A Supplementary Information

This supplementary material provides additional information and results to complement the main manuscript. It includes detailed performance results in terms of RMSE and R^2 for each traditional baseline model, presented for both the FT-Raman and InGaAs-truncated datasets. Additionally, this document contains further details on the experimental setups, data preprocessing techniques, and ablation studies not covered in the main manuscript.

B Preprocessing Design Matrix

Table 5 illustrates all 64 preprocessing procedures conducted in our experiments. It contains one raw data that has not been subjected to any preprocessing methods, and other procedures that have been subjected to baseline correction, scatter correction, derivative transformations, and scaling preprocessing methods. Each procedure is applied in a predefined order based on [9] to ensure consistency and effectiveness, avoiding the complexity associated with testing all possible permutations.

Table 5: The experimental design matrix of 63 preprocessing procedures and one raw data. The preprocessing procedure selection has been used for both FT-Raman and InGaAs datasets. We use "-" to denote this step is not applied. Abbreviations: LB: Linear Baseline [9]; GS: Global Scaling [2]; SNV: Standard Normal Variate [9];

ID	Baseline	Scatter	Derivative (D)	Scaling (Sg)
1(raw)	-	-	-	-
2	LB	-	-	-
3	LB	SNV	-	-
4	LB	SNV	1st, w=5	-
5	LB	SNV	1st, w=9	-
6	LB	SNV	1st, w=13	-
7	LB	SNV	1st, w=17	-
8	LB	SNV	1st, w=21	-
9	LB	SNV	1st, w=25	-
10	LB	SNV	2nd, w=13	-
11	LB	SNV	2nd, w=15	-
12	LB	SNV	2nd, w=17	-
13	LB	SNV	2nd, w=19	-
14	LB	SNV	2nd, w=21	-
15	LB	SNV	2nd, w=23	-
16	LB	SNV	2nd, w=25	-
17	LB	SNV	2nd, w=31	-
18	-	SNV	-	-
19	-	SNV	1st, w=5	-
20	-	SNV	1st, w=9	-

Table 5: The experimental design matrix of 63 preprocessing procedures and one raw data. The preprocessing procedure selection has been used for both FT-Raman and InGaAs datasets. We use "-" to denote this step is not applied. Abbreviations: LB: Linear Baseline [9]; GS: Global Scaling [2]; SNV: Standard Normal Variate [9];

ID	Baseline	Scatter	Derivative (D)	Scaling (Sg)
21	-	SNV	1st, w=13	-
22	-	SNV	1st, w=17	-
23	-	SNV	1st, w=21	-
24	-	SNV	1st, w=25	-
25	-	SNV	2nd, w=13	-
26	-	SNV	2nd, w=15	-
27	-	SNV	2nd, w=17	-
28	-	SNV	2nd, w=19	-
29	-	SNV	2nd, w=21	-
30	-	SNV	2nd, w=23	-
31	-	SNV	2nd, w=25	-
32	-	SNV	2nd, w=31	-
33	LB	-	-	GS
34	LB	SNV	-	GS
35	LB	SNV	1st, w=5	GS
36	LB	SNV	1st, w=9	GS
37	LB	SNV	1st, w=13	GS
38	LB	SNV	1st, w=17	GS
39	LB	SNV	1st, w=21	GS
40	LB	SNV	1st, w=25	GS
41	LB	SNV	2nd, w=13	GS
42	LB	SNV	2nd, w=15	GS
43	LB	SNV	2nd, w=17	GS
44	LB	SNV	2nd, w=19	GS
45	LB	SNV	2nd, w=21	GS
46	LB	SNV	2nd, w=23	GS
47	LB	SNV	2nd, w=25	GS
48	LB	SNV	2nd, w=31	GS
49	-	SNV	-	GS
50	-	SNV	1st, w=5	GS
51	-	SNV	1st, w=9	GS
52	-	SNV	1st, w=13	GS
53	-	SNV	1st, w=17	GS
54	-	SNV	1st, w=21	GS
55	-	SNV	1st, w=25	GS
56	-	SNV	2nd, w=13	GS
57	-	SNV	2nd, w=15	GS
58	-	SNV	2nd, w=17	GS
59	-	SNV	2nd, w=19	GS
60	-	SNV	2nd, w=21	GS
61	-	SNV	2nd, w=23	GS

