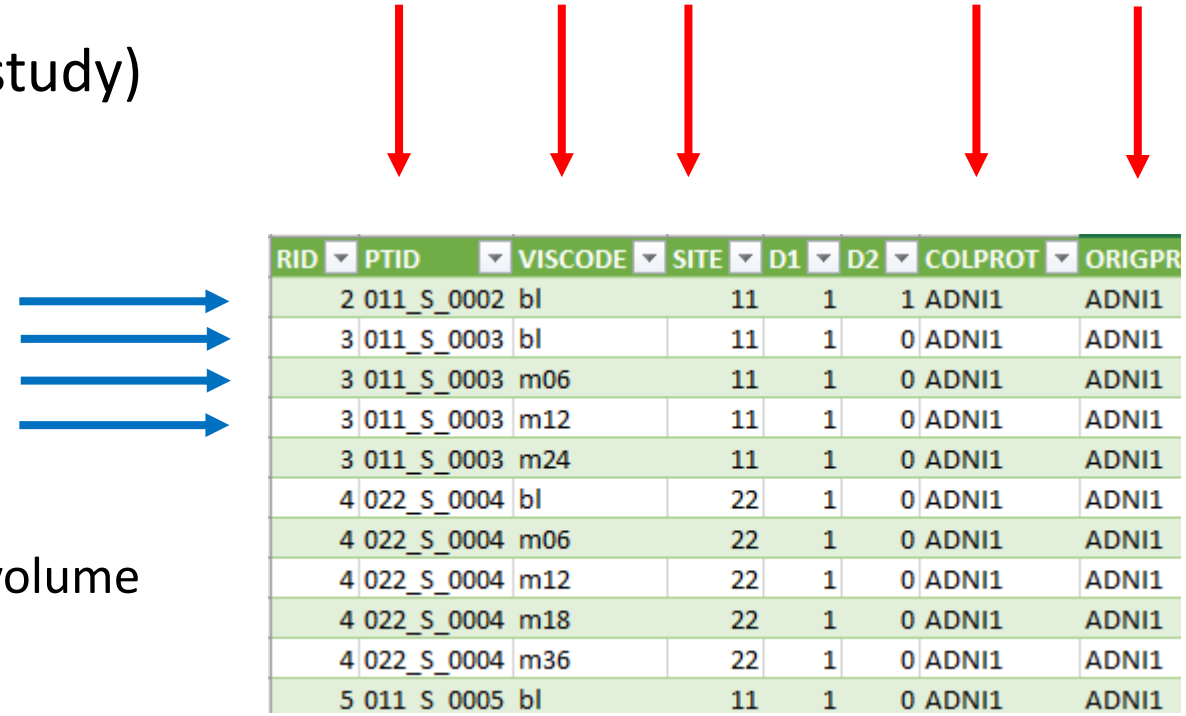
BrainHack 2019

#Dementia_Forecast_Group

23/03/2019

Understanding the data!!

- 1737 patients
- 12741 instances (longitudinal study)
- 1906 features
- Goal:
  - Diagnosis
  - ADAS-13
  - Ventricles volume/intracranial volume



RID	PTID	VISCODE	SITE	D1	D2	COLPROT	ORIGPR
2	011_S_0002	bl	11	1	1	ADNI1	ADNI1
3	011_S_0003	bl	11	1	0	ADNI1	ADNI1
3	011_S_0003	m06	11	1	0	ADNI1	ADNI1
3	011_S_0003	m12	11	1	0	ADNI1	ADNI1
3	011_S_0003	m24	11	1	0	ADNI1	ADNI1
4	022_S_0004	bl	22	1	0	ADNI1	ADNI1
4	022_S_0004	m06	22	1	0	ADNI1	ADNI1
4	022_S_0004	m12	22	1	0	ADNI1	ADNI1
4	022_S_0004	m18	22	1	0	ADNI1	ADNI1
4	022_S_0004	m36	22	1	0	ADNI1	ADNI1
5	011_S_0005	bl	11	1	0	ADNI1	ADNI1

