UPDATES

July 11, 2024

Notes

- final report removed, contacting staff
- qualitative data is ordinal now - better results

	index	aquifer_vulnerability	drainage_class	gus	tgus	soil_hal...	pcoef	offset	detected
0	0	0	0	-0.143	0.122	1.953	-0.02	1	-1
1	1	0	0	-0.465	0.553	2.402	-0.163	1	-1
2	2	0	0	-0.143	0.122	1.953	-0.02	1	-1
3	3	0	0	-0.465	0.553	2.402	-0.163	1	-1
4	4	0	0	-0.143	0.122	1.953	-0.02	1	-1
5	5	0	0	-0.465	0.553	2.402	-0.163	1	-1
6	6	2	0	-1.206	-0.737	-0.598	2.97	1	-1
7	7	2	0	-0.474	-0.831	2.058	1.346	1	-1
8	8	2	0	-1.206	-0.737	-0.598	2.97	1	-1

Old dataset, with detection limit, simple TGUS equation (3.4 assumption)

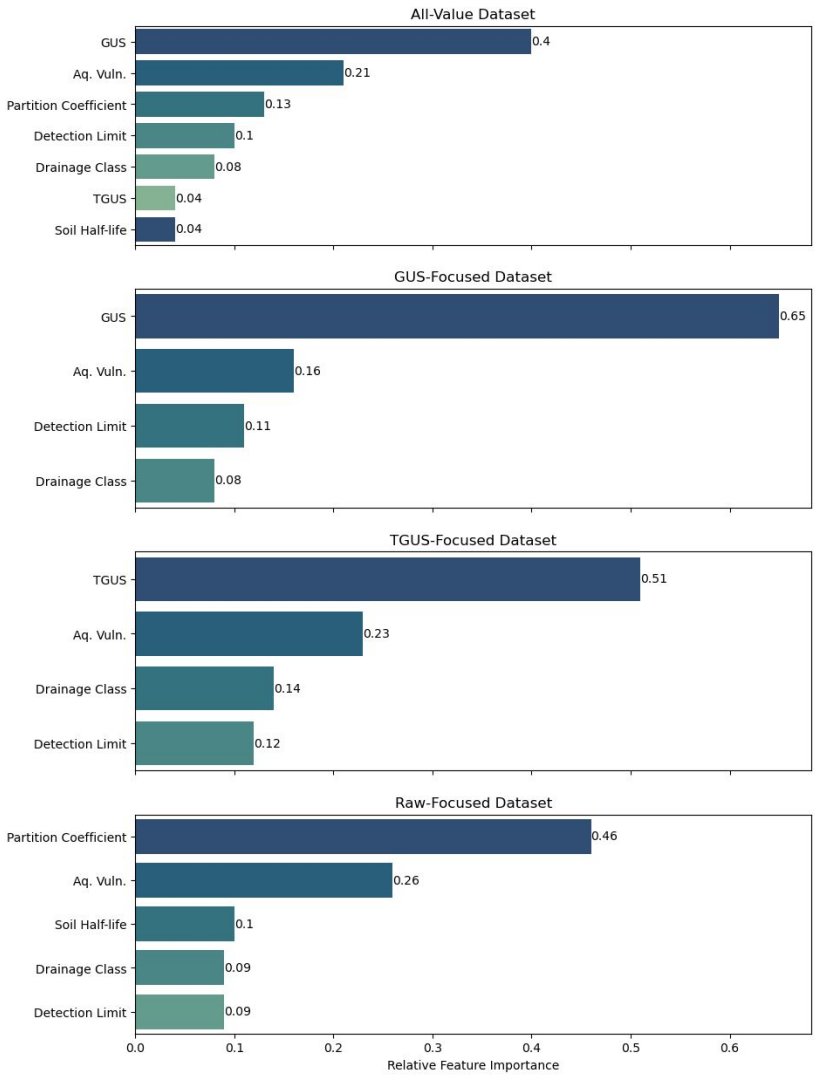
Results - binary classification accuracy (detected or nondetected)

