among these, the predominant pigment contributing to the orange coloration of the foot muscle was zeaxanthin. As shown in Fig. 3a, the concentration of zeaxanthin in the dried foot muscle of orange-colored abalone was 20.314 μg/g, significantly higher than the other carotenoids, all of which had concentrations below 0.5 μg/g. Comparing the zeaxanthin levels across the different foot muscle colors, the orange muscle contained the highest amount, followed by the green muscle, and the white muscle had the lowest amount. Specifically, zeaxanthin content in the orange muscle was 12 times higher than that in the white muscle and 7 times higher than that in the green muscle, while the green muscle had approximately double the zeaxanthin content of the white muscle.

The cluster analysis shows (see Fig. 3b) that carotenoid accumulation, such as zeaxanthin and *β*-carotene, is significantly higher in the DD-ReddishOrange group, with zeaxanthin and its palmitate ester being particularly abundant in this group compared to other samples. In contrast, the carotenoid accumulation in the DD-White group is relatively low, as indicated by the predominance of blue areas, reflecting lower levels of these metabolites in white muscle samples. The DD-DarkGreen group exhibits moderate levels of metabolite accumulation,

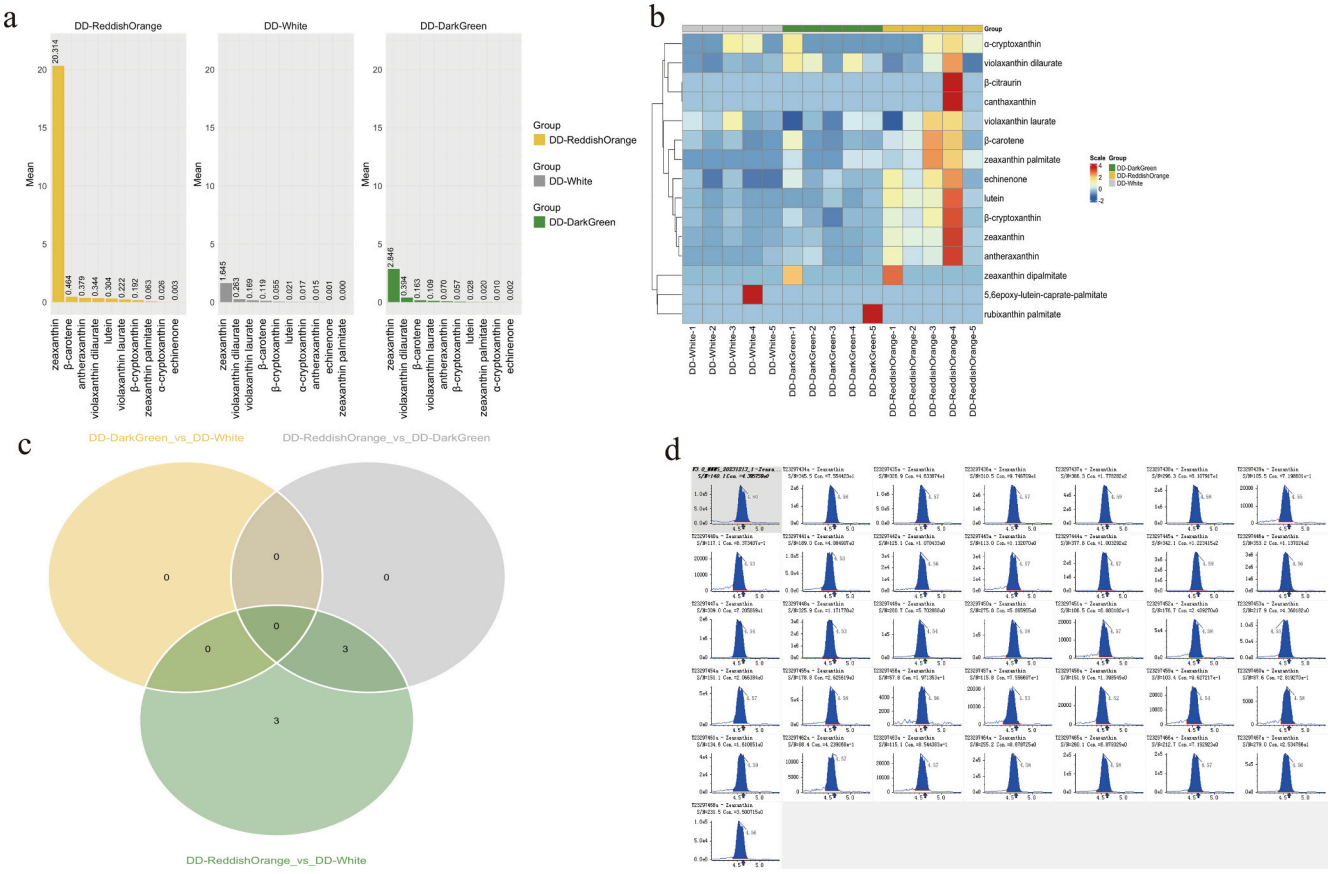


Fig. 3. Comparison of Carotenoid Content in Abalone Foot Muscle Across Different Pedal Sole Colors (Reddish orange, White, Dark green). (a) Bar charts showing the carotenoid content in foot muscle across three color groups (Reddish Orange, White, Dark Green). (b) Heatmap comparing carotenoid content across samples. (c) The Venn diagram shows the overlap and unique upregulated carotenoid metabolites across comparisons of reddish-orange, white, and dark green foot muscles. (d) Chromatographic peak integration and calibration for zeaxanthin across different samples. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

with certain compounds, such as violaxanthin dilaurate, showing higher abundance in green samples. Overall, there is a notable correlation between muscle color and carotenoid content across different samples.

3.1.2.2. Carotenoid differences in foot muscles of different colors. By analyzing the carotenoid metabolites in foot muscles of different colors in Pacific abalone, we first compared the differences between the red foot muscle and the white foot muscle (see Fig. 4). In the comparison of DD-ReddishOrange vs DD-White, a total of six upregulated carotenoid metabolites were identified: zeaxanthin, antheraxanthin, lutein, β-cryptoxanthin, β-carotene and echinenone. All of these metabolites had p-values less than 0.05, VIP values greater than or equal to 1, and Fold Change (FC) ≥ 2, indicating significant differences in the metabolites between the two foot muscle colors. Notably, zeaxanthin had the highest content (Fig. 3a) among all foot muscle colors, and its p-value in the comparison between the reddish-orange foot muscle and the white foot muscle was 0.0079 (Fig. 4a), showing a significant difference. Next, we compared the differences between the reddish orange foot muscle and the dark green foot muscle (see Fig. 4a). In the comparison of DD-ReddishOrange vs DD-DarkGreen, a total of three upregulated carotenoid metabolites were identified, including zeaxanthin, lutein and β-cryptoxanthin, which all had significantly increased content in the reddish-orange foot muscle, with p-values also showing significance. Zeaxanthin again had the most significant p-value of 0.0079, further supporting its key role in foot muscle color changes.

Finally, we compared the differences between the dark green foot muscle and the white foot muscle. In the comparison between DD-DarkGreen and DD-White, no significant differential carotenoid metabolites were identified, indicating that there were no notable changes in carotenoid content between the green and white foot muscles. Fig. 4b shows the OPLS-DA model's ability to distinguish different foot muscle colors based on targeted carotenoid content. The results of the three group comparisons are as follows: RedishOrange vs White with $R^2Y = 0.861$ and $Q^2 = 0.50$; RedishOrange vs DarkGreen with $R^2Y = 0.866$ and $Q^2 = 0.655$; DarkGreen vs White with $R^2Y = 0.961$ and $Q^2 = 0.577$. The R^2Y and Q^2 values for each group indicate that the model has strong explanatory power and predictive ability. Overall, the results in the figure demonstrate that the metabolomic data effectively differentiate between foot muscle color groups, and the established model has high predictive ability and stability.

Therefore, zeaxanthin, as a key metabolite, especially in the reddish-orange foot muscle, has a significantly higher content than in the other foot muscle colors, confirming its importance in foot muscle color changes and further illustrating its central role in carotenoid accumulation in abalone foot muscle. The Venn diagram (see Fig. 3c) illustrates the overlap and uniqueness of carotenoid metabolites between different foot muscle colors, highlighting that the reddish-orange foot muscle has significantly higher levels and a distinct composition of carotenoid metabolites compared to other foot muscle colors, which is also the primary color for our future breeding selection.

