Lithofacies classification using machine learning

Xin Zhao

Problem statement

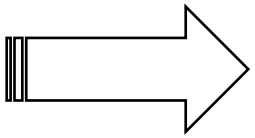
- An accurate identification of rock class is critical in oil and gas industry

- This is a **Classification** problem:
  - Target: rock facies
  - Features: log measurements

Data used for this research

- Real well logs from GitHub

Features	
1	Gamma ray (GR)
2	Resistivity (ILD_log10)
3	Photoelectric effect (PE)
4	Neutron-density porosity difference (DeltaPHI)
5	Average neutron-density porosity (PHIND)
6	Non-marine/marine indicator (NM_M)
7	Relative position (RELPOS)

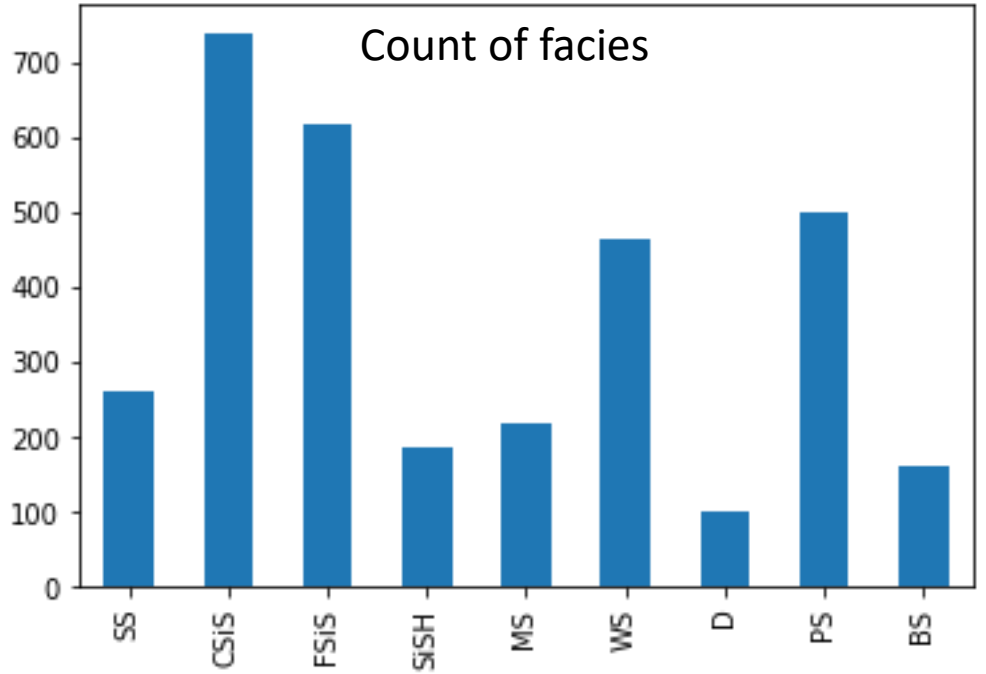


Target			
Facies	**Rock class**	**Label**	**Adjacent Facies**
1	Non-marine sandstone	SS	2
2	Non-marine coarse siltstone	Csis	1, 3
3	Non-marine fine siltstone	Fsis	2
4	Marine siltstone and shale	SiSh	5
5	Mudstone	MS	4, 6
6	Wackestone	WS	5, 7, 8
7	Dolomite	D	6, 8
8	Packstone-grainstone	PS	6, 7, 9
9	Phylloid-algal bafflestone	BS	7, 8

