Supporting Information

Accuracy Enhancement of Glioma Boundary Tissue Identification by Polarization-resolved LIBS Spectral Fusion

Xiangjun Xu,[a, d, †] Geer Teng,[a, f, †] Qianqian Wang,[a, d, e, *] Haifeng Yang,[b, *] Haiyang Yang,[c] Zhifang Zhao,[a, d] Bingheng Lu,[a, d] Mengyu Bao,[a, d] Yongyue Zheng,[a, d] and Tianzhong Luo [a, d, e]

[a] School of Optics and Photonics, Beijing Institute of Technology, Beijing 100081, P. R. China

[b] Department of Neuro-Oncology, Chongqing University Cancer Hospital, Chongqing 400030, P. R. China

[c] Department of Neurosurgery, Kunming Sanbo Brain Hospital, Kunming 650100, P. R. China

[d] Key Laboratory of Photonic Information Technology, Ministry of Industry and Information Technology, Beijing Institute of Technology, Beijing 100081, P. R. China

[e] Yangtze Delta Region Academy of Beijing Institute of Technology, Jiaxing 314033, P. R. China

[f] Institute of Biomedical Engineering, Department of Engineering Science, University of Oxford, Oxford OX3 7LD, United Kingdom

[†] These authors contributed to the work equally and should be considered co-first authors

Corresponding Author

Email address: qqwang@bit.edu.cn (*Q. Q. Wang); yanghaifeng@cqu.edu.cn (*H. F. Yang)

1.Methods of dividing the training and test sets

Owing to the diversity of individual dietary habits, environments, and upbringings, the physiological environments of some individuals may considerably differ from those of others, causing significant changes in the spectral characteristics. If the spectral distributions of the training set and the test set are significantly different, some intensity distribution information of the test set may lack in the training set, resulting in poor accuracy of the model on the test set. Therefore, the spectral distribution of the training set should not be significantly different from the test set. Specifically, the training and test sets are divided by the following two steps. (1) The test set should be randomly selected to ensure that the predicted results are representative. From the total sample, 25% is randomly selected to form the test set and the remaining 75% to form the training set. (2) Wilcoxon tests are performed on the training set spectra and test set spectra in the three tissues, respectively. The goal of the test is to determine whether the two sets of spectra differed from each other in a statistically significant manner. If the *p*-value of the Wilcoxon test that divides the training and test sets is less than 0.05, the training and test set data will be randomly divided again. For both the training and test sets for the three tissues, the *p*-values of the Wilcoxon test are greater than 0.05 indicating that the spectra of the training set are not significantly different from the test set. In this research, the *p*-value of Wilcoxon test between the training and test sets is 0.1068 for normal tissue, 0.0650

for boundary tissue, and 0.0938 for tumor tissue. A detailed description can be found in our published paper[1].

Table S1. Division of the dataset

Types of tissues	Normal tissue	Boundary tissue	Tumor tissue
All Samples	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14	15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35	36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56
Training set	1, 2, 3, 4, 6, 8, 9, 11, 12, 13	15, 16, 17, 18, 19, 20, 22, 23, 26, 27, 29, 31, 32, 33, 34, 35	36, 37, 38, 39, 40, 41, 43, 45, 47, 48, 50, 51, 52, 53, 54, 55,
Test set	5 7 10 14	21, 24, 25, 28, 30	42, 44, 46, 49, 56

2.Wavelength calibration of spectrometers

The spectrometer has been calibrated at the manufacturer to ensure accurate wavelength assignment for each pixel point. However, due to factors such as environment, temperature and mechanical vibration, the optics of the spectrometer may undergo small changes, resulting in drift in the measurement results. Therefore, wavelength calibration of each spectrometer is required before the experiment to ensure the accuracy of the measurement results. Specifically, the spectrometer that needed to be calibrated was connected by optical fiber to a mercury-argon lamp (AvaLight-CAL). Then, the integration time was set to 1 ms in the software for spectral acquisition. As shown in Fig. S1(a), the spectrum of the Hg-Ar lamp collected by spectrometer, from which all the atomic characteristic spectral lines between 350nm and 950nm can be clearly observed. From all the characteristic spectral lines, the stronger ones were selected as calibration points, and the position of the pixel point corresponding to each calibration point was recorded. The standard wavelength of each calibration point was obtained by referring to the factory manual of the light source, and a third-degree polynomial was used to fit the function between the pixel position and the wavelength.