(a) Target carotenoid differential metabolites were classified as 'up' if they simultaneously met the following criteria: fold change (FC) ≥ 2, Log2FC ≥ 1, p-value ≤ 0.05, and variable importance in projection (VIP) score ≥ 1. Target carotenoid differential metabolites were classified as 'down' if they simultaneously met the following criteria: FC ≤ 0.5, Log2FC ≤ -1, p-value ≤ 0.05, and VIP score ≥ 1. Otherwise, they were classified as 'non-significant.' (b) The permutation testing (right) evaluates the stability and predictive power of the model, with R2Y and Q2 values indicating model performance.

In the comparison between DD-ReddishOrange and DD-White, based on the differential analysis, we annotated and identified three major metabolic pathways—retinol metabolism, metabolic pathways, and biosynthesis of cofactors. Multiple metabolites involved in these pathways were upregulated, as indicated by the red markers in the pathway map (Fig. 5a). In the comparison between DD-DarkGreen and DD-White, most metabolites showed no significant changes (indicated by blue), as shown in Supplementary Fig. 1. A further classification of these metabolic pathways is provided in Fig. 5b, where the percentage of metabolites involved in each pathway category is displayed. Among the significantly enriched pathways, metabolic pathways showed the highest percentage of enriched metabolites (100 %), followed by retinol metabolism and biosynthesis of cofactors (both at 25 %).

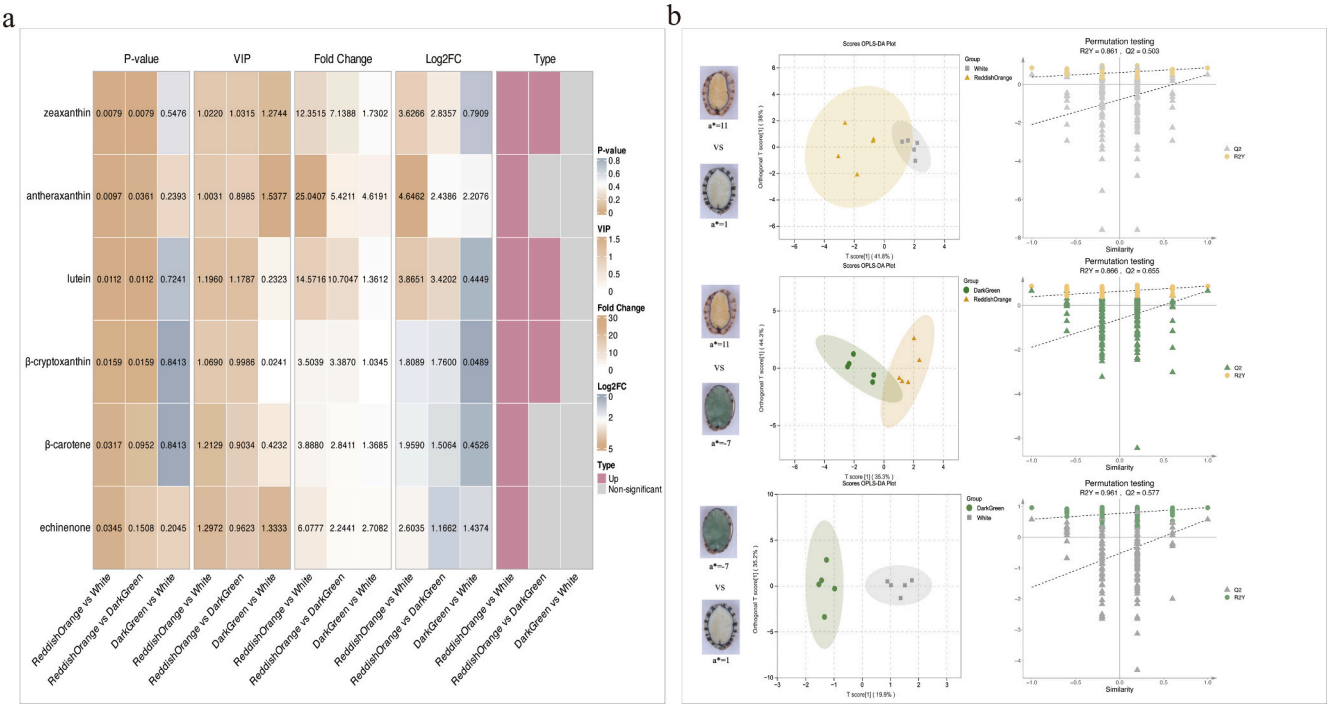


Fig. 4. Differential Analysis of Targeted Carotenoid Metabolites and OPLS-DA Classification Between Foot Muscle Color Groups of Pacific Abalone.

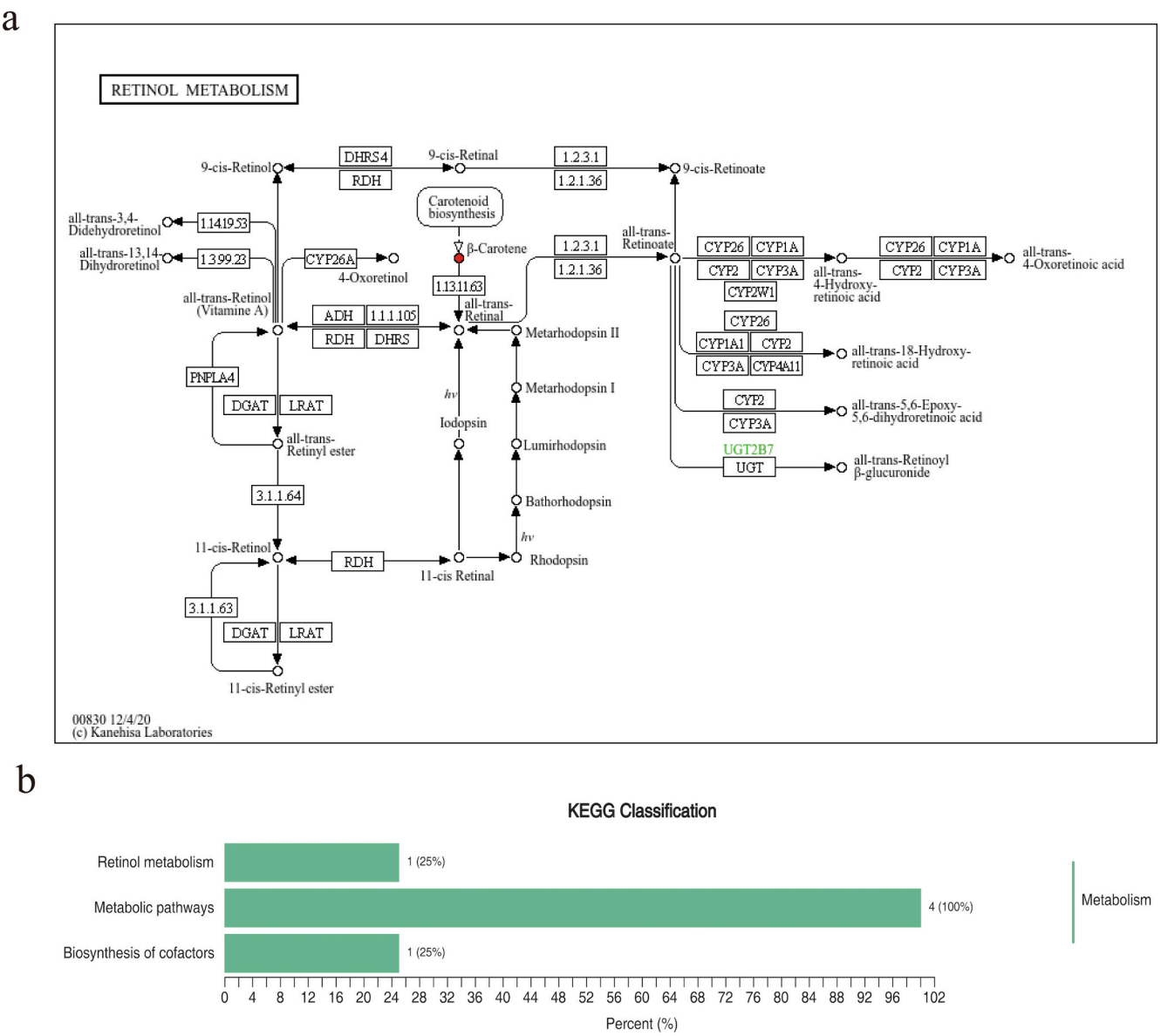


Fig. 5. KEGG Pathway Enrichment Analysis of Differential Metabolites between DD-ReddishOrange and DD-White Foot Muscles. (a) Pathway map highlighting upregulated metabolites in the DD-ReddishOrange foot muscle group, with red marking increased activity in pathways such as retinol metabolism and biosynthesis of cofactors. (b) Bar chart classifying the enriched pathways, showing the percentage of metabolites involved in each. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