Supervised learning experiment

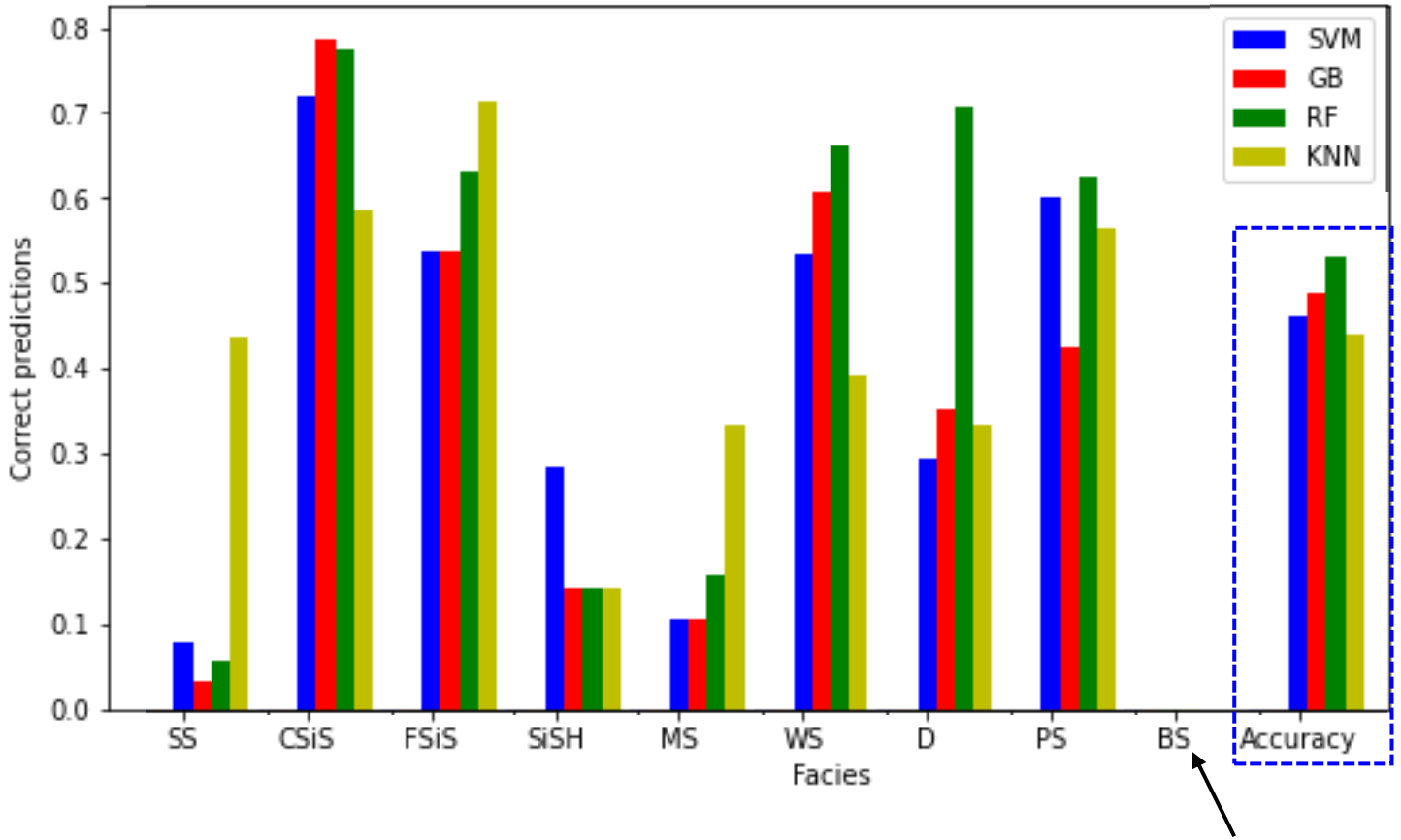
- Total 9 wells:
  - 8 → modelling
  - 1 → blind test

- Classification algorithms:
  - Support vector machine (SVM)
  - Gradient boosting (GB)
  - Random forest (RM)
  - Nearest neighbors (KNN)



- Significance in difference of prediction accuracy
  1. Repeat each algorithm 100 times using randomly split training data
  2. Get the 100 F1 scores of prediction on the blind data
  3. T-test
  4. Choose the best model

Algorithm evaluation



All four algorithms did not predict BS, which is true for the blind test well.

P-values

	RF
GB	1.3e-14
KNN	1.3e-92
SVM	0.09

There is **95%** confidence to accept the difference in accuracy between **RF** and **GB/KNN** is **significant**.

