Layer	Input	Operation	Output	Parameters
1	$E \times T$	$25 \times$ Conv2D (1×10)	$E \times 1015 \times 25$	275
2	$E \times 1015 \times 25$	Reshape	$E \times 1015 \times 25 \times 1$	-
	$E \times 1015 \times 25 \times 1$	$25 \times$ Conv3d $(3 \times 1 \times 25)$	$1 \times 1015 \times 1 \times 25$	1,900
	$1 \times 1015 \times 1 \times 25$	BatchNorm	$1 \times 1015 \times 1 \times 25$	100
	$1 \times 1015 \times 1 \times 25$	ELU	$1 \times 1015 \times 1 \times 25$	-
	$1 \times 1015 \times 1 \times 25$	MaxPool2D (3×1)	$338 \times 25 \times 1$	-
	$338 \times 25 \times 1$	Dropout (0.5)	$338 \times 25 \times 1$	-
3	$338 \times 25 \times 1$	$50 \times$ Conv2D (10×25)	$329 \times 1 \times 50$	12,550
	$329 \times 1 \times 50$	BatchNorm	$329 \times 1 \times 50$	200
	$329 \times 1 \times 50$	ELU	$329 \times 1 \times 50$	-
	$329 \times 1 \times 50$	MaxPool2D (3×1)	$109 \times 1 \times 50$	-
	$109 \times 1 \times 50$	Dropout (0.5)	$109 \times 1 \times 50$	-
4	$109 \times 1 \times 50$	Reshape	$109 \times 50 \times 1$	-
	$109 \times 50 \times 1$	$100 \times$ Conv2D (10×50)	$100 \times 1 \times 100$	50,100
	$100 \times 1 \times 100$	BatchNorm	$100 \times 1 \times 100$	400
	$100 \times 1 \times 100$	ELU	$100 \times 1 \times 100$	-
	$100 \times 1 \times 100$	MaxPool2D	$33 \times 1 \times 100$	-
	$33 \times 1 \times 100$	Dropout (0.5)	$33 \times 1 \times 100$	-
5	$33 \times 1 \times 100$	Reshape	$33 \times 100 \times 1$	-
	$33 \times 100 \times 1$	$200 \times$ Conv2D (10×100)	$24 \times 1 \times 200$	200,200
	$24 \times 1 \times 200$	BatchNorm	$24 \times 1 \times 200$	800
	$24 \times 1 \times 200$	ELU	$24 \times 1 \times 200$	-
	$24 \times 1 \times 200$	MaxPool2D (3×1)	$8 \times 1 \times 200$	-
6	$8 \times 1 \times 200$	Flatten	1600	-
	1600	Softmax	K	3,202
Total				$268,977$

Table 2.1: Deep CNN architecture as proposed by Schirrmeister et al. [60] (re-implemented in this work), where E is the number of channels, T is the number of timesteps and K is the number of classes. Input and Output sizes are shown for cropped training with $E = 3$ (electrodes C3, C4 and Cz) and $T = 1024$ for window size of 4 seconds; binary classification with two classes for $K = 2$.