## Gracia Gracia

Method 1: Not using – per subject model

Method 2:

* PVT1 vs PVT2 with 1000 images per trial for training, 15% of this for validation. Remaining data (most of the data) for testing
* Tested on PVT 3 (77.82%)

Method 3:

* PVT1 vs PVT2 for training (same as method 2 dataset?)
* Evaluate on first, second and third trials for subjects 3, 8 and 14

Our method correctly labels subjects 3, 8 and 14 using the leave-one-out approach.

## Maior

* Used a different dataset to develop model. On DROZY, only tested on KSS <=3 for PVT1 and KSS >=7 for PVT3 and if a warning is given.
* Method and Results are in the relevant behavioural segment as this model is not altered from our method of LOOCV we cannot build on an external dataset.

Our results on relevant trials with a model built for all subjects KSS789-6:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Subject | Trial 1 |  |  | Trial 3 |  |  |
|  | KSS | Maior | Correct? | KSS | Maior | Correct? |
| 1 | 3 | Y | Y | 7 | Y | Y |
| 2 | 3 | Y | Y | - | - |  |
| 3 | 2 | Y | Y | - | - |  |
| 4 | - | - | - | 9 | Y | Y |
| 5 | 3 | Y | Y | 8 | N | Y |
| 6 | 2 | Y | N | 7 | Y | Y |
| 7 | - | - | - | 9 | Y | Y |
| 8 | 2 | Y | Y | 8 | Y | Y |
| 9 | - | - | - | 8 | Y | Y |
| 10 | 3 | Y | Y | 7 | Y | Y |
| 11 | - | - | - | 7 | Y | N |
| 12 | 2 | Y | Y | - | - | - |
| 13 | - | - | - | - | - | - |
| 14 | - | - | - | 8 | Y | Y |

## Nxgande

* Random held out data, labels not specified

## Tested results:

Nxgande: PVT1 vs PVT2: Using 30% random held-out data, then 15% of the remaining data for validation

|  |  |  |  |
| --- | --- | --- | --- |
| Test Number | Confusion Matrix | Accuracy on 30% held-out | PVT3 Accuracy |
| 1 | [[2728 5]  [ 1 2122]] | 99.88% | 82.15% |
| 2 | [[2763 9]  [ 20 2064]] | 99.40% | 83.30% |
| 3 | [[2725 1]  [ 6 2124]] | 99.86% | 78.13% |
| 4 | [[2761 7]  [ 1 2087]] | 99.84% | 86.78% |
| 5 | [[2768 6]  [ 0 2082]] | 99.88% | 85.96% |
| Average |  | >99% | 83.26% \pm 3.44% |

Results using KSS labelling splits

|  |  |  |
| --- | --- | --- |
| Test Number | Confusion Matrix | Accuracy on 30% held-out |
| 1 | [[4473 9]  [ 33 2592]] | 99.41% |
| 2 | [[4480 8]  [ 25 2594]] | 99.54% |
| 3 | [[4411 13]  [ 24 2659]] | 99.48% |
| 4 |  |  |
| 5 |  |  |
| Average |  |  |

Gracia: PVT1 vs PVT2: Using 100 sequences per subject for training (85 for training and 15 for validation: 1 sequence is equivalent to 10 frames). The remaining sequences for testing. (Test on PVT3 also).

|  |  |  |  |
| --- | --- | --- | --- |
| Test Number | Accuracy on remaining PVT1 | Accuracy on remaining PVT2 | PVT3 Accuracy |
| 1 | 99.78% | 99.90% | 84.59% |
| 2 | 99.95% | 99.75% | 76.70% |
| 3 | 99.86% | 99.43% | 76.93% |
| 4 | 99.41% | 99.73% | 90.15% |
| 5 | 99.50% | 99.80% | 89.61% |
| Average | >99% | >99% | 83.60% \pm 6.56% |

Maior: Leave-one-out approach is complex to implement with specific trials. Therefore we look at first versus last, regardless of KSS. (Half-trial, ResNet)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Subject | Trial 1 |  |  | Trial 3 |  |  |
|  | KSS | Maior | Correct? | KSS | Maior | Correct? |
| 1 | 3 | Y | Y | 7 | Y | Y |
| 2 | 3 | Y | Y | - | - | - |
| 3 | 2 | Y | Y | - | - | - |
| 4 | - | - | - | 9 | Y | Y |
| 5 | 3 | Y | Y | 8 | N | Y |
| 6 | 2 | Y | Y | 7 | Y | Y |
| 7 | - | - | - | 9 | Y | Y |
| 8 | 2 | Y | N | 8 | Y | Y |
| 9 | - | - | - | 8 | Y | Y |
| 10 | 3 | Y | Y | 7 | Y | Y |
| 11 | - | - | - | 7 | Y | Y |
| 12 | 2 | Y | Y | - | - | - |
| 13 | - | - | - | - | - | - |
| 14 | - | - | - | 8 | Y | Y |