Table 5: The experimental design matrix of 63 preprocessing procedures and one raw data. The preprocessing procedure selection has been used for both FT-Raman and InGaAs datasets. We use "-" to denote this step is not applied. Abbreviations: LB: Linear Baseline [9]; GS: Global Scaling [2]; SNV: Standard Normal Variate [9];

ID	Baseline	Scatter	Derivative (D)	Scaling (Sg)
62	-	SNV	2nd, w=25	GS
63	-	SNV	2nd, w=31	GS
64	-	-	-	GS

C RMSE results

For this part, we present the same format as the main paper content, but for the RMSE metric. In particular, the results are divided into two parts. First, the overall RMSE performance averaging the results across all three targets (Water, Protein, Lipids yield) to provide a general overview of the model's effectiveness. Subsequently, the individual target performance, which examines each target separately, will be discussed to provide a more detailed analysis of the model's performance on each target.

To understand the multioutput $RMSE$ performance, we use the formula: $RMSE$:

$$RMSE_{\text{overall}} = \frac{1}{N} \sum_{i=1}^{N} RMSE_i$$

, where $RMSE_{\text{overall}}$ is the overall $RMSE$ value, N is the number of targets, and $RMSE_i$ is the $RMSE$ value for each target i (Water, Protein, Lipids yield). We calculate the $RMSE_{\text{overall}}$ scores for each of the 6 folds in the cross-validation process. The reported $RMSE_{\text{CV}}$ is expressed as the mean ± standard deviation of these 6 $RMSE_{\text{overall}}$ scores. This allows us to evaluate the multioutput predictive performance while accounting for variance across the cross-validation folds.

C.1 Overall Performance Analysis

Table 6: Overall RMSE performance comparison of baseline, two benchmark CNNs, and our CNN. Both FT-Raman and InGaAs-trun are SNV [9] preprocessed. The p-values are obtained using the pairwise Mann-Whitney U test based on 10 individual runs, while deterministic models and FishCNN itself are denoted by "-". The significantly better results are highlighted in bold, values before and after ± indicate mean±standard deviation.

Data	Model	$RMSECV$(mean±std)	*p-value*
FT-Raman	**FishCNN**	**0.788±0.230**	-
	1CLNN [7]	0.900±0.340	5.4e-02
	3CLNN [16]	0.961±0.201	5.4e-06
	KNN	0.901±0.286	-
InGaAs-trun	**FishCNN**	**0.650±0.120**	-
	1CLNN [7]	0.750±0.174	7.3e-05
	3CLNN [16]	0.870±0.200	2.9e-10
	LS-SVM(poly)	0.840±0.183	-

As shown in Table 6, on both FT-Raman and InGaAs truncated datasets, our CNN model achieves significantly better $RMSECV$ scores than all traditional models as well as 1CLNN and 3CLNN. The p-values of the relative pairwise

Mann-Whitney U test [21] are consistently lower than 0.05, indicating statistically significant improvement. This findings aligns with the results of the R^2CV metric presented in the main text.

C.2 Individual Target Performance

Table 7: Each individual target performance comparison. For lipids yield, the InGaAs data is preprocessed by SNV followed by second order derivative with a 19-point window size, while others are all preprocessed by SNV only. The p-values are obtained using the pairwise Mann-Whitney U test based on 10 individual runs, while deterministic models and FishCNN itself are denoted by "-". The significantly better results are highlighted in bold, values before and after ± indicate mean±standard deviation.

Target	Data	Model	$RMSECV$(mean±std)	*p-value*
Water	FT-Raman	**FishCNN**	1.104±0.520	-
		1CLNN[7]	1.150±0.612	0.46
		3CLNN[16]	1.428±0.404	3.9e-06
		RF	1.216±0.413	0.0065
	InGaAs-trun	**FishCNN**	**0.847±0.246**	-
		1CLNN[7]	0.999±0.340	1.6e-03
		3CLNN[16]	1.256±0.390	1.7e-09
		LS-SVM(poly)	1.247±0.215	-
Protein	FT-Raman	FishCNN	0.842±0.257	-
		1CLNN[7]	0.794±0.247	8.4e-01
		3CLNN[16]	0.967±0.258	3.0e-03
		KNN	0.879±0.287	-
	InGaAs-trun	**FishCNN**	**0.648±0.215**	-
		1CLNN[7]	0.757±0.255	5.3e-03
		3CLNN[16]	0.895±0.258	1.3e-07
		KNN	0.817±0.200	-
Lipids yield	FT-Raman	**FishCNN**	**0.418±0.183**	-
		3CLNN	0.757±0.326	1.5e-10
		1CLNN	-0.488±0.130	9.4e-03
		LS-SVM(linear)	0.420±0.158	-
	InGaAs-trun	**FishCNN**	**0.304±0.090**	-
		1CLNN[7]	0.495±0.110	7.4e-03
		3CLNN[16]	0.459±0.092	2.9e-01
		SVR(linear)	0.389±0.103	-

As illustrated in Table 7, for RMSE performance, the performance of our FishCNN model is aligned with the R^2CV results. This suggests that InGaAs-trun data is a better spectroscopic technique than FT-Raman in predicting Water, Protein, and Lipids yield. And our FishCNN model is more effective in predicting these targets than the benchmark models.

D Detailed Baseline Models and Performance Results

In this appendix, we provide detailed performance results for all baseline models used in the study. While the main manuscript highlights the best-performing models, this appendix includes complete RMSE and R^2 results for all methods tested, allowing for a comprehensive evaluation.

Table 8: RMSE and R^2 performance results for all baseline models. Both FT-Raman and InGaAs-trun are SNV [9] preprocessed. Values before and after ± indicate mean±standard deviation.

Target	Data	Model	R^2CV(mean±std)	$RMSECV$(mean±std)
Overall	FT-Raman	LS-SVM(poly)	0.594±0.197	1.023±0.198
		LS-SVM(linear)	0.657±0.145	0.967±0.158
		LS-SVM(rbf)	0.368±0.174	1.367±0.252
		PLSR1	0.569±0.174	1.024±0.148
		ENet	0.675±0.137	0.958±0.160
		LassoLars	0.642±0.124	1.043±0.172
		LGBM	0.633±0.132	0.952±0.146
		KNN	0.680±0.125	0.901±0.286
		RF	0.598±0.145	1.024±0.195
	InGaAs-trun	LS-SVM(poly)	0.752±0.091	0.840±0.183
		LS-SVM(linear)	0.711±0.078	0.957±0.094
		LS-SVM(rbf)	0.684±0.085	0.931±0.174
		PLSR1	0.732±0.137	0.885±0.145
		ENet	0.706±0.076	0.966±0.094
		LassoLars	0.678±0.092	1.001±0.081
		LGBM	0.650±0.108	1.020±0.169
		KNN	0.711±0.067	0.864±0.209
		RF	0.624±0.112	1.008±0.191
Water	FT-Raman	LS-SVM(poly)	0.742±0.247	1.353±0.377
		LS-SVM(linear)	0.752±0.191	1.374±0.282
		LS-SVM(rbf)	0.528±0.196	2.020±0.333
		PLSR1	0.761±0.190	1.344±0.310
		ENet	0.749±0.196	1.379±0.293
		LassoLars	0.697±0.231	1.518±0.332
		LGBM	0.827±0.053	1.249±0.256
		KNN	0.785±0.070	1.393±0.309
		RF	0.832±0.080	1.216±0.413
	InGaAs-trun	LS-SVM(poly)	0.824±0.058	1.247±0.215
		LS-SVM(linear)	0.781±0.054	1.395±0.169
		LS-SVM(rbf)	0.799±0.072	1.337±0.367
		PLSR1	0.794±0.082	1.332±0.234
		ENet	0.781±0.054	1.397±0.170
		LassoLars	0.766±0.038	1.453±0.142
		LGBM	0.725±0.115	1.548±0.372
		KNN	0.846±0.081	1.167±0.406
		RF	0.748±0.122	1.458±0.471
Protein	FT-Raman	LS-SVM(poly)	0.542±0.250	1.179±0.266

Table 8: RMSE and R^2 performance results for all baseline models. Both FT-Raman and InGaAs-trun are SNV [9] preprocessed. Values before and after ± indicate mean±standard deviation.

Target	Data	Model	R^2CV(mean±std)	$RMSECV$(mean±std)
		LS-SVM(linear)	0.644±0.195	1.043±0.260
		LS-SVM(rbf)	0.490±0.184	1.282±0.275
		PLSR1	0.599±0.180	1.121±0.230
		ENet	0.644±0.196	1.043±0.259
		LassoLars	0.573±0.227	1.151±0.302
		LGBM	0.682±0.146	0.999±0.205
		KNN	0.764±0.112	0.879±0.287
		RF	0.664±0.173	1.022±0.251
	InGaAs-trun	LS-SVM(poly)	0.708±0.132	0.942±0.188
		LS-SVM(linear)	0.658±0.103	1.042±0.151
		LS-SVM(rbf)	0.768±0.085	0.875±0.226
		PLSR1	0.701±0.204	0.908±0.225
		ENet	0.647±0.106	1.059±0.156
		LassoLars	0.641±0.097	1.073±0.151
		LGBM	0.711±0.091	0.965±0.175
		KNN	0.798±0.065	0.817±0.200
		RF	0.730±0.096	0.932±0.216
Lipids yield	FT-Raman	LS-SVM(poly)	0.497±0.327	0.535±0.132
		LS-SVM(linear)	0.643±0.331	0.420±0.158
		LS-SVM(rbf)	0.087±0.259	0.800±0.178
		PLSR1	0.346±0.431	0.608±0.123
		ENet	0.632±0.274	0.451±0.114
		LassoLars	0.655±0.204	0.460±0.090
		LGBM	0.391±0.345	0.609±0.146
		KNN	0.403±0.193	0.640±0.123
		RF	0.349±0.419	0.643±0.162
	InGaAs-trun	LS-SVM(poly)	0.727±0.200	0.392±0.101
		LS-SVM(linear)	0.693±0.180	0.434±0.081
		LS-SVM(rbf)	0.483±0.252	0.581±0.145
		PLSR1	0.702±0.211	0.414±0.111
		ENet	0.690±0.168	0.441±0.075
		LassoLars	0.628±0.228	0.477±0.090
		LGBM	0.515±0.248	0.548±0.163
		KNN	0.494±0.165	0.594±0.144
		RF	0.395±0.237	0.635±0.129

E Ablation Study

In this section, we present the ablation study results to evaluate the effectiveness of the individual components of our proposed framework, including the order of applying Data Augmentation (DA) and preprocessing, and the impact of the augmentation factor.

We conduct the ablation study on the InGaAs-truncated dataset with the focus on the overall R^2CV performance. InGaAs data is chosen for this study due to its better performance compared to FT-Raman data, as shown in the main text. While, the overall R^2CV is used since the multioutput regression task is the main focus of this study.

