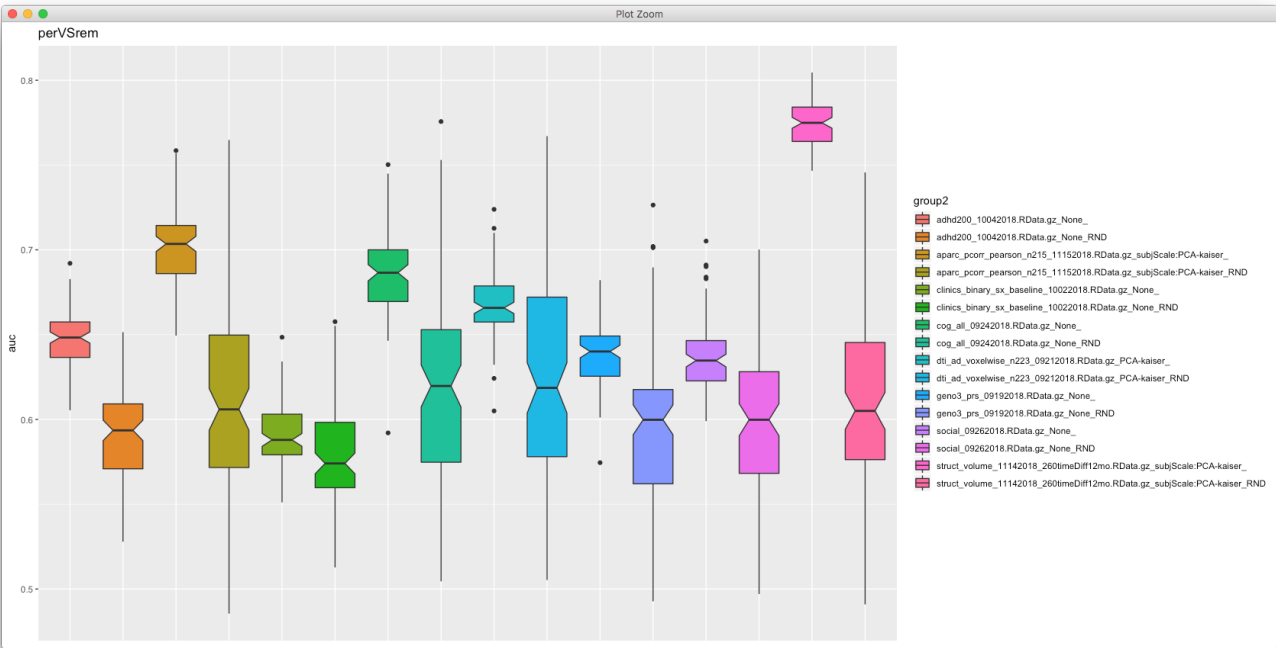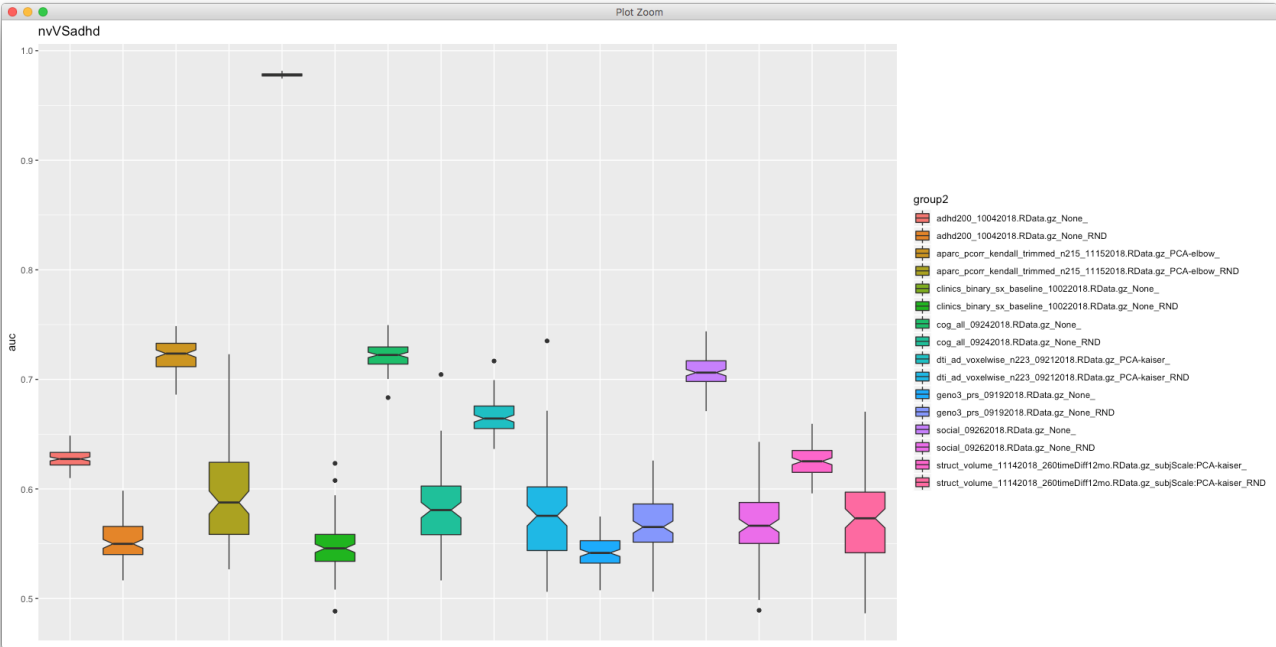