

# CREWES solutions for Machine Learning competitions in 2020

Marcelo Guarido, David Emery, Daniel Trad, and Kris Innanen

CREWES Sponsors Meeting, December 4<sup>th</sup>, 2020



**NSERC  
CRSNG**




**UNIVERSITY OF CALGARY**  
FACULTY OF SCIENCE  
Department of Geoscience



2020 SEG Annual  
Meeting Machine  
Learning Interpretation  
Workshop

FORCE: Seismic Fault  
Mapping

FORCE: Machine  
Predicted Lithology



# 2020 SEG Annual Meeting Machine Learning Interpretation Workshop

Sun, J., Zhang, T., Niu, Z., Emery, D. J.,  
Guarido, M., Trad, D. O., and Innanen, K. A.  
H., 2020, Deep learning for seismic facies  
classification in Parihaka: *CREWES  
Research Report*, **32**, 51.



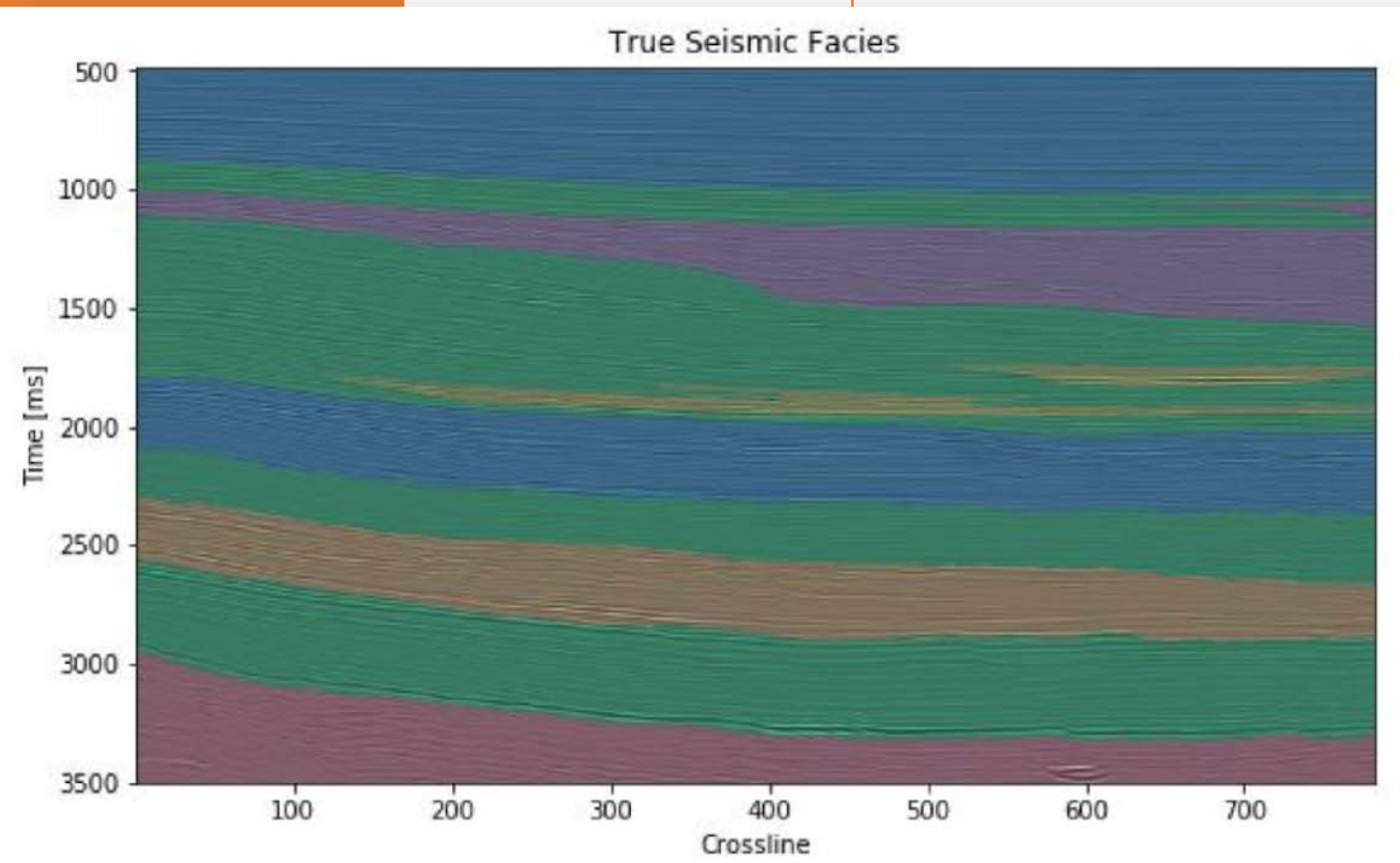
## Seismic facies classification

3D Parihaka, NZ

Training data – labels provided by Chevron

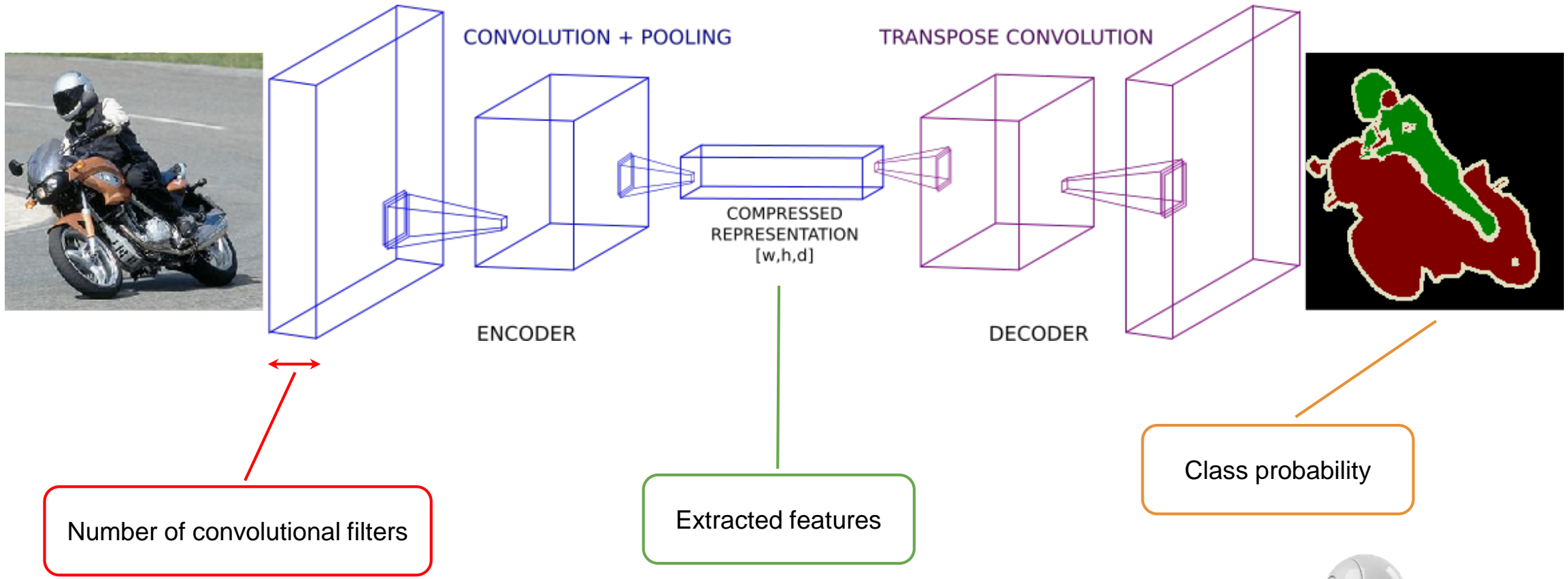
Six different classes (facies)

