Classifying Lithofacies in Wireline Logs with Deep Learning

Kyle Rosa, Margarita Kongawoin and Fabian Guisset, DUG Technology

Introduction

Identifying lithofacies from subsurface well data is a foundational task in geoscience. However, traditional rock physics workflows typically rely on a narrow set of well logs and often fail to incorporate the full breadth of available subsurface information. These methods are further constrained by their use of simplified heuristics and manual interpretation, which can be time-consuming and error-prone, limiting scalability and consistency across large datasets. Automated lithology prediction can provide rapid, reproducible insight about the subsurface, reducing the turn-around time of interpretation workflows.

We present a neural network–based method for predicting lithofacies directly from wireline well logs, generating probability distributions over nine distinct lithology types. Our model achieves a top-1 classification accuracy of 90.9%, and a top-2 classification accuracy of 97.5%, with F1 scores exceeding 90% on the most common lithologies.

Method

Our training data is derived from DUG's multi-client statistical rock physics studies, including wells from the Browse, Bonaparte, Canning, Perth, and Carnarvon basins, and the Exmouth Plateau. Collectively, these studies cover over 300 onshore and offshore wells and provide a depth-calibrated, petrophysically consistent, and geologically informed library of end-member picks. These end-member picks are derived through interpreter-guided analysis of multiple well logs and cuttings descriptions. Representative intervals are manually selected based on log responses and geological context, and their lithologies are identified as the cleanest examples of each facies, as described in Lamont et al., 2008.

To produce labels for training, we mapped the interpreter-provided lithology annotations onto one of nine lithology classes, and rendered the end-member intervals into a dense array of class labels with sampling aligned to the well curve data. Our lithology classes include base lithologies such as claystone, sandstone, and limestone, as well as variants like argillaceous sandstone and hot sandstone. We also include a "high density" class as a catch-all bucket for intervals where logs are impacted by high-density phases, regardless of host rock.

Our model predicts these classes from nine raw well curves, chosen for their suitability in differentiating lithologies: gamma ray, caliper, spontaneous potential, resistivity, compressional slowness, shear slowness, density, neutron porosity, and photoelectric factor. These inputs are typically available in most well logs, making the model widely applicable in real-world scenarios. Subsets of these curves have previously been successfully used for similar classification tasks (see Xie et al, 2024 and Prajapati et al, 2024 for examples). Note that we exclude sample depth from our model's input to avoid overfitting to any well-specific geological trends.

The model architecture combines a transformer-based embedding module (Vaswani et al., 2017) and a 1D CNN. The transformer maps raw measurements into an embedding space and neatly handles missing data through attention masking, after which the CNN aggregates contextual information from nearby depth samples and produces final class scores.

Our dataset exhibits a strong class imbalance, reflecting the natural abundance of different lithologies. To address this, we use the focal loss described in Lin et al., 2020 to down-weight the gradient contribution of well-classified samples during training and improve performance on minority classes.

On our validation data, our model achieves 90.9% top-1 accuracy, and a top-2 classification accuracy of 97.7%. Imbalanced datasets such as ours can skew these metrics, so to put these numbers into context,

blindly predicting the two most common lithologies would achieve 54.4% top-1 and 85.3% top-2 accuracy. To evaluate the model's performance across different lithology classes, we calculate a confusion matrix and derive class-specific precision, recall, and F1 score metrics, which are tabulated in Figure 1.

Class	Lithology	label%	pred%	recall	precision	F1 score
0	Claystone	54.42%	55.60%	95.53%	93.50%	94.51%
1	Sandstone	30.89%	27.44%	85.98%	96.78%	91.06%
2	Sandstone - Argillaceous	5.01%	5.16%	79.23%	76.88%	78.04%
3	Limestone	6.90%	8.18%	93.27%	78.64%	85.34%
4	Siltstone	2.34%	2.81%	68.50%	57.05%	62.25%
5	High Density	0.15%	0.45%	93.43%	30.41%	45.88%
6	Sandstone - Hot	0.21%	0.25%	86.99%	73.23%	79.52%
7	Dolomite	0.03%	0.04%	98.15%	72.60%	83.46%
8	Basement	0.06%	0.06%	100.00%	92.56%	96.14%

Figure 1 *Class-specific validation metrics, including the prevalence of each class in the labels and predictions, as well as precision, recall, and F1 scores.*

Well-represented lithology types such as claystone, sandstone, and limestone can be reliably identified, with class-specific F1 scores around 90%. We achieve moderate F1 scores on hot sandstone (79.5%), argillaceous sandstone (78.0%), and siltstone (62.25%), however the latter two improve significantly when considering top-2 recall. Reviewing the confusion matrices in Figure 2, we find that the model often predicts claystone instead of siltstone and plain sandstone instead of argillaceous sandstone. The high density class has great recall but is over-predicted, leading to poor precision and an F1 score of only 45.9%. The dolomite and basement classes also show great results, but the small number of samples in the training and validation data make it difficult to draw strong conclusions about how well these results will generalise to more diverse wells.

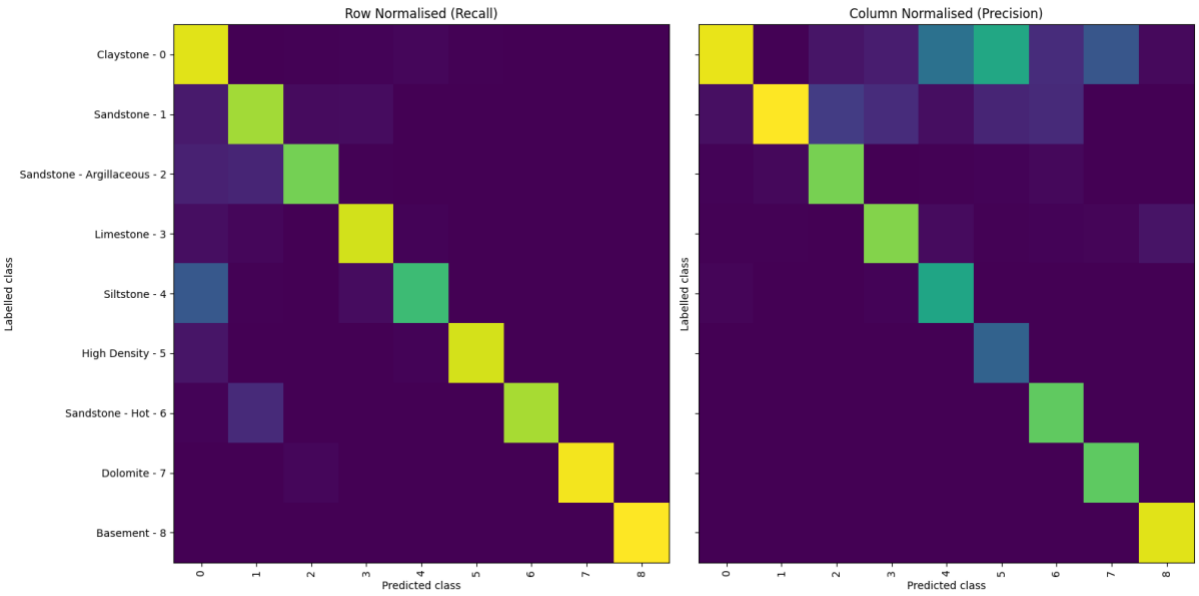

Figure 2 *Normalised confusion matrices, with labels on the y axis and predictions on the x axis. The left plot is normalised so that rows sum to one to show the distribution of predictions for each label,*

and the right has normalised columns to show the distribution of labels for each prediction. The diagonals of these matrices are the class-specific recall and precision values in Figure 1.

Examples

For a qualitative evaluation, we compare our model's results to a traditional petrophysical interpretation track in Figure 3.