3.1.3. Liquid chromatography

To accurately quantify the zeaxanthin content in all 344 abalone foot muscle samples, we employed high-performance liquid chromatography (HPLC) as described in Section 2.3.1. A zeaxanthin standard solution at a concentration of 50 μg/mL was prepared to establish a calibration curve and determine the compound's retention time under the specified chromatographic conditions. The standard exhibited a retention time of 6.847 min.

Each abalone foot muscle sample was individually processed and analyzed using HPLC. The chromatograms consistently showed a major peak corresponding to the retention time of the zeaxanthin standard, confirming the presence of zeaxanthin in all samples. Representative chromatograms displayed main peaks at retention times around 6.6 min (Fig. 6a), closely matching the standard and indicating reliable detection across the entire sample set.

We measured the area under the curve (AUC) for the zeaxanthin peak in each sample's chromatogram. Using the established calibration curve ($y = 1.3685\times + 0.6003$, $R^2 = 0.9985$; Fig. 6b), we calculated the zeaxanthin concentration in each sample. The calibration curve demonstrated excellent linearity over the concentration range of 0.315 μg/mL to 45 μg/mL, ensuring accurate quantification.

The zeaxanthin content among the 344 samples varied significantly, ranging from 0.03 μg/g to 60.25 μg/g on a dry weight basis. This wide range reflects the diverse foot muscle coloration observed in the samples.

3.1.4. Carotenoid concentration in different foot muscle colors

To account for samples with very low zeaxanthin concentrations, absolute values were taken to ensure that even minimal concentrations were represented accurately. The final zeaxanthin concentrations for all 344 abalone samples are provided in the Abs_Carotenoid_Content_μg/g column in Supplementary Table 1. The concentrations vary widely, ranging from 0.03 μg/g to 60.25 μg/g, indicating significant differences across the samples. Fig. 7 illustrates the relationship between chromaticity values (L^*, a^*, and b^*) and carotenoid content in 344 abalone foot muscle samples. Fig. 7a shows significant variation in the distribution of

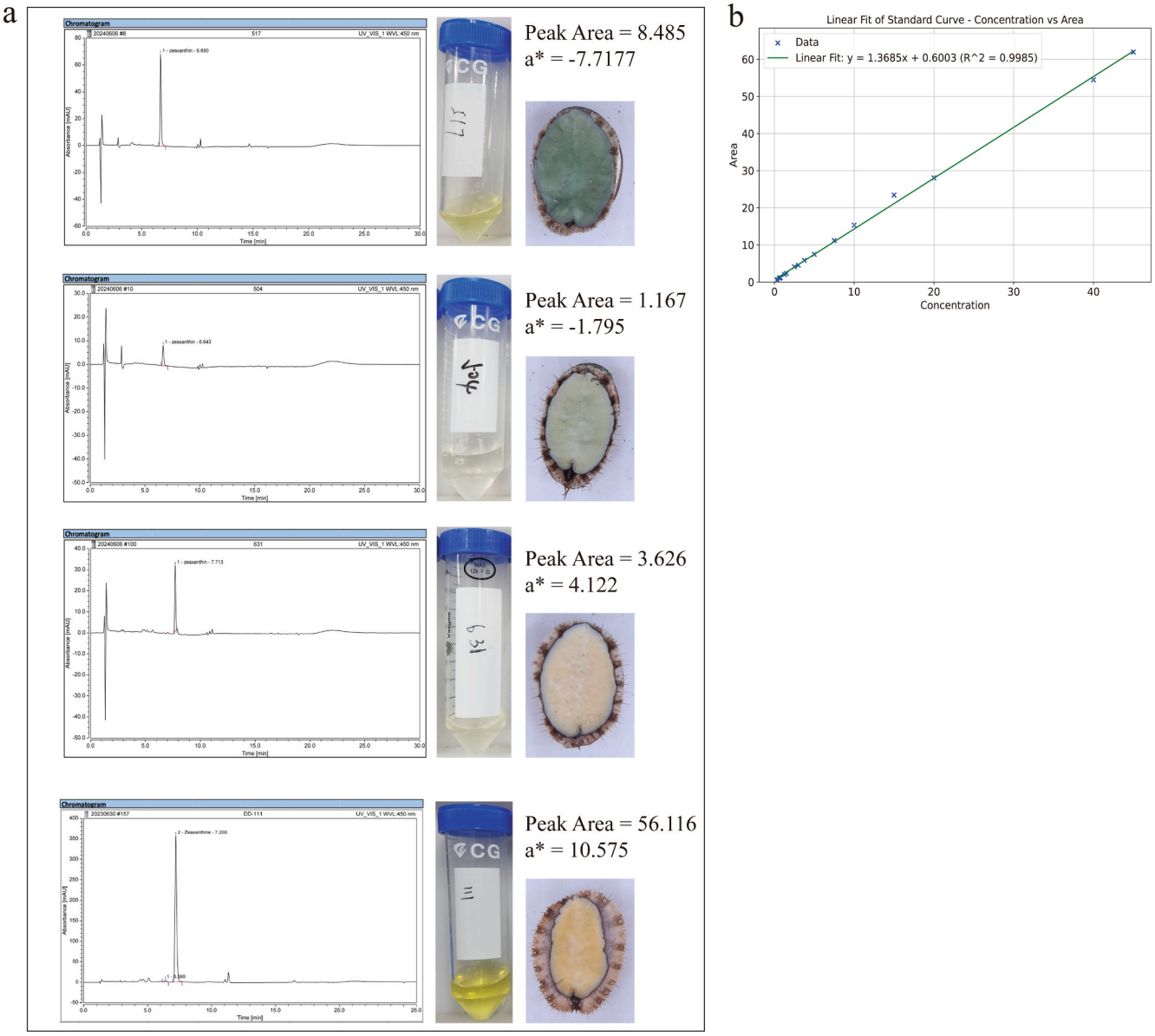


Fig. 6. Comprehensive Visualization of Carotenoid Quantification: HPLC Analysis, Foot Muscle Coloration, and Zeaxanthin Standard Curve. (a) High-performance liquid chromatography (HPLC) analysis quantifying zeaxanthin content in abalone foot muscles categorized into 11 distinct color groups based on a^* values, ranging from dark green to reddish orange. (b) Zeaxanthin standard curve. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

L^*, a^*, and b^* values, reflecting the diversity in the color of the abalone foot muscles, providing valuable data support for subsequent non-destructive models. Fig. 7b demonstrates a positive correlation between L^* (lightness) and carotenoid content, with higher lightness values generally corresponding to higher carotenoid concentrations. Fig. 7c indicates that a^* values (red-green axis) are also associated with carotenoid content, especially in the higher a^* range (redder), where carotenoid levels are elevated. Fig. 7d shows that b^* values (yellow-blue axis) correlate with carotenoid content, with higher b^* values (more yellow) linked to increased carotenoid concentrations. These results suggest that color variations in abalone foot muscles are closely related to carotenoid accumulation, with lighter, more reddish, and more yellowish foot muscles typically having higher carotenoid content. This is also validated in Fig. 8a. To improve the model's ability to predict carotenoid content related to foot muscle color, we aimed to include a wide variety of colors found in the Pacific abalone population, ranging from green to white to orange. As shown in Fig. 7a, the frequency distribution of a* values capture this range of colors. This approach ensures the model can better fit the data, thereby improving its accuracy and generalizability.