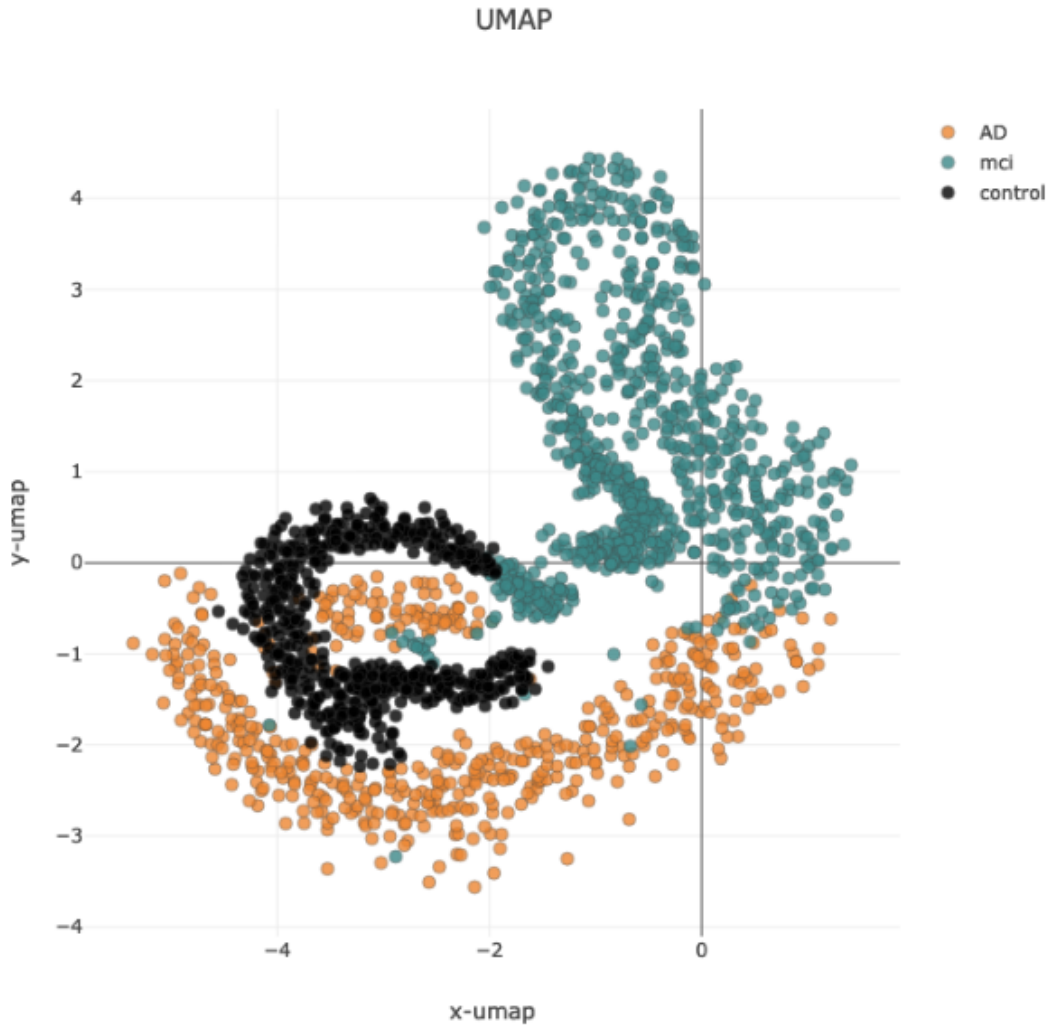
Group	Var	description
Demo	AGE	Age at BASELINE, not current age
	PTGENDER	Sex
Genotype	APOE4	APOE4 minor allele load
sMRI	ST*NNN*CV_*	Regional Volume
	ST*NNN*SA_*	Regional surface area
	ST*NNN*TA_*	Regional thickness average
	ST*NNN*SD_*	Regional thickness standard deviation
PET	*UCBERKELEYAV1451*	Regional PET measures of abnormal Tau
	UCBERKELEYAV45	Regional PET measures of abnormal Abeta
	BAIPETNMRC	Regional PET measures of glucose metabolism
	FDG / PIB / AV45	Average relative PET measures
DTI	FA_*	Regional fractional anisotropy
	MD_*	Regional mean diffusivity
	AD_*	Regional axial diffusivity
	RD_*	Regional radial diffusivity
CSF	ABETA_*	CSF Amyloid Beta
	TAU_*	CSF Tau
	PTAU_*	CSF phospho-Tau

!

Data handling

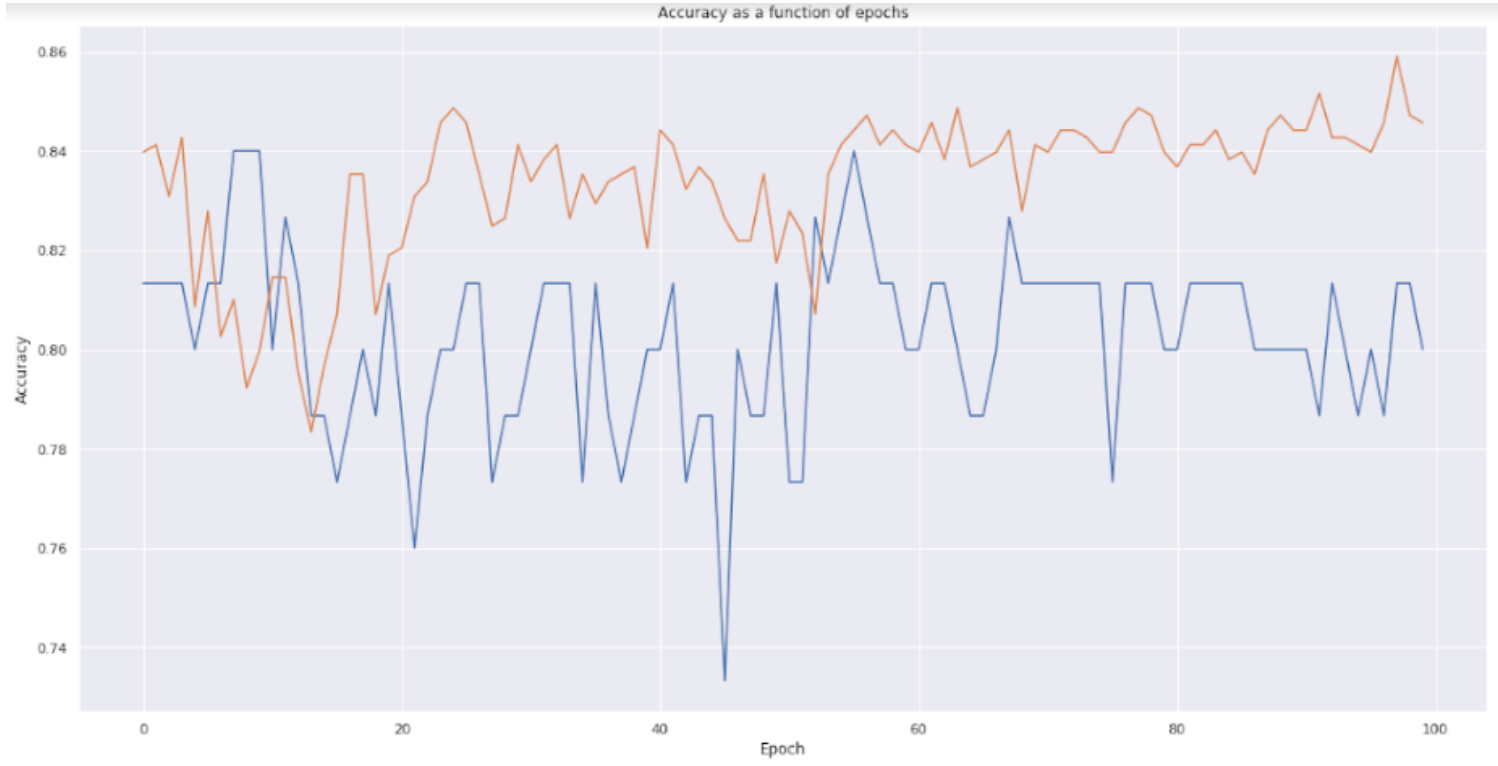
- **Training data**: up to April, 2010
- **Test data:** from April 2010, onwards
- **Missing values**: e.g. for structural MRI:
  - Total = 442 features
  - 19 features not available (unavailable for all patients)
  - 225 incomplete features (available only for some patients)

