Automating Well Log Correlation with Soft Attention Convolutional Neural Networks

Nicholas Bivens Corey Dillard Colton Baker

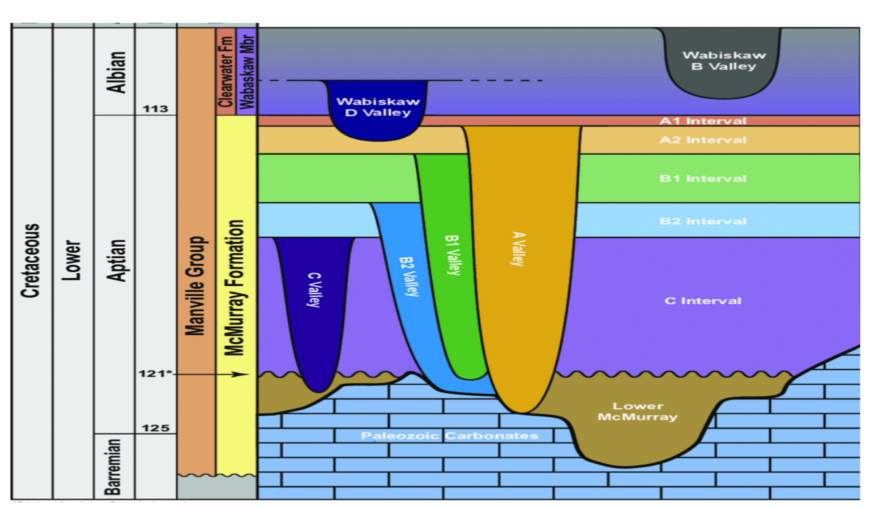
Abstract

This study applies advanced machine learning tools to predict formation tops in well logging within the Athabasca Oil Sands Area, focusing on the McMurray Formation and the overlying Wabiskaw Member of the Clearwater Formation. Utilizing well log data rich in geological details, our approach centers on employing Convolutional Neural Networks (CNN) and Random Forest models. These models were chosen for their ability to handle complex patterns and robustness against overfitting, respectively, making them well-suited for analyzing the intricate spatial characteristics of subsurface geological data. The integration of CNN and Random Forest aims to enhance the accuracy and efficiency of determining formation boundaries, crucial in well logging. This approach represents a significant advancement in the use of machine learning for geological predictions in the oil and gas industry, offering a more precise and effective method for analyzing and interpreting geological formations.

Background

Determining formation tops are traditionally a subjective process performed by petrophysicists and geologists by comparing logs from the target well with logs from neighboring. This subjectivity can lead to inconsistencies in marker identification in that different geologists may make different picks for the same formations. Identifying the tops of formations is a key part of determining the pay capability of wells as well as optimizing the techniques used in order to maximize said pay capability. The first step towards finding the reservoir properties of a basin is understanding acquiring the various log measurements such as Gamma Ray(GR), Resistivity(RES), Density Porosity(DPHI), etc. from the specified well. All logs are initially checked for the presence of washout zones, outliers, missing data, and other issues that could be resulting from limitations and errors in data collection. After processing the data for known imperfections, the next step is to perform log correlation to identify geological formations in the field or basin using the logs and align the wells according to the placement of the markers or tops of these formations.

In this project, data from the McMurray formation and the Wabiskaw Member of the Clearwater Formation in the Athabasca Oil Sands Area, Canada, was utilized. The preprocessed dataset included tops for 14 formations, provided in CSVs for picks, Log ASCII Standard (LAS) files, and a corresponding CSV mapping well IDs to LAS files. Focusing on Gamma Ray, Resistivity, Density/Porosity, and later the Shaliness Index, well logs were filtered for each formation with desired input curves and a formation top within the specified depth range. Quality picks were prioritized, and data lacking quality picks for a formation were excluded to ensure dataset integrity.



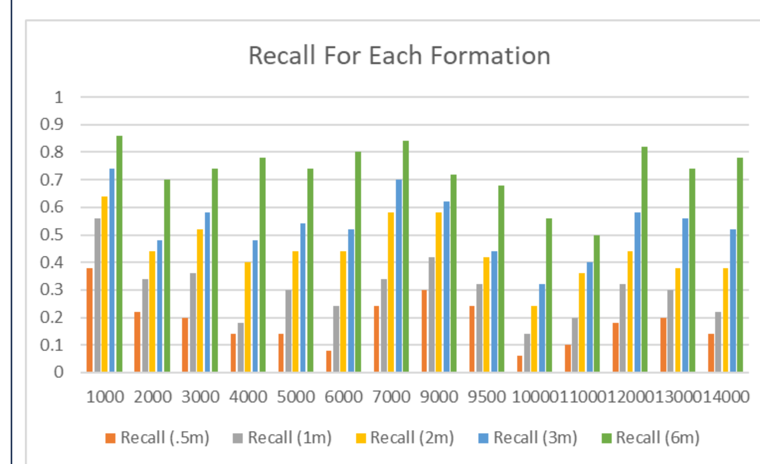
Our dataset initially presented a notable challenge due to data imbalance, where each well log featured only one expert pick for each formation. Consequently, the majority of data points were categorized as '0,' resulting in a limited range of success for the models. To address this issue, a Gaussian kernel was employed, creating a normal distribution around the expert pick. This adjustment provