Supporting Information

Robust methodology for PEC performance analysis of photoanodes using machine learning and analytical data

Moeko Tajima,[1†] Yuya Nagai,[1†] Cyan Chen,[1] Zhenhua Pan[1] and Kenji Katayama[1*]

1 Department of Applied Chemistry, Chuo University, Tokyo 112-8551, Japan;

† Equally contributed

*Corresponding authors:

K. Katayama, Phone: +81-3-3817-1913, E-mail: kkata@kc.chuo-u.ac.jp

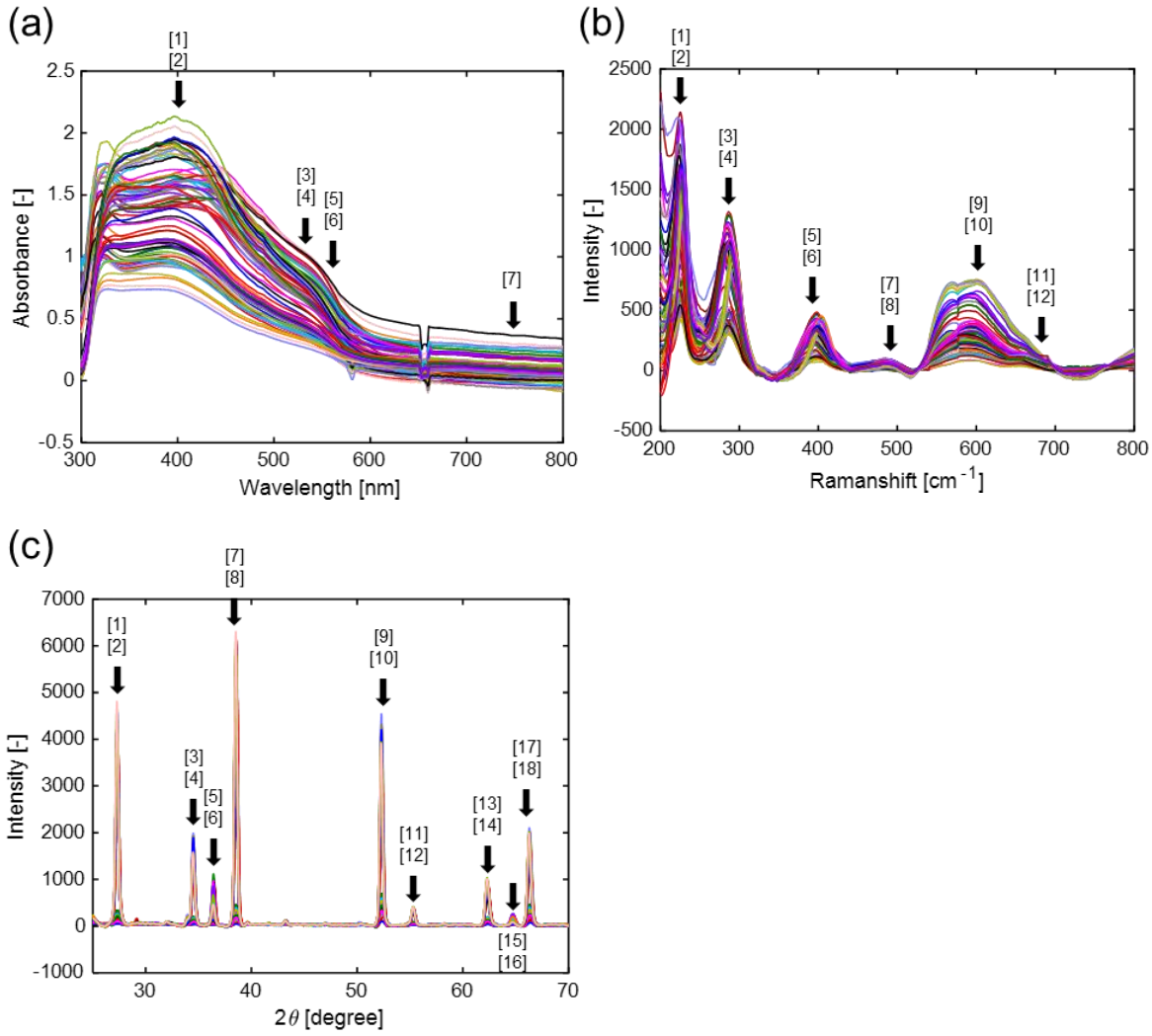

Figure S1 The analytical data for the hematite electrodes; (a) UV/vis spectra, (b) Raman spectra, and (c) XRD patterns. Each separated region is shown in black arrows with labels. The peak intensities and positions, the slope and positions were used as descriptors. The name and explanation of these descriptors from the analytical data were shown in the following table.

Table S1 The abbreviations of the descriptor names for the analytical data in Figure S1 and their explanation.