Dimensionality reduction



Training a Neural Net

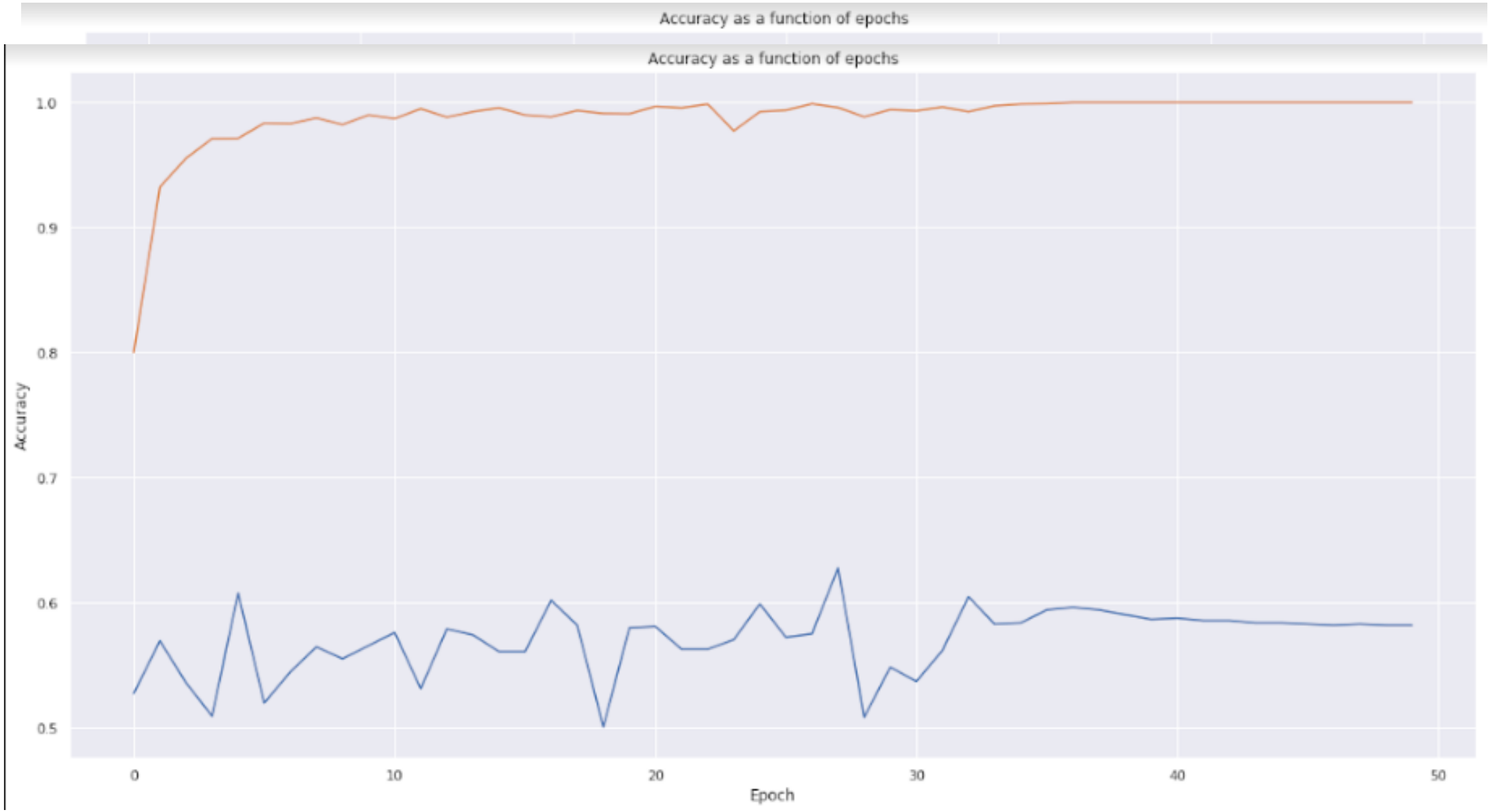
- Trying to predict the Umap embeddings using a neural network on the test set