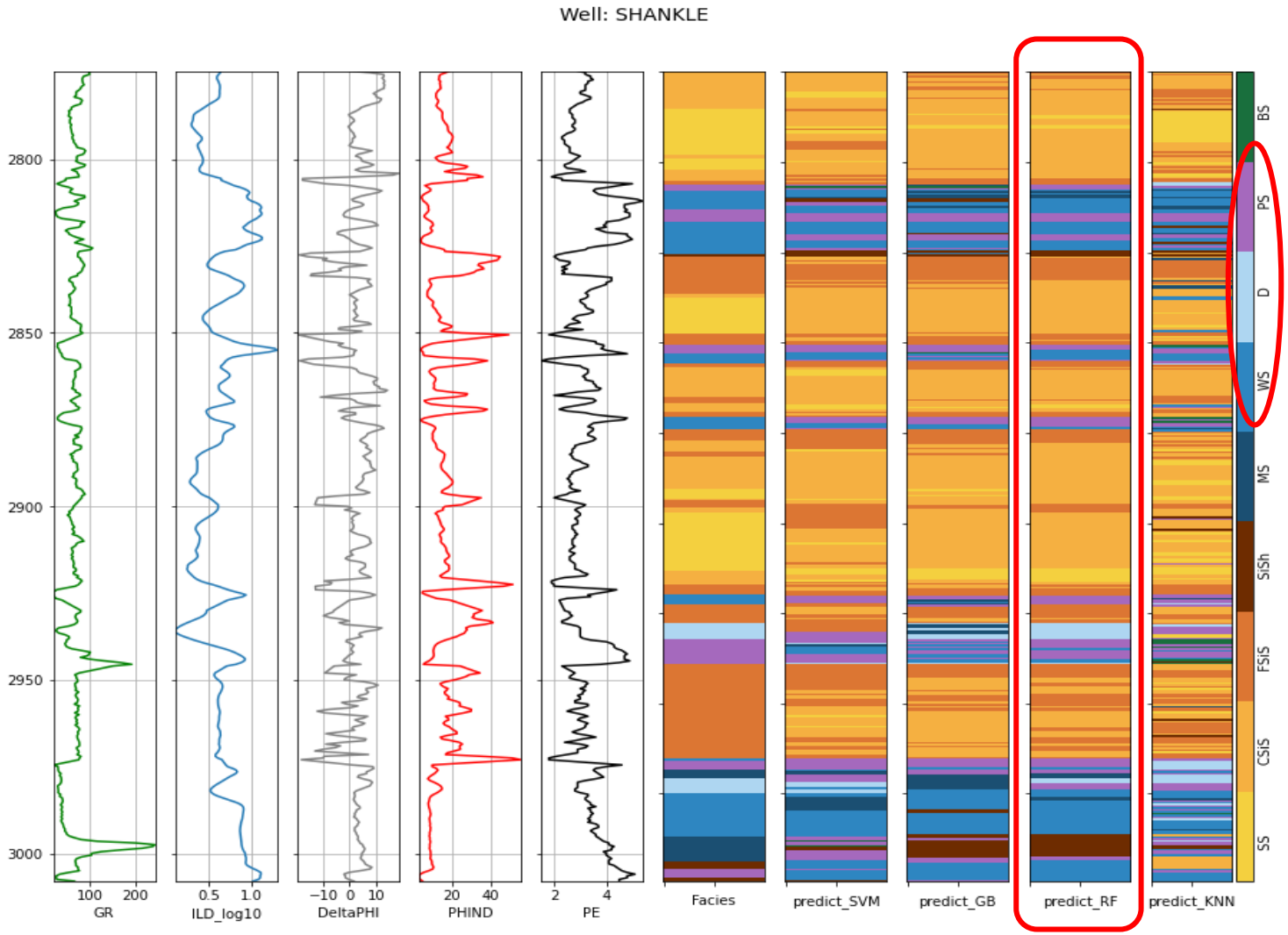
RF is better than SVM

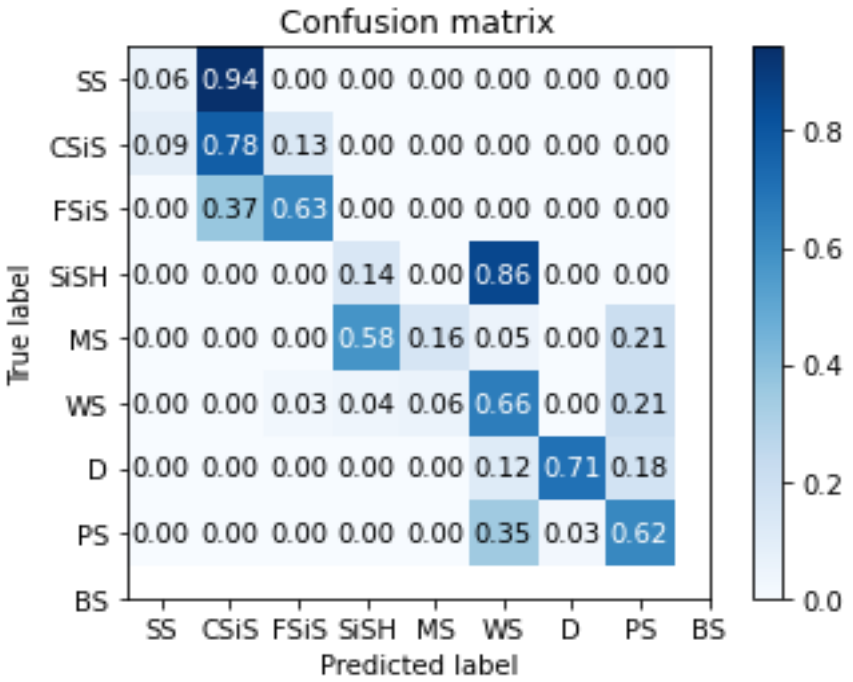
RF				
Facies	precision	recall	f1-score	support
SS	0.38	0.06	0.1	89
CSiS	0.36	0.79	0.49	89
FSiS	0.85	0.63	0.73	117
SiSH	0.07	0.14	0.09	7
MS	0.38	0.16	0.22	19
WS	0.67	0.65	0.66	71
D	0.92	0.71	0.8	17
PS	0.53	0.62	0.57	40
avg	0.57	0.53	0.5	449

SVM				
Facies	precision	recall	f1-score	support
SS	0.3	0.08	0.12	89
CSiS	0.36	0.72	0.48	89
FSiS	0.62	0.54	0.58	117
SiSH	0.25	0.29	0.27	7
MS	0.17	0.11	0.13	19
WS	0.66	0.54	0.59	71
D	0.71	0.29	0.42	17
PS	0.41	0.6	0.49	40
