



(a) Attention weight map in last layer. The attention weights (red) of 6 granularity's routers are plotted against the raw input (dark gray, 1s at 256Hz). X-axis: time/token index, y-axis: channel amplitude (with offset for visualization). Deeper red means more attention weight for the according token.

(b) Classification performance with masked brain regions. The lower plot shows the brain regions that are available (deep red) and unavailable (light pink) in each setting. The upper plot shows the according mean prediction accuracy and F1 score with standard deviation over 5 independent runs. Central sulcus, Frontal, Occipital and Parietal lobes contribute to the accuracy with descending importance.

Figure R.1: Explainability analysis on TDBrain dataset

**Table R.1: New Datasets and Methods** We selected three old baselines in our paper that showed strong performance: Crossformer, Reformer, and Transformer. Additionally, we introduced three new baselines: TCN, ModernTCN, and Mamba. These six baselines were evaluated on one old dataset in our paper, TDBrain(6,240 samples, 2 classes), and three new datasets: two human activity recognition datasets, FLAAP (13,123 samples, 10 classes) and UCI-HAR (10,299 samples, 6 classes), and one multi-modal vital signs dataset, MIMIC-PERform-AF (20,400 samples, 2 classes), which includes PPG and ECG data. The bold number denotes the best result, and the underlined number denotes the second best.

Datasets	TDBrain (2 Classes)		FLAAP (10 Classes)		UCI-HAR (6 Classes)		MIMIC-PERform-AF (2 Classes)	
Metrics	Accuracy	F1 Score	Accuracy	F1 Score	Accuracy	F1 Score	Accuracy	F1 Score
Models								
Crossformer	81.56 $\pm$ 2.19	81.50 $\pm$ 2.20	<u>75.84<math>\pm</math>0.52</u>	<u>75.52<math>\pm</math>0.66</u>	89.74 $\pm$ 1.08	89.70 $\pm$ 1.10	<u>73.93<math>\pm</math>2.98</u>	<u>73.87<math>\pm</math>2.99</u>
Reformer	87.92 $\pm$ 2.01	87.85 $\pm$ 2.08	71.65 $\pm$ 1.27	71.14 $\pm$ 1.45	88.44 $\pm$ 2.02	88.34 $\pm$ 1.98	70.12 $\pm$ 5.14	69.68 $\pm$ 5.47
Transformer	87.17 $\pm$ 1.67	87.10 $\pm$ 1.68	74.96 $\pm$ 1.25	74.49 $\pm$ 1.39	88.86 $\pm$ 1.65	88.80 $\pm$ 1.67	65.00 $\pm$ 4.17	63.81 $\pm$ 3.35
TCN	80.92 $\pm$ 2.94	80.82 $\pm$ 3.03	66.48 $\pm$ 1.66	65.29 $\pm$ 1.74	<b>93.08<math>\pm</math>0.95</b>	<b>93.19<math>\pm</math>0.88</b>	72.37 $\pm$ 1.59	71.30 $\pm$ 1.27
ModernTCN	81.96 $\pm$ 2.12	81.79 $\pm$ 2.23	74.80 $\pm$ 0.96	74.35 $\pm$ 0.85	91.44 $\pm$ 1.01	91.47 $\pm$ 0.98	73.26 $\pm$ 3.65	72.21 $\pm$ 4.60
Mamba	<u>89.58<math>\pm</math>0.74</u>	<u>89.58<math>\pm</math>0.73</u>	64.87 $\pm$ 2.78	64.14 $\pm$ 2.70	87.78 $\pm$ 1.10	87.72 $\pm$ 1.10	60.53 $\pm$ 60.53	60.53 $\pm$ 1.99
Medformer	<b>89.62<math>\pm</math>0.81</b>	<b>89.62<math>\pm</math>0.81</b>	<b>76.44<math>\pm</math>0.64</b>	<b>76.25<math>\pm</math>0.65</b>	<u>91.65<math>\pm</math>0.74</u>	<u>91.61<math>\pm</math>0.75</u>	<b>77.47<math>\pm</math>2.24</b>	<b>76.85<math>\pm</math>2.46</b>

**Table R.2: Running Time.** The running time for five seed experiments of 6 baselines and our method on FLAAP datasets. Experiments run on RTX 4090 GPU. The m denotes minute, and h denotes hour.

Datasets	FLAAP		
Metrics	Run Time	Accuracy	F1 Score
Models			
Crossformer	17m	75.84 $\pm$ 0.52	75.52 $\pm$ 0.66
Reformer	21m	71.65 $\pm$ 1.27	71.14 $\pm$ 1.45
Transformer	8m	74.96 $\pm$ 1.25	74.49 $\pm$ 1.39
TCN	2m	66.48 $\pm$ 1.66	65.29 $\pm$ 1.74
ModernTCN	7m	74.80 $\pm$ 0.96	74.35 $\pm$ 0.85
Mamba	3h 42m	64.87 $\pm$ 2.78	64.14 $\pm$ 2.70
Medformer	24m	<b>76.44<math>\pm</math>0.64</b>	<b>76.25<math>\pm</math>0.65</b>