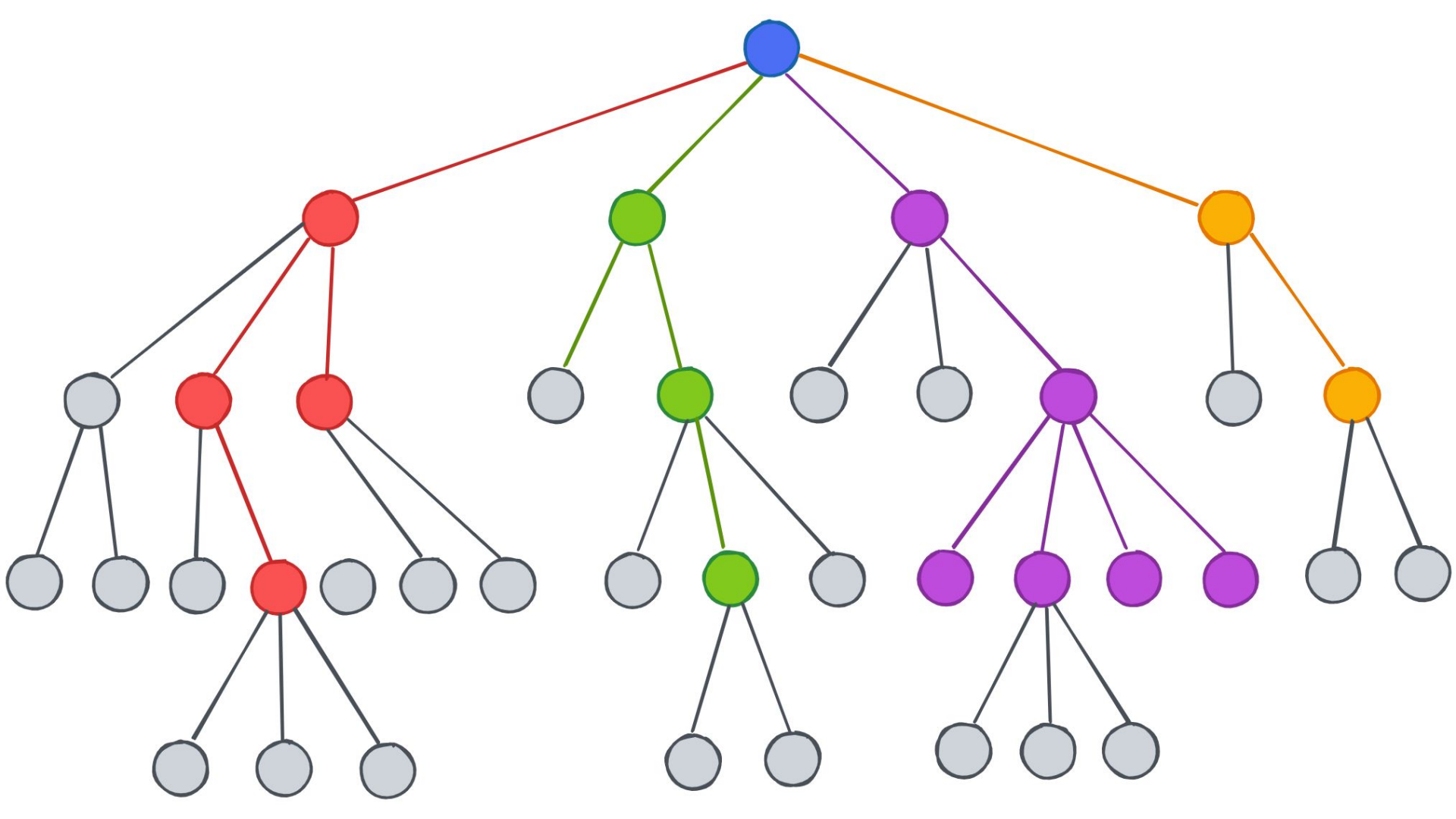
	Avg. Train %	Avg. Validation %	Avg. Test %	Best Sc. Train %	Best Sc. Validation %	Best Sc. Test %
All	98.3	99.2	96.7	98	100	99.1
GUS Focus	98.3	99.1	96.2	98	98.6	100
TGUS Focus	98.3	98.7	95.9	98	98.6	99.1
Raw Focus	98.2	99	96.5	98	97.2	100