E.1 Effectiveness of the Framework Components

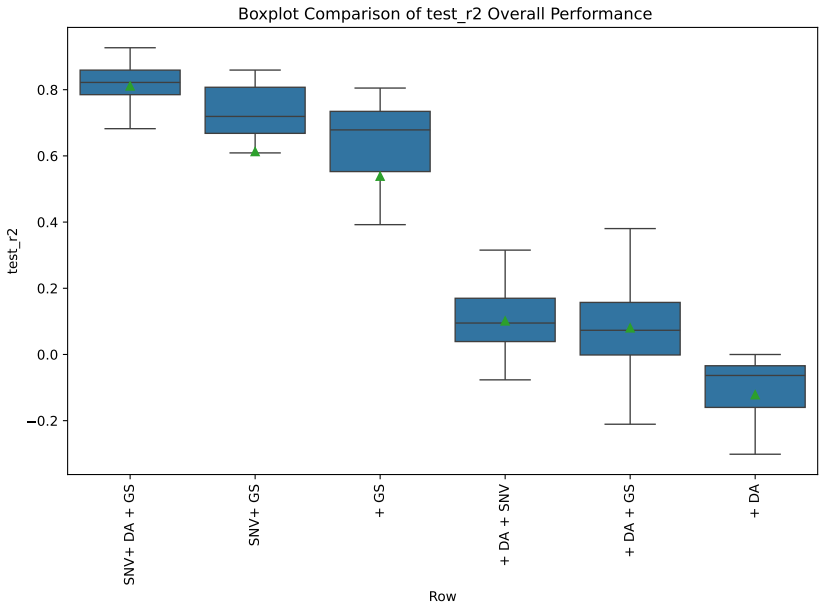

Fig. 5: Boxplot of Effectiveness of Framework Components. The mean value is denoted as the green triangle, and the median is denoted as the line that split the box in two. X-axis: Different DA Factors, Y-axis: R^2 Performance

As illustrated in Table 9 and Fig.5, we have found that the sequential application of preprocessing and data augmentation (SNV+DA+GS, SNV+DA) perform significantly better than others, especially when compared to the common practice [2, 16] of applying data augmentation before preprocessing (DA+SNV). The p-values of the relative pairwise Mann-Whitney U test [21] are consistently lower than 0.05, indicating statistically significant improvement. The reason behind this is that data augmentation introduces noise and irrelevant information

to the raw unpreprocessed data, which further amplifies the noise and makes it difficult for the model to learn meaningful patterns.

In addition to the common practise, we have investigated the impact of not applying the SNV preprocessing method and data augmentation (DA) on the model's performance. And the results clearly show a significant drop in performance when the SNV preprocessing method is excluded (DA+GS, DA) and when data augmentation is excluded (SNV+GS, GS).

This ablation study underscores the contribution of each component in preparing spectral data for predicting multiple fish biochemical contents effectively. The results demonstrate that the SNV preprocessing method and data augmentation must be applied sequentially to enhance the model's performance and generalization ability, especially when dealing with ultra-small spectral datasets.

Table 9: Ablation Study: Effectiveness of Framework Components in InGaAs-truncated Data. The p-values are obtained using the pairwise Mann-Whitney U test between SNV+DA+GS and others
Abbreviations: SNV: Standard Normal Variate preprocessing method, DA: Data Augmentation, GS: Global Scaling, +: sequential application

Procedure	R^2CV	*p-value*
SNV+DA+GS	0.811±0.105	-
SNV+GS	0.613±0.415	1.6e-08
GS	0.539±0.372	2.5e-16
DA+SNV	0.101±0.107	6.5e-21
DA+GS	0.080±0.115	5.9e-21
DA	-0.122±0.132	1.8e-21

E.2 Effectiveness of Different Data Augmentation Factor

The data augmentation factor determines the number of synthetic samples generated from each original sample, thereby increasing the diversity of the training data. A higher augmentation factor exposes the model to a wider range of variations during training, which can potentially enhance its robustness and generalization capabilities. To investigate the effect of the augmentation factor, we evaluate the performance of our framework using augmentation factors of 10, 30, 50, and 60 on the InGaAs-truncated dataset.

Table 10 presents the results of this ablation study. As shown, the model's performance is significantly improved when the augmentation factor is increased from 10 to 50. The p-values obtained from the pairwise Mann-Whitney U test between the augmentation factor of 50 and other factors indicate that the improvement is statistically significant (p-value < 0.05) for the factor of 10, but not for factors of 30 and 60.

To determine which factor should be chosen, the boxplot, as shown in Fig. 6, is used to explore their performance distribution. We can clearly observe that

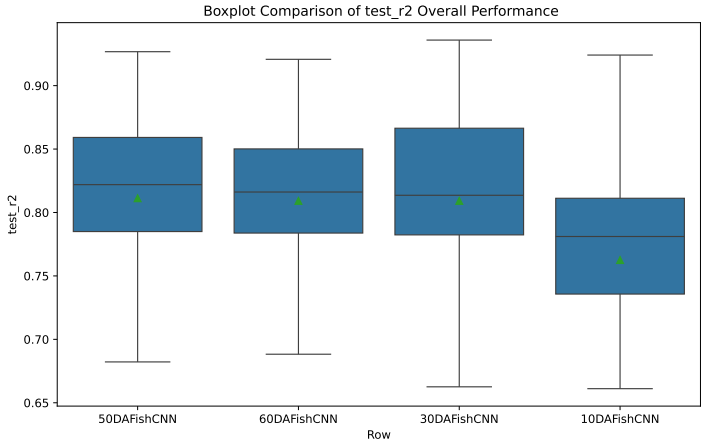

Fig. 6: DA Factor Boxplot. The mean value is denoted as the green triangle, and the median is denoted as the line that split the box in two.
X: Different DA Factors, Y: R^2 Performance

