Supporting Information for:

Development of robust models for rapid classification of microplastic polymer types based on near infrared hyperspectral images

Tomo Kitahashi,*[a] Ryota Nakajima,[a] Hidetaka Nomaki,[b] Masashi Tsuchiya,[a] Akinori Yabuki,[a] Sojiro Yamaguchi,[c] Chunmao Zhu,[d] Yugo Kanaya,[d] Dhugal J. Lindsay,[e] Sanae Chiba[a] and Katsunori Fujikura[a]

[a] Marine Biodiversity and Environmental Assessment Research Center (BioEnv), Research Institute for Global Change (RIGC), Japan Agency for Marine-Earth Science and Technology (JAMSTEC), 2-15 Natsushima-cho, Yokosuka, Kanagawa 237-0061, Japan

[b] SUper-cutting-edge Grand and Advanced Research (SUGAR) Program, Institute for eXtra-cutting-edge Science and Technology Avant-garde Research (X-star), Japan Agency for Marine-Earth Science and Technology (JAMSTEC), 2-15 Natsushima-cho, Yokosuka, Kanagawa 237-0061, Japan

[c] JFE Techno Research, 1 Kawasaki-cho, Chuo-ku, Chiba 260-0835, Japan

[d] Earth Surface System Center (ESS), Research Institute for Global Change (RIGC), Japan Agency for Marine-Earth Science and Technology (JAMSTEC), 3173-25, Showa-machi, Kanazawa-ku, Yokohama, Kanagawa 236-0001, Japan

[e] Advanced Science and TEchnology Research (ASTER) Program, Institute for eXtra-cutting-edge Science and Technology Avant-garde Research (X-star), Japan Agency for Marine-Earth Science and Technology (JAMSTEC), 2-15 Natsushima-cho, Yokosuka, Kanagawa 237-0061, Japan

Number of pages: 21
Number of figures: 20

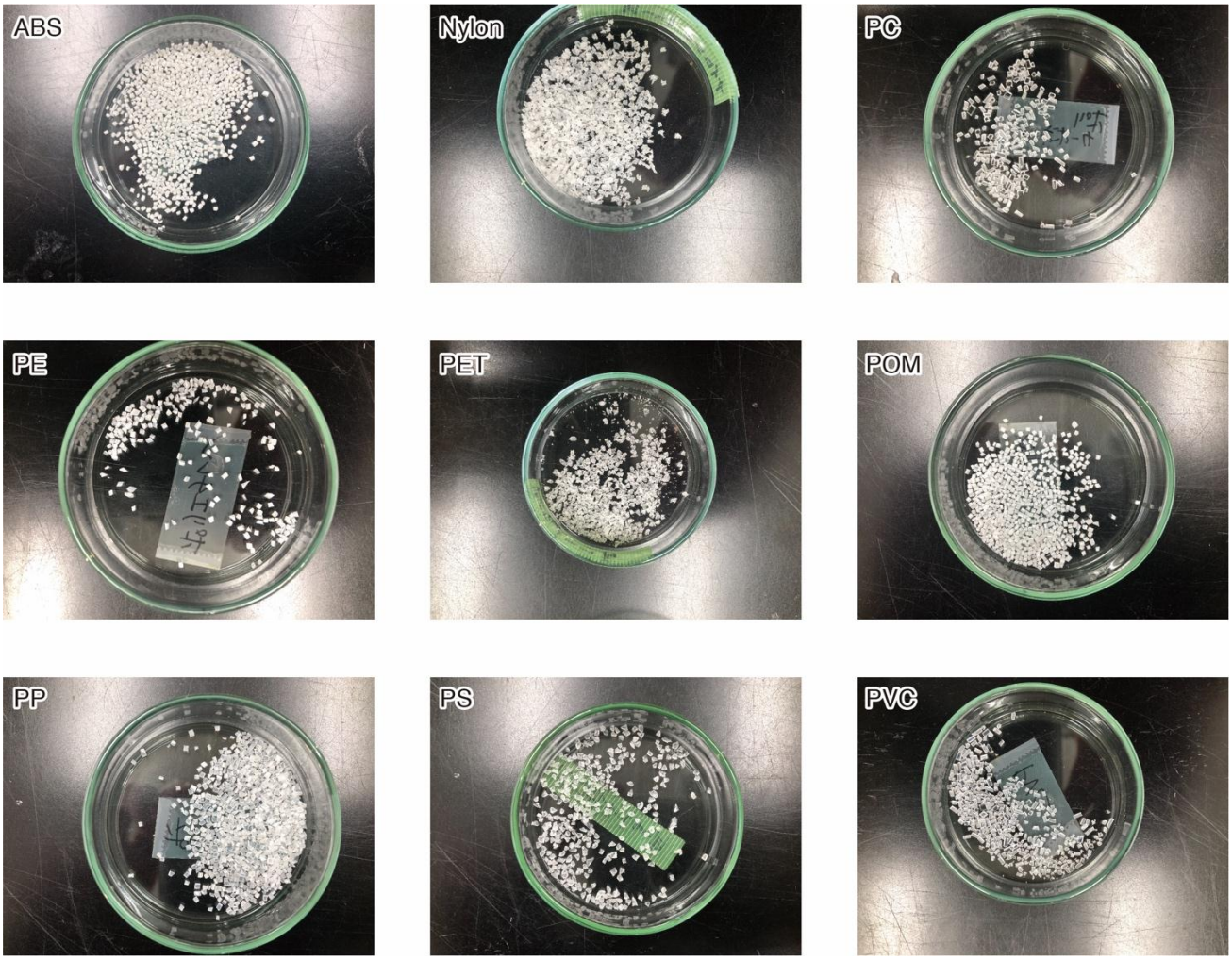

Figure S1. Pictures of polymer particles of 1 mm used in this study. Diameter of each petri dish is 5 cm.

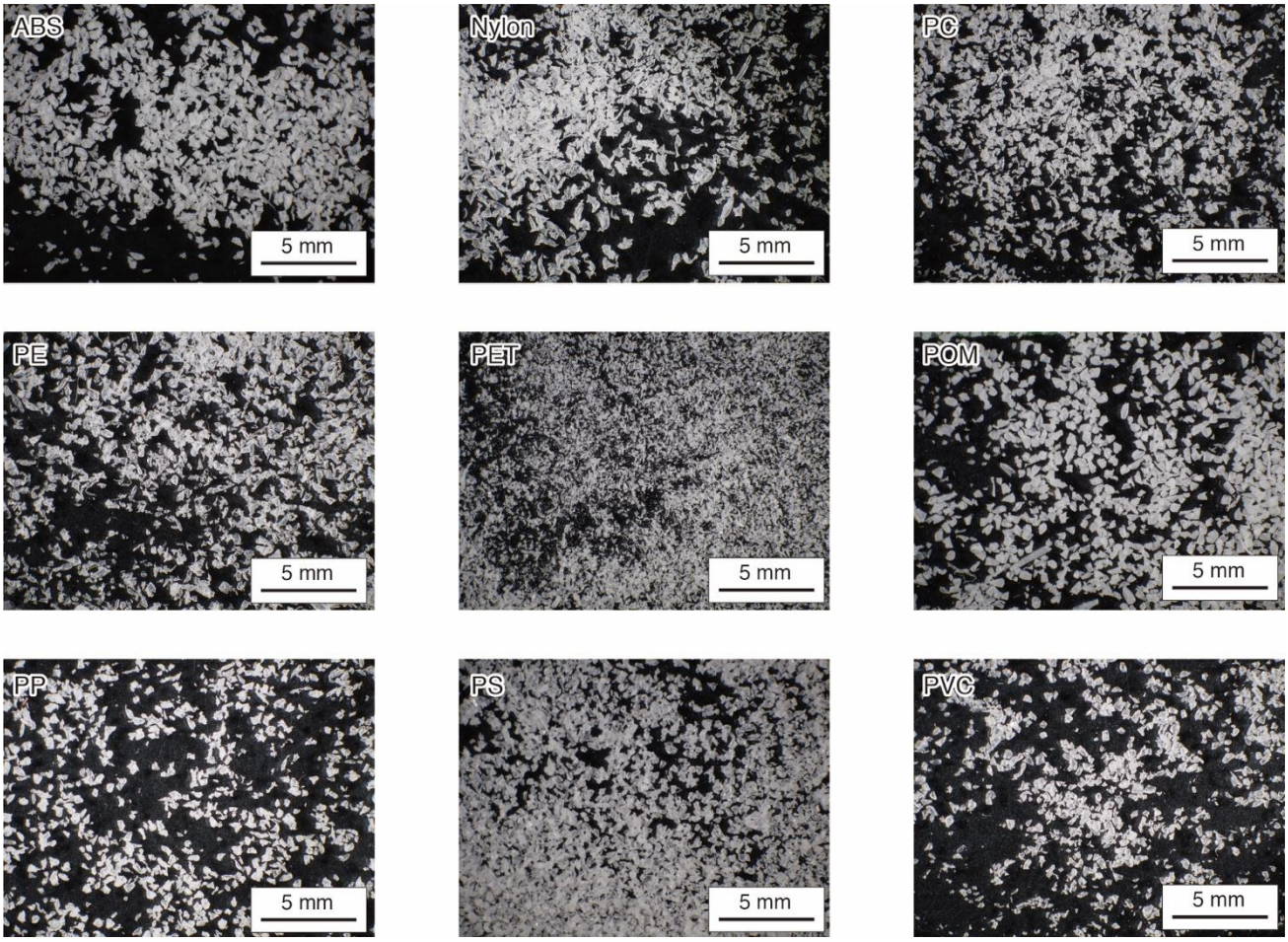

Figure S2. Pictures of polymer particles of 100–500 µm used in this study.