- no significant differences in t-tests

Feature Importances

- averaged over 50 iterations -> consistent pattern
- GUS vs TGUS
  - correlation with aq. vuln. & drainage class
  - detection limit adds more info consistently
  - **more complicated TGUS could be super powerful predictor**
- All-value dataset
  - partition coefficient still valuable alone
  - not better than any other model…redundancy
- Raw-focused dataset
  - soil half-life outweighed by partition coefficient





Old dataset, without detection limit, simple TGUS equation (3.4 assumption)

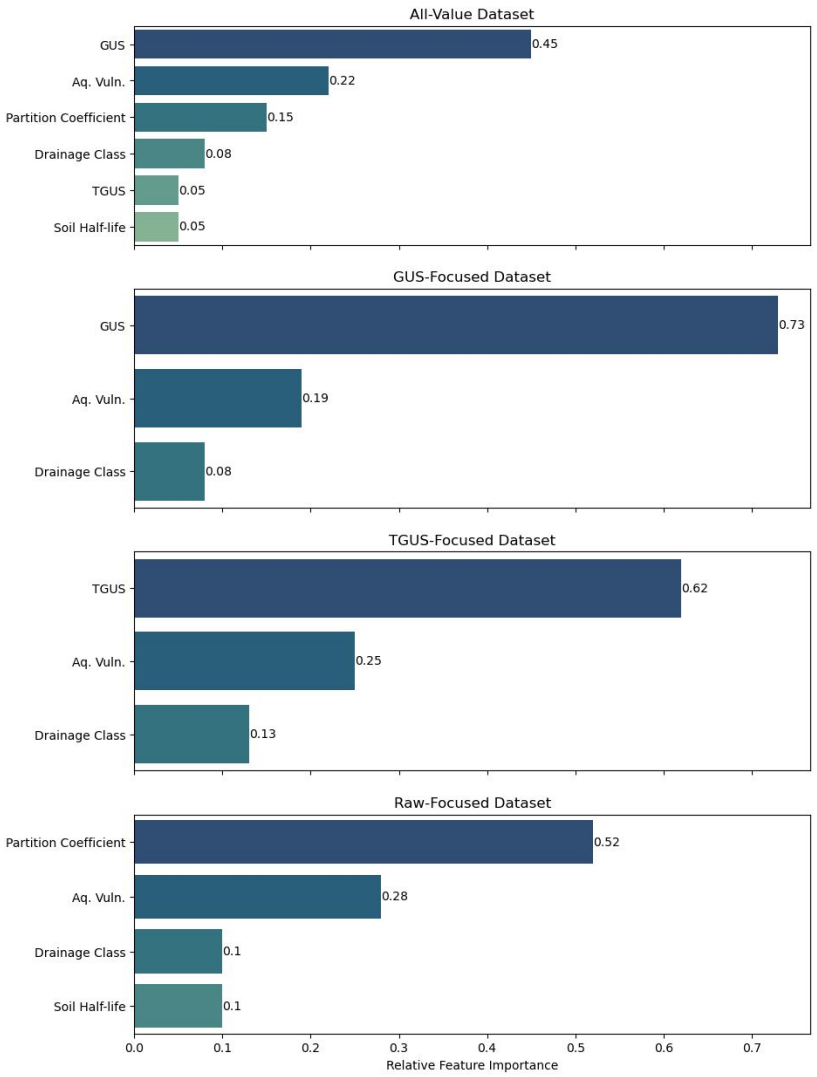
Results - binary classification accuracy (detected or nondetected)

