TABLE I

SUMMARY OF THE HYPERPARAMETER SEARCH SPACE USED FOR THE XGBOOST REGRESSOR, WHEN RUNNING SMAC3 [**?**] AND THE SELECTED CONFIGURATION FOR ALL INPUT COMBINATIONS.

Hyperparameters	Our XGBoost Regressor								
	Search Space	Input 1	Input 2	Input 3	Input 4	Input 5	Input 6	Input 7	Input 8
eta	Uniform Float $\in$ [0.0001, 0.5]	0.0431	0.06818	0.04942	0.04457	0.07178	0.06647	0.04901	0.04798
gamma	Uniform Integer $\in$ [0, 10]	1	7	2	1	7	3	3	2
number estimate	Uniform Integer $\in$ [100, 1000]	972	155	932	880	230	913	715	949
max depth	Uniform Integer $\in$ [1, 10]	10	10	10	10	10	10	10	10
min child weight	Uniform Integer $\in$ [1, 100]	95	6	87	67	59	83	84	84
max delta step	Uniform Integer $\in$ [0, 10]	7	10	0	1	5	3	10	1
subsample	Uniform Float $\in$ [0.5, 1]	0.647	0.9632	0.5042	0.5068	0.6114	0.5486	0.507	0.5038
colsample bytree	Uniform Float $\in$ [0.5, 1]	0.9825	0.9517	0.8671	0.868	0.7449	0.8242	0.9927	0.7814
colsample bylevel	Uniform Float $\in$ [0.5, 1]	0.9819	0.9223	0.8572	0.839	0.9264	0.7642	0.9479	0.8407
colsample bynode	Uniform Float $\in$ [0.5, 1]	0.8042	0.9155	0.6173	0.7416	0.9802	0.924	0.8621	0.9989

TABLE II

TABLE SUMMARIZING THE HYPERPARAMETER SEARCH SPACE USED FOR THE FEED-FORWARD DEEP NEURAL NETWORK, WHEN RUNNING SMAC3 [**?**] AND THE SELECTED CONFIGURATION FOR ALL INPUT COMBINATIONS.

Hyperparameters	Our Feed-Forward Neural Network								
	Search Space	Input 1	Input 2	Input 3	Input 4	Input 5	Input 6	Input 7	Input 8
Batch Size	Categorical [256, 512]	256	512	256	512	512	512	512	512
Learning Rate	Categorical $a \times e^{-c}$ for $a \in \mathbb{N}^+$ and $\in [1, 9]$ $c \in \mathbb{N}^+$ and $\in [2, 5]$	0.0003	0.0002	0.0005	0.0006	0.0003	0.0006	0.0004	0.0005
Number of Layers (L)	Uniform Int Lower: 4 Upper: 12	7	10	7	7	7	9	7	8
Numebr of Neurons in Layer i for $i \in [0, L]$	Uniform Int Lower: 50 Upper: 1000 Step: 10	[860, 670, 160, 580, 900, 1000, 440]	[590,300, 820,520, 90,670, 850,120, 330,570]	[620, 470, 120, 620, 830, 890, 350]	[730, 390, 120, 630, 770, 720, 380]	[820,740, 190,740, 1000,850, 430]	[580,710, 600,170, 270,350, 70,780, 690]	[580,580, 160,450, 920,920, 380]	[470,600, 440,830, 790,900, 190,270]
Activation function in Layer i for $i \in [0, L]$	Categorical [sigmoid, relu, hardtanh, tanh, leakyrelu, elu]	[hardtanh, tanh, elu, relu, leakyrelu, relu, leakyrelu]	[elu, relu, hardtanh, elu, leakyrelu, relu, elu, elu, elu, leakyrelu]	[tanh, elu, elu, leakyrelu, leakyrelu, elu, leakyrelu]	[tanh, tanh, elu, hardtanh, sigmoid, leakyrelu, leakyrelu]	[tanh, relu, elu, leakyrelu, elu, relu, leakyrelu]	[tanh, relu, leakyrelu, elu, leakyrelu, leakyrelu, leakyrelu, hardtanh, leakyrelu]	[tanh, relu, leakyrelu, hardtanh, sigmoid, relu, leakyrelu]	[tanh, relu, elu, leakyrelu, hardtanh, leakyrelu, elu, leakyrelu]
Negative Slope for Leakyrelu	Categorical $a \times e^{-1}$ for $a \in \mathbb{N}^+$ and $\in [1, 9]$	0.5	0.7	0.4	0.2	0.1	0.4	0.3	0.6

TABLE III

TABLE SUMMARIZING THE HYPERPARAMETER SEARCH SPACE USED FOR THE TRANSFORMER ENCODER, WHEN RUNNING SMAC3 [**?**] AND THE SELECTED CONFIGURATION FOR ALL INPUT COMBINATIONS.