Blind volumes for testing



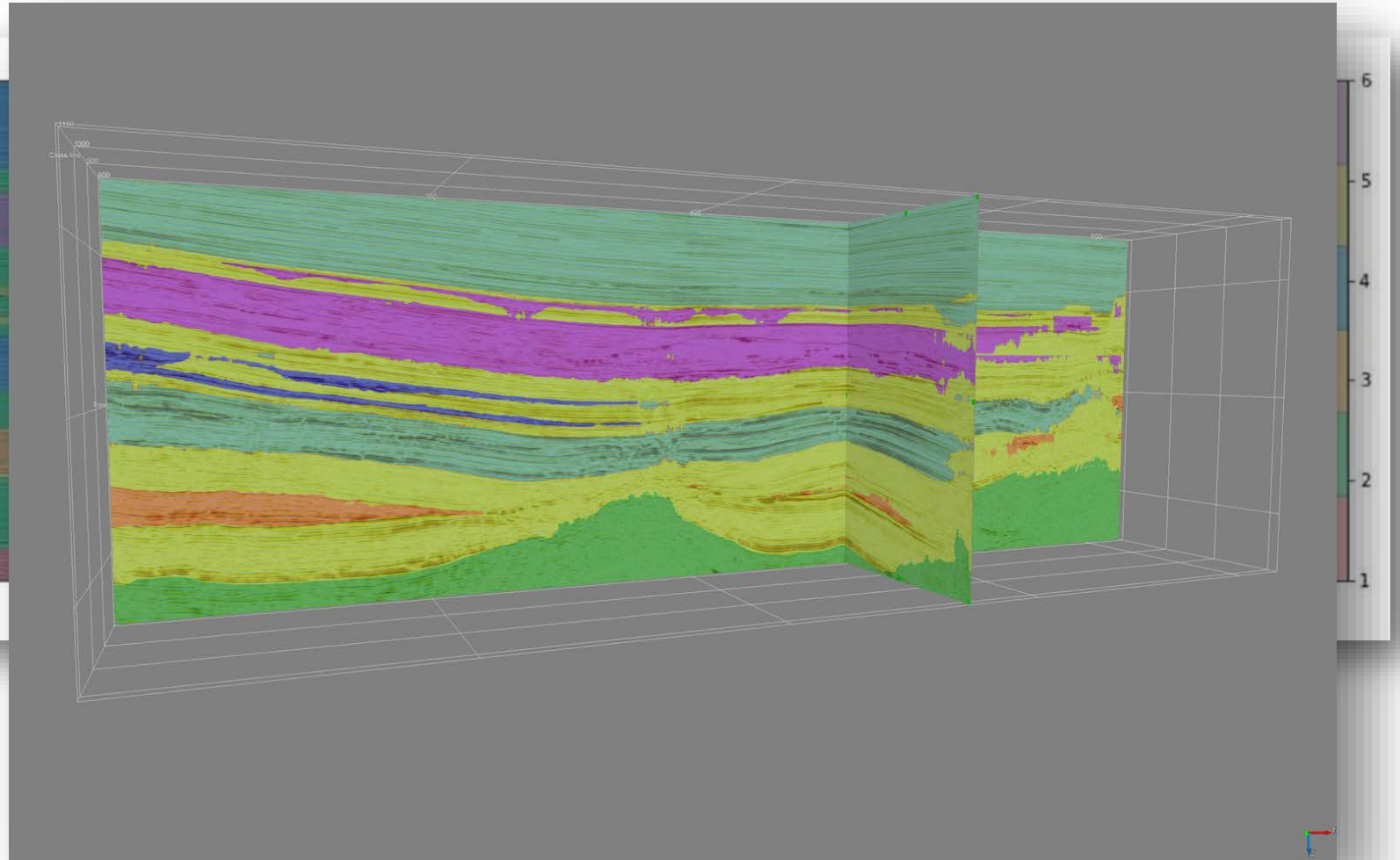
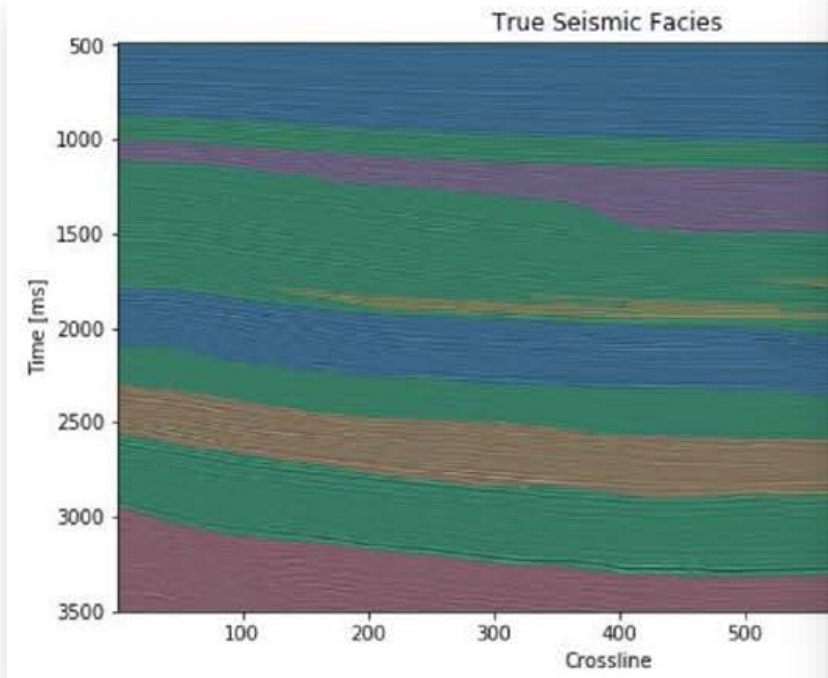


# Image Segmentation





# Results





Team	Pixel Acc.	SSIM	Mean IoU	Channel IoU	Geology
🏆 A	3	2	5	3	3
B	6	6	6	6	4
🏆 C	5	5	3	2	2
🏆 D	4	3	3	4	4
E	8	8	8	7	7
F	7	7	7	8	8
G	9	9	9	9	9
H	10	10	10	10	10
🏆 I	2	3	2	5	4
🏆 K	1	1	1	1	1




# Force: Seismic Fault Mapping

Wozniakowska, P., Guarido, M., Fathalian, A., Trad, D. O., and Emery, D. J., 2020, A 2.5D deep learning approach to identify faults in seismic sessions: CREWES Research Report, 32, 55.





X E E K CHALLENGES SUPPORT SIGN UP SIGN IN



## FORCE: Seismic Fault Mapping

Builds the best machine learning based fault mapping algorithm for seismic data

ACTIVE | \$0K IN PRIZES | 8/10/2020-10/16/2020 | 77

SEISMIC DATA

[OVERVIEW](#) DATA RULES LEADERBOARD DISCUSSION

Join

Seismic fault detection

3D volume

Training on synthetic data

2.5D approach

## Overview

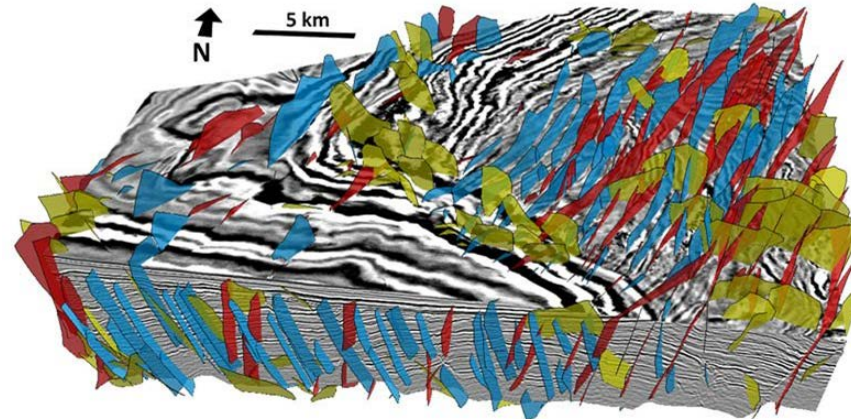
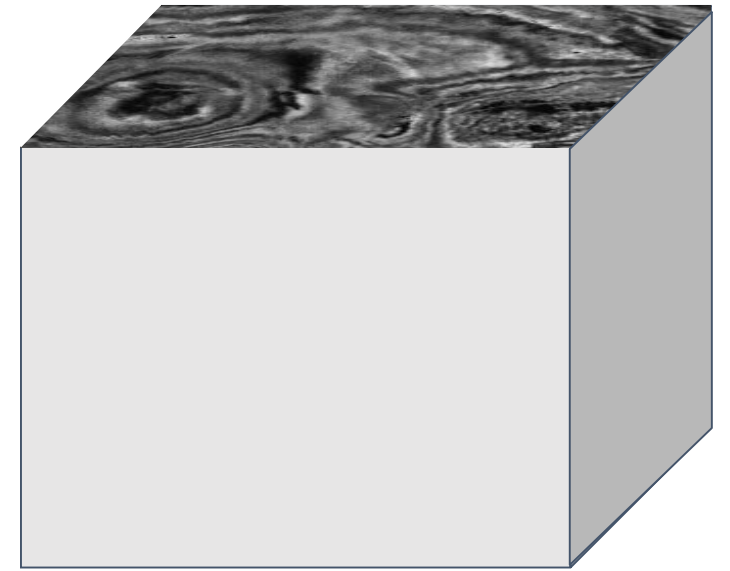
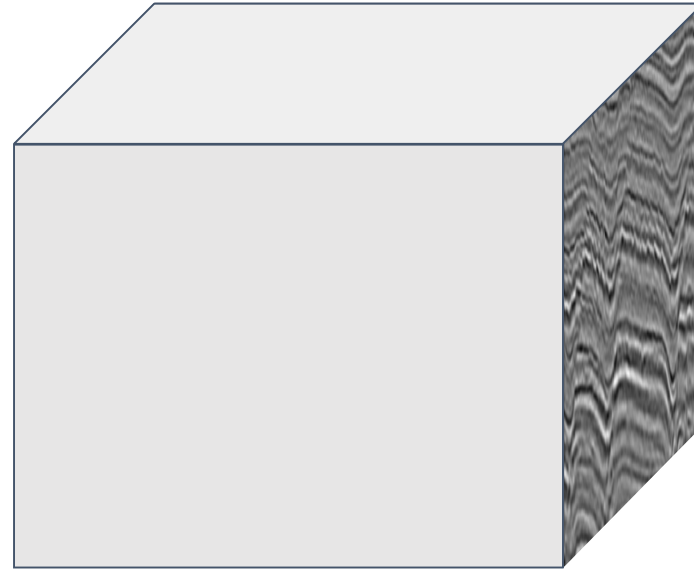
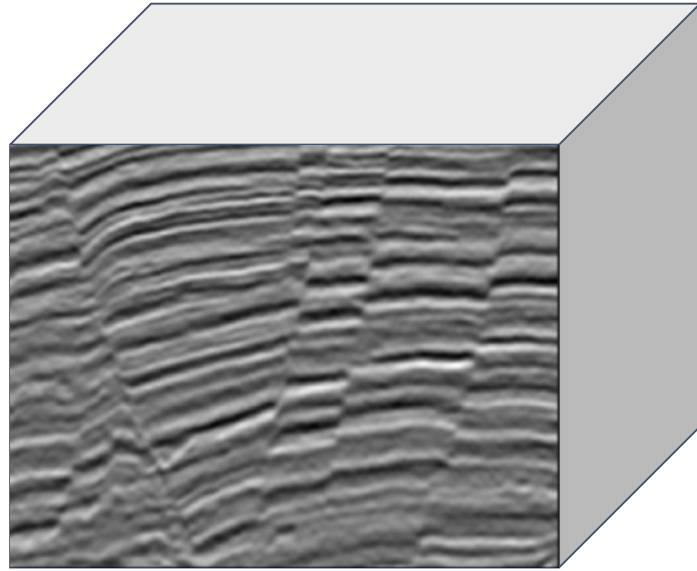
There is no leaderboard for this competition. If you want to share screenshots of your current algorithm please send them to the organizers who will publish them on a gallery page.

This contest is developed in collaboration with [FORCE](#).

[Competition link](#)



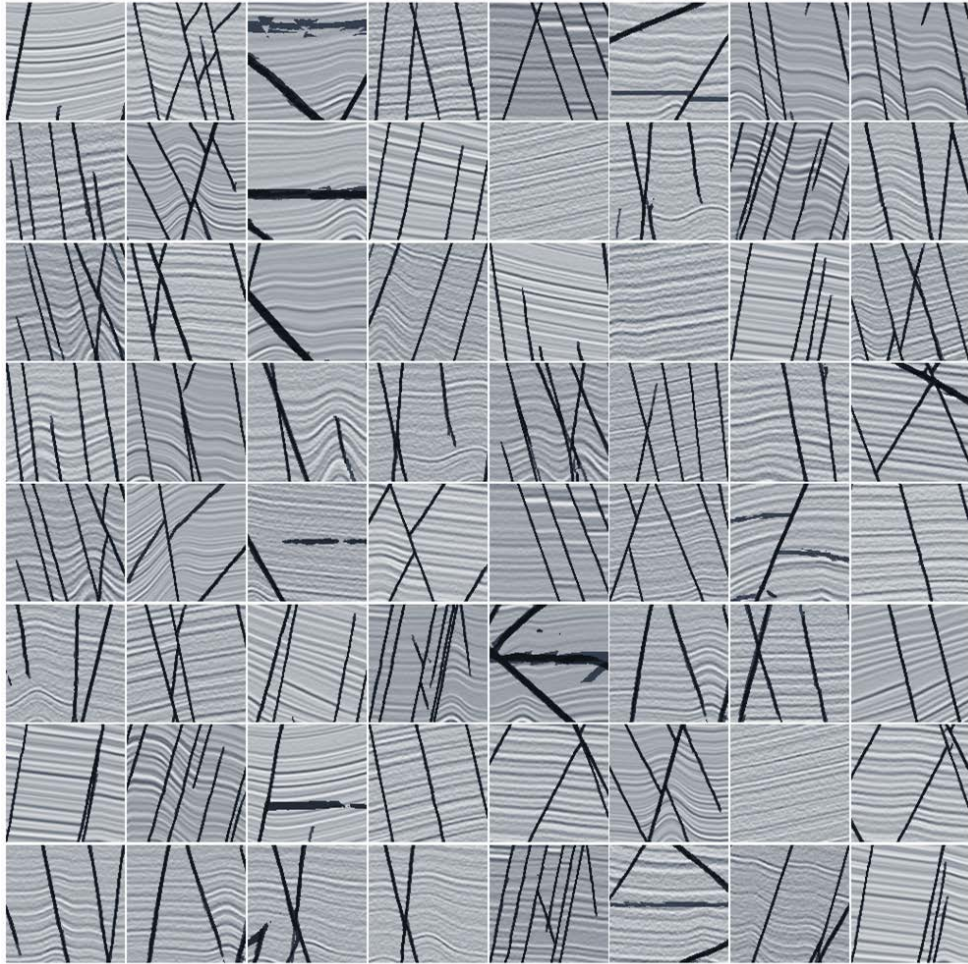
# 2.5D Approximation







# Synthetic Data





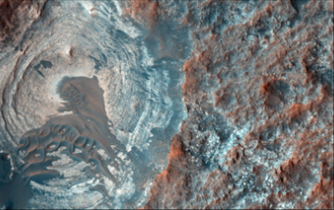
# Force: Machine Predicted Lithology

Guarido, M., Emery, D. J., Macquet, M.,  
Trad, D. O., and Innanen, K. A. H.,  
2020, The Pitfalls and Insights of Log  
Facies Classification for a Machine  
Learning Contest: CREWES Research  
Report, 32, 18.



X E E K

CHALLENGES SUPPORT SIGN UP SIGN IN



## FORCE: Machine Predicted Lithology

Create a machine learning model that has the highest accuracy in prediction lithology from a suite of wireline logs. A training dataset with hand interpreted and QC'ed wellbore lithology is available.

ACTIVE | \$0K IN PRIZES | 8/10/2020-10/16/2020 | 285

SUPERVISED LEARNING, RECURRENT NETWORKS, WELL LOGS, CNN, LITHOLOGY, NPD, MACHINE LEARNING

OVERVIEW DATA RULES LEADERBOARD DISCUSSION

Join

Facies classification

Well logs

Raw data

Tricky data and goals

## Overview

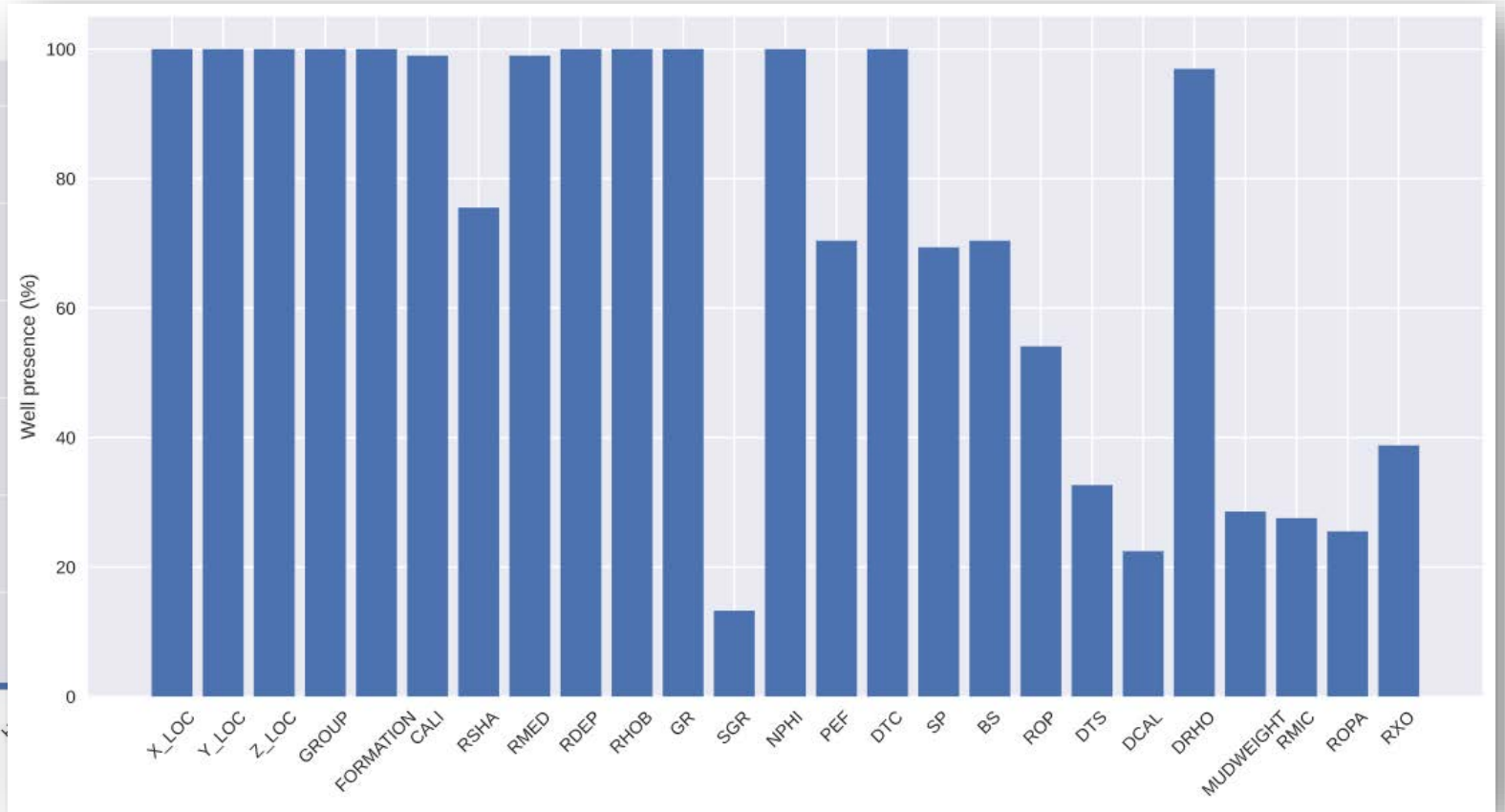
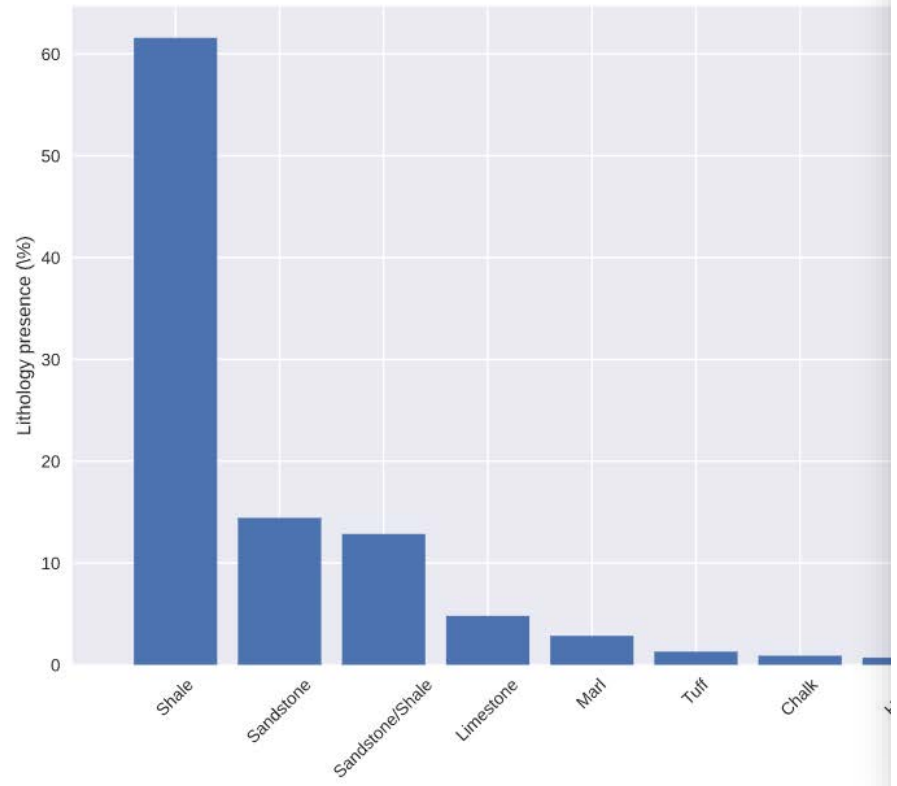
This contest is developed in collaboration with [FORCE](#).

The objective of the competition is to correctly predict lithology labels for provided well logs, provided NPD lithostratigraphy and well X, Y position.

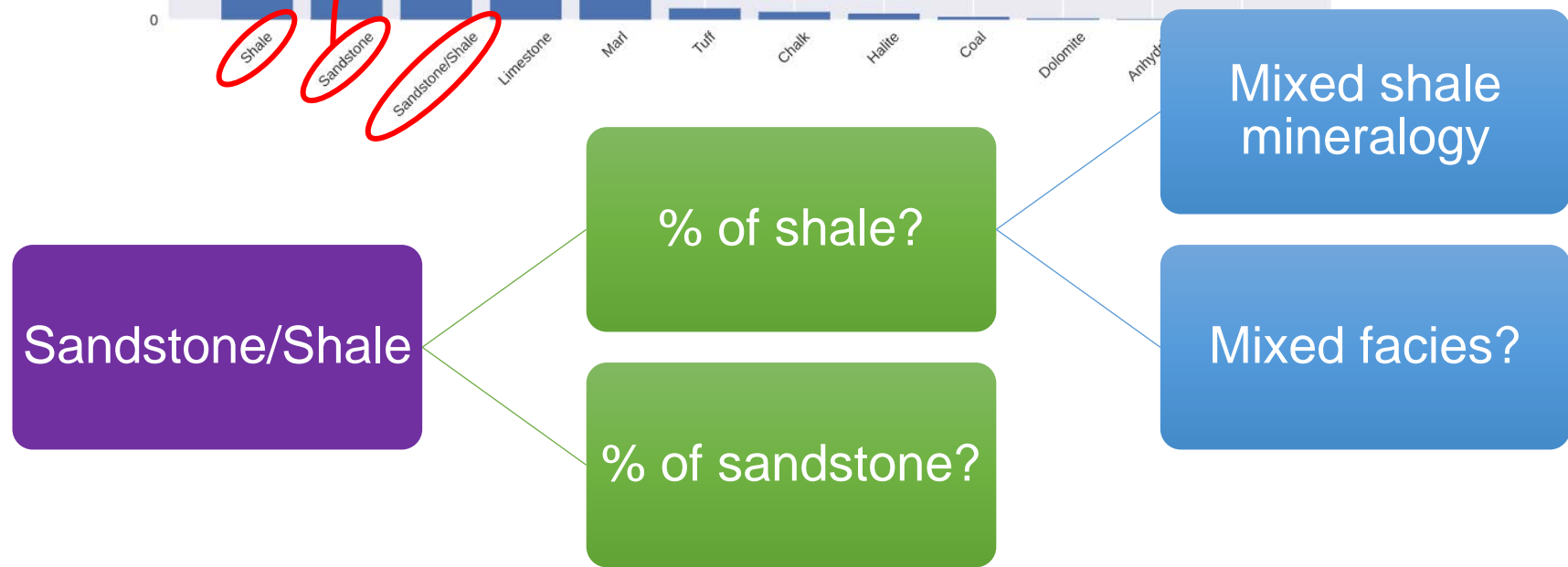
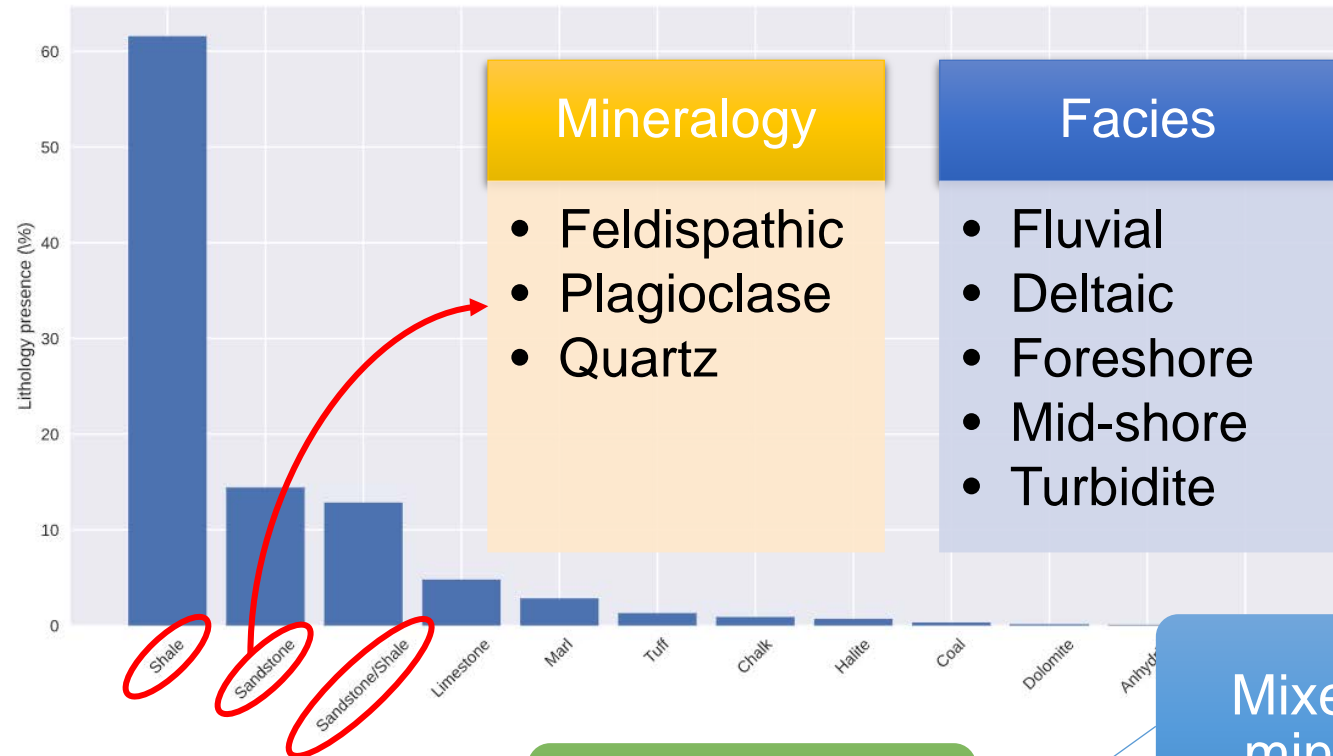
[Competition link](#)



98 Wells

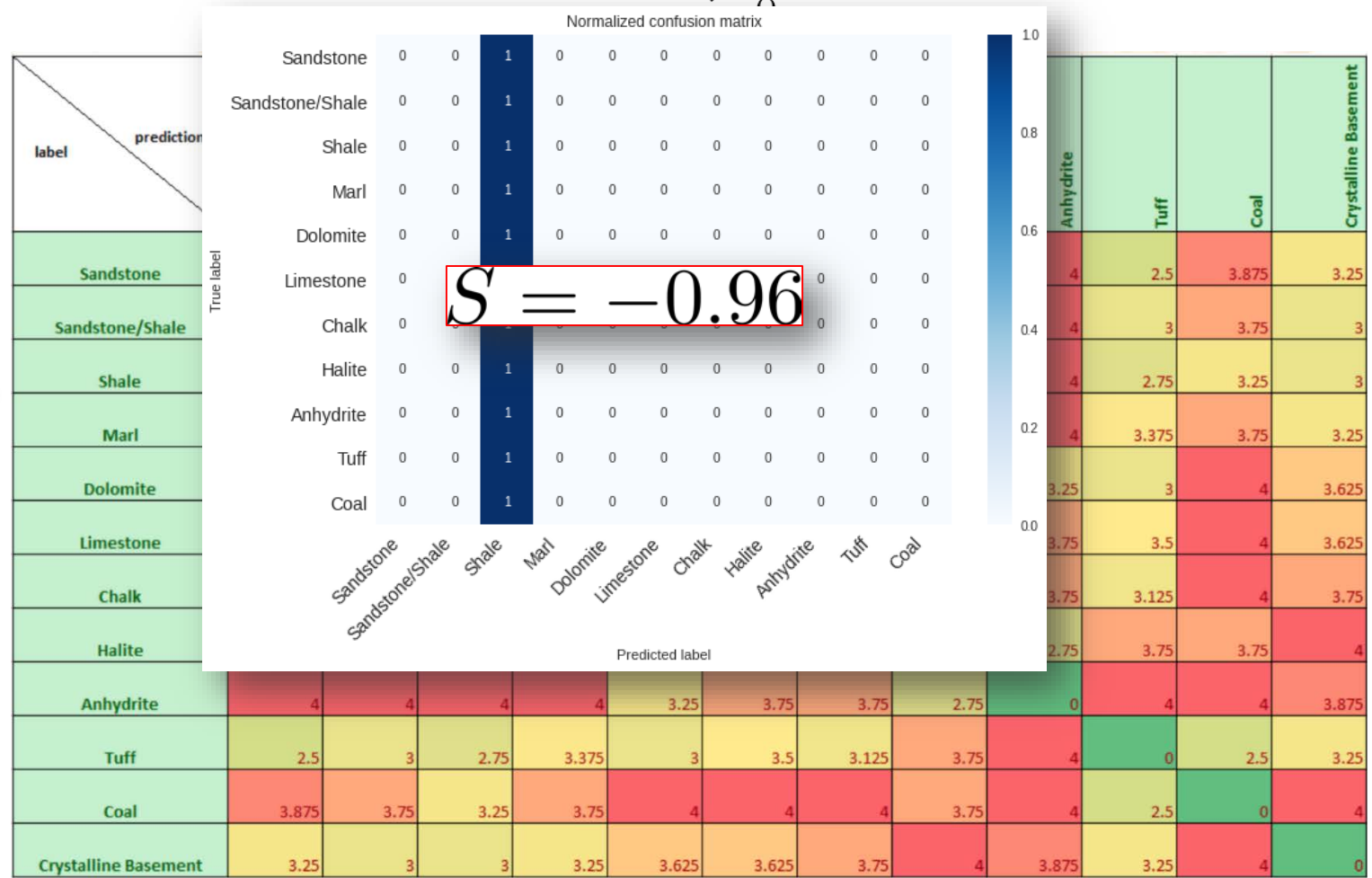






$$S = -\frac{1}{N} \sum_{i=1}^N A_{\hat{y}_i y_i}$$

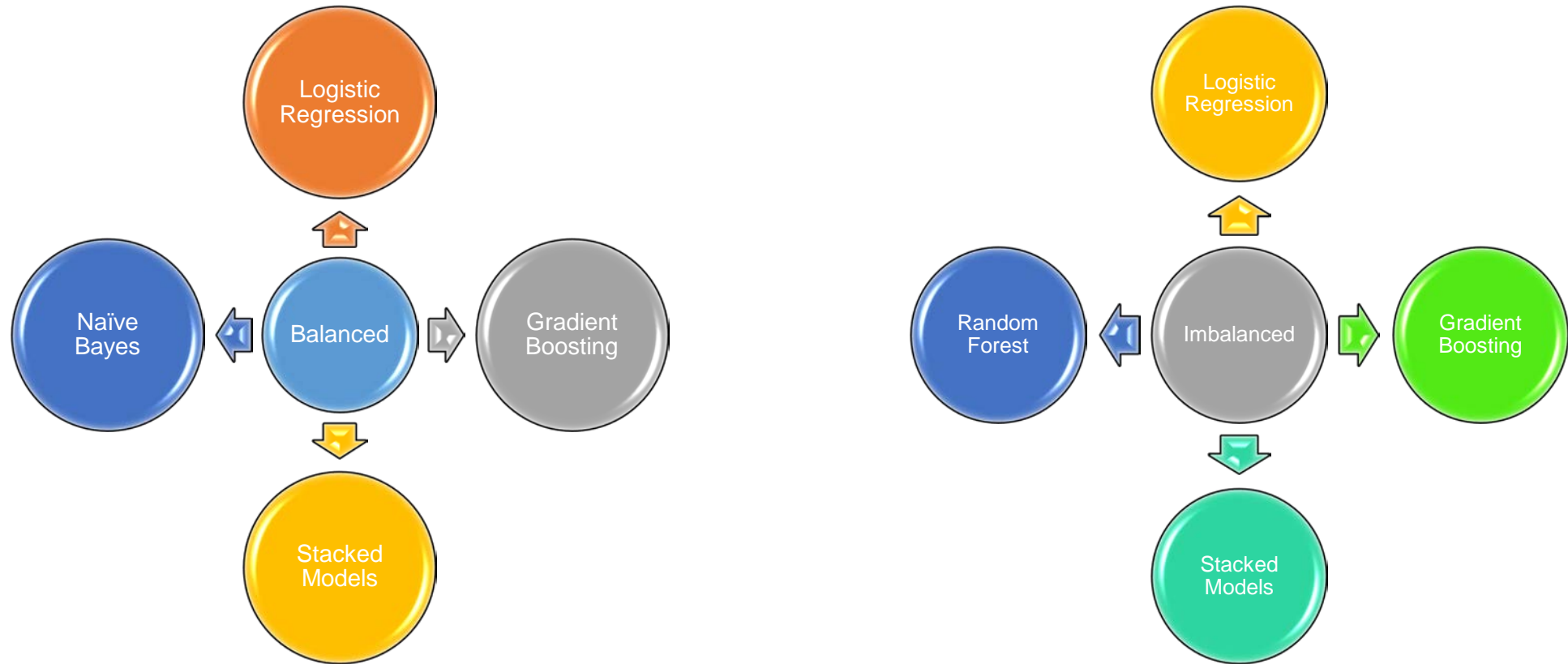
Penalty Matrix







# Models





# Balanced Models

## Metrics



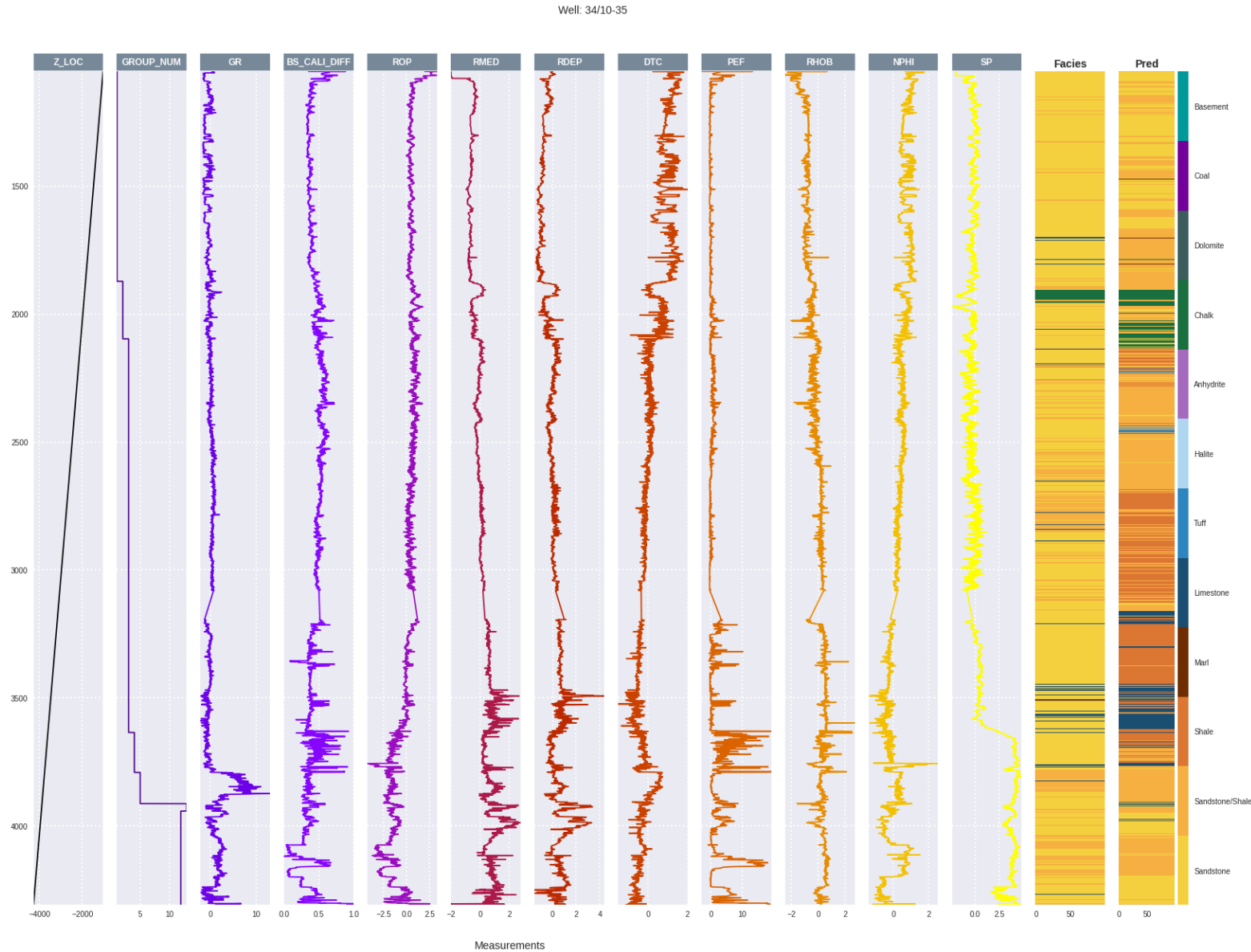
Balanced accuracy: 0.56



Contest's metric: -1.35

Normalized confusion matrix

True label \ Predicted label	Sandstone	Sandstone/Shale	Shale	Marl	Dolomite	Limestone	Chalk	Halite	Anhydrite	Tuff	Coal
Sandstone	0.52	0.22	0.03	0.03	0.0081	0.009	0.0027	0.003	0	0.12	0.047
Sandstone/Shale	0.07	0.24	0.29	0.28	0.0079	0.0093	0.017	9.7e-05	0	0.046	0.045
Shale	0.012	0.15	0.49	0.16	0.0044	0.0023	0.00047	0.0081	0.00097	0.12	0.051
Marl	0.018	0.031	0.2	0.63	0.0052	0.04	0.067	0.012	0	0	0
Dolomite	0	0.3	0	0.53	0.023	0	0	0.017	0.13	0	0.0057
Limestone	0.053	0.088	0.068	0.33	0.046	0.2	0.11	0.013	0	0.071	0.023
Chalk	0.0023	0.016	0.078	0.031	0	0	0.87	0	0	0	0
Halite	0	0	0	0	0	0	0	0.95	0.037	0	0.017
Anhydrite	0.0088	0.057	0	0.031	0	0	0	0.12	0.75	0	0.031
Tuff	0.0026	0	0.41	0	0	0	0	0	0	0.59	0
Coal	0.0051	0.025	0.02	0.02	0	0	0	0.0051	0	0	0.92