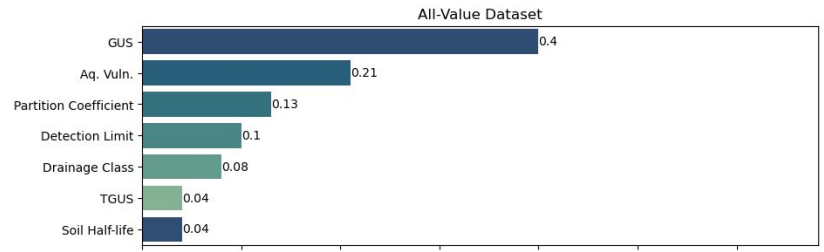
	Avg. Train %	Avg. Validation %	Avg. Test %	Best Sc. Train %	Best Sc. Validation %	Best Sc. Test %
All	96.5	97.2	94.2	96	95.7	98.1
GUS Focus	96.5	97.6	94.7	96.6	97.2	98.1
TGUS Focus	96.6	97.5	93.7	95.7	97.2	98.1
Raw Focus	96.5	97.6	94.2	95.7	97.2	98.1

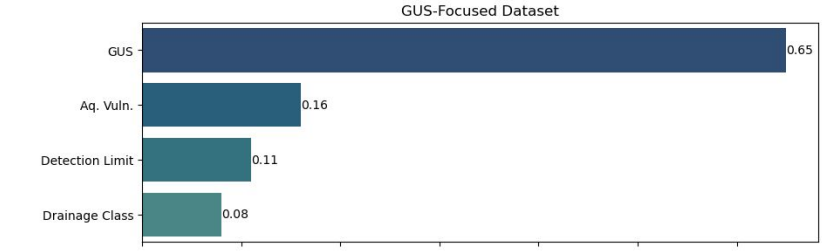
- no significant differences in t-tests
- slight decrease in performance

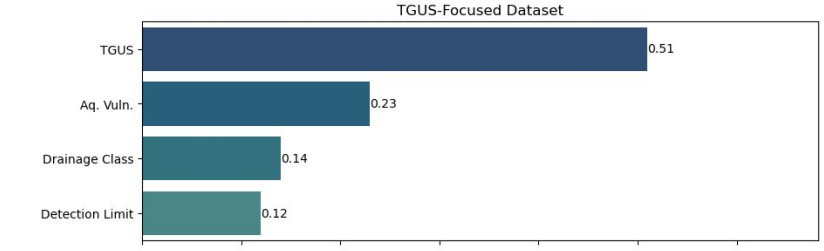
Feature Importances

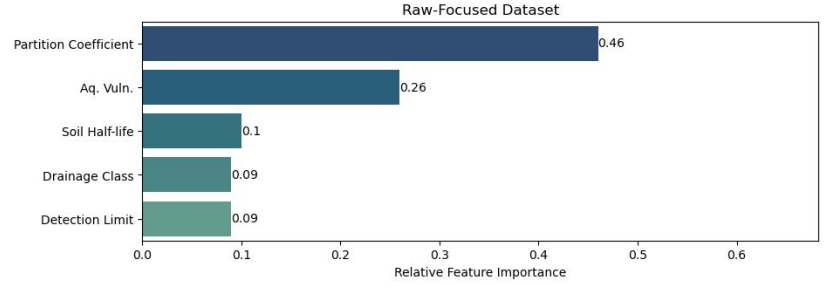
- averaged over 50 iterations -> consistent pattern
- similar patterns as with detection limit