Hyperparameters	Our Transformer Encoder								
	Search Space	Input 1	Input 2	Input 3	Input 4	Input 5	Input 6	Input 7	Input 8
Batch Size	Categorical [256, 512]	512	512	512	256	512	256	512	256
Learning Rate	Categorical $a \times e^{-c}$ for $a \in \mathbb{N}^+$ and $\in [1, 9]$ $c \in \mathbb{N}^+$ and $\in [2, 5]$	0.00009	0.00004	0.00009	0.0004	0.004	0.0003	0.00004	0.002
Number of Layers (L)	Uniform Int Lower: 2 Upper: 8	3	8	3	2	7	4	7	2
Number of Multi-Heads	Categorical [1, 2, 4, 8]	4	8	4	8	1	4	1	8
Dropout Rate	Categorical [0.0, 0.1, 0.2, 0.3, 0.4, 0.5]	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Embedding Dimensions	Categorical [32, 64, 128, 256, 512, 1024]	128	128	128	128	64	32	128	64
Hidden Dimension	Categorical [64, 128, 256, 512, 1024]	512	512	512	512	128	256	128	256

TABLE IV

RESULTS OF ALL TRAINED NETWORKS OVER ALL 10 FOLDS FOR ALL INPUT COMBINATIONS. THE VALUE IN PARENTHESES REPRESENTS THE STANDARD DEVIATION. THE METRICS ARE CALCULATED OVER ALL CHANNELS AND OVER ALL ELECTRODES. MAE IS CALCULATED ON THE OUTPUT VALUES IN THE ORIGINAL SCALE. NORM. MAE IS CALCULATED ON THE NORMALIZED OUTPUT OF THE NETWORK; THE LOWER, THE BETTER.

		Nb. Param.	MSE	Norm. MSE	MSE	Norm. MSE
			Over all Channels		Electrodes Only	
Ruppel et al. [**?**]	1	806K	18.651 (1.732)	0.223 (0.022)	16.804 (1.645)	0.227 (0.023)
	2	805K	18.818 (1.583)	0.223 (0.021)	16.720 (1.618)	0.226 (0.023)
	3	804K	20.100 (1.653)	0.227 (0.021)	16.784 (1.606)	0.227 (0.023)
	4	807K	18.209 (1.653)	0.221 (0.022)	16.777 (1.657)	0.227 (0.023)
	5	819K	18.223 (1.640)	0.223 (0.022)	16.958 (1.639)	0.229 (0.023)
	6	812K	18.424 (1.637)	0.222 (0.022)	16.823 (1.652)	0.227 (0.023)
	7	809K	18.421 (1.565)	0.221 (0.021)	16.618 (1.562)	0.225 (0.022)
	8	822K	18.065 (1.632)	0.221 (0.022)	16.803 (1.640)	0.227 (0.023)
Our XGBoost Regressor	1	1584K	13.368 (1.340)	0.150 (0.015)	11.446 (1.204)	**0.150** (0.015)
	2	1041K	14.372 (1.446)	0.159 (0.016)	12.002 (1.298)	0.158 (0.016)
	3	1209K	15.994 (1.622)	0.168 (0.017)	12.373 (1.340)	0.163 (0.017)
	4	1694K	13.275 (1.430)	0.152 (0.016)	11.638 (1.335)	0.153 (0.016)
	5	840K	13.407 (1.393)	0.153 (0.015)	11.774 (1.277)	0.154 (0.016)
	6	1046K	13.785 (1.385)	0.156 (0.015)	11.866 (1.277)	0.156 (0.016)
	7	1239K	13.299 (1.330)	0.150 (0.015)	11.375 (1.182)	**0.150** (0.015)
	8	1331K	**12.873** (1.255)	**0.148** (0.014)	11.385 (1.156)	**0.150** (0.014)