Analytical method	Descriptor name	Descriptor explanation
UV/Vis	[1] UV Vis pks abs [2] UV Vis pks locs	Peak intensity and position around 400 nm
	[3] UV_Vis_shol_posi_abs [4] UV_Vis_shol_posi_locs	Absolute intensity value at the 1st derivative maximum and its location in 500 − 573 nm
	[5] UV_Vis_shol_neg_abs [6] UV_Vis_shol_neg_locs	Absolute intensity value at the 1st derivative minimum and its location in 500 − 573 nm
	[7] UV_Vis_average_abs	Average intensity in 680~800 nm
Raman	[1] Raman_pk1_posi_int [2] Raman_pk1_posi_loc	Intensity and position of peak 1
	[3] Raman_pk2_posi_int [4] Raman_pk2_posi_loc	Intensity and position of peak 2
	[5] Raman_pk3_posi_int [6] Raman_pk3_posi_loc	Intensity and position of peak 3
	[7] Raman_pk4_posi_int [8] Raman_pk4_posi_loc	Intensity and position of peak 4
	[9] Raman_pk5_posi_int [10] Raman_pk5_posi_loc	Intensity and position of peak 5
	[11] Raman_pk_shol_int [12] Raman_pk_shol_loc	Intensity and position of the shoulder peak
XRD	[1] XRD_pk1_int [2] XRD_pk1_loc	Intensity and position of peak 1
	[3] XRD_pk2_int [4] XRD_pk2_loc	Intensity and position of peak 2
	[5] XRD_pk3_int [6] XRD_pk1_3oc	Intensity and position of peak 3
	[7] XRD_pk4_int [8] XRD_pk4_loc	Intensity and position of peak 4
	[9] XRD_pk5_int	Intensity and position of peak 5

	[10] XRD_pk5_loc	
	[11] XRD_pk6_int [12] XRD_pk6_loc	Intensity and position of peak 6
	[13] XRD_pk7_int [14] XRD_pk7_loc	Intensity and position of peak 7
	[15] XRD_pk8_int [16] XRD_pk8_loc	Intensity and position of peak 8
	[17] XRD_pk9_int [18] XRD_pk9_loc	Intensity and position of peak 9

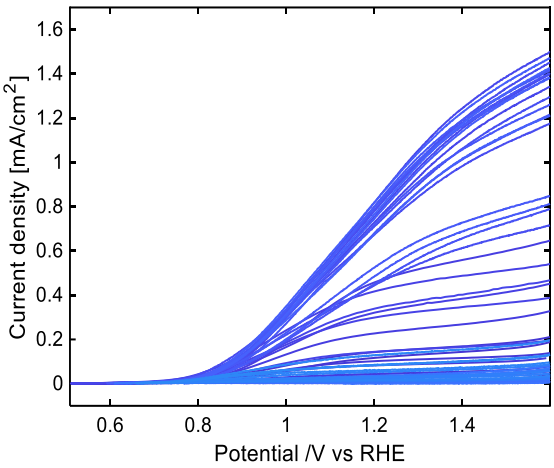

Figure S2 The current–voltage curves for 75 samples of hematite photoanodes. The potential is given versus RHE.

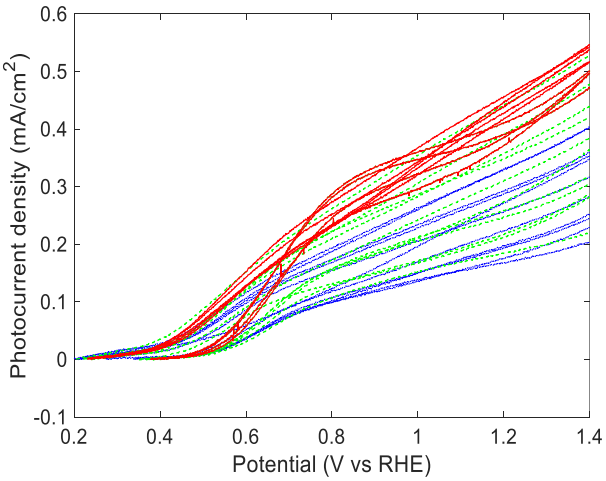

Figure S3 The current–voltage curves for thirty-two samples of $BiVO_4$ photoanodes are shown. The potential is indicated versus reversible hydrogen electrode (RHE). Blue, green, and red curves show the current–voltage curves for the samples with the repeated number of the spin coating of 9, 11, and 13 times.