the mean and median of an augmentation factor of 50 is slightly better than others. Hence, 50 is chosen as the optimal value for our framework.

Table 10: Ablation Study: Comparsions of 10, 30, 50, 60 Data Augmentation Factor in InGaAs-truncated Data. The p-values are obtained using the pairwise Mann-Whitney U test between 50 DA and others

Procedure	R^2CV	*p-value*
SNV+ 50DA + GSX	0.811±0.105	-
SNV+ 60DA + GSX	0.809±0.085	0.25
SNV+ 30DA + GSX	0.809±0.092	0.38
SNV+ 10DA + GSX	0.763±0.163	0.00055

E.3 Effect of Different Convolutional Kernel Size

Based on the comparative analysis of various kernel sizes illustrated in Table 11, it is evident that the choice of kernel size significantly impacts performance metrics across different targets. According to the results, the 64 kernel size demonstrates consistently high performance across key metrics, particularly in water and protein prediction performance. However, the prediction accuracy for the lipids yield target with a 64 kernel size is not optimal. As discussed in Section 5.2, predicting lipids yield is more challenging and requires additional derivative transformations. Therefore, we conducted further experiments using SNV + second derivative transformation with a 19-window size, as illustrated in Table 12. These results show that the 64 kernel size consistently outperforms other sizes in this context, reaffirming its suitability as the optimal choice for balancing model complexity and predictive accuracy.

Table 11: Ablation Study: Comparsion Results of 64, 16, 8, 4 convolutional Kernel sizes on InGaAs-truncated data. The data preparation procedure is SNV + DA + GS. The p-values are obtained using the pairwise Mann-Whitney U test between 64 sizes and others

Performance & *p-value*	64Size	16Size	8Size	4Size
Overall R^2CV	0.811±0.098	0.786±0.149	0.808±0.080	0.809±0.066
p-value	-	9.5e-02	1.4e-01	1.3e-01
Water R^2CV	0.917±0.038	0.910±0.041	0.905±0.039	0.897±0.056
Water *p-value*	-	1.4e-01	8.8e-03	5.4e-03
Protein R^2CV	0.867±0.070	0.802±0.279	0.846±0.089	0.840±0.097
Protein *p-value*	-	3.1e-02	1.0e-01	5.6e-02
Lipids yield R^2CV	0.649±0.291	0.647±0.257	0.675±0.236	0.690±0.176
Lipids yield *p-value*	-	4.9e-01	7.1e-01	7.4e-01

Table 12: Ablation Study: Comparsion Results of 64, 16, 8, 4 convolutional Kernel sizes on InGaAs-truncated data. The data preparation procedure is SNV + 2nd derivative with 19 window sizes + DA + GS. The p-values are obtained using the pairwise Mann-Whitney U test between 64 sizes and others

Hyperparameters	64Size	16Size	8Size	4Size
Overall R^2CV	0.758±0.095	0.738±0.117	0.738±0.104	-0.874±14.851
Overall *p-value*	-	1.1e-01	1.0e-01	3.5e-02
Water R^2CV	0.565±0.268	0.540±0.257	0.542±0.250	-4.205±44.469
Water *p-value*	-	1.4e-01	1.4e-01	1.1e-01
Protein R^2CV	0.862±0.101	0.845±0.165	0.842±0.151	0.807±0.261
Protein *p-value*	-	5.6e-01	4.4e-01	4.6e-01
Lipids yield R^2CV	0.846±0.102	0.828±0.151	0.831±0.119	0.777±0.231
Lipids yield *p-value*	-	2.5e-01	2.5e-01	6.0e-03