3.2. Non-invasive carotenoid prediction model selection

3.2.1. Comparison of single-trait and multi-trait models

We tested nine models to predict carotenoid content in abalone foot muscle using chromaticity values (L^*, a^*, and b^*). Among them, two were deep learning algorithms (CNN and LSTM), two were models based on non-linear activation functions (Softplus and ReLU), and the remaining were linear models, including polynomial regression and multiple linear regression. The performance of these models was evaluated using metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), correlation coefficients, and the coefficient of determination (R^2) values. In Fig. 8, we compared the predictive performance of single-dimension models, where each

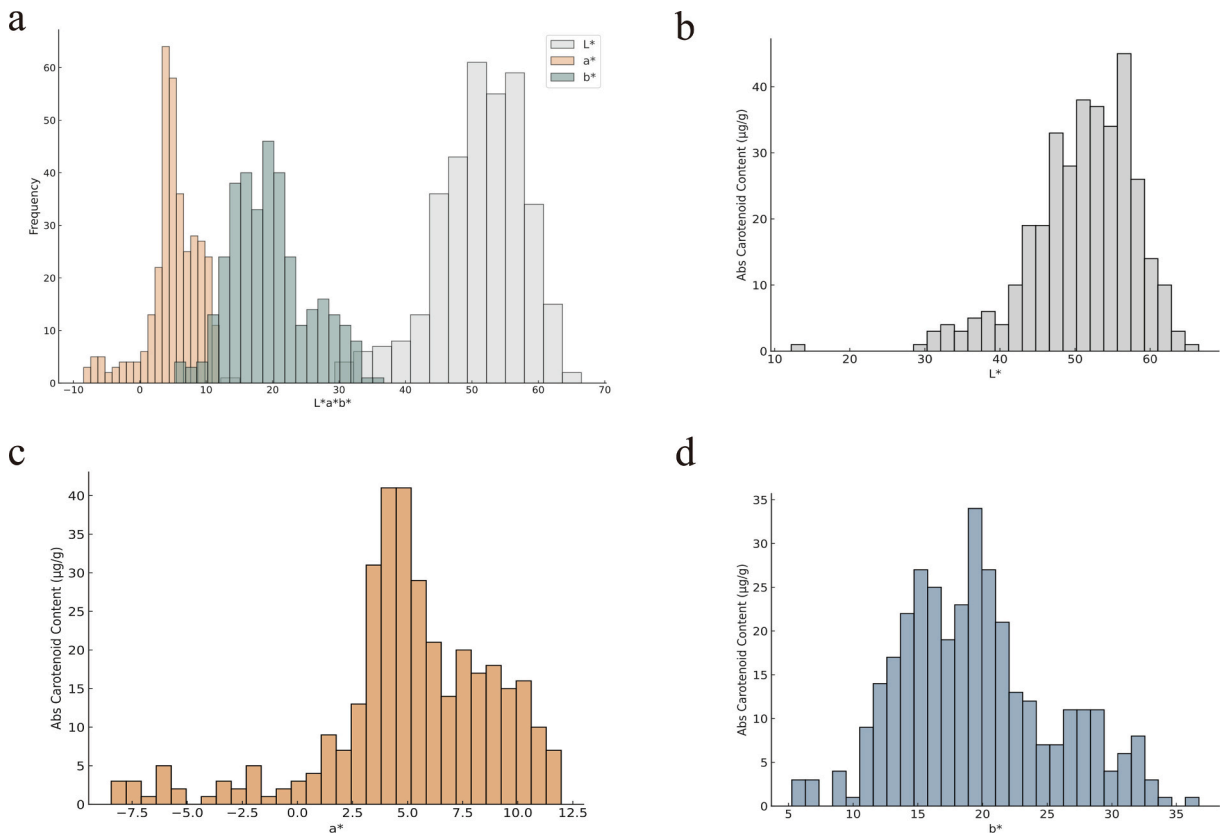


Fig. 7. Distribution of Carotenoid Content and Chromaticity (L^*, a^*, and b^*) in Abalone Foot Muscle Samples. (a) Frequency Distribution Histograms of L^*, a^*, and b^* Chromaticity in 344 Samples. (b) Frequency Distribution of L^* Chromaticity Values and Corresponding Carotenoid Content. (c) Frequency Distribution of a^* Chromaticity Values and Corresponding Carotenoid Content. (d) Frequency Distribution of b^* Chromaticity Values and Corresponding Carotenoid Content.

chromaticity value (L^*, a^*, or b^*) was considered individually, and multi-dimension models, which incorporated all three chromaticity dimensions ($L^*a^*b^*$) simultaneously. This comparison allowed us to assess how well models perform when using individual chromatic dimensions versus combining multiple dimensions for prediction. The single-dimension models included linear (first-degree polynomial), quadratic (second-degree polynomial), cubic (third-degree polynomial), quartic (fourth-degree polynomial) models, as well as Softplus and ReLU. The multi-dimension models included CNN, LSTM, and Multiple Linear Regression. Fig. 8b shows the comparison between the single L^* dimension and the multi-dimension $L^*a^*b^*$; Fig. 8c compares the single a^* dimension and the multi-dimension $L^*a^*b^*$; and Fig. 8d compares the single b^* dimension and the multi-dimension $L^*a^*b^*$. In Fig. 8a, the a^* shows the highest correlation with carotenoid content, and the scatter plot reveals a clear positive relationship between a^* and measured carotenoid content. This trend is not observed in the L^* or b^* dimensions, where the relationship with carotenoid content is weaker. This result is further confirmed in the model comparisons (Fig. 8b-d), where models using a^* as input consistently outperform those using L^* or b^*.

This strong relationship between a^* and carotenoid content translates into better overall performance for models based on the a* dimension. For example, in the Softplus model (Fig. 8c), the correlation coefficient for a^* is 0.67, while the coefficients for b^* and L^* are 0.57 and 0, respectively. In the multi-dimension models, LSTM and CNN, which integrate the $L^*a^*b^*$ dimensions, slightly outperform most single-dimension models, though in some cases, single-dimension models still achieve better results. Notably, in Fig. 8c, the single-dimension cubic polynomial model (third-degree polynomial) outperforms the multi-dimension models—CNN, LSTM, and Multiple Linear Regression—in several metrics, including correlation, MSE, RMSE, and R^2. For instance, the cubic polynomial model achieves a correlation coefficient of 0.69, which is slightly better than the 0.68 of the multi-dimension CNN model. However, despite its strong performance, the cubic polynomial model may be overfitting.

To address the potential overfitting issue, we subsequently conducted a 10-fold cross-validation on both the a^* single- dimension models and the $L^*a^*b^*$ multi-dimension models. This helped to further evaluate the models' generalization ability, ensuring that the selected model performs well not only on the training data but also on unseen data. This validation process is essential for determining the optimal model for non-invasive prediction of carotenoid content.

3.2.2. *Robustness testing of a* and multi-trait models*

After 10-fold cross-validation, the results (see Table 1) showed that the LSTM model performed best across four key indicators: MAE, MSE, RMSE, and R^2. The reported values are averages over the 10 folds. Specifically, the LSTM model achieved the lowest MAE of 4.079 among all models, indicating the smallest average difference between its predicted values and the actual measurements. Its MSE was 49.198 and RMSE was 6.692, both the lowest values, further demonstrating its high predictive accuracy. In terms of the coefficient of determination, the LSTM model attained the highest R^2 value of 0.415, close to 1, indicating a high degree of fit between the model and the data.

Although the third-degree polynomial model showed the best predictive results on the entire sample data, its performance significantly declined during the 10-fold cross-validation, exhibiting overfitting. The model's mean MAE increased to 5.292, the mean MSE sharply rose to 342.756, the mean RMSE was 11.951, and the mean R^2 even dropped to a negative value (−11.811). These results indicate that the model's predictive performance on new data significantly decreased, with poor robustness, making it unsuitable for practical applications.

