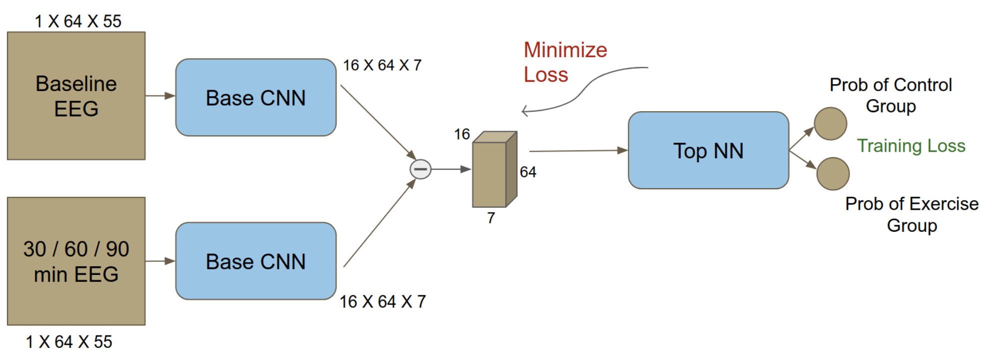
Deep Semantic Architecture with discriminative feature visualization for neuroimage analysis

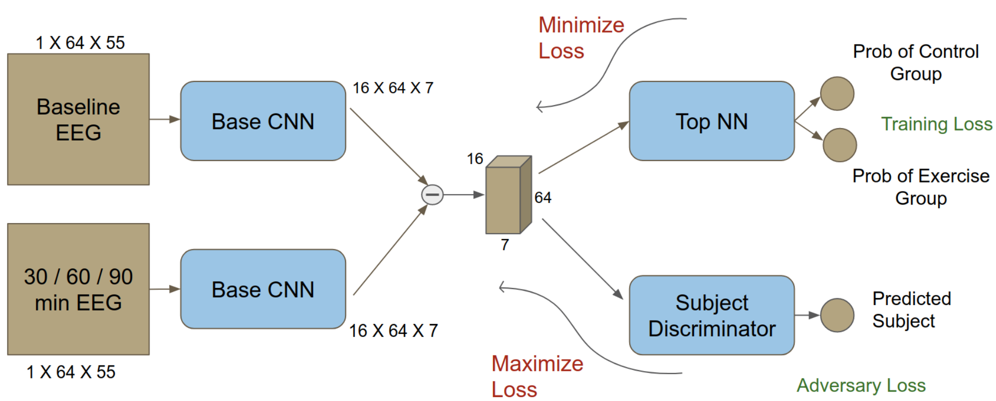
Anonymous Author(s)
Affiliation
Address
`email`

Supplementary Material

1 Network Architecture



(a) Basic Architecture without adversary



(b) Modified Architecture with adversary to avoid subject discrimination

Figure S1: Deep Network Architecture. The initial choice of architecture (without any adversary) gives good subject prediction accuracy from features extracted by the Base CNN. Therefore, a subject discriminator of roughly the same model capacity as the Top NN is added. The subject discrimination acts as a regularizer while training and avoids the Base CNN from learning subject specific features.

Submitted to 32nd Conference on Neural Information Processing Systems (NIPS 2018). Do not distribute.

Notation:- *Conv* denotes the 2D Spatial Convolutional layer. *ReLU* denotes the Rectified Linear Unit Layer that adds non-linearity to the network. *MaxPool* denotes the 2D Spatial Max Pooling layer. *FullyConn* denotes a Fully Connected layer, also known as the linear layer of the network.

1.1 TF maps

Layer	Type	Maps and Neurons	Filter Size
0	Input	1M × 64 × 55N	-
1	Conv	6M × 64 × 28N	1×5
2	ReLU	6M × 64 × 28N	-
3	MaxPool	6M × 64 × 14N	1×2
4	Conv	16M × 64 × 14N	1×5
5	ReLU	16M × 64 × 14N	-
6	MaxPool	16M × 64 × 7N	1×2

Table S1: Network architecture used for EEG feature extraction network (Base CNN). The output of the network is a tensor of dimensions $16 \times 64 \times 7$.

Layer	Type	Maps and Neurons	Filter Size
0	Input	16M × 64 × 7N	-
1	Flatten	7168N	-
2	Dropout (p=0.5)	-	-
3	FullyConn	8N	1×1
4	ReLU	8N	-
5	FullyConn	2N	1×1

Table S2: Network architecture used for group discrimination network (Top NN). The output of the network is a vector of dimension 2, values corresponding to the probability that the data tuple belongs to particular class.

Layer	Type	Maps and Neurons	Filter Size
0	Input	16M × 64 × 7N	-
1	Flatten	7168N	-
2	Dropout (p=0.5)	-	-
3	FullyConn	8N	1×1
4	ReLU	8N	-
5	FullyConn	25N	1×1

Table S3: Network architecture used for subject discrimination network (adversary). The output of the network is a vector of dimension 25, values corresponding to the probability that the data tuple belongs to particular subject.

