Supporting information

Advanced Algorithm for Step Detection in Single-Entity Electrochemistry: A Comparative Study of Wavelet Transforms and Convolutional Neural Networks

Ziwen Zhao,[a*] Arunava Naha [b] , Nikolaos Kostopoulos[a] and Alina Sekretareva[a*]

[a.] *Department of Chemistry - Ångström, Uppsala University, 75120 Uppsala, Sweden*
[b.] *Department of Electrical Engineering, Uppsala University, 75120, Uppsala, Sweden*
*Corresponding author: ziwen.zhao@kemi.uu.se, alina.sekretareva@kemi.uu.se

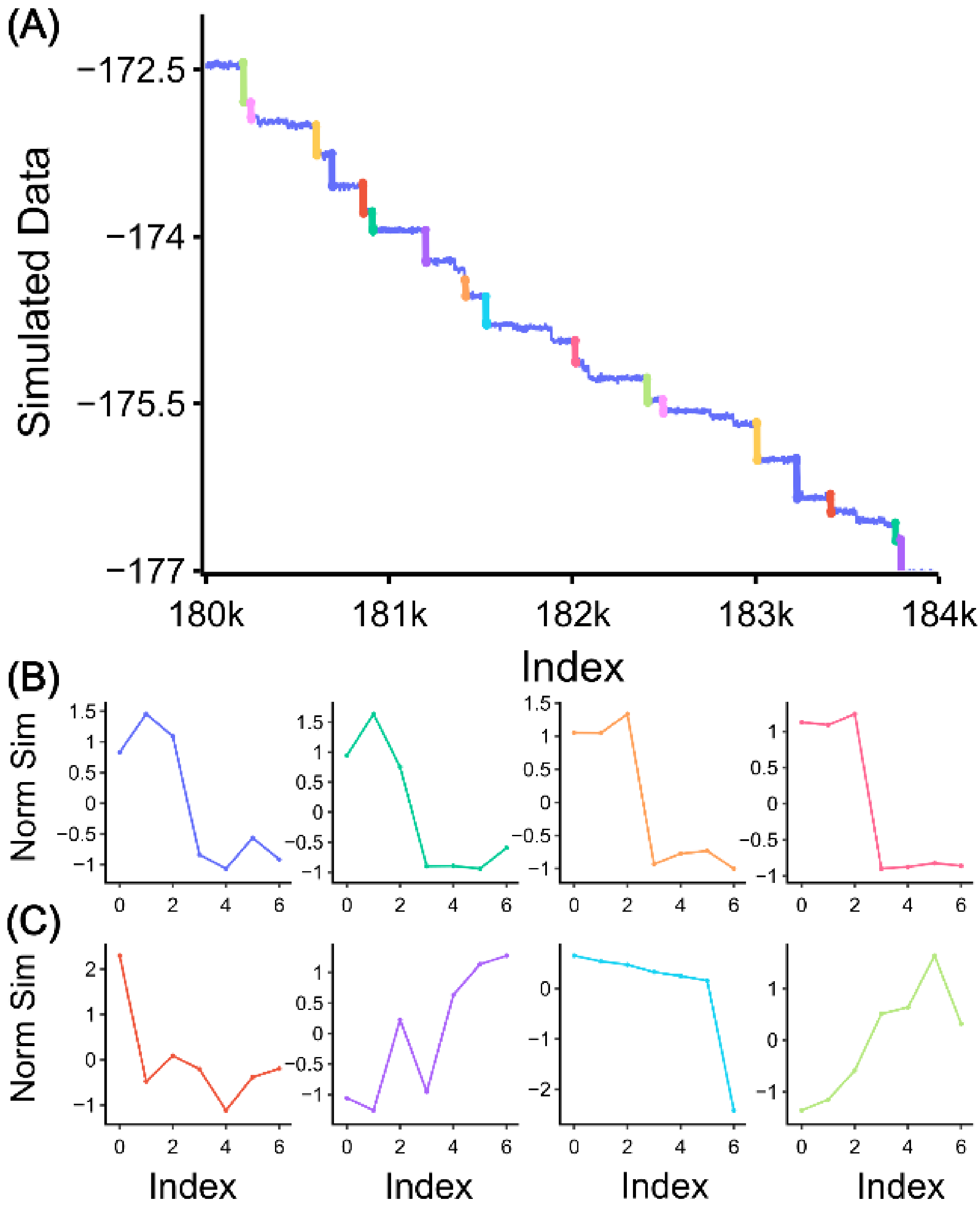

Figure S1. The training set for the Convolutional Neural Network. (A) Mathematical simulation of the signal, the coloured intervals are simulated steps. (B) Positive training set, each step signal is normalized. (C) Negative training set, each interval signal is normalized.