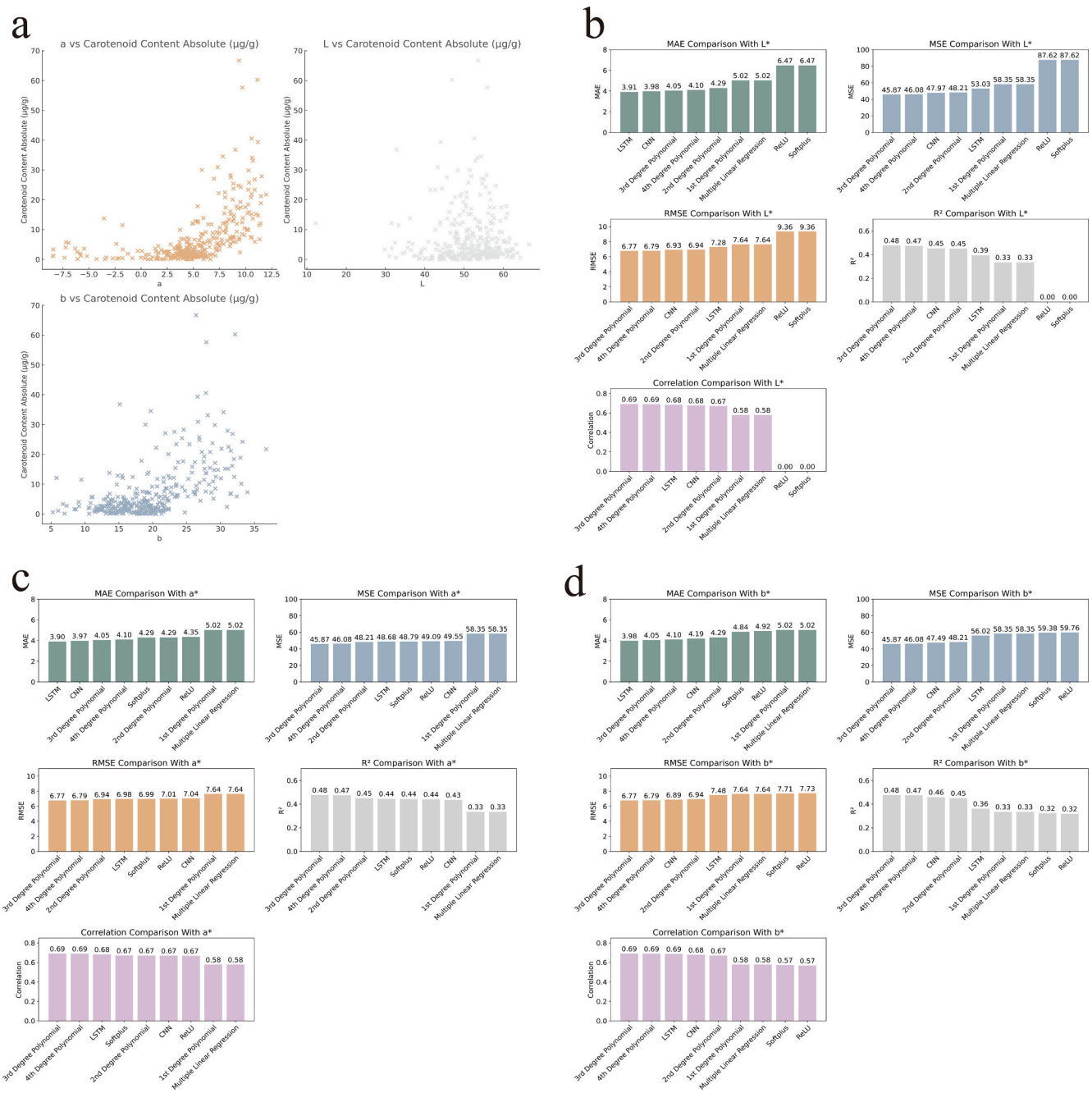


Fig. 8. Performance Comparison of Single-Dimension and Multi-Dimension Models in Carotenoid Content Prediction. (a) Scatter plots showing the relationship between individual chromaticity dimensions (L^*, a^*, b^*) and carotenoid content. (b-d) Model performance comparison using individual chromaticity dimensions (L^*, a^*, b^*) versus the combined $L^*a^*b^*$ dimensions.

Table 1
Average Performance Metrics of Different Models in 10-Fold Cross-Validation Predicting Carotenoid Content.

Model	Input Features	MAE	MSE	RMSE	R^2	Correlation
First-degree polynomial model	a^*	5.078	59.671	7.354	0.306	0.611
Second-degree polynomial model	a^*	4.589	61.358	7.447	−0.0337	0.667
Third-degree polynomial model	a^*	5.292	342.756	11.951	−11.811	0.636
Fourth-degree polynomial model	a^*	10.247	9310.093	38.421	−388.889	0.508
Softplus	a^*	4.324	49.962	6.732	0.404	**0.704**
ReLU	a^*	4.394	50.409	6.765	0.397	0.702
Multiple Linear Regression	$L^*a^*b^*$	5.078	59.671	7.354	0.306	0.611
CNN	$L^*a^*b^*$	4.169	50.410	6.800	0.385	0.690
LSTM	$L^*a^*b^*$	**4.079**	**49.198**	**6.692**	**0.415**	0.698

Note: The best model's values are highlighted in bold. MAE, Mean Absolute Error; MSE, Mean Squared Error; RMSE, Root Mean Squared Error; R^2, Coefficient of Determination; Correlation, Pearson correlation coefficient between predicted and actual values.

The CNN model, while slightly inferior to the LSTM model, still outperformed models based on a single chromaticity value. The CNN model's mean MAE was 4.169, mean MSE was 50.410, mean RMSE was 6.800, and mean R^2 was 0.385. These indicators show that the CNN model has good performance in predictive accuracy and model fitting and is superior to models that use only single-dimensional data. Therefore, the CNN model is also a reliable choice.

Based on the above analysis, we selected the LSTM and CNN models to develop user-friendly software for the non-invasive and high-precision prediction of carotenoid content in Pacific abalone. Since the LSTM model performed best in multiple key indicators and the CNN model also outperformed single-dimension models with good generalization ability, these two models were determined to be the optimal choices for subsequent software development and practical applications.

3.2.3. *Software development and implementation*

We have developed a Color-Based Carotenoid Estimation Suite (CCES) with a graphical user interface that includes three main modules: CCES-Train_app.py, CCES-Sample_app.py, and CCES-Predict_app.py, simplifying the operational process. Users can train Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM) models through simple clicks using the CCES-Train module, generating .h5 model files for carotenoid content prediction. The CCES-Sample module allows for random sampling from actual datasets for testing and validation, ensuring the model's accuracy and reliability. With the CCES-Predict module, users only need to input $L^*a^*b^*$ chromaticity values to quickly estimate the carotenoid content in samples based on the trained model, achieving non-destructive and precise measurement.

The suite's user-friendly interface design enhances research efficiency and is suitable for various experimental data and breeding needs. It can be applied not only to carotenoid research in abalone but also extended to other aquatic and terrestrial species, offering broad application prospects. As shown in Fig. 9, Fig. 9a demonstrates rapid prediction for a single sample; Fig. 9b displays the batch input function for multiple samples; and Fig. 9c presents an intuitive output of batch sample prediction results. Detailed introduction and software downloads are available on GitHub (https://github.com/heidi-liu/CCESuite). The software is limited to non-commercial use, providing flexibility and facilitating further customization and adaptation to different species and datasets.