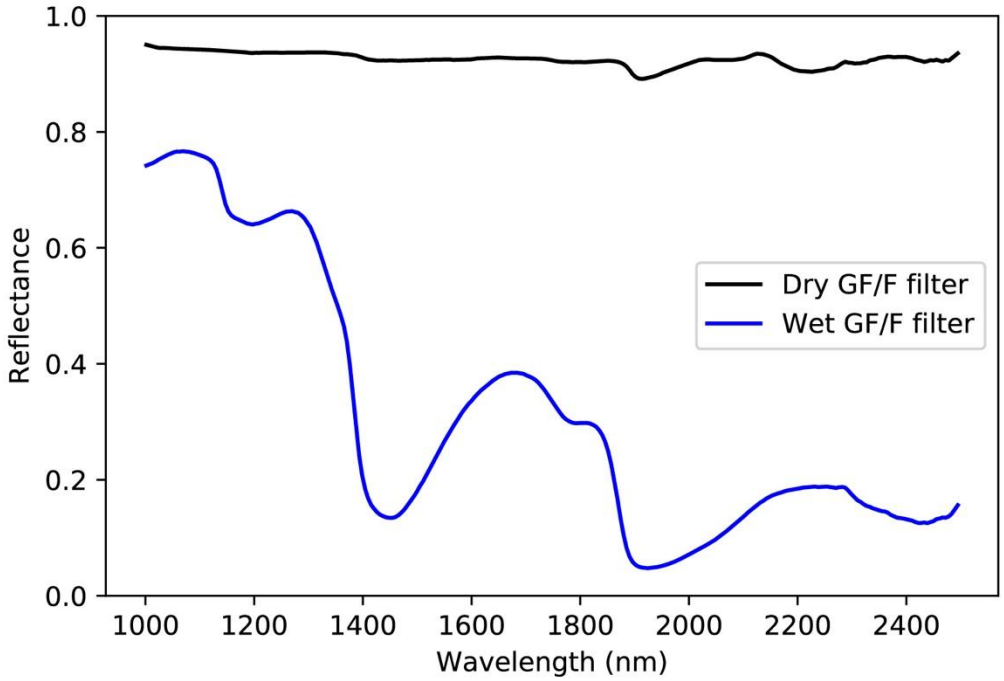

Figure S3. The spectral pattern of the dry and wet GF/F glass filter.

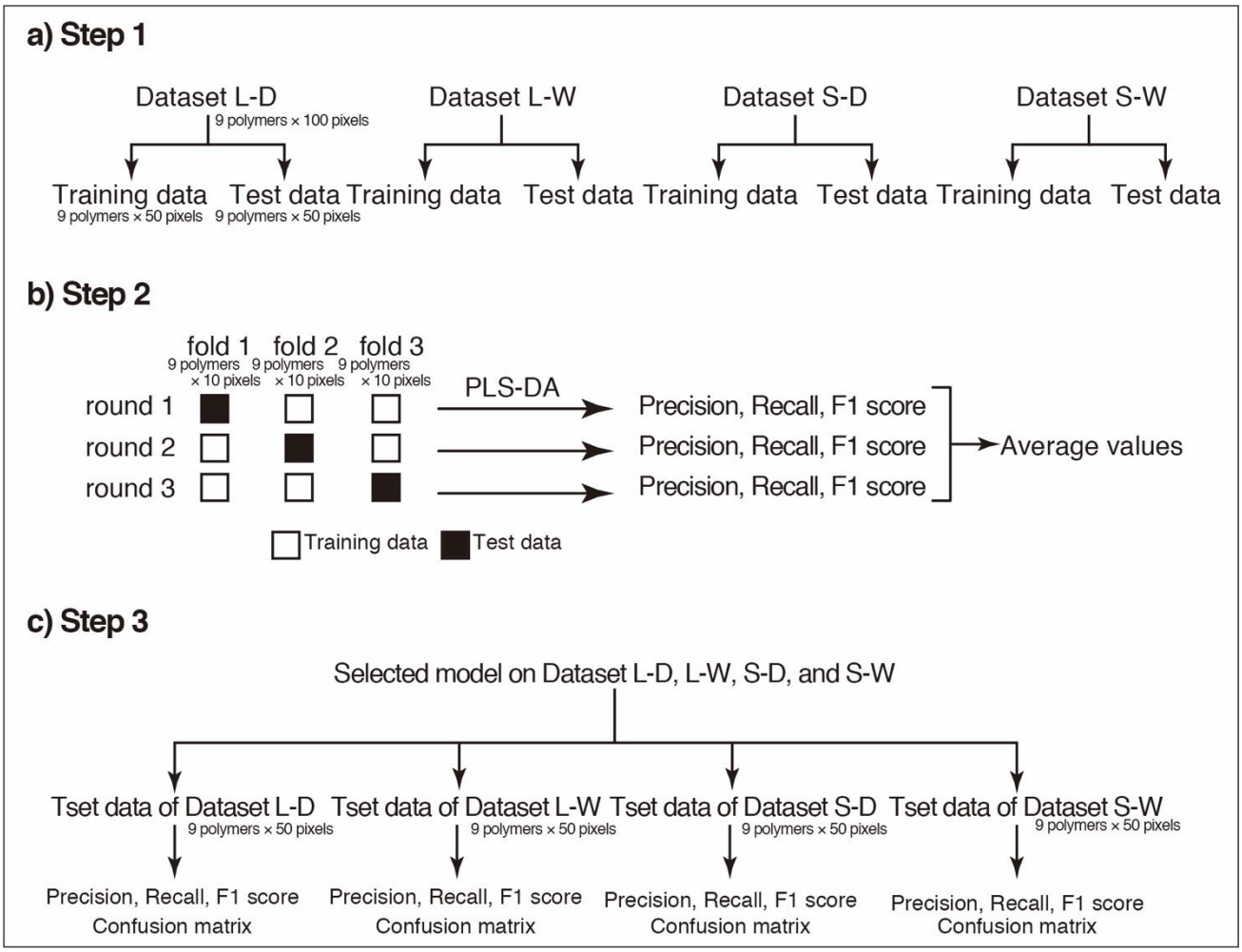

Figure S4. Schematic flow of model construction in this study. a) Each dataset was split into two subsets: training and test data. The training data were used to constrain the classification model. b) For the construction of classification models, we adopt K-fold cross validation. The training data were split into three subsets (folds) and the model was constrained with PLS-DA methods. For each round, F1 scores were calculated and averaged. In this step, the models were constrained at PLS ranges from 1 to 20. The "best" model was selected according to the criteria (see text). c) The selected models in step b) were validated on the test data of all datasets and F1 scores were calculated.

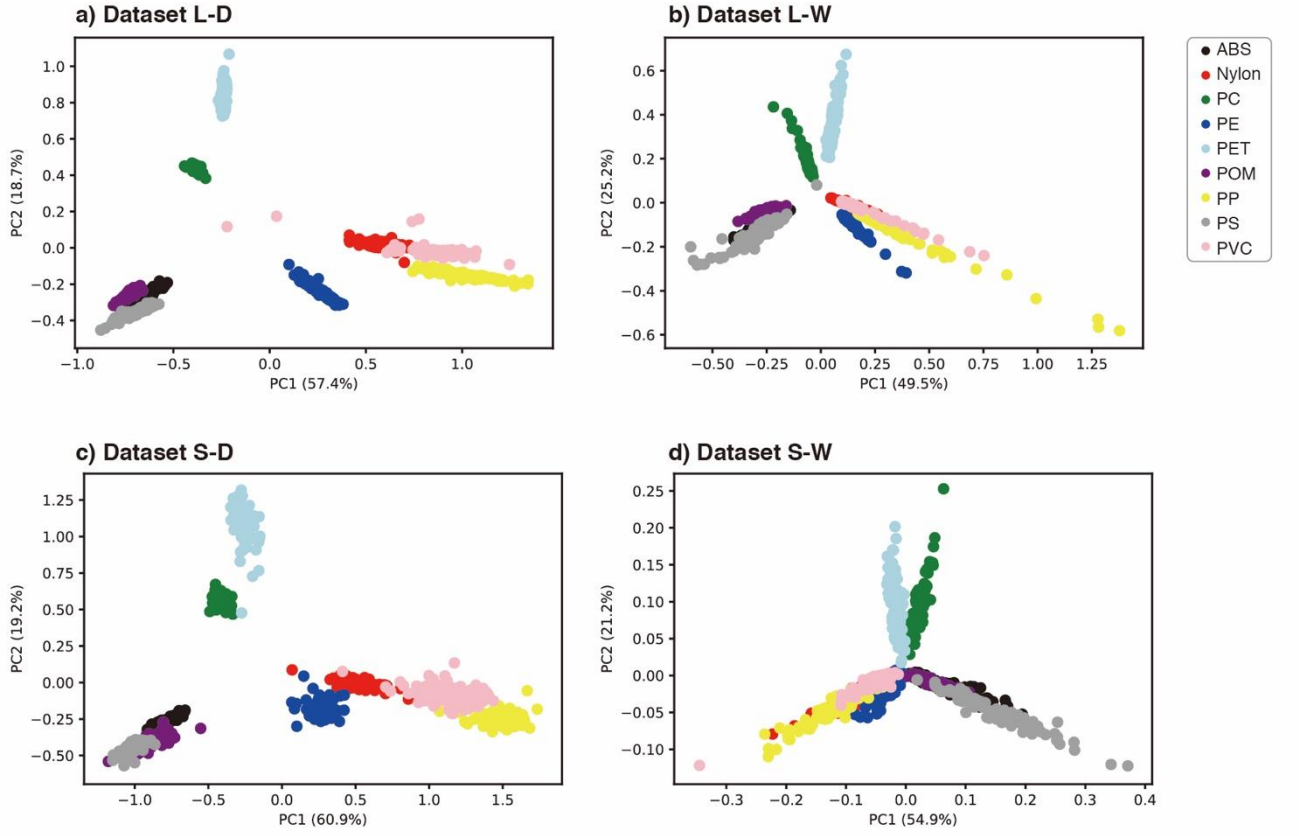

Figure S5. PCA plot based on the spectral data in the range of 1000–1700 nm wavelength for the 9 polymer types. a) data of 1 mm particles measured on a dry filter (Dataset L-D). b) data of 1 mm particles measured on a wet filter (Dataset L-W). c) data of 100–500 µm particles measured on a dry filter (Dataset S-D). d) data of 100–500 µm particles measured on a wet filter (Dataset S-W).

