SUPPLEMENTARY MATERIALS

A. Additional description of dataset

TABLE S1

NUMBERS OF SCANS AND ICS OF EACH DATASET.

Dataset	# of Scans	# of ICs (Signal ICs / Noise ICs)
BCP [46]	99	13,800 (3,284 / 10,516)
HCP [3]	100	22,877 (2,326 / 20,551)
WHII-MB6 [47]	39	5,143 (795 / 4,348)
WHII-STD [47]	45	2,585 (422 / 2,783)

B. Effectiveness of learnable wavelet transform module

Since 1D IC time series have limited information compared to the 3D IC spatial maps resulting in lower model performance, we believe that extracting key features from 1D IC time series plays an important role in improving the overall model performance like other previous works [27]. Thus, in order to demonstrate the effectiveness of CNN_{lwt}, we have compared the results of our CNN_{lwt} with those of our previous models [27] on time series IC (*ts*) and power spectrum (*ps*) and their combination, i.e., CNN_{ts}^{AM}, CNN_{ps}, CNN_{ts+ps}, respectively, as well as those of other previous model combining CNN and LSTM in parallel, i.e., $CNN + LSTM$ [28], illustrated in the literature [27]. In addition, we have also implemented an ablation model based on basic wavelet transform (*wt*) of 1D IC time series without learnable parameters, i.e., CNN_{wt}.

Tables S2-S5 show the comparative results on each dataset, i.e., BCP, HCP, WHII-MB6, and WHII-STD, respectively. First, it is worth emphasizing that, as shown in Table S2, CNN_{lwt} demonstrated the best performance in all the evaluation metrics on the BCP dataset. Specifically, it achieved significant improvement (*p<0.05 or **p<0.0001) compared to CNN_{ts}^{AM}, CNN_{ps}, and $CNN + LSTM$ in most of the metrics, and showed slightly improved performance (p>0.05) compared to CNN_{ps+ts}. In addition, our CNN_{lwt} also achieved excellent performance on the HCP dataset, especially, in SPEC with significant improvement (*p<0.05 or **p<0.0001) compared to the others as shown in Table S3. On the WHII-MB6 and WHII-STD datasets, it showed similar performance (p>0.05) compared to CNN_{ts}^{AM}, CNN_{ps+ts}, and $CNN + LSTM$, but achieved meaningful improvement (*p<0.05 or **p<0.0001) compared to CNN_{ps} and CNN_{wt} as shown in Tables S4 and S5. Consequently, according to the comparison results on the various datasets, we believe CNN_{lwt}, which is adopted in our framework, achieved meaningful performance improvement compared to the existing methods.

TABLE S2

PERFORMANCE COMPARISON OF THE PROPOSED AND COMPETING METHODS. ALL METHODS WERE EVALUATED BASED ON ACCURACY (ACC), SENSITIVITY (SEN), SPECIFICITY (SPEC), F-MEASURE (F1), AND G-MEASURE (G-MEAN) ON THE BCP DATASET. THE BEST PERFORMANCE ON EACH EVALUATION METRICS ACROSS THE DIFFERENT METHODS IS EXPRESSED IN BOLD. WILCOXON SIGNED-RANK TEST IS CONDUCTED TO COMPARE THE PREDICTIVE PERFORMANCE OF OUR LEARNABLE WAVELET TRANSFORM AND THOSE OF OTHER COMPETING METHODS, RESPECTIVELY (*p < 0.05, **p < 0.0001).

Model	ACC (%)	SEN (%)	SPEC (%)	F1 (%)	G-Mean (%)
CNN_{ts}^{AM} [27]	95.73 ±2.11*	89.61 ±10.55	97.15 ±2.32*	89.56 ±8.19*	93.09 ±6.03*
CNN_{ps} [27]	94.44 ±2.92**	87.11 ±11.01**	96.26 ±3.15**	86.53 ±10.19**	91.35 ±6.50**
CNN_{ps+ts} [27]	95.95 ±2.26	90.09 ±11.15	97.24 ±2.39	89.76 ±8.83	93.36 ±6.35
$CNN + LSTM$ [28]	95.69 ±2.05*	89.77 ±10.40	97.04 ±2.30*	89.57 ±7.97*	93.12 ±5.92*
CNN_{wt}	95.50 ±2.39*	89.18 ±10.84	97.02 ±2.74*	89.15 ±8.40*	92.78 ±6.16*
CNN_{lwt} (ours)	**96.22±2.10**	**90.17 ±10.50**	**97.60 ±2.15**	**90.62 ±8.18**	**93.61 ±5.87**

TABLE S3

PERFORMANCE COMPARISON OF THE PROPOSED AND COMPETING METHODS. ALL METHODS WERE EVALUATED BASED ON ACCURACY (ACC), SENSITIVITY (SEN), SPECIFICITY (SPEC), F-MEASURE (F1), AND G-MEASURE (G-MEAN) ON THE HCP DATASET. THE BEST PERFORMANCE ON EACH EVALUATION METRICS ACROSS THE DIFFERENT METHODS IS EXPRESSED IN BOLD. WILCOXON SIGNED-RANK TEST IS CONDUCTED TO COMPARE THE PREDICTIVE PERFORMANCE OF OUR LEARNABLE WAVELET TRANSFORM AND THOSE OF OTHER COMPETING METHODS, RESPECTIVELY (*p < 0.05, **p < 0.0001).