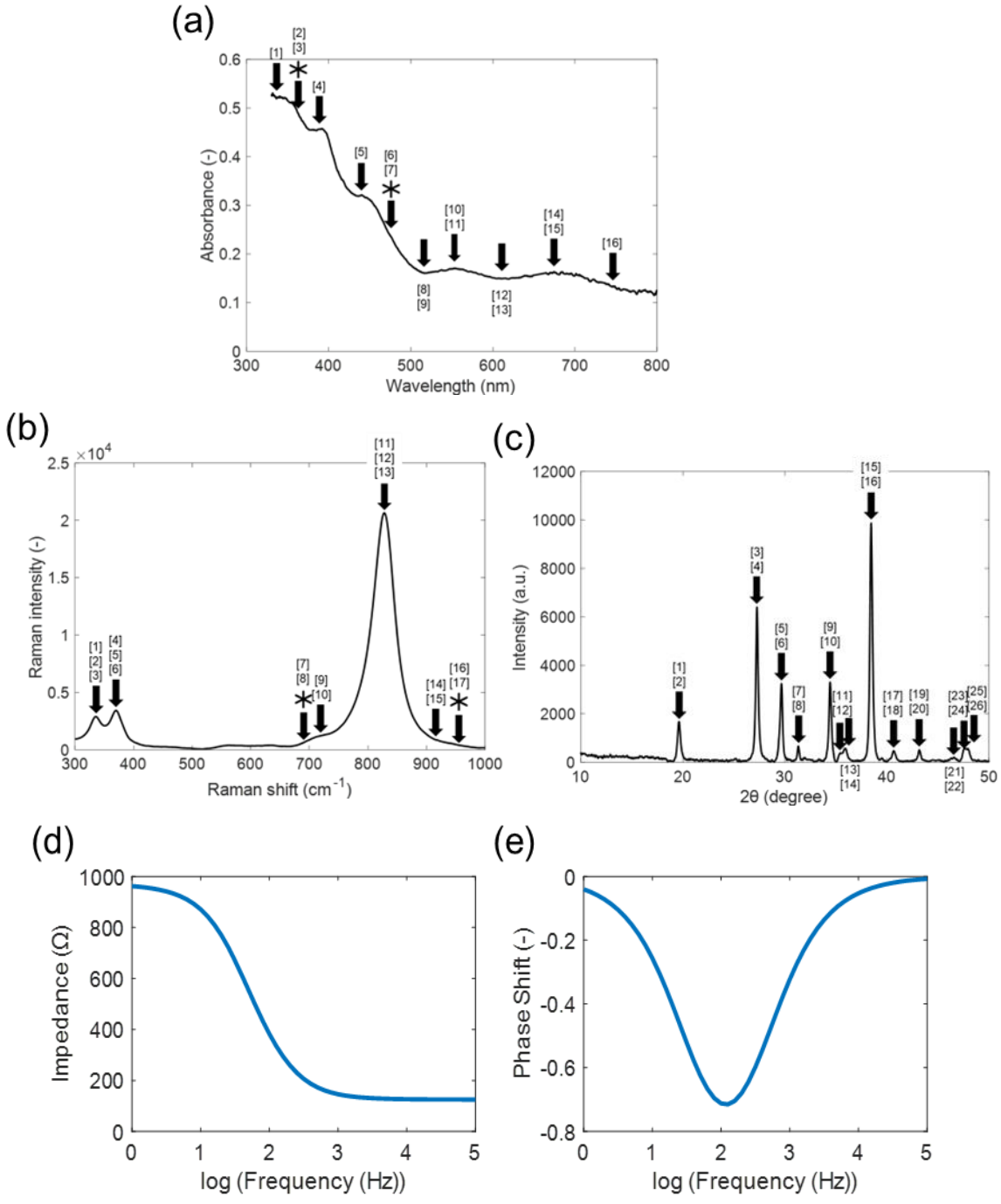

Figure S4 An example of the analytical data for $BiVO_4$ electrode is shown with an indication of features in the spectra and patterns; (a) UV/Vis spectrum, (b) XRD pattern, and (c) Raman spectrum, (d)(e) impedance and phase shift of PEIS. The intensities, positions and width were taken as feature values at the arrows and the slope was taken at (*). The name and explanation of these descriptors from the analytical data were shown in the following table.

Table S2 The abbreviations of the descriptor names for the analytical data in Figure S4 and their explanation. −

Analytical method	Descriptor name	Descriptor explanation
UV Vis	[1] UV_Vis_average_330_350	Average intensity in 330－350 nm
	[2] UV_pk_slope1_int [3] UV_pk_slope1_loc	Maximum value and its location of 1st derivative spectrum in 350－380 nm
	[4] UV_Vis_average_380_390	Average intensity in 380－390 nm
	[5] UV_Vis_average_430_440	Average intensity in 430－440 nm
	[6] UV_pk_slope2_int [7] UV_pk_slope2_loc	Maximum value and its location of 1st derivative spectrum in 440－500 nm
	[8] UV_pk_nega1_int [9] UV_pk_nega1_loc	1st negative peak intensity and position in 500－700 nm
	[10] UV pk1_int [11] UV pk1 loc	1st peak intensity and position in 500－700 nm
	[12] UV_pk_nega2_int [13] UV_pk_nega2_loc	2nd negative peak intensity and position in 500－700 nm
	[14] UV_pk2_int [15] UV_pk2_loc	2nd peak intensity and position in 500－700 nm
	[16] UV_Vis_average_700_800	Average intensity in 700－800 nm
Raman	[1] Raman_pk1_int [2] Raman_pk1_loc [3] Raman_pk1_wid	Intensity, position, and width of peak 1
	[4] Raman_pk2_int [5] Raman_pk2_loc [6] Raman_pk2_wid	Intensity, position, and width of peak 2
	[7] Raman_shoulder_slope_loc [8] Raman_sholder_slope	Maximum value of slope and its location in 650－800 cm^{-1}
	[9] Raman_shoulder_end_loc	Absolute intensity value at the