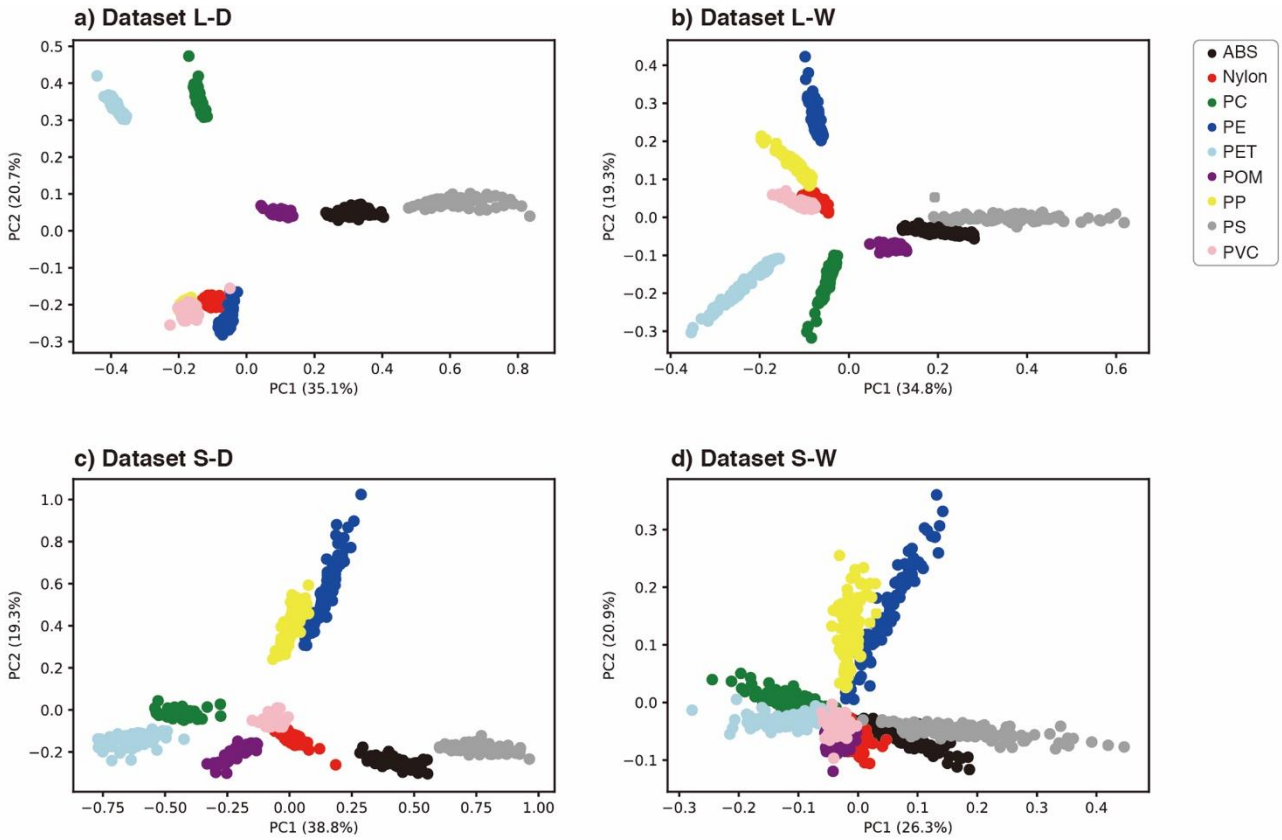

Figure S6. PCA plot based on the spectral data in the range of 1000–2500 nm wavelength for the 9 polymer types. a) data of 1 mm particles measured on a dry filter (Dataset L-D). b) data of 1 mm particles measured on a wet filter (Dataset L-W). c) data of 100–500 µm particles measured on a dry filter (Dataset S-D). d) data of 100–500 µm particles measured on a wet filter (Dataset S-W).

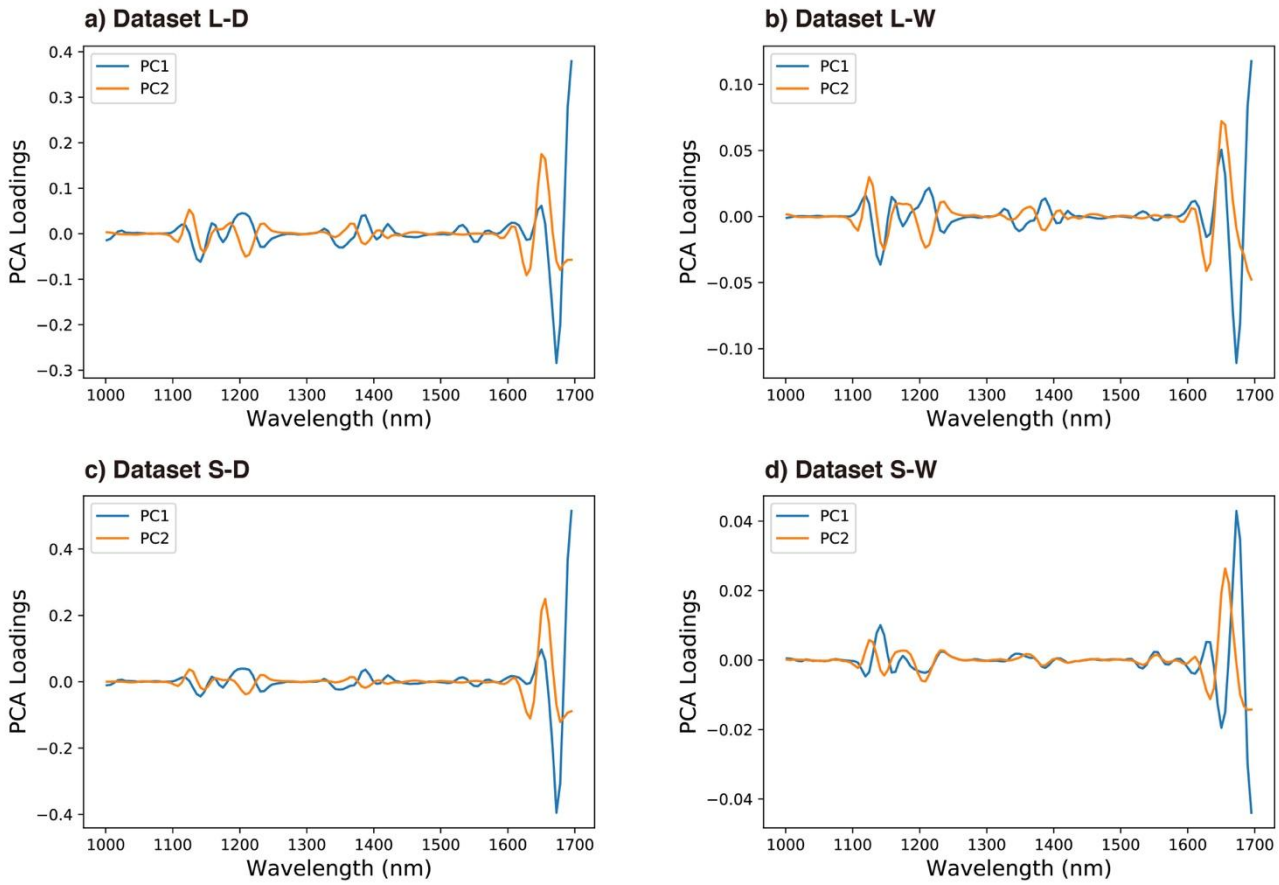

Figure S7. PCA loading plots of PC1 and PC2 of PCA analyses based on the spectral data in the range of 1000–1700 nm wavelengths.