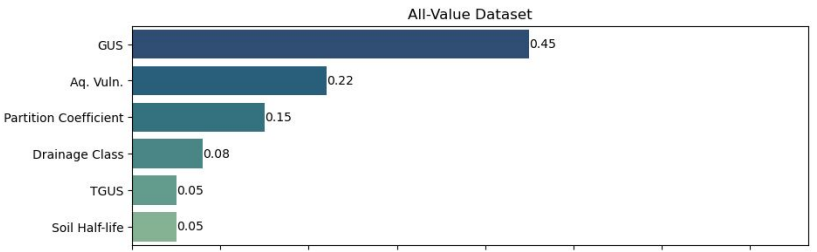


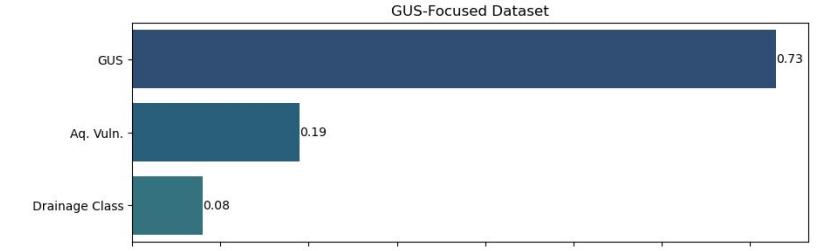

All-Value Dataset
GUS
Aq. Vuln.
Partition Coefficient
Detection Limit
Drainage Class
TGUS
Soil Half-life
0.4
0.21
0.13
0.1
0.08
0.04
0.04

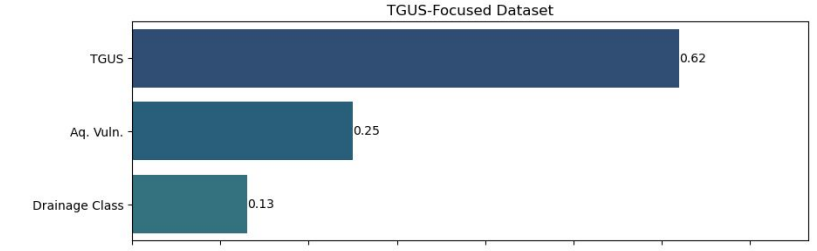

GUS-Focused Dataset
GUS
Aq. Vuln.
Detection Limit
Drainage Class
0.65
0.16
0.11
0.08

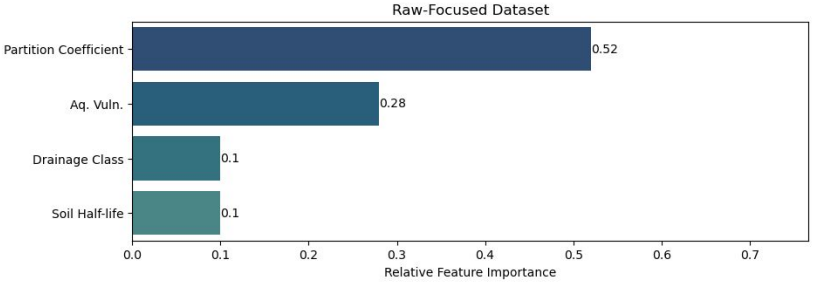

TGUS-Focused Dataset
TGUS
Aq. Vuln.
Drainage Class
Detection Limit
0.51
0.23
0.14
0.12


Raw-Focused Dataset
Partition Coefficient
Aq. Vuln.
Soil Half-life
Drainage Class
Detection Limit
0.46
0.26
0.1
0.09
0.09
0.0
0.1
0.2
0.3
0.4
0.5
0.6
Relative Feature Importance


All-Value Dataset
GUS
Aq. Vuln.
Partition Coefficient
Drainage Class
TGUS
Soil Half-life
0.45
0.22
0.15
0.08
0.05
0.05


GUS-Focused Dataset
GUS
Aq. Vuln.
Drainage Class
0.73
0.19
0.08


TGUS-Focused Dataset
TGUS
Aq. Vuln.
Drainage Class
0.62
0.25
0.13


Raw-Focused Dataset
Partition Coefficient
Aq. Vuln.
Drainage Class
Soil Half-life
0.52
0.28
0.1
0.1
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
Relative Feature Importance

Moving Forward

- more feature importance research
  - more robust methods
  - different combinations to see overlaps
- fine-tune function arguments
- fix TGUS equation
- new data