		Nb. Param.	MSE	Norm. MSE	MSE	Norm. MSE
			Over all Channels		Electrodes Only	
Our Feed-Forward Neural Network	1	2233K	14.693 (1.386)	0.168 (0.015)	12.724 (1.252)	0.169 (0.016)
	2	2881K	15.624 (1.536)	0.173 (0.017)	12.965 (1.388)	0.171 (0.018)
	3	1701K	17.010 (1.314)	0.181 (0.014)	13.265 (1.105)	0.175 (0.014)
	4	1478K	14.712 (1.317)	0.171 (0.015)	13.109 (1.247)	0.173 (0.016)
	5	2554K	14.056 (1.213)	0.166 (0.013)	12.721 (1.100)	0.169 (0.014)
	6	1124K	14.476 (1.150)	0.167 (0.013)	12.676 (1.043)	0.168 (0.013)
	7	1794K	15.231 (1.344)	0.177 (0.016)	13.351 (1.242)	0.178 (0.016)
	8	2490K	14.517 (1.450)	0.172 (0.017)	13.204 (1.404)	0.175 (0.017)
Our Transformer Encoder	1	599K	13.760 (1.400)	0.156 (0.016)	11.736 (1.280)	0.155 (0.016)
	2	203K	14.534 (1.482)	0.162 (0.017)	12.102 (1.374)	0.160 (0.017)
	3	598K	15.229 (1.617)	0.160 (0.017)	11.595 (1.328)	0.154 (0.017)
	4	401K	13.204 (1.363)	0.152 (0.016)	11.564 (1.266)	0.153 (0.016)
	5	237K	13.441 (1.241)	0.155 (0.014)	11.824 (1.147)	0.156 (0.015)
	6	114K	13.669 (1.505)	0.155 (0.018)	11.719 (1.376)	0.155 (0.018)
	7	701K	13.927 (1.614)	0.158 (0.019)	11.838 (1.468)	0.157 (0.019)
	8	103K	12.984 (1.552)	0.149 (0.018)	**11.334** (1.473)	**0.150** (0.019)

TABLE V

EXTENDED SIGNIFICANCE TEST WITH THE CORRECTED PAIRED t-TEST [**?**] CONDUCTED ON DIFFERENT INPUT COMBINATIONS FOR ALL NETWORKS. THE FIRST VALUE DEPICTS THE PAIRED NORMALIZED MAE DIFFERENCE IN PERCENT OVER THE TEN FOLDS, THE SECOND VALUE REPRESENTS t-STATISTIC, AND THE THIRD VALUE BETWEEN PARENTHESIS REPRESENTS THE p-VALUE.

	1 vs 2	1 vs 3	1 vs 4	1 vs 5	1 vs 6	7 vs 1	8 vs 1	5 vs 6	8 vs 5
Ruppel et al. [**?**]	-0.944% -3.723 **(0.002)**	-1.809% -7.624 **(0.000)**	-0.514% -2.132 **(0.031)**	-0.667% -2.975 **(0.008)**	-0.715% -3.107 **(0.006)**	0.715% 3.107 (0.994)	0.584% 2.058 (0.965)	-0.048% -0.575 (0.290)	-0.083% -0.754 (0.235)
Our XGBoost Regressor	-0.859% -8.660 **(0.000)**	-1.773% -9.979 **(0.000)**	-0.113% -1.206 (0.129)	-0.282% -4.143 **(0.001)**	-0.539% -5.998 **(0.000)**	-0.055% -0.923 (0.190)	-0.208% -2.889 **(0.009)**	-0.258% -2.860 **(0.009)**	-0.490% -4.046 **(0.001)**
Our Feed-Forward Neural Network	-0.441% -2.222 **(0.027)**	-1.267% -5.436 **(0.000)**	-0.324% -2.184 **(0.028)**	0.217% 0.589 (0.715)	0.133% 0.373 (0.641)	0.862% 1.995 (0.961)	0.402% 0.729 (0.758)	-0.084% -0.226 (0.413)	0.620% 1.296 (0.886)
Our Transformer Encoder	-0.587% -2.479 **(0.018)**	-0.391% -1.400 (0.097)	0.385% 1.074 (0.845)	0.077% 0.344 (0.631)	0.083% 0.438 (0.664)	0.188% 0.461 (0.672)	-0.662% -1.870 **(0.047)**	0.007% 0.019 (0.507)	-0.585% -1.537 (0.079)