avg	0.47	0.46	0.43	449

RF has overall better precision in prediction than SVM – 6.5:1.5
RF has overall better recall in prediction than SVM – 6:2

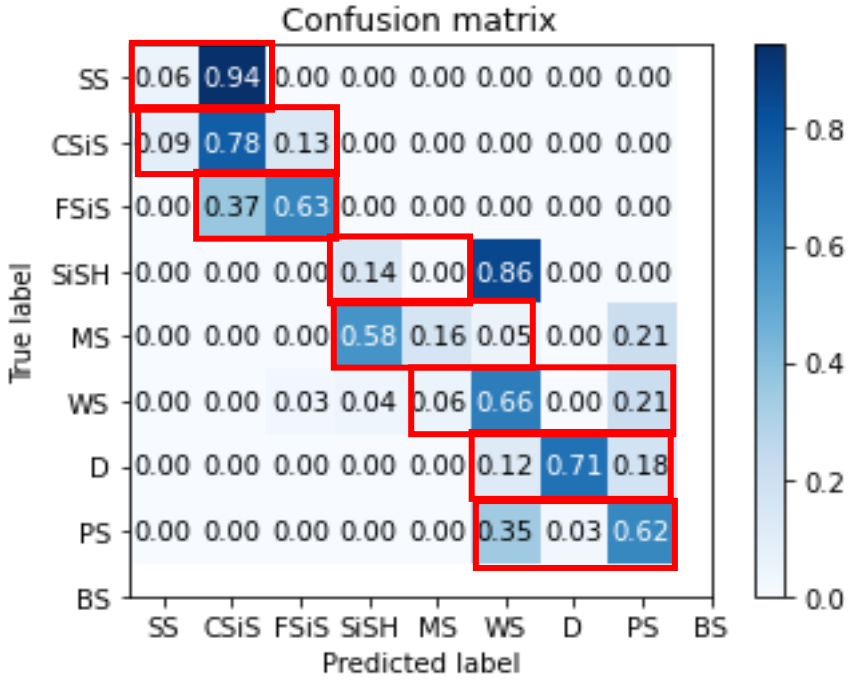
Random forest is the most optimal algorithm





Lithofacies classification accuracy

- Current best prediction is from random forest, the F1 score is ***0.53***
- The F1 score ***0.87*** if considering the adjacent facies prediction



More thinks…

- Some facies are not well precited as others
  - More well information could help

- Labeling the importance of rocks