Model	ACC (%)	SEN (%)	SPEC (%)	F1 (%)	G-Mean (%)
CNN_{ts}^{AM} [27]	97.53 ±1.09	90.90 ±6.70	98.28 ±1.22*	88.24 ±4.78	94.44 ±3.46
CNN_{ps} [27]	96.02 ±2.09**	82.86 ±12.78**	97.51 ±2.41**	80.69 ±10.09**	89.52 ±7.49**
CNN_{ps+ts} [27]	97.55 ±1.45	**91.65 ±7.52**	98.22 ±1.56*	**88.55 ±5.91**	**94.79 ±3.88**
$CNN + LSTM$ [28]	97.55 ±1.16	90.99 ±6.98	98.29 ±1.27*	88.31 ±4.85	94.49 ±3.61
CNN_{wt}	97.53 ±1.18	90.20 ±8.26	98.37 ±1.19*	88.05 ±5.50	94.08 ±4.40
CNN_{lwt} (ours)	**97.64 ±1.31**	90.03 ±8.93	**98.51 ±1.11**	88.49 ±6.29	94.06 ±4.98

TABLE S4

PERFORMANCE COMPARISON OF THE PROPOSED AND COMPETING METHODS. ALL METHODS WERE EVALUATED BASED ON ACCURACY (ACC), SENSITIVITY (SEN), SPECIFICITY (SPEC), F-MEASURE (F1), AND G-MEASURE (G-MEAN) ON THE WHII-MB6 DATASET. THE BEST PERFORMANCE ON EACH EVALUATION METRICS ACROSS THE DIFFERENT METHODS IS EXPRESSED IN BOLD. WILCOXON SIGNED-RANK TEST IS CONDUCTED TO COMPARE THE PREDICTIVE PERFORMANCE OF OUR LEARNABLE WAVELET TRANSFORM AND THOSE OF OTHER COMPETING METHODS, RESPECTIVELY ($*p < 0.05$, $**p < 0.0001$).

Model	ACC (%)	SEN (%)	SPEC (%)	F1 (%)	G-Mean (%)
CNN_{ts}^{AM} [27]	94.96 ±3.38	**87.17 ±7.77**	96.14 ±4.05*	**82.03 ±15.27**	**91.41 ±4.11**
CNN_{ps} [27]	91.86 ±3.44**	63.03 ±24.70**	95.77 ±4.59	65.73 ±21.48**	75.23 ±18.67**
CNN_{ps+ts} [27]	94.93 ±3.82	83.62 ±12.93	96.30 ±4.52	81.37 ±16.88	89.39 ±7.90
$CNN + LSTM$ [28]	94.96 ±3.21	86.02 ±9.32	96.22 ±3.80	81.75 ±14.99	90.80 ±5.11
CNN_{wt}	94.84 ±3.75	83.09 ±11.80	96.56 ±4.56	80.71 ±15.52	89.25 ±6.17*
CNN_{lwt} (ours)	**95.24 ±2.80**	85.04 ±13.55	**96.66 ±3.12**	81.56 ±14.94	90.28 ±7.82

TABLE S5

PERFORMANCE COMPARISON OF THE PROPOSED AND COMPETING METHODS. ALL METHODS WERE EVALUATED BASED ON ACCURACY (ACC), SENSITIVITY (SEN), SPECIFICITY (SPEC), F-MEASURE (F1), AND G-MEASURE (G-MEAN) ON THE WHII-STD DATASET. THE BEST PERFORMANCE ON EACH EVALUATION METRICS ACROSS THE DIFFERENT METHODS IS EXPRESSED IN BOLD. WILCOXON SIGNED-RANK TEST IS CONDUCTED TO COMPARE THE PREDICTIVE PERFORMANCE OF OUR LEARNABLE WAVELET TRANSFORM AND THOSE OF OTHER COMPETING METHODS, RESPECTIVELY ($*p < 0.05$, $**p < 0.0001$).