	[10] Raman_sholder_end_int	1st derivative maximum and its location in 650－800 cm^{-1}
	[11] Raman_pk3_int [12] Raman_pk3_loc [13] Raman_pk3_wid	Intensity and position of peak 3
	[14] Raman_shoulder_2_slope_loc [15] Raman_shoulder_2_slope	Absolute intensity value at the 1st derivative minimum and its location in 915－960 cm^{-1}
	[16] Raman_shoulder_2_loc [17] Raman_shoulder_2_int	Maximum negative value of slope and its location at 915－960 cm^{-1}
XRD	[1] XRD_pk1_int [2] XRD_pk1_loc	Intensity and position of peak 1
	[3] XRD_pk2_int [4] XRD_pk2_loc	Intensity and position of peak 2
	[5] XRD_pk3_int [6] XRD_pk3_loc	Intensity and position of peak 3
	[7] XRD_pk4_int [8] XRD_pk4_loc	Intensity and position of peak 4
	[9] XRD_pk5_int [10] XRD_pk5_loc	Intensity and position of peak 5
	[11] XRD_pk6_int [12] XRD_pk6_loc	Intensity and position of peak 6
	[13] XRD_pk7_int [14] XRD_pk7_loc	Intensity and position of peak 7
	[15] XRD_pk8_int [16] XRD_pk8_loc	Intensity and position of peak 8
	[17] XRD_pk9_int [18] XRD_pk9_loc	Intensity and position of peak 9
	[19] XRD_pk10_int [20] XRD_pk10_loc	Intensity and position of peak 10
	[21] XRD_pk11_int [22] XRD_pk11_loc	Intensity and position of peak 11
	[23] XRD_pk12_int [24] XRD_pk12_loc	Intensity and position of peak 12
	[25] XRD_pk13_int	Intensity and position of peak

	[26] XRD_pk13_loc	13
PEIS	EIS_R_1, EIS_R_2, EIS_C_2, EIS_n_2	Fitting parameters analyzed by an equivalent circuit model