TABLE VI

EXTENDED SIGNIFICANCE TEST WITH THE CORRECTED PAIRED t-TEST [**?**] CONDUCTED FOR ALL NETWORK PAIRS. THE FIRST VALUE DEPICTS THE PAIRED NORMALIZED MAE DIFFERENCE IN PERCENT OVER THE TEN FOLDS, THE SECOND VALUE REPRESENTS t-STATISTIC, AND THE THIRD VALUE BETWEEN PARENTHESIS REPRESENTS THE p-VALUE.

vs	Our XGBoost Ruppel et al. [**?**]	Our FFNN Ruppel et al. [**?**]	Our Transformer Ruppel et al. [**?**]	Our XGBoost Our FFNN	Our XGBoost Our Transformer	Our Transformer Our FFNN
1	-7.265% -8.688 **(0.000)**	-5.487% -6.240 **(0.000)**	-6.734% -8.053 **(0.000)**	-1.778% -5.188 **(0.000)**	-0.531% -1.581 (0.074)	-1.247% -3.690 **(0.002)**
2	-7.350% -9.037 **(0.000)**	-5.990% -7.084 **(0.000)**	-7.091% -8.032 **(0.000)**	-1.360% -4.284 **(0.001)**	-0.259% -0.704 (0.250)	-1.101% -5.338 **(0.000)**
3	-7.302% -8.965 **(0.000)**	-6.029% -7.134 **(0.000)**	-8.152% -10.181 **(0.000)**	-1.272% -3.623 **(0.003)**	0.851% 3.407 (0.996)	-2.123% -6.570 **(0.000)**
4	-7.666% -8.550 **(0.000)**	-5.677% -6.088 **(0.000)**	-7.633% -7.694 **(0.000)**	-1.989% -5.653 **(0.000)**	-0.032% -0.106 (0.459)	-1.956% -11.260 **(0.000)**
5	-7.651% -8.912 **(0.000)**	-6.371% -7.707 **(0.000)**	-7.478% -10.302 **(0.000)**	-1.279% -3.888 **(0.002)**	-0.173% -0.477 (0.322)	-1.107% -3.810 **(0.002)**
6	-7.441% -9.045 **(0.000)**	-6.335% -7.078 **(0.000)**	-7.533% -9.174 **(0.000)**	-1.106% -3.610 **(0.003)**	0.092% 0.303 (0.616)	-1.197% -3.918 **(0.002)**
7	-8.036% -9.613 **(0.000)**	-5.340% -7.140 **(0.000)**	-7.261% -7.370 **(0.000)**	-2.696% -5.235 **(0.000)**	-0.774% -1.347 (0.105)	-1.921% -5.899 **(0.000)**
8	-8.058% -10.024 **(0.000)**	-5.668% -5.135 **(0.000)**	-7.980% -9.696 **(0.000)**	-2.389% -4.242 **(0.001)**	-0.077% -0.157 (0.439)	-2.312% -4.004 **(0.002)**

TABLE VII

NUMBER OF PARAMETERS IN THOUSANDS, INFERENCE TIME IN MILLISECONDS, AND FLOATING-POINT OPERATIONS PER SECOND IN MILLIONS FOR ALL NEURAL NETWORKS ACROSS ALL INPUT COMBINATIONS.

	Our XGBoost			Our FFNN			Our Transformer		
	Num. Param.	Inference (ms)	Num. FLOPS	Num. Param.	Inference (ms)	Num. FLOPS	Num. Param.	Inference (ms)	Num. FLOPS
1	806K	0.628	1.61M	2233K	0.876	4.46M	599K	1.618	5.97M
2	805K	0.539	1.61M	2881K	1.173	5.75M	203K	2.982	1.63M
3	804K	0.539	1.61M	1701K	0.741	3.40M	598K	1.396	3.58M
4	807K	0.555	1.61M	1478K	0.757	2.95M	401K	1.005	5.59M
5	819K	0.572	1.65M	2554K	0.833	5.10M	237K	2.995	11.64M
6	812K	0.552	1.62M	1124K	0.613	2.24M	114K	1.427	2.98M
7	809K	0.572	1.62M	1794K	0.760	3.58M	701K	2.939	9.89M
8	822K	0.607	1.65M	2490K	1.002	4.98M	103K	0.988	5.30M