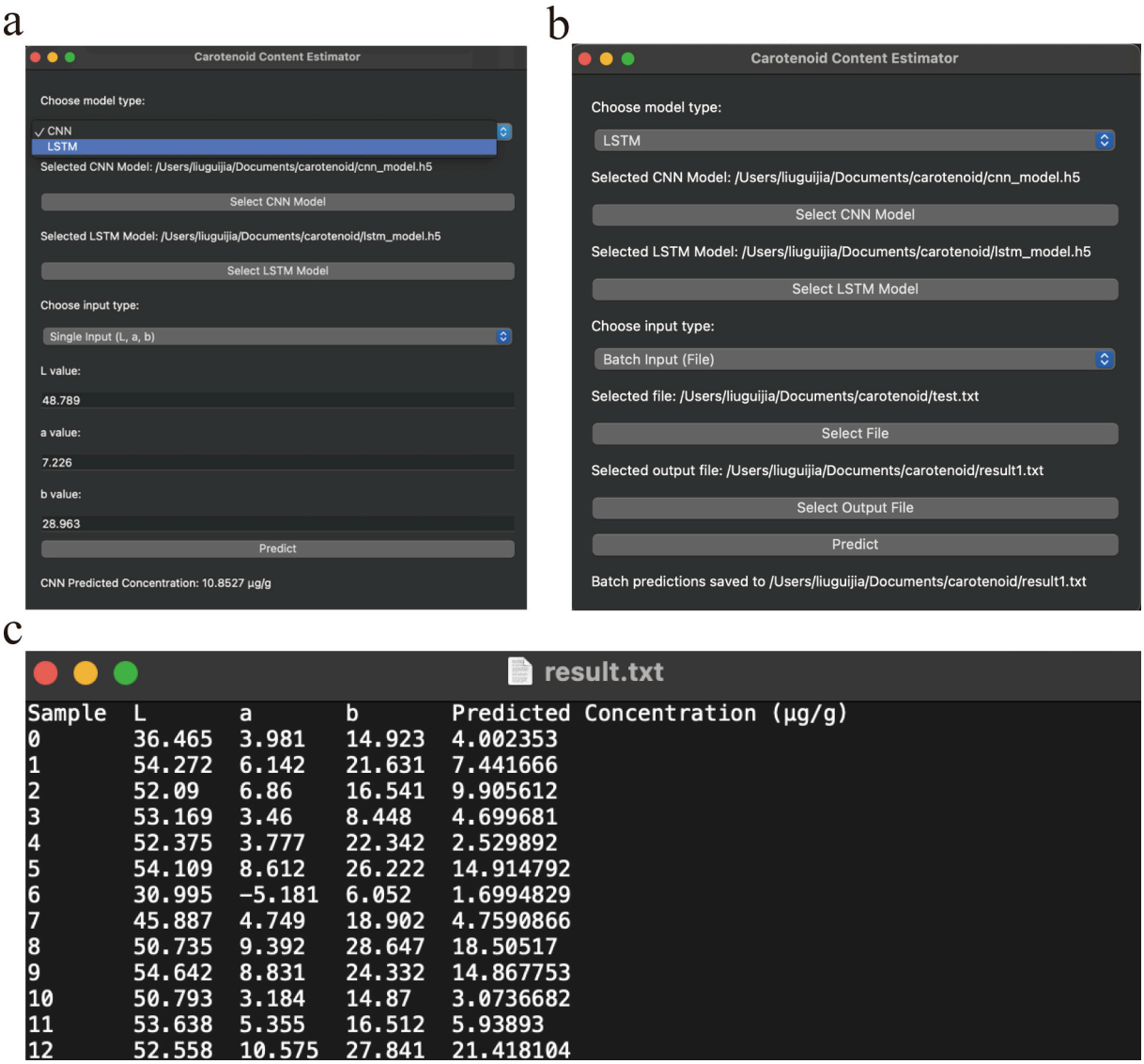


Fig. 9. Graphical Interface and Batch Prediction Features of the CCES Software. (a) Single-sample carotenoid content prediction interface using L^*, a^*, and b^* values. (b) Batch prediction interface for estimating carotenoid content from multiple input samples. (c) Output file showing the predicted carotenoid concentrations for multiple samples in a tabular format.

4. Discussion

In this study, we developed a non-invasive model that rapidly predicts carotenoid content in the foot muscle of Pacific abalone by integrating L^*, a^*, and b^* chromaticity values with deep learning algorithms. Compared to traditional UPLC methods, this model demonstrates significant advantages in terms of efficiency and non-destructive detection. These improvements provide a practical advantage for high-throughput analysis without the need for sacrificing the samples. Especially in large-scale aquaculture breeding programs, this model provides a fast and non-destructive way to estimate carotenoid content, particularly zeaxanthin, offering effective support for selective breeding. Zeaxanthin is highly valued for its health benefits, particularly in preventing eye diseases such as macular degeneration (Abdel-Aal et al., 2013).

Moreover, our research reveals significant individual differences in carotenoid content among abalone (see Fig. 7a), with the data showing a normal distribution (see Fig. 7c). This suggests that selecting individuals with higher carotenoid levels for breeding is a practical approach. The application of this model allows for more precise selection of superior individuals in the breeding process, while also significantly reducing costs and improving overall efficiency. This makes the model a practical tool for enhancing health and pigment-related traits in aquaculture.

4.1. Carotenoid composition in foot muscles of different colors

Zeaxanthin is the dominant carotenoid in the orange foot muscle of abalone, which explains the characteristic orange coloration observed. The higher levels of zeaxanthin in the orange samples align with its known role as a primary pigment in organisms exhibiting orange-red coloration (Bunluephan et al., 2023). The biochemical basis for this high accumulation may be linked to the preferential biosynthesis and storage of zeaxanthin in certain tissues, driven by the expression of key enzymes in the carotenoid biosynthetic pathway, such as phytoene synthase and carotenoid hydroxylases. Additionally, zeaxanthin's antioxidant properties likely contribute to its higher accumulation, offering protection against oxidative stress caused by environmental factors, such as temperature fluctuations (Guo et al., 2021).

In contrast, although the green and white foot muscles contain lower levels of carotenoids, zeaxanthin remains the most abundant, suggesting that carotenoid biosynthesis in these tissues may still favor zeaxanthin, albeit at lower rates. This could reflect differential expression of carotenoid-related genes or variations in carotenoid uptake and storage mechanisms across different tissue types. Future studies could focus on investigating the gene expression profiles of key carotenoid biosynthesis enzymes in different color morphs of abalone and the environmental factors that influence carotenoid accumulation.

It is important to note that KEGG annotations currently only include β-carotene as a metabolite in the retinol biosynthesis pathway, despite the identification of several other upregulated carotenoid metabolites in our comparisons. For example, in the DD-ReddishOrange vs DD-White comparison, six upregulated carotenoids were found: zeaxanthin, antheraxanthin, lutein, β-cryptoxanthin, β-carotene, and echinenone, while in the DD-ReddishOrange vs DD-DarkGreen comparison, three upregulated carotenoids—zeaxanthin, lutein, and β-cryptoxanthin—were identified, all with significantly higher content in the reddish-orange foot muscle. Although KEGG primarily focuses on β-carotene in the retinol biosynthesis pathway, zeaxanthin also plays a crucial role in this pathway, particularly for eye health (Abdel-Aal et al., 2013). Both lutein and zeaxanthin are well known for their protective effects in the retina, helping prevent oxidative stress and maintain visual function. Furthermore, the targeted metabolomics analysis of the retinol biosynthesis pathway has provided valuable candidate metabolites. Combining these results with transcriptomic and genomic data, particularly focusing on genes associated with lutein and zeaxanthin, will help narrow down gene localization. This integrated approach could enhance the precision of gene identification and contribute to a more comprehensive understanding of carotenoid biosynthesis in abalone.

4.2. Cost efficiency and practicality

Although the HPLC method is widely used for the quantitative analysis of carotenoids due to its high precision and sensitivity (Balasubramaniam et al., 2020; Guo et al., 2021), it has disadvantages such as being time-consuming and operationally complex. Traditional HPLC analysis requires a complex sample preparation process, including dissection, freeze-drying, and grinding of samples, with each sample taking approximately 30 min for analysis. When processing large batches of samples, the efficiency is low, and the time consumption is substantial. In this study, to complete the measurement of 344 samples, the traditional method would require about 2 days for sample preparation and approximately 7.17 days for HPLC analysis, totaling nearly 9 days (216 h).

In contrast, the non-invasive chromaticity-based method estimates carotenoid content by quickly measuring the L^*, a^*, and b^* chromaticity values of the samples. The data collection time for each sample is only 5 s, and the total time for all 344 samples is about 28.7 min, improving efficiency by approximately 450 times. Moreover, some studies have utilized near-infrared spectroscopy (NIRS) and colorimeter equipment to predict carotenoid content, significantly reducing processing time (Sánchez et al., 2013). However, these methods still require organic solvent extraction, leading to sample destruction.