1.2 Topographical maps

Layer	Type	Maps and Neurons	Filter Size
0	Input	3M $\times$ 64 $\times$ 64N	-
1	Conv	16M $\times$ 32 $\times$ 32N	5$\times$5
2	ReLU	16M $\times$ 32 $\times$ 32N	-
3	MaxPool	16M $\times$ 16 $\times$ 16N	2$\times$2
4	Conv	32M $\times$ 16 $\times$ 16N	5$\times$5
5	ReLU	32M $\times$ 16 $\times$ 16N	-
6	MaxPool	32M $\times$ 8 $\times$ 8N	2$\times$2
7	Conv	64M $\times$ 8 $\times$ 8N	3$\times$3
8	ReLU	64M $\times$ 8 $\times$ 8N	-
9	MaxPool	64M $\times$ 4 $\times$ 4N	2$\times$2

Table S4: Network architecture used for EEG feature extraction network (Base CNN). The output of the network is a tensor of dimensions $64 \times 4 \times 4$.

Layer	Type	Maps and Neurons	Filter Size
0	Input	64M $\times$ 4 $\times$ 4N	-
1	Flatten	1024N	-
2	Dropout (p=0.5)	-	-
3	FullyConn	8N	1$\times$1
4	ReLU	8N	-
5	FullyConn	2N	1$\times$1

Table S5: Network architecture used for group discrimination network (Top NN). The output of the network is a vector of dimension 2, values corresponding to the probability that the data tuple belongs to particular class.

Layer	Type	Maps and Neurons	Filter Size
0	Input	64M $\times$ 4 $\times$ 4N	-
1	Flatten	1024N	-
2	Dropout (p=0.5)	-	-
3	FullyConn	8N	1$\times$1
4	ReLU	8N	-
5	FullyConn	25N	1$\times$1

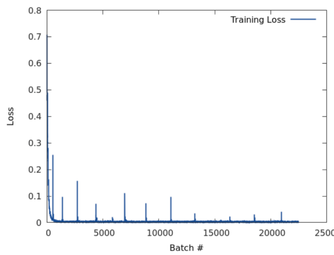
Table S6: Network architecture used for subject discrimination network (adversary). The output of the network is a vector of dimension 25, values corresponding to the probability that the data tuple belongs to particular subject.

2 Training curves

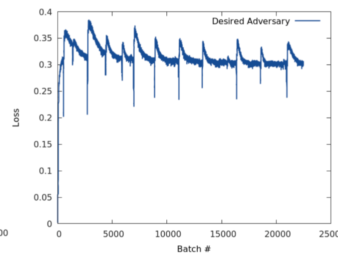
2.1 Time-Frequency Maps

Hyperparameter	Value
Learning Rate	0.001
Learning Rate Decay	0.0001
Weight Decay	0.001

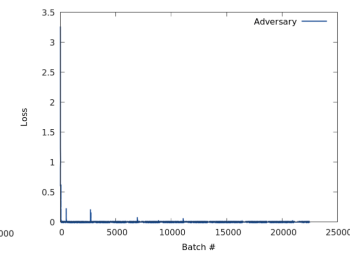
Table S7: List of hyperparameters used for training the networks on TF maps.



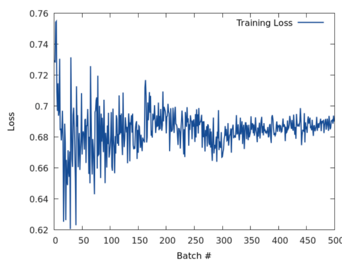
(a) Group prediction loss ($\lambda = 0$)



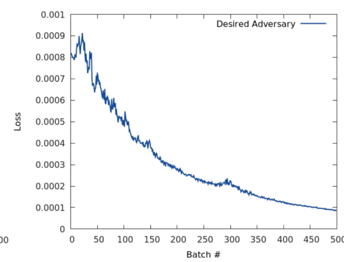
(b) KL divergence ($\lambda = 0$)



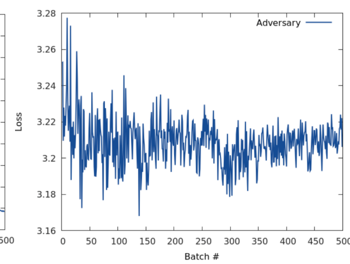
(c) Subject prediction loss ($\lambda = 0$)



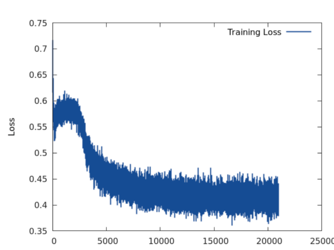
(d) Group prediction loss ($\lambda = 15$)