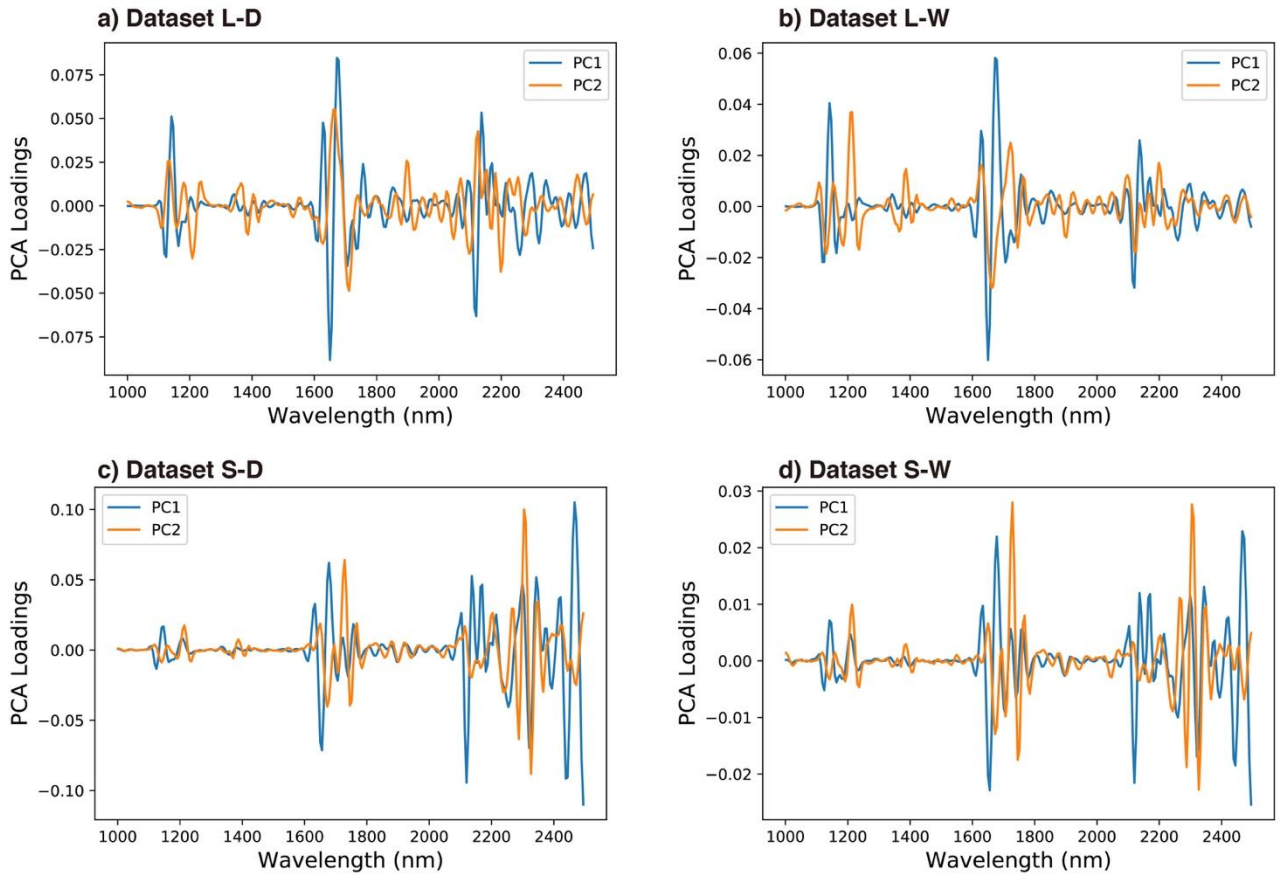

Figure S8. PCA loading plots of PC1 and PC2 of PCA analyses based on the spectral data in the range of 1000–2500 nm wavelengths.

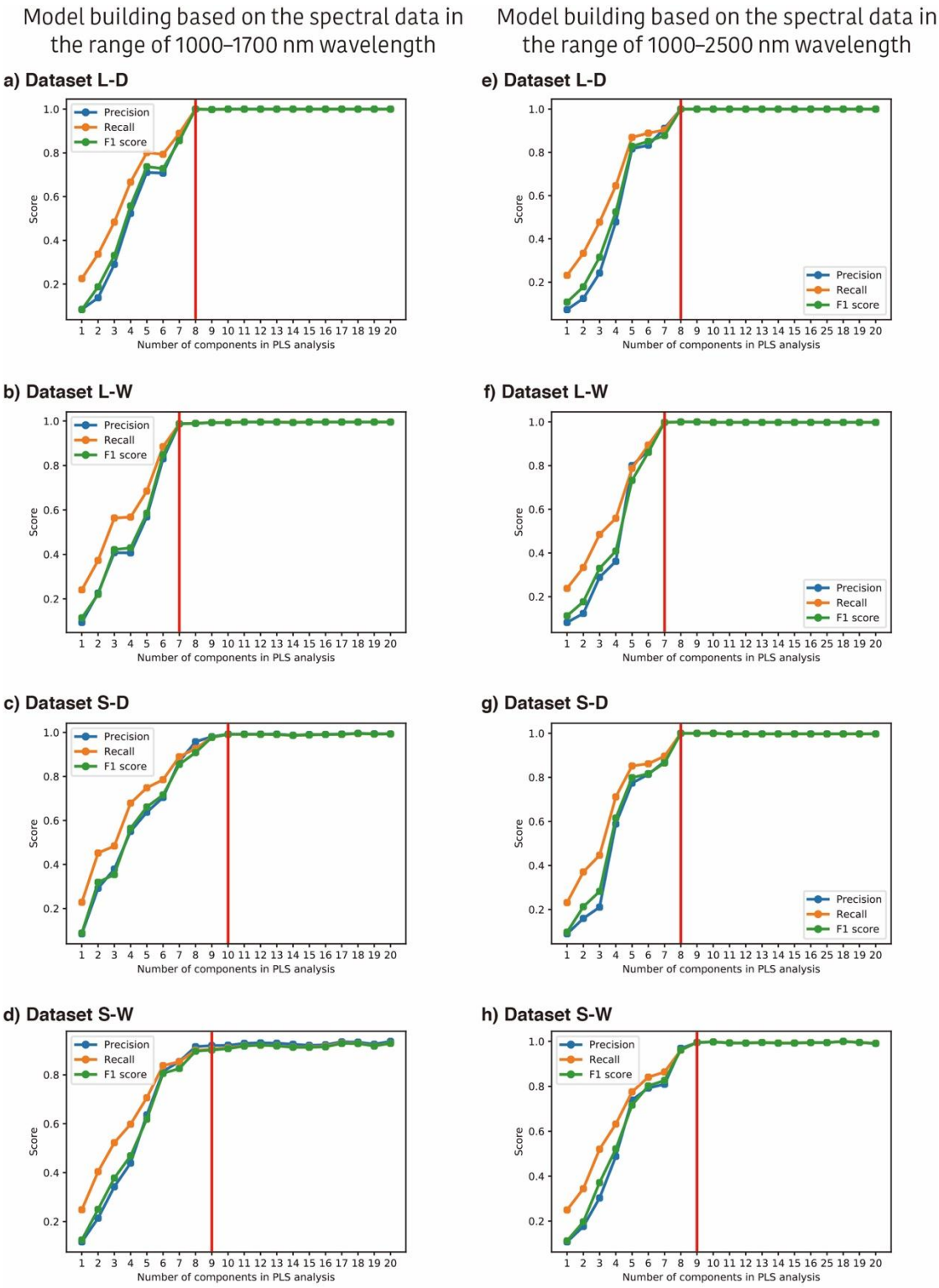

Figure S9. Relationships between the number of components in the PLS analysis and evaluating indices (Precision, Recall, and F1 score) for each model construction process. Left and right panels show model construction processes based on the spectral data in the ranges of 1000–1700 nm and 1000–2500 nm wavelengths, respectively. The adopted number of components in the PLS-DA analysis are indicated by a red line.