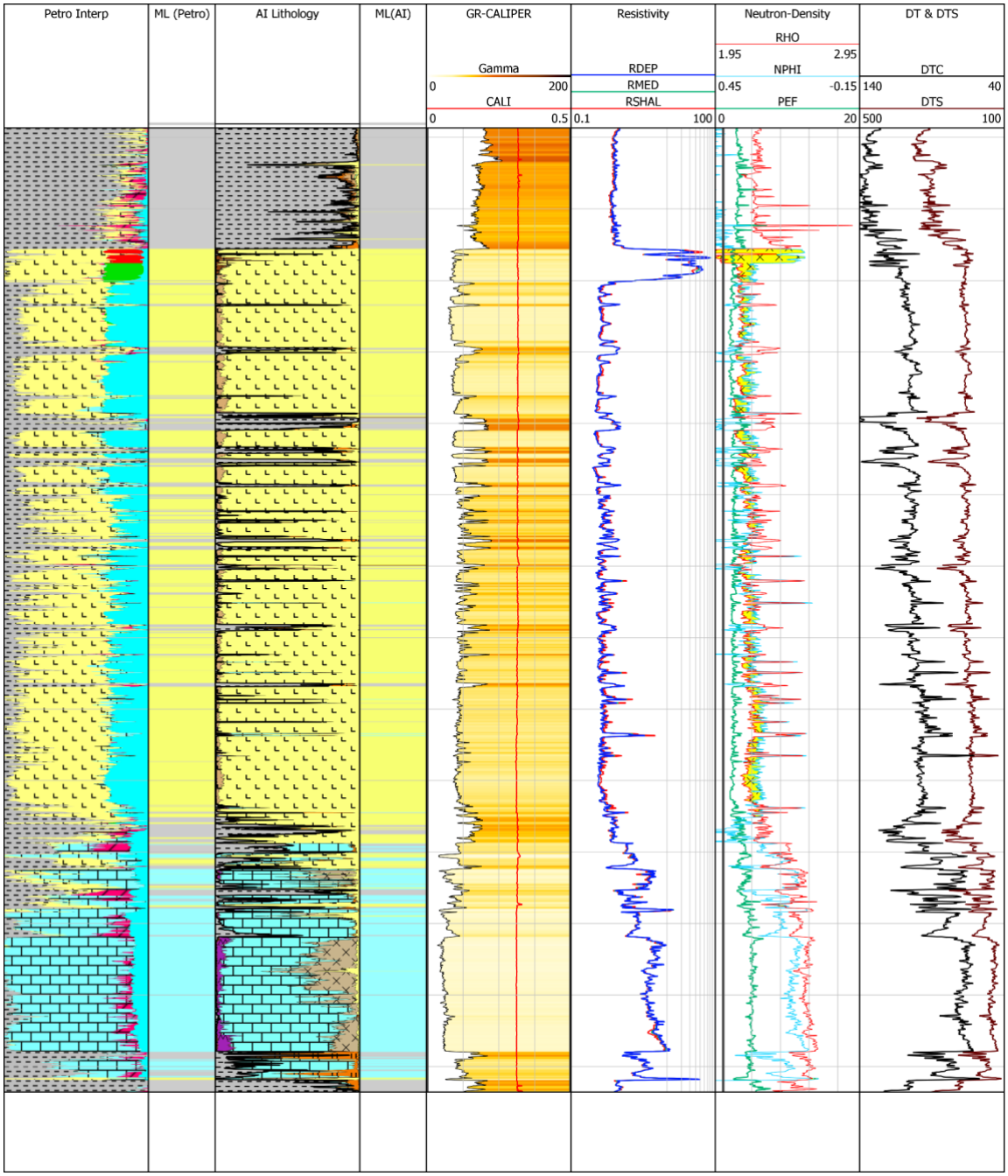

Figure 3 *Comparison of petrophysical interpretation results and our inference model, with input well curves displayed to the right. Most likely (ML) lithology tracks are displayed to the immediate right of the Petro Interp and AI Lithology density tracks.*

The petrophysical interpretation track displays the relative volumetric composition as assessed by a petrophysicist, and is therefore influenced by interpreter judgement and methodological assumptions. In contrast, the AI lithology track predicts how an end-member pick at that location would be annotated using only the well log inputs. The class probabilities can be used to inform an interpreter of several possible alternatives, or flattened into a single Most Likely (ML) prediction, as shown in the ML tracks

of Figure 3. While the two tracks show a high level of agreement between majority composition and predicted lithology, the machine learning approach offers a scalable, reproducible alternative that can be applied consistently across multiple wells.

Conclusions

Our model learned to predict lithologies from wireline logs, and largely agrees with traditional petrophysical interpretation results when generalised to entire wells despite only training on end-member picks. The pre-trained model can be applied to novel well data to provide an immediate overview of its facies. This will work best for wells with similar geological features to our training dataset, but can still provide useful insight at a glance and aid initial interpretation efforts for more diverse wells.

End-member picks are conceptually simple and user-friendly to create, making it straightforward to produce additional annotations and fine-tune the base model for application to specific wells or basins. The interpreter can highlight and correct a few key mistakes, and have that correction propagate to the rest of the lithology predictions for that well and others. This enables faster, more reproducible workflows that are still adaptable to expert judgement.

Acknowledgements

The authors would like to thank Mehdi Labani for his subject matter expertise in petrophysics and end-member picking. We also gratefully acknowledge DUG Multi-Client for providing access to their extensive rock physics data library and supporting the use of multi-client datasets in this study.

References

Hall, B. [2016] Facies classification using machine learning. *The Leading Edge*, **35**(10), 906–909.

Lamont, M.G., Thompson, T.A. and Bevilacqua, C. [2008] Drilling success as a result of probabilistic lithology and fluid prediction: A case study in the Carnarvon Basin, WA. *APPEA Journal,* **48**(1), 273–288.

Lin, T.-Y. et al. [2020] Focal loss for dense object detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **42**(2), 318–327.

Prajapati, R. et al. [2024] Machine learning assisted lithology prediction using geophysical logs: A case study from Cambay basin. *J Earth Syst Sci*, **133**, 108.

Vaswani, A. et al. [2017] Attention is all you need. *Advances in Neural Information Processing Systems*, **30**, 5998–6008.

Xie, D. et al. [2024]. A Transformer and LSTM-Based Approach for Blind Well Lithology Prediction. *Symmetry*, ***16***(5), 616.