The non-invasive chromaticity method significantly outperforms traditional HPLC in terms of time efficiency and operational simplicity, while also avoiding sample destruction, thus preserving valuable cultured individuals. This method not only saves substantial time but also offers significant cost advantages. For instance, in China, the colorimeter (3NH-NR110, Shenzhen 3nh Technology Co., Ltd.) costs approximately 3900 CNY, making it an economical and practical solution for aquaculture. In contrast, more advanced instruments such as Fourier Transform Near-Infrared Spectroscopy (FT-NIR) systems, like the Antaris MX (Thermo Fisher), cost around 185,250 CNY in China, while the Finnish Specim FX10 portable hyperspectral imaging system is priced at approximately 300,000 CNY in China. The price difference between the FT-NIR system and the colorimeter is 181,350 CNY, about 47 times the cost of the colorimeter, and the price of the Specim FX10 is about 77 times the cost of the colorimeter. These significant price differences highlight the advantages of the chromaticity-based method as an economical, efficient, and easy-to-operate monitoring tool in aquaculture. Additionally, the chromaticity method is more portable and easier to use in field conditions, making it ideal for large-scale aquaculture breeding programs. It enables rapid, non-destructive carotenoid monitoring without the need for complex equipment or time-consuming processes, making it a practical alternative for carotenoid monitoring in aquaculture.

4.3. Improvement in carotenoid measurement accuracy based on multidimensional chromaticity analysis

Previous studies have predominantly focused on using the single-dimensional a^* to predict carotenoid content. However, relying solely on the a^* has limitations. For example, it has been noted in the analysis of carotenoid content in apricots that although the a^* value shows a strong correlation with carotenoid levels (Vandeputte et al., 2020), combining the L^*, a^*, and b^* chromaticity values can provide more accurate predictions. The b^* and L^* values also offer additional information, with b^* related to yellow hues and L^* associated with brightness, which is particularly relevant in cases with varying pigment deposition types.

Similarly, studies have emphasized that using only the a^* cannot fully capture the complexity of color (Butler et al., 2011). By integrating L^*, a^*, and b^* in a multidimensional analysis, it is possible to more

comprehensively reflect color variations and improve the accuracy of product quality assessments. By incorporating L^* and b^* values, a more thorough characterization of overall color changes was achieved, enhancing the predictive model's accuracy. In many previous studies, $L^*a^*b^*$ chromaticity values have often been used individually (Bunluephan et al., 2023; Itle & Kabelka, 2009), neglecting the potential contributions of other dimensions. However, carotenoids are complex natural pigments that encompass multiple hues and pigment types, making it challenging for single-dimensional analyses to capture all color information (Wen et al., 2023). In this study, we constructed multidimensional predictive models by combining L^*, a^*, and b^*, which significantly improved predictive performance.

By applying deep learning algorithms, such as the LSTM model, we were able to capture the complex nonlinear relationships between chromaticity values and carotenoid content. The LSTM model demonstrated the best predictive accuracy during 10-fold cross-validation, with a root mean square error (RMSE) of 6.692, a coefficient of determination (R^2) of 0.415, and a correlation coefficient of 0.698. These results indicate that multidimensional chromaticity analysis enhances the robustness and predictive accuracy of the model, overcoming the limitations associated with previous single-dimensional analyses.

4.4. Limitations and future improvement directions

Although our study has demonstrated the advantages of multidimensional chromaticity prediction models, there are still certain limitations. First, the sample data are derived only from a specific population, which may limit the applicability of the model to other populations or under different growth conditions. Second, although the multidimensional chromaticity method has improved prediction accuracy, there remains unexplained variance in the model, indicating room for further improvement.

Future research should expand the diversity of the dataset to include samples from different populations and growth stages, ranging from juvenile to adult. Additionally, including samples from various environmental conditions, such as different diets, may help to explore how carotenoid accumulation and pigmentation are influenced by diet. This research will contribute to validating the model across a broader range of conditions, improving its generalizability, and enhancing its applicability to different populations and environments. The software also provides a training interface, allowing users to train the model based on their own data. Furthermore, applying this method to studies of other aquatic species—such as salmon (Lehnert et al., 2019), which also exhibit variations in flesh color gradients—could help validate its universality and assess its potential for broader application across different species.

5. Conclusions

This study demonstrates that non-invasive deep learning models utilizing multi-dimensional chromaticity data (L^*, a^*, and b^*) can accurately estimate carotenoid content in Pacific abalone, significantly enhancing efficiency by approximately 450 times compared to traditional methods like UPLC. By integrating these chromaticity values into models such as the Long Short-Term Memory (LSTM) network, we achieved robust predictions, and the development of the user-friendly Color-Based Carotenoid Estimation Suite (CCES) software facilitates practical application and customization, allowing users to train their own models using their datasets. Our approach is more economical in terms of time and cost, avoids destructive sampling—thus preserving valuable cultured individuals—and empowers researchers and practitioners to implement efficient and precise carotenoid estimation. This not only provides important technical support for quality control and selective breeding in aquaculture but also promotes innovation and progress in the field, offering a valuable tool with potential applications extending to other species where pigmentation is of interest.

CRediT authorship contribution statement

Guijia Liu: Writing – review & editing, Writing – original draft, Validation, Software, Methodology, Formal analysis, Data curation, Conceptualization. **Xiaoyong Wu:** Writing – review & editing, Validation, Investigation, Data curation. **Yiming Wei:** Validation, Investigation, Data curation. **Tian Xu:** Investigation, Data curation. **Dongchang Li:** Resources. **Xuan Luo:** Supervision, Resources, Project administration. **Weiwei You:** Supervision, Resources, Funding acquisition. **Caihuan Ke:** Writing – review & editing, Supervision, Resources, Project administration, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Caihuan Ke reports financial support was provided by National Natural Science Foundation of China. Weiwei You reports financial support was provided by Seed Industry Innovation and Industrialization in Fujian Province, China. Caihuan Ke reports financial support was provided by Major Science and Technology Project in Fujian Province, China. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by grants from National Natural Science Foundation of China (U22A20530), Seed Industry Innovation and Industrialization in Fujian Province (No.2021FJSCZY02), Earmarked Fund for CARS (No. CARS-49), Fujian Special Fund for the Development of Marine and Fishery (FJHYF-ZH-2023-04). Thanks for the support from the Germplasm resources sharing platform of aquatic species in Fujian Province. Carotenoid contents were detected using the AB Sciex QTRAP 6500 LC-MS/MS platform, with the analysis performed by MetWare Biotechnology.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.foodchem.2025.143913.

Data availability

I have shared the link to my data/code at the Attach File step.

References

Abdel-Aal, E.-S., Akhtar, H., Zaheer, K., & Ali, R. (2013). Dietary sources of lutein and zeaxanthin carotenoids and their role in eye health. *Nutrients, 5*, 1169–1185. https://doi.org/10.3390/nu5041169

Balasubramaniam, V., June Chelyn, L., Vimala, S., Mohd Fairulnizal, M. N., Brownlee, I. A., & Amin, I. (2020). Carotenoid composition and antioxidant potential of Eucheuma denticulatum. *Sargassum polycystum and Caulerpa lentillifera. Heliyon, 6*, Article e04654. https://doi.org/10.1016/j.heliyon.2020.e04654

Bunluephan, M., Chuenduang, C., Suamuang, S., Amkul, K., Laosatit, K., Terdwongworakul, A., & Tanadul, O. (2023). Non-destructive estimation of anthocyanin content in yardlong bean based on tristimulus values and reflectance spectra. *Crop Breeding and Applied Biotechnology, 23*, Article e46112341. https://doi.org/10.1590/1984-70332023v23n4a36

Butler, M. W., Toomey, M. B., & McGraw, K. J. (2011). How many color metrics do we need? Evaluating how different color-scoring procedures explain carotenoid pigment content in avian bare-part and plumage ornaments. *Behavioral Ecology and Sociobiology, 65*, 401–413. https://doi.org/10.1007/s00265-010-1074-1

Dall, W., Li, X., Wang, S., Xun, X., Zhang, M., Wang, S., … Bao, Z. (2019). A carotenoid oxygenase is responsible for muscle coloration in scallop. Biochimica et Biophysica Acta (BBA) - molecular and cell biology of. *Lipids, 1864*, 966–975. https://doi.org/10.1016/j.bbalip.2019.03.003