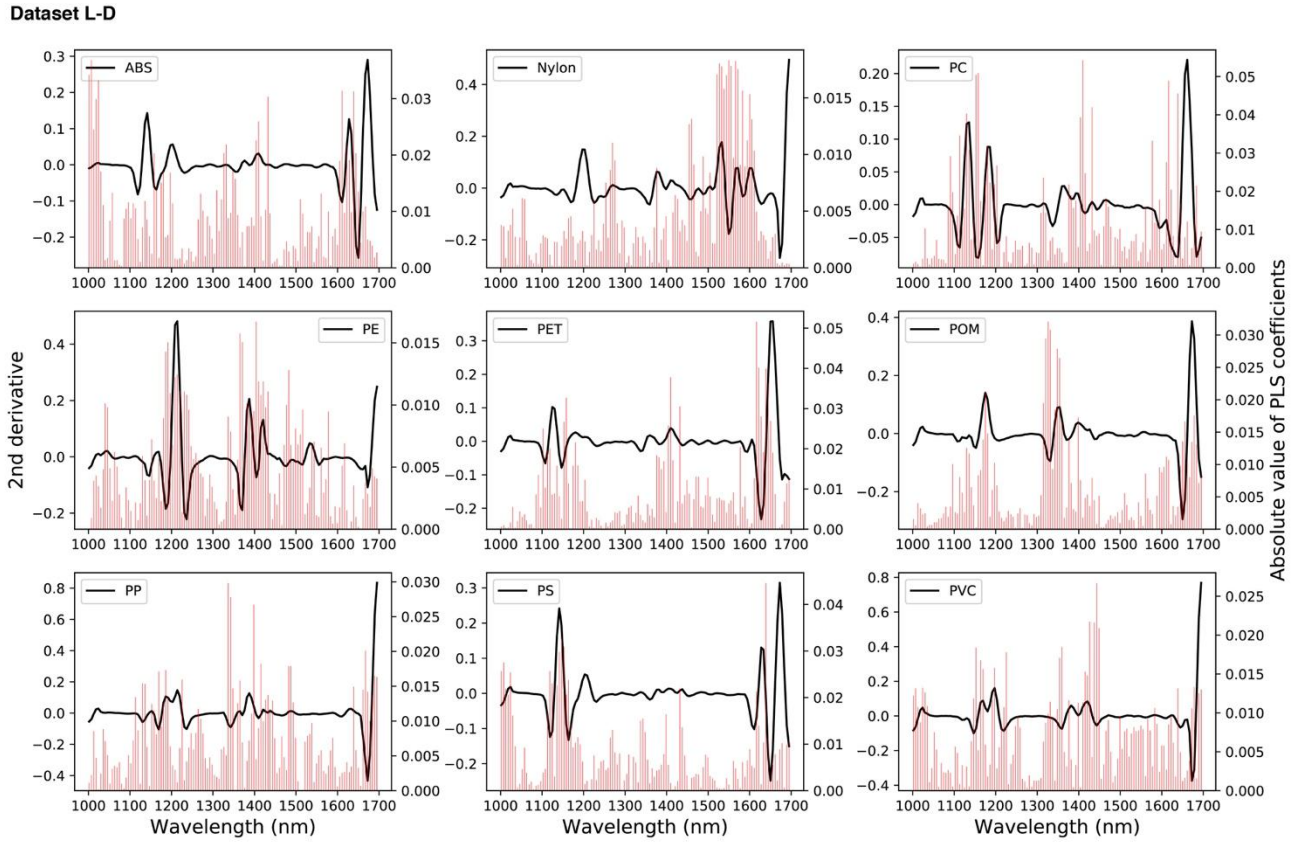

Figure S10. The relationships between the absolute value of the PLS coefficients of the model based on Dataset L-D and the preprocessed spectra for each polymer type in the range of 1000–1700 nm.

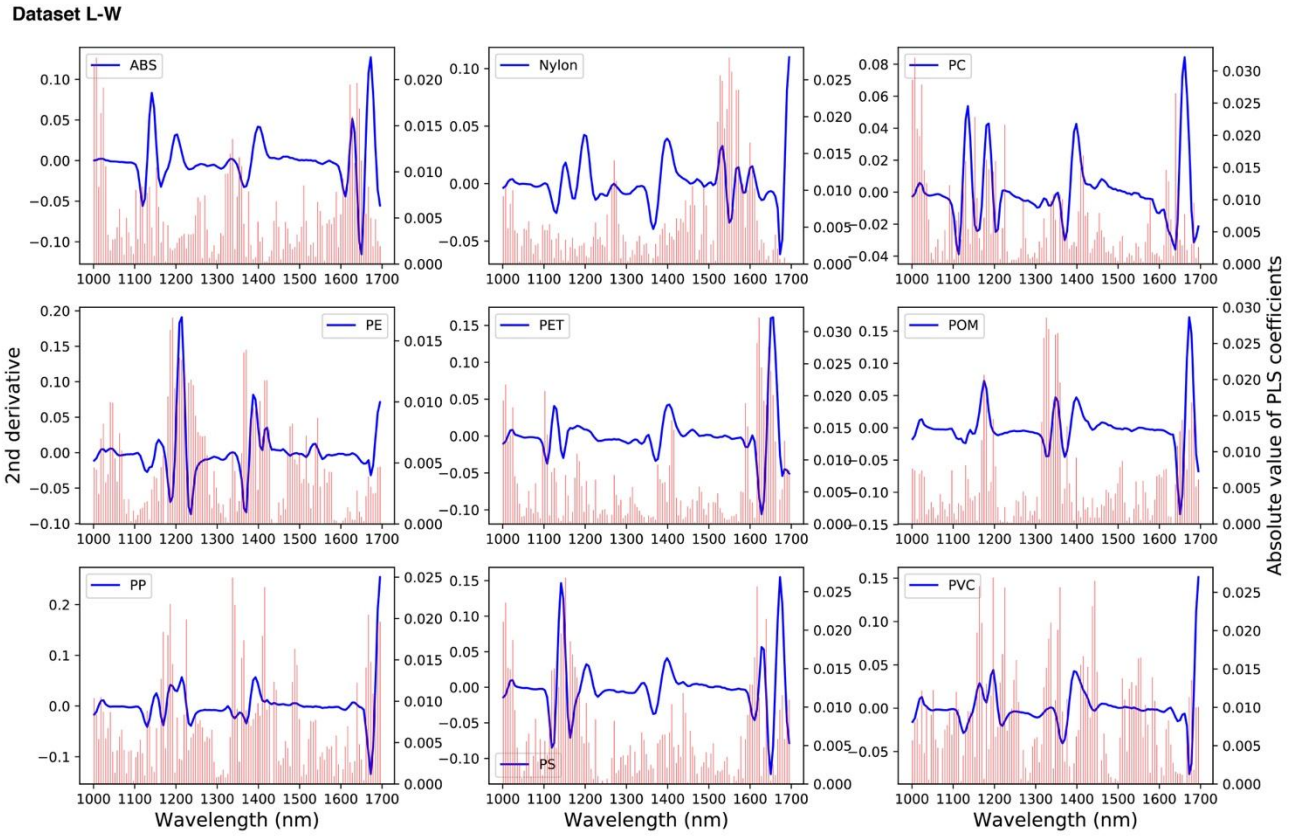

Figure S11. The relationships between the absolute value of the PLS coefficients of the model based on Dataset L-W and the preprocessed spectra for each polymer type in the range of 1000–1700 nm.

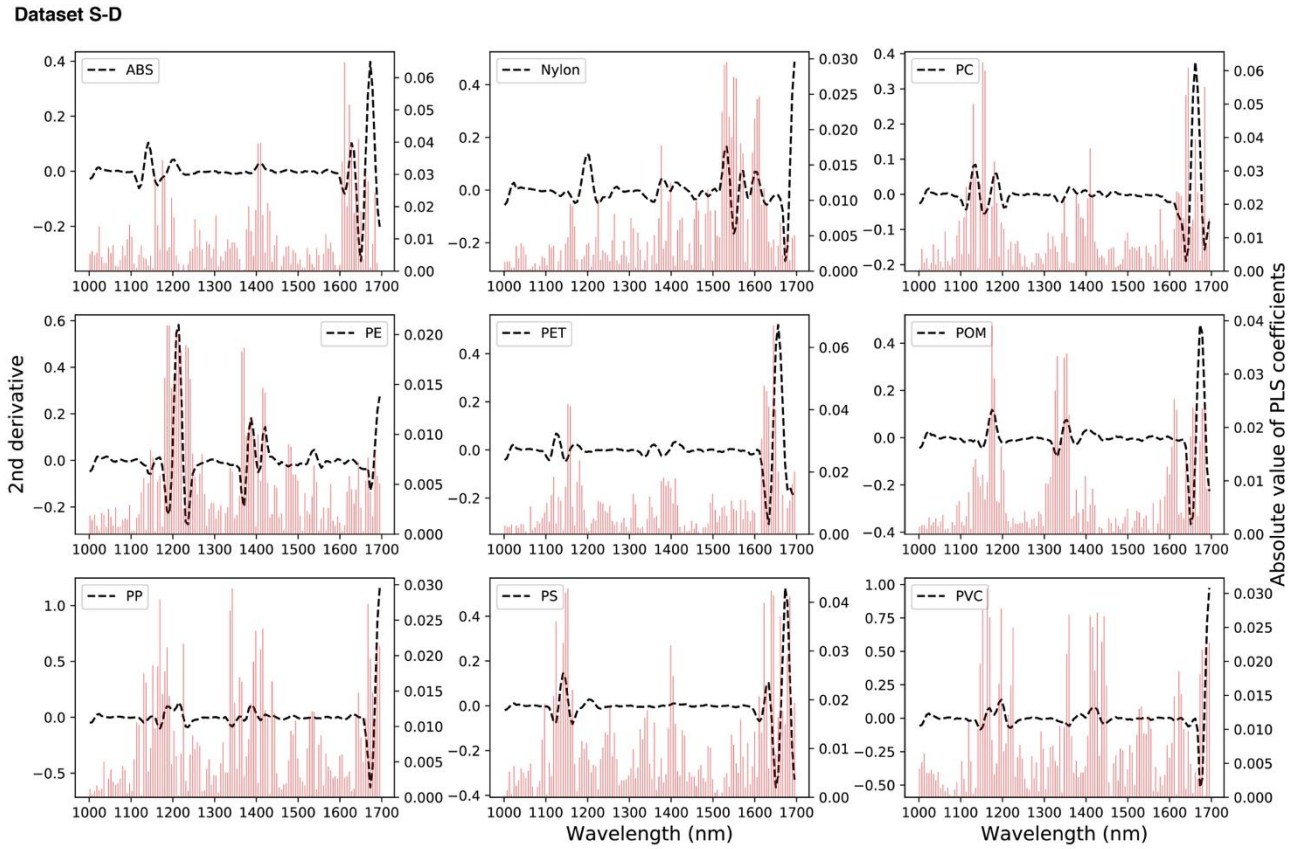

Figure S12. The relationships between the absolute value of the PLS coefficients of the model based on Dataset S-D and the preprocessed spectra for each polymer type in the range of 1000–1700 nm.