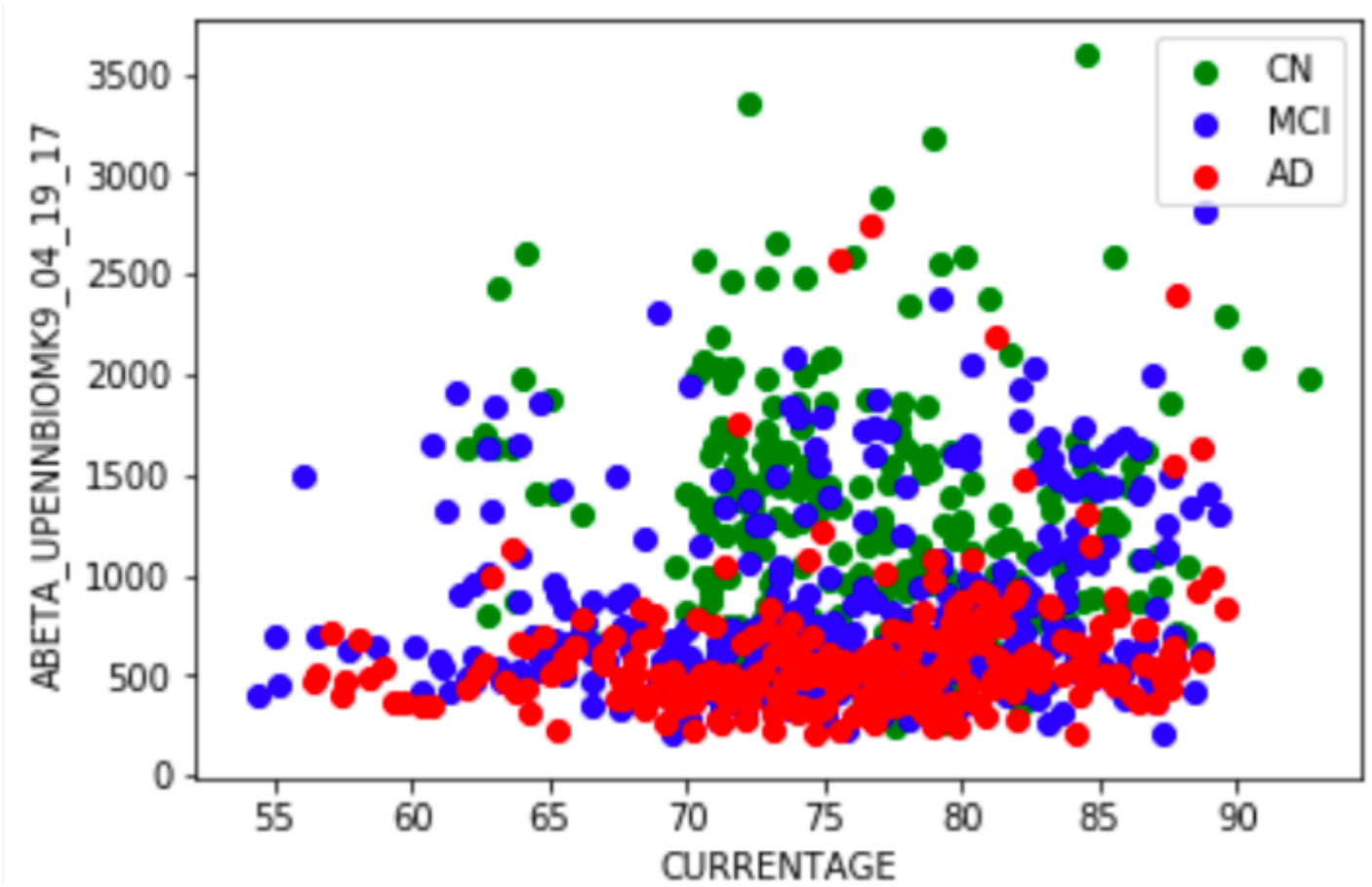
Trying out classifying the diagnosis

- Using **only** CSF features with FFNN: overall diagnosis accuracy 57%

- Using the template SVM file provided by challenge organizers:
  - Be careful of dropping the feature to classify!
  - Code not optimized
  - Recommend writing one's own code


Accuracy as a function of epochs
Accuracy as a function of epochs
1.0
0.9
0.8
0.7
0.6
0.5
Accuracy
0
10
20
30
40
50
Epoch

Example of variable distribution



Predictions with and without age extrapolation

- Classification: random forest
- Regression: SVR

```
AGE
ADAS13
Ventricles
ICV_bl
ST44CV_UCSFFSL_02_01_16_UCSFFSL51ALL_08_01_16
ST105CV_UCSFFSL_02_01_16_UCSFFSL51ALL_08_01_16
ST121TS_UCSFFSL_02_01_16_UCSFFSL51ALL_08_01_16
ABETA_UPENNBIOMK9_04_19_17
TAU_UPENNBIOMK9_04_19_17
PTAU_UPENNBIOMK9_04_19_17
CURRENTAGE
Ventricles_ICV
```

with

```
(9240,)
Evaluate predictions
Diagnosis:
mAUC = 0.679
BAC = 0.621
ADAS:
MAE = 10.859
WES = 9.190
CPA = 0.449
VENTS:
MAE = 8.109e-03
WES = 8.023e-03
CPA = 0.500
```

without

```
mAUC = 0.665
BAC = 0.624
ADAS:
MAE = 8.171
WES = 8.217
CPA = 0.431
VENTS:
MAE = 7.058e-03
WES = 6.898e-03
CPA = 0.500
```

Trying out XGBOOST

Explore hyperparameter optimization with boosting