# Imbalanced Models

## Metrics



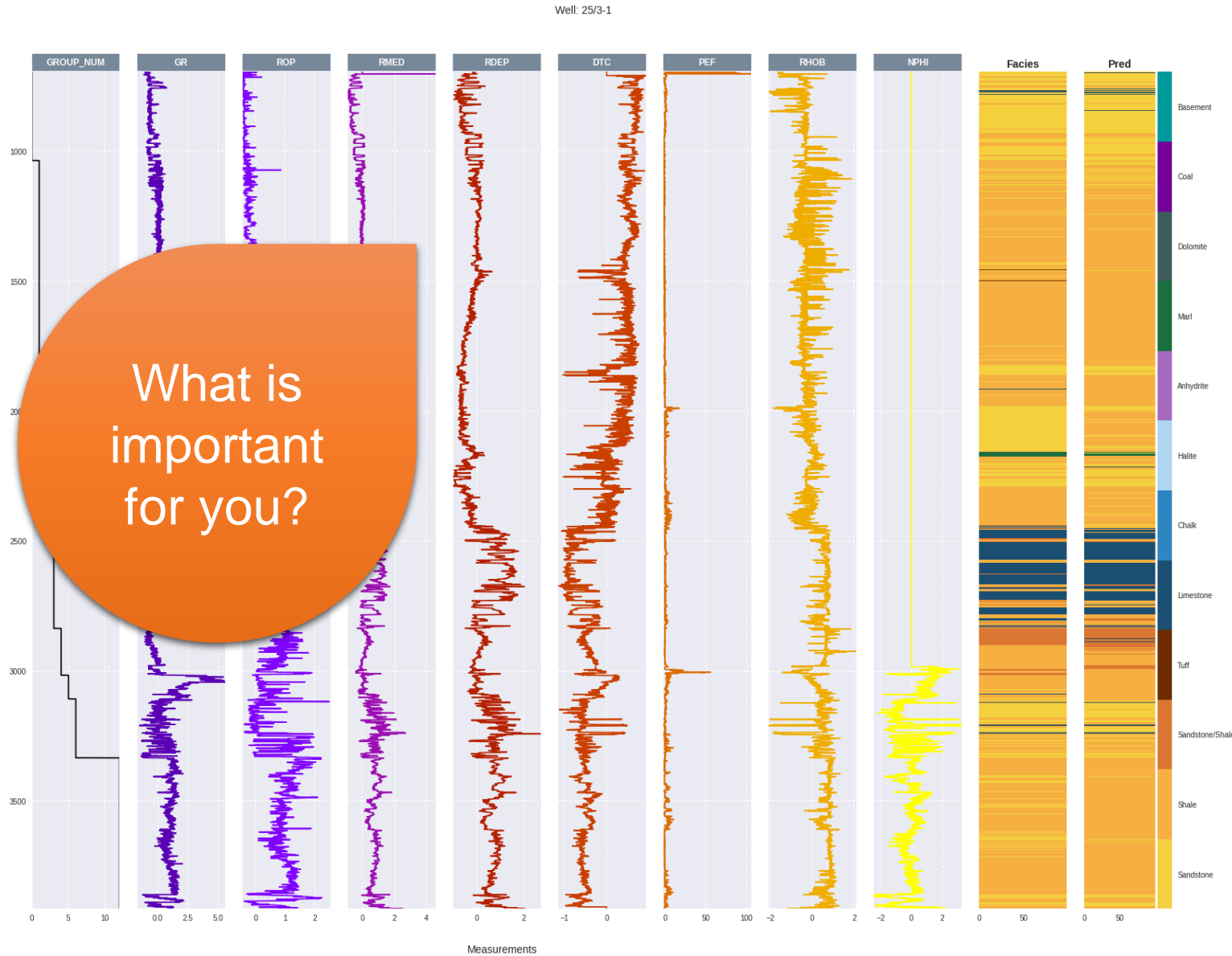
Balanced accuracy: 0.41



Contest's metric: -0.58

Normalized confusion matrix

True label \ Predicted label	Sandstone	Sandstone/Shale	Shale	Marl	Dolomite	Limestone	Chalk	Halite	Anhydrite	Tuff	Coal
Sandstone	0.64	0.097	0.23	0.0018	3e-05	0.017	0.0098	0	3e-05	0.0055	0.0026
Sandstone/Shale	0.26	0.22	0.47	0.025	8e-05	0.01	8e-05	0	0.0038	0.0043	0.0008
Shale	0.013	0.063	0.91	0.0041	5.7e-06	0.0052	0	0	3.4e-05	0.0092	0.00036
Marl	0.0074	0.1	0.23	0.38	0	0.27	0.00018	0	0	0.011	0
Dolomite	0.43	0.1	0.3	0	0.006	0.16	0	0	0	0	0
Limestone	0.035	0.037	0.17	0.039	0.00065	0.7	0.0034	0.00022	0.0018	0.016	7.3e-05
Chalk	0	0	0	0	0	1	0	0	0	0	0
Halite	0.8	0	0.05	0	0	0	0	0	0	0	0.15
Anhydrite	0.94	0.059	0	0	0	0	0	0	0	0	0
Tuff	0.041	0.032	0.4	0	0	0.028	0.0013	0	0	0.5	0
Coal	0.18	0.11	0.17	0	0	0.0018	0	0	0	0	0.53





# Models Performance

Model	Balanced Accuracy	Contest Metric
Gradient Boosting (balanced)	0.56	-1.35
Naïve Bayes	0.40	-1.86
Logistic Regression (balanced)	0.32	-2.17
Stacked Models (balanced)	0.56	-1.38
<b>Logistic Regression</b>	<b>0.08</b>	<b>-0.96</b>
Gradient Boosting	0.42	-0.59
Random Forest (balanced)	0.40	-2.00
Stacked Models	0.41	-0.58



# Closing Notes

What's next?



## Competitions

Source of data  
Keeps motivation  
Not easy



## Insights

What do you want from your data?  
What is your goal?



CREWES Sponsors

NSERC CRDPJ 461179-13 and CRDPJ 543578-19

CFREF

CREWES Staff and Students