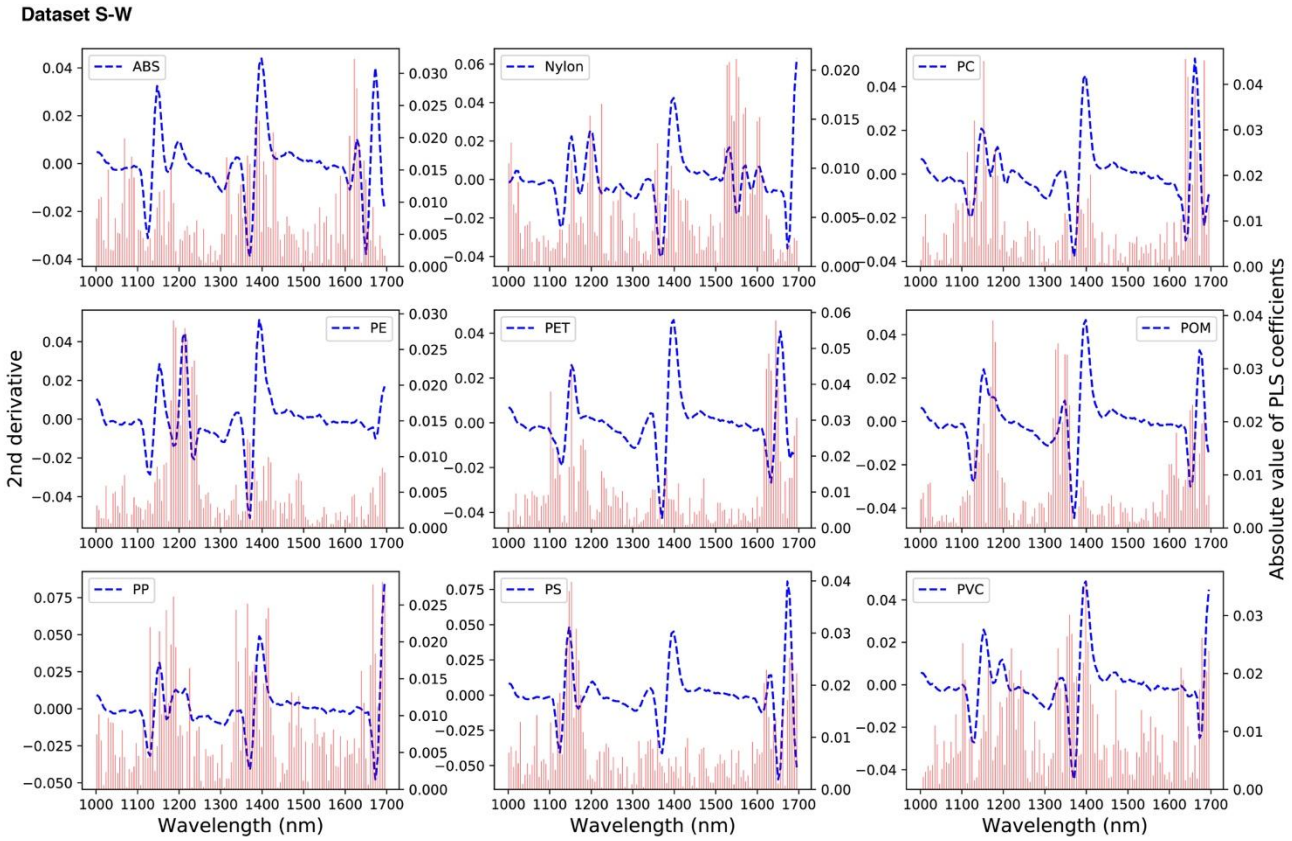

Figure S13. The relationships between the absolute value of the PLS coefficients of the model based on Dataset S-W and the preprocessed spectra for each polymer type in the range of 1000–1700 nm.

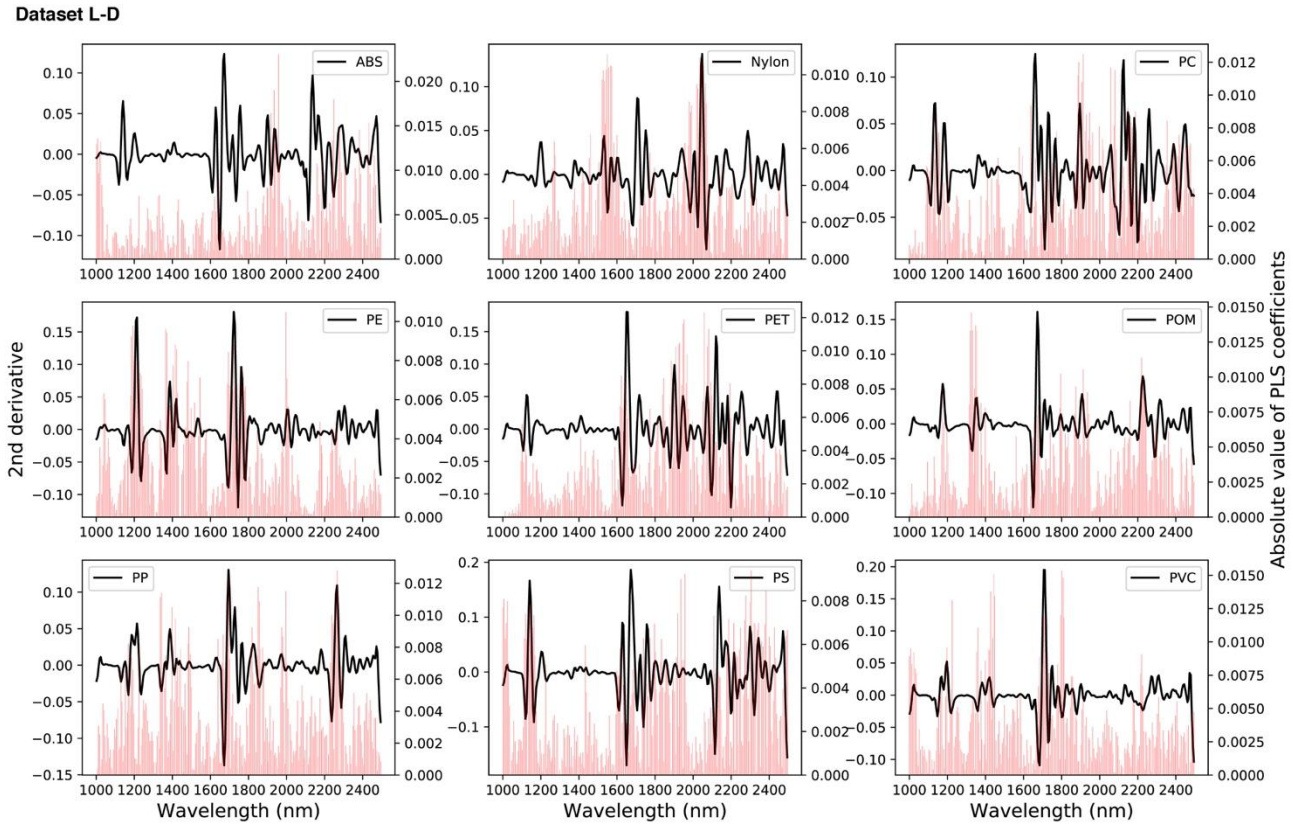

Figure S14. The relationships between the absolute value of the PLS coefficients of the model based on Dataset L-D and the preprocessed spectra for each polymer type in the range of 1000–2500 nm.

