**MTBI NP Progress Report**

**Number of Available Data:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | T1 | T2 | T3 | T2 & T3 Shared |
| MTBI | 84 | 26 | 32 | 13 |
| Control | 70 | 26 | 27 | 11 |
| Total | 154 | 52 | 59 | 24 |

Table 1: Available subject count for each time spot.

**Input Features:**

Two different feature sets have been used for regression. First set is the statistical and the second set is the Adversarial BoW features.

The list of metrics used for statistical features are AK, MD, AWF, Deperp, Depar, RK, MK, RD, Tort, AD, DA and FA. For each metric mean and standard deviation value were extracted for 5 different components: left and right thalamus, CC body, genu and splenium. So for each metric we have a total 2x5 = 10 features. In total feature space is of dimension 120. After extracting these metrics I have concatenated the metrics together and used the greedy forward feature selection followed by support vector regression. I have used these metrics to fit only the T1 Digit span backward results.

The deep words are the same words that we have previously obtained using a deep adversarial network on patches followed by bag-of-words. The metrics are AK, MD, AWF, Deperp, Depar, RK, MK, DA and FA. But for these words we are not using CC body and CC genu regions. The metrics AK, FA, MD, MK, RK are used in both thalamus and CCS regions and the rest are used only in CCS. So in total we have 280 features as every metric-region pair is a histogram of size 20.

**Train - Test Split and Validation:**

I have used 20 test subjects out of the available 154 and the rest were used as training. During training for every feature set the train-test split data was shuffled 25 times. And the model performance is evaluated as an average across these 25 shuffling. During greedy forward selection the feature with the highest R2 value across 25 shuffling is added to the selected feature set. Currently I am not using a held-out a set but plan to do so after I establish a framework with good results.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Selected Features | Mean Square Error | R2 (1- MSE/var) |
| **Statistical Features SVR (Patient and Control Combined Model):** | AWF\_CC\_genu\_mean  AD\_RThal\_std  RD\_CC\_genu\_std  Deperp\_RThal\_mean  Deperp\_CC\_splen\_mean  RK\_CC\_body\_std  AK\_CC\_body\_mean  Depar\_CC\_body\_std  RD\_CC\_genu\_mean  RD\_RThal\_mean | 0.726 | 0.39 |
| **Statistical Features SVR (Only Patients Model):** | RD\_LThal\_mean  RK\_LThal\_mean  Depar\_RThal\_mean  AK\_CC\_Genu\_std  MD\_CC\_Body\_mean  AWF\_CC\_Body\_mean  RD\_CC\_splen\_std  AWF\_CC\_Genu\_mean  Tort\_CC\_Body\_mean  AD\_RThal\_std | 0.801 | 0.29 |
| **Statistical Features SVR (Only Control Model):** | FA\_LThal\_mean  DA\_CC\_Splen\_mean  MD\_CC\_Body\_mean  MK\_CC\_Body\_mean  RD\_CC\_Splen\_mean  AK\_CC\_Splen\_mean  RD\_CC\_Body\_mean  RD\_CC\_Genu\_mean | 0.603 | 0.33 |
| **Deep BOW Features SVR (Combined Model):** | CCS\_AK, CCS\_Depar, CCS\_RK, Thal\_FA, Thal\_RK, CCS\_MD, CCS\_depar, Thal\_AK, CCS\_AK, CCS\_AWF | 0.637 | 0.46 |

Table 2: Selected features and performances of different frameworks. Reported results are average test error across 25 shuffling of train-test data splits.

I have also tried a simple linear regression using greedy feature selection for Deep BOW but I had negative R2 value, indicating that the model performed worse than a constant fit. So I chose not to follow this approach further with the statistical features as well.

**Future work / questions:**

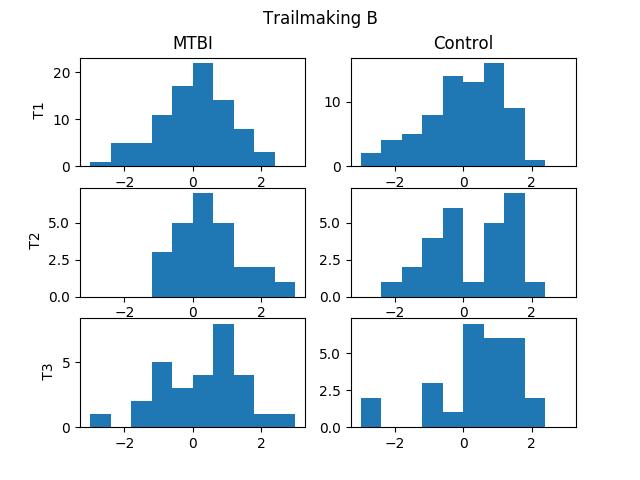
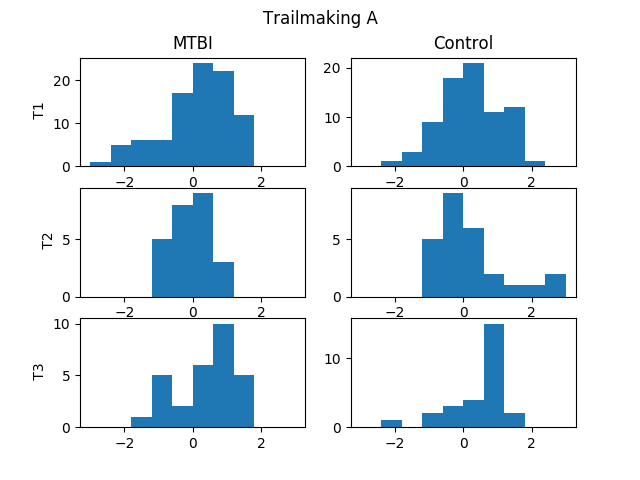
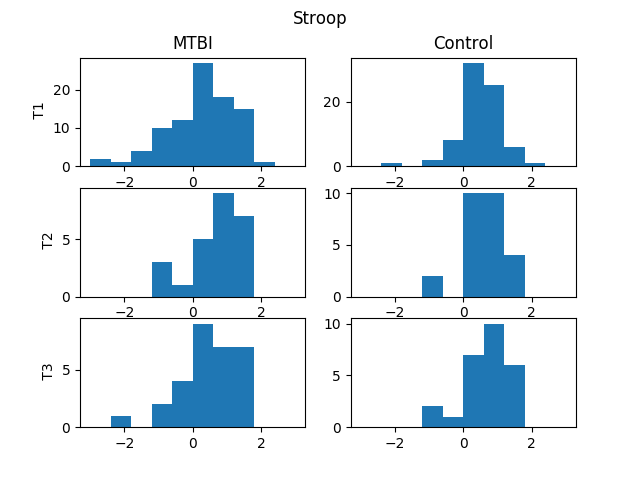
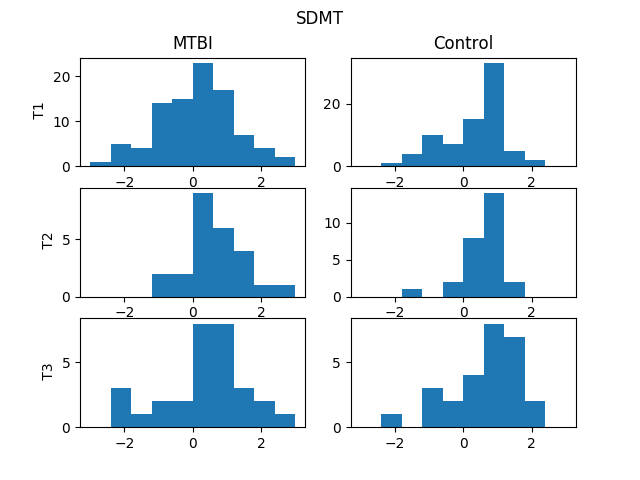
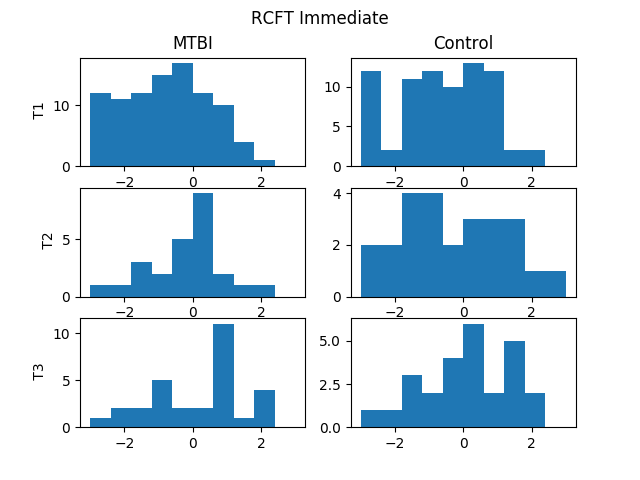
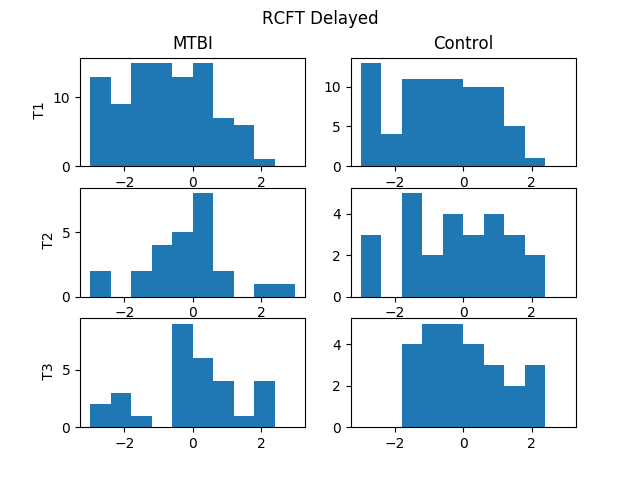
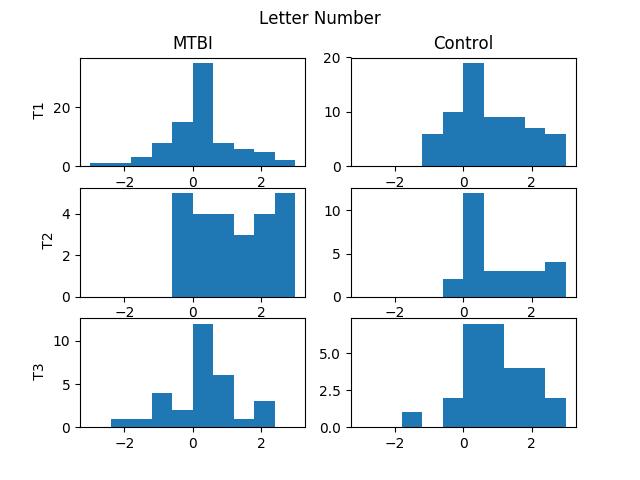
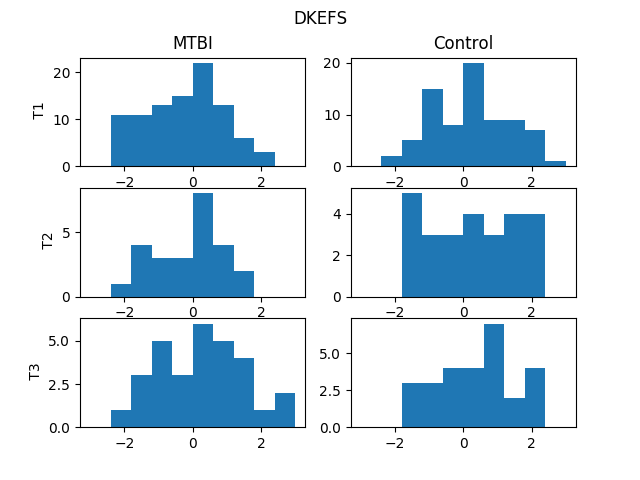
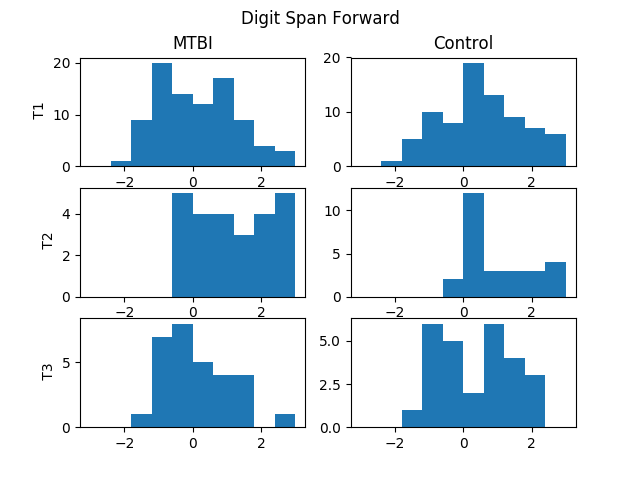
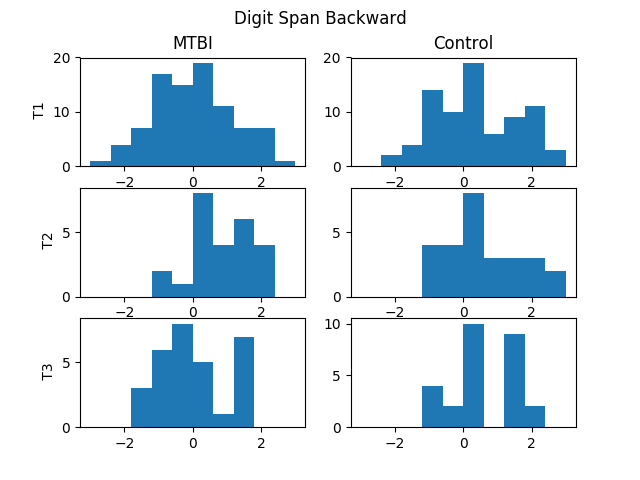
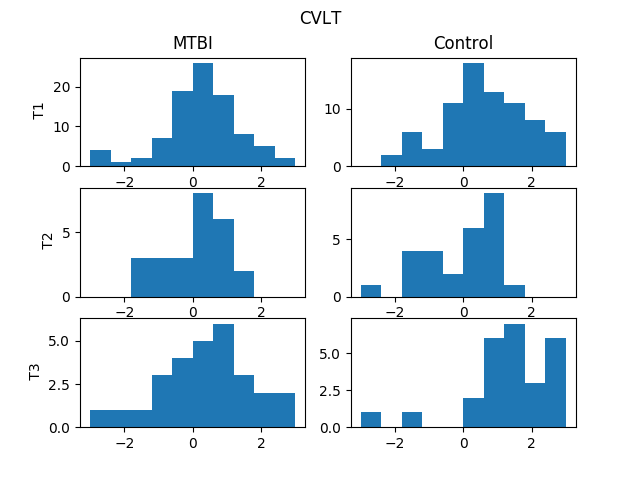
1. The R2 values are obviously very low, I will try the exact same hybrid feature reduction algorithm that Shervin has used in his 2013 paper.
2. I believe the best metric for T2 will be the T1 scores. So I am planning on using T1 scores for T2 prediction as well as MR metrics for the next task.
3. I will add 2nd and 3rd moments in addition to mean and variance for the training.
4. I have splitted left and right thalamus thinking there could be some asymmetries present that the algorithm can detect. However algorithm never used the same metric from both the left and right thalamus at the same time. So I wonder if it is a better idea to use the thalamus as a whole
5. I have used 3 additional metrics compared to the previous MTBI classification (Tort, AD, RD) simply because the data was available. Does it make more sense to lose them?
6. I have also used every metric on thalamus region but in the previous work we have only used AK, FA, MD, MK, RK in the thalamus. Rest of the metrics were CC only. Should I switch to the CC only model?
7. Likewise we have focused only on the splenium part of the CC in the previous work. Do you think it is better to use only genu and body?

**Meeting Notes July 23:**

* Get white matter data and mask from Hugh (old/new)
* Export a merged data across cycles (you can make a histogram to check if two cycles have similar distributions)
* Other tests (like digit span)
* Cluster/BoW the entire white matter over different sub-regions and consider entire white matter overall
* Try to find locations from correlation then use these for prediction
* Read Sohae’s paper on white matter region correlation with NP tests
* Make sure if your data is subject space or template space
* Wed August 1. 12.15

**July 28**

The data has been merged across all time points and NP score histograms are generated for easier visualization.



**August 22**

Currently we have 36 regional masks and 4 white matter metrics (AWF, Depar, Deperp, DA). I have limited my dataset into these metrics as these were the metrics that were used for the white matter analysis in Sohae’s paper. The data is in template space.

Using these metric and regions I have constructed two separate feature sets: BoW Entire White Matter and Statistical Features for Each Region.

**3D BoW Entire White Matter:**

In this approach I have considered the union of the 36 white matter regions and made a 3D bag-of-words of 20 bins for each metric, resulting in 80 features for each subject. The BoW were constructed without separating the control and MTBI patients.

**Regional Statistical Features:**

For each region and each metric following statistical features are obtained: mean, std, 2nd, 3rd and 4th moments. So the feature space size is 4\*5\*36 = 720.

**Regression and Feature Selection**

I have used linear regression instead of SVR for these datasets as it runs slightly faster and the feature size was significantly larger in this case. Greedy forward selection algorithm was used to reduce dimensionality as before and T1 test scores were predicted on the digit span, CVLT, letter-number and RCFT tests. However for the Regional Statistical Features size was very massive so I have limited my work to only the letter-number scores for regional statistics.

The dataset was not separated during regression. MTBI and control samples were trained together to obtain a single regression model.

Out of the 154 subjects for T1 data points I have used 25 for the validation set and shuffled the validation set 30 times for BoW method. Due to massive feature set size of regional statistical features instead of shuffling the validation dataset 30 times I have used 10 folds cross-validation.

Table 1: Validation R^2 scores averaged over different shufflings.

|  |  |  |
| --- | --- | --- |
|  | Entire WM BoW | Regional Statistics |
| CVLT | 0.53 | - |
| RCFT | 0.47 | - |
| Letter-num | 0.61 | 0.27 |
| Digit Span (backward) | 0.57 | - |

Table 2: Validation MSE scores averaged over different shufflings.

|  |  |  |
| --- | --- | --- |
|  | Entire WM BoW | Regional Statistics |
| CVLT | 0.46 | - |
| RCFT | 0.62 | - |
| Letter-num | 0.38 | 0.89 |
| Digit Span (backward) | 0.39 | - |

**Selected Regions for Regional Statistics:**

|  |
| --- |
| **Genu of corpus callosum std**  **Retrolenticular part of internal capsule R 2nd moment**  **Body of corpus callosum mean**  **Posterior corona radiata R mean**  **Sagittal stratum (include inferior longitidinal fasciculus and inferior fronto-occipital fasciculus) R 3rd moment**  **Posterior limb of internal capsule L std**  **Fornix (cres) / Stria terminalis (can not be resolved with current resolution) L std**  **Splenium of corpus callosum 2nd moment**  **Posterior limb of internal capsule R 3rd moment**  **Body of corpus callosum std** |

**Remarks:**

* Entire BoW significantly outperforms other models but I believe it would yield even better results if different BoW regions were used but not as much as 36. I am trying to come up with a way to merge some of these 36 regions (or break-up the entire WM region) into sub-regions. Maybe 8-10 regions in total can still be manageable with BoW. I could simply merge the nearest regions but I am trying to come up with a way to merge in a statistically significant manner. After merging all left-right components I am still left with 20 regions.
* During the BoW algorithm would it be better if I tried to identify the most significant words instead of k-means clustering the data to come up with the words?

**August 22, 2018 Meeting**

* Clustering code is not correct, check to see what was actually calculated in the code
* Draw histogram of each statistical feature
* Use entropy statistical feature as well (check Shervin’s paper again for all statistical features)
* Quantize the score and draw the different statistical features between the groups to identify which features are effective
* FA, MD, MK, AWF, AK, Depar, Deperp, DA are the metrics to be used
* Try statistical feature extraction from the bigger regions
* If we are to use BoW start with patch size 5x5x5 and cluster the regions into sub-regions
* Look at different clustering algorithms to see if you can find outliers betters (not detecting generic classes, automatic cluster size, region detection )
* Do seperate training for MTBI and control (focus on control subjects first)
* Focus only on letter-num. Then add digit span (both backward and forward) as 2nd option for now