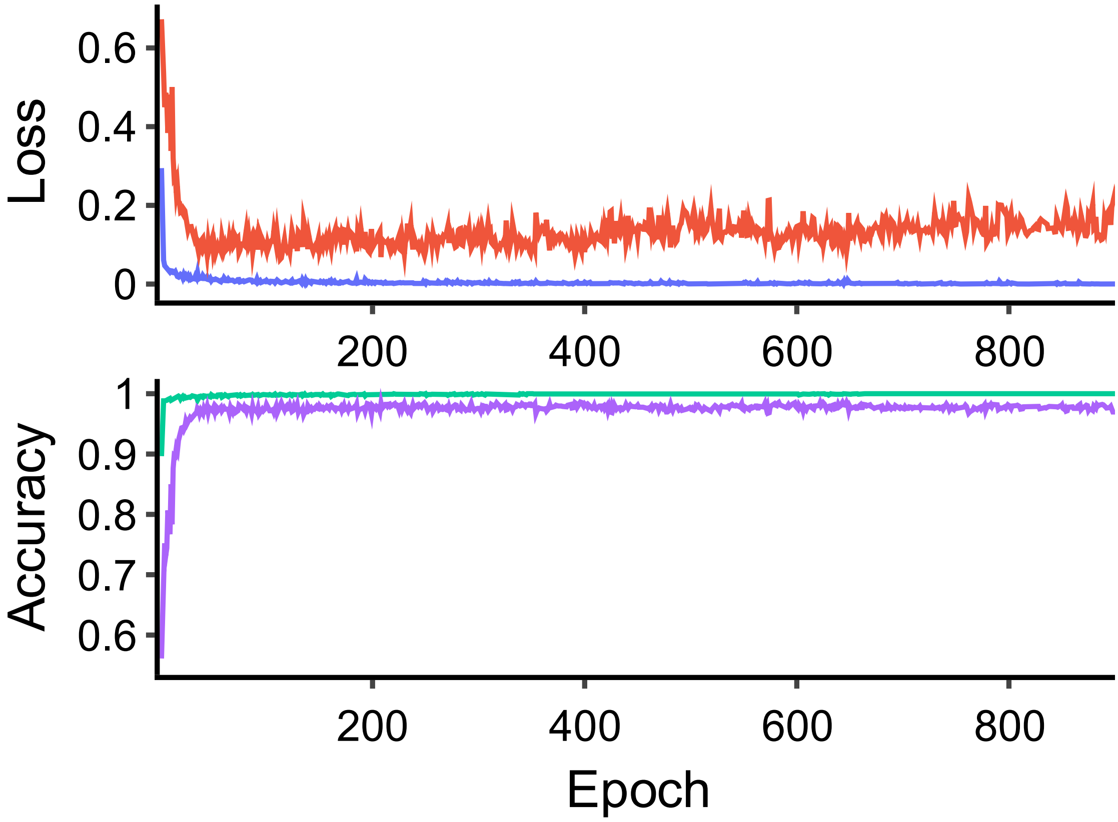

Figure S2. Training progress of the CNN model for the low-frequency data. The blue line indicates the loss on the training set, while the red line represents the loss on the cross-validation set. Similarly, the green line shows the accuracy on the training set, and the purple line shows the accuracy on the cross-validation set.