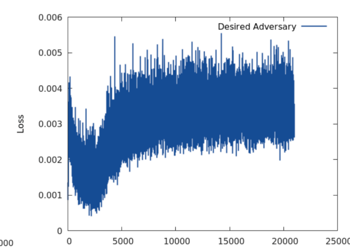
(e) KL divergence ($\lambda = 15$)



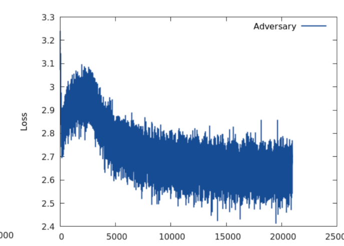
(f) Subject prediction loss ($\lambda = 15$)



(g) Group prediction loss ($\lambda = 13$)



(h) KL divergence ($\lambda = 13$)



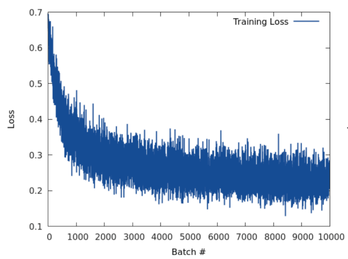
(i) Subject prediction loss ($\lambda = 13$)

Figure S2: Time-Frequency Maps Training curves for three different weight values to the subject predictor regularizer.

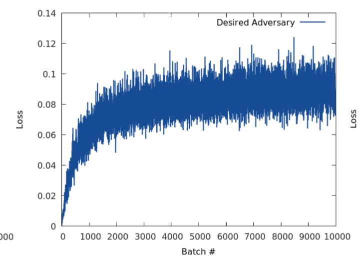
2.2 Topographical Maps

Hyperparameter	Value
Learning Rate	0.001
Learning Rate Decay	0.001
Weight Decay	0.03

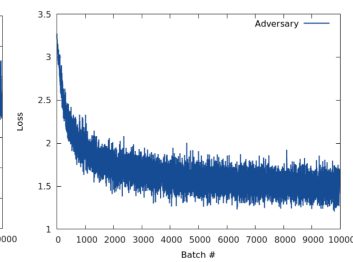
Table S8: List of hyperparameters used for training the networks on Topographical maps.



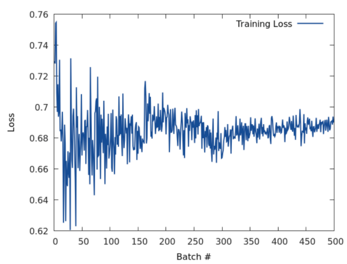
(a) Group prediction loss ($\lambda = 0$)



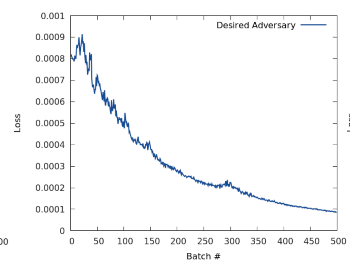
(b) KL divergence ($\lambda = 0$)



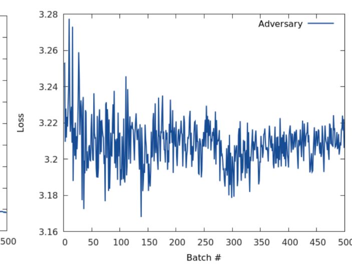
(c) Subject prediction loss ($\lambda = 0$)



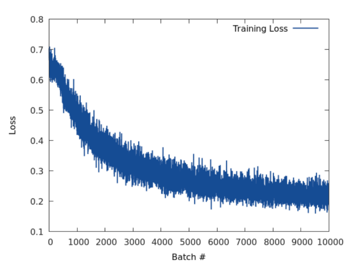
(d) Group prediction loss ($\lambda = 10$)



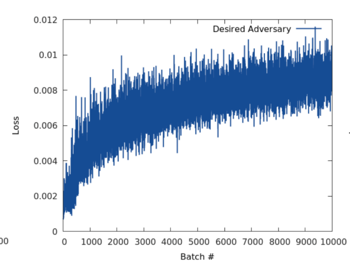
(e) KL divergence ($\lambda = 10$)



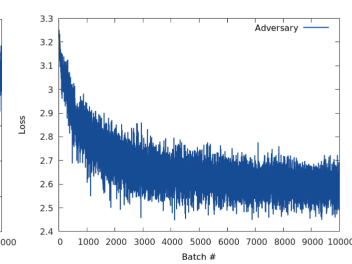
(f) Subject prediction loss ($\lambda = 10$)



(g) Group prediction loss ($\lambda = 5$)



(h) KL divergence ($\lambda = 5$)



(i) Subject prediction loss ($\lambda = 5$)

Figure S3: Topographical Maps Training curves for three different weight values to the subject predictor regularizer.