Supplementary Material of

High-precision Identification of Breast Cancer Based on End-to-end Parallel Spectral Convolutional Neural Network Assisted Laser-induced Breakdown Spectroscopy

Shengqun Shi, [a] Lingling Pi, [b] Lili Peng, [b] Deng Zhang, [d] Honghua Ma, [a] Yuanchao Liu, [e] Nan Deng, [a] Xiong Wang *[c] and Lianbo Guo *[a]

[a] *Wuhan National Laboratory for Optoelectronics (WNLO), Huazhong University of Science and Technology (HUST), Wuhan, Hubei 430074, P.R. China.*

[b] *School of Optical and Electronic Information, Huazhong University of Science and Technology, Wuhan, Hubei 430074, P.R. China.*

[c] *Department of Laboratory Medicine, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China.*

[d] *School of Computer and Electronic Information, Nanjing Normal University, Nanjing, Jiangsu 210023, P.R. China.*

[e] *Department of Physics, City University of Hong Kong, Kowloon 999077, Hong Kong SAR, China.*

1. PCA parameter optimization for traditional machine learning algorithms

To optimize the classification performance of the machine learning (ML) algorithm, we optimized the latent variables of the feature extraction algorithm principal component analysis (PCA). Specifically, we selected two ML algorithms, Random Forest (RF) and Linear Discriminant Analysis (LDA), as PCA-optimized classifiers. The optimizing process was conducted in two stages to ensure both a thorough search and precise optimization.

Initially, we explored a broad tuning range for the number of PCA components, spanning from 10 to 300. This wide range was chosen to cover a comprehensive set of potential values and to identify the general region where optimal performance might be found. To refine our search, we analyzed the results from the learning curves generated within this broad range. The learning curves provided insights into how different numbers of PCA components affected model performance. Based on analysis, we identified a more promising narrower range for further optimization, specifically from 10 to 75. Within this refined range, we conducted a more focused tuning process. By systematically evaluating the performance of the RF and LDA models with various numbers of PCA components within this range, we determined that the optimal number of PCA components was 30. This value provided the best balance between model complexity and performance, as indicated by the evaluation metrics. The results of the learning curve are shown in Figure S1

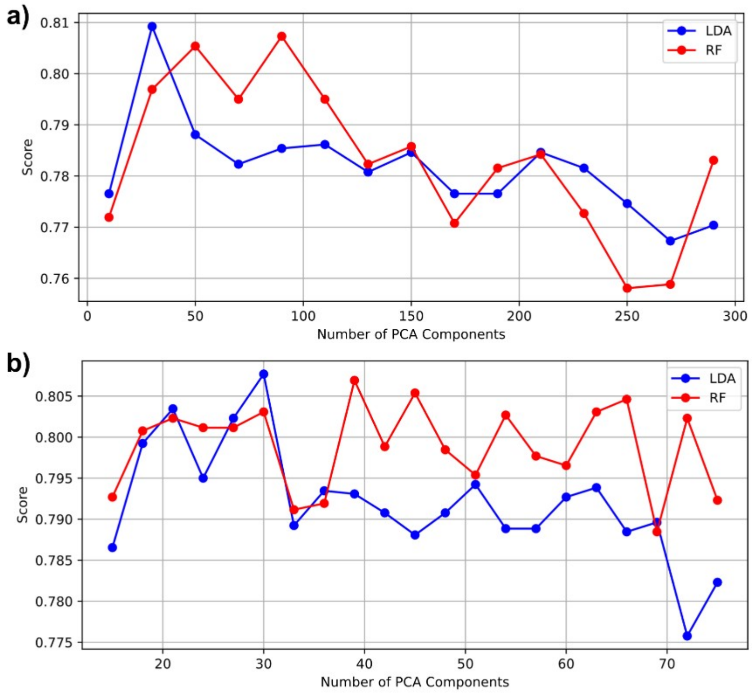

Figure S1. PCA optimizing parameter learning curve, the parameter ranges are a) [10,300], b) [10,75].

2. Single-task convolutional neural network classifier

In addition, as a classifier with better automatic feature extraction ability, we also built a single-task convolutional neural network (CNN) classifier for breast cancer identification. Table S1 is the specific layer structure of single-task CNN.

Table S1. The architecture and parameters of single-task CNN.

No.	Module	Layer type	Output shape
1	Classification module	Conv-1D	(None, 16, 12793)
2		BatchNorm	(None, 16, 12793)
3		ReLU	(None, 16, 12793)
4		MaxPool-1D	(None, 16, 4264)
5		Conv-1D	(None, 32, 2131)
6		BatchNorm	(None, 32, 2131)
7		ReLU	(None, 32, 2131)
8		MaxPool-1D	(None, 32, 710)
9		Conv-1D	(None, 8, 354)
10		ReLU	(None, 8, 354)
11		Flatten	(None, 2832)
12		Linear	(None, 64)
13		ReLU	(None, 64)
14		Dropout	(None, 64)
15		Linear	(None, 32)
16		ReLU	(None, 32)
17		Linear	(None, 2)

3. The specific layer structure of our PSCNN

Similarly, the specific layer structure of the PSCNN architecture proposed in this work is shown in Table S2.

Table S2. The architecture and parameters of PSCNN.

No.	Module	Layer type	Output shape
1	Preprocessing shared module of broadband spectra	Conv-1D	(None, 16, 24564)
2		BatchNorm	(None, 16, 24564)
3		ReLU	(None, 16, 24564)
4		Conv-1D	(None, 64, 24564)
5		BatchNorm	(None, 64, 24564)
6		ReLU	(None, 64, 24564)
7		Conv-1D	(None, 8, 24564)
8		ReLU	(None, 8, 24564)
9		Conv-1D	(None, 64, 24564)
10		ReLU	(None, 64, 24564)
11		Conv-1D	(None, 1, 24564)
12		ReLU	(None, 1, 24564)
13		Flatten	(None, 24564)
14	Preprocessing shared module of narrowband spectra	Conv-1D	(None, 16, 1024)
15		BatchNorm	(None, 16, 1024)
16		ReLU	(None, 16, 1024)
17		Conv-1D	(None, 32, 1024)
18		BatchNorm	(None, 32, 1024)
19		ReLU	(None, 32, 1024)
20		Conv-1D	(None, 8, 1024)
21		ReLU	(None, 8, 1024)
22		Conv-1D	(None, 16, 1024)
23		ReLU	(None, 16, 1024)
24		Conv-1D	(None, 1, 1024)

25		ReLU	(None, 1, 1024)
26		Flatten	(None, 1024)
27	Breast cancer qualitative classification module	Conv-1D	(None, 16, 12794)
28		BatchNorm	(None, 16, 12794)
29		ReLU	(None, 16, 12794)
30		MaxPool-1D	(None, 16, 4264)
31		Conv-1D	(None, 8, 2132)
32		BatchNorm	(None, 8, 2132)
33		ReLU	(None, 8, 2132)
34		MaxPool-1D	(None, 8, 710)
35		Flatten	(None, 5680)
36		Linear	(None, 128)
37		ReLU	(None, 128)
38		Dropout	(None, 128)
39		Linear	(None, 2)