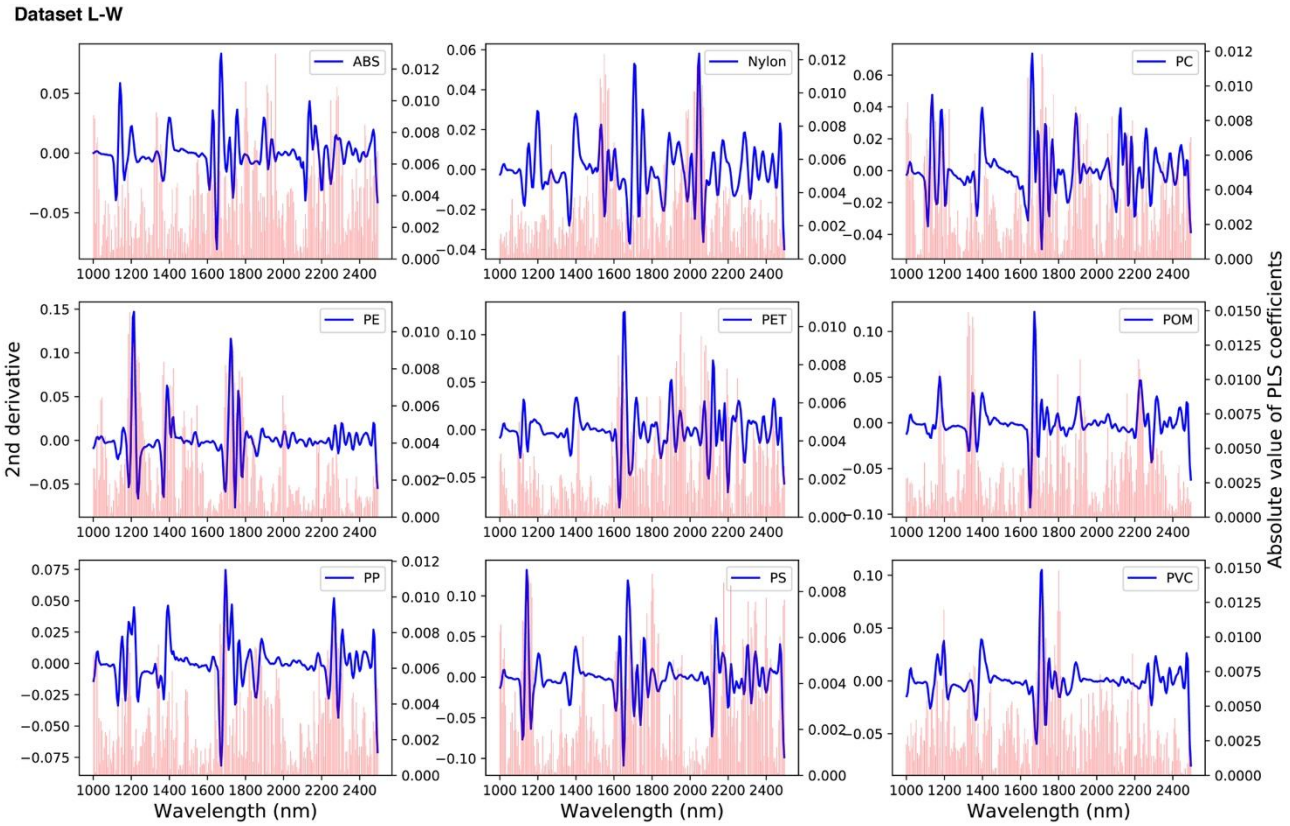

Figure S15. The relationships between the absolute value of the PLS coefficients of the model based on Dataset L-W and the preprocessed spectra for each polymer type in the range of 1000–2500 nm.

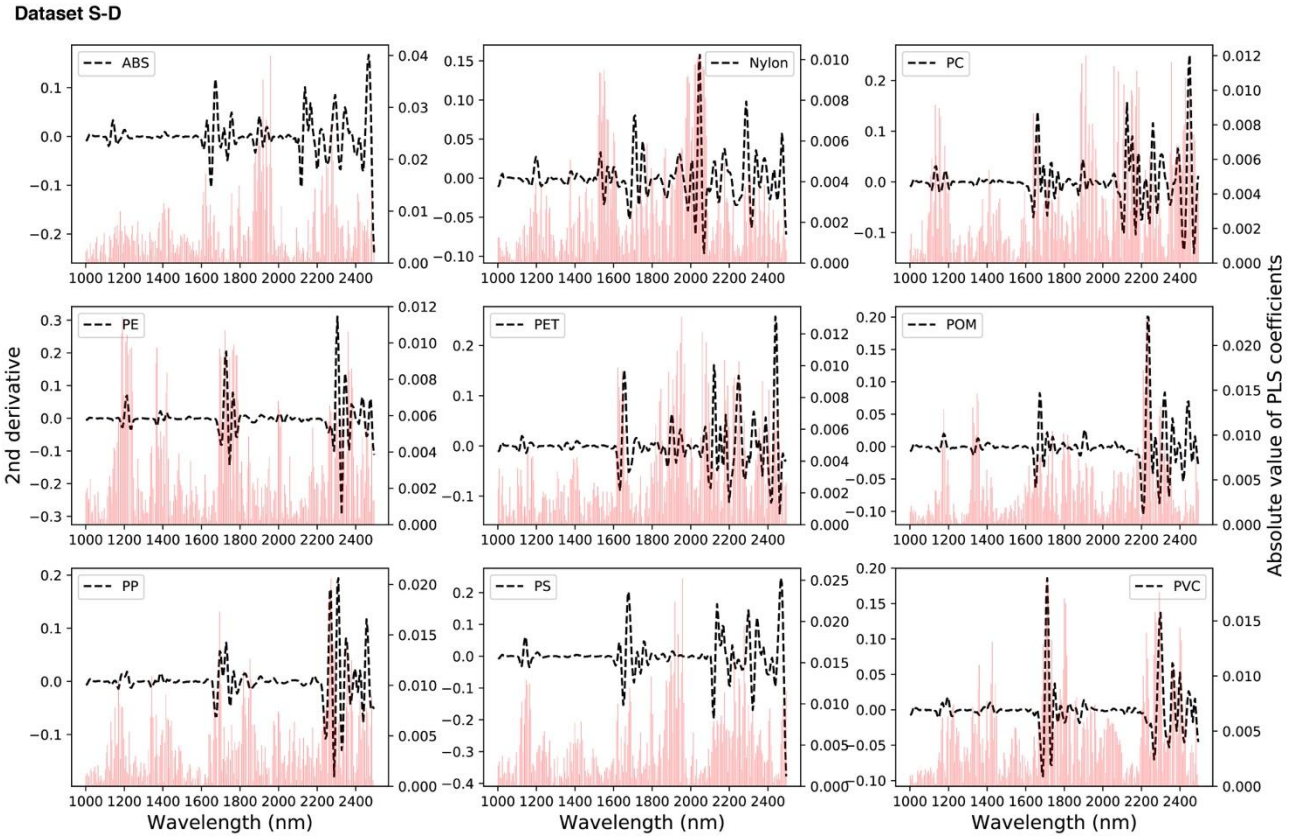

Figure S16. The relationships between the absolute value of the PLS coefficients of the model based on Dataset S-D and the preprocessed spectra for each polymer type in the range of 1000–2500 nm.

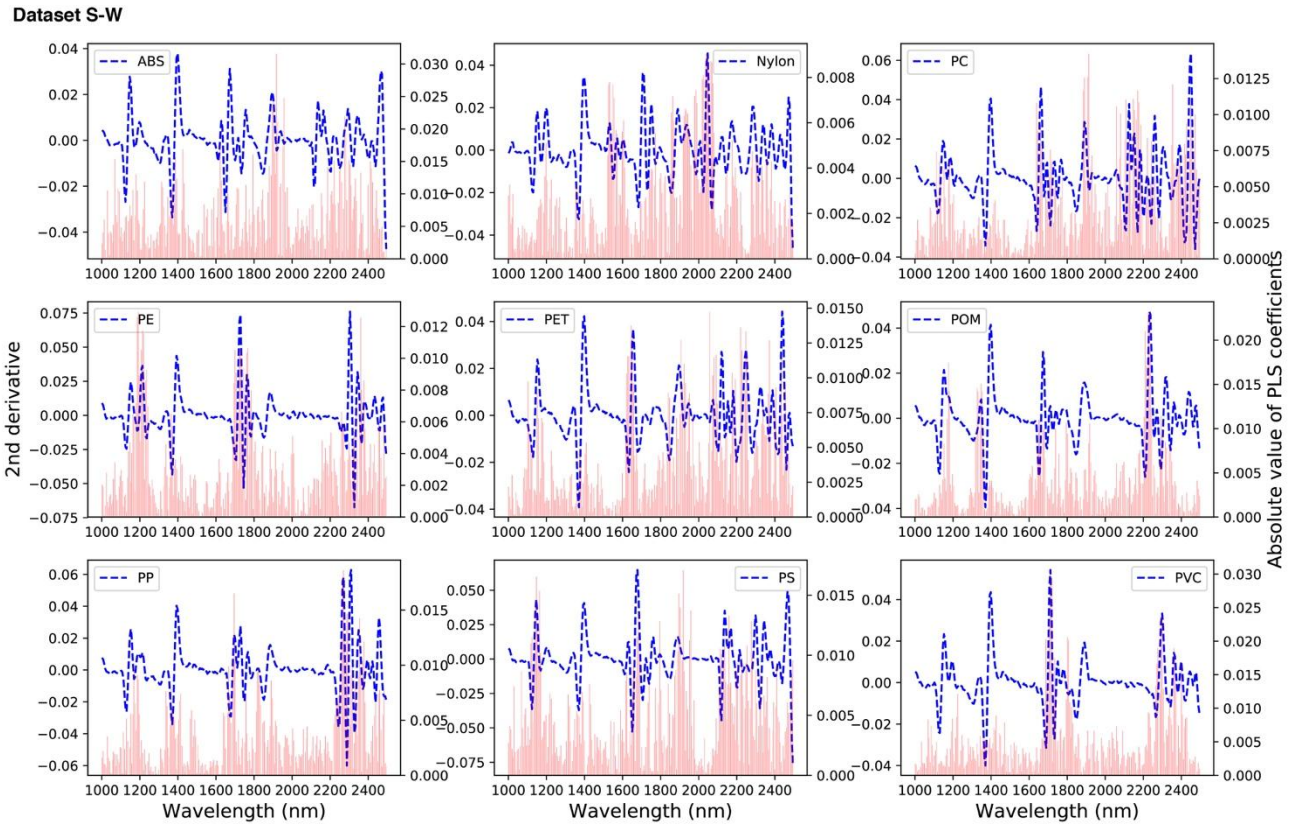

Figure S17. The relationships between the absolute value of the PLS coefficients of the model based on Dataset S-W and the preprocessed spectra for each polymer type in the range of 1000–2500 nm.

