A Probabilistic Machine Learning based Framework for Lithology Prediction on Well Log Data

Mordekhai[1]*, Taufan Rusady[1], Andrean Satria[1]
[1]Halliburton Landmark, Kawasan Industri & Pergudangan Taman Tekno No. 1 Sektor XI, BSD City Serpong, Tangerang Selatan. Tangerang 15314

**Corresponding author's email: mordekhai.mordekhai@halliburton.com*

Abstract. Lithology prediction is one of the most important processes in petrophysical workflow since it is useful for knowing the prospective reservoir zone in the target well. Unfortunately, this process sometimes takes a long time and results in inaccurate interpretations due to the well data that has various mnemonics, the massive amount of data, and the inconsistency in manual interpretation. We present an assisted lithology interpretation framework with additional feature to compute prior and posterior probabilities during lithology prediction to help geoscientists if there are some irrelevant prediction results. We use various references to oil and gas basins data around the world resulting more than 60 pre-built models included in this framework. The framework can select model automatically based on the similarity between the well log curve in test data and references data. Besides being able to provide reliable and accurate results, this framework is a cloud based and has a centralized database. These features can accelerate collaboration and integration between users in predicting lithology. The data used in this study is from one of the most productive oil and gas field that has a varied number of wells and lithology. Based on these characteristics, this field is considered suitable for providing an objective judgement. This application is proven to be able to provide reliable results by producing prediction accuracy and F1 score above 0.6 using an automated model. This framework can also assist geoscientists to interpret exploration wells that do not yet have a valid lithology label.
Keywords: Probabilistic, Machine Learning, Lithology Prediction

1. Introduction

Lithology interpretation is an important workflow in oil and gas exploration. In well data, this process is carried out to characterize the reservoir based on the wireline log contained in each well. However, this process takes a long time (manual lithology interpretation for a 1km well with 3 tracks occurs around 4 hours) and sometimes results in inaccurate interpretations. The inaccurate interpretations can be caused by a human bias, inconsistency of various mnemonics/massive amount of data, and a missed pattern/ relationships in the well log data.

Machine learning assisted lithology interpretation framework is implemented to get a standardize, scalable, and seamless integration between petrophysicist/geoscientist. The framework offers an uncertainty measurement by using prior and posterior probability. These probability parameters help users to re-check the lithology whether it is still not producing a good confidence level.

2. Data and Methodology

The Dataset is taken from offshore Norway (Figure 1) and dominated by Shale and Sandstone lithology. This area has a complex geological setting since there is a lot of lithology class consists of Shale, Sandstone, Sandy Shale, Limestone, Marl, Tuff, Coal, Chalk, Dolomite, Anhydrite (Figure 2). Furthermore, many large structural closures have been mapped in anticlines and rotated fault blocks, in addition to stratigraphic traps in clastic environments and in carbonate build-ups and associated facies (Bjørlykke, 2019).

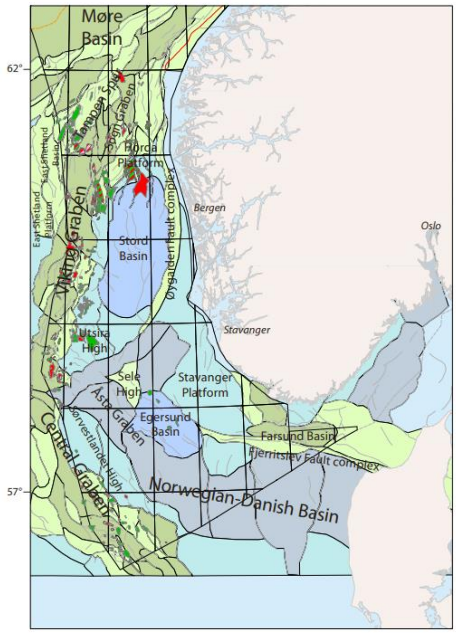

Figure 1. Regional geology setting of the research area (Norwegian Petroleum Directorate, 1996).

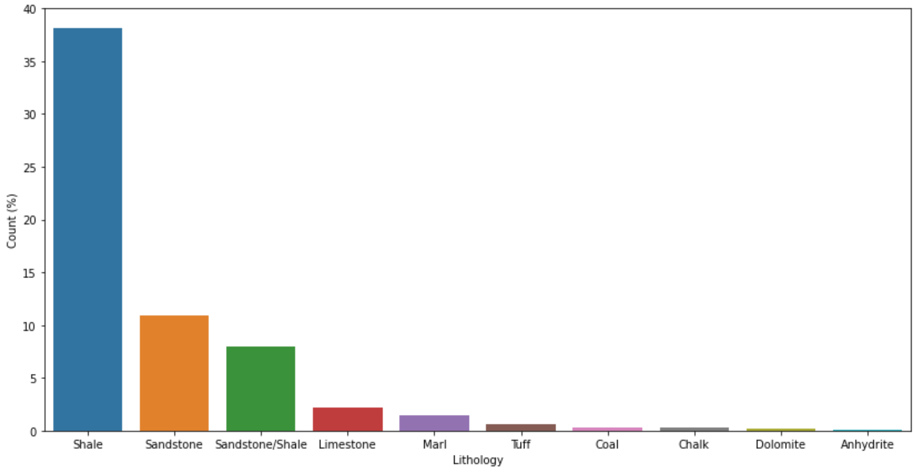

Figure 2. Lithology class percentage in the dataset.

10 well log data (Figure 3) are acquired from blind well data FORCE 2020 lithofacies competition. Each well log data consists of wireline log and lithology label. The wireline logs used are Gamma Ray (GR), Porosity (NPHI), Density (RHOB), Sonic (DTC), and Resistivity (DRES). Lithology labels (ground truth) are determined based on manual interpretation from completion logs, mud logs, & wireline curves. We implement machine learning framework with cloud native technology that uses supervised machine learning to predict lithology from wireline or LWD (Logging While Drilling) log data.

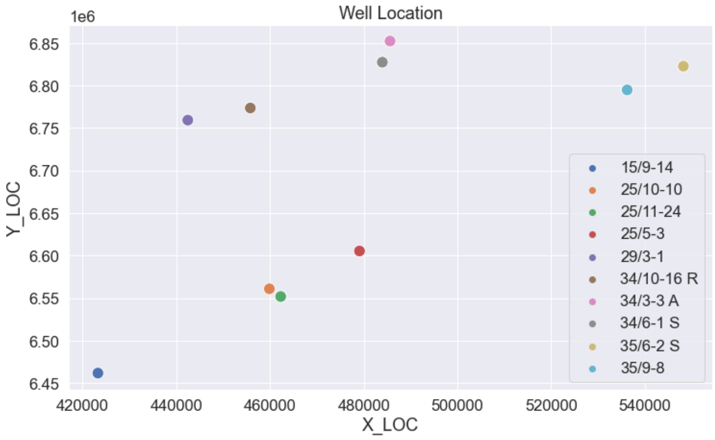

Figure 3. Well location of the research area.

In general, machine learning workflow (Figure 4) is divided into two pipelines: training and prediction pipeline.

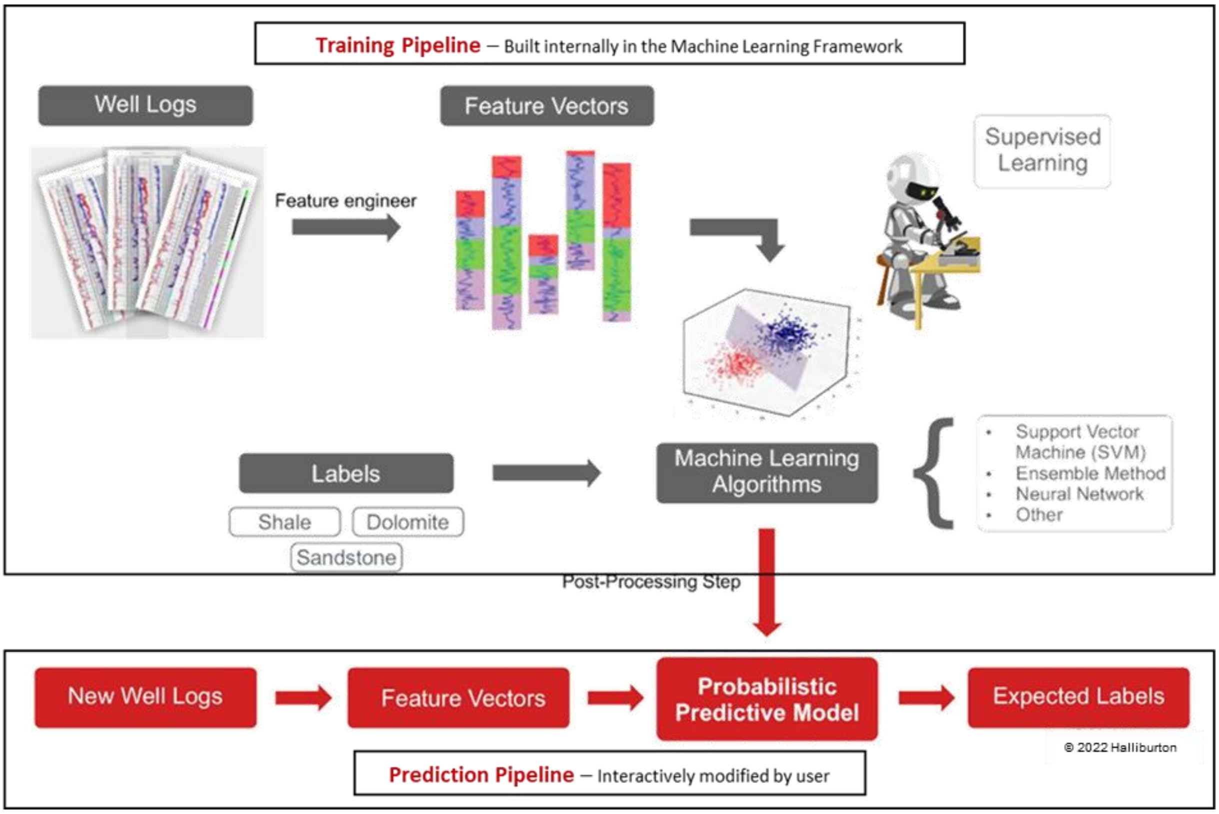

Figure 4. Training & Prediction pipeline of the machine learning framework.

Training pipeline is built internally in the machine learning framework. The algorithm consists of 64 pre-trained model that covers a wide array of wireline log combinations and lithologies. The supervised machine learning algorithm ensured that the wells were interpreted consistently by removing interpreter biases and inconsistencies (Popescu, 2021). The generated models can be chosen manually or automatically without modifying the existing algorithm in the prediction pipeline.

Prediction pipeline is a workflow that is can be fully utilized by user. This pipeline covers from uploading single/multiple well log until lithology prediction. We can choose which well log parameter that we want to use as the feature vector for predicting the lithology.

2.1. Prior Probability

This machine learning framework uses a probabilistic measurement to determine the most optimum selected model. The probability that is used as an indicator before prediction is prior probability. Prior probability measures the wireline data similarity between the training and test data before making predictions. If the algorithm recognizes the features in the incoming wireline, it can confidently predict a lithology classification and Prior Probability will be high (Figure 5).

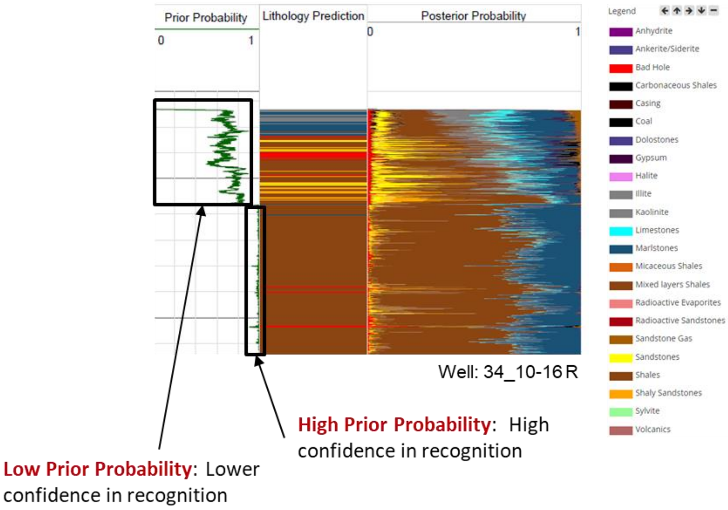

Figure 5. Prior probability in Well 34_10-16R.

2.2. Posterior Probability

Another probability that is produced to quantify uncertainty is posterior probability. Posterior probability distributions can provide additional information for geoscientist when validating or refining the initial prediction results. Figure 6 shows the posterior probability result after lithology prediction.

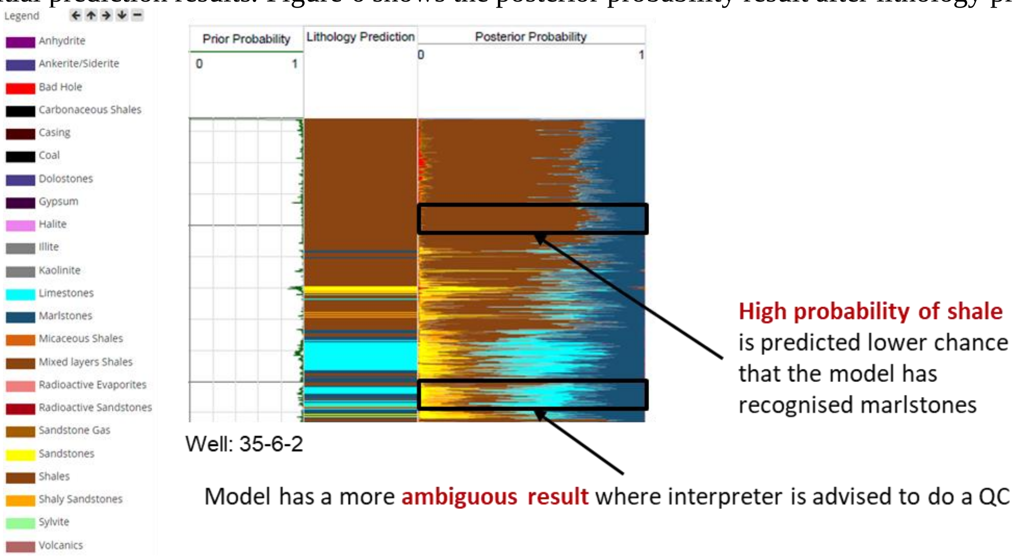

Figure 6. Posterior probability in Well 35-6-2.

These two probabilities are not only providing a confidence estimate for the output classification but also suggests possible alternative classifications if users don't agree with the algorithm's preferred classification. The analogy of the Prior and Posterior is illustrated in the Figure 7 using the example of binary (Sandstone and Shale) classification.

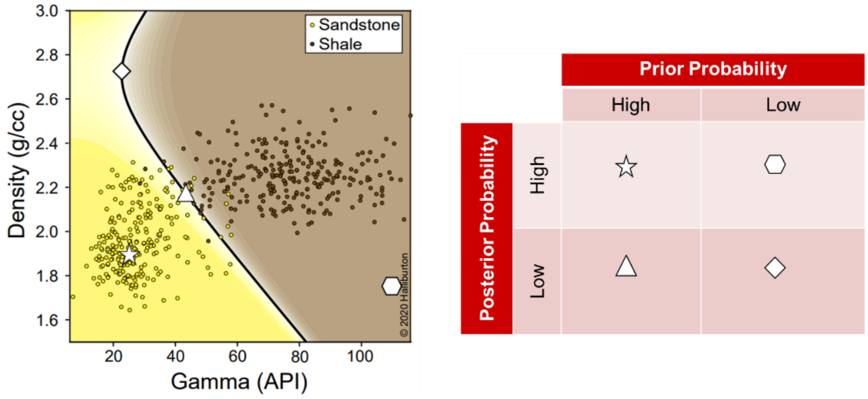

Figure 7. Illustration of prior and posterior probability.

2.3. Evaluation Metrics

After we get all the prediction results, we calculate various evaluation metric to ensure the credibility of the machine learning framework. Precision, Recall, and F1 Score is calculated in 10 wells by comparing the ground truth and prediction results.

$$Precision = \frac{TP}{TP + FP} \qquad Recall = \frac{TP}{TP + FN}$$

$$F1 = \frac{TP}{TP + \frac{1}{2}(FP + FN)} \rightarrow \frac{2}{\frac{1}{Recall} + \frac{1}{Precision}}$$

Figure 8. Precision, Recall, and F1 Score equation.

Recall is the proportion of Real Positive cases that are correctly Predicted Positive. This measures the Coverage of the Real Positive cases by the +P (Predicted Positive) rule. Conversely, Precision denotes the proportion of Predicted Positive cases that are correctly Real Positives (Powers, 2011). The weighted harmonic means of precision and recall, the F-measure, also known as the F1-score is a scale of testing accuracy for an input dataset (Srivastava, 2016).
The equation in Figure 8 is determined based on the TP (True Positive), FP (False Positive), TN (True Negative), and FN (False Negative). The illustration of these metrics is shown in Figure 9 when we assume there are 2 classes of lithologies (Sand & Shale) in our dataset.

		Actual	
		POSITIVE (Sand)	NEGATIVE (Shale)
Predicted	POSITIVE (Sand)	TP (25)	FP (20)
	NEGATIVE (Shale)	FN (13)	TN (225)

Figure 9. Illustration of TP, FP, TN, & FN in binary classification (Sand & Shale).

True Positive: The number of positive data (sand) that is predicted to be true
False Positive: The number of negative data (shale) that is predicted as positive data (sand)
True Negative: The number of negative data (shale) that is predicted to be true
False Negative: The number of positive data (sand) that is predicted as negative data (shale)

3. Results and Discussion

The prediction result was obtained by importing all the wireline log data (without the ground truth lithology data) into the machine learning framework. This framework predicts the lithology based on the wireline log and the selected pre-trained model in the framework. After we get the lithology results, we calculate the precision, recall, and F1 score between the ML prediction result and the ground truth from all wells.

Table 1. Precision, Recall, and F1 Score for 10 tested wells. Yellow highlight indicates best prediction. Orange highlight indicates worst prediction.

Well	Precision	Recall	F1
15_9-14	0.57	0.47	0.48
25_5-3	0.74	0.69	0.69
25_10-10	0.88	0.71	0.77
25_11-24	0.42	0.57	0.45
29_3-1	0.7	0.69	0.64
34_3-3 A	0.91	0.91	0.91
34_6-1 S	0.65	0.63	0.6
34_10-16 R	0.79	0.72	0.74
35_6-2 S	0.72	0.56	0.61
35_9-8-9	0.44	0.22	0.15
AVERAGE	**0.682**	**0.617**	**0.604**

The overall evaluation metrics calculated in Table 1 are producing good results. The average score for Precision, Recall, & F1 are above 0.6. The highest average score obtained by well 34_3-3 A by achieving average score above 0.9 (yellow highlight in Table 1), where only two wells that having average score below 0.5 (Well 25_11-4 & Well 35_9-8-9) highlighted in orange color (Table 1). The Machine Learning framework providing a good result by only acquiring the wireline log data without having an information related to ground truth lithology beforehand.

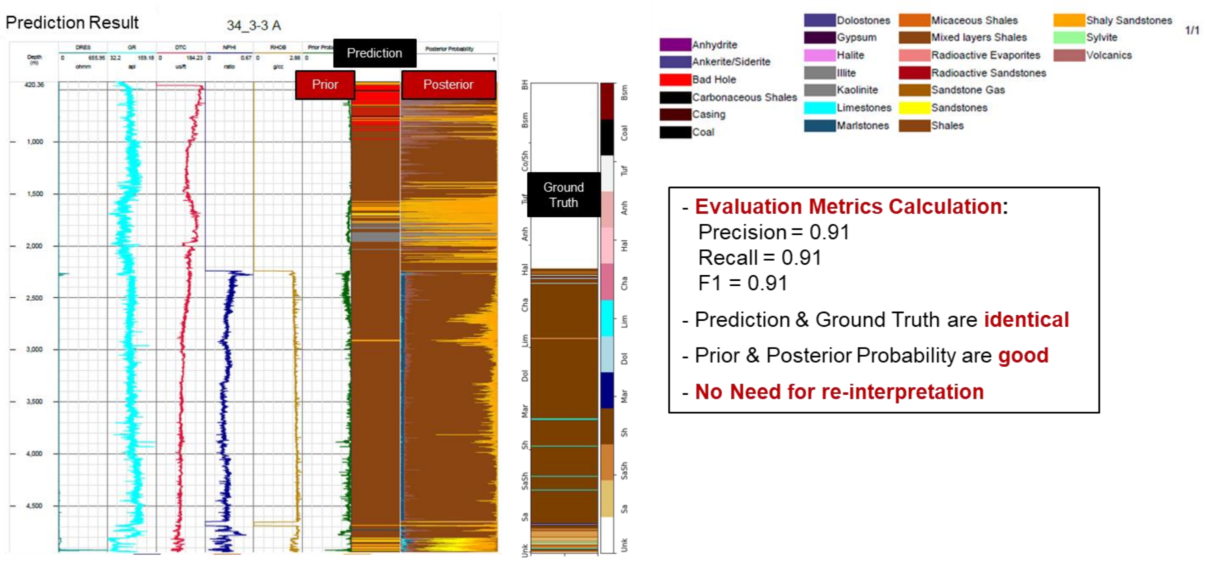

Figure 10. Example on Well 34_3_3 that has a good prediction. This well is not needed for QC.

Figure 10 shows the prediction result of Well 34_3_3 compare with the ground truth. The evaluation metrics show a good result as well as the uncertainty quantification (prior & posterior probability). The prior probability shows ~1 value along the well and the posterior probability shows a good confidence rating of the predicted lithology.

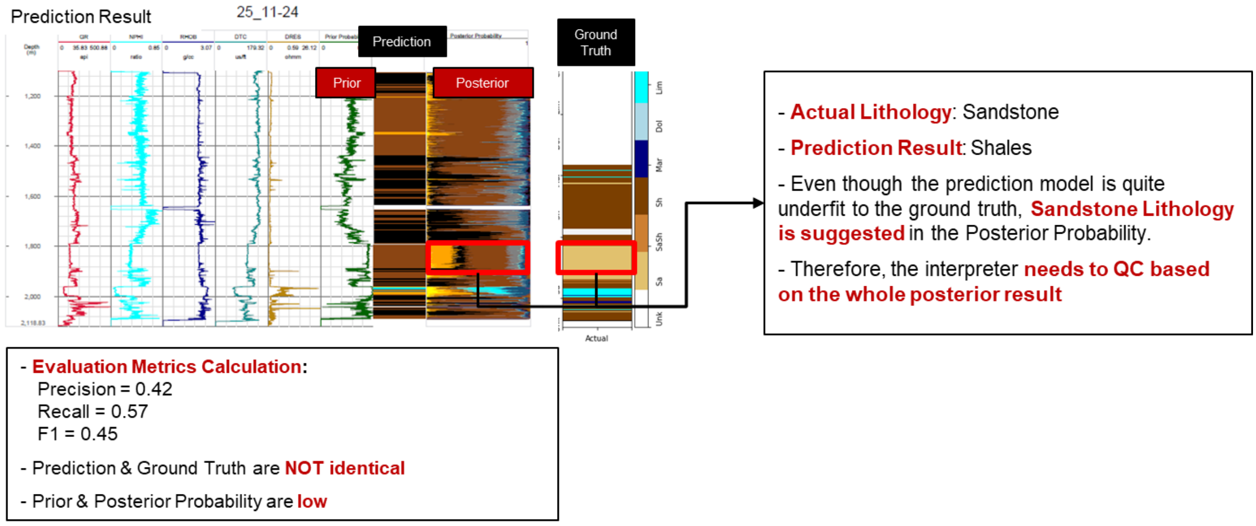

Figure 11. Example on Well 25-11-24 that has a bad prediction. This well is needed for QC by looking at the prior and posterior probability.

Figure 11 shows the prediction result of Well 25_11_24. By comparing it with the ground truth, the prediction shows a non-identical result. This is also shown from the evaluation metrics which are not very good when compared to other wells. This non-accurate prediction is coherent with the high uncertainty results. We can see that along the well, prior probability has the value around 0.5, while the posterior probability indicates other high probability lithology alternatives.

Figure 10 and Figure 11 show us that the accuracy of the lithology prediction results can be determined from the level of uncertainty. if the uncertainty (prior & posterior probability) result shows a low value, it is possible that the prediction results also show good accuracy (geologists do not need to re-interpret). On the other hand, if the uncertainty results show a high value, the prediction results may show poor accuracy (geologists need to re-interpret).

The uncertainty feature in this machine learning framework can help geologists determine whether the prediction results are good or not, especially when we don't have the ground truth. For example, in Figure 11 there is a false prediction in depth interval 1800-1900 (red square). The ground truth is sandstone, but the prediction is Shale. Even though the prediction model is quite underfit to the ground truth, there is Sandstone probability in the posterior result. Geologist is advised to re-interpret this interval using posterior probability alternatives lithology suggestion to get a comprehensive judgment.

Lithologically speaking, the best prediction was obtained on Shale prediction while the worst was obtained on Sandy Shale prediction (Table 2). In terms of geological interpretation, Sandy shale is often misinterpreted due to the nature similarity with Sandstone or Shale. The complexity of the high content of clay minerals in the shaly-sand lamination layer (thin bed) is accompanied by mineral clay distribution which inherently affects the log data response in the form of high gamma ray values and low resistivity (Brandsen, 2016). This condition has an impact on ground truth data labeling. The ambiguity of Sandy Shale manual interpretation makes the ground truth label becoming uncertain (the interpreted label could be Sandstone or Shale lithology).

Table 2. Shale and Sandy Shale Evaluation Metrics.

	Shale	Sandy Shale
Precision	0.697	0.276
Recall	0.86	0.202
F1 Score	0.742	0.195

The Sandy Shale ambiguity is also shown at the FN and FP percentage comparison on Sandy Shale prediction at Table 3. The misclassification of Sandy Shale among 10 wells was dominated by Sand or Shale label. The error percentage of Sandy Shale misclassification to Sand or Shale is varied from 45% to 100%.

Table 3. (a) Proportion of Sandy Shale False Negative (FN) that classified as Sand and Shale. (b) Proportion of Sandy Shale False Positive (FP) that classified as Sand and Shale.

	FN proportion to total FN of Sandy Shale		
Well	**Sand**	**Shale**	**Combine**
15-9-14	46%	6%	**52%**
25-5-3	61%	4%	**65%**
25_10-10-1	85%	6%	**92%**
25_11-24	100%	0%	**100%**
29_3-1-4	48%	5%	**53%**
34_3-3 A	73%	2%	**75%**
34_6-1 S	95%	3%	**98%**
34_10-16 R	42%	3%	45%
35_6-2 S	71%	4%	**75%**
35_9-8-9	100%	0%	**100%**

(a)

	FP proportion to total FP of Sandy Shale		
Well	**Sand**	**Shale**	**Combine**
15-9-14	50%	37%	**87%**
25-5-3	14%	85%	**98%**
25_10-10-1	38%	49%	**87%**
25_11-24	0%	0%	0%
29_3-1-4	46%	43%	**89%**
34_3-3 A	9%	79%	**89%**
34_6-1 S	11%	84%	**95%**
34_10-16 R	22%	74%	**96%**
35_6-2 S	42%	41%	**83%**
35_9-8-9	10%	52%	**61%**

(b)

4. Conclusions

- This research shows that the machine learning framework is useful to accelerate geoscientists in performing lithological interpretations. This application is proven to be able to provide reliable results by producing F1 score above 0.6 (Table 1).
- This machine learning framework can quantify the confidence rating of the lithology prediction by using the quantitative uncertainty algorithm. This uncertainty feature can help geologists determine the quality of prediction results especially when ground truth lithology data is not available, which represent the reality.
- High uncertainty can be caused by low prior probability and/or low posterior probability. If these conditions occur at certain log intervals, geoscientists need to do a quality control and reinterpretation using the posterior alternative suggestions.
- Based on the lithology evaluation metrics, the best prediction was obtained by Shale while the worst was obtained by Sandy Shale. The misclassified of Sandy Shale is dominated by Sand and Shale label.

Acknowledgments

We would like to express a special gratitude to Halliburton Landmark for granting the permission to access the Assisted Lithology Interpretation demo environment for this research paper.

References

[1] Brandsen, H. (2016), Shale and Clay Part II. NTNU Lecture Notes: Petrophysics, Interpretation of Well Data, Advanced Course.

[2] Bjørlykke, Knut. (2019). Early history of petroleum exploration offshore Norway and its impact on geoscience teaching and research. Norwegian Journal of Geology. 10.17850/njg99-3-2.

[3] Norwegian Petroleum Directorate. 1996. Geology and Petroleum Resources in the Barents Sea.

[4] Popescu, Marian, Head, Rebecca, Ferriday, Tim, Evans, Kate, Montero, Jose, Zhang, Jiazuo, Jones, Gwynfor, and Geovani Christopher Kaeng. "Using Supervised Machine Learning Algorithms for Automated Lithology Prediction from Wireline Log Data." Paper presented at the SPE Eastern Europe Subsurface Conference, Kyiv, Ukraine, November 2021. doi: https://doi.org/10.2118/208559-MS.

[5] Powers, David & Ailab. (2011). Evaluation: From precision, recall and F-measure to ROC, informedness, markedness & correlation. J. Mach. Learn. Technol. 2. 2229-3981. 10.9735/2229-3981.

[6] Srivastava, Saurabh & Singh, Girdhari. (2016). F1 Score Analysis of Search Engines. S.K.I.T Research Journal. 6. 1-6.