Baseline Selection:

We select “Shallow and Deep ConvNet” from [] as our baseline because they render similar performance in terms of accuracy with the least computational cost in []. Shallow ConvNext consists of two convolution layers which convolve the temporal and spatial dimensions respectively. Then, there is a single mean pooling layer followed by a classification layer which consists of a dense linear layer and a Softmax layer. On the other hand, Deep ConvNet also consists of temporal and spatial convolution layers which is the same as Shallow ConvNet. After that, Deep ConvNet has three convolution blocks, each consists of a convolution layer and a max pooling layer. The input size is down sampled to 1/3 after each convolution block. Finally, there is a linear classification layer and a LogSoftmax layer. It is noteworthy that these architectures explicitly attempted to replicate the behaviors of the Filter bank common spatial patterns (FBCSP) pipeline, which was the state-of-the-art EEG signal interpretation technique at the time. Since these two models were not designed for analyzing EEG data of MBT-42 dataset in the original paper, we need to modify some of their parameters such as kernel size and stride to fit the data shape of MBT-42 dataset. The modified Shallow and Convnets will be used as our baseline models.