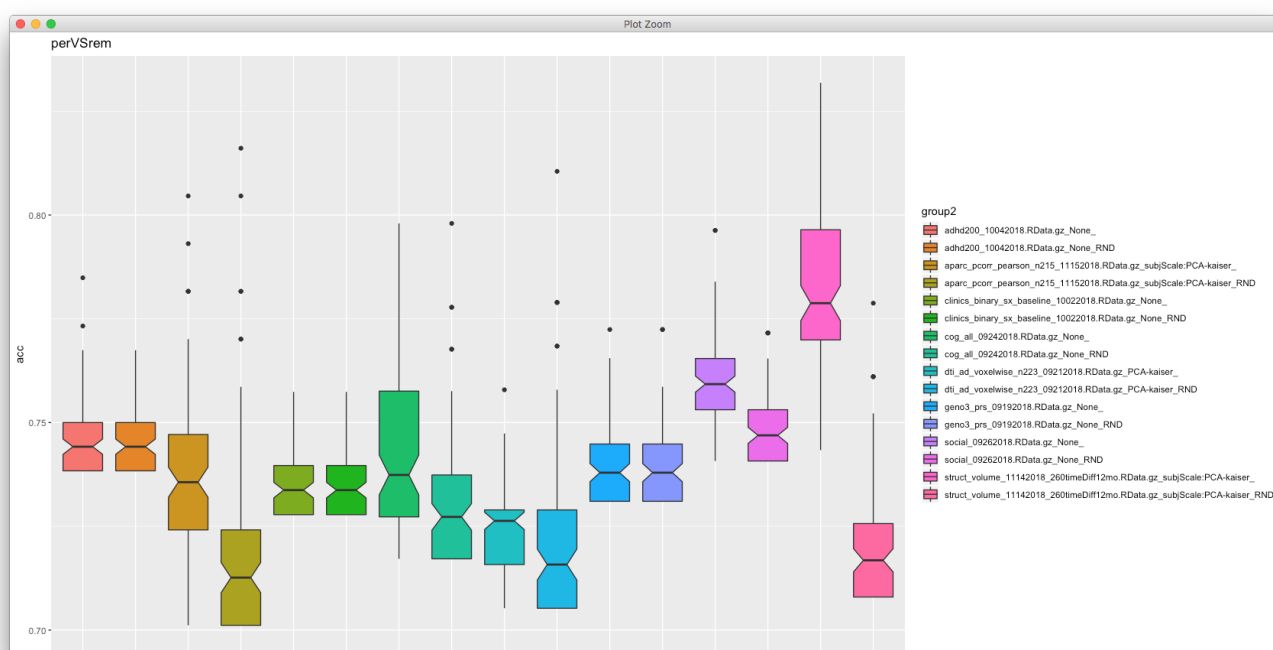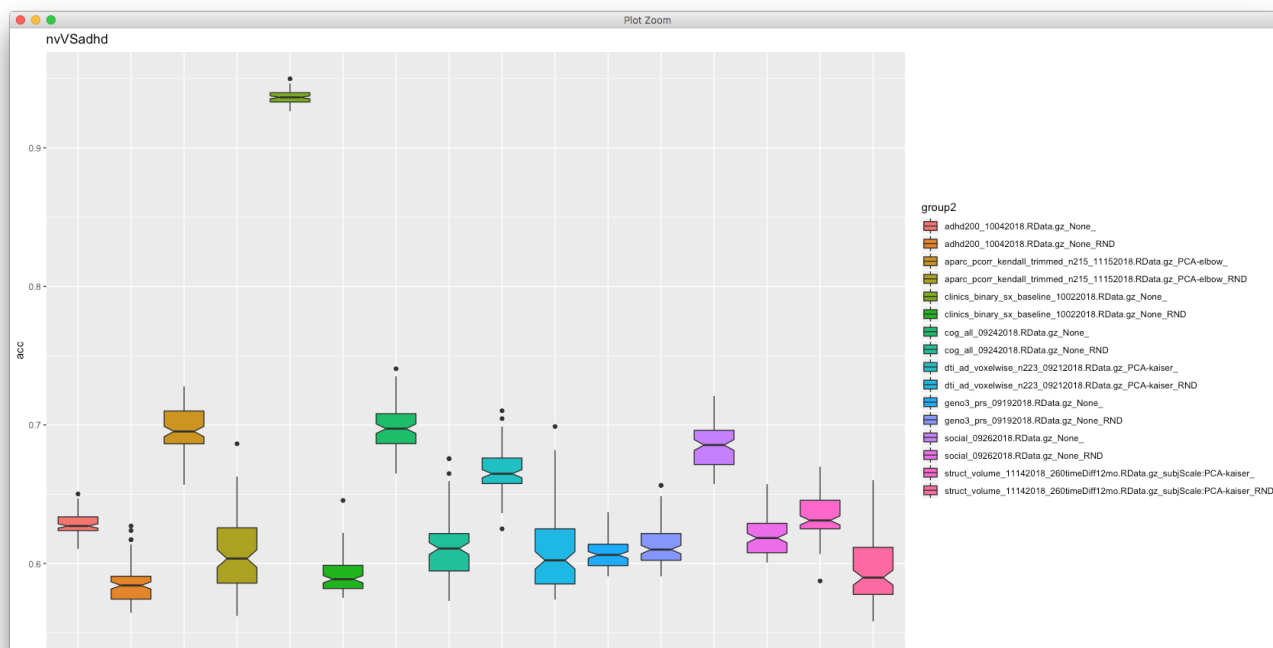# 2019-06-17 13:33:02