Falcioni, R., Antunes, W. C., Berti de Oliveira, R., Chicati, M. L., Demattê, J. A. M., & Nanni, M. R. (2024). Hyperspectral and chlorophyll fluorescence analyses of

comparative leaf surfaces reveal cellular influences on Leaf Optical properties in Tradescantia plants. *Cells, 13*, 952. https://doi.org/10.3390/cells13110952

Goforth, K. M., Lohmann, C. M. F., Gavin, A., Henning, R., Harvey, A., Hinton, T. L., Lim, D. S., Lohmann, K. J., Goforth, K. M., Lohmann, C. M. F., Gavin, A., Henning, R., Harvey, A., Hinton, T. L., Lim, D. S., & Lohmann, K. J. (2025). Learned magnetic map cues and two mechanisms of magnetoreception in turtles. *Nature, 638*, 1015–1022. https://doi.org/10.1038/s41586-024-08554-y

Guo, J., Khan, J., Pradhan, S., Shahi, D., Khan, N., Avci, M., Mcbreen, J., Harrison, S., Brown-Guedira, G., Murphy, J. P., Johnson, J., Mergoum, M., Esten Mason, R., Ibrahim, A. M. H., Sutton, R., Griffey, C., Babar, M. A., Wojdyło, A., Nowicka, P., … Turkiewicz, I. P. (2021). Fruit tree leaves as unconventional and valuable source of chlorophyll and carotenoid compounds determined by liquid chromatography-photodiode-quadrupole/time of flight-electrospray ionization-mass spectrometry (LC-PDA-qTof-ESI-MS). *Food Chemistry, 349*, Article 129156. https://doi.org/10.1016/j.foodchem.2021.129156

Itle, R. A., & Kabelka, E. A. (2009). Correlation between L*a*b* color space values and carotenoid content in pumpkins and squash (*Cucurbita spp.*). *horts, 44*, 633–637. https://doi.org/10.21273/HORTSCI.44.3.633

Lehnert, S. J., Christensen, K. A., Vandersteen, W. E., Sakhrani, D., Pitcher, T. E., Heath, J. W., Koop, B. F., Heath, D. D., & Devlin, R. H. (2019). Carotenoid pigmentation in salmon: Variation in expression at *BCO2-l* locus controls a key fitness trait affecting red coloration. *Proceedings of the Royal Society B: Biological Sciences, 286*, 206–211. https://doi.org/10.1098/rspb.2019.1588

Liu, M., Kang, B., Wu, H., Aranda, M. A., Peng, B., Liu, L., Fei, Z., Hong, N., Gu, Q., Liu, M., Kang, B., Wu, H., Aranda, M. A., Peng, B., Liu, L., Fei, Z., Hong, N., & Gu, Q. (2023). Transcriptomic and metabolic profiling of watermelon uncovers the role of salicylic acid and flavonoids in the resistance to cucumber green mottle mosaic virus. *Journal of Experimental Botany, 74*, 5218–5235. https://doi.org/10.1093/jxb/erad197

Meier, A. B., Zawada, D., De Angelis, M. T., Martens, L. D., Santamaria, G., Zengerle, S., Nowak-Imialek, M., Kornherr, J., Zhang, F., Tian, Q., Wolf, C. M., Kupatt, C., Sahara, M., Lipp, P., Theis, F. J., Gagneur, J., Goedel, A., Laugwitz, K.-L., Dorn, T., … Ke, C. (2021). Heat adhesion duration: A new high-throughput abalone thermal tolerance assessment method. *Aquaculture, 545*, Article 737226. https://doi.org/10.1016/j.aquaculture.2021.737226

Porturas Olaechea, R., Ushio, H., Watabe, S., Takada, K., & Hatae, K. (1993). Toughness and collagen content of abalone muscles. *Bioscience, Biotechnology, and Biochemistry, 57*, 6–11. https://doi.org/10.1271/bbb.57.6

Sachindra, N. M., Bhaskar, N., & Mahendrakar, N. S. (2005). Carotenoids in different body components of Indian shrimps. *Journal of the Science of Food and Agriculture, 85*, 167–172. https://doi.org/10.1002/jsfa.1977

Thévenot, E. A., Roux, A., Xu, Y., Ezan, E., & Junot, C. (2015). Analysis of the human adult urinary metabolome variations with age, body mass index, and gender by implementing a comprehensive workflow for univariate and OPLS statistical analyses. *Journal of Proteome Research, 14*, 3322–3335. https://doi.org/10.1021/acs.jproteome.5b00354

Vandeputte, M., Haffray, P., Zhou, W., Niu, Y., Ding, X., Zhao, S., Li, Y., Fan, G., Zhang, S., & Liao, K. (2020). Analysis of carotenoid content and diversity in apricots (Prunus armeniaca L.) grown in China. *Food Chemistry, 330*, Article 127223. https://doi.org/10.1016/j.foodchem.2020.127223

Wade, N. M., Gabaudan, J., & Glencross, B. D. (2017). A review of carotenoid utilisation and function in crustacean aquaculture. *Reviews in Aquaculture, 9*, 141–156. https://doi.org/10.1111/raq.12109

Wei, X., Zeng, W., Tang, B., He, J., Chen, N., Luo, X., Feng, D., You, W., & Ke, C. (2019). Comparative analysis of the predominant carotenoids and chemical components in the common and orange-muscle mutant of *Haliotis gigantea*. *Aquaculture Research, 50*, 2938–2950. https://doi.org/10.1111/are.14248

Wen, Y., Lin, S., Li, X., Zhang, J., Zhao, Y., Ma, D., Li, M., Ren, X., & Zhang, W. (2023). Relationship between wheat flour's quality characteristics and color of fresh wet noodles. *International Journal of Food Properties, 26*, 290–300. https://doi.org/10.1080/10942912.2022.2161565

Xu, J., You, F., Wu, X., Zhang, P., Lin, Y., Jiang, H., Zheng, C., Xue, Y., Zhang, H., Tan, K., Ma, H., Li, S., & Zheng, H. (2022). Identification of a key gene StAR-like-3 responsible for carotenoids accumulation in the noble scallop Chlamys nobilis. *Food Chemistry: Molecular Sciences, 4*, Article 100072. https://doi.org/10.1016/j.fochms.2021.100072

Yang, L., Liu, Y., Yu, H., Fang, X., Song, L., Li, D., & Chen, Y. (2021). Computer vision models in intelligent aquaculture with emphasis on fish detection and behavior analysis: A review. *Arch Computat Methods Eng, 28*, 2785–2816. https://doi.org/10.1007/s11831-020-09486-2

Yu, W., Liu, J., Yu, F., Shen, Y., Gong, S., Lu, Y., Peng, W., Wang, Y., Gan, Y., Xiao, Q., Luo, X., You, W., & Ke, C. (2022). Heritability and genetic correlation for residual feed intake of Pacific abalone (*Haliotis discus hannai*). *Aquaculture, 553*, Article 738060. https://doi.org/10.1016/j.aquaculture.2022.738060

Zhao, J., Zhu, T., Qiu, Z., Li, T., Wang, G., Li, Z., & Du, H. (2023). Hyperspectral prediction of pigment content in tomato leaves based on logistic-optimized sparrow search algorithm and back propagation neural network. *J. Agric. Eng.*. https://doi.org/10.4081/jae.2023.1528

Zheng, H., Liu, H., Zhang, T., Wang, S., Sun, Z., Liu, W., & Li, Y. (2010). Total carotenoid differences in scallop tissues of Chlamys nobilis (bivalve: Pectinidae) with regard to gender and shell colour. *Food Chemistry, 122*, 1164–1167. https://doi.org/10.1016/j.foodchem.2010.03.109

Zhou, M., Chen, J., Peng, W., Liu, J., Yu, F., Lin, W., Xie, Q., You, W., Ke, C., & Luo, X. (2022). Estimation of genetic parameters for growth relevant traits in adult Pacific abalone (*Haliotis discus hannai*). *Aquaculture Research, 53*, 5018–5028. https://doi.org/10.1111/are.15987