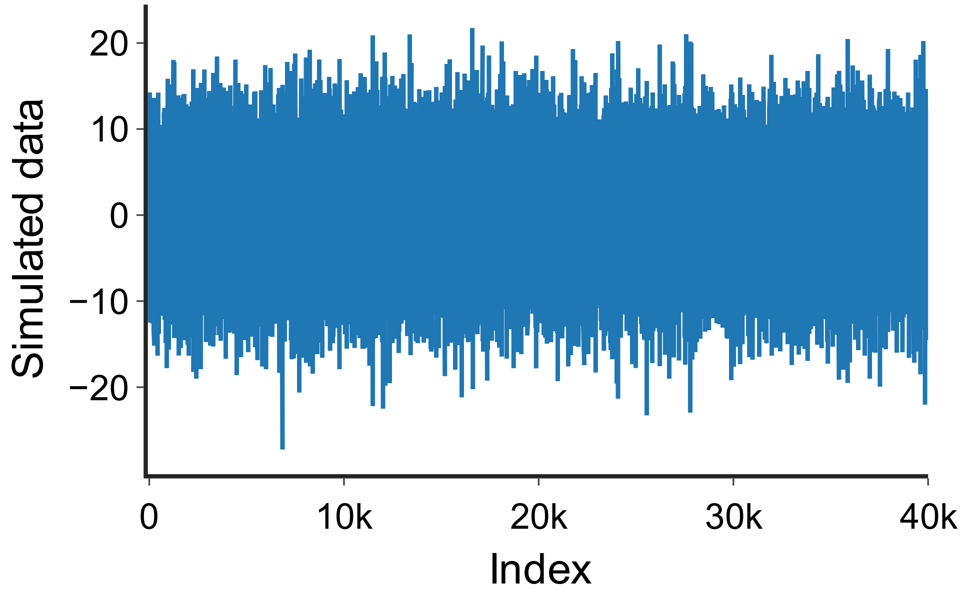

Figure S3. Unit derivate (sequential discrete difference of the data) of the data in Fig. 6(A) in the main text. No distinguishable spikes can be observed to use for step detection.

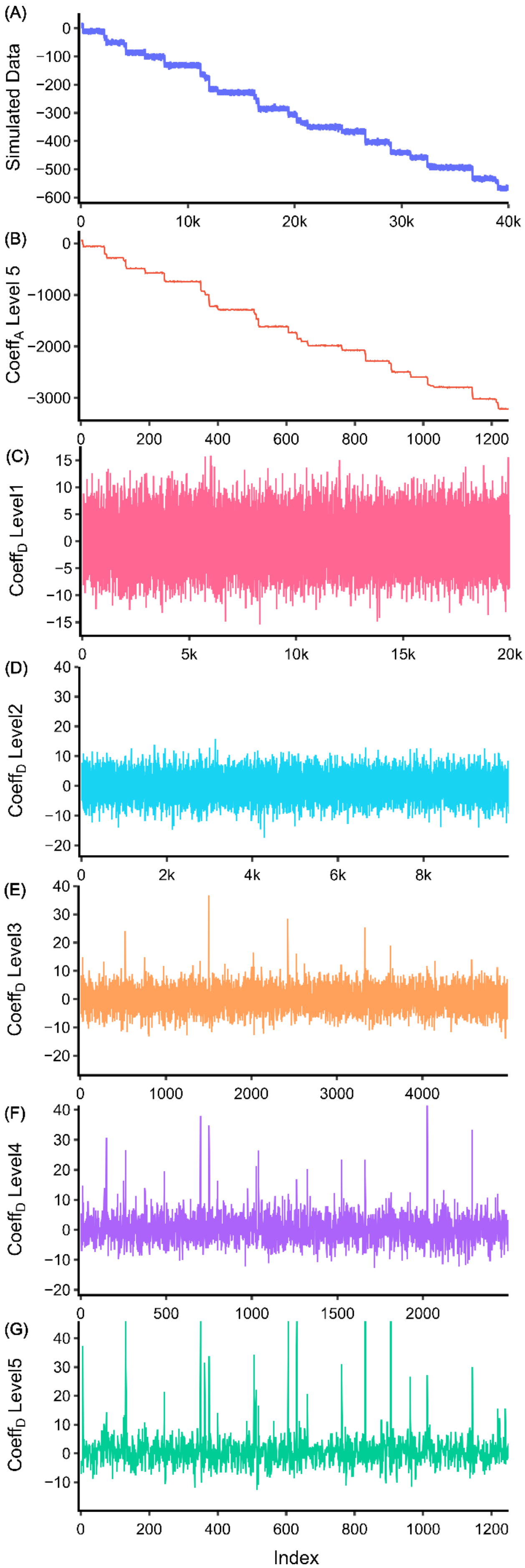

Figure S4. Five iterations of the DWT applied to the validation simulated data. (A) Simulated data. (B) Approximation coefficient after five iterations of the transform. (C-G) Detailed coefficients after the 1st to the 5th DWT transform.