The overall goal is to correctly predict ADHD persistence or remission in a future date based on baseline data.

Our best results can be summarized on these AUC pictures:

And now we do accuracy as well:

Each pair of consecutive bars in the plots show a data domain and its corresponding results using random data. By random data, it's just data generated from a uniform distribution using feature-specific max and min as boundaries. That was just a benchmark to check against overfitting, especially when using some of the more complex algorithms.

I'm not sure how familiar you are with the types of data we're using, but here's a quick legend including the total N in each dataset for persistent vs remission, then (NVs + ADHDs), and the number of variables:

- adhd200_10042018: Variables used in the ADHD-200 challenge (Sex, age, Handedness, IQ): N=219 (380); M=4
- aparc_pcorr_pearson_n215_11152018: rsFMRI: N=118 (215); M=2278
- clinics_binary_sx_baseline_10022018: Binary symptoms: N=215 (375); M=20

- cog_all_09242018: Neuropsychiatric tests: N=133 (234); M=25
- dti_ad_voxelwise_n223_09212018: DTI-AD: N=127 (223); M=12106
- geno3_prs_09192018: Polygenic Risk Score: N=186 (328); M=13
- social_09262018: Socio-economic: N=206 (355), M=18
- struct_volume_11142018_260timeDiff12mo: Cortical volume: N=149 (260); M=5124

We can see that in most cases we do considerably better using the actual data, which is good. In those plots I'm only showing the results for persistent versus remitted classes, and also for normal volunteers (NVs) versus ADHDs (persistent pus remission). The former throws away almost half of the data (i.e. all the NVs), and it's just one of the many questions we can ask. Including the NVs (3-class classification) makes the problem harder, but it could be useful to answer some biological questions (e.g. does remission look more like NVs? or is it a completely different process?)

A few extra observations on these results:

- it makes sense that the baseline binary symptoms would have nearly perfect nvVSadhd classification, but poor perVSrem. After all, ADHD diagnosis is simply based on adding those binary symptoms, but whether those symptoms will persist in the future cannot be determined by that alone.
- we do get some decent AUC values for nvVSadhd besides that, such as rsFMRI (aparc), neuropsych (cog), and social/demographic metrics (social), all above .7
- for perVSrem we also get some interesting results, especially for rsFMRI and volume.

You can also see in the plot that each predicion is done within-data domain. For example, predicitons use fMRI data only, or DTI data only. Anything I tried across domain, either using imputation or methods that can deal with missing data, didn't perform as well.

Methods

I used H2O's autoML methods for this work, using the R API. I played with carret and scikit-learn as well, which I'm actually more familiar with. But at the time H2O's Java implementation was making better use of all the computation power I could allocate in Biowulf, so I went with that.

I started by using their autoML framework, but not after long it was clear that DeepLearning was doing better than all other algorithms in every category, so I stuck with that. No stacked methods worked better, though.

All tests were done using a cross-validation framework, with 90% of data for training and validation, and 10% for testing. There was a 5-fold cross validation, repeated multiple times, in the train+validation set.

Targets

One of the many challenges in this project is to define the targets for prediction. I showed the best results we got in the 2-group classification, but one could also consider a 3-group classification (including NVs).

Another way I tried was to predict symptom change rate, which is a more continuous metric. There, one needs to be careful because a rate of 0 is different (?) between NVs that don't change symptoms and ADHDs that remain symptomatic. However, the results weren't as good.

Finally, we also explored predicting latent classes, created by looking just at the symptom trajectories of each subject. Ideally this would be done in a cross-validation framework as well. But in any case, we could classify

each subject in one of 3 classes (e.g. improvers, deterioration, stable) and just use those classes as targets for the other datasets.

Data processing

There are always many things one can do to the data in order to make features better. Especially in the neuroimaging side, different pre-processing pipelines will generate different numbers of features, of varying quality as well. As they say, gargage-in, garbage-out. We can also tweak our quality control thresholds to include more data at the risk of including more noise (especially due to movement, as it's neuroimaging data, in kids, who have ADHD). If we go in that direciton, we'd need to be careful to show that any results are not just picking up movement...

PCA or other dimensionality reduction techniques, compared to simply downsampling the data, sometimes helped as well.

The results I presented are the best ones we got so far, but we are currently working on new DTI and rsFMRI preprocessing pipelines that might yield better prediction results.

Going forward

I keep all my code in Github (gsudre) and my lab notes as well, in Markdown format. For this project I didn't use Jupyter notebooks, but I'm open to going back to it. I'm also open to playing with other tools to make collaboration for this project a bit easier.