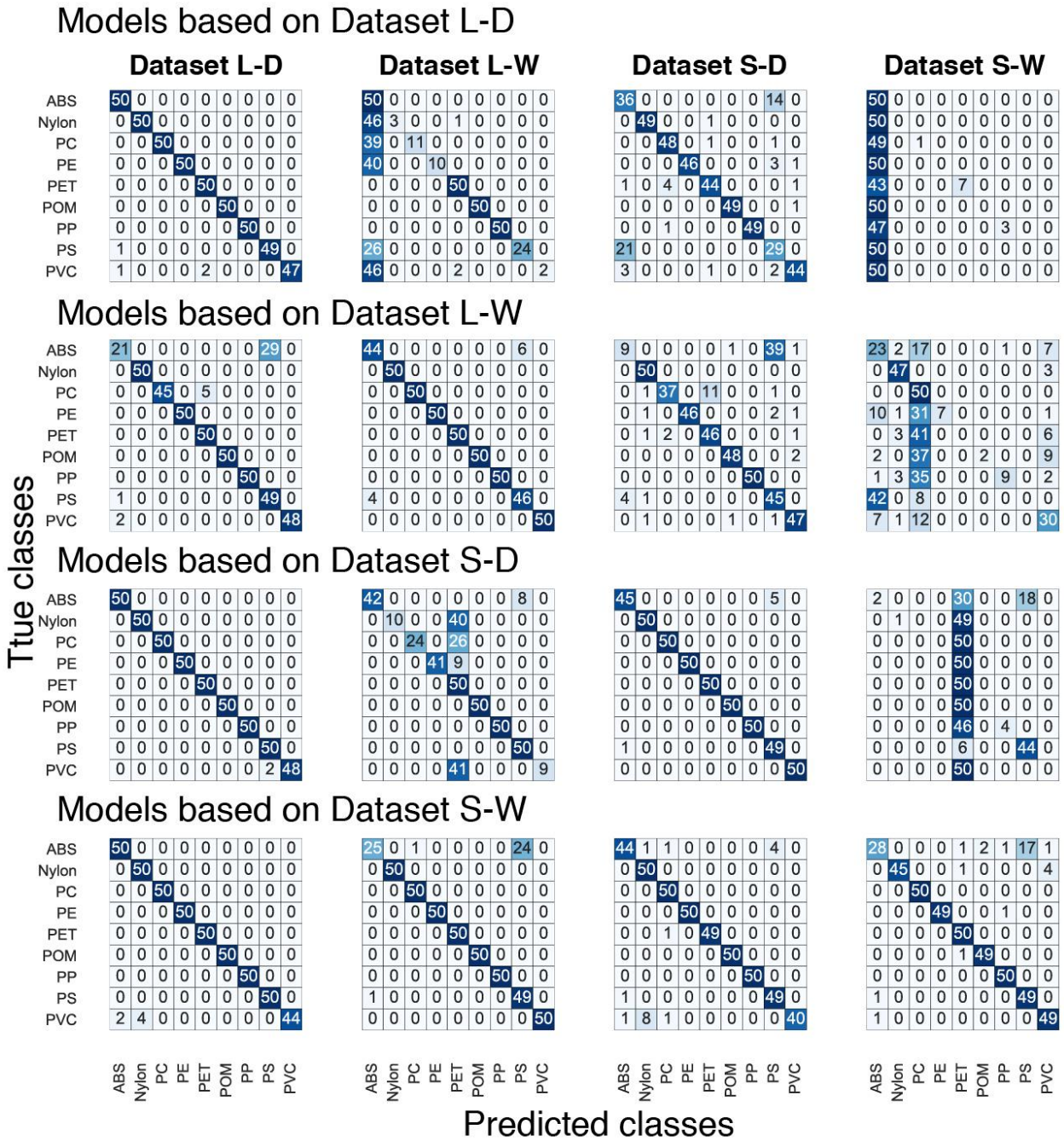

Figure S18. Confusion matrices of the classification models based on the spectral data in the range of 1000–1700 nm.

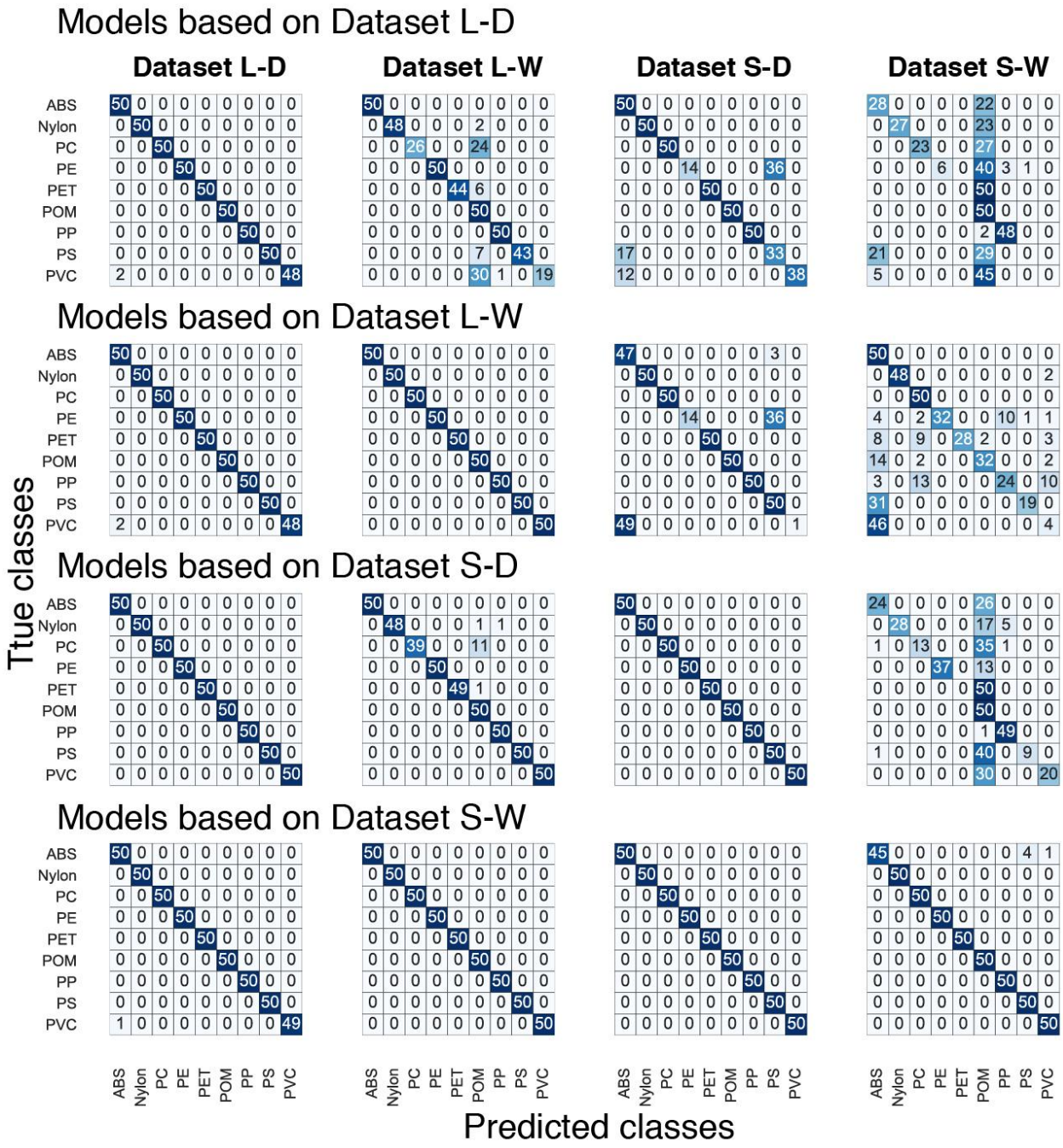

Figure S19. Confusion matrices of the classification models based on the spectral data in the range of 1000–2500 nm.

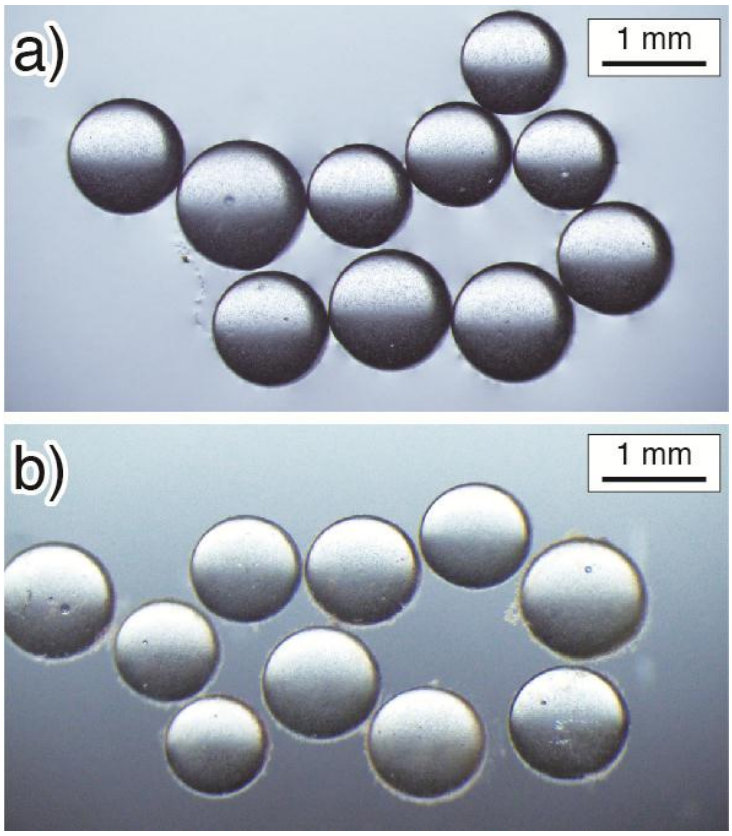

Figure S20. Picture of PS beads without microalgae (a) and covered with microalgae (b).