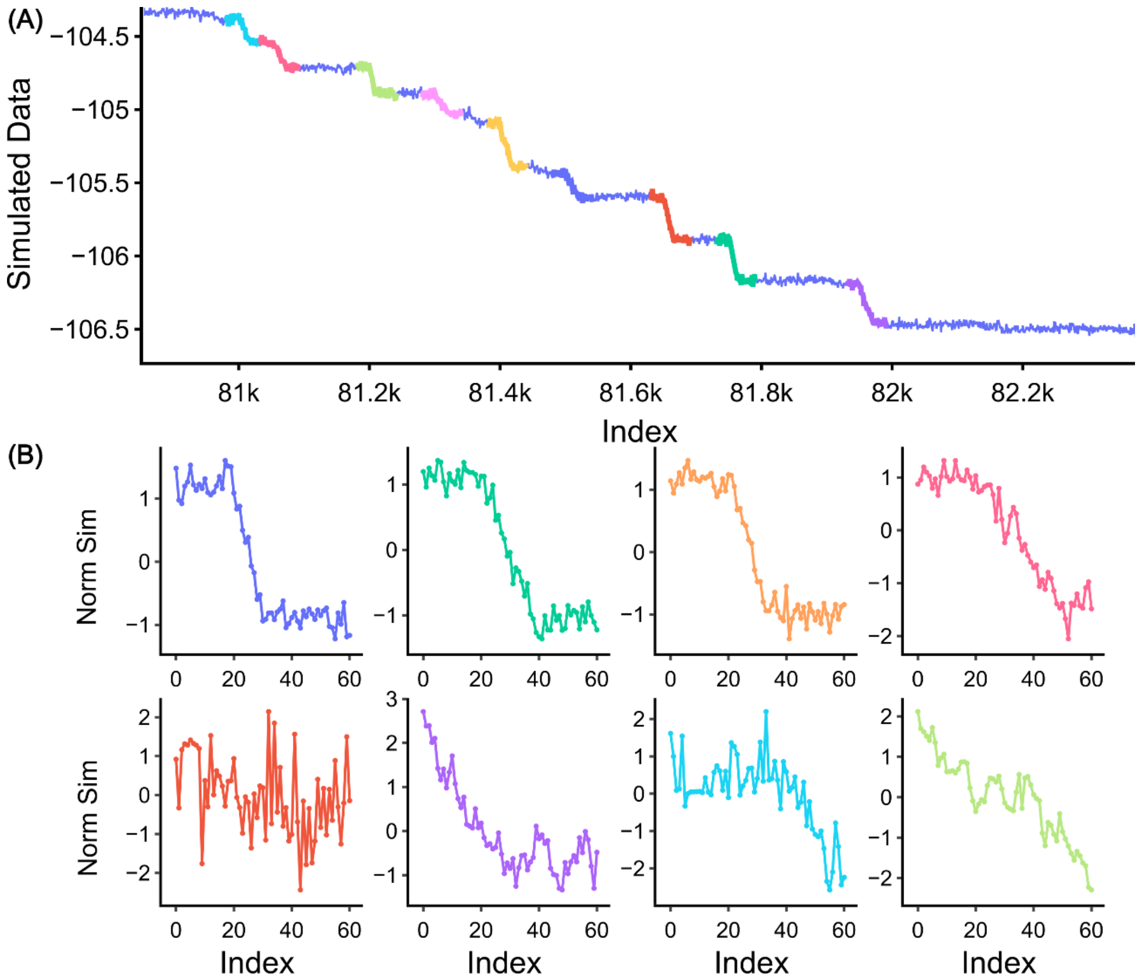

Figure S5. (A) Simulated data used as the training set for the CNN model. Each simulated step is colour-coded. (B) The training set, normalized to zero mean and unit variance. The first row serves as the positive training set, while the second row represents different types of negative signals: The first type is Gaussian noise. The second and third types are examples where the signal value drops at the beginning or end of the interval. The fourth type is slow decay, which is needed to prevent the CNN from identifying any value drop as a step.