Here, 21 characteristic spectral lines were selected for calibration, specifically 365.01nm, 404.66nm, 435.83nm, 546.08nm, 576.96nm, 579.07nm, 696.54nm, 706.72nm, 727.29nm, 738.40nm, 750.39nm, 763.51nm, 772.42nm, 794.82nm, 800.62nm, 811.53nm, 826.45nm, 842.45nm, 852.14nm, 912.30nm and 922.45nm. The functional relationship between the wavelength λ and the pixel number x of CCD is calculated as follows:

$$\lambda = a_0 + a_1 \cdot x + a_2 \cdot x^2 + a_3 \cdot x^3 \quad (1)$$

In formula: a_0=344.5, a_1=0.3778, a_2=-1.506×10^{-5}, a_3=-2.195×10^{-9}

Subsequently, the calibrated wavelengths can be obtained by bringing pixel points 1 to 2048 into Eq. (1), respectively. After wavelength calibration, the fitting relationship between the wavelengths and pixel points of the 21 characteristic spectral lines is shown in Fig. S1(b), and the fitting coefficient R^2= 0. 999 indicates that the method has achieved good results for the wavelength calibration of the spectrometer. Similarly, wavelength calibration is performed for other spectrometers.

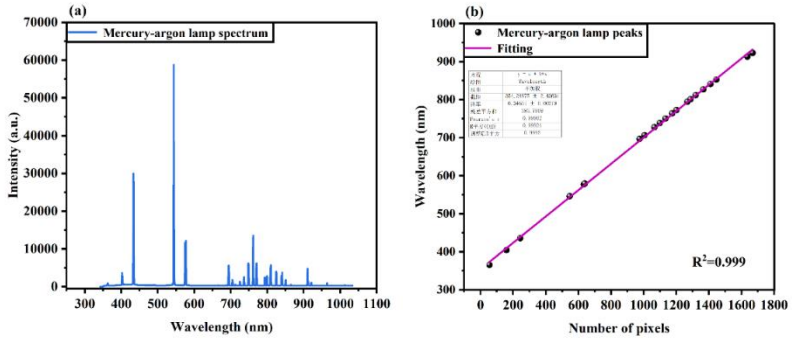

Fig. S1. Spectrometer wavelength calibration process (**a**) atomic spectrum of mercury-argon lamp, (**b**) linear fitting relationship between pixel points and wavelengths

3.Absolute radiometric calibration of spectrometers

In the polarization spectrum collection system, although four fiber optic spectrometers of the same model are used, the output intensity of each spectrometer cannot directly represent the absolute radiation intensity of the collected polarization spectrum. This is due to the different transmittances of the optics and the different quantum efficiencies of the CCD sensors in the different spectrometers. Hence, the absolute radiation intensity of the spectrometer needs to be calibrated before the experiment using a standard radiation source. Assuming that the spectral transmittance of the optical fiber and grating is $F(\lambda)$, the formula between the intensity of the fiber optic spectrometer $I(\lambda)$ and the incident irradiance of the optical fiber $E(\lambda)$ is as follows:

$$I(\lambda) = GF(\lambda)\eta(\lambda)\frac{\lambda A}{hc}E(\lambda)t_{exp} \tag{2}$$

where G is the system gain, $F(\lambda)$ is the spectral transmittance, $\eta(\lambda)$ is the quantum efficiency of the CCD sensor, A is the area of the pixel, h is Planck's constant, c is the speed of light, λ is the wavelength, $E(\lambda)$ is the incident irradiance at the fiber port, and t_{exp} is the integration time. Defining the product of the three parameters G, $F(\lambda)$ and $\eta(\lambda)$ as the absolute spectral response efficiency $R_\lambda(\lambda)$ of the fiber optic spectrometer, formula (2) can be expressed as:

$$I(\lambda) = R_\lambda(\lambda)\frac{\lambda A}{hc}E(\lambda)t_{\exp} \tag{3}$$

The output spectral irradiance $E(\lambda)$ of the standard radiation source is known (the calibration file is traceable to the Chinese Academy of Science and Metrology). During the calibration process, the standard radiation light source was first preheated for 30 min, and then the light source was connected to the fiber optic spectrometer by a cosine corrector. The integration time of the spectrometer is set to 50 ms and the intensity $I(\lambda)$ of the spectrometer is measured. Finally, the absolute spectral response efficiency $R_\lambda(\lambda)$ of the spectrometer can be calculated by formula (3). Similarly, when the integration time is determined and the intensity $I(\lambda)$ of the spectrometer is obtained, and the measured spectral irradiance (absolute radiation intensity) can be calculated by bringing $R_\lambda(\lambda)$ into formula (3).