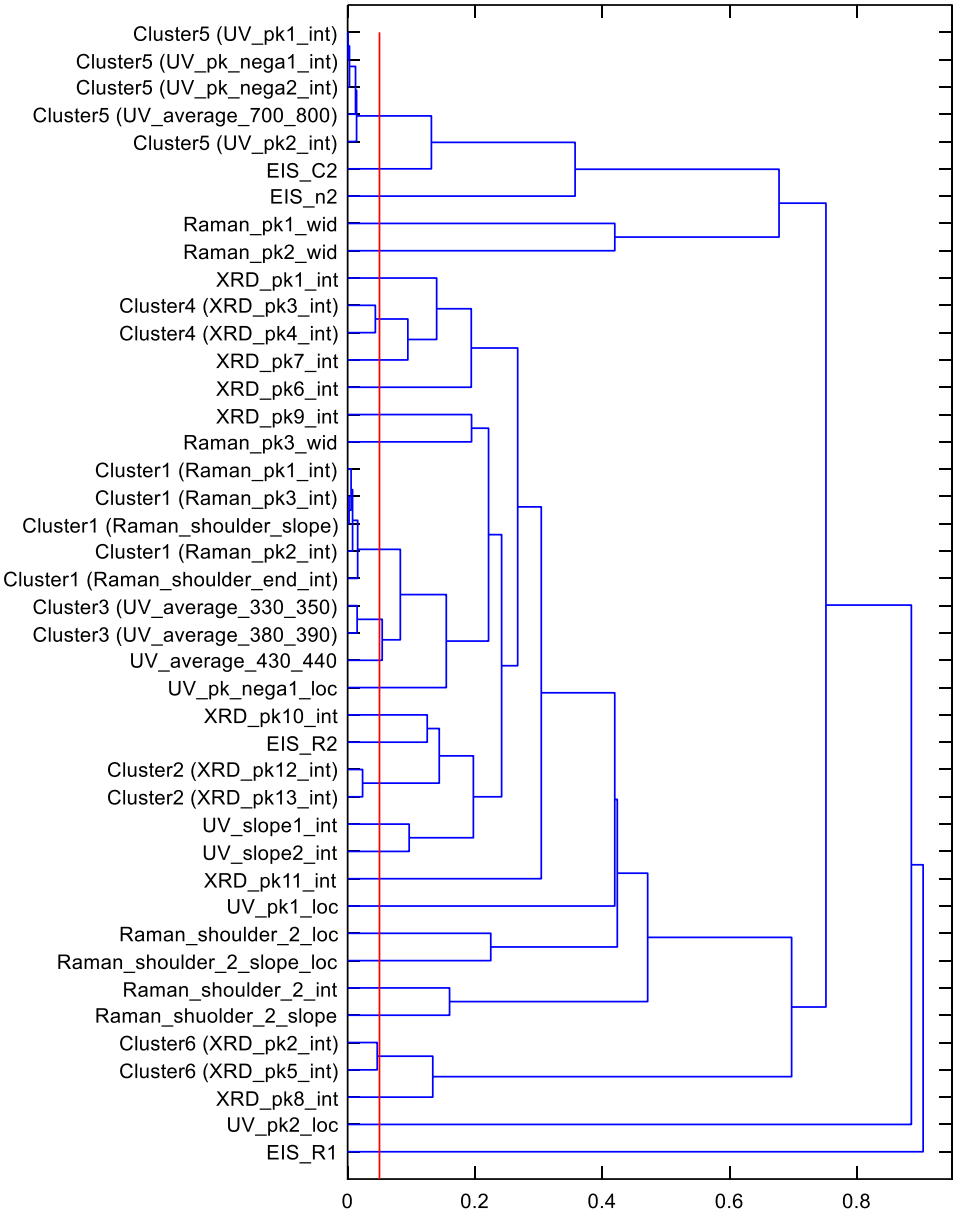

Figure S5 The dendrogram illustrates feature clustering for the analytical data features obtained from $BiVO_4$ photoanodes. The horizontal axis depicts the absolute cosine distance, while the vertical axis displays the name of features. The red line represents the clustering threshold.

Table S3 The list of the clustered features in the clustering analysis shown in Figure S5.

Features name	Cluster index
Raman_pk1_int	1
Raman_pk2_int	1
Raman_pk3_int	1
Raman_shoulder_slope	1
Raman_shoulder_end_int	1
XRD_pk12_int	2
XRD_pk13_int	2
UV_average_330_350	3
UV_average_380_390	3
XRD_pk3_int	4
XRD_pk4_int	4
UV_pk1_int	5
UV_pk2_int	5
UV_pk_nega1_int	5
UV_pk_nega2_int	5
UV_average_700_800	5
XRD_pk2_int	6
XRD_pk5_int	6

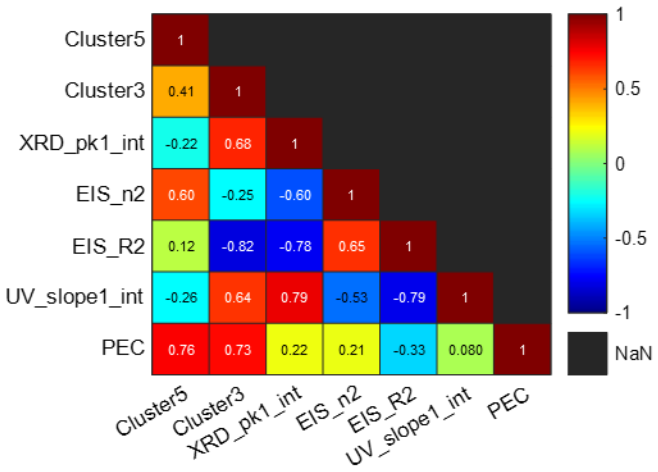

Figure S6 The correlation matrix of $BiVO_4$ photoanode data. The color represents the value of the correlation coefficients (cosine similarity).

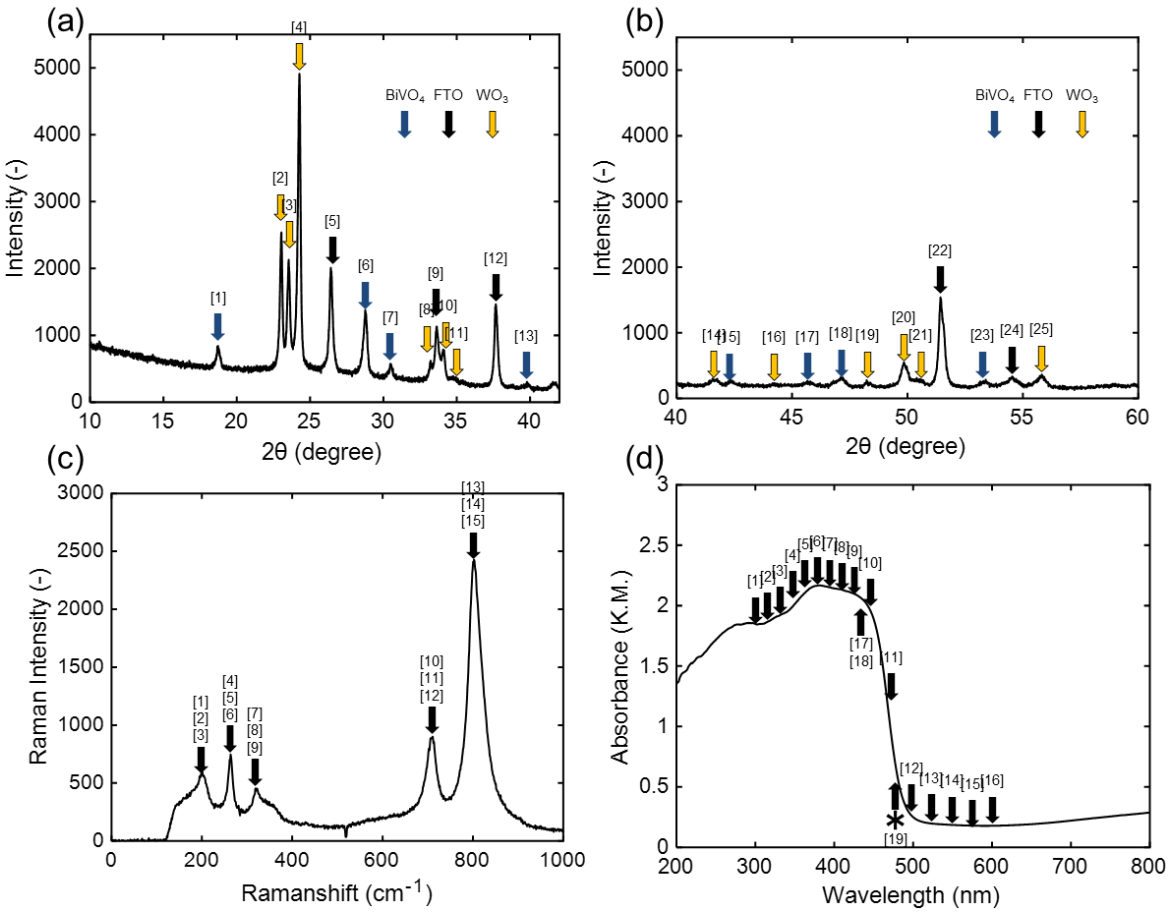

Figure S7 An example of the analytical data for WO_3/ $BiVO_4$ electrode is shown with an indication of features in the spectra and patterns; (a) XRD pattern with an angle of diffraction of between 10 and 42 degrees (b) XRD pattern with an angle of diffraction of between 42 and 60 degrees. (c) Raman spectrum, (d) DRS spectrum. The intensities, positions, and width were taken as feature values at the arrows and the slope was taken at (*). For the Raman intensity, the ratios of the individual peaks in all the combinations have been taken. The name and explanation of these descriptors from the analytical data were shown in the following table.

Table S4 The abbreviations of the descriptor names for the analytical data in Figure S7 and their explanation.

Analytical method	Descriptor name	Descriptor explanation
UV vis DRS	[1] DRS_300nm	Peak intensity at 300 nm
	[2] DRS_320nm	Peak intensity at 320 nm
	[3] DRS_340nm	Peak intensity at 340 nm
	[4] DRS_360nm	Peak intensity at 360 nm
	[5] DRS_380nm	Peak intensity at 380 nm
	[6] DRS_400nm	Peak intensity at 400 nm
	[7] DRS_420nm	Peak intensity at 420 nm
	[8] DRS_440nm	Peak intensity at 440 nm
	[9] DRS_460nm	Peak intensity at 460 nm
	[10] DRS_480nm	Peak intensity at 480 nm
	[11] DRS_500nm	Peak intensity at 500 nm
	[12] DRS_520nm	Peak intensity at 520 nm
	[13] DRS_540nm	Peak intensity at 540 nm
	[14] DRS_560nm	Peak intensity at 560 nm
	[15] DRS_580nm	Peak intensity at 580 nm
	[16] DRS_600nm	Peak intensity at 600 nm
	[17] DRS_shoulder_pk [18] DRS_shoulder_loc	Intensity and position of shoulder peak in 400 −500 nm
	[19] DRS_slope_ave	Averaged slope in 450 −500 nm
Raman	[1] Raman_pk1_int [2] Raman_pk1_loc [3] Raman_pk1_w	Intensity, position and width of peak 1
	[4] Raman_pk2_posi_int [5] Raman_pk2_posi_loc [6] Raman_pk2_w	Intensity, position and width of peak 2
	[7] Raman_pk3_posi_int [8] Raman_pk3_posi_loc [9] Raman_pk3_w	Intensity, position and width of peak 3

	[10] Raman_pk4_posi_int [11] Raman_pk4_posi_loc [12] Raman_pk4_w	Intensity, position and width of peak 4
	[13] Raman_pk5_posi_int [14] Raman_pk5_posi_loc [15] Raman_pk5_w	Intensity, position and width of peak 1
	[16] Raman_ratio1	Ratio: (peak1/peak2)
	[17] Raman_ratio2	Ratio: (peak1/peak3)
	[18] Raman_ratio3	Ratio: (peak1/peak4)
	[19] Raman_ratio4	Ratio: (peak1/peak5)
	[20] Raman_ratio5	Ratio: (peak2/peak3)
	[21] Raman_ratio6	Ratio: (peak2/peak4)
	[22] Raman_ratio7	Ratio: (peak2/peak5)
	[23] Raman_ratio8	Ratio: (peak3/peak4)
	[24] Raman_ratio9	Ratio: (peak3/peak5)
	[25] Raman_ratio10	Ratio: (peak4/peak5)
XRD	[1] XRD_pk1_int	Intensity at peak 1
	[2] XRD_pk2_int	Intensity at peak 2
	[3] XRD_pk3_int	Intensity at peak 3
	[4] XRD_pk4_int	Intensity at peak 4
	[5] XRD_pk5_int	Intensity at peak 5
	[6] XRD_pk6_int	Intensity at peak 6
	[7] XRD_pk7_int	Intensity at peak 7
	[8] XRD_pk8_int	Intensity at peak 8
	[9] XRD_pk9_int	Intensity at peak 9
	[10] XRD_pk10_int	Intensity at peak 10
	[11] XRD_pk11_int	Intensity at peak 11
	[12] XRD_pk12_int	Intensity at peak 12
	[13] XRD_pk13_int	Intensity at peak 13
	[14] XRD_pk14_int	Intensity at peak 14
	[15] XRD_pk15_int	Intensity at peak 15
	[16] XRD_pk16_int	Intensity at peak 16
	[17] XRD_pk17_int	Intensity at peak 17
	[18] XRD_pk18_int	Intensity at peak 18
	[19] XRD_pk19_int	Intensity at peak 19

	[20] XRD_pk20_int	Intensity at peak 20
	[21] XRD_pk21_int	Intensity at peak 21
	[22] XRD_pk22_int	Intensity at peak 22
	[23] XRD_pk23_int	Intensity at peak 23
	[24] XRD_pk24_int	Intensity at peak 24
	[25] XRD_pk25_int	Intensity at peak 25
PEIS	PEIS_R_1, PEIS_R_2, PEIS_C_2, PEIS_n_2	Fitting parameters analyzed by the equivalent circuit model

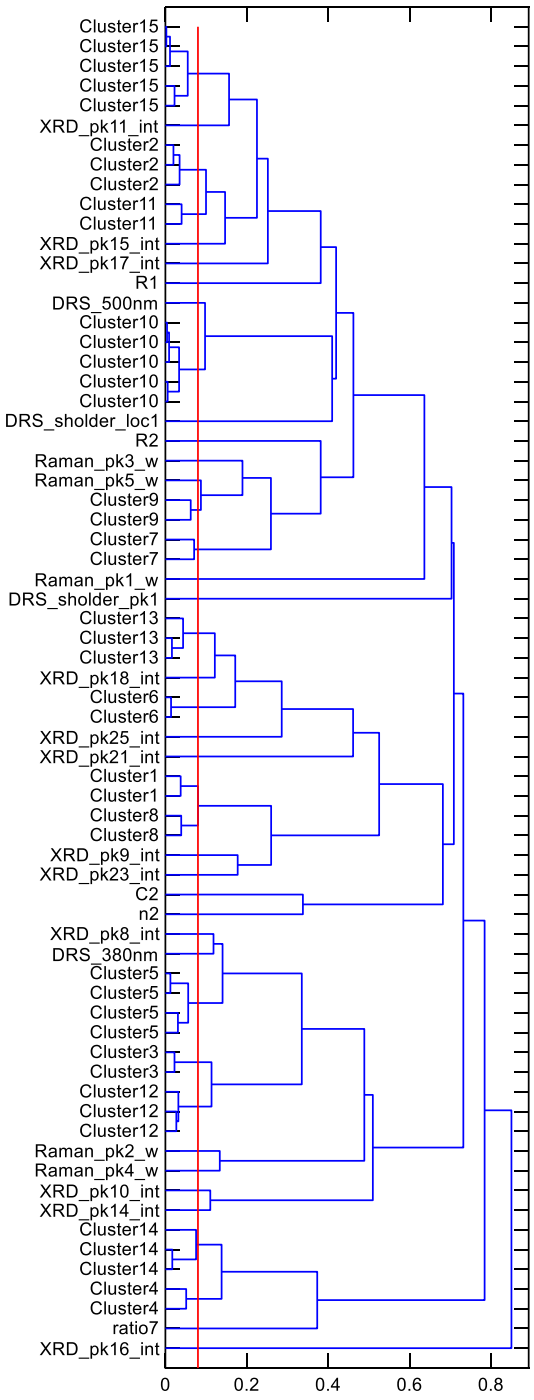

Figure S8 The dendrogram illustrates feature clustering for the analytical data features obtained from a WO_3/ $BiVO_4$ heterojunction photoanode. The horizontal axis depicts the absolute cosine distance, while the vertical axis displays the name of features. The red line represents the clustering threshold.

Table S5 The list of the clustered features for the analytical data features obtained from a WO_3/ $BiVO_4$ heterojunction photoanode.

Features name	Cluster index
XRD_pk5_int ,XRD_pk12_int	1
XRD_pk1_int, XRD_p6_int, DRS_480nm	2
Raman_pk1_int, Raman_pk3_int	3
Raman_ratio6, Raman_ratio9	4
DRS_300nm, DRS_320nm, DRS_340nm, DRS_360nm	5
XRD_pk4_int, XRD_pk20_int	6
Raman_ratio1, Raman_ratio5	7
XRD_pk22_int, XRD_pk24_int	8
Raman_ratio2, Raman_ratio10	9
DRS_520nm, DRS_540nm, DRS_560nm, DRS_580nm, DRS_600nm,	10
XRD_pk7_int, XRD_pk13_int	11
Raman_pk2_int, Raman_pk4_int, Raman_pk5_int	12
XRD_pk2_int, XRD_pk3_int, XRD_pk19_int	13
Raman_ratio3, Raman_ratio4, Raman_ratio8	14
DRS_400nm, DRS_420nm, DRS_440nm, DRS_460nm,DRS_slope_ave1	15

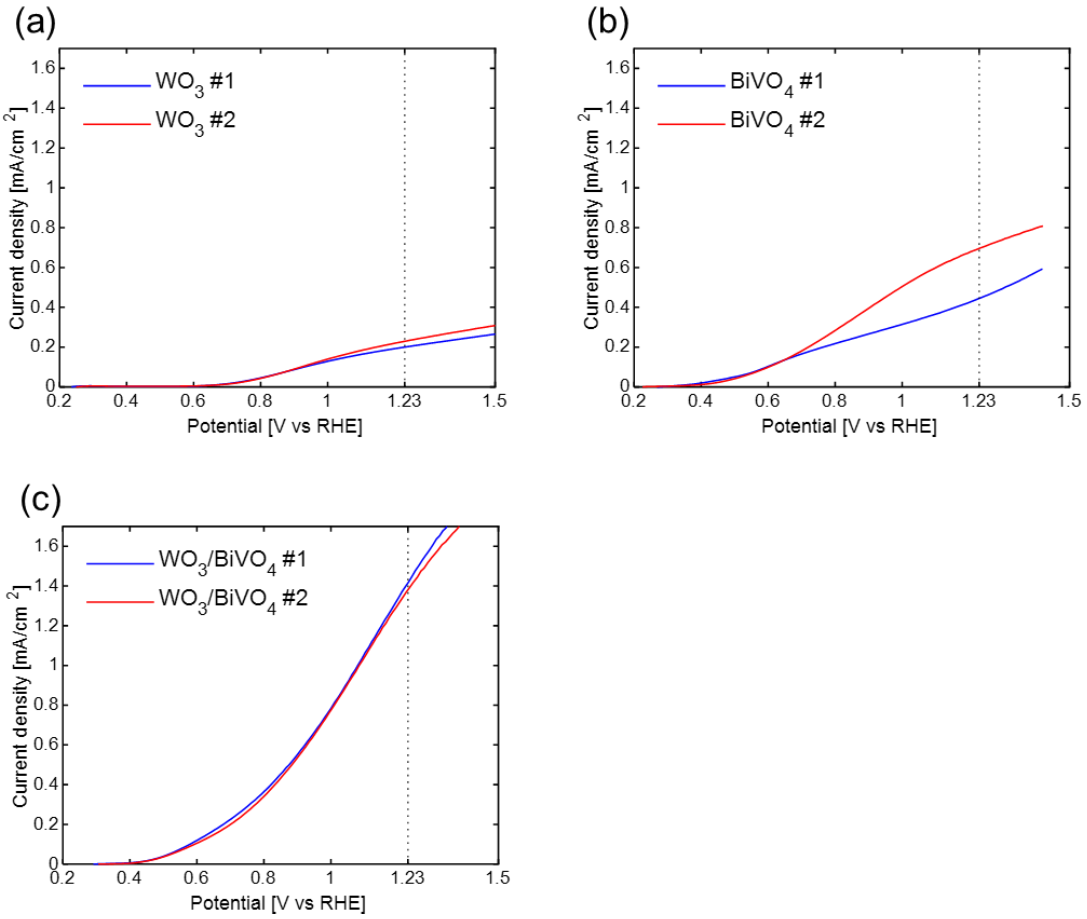

Figure S9 The current–voltage curves for WO_3-only (a), $BiVO_4$-only (b) and heterojunction samples (c) are shown. Two examples for each sample type are shown. The potential is indicated versus reversible hydrogen electrode (RHE).