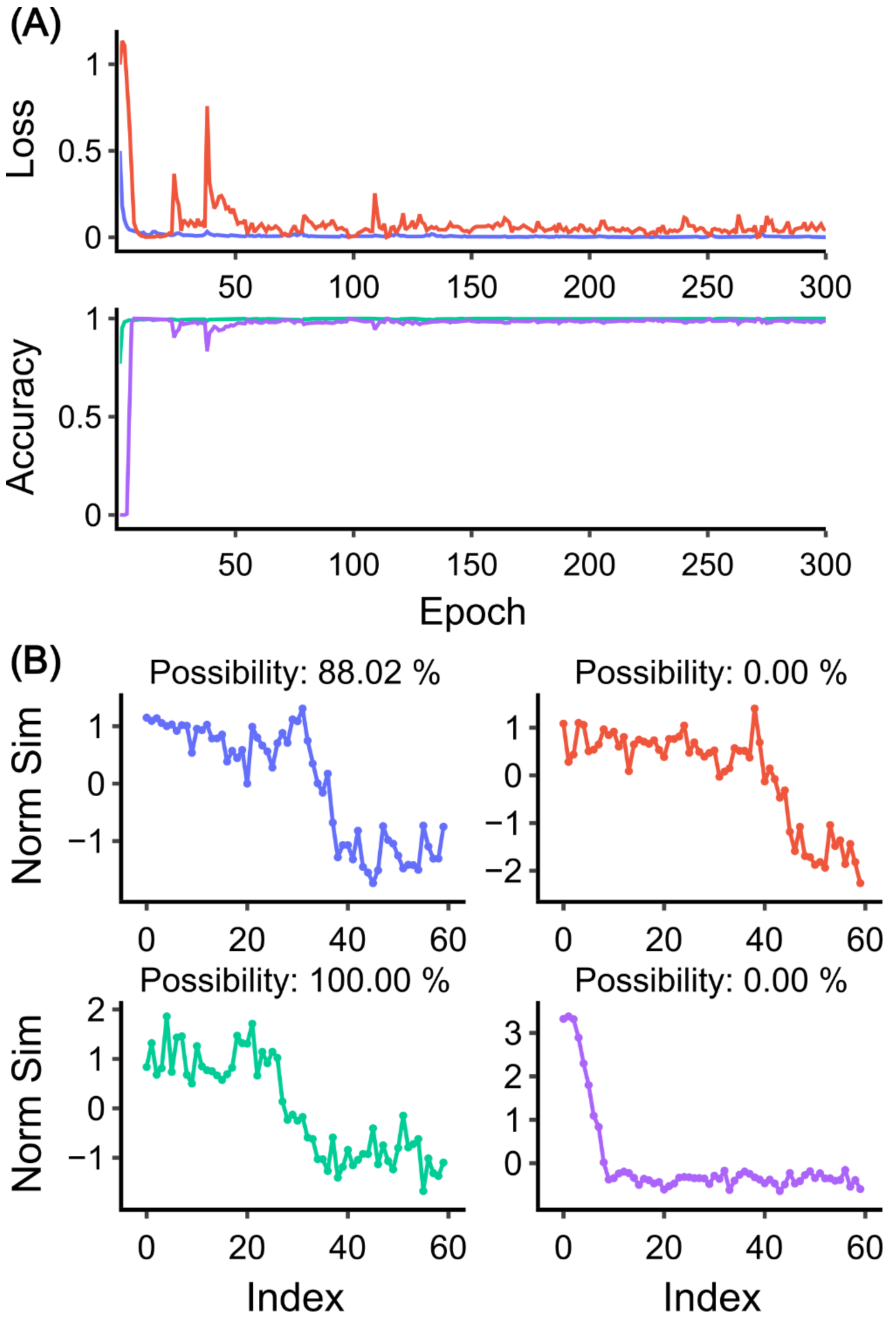

Figure S6. (A)Training progress of the CNN for the high-frequency data. The blue line indicates the loss on the training set, while the red line represents the loss on the cross-validation set. Similarly, the green line shows the accuracy on the training set, and the purple line shows the accuracy on the cross-validation set. (B) Preliminary validation of the CNN model. The probability that the CNN identifies a step is labeled at the top of each plot. Only changes occurring in the middle of the window are considered as steps.

Table S1. Parameters extracted from the step intervals of the first 32 seconds of the digitized data

A. The DWT method.

UpperSlope	LowerSlope	StepHeight [pA]
-0.014550	-0.018069	39.085
-0.016059	-0.011372	35.230
-0.004017	-0.004416	11.560
-0.019573	-0.020075	46.240
-0.012544	-0.013384	28.630
-0.024091	-0.022754	52.850

B. The CNN method.

UpperSlope	LowerSlope	StepHeight [pA]
-0.018063	-0.013552	39.640
-0.016059	-0.011372	35.230
-0.019573	-0.030112	44.590
-0.014721	-0.009536	26.980
-0.024091	-0.022754	52.850

Table S2. Parameters extracted from the first 5 step intervals of the validation data set.

A. The DWT method.

UpperSlope	LowerSlope	StepHeight [unit height]
-0.279535	0.097431	21.121519
-0.414436	-0.451027	25.322703
-0.225344	-0.177476	17.235023
-0.856368	-0.350936	10.308423
-0.569017	-0.597818	29.220601

B. The CNN method.

UpperSlope	LowerSlope	StepHeight [unit height]
0.164082	-0.496639	20.228
0.119585	-0.535077	22.990
0.392269	-0.194749	15.322
0.297074	0.118258	8.356
-0.260250	-0.318488	27.333