Model	ACC (%)	SEN (%)	SPEC (%)	F1 (%)	G-Mean (%)
CNN_{ts}^{AM} [27]	92.17 ±4.48	73.13 ±17.93	**95.03 ±4.97**	71.61 ± 12.77	**82.48 ±10.63**
CNN_{ps} [27]	89.11 ±6.74**	63.66 ±20.93**	92.97 ±7.56*	61.28 ±16.21**	75.30 ±13.89**
CNN_{ps+ts} [27]	91.60 ±5.59	72.99 ±23.73	94.19 ±6.66	70.04 ±17.52	80.73 ±17.18
$CNN + LSTM$ [28]	92.00 ±4.77	72.03 ±19.64	95.00 ±4.99	70.60 ±13.98	81.64 ±12.13
CNN_{wt}	91.09 ±4.98*	67.08 ±21.18*	94.86 ±5.26	66.38 ±14.95*	78.28 ±13.92*
CNN_{lwt} (ours)	**92.18 ±5.35**	**73.94 ±22.58**	94.80 ±6.16	**72.05 ±15.13**	82.33 ±13.22

C. Effectiveness of D-MMTM on BCP, WHII-MB6, and WHII-STD datasets

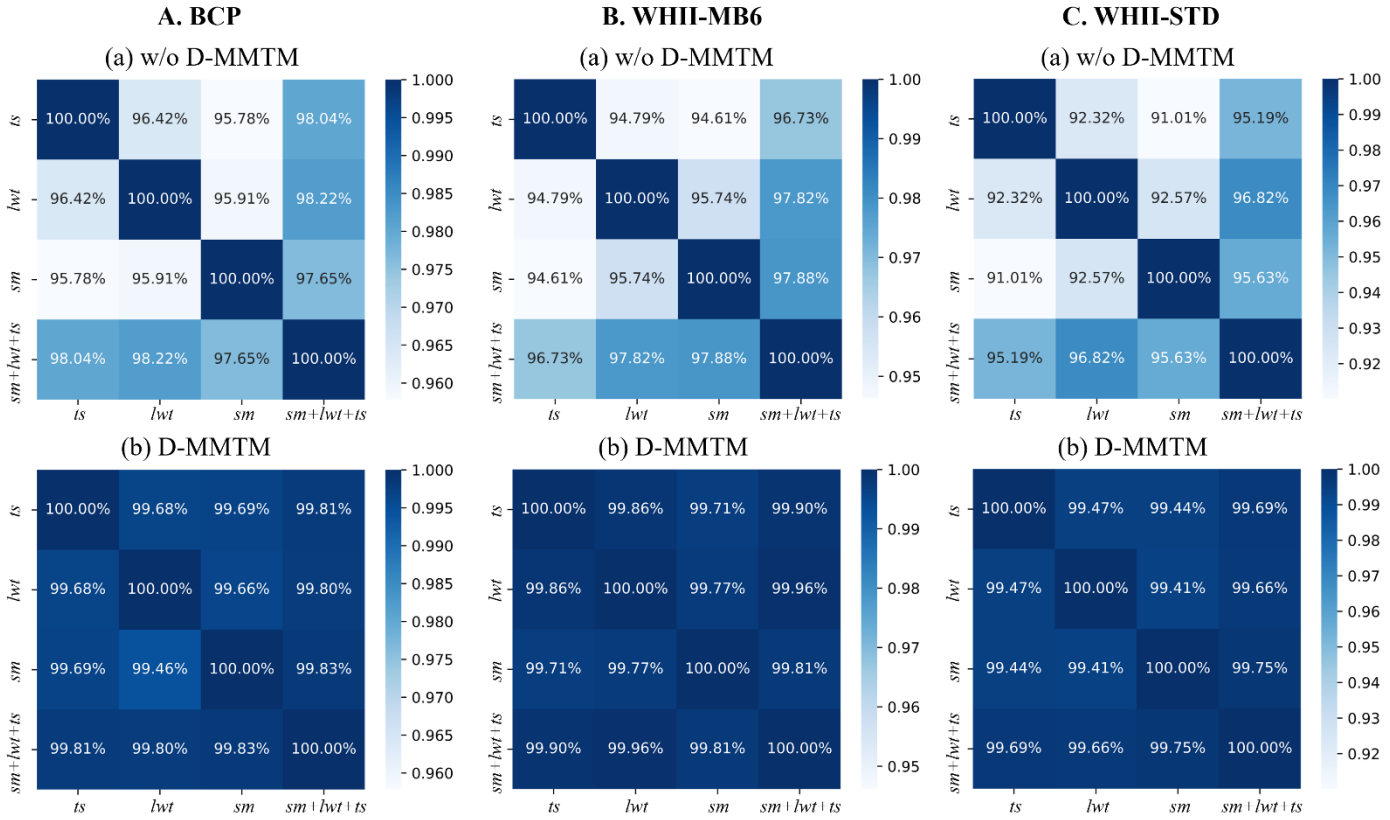

Fig. S1. Consistency among the decisions made by each individual modal stream (i.e., *sm*, *lwt*, and *ts*) and the multi-modality fusion model (i.e., *sm+lwt+ts*) from $CNN_{sm+lwt+ts}$ on the (A) BCP, (B) WHII-MB6, and (C) WHII-STD datasets, (a) without and (b) with D-MMTM.