Before the experiment, the four spectrometers were calibrated for absolute radiation according to the above method. In order to verify the calibration effect, the calibrated spectrometers were used for the measurement of LED and mercury-argon light spectra. As shown in Fig. S2(a) and (b), the absolute radiation spectra of LED and Hg-Ar lamps were collected respectively, from which it can be clearly seen that the black line and the red line almost coincide, indicating that the absolute radiation calibration can effectively eliminate the differences between different spectrometers. Similarly, following absolute radiometric calibration, spectrometers 3 and 4 are perfectly calibrated.

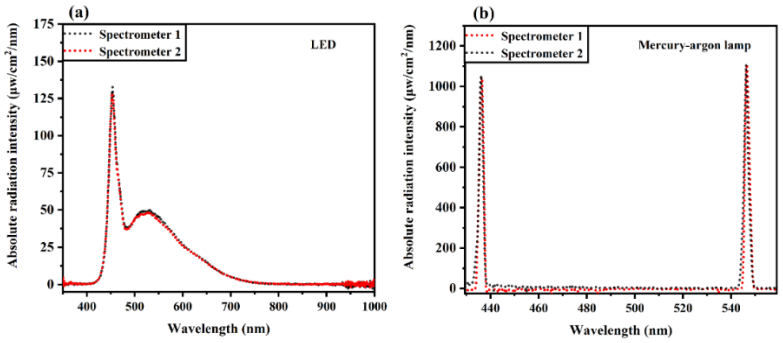

Fig. S2. Measurements of two spectrometers after absolute radiometric calibration. (**a**) LED, (**b**) mercury-argon lamps

4.Polarization parameters of the characteristic spectral lines

Below are the results of the calculations for the polarization parameters of the characteristic spectral lines. For the characteristic spectral lines of the tumor tissue in Fig. 4, specific Stokes parameters, I_{pol} and DOP values are listed in Table S2. S_0 represents the total spectral intensity of the characteristic spectral lines. S_1 is obtained by subtracting I(0°,0) and I(90°,0). I(90°,0) is larger for most of the characteristic spectral lines, whereas I(0°,0) is larger for C_2 517.9 nm and H 656.3 nm. Similarly, S_2 is obtained by subtracting the intensities of I(45°,0) and I(135°,0). I(45°,0) has a larger value for the characteristic spectral line C_2 517.9 nm. I(135°,0) has a larger value for the remaining six characteristic spectral lines. S_3 is obtained by subtracting the intensities of I(135°, π/2) and I(45°, π/2), and the value of I(135°,π/2) is larger for all characteristic spectral lines. In addition, the polarization effect of the emitted light of the characteristic spectral lines from the plasma is analyzed. The emission light of most characteristic spectral lines has DOP values between 0.2 and 0.3, with the exception of H 656.3 nm. Among them, Na 589.0 nm has the smallest DOP value of 0.201, while H 656.3 nm has a DOP value of 0.525, showing a more obvious polarization effect. For metallic and non-metallic elements, the difference in the DOP of the emitted light may be due to the anisotropy of the particle distributions of the different elements in the plasma.

Table S2. Polarization parameters of the main characteristic spectral lines of tumor tissue.

Element	S_0	S_1	S_2	S_3	I_{pol}	DOP
CN 387.7 nm	77144.7	-15887.5	-8287.3	4810.1	18553.4	0.241
Ca 393.4 nm	39442.7	-7677.3	-6133.5	3728.3	10510.0	0.266
Ca 396.8 nm	37895.2	-857.3	-7112.7	5384.6	8962.1	0.236
Ca 422.7 nm	47627.9	-9063.4	-8577.8	6124.5	13900.8	0.292
C_2 517.9 nm	6845.1	210.5	385.7	1982.5	2030.6	0.297
Na 589.0 nm	192305.3	-16296.3	-34548.2	6294.4	38713.9	0.201
H 656.3 nm	41712.6	3946.5	-7551.0	20169.9	21895.6	0.525

5.Average DOPs of characteristic spectral lines of three different tissues (normal, border and tumor).

Table S3. The average DOPs of the characteristic spectral lines of normal tissue.

Number of the tissue samples	Ca 393.4 nm	Ca 422.7 nm	Na 589.0 nm
No.1	0.596	0.485	0.203
No.2	0.354	0.403	0.190
No.3	0.307	0.452	0.189

No.4	0.322	0.389	0.173
No.5	0.350	0.408	0.320
No.6	0.302	0.353	0.213
No.7	0.397	0.361	0.190
No.8	0.302	0.340	0.198
No.9	0.336	0.386	0.348
No.10	0.343	0.396	0.189
No.11	0.327	0.410	0.234
No.12	0.353	0.386	0.351
No.13	0.307	0.408	0.200
No.14	0.288	0.361	0.182
Average values	0.349	0.396	0.227

Table S4. The average DOPs of the characteristic spectral lines of the boundary tissue.

Number of the tissue samples	Ca 393.4 nm	Ca 422.7 nm	Na 589.0 nm
No.15	0.294	0.340	0.149
No.16	0.262	0.297	0.204
No.17	0.363	0.408	0.190
No.18	0.332	0.313	0.247
No.19	0.410	0.424	0.411
No.20	0.501	0.439	0.195
No.21	0.306	0.400	0.179
No.22	0.311	0.292	0.212
No.23	0.382	0.482	0.504
No.24	0.373	0.391	0.398
No.25	0.401	0.504	0.352
No.26	0.570	0.524	0.251
No.27	0.466	0.485	0.461
No.28	0.271	0.311	0.189
No.29	0.434	0.383	0.194
No.30	0.279	0.339	0.189
No.31	0.328	0.377	0.175
No.32	0.305	0.425	0.333
No.33	0.219	0.344	0.198
No.34	0.266	0.306	0.204
No.35	0.239	0.383	0.183
Average values	0.348	0.389	0.258

Table S5. The average DOPs of the characteristic spectral lines of the tumor tissues.

Number of the tissue samples	Ca 393.4 nm	Ca 422.7 nm	Na 589.0 nm
No.36	0.288	0.324	0.204
No.37	0.856	0.790	0.227

No.38	0.436	0.456	0.215
No.39	0.296	0.430	0.188
No.40	0.401	0.407	0.161
No.41	0.657	0.507	0.387
No.42	0.551	0.484	0.235
No.43	0.253	0.303	0.200
No.44	0.358	0.395	0.158
No.45	0.320	0.446	0.452
No.46	0.332	0.433	0.152
No.47	0.250	0.312	0.215
No.48	0.411	0.414	0.220
No.49	0.505	0.436	0.236
No.50	0.262	0.359	0.206
No.51	0.287	0.310	0.205
No.52	0.533	0.469	0.419
No.53	0.307	0.416	0.244
No.54	0.276	0.304	0.190
No.55	0.293	0.420	0.231
No.56	0.248	0.379	0.199
Average values	0.387	0.419	0.235

6.The best parameters of each classification model are optimized using the 10-fold cross-validation method.

Table S6. Optimal parameters for each classification model.

Machine learning models	SVM		PLS-DA	RF	KNN
Hyperparameters	*c*	*g*	Number of components	Number of trees	*k*
LIBS (All spectral peaks)	100	102.14	19	90	6
LIBS (Selected spectral peaks)	100	1000	16	100	8
Polarization parameters	86.87	16.71	11	200	9
PRLIBS fusion	100	1000	18	200	7

7.Classification results of gliomas based on polarization parameters

In Section 3.3.1, the seven characteristic spectral lines of the LIBS spectrum were selected to build classification models, which can enhance the classification performance of the model. However, traditional LIBS classification models do not take into account the polarization properties of the plasma. Hence, this section attempts to use the polarization parameters of the characteristic spectral lines to build the classification model. From the deduction in Section 3.2, it can be seen that 6 polarization parameters can be calculated for each spectral peak, including: the four Stokes parameters (S_0, S_1, S_2, and S_3), the total intensity of the polarization component (I_{pol}), and the degree of polarization (DOP). It is possible to calculate a total of 42 polarization parameters for the seven spectral peaks in the PRLIBS spectrum. Polarization parameters were calculated for the entire training and test sets, and the final dataset dimensions were 2100×42 and 700×42, respectively. Subsequently, maximum-minimum normalization (in the range

between [0,1]) was applied to each one-dimensional variable, and the normalized data were used to build SVM, PLS-DA, RF and KNN classification models, respectively.

The prediction results of the models are shown in Fig. 7. For the training set, both the SVM and RF models had 100% prediction accuracy, PLS-DA and KNN showed a slight decrease in prediction accuracy compared to the models based on LIBS intensity spectra. However, for the test set, the prediction accuracies of all the models are less than 50%, which indicates a poor classification performance of the models. In particular, for the SVM and PLS-DA models, the prediction accuracies are 36.86% and 30.86%, respectively. The poor classification performance of the model established using polarization parameters is mainly caused by the following factors. In this research, isolated fresh tissue was used, and the matrix effect was more obvious between different ablation points. As a result, the RSD of polarization parameters is large. For example, for the characteristic spectral line Ca 393.4 nm, the RSD of the polarization spectrum PRLIBS (0°,0) is 26.51 %, but the RSD of the Stokes parameter S_1 is as high as 304.84 %. The larger the RSD of the polarization parameters, the machine learning model will not be able to effectively learn the patterns between the spectra of different types of tissues, resulting in a poorer accuracy of the test set.

8.Classification results of gliomas based on single-channel polarization spectra

As a comparison, the polarization spectra of the four channels in Fig. 6 are individually input into SVM, PLS-DA, RF and KNN to build the classification model. For the single-channel polarization spectra model, the data sizes of the training set and test set are 2100×155 and 700×155, respectively. Before building the classification model, the spectral variables were normalized using the total intensity normalization method. The same 10-fold cross-validation was used to optimize the parameters of the machine learning models.

The prediction results of each model are shown in Fig. S3. For the training set, the SVM and RF models have higher prediction accuracies, but the PLS-DA and KNN models have prediction accuracies less than 90%. For the test set, in the SVM model, the highest prediction accuracy of the model using PRLIBS (45°, 0) was 57.29%. Similarly, in the PLS-DA model, the highest model prediction accuracy of 56.43% was obtained using PRLIBS (45°, π/2). In the RF model, the prediction accuracy of the models using PRLIBS (45°,0) and PRLIBS (45°, π/2) were both about 60%. In the KNN model, the highest prediction accuracy of the model using PRLIBS (45°, π/2) was 54.43%. Compared to the prediction accuracy of the PRLIBS fusion model in Fig. 7, the single-channel polarization spectral model has poor classification performance. This is mainly because the single-channel polarization spectral model does not effectively utilize the polarization effect of the plasma. The variable dimension of the single-channel polarized spectral model is reduced to 1/4 of the PRLIBS fusion model, which is under-learned for spectral differences between different types of tissues.

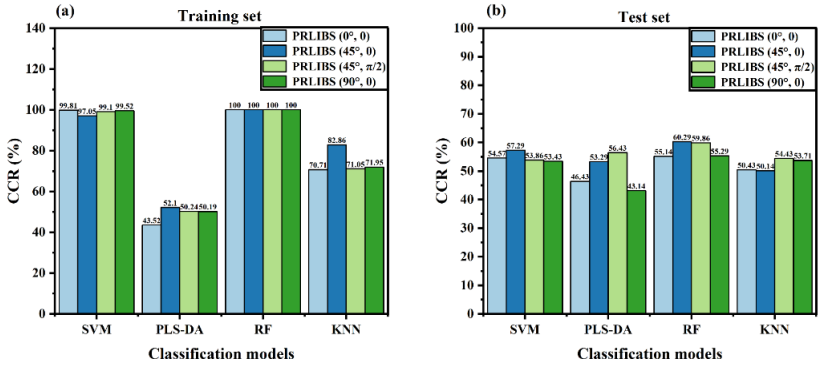

Fig. S3. Prediction results of modeling using single-channel polarization spectra. (a) Training set and (b) test set.

REFERENCES

1. Z. F. Zhao, W. P. Ma, G. Teng, X. J. Xu, K. Wei, G. Y. Chen, Q. Q. Wang, and W. S. Xu, *Spectrochim. Acta B*, 2023, **202**. https://doi.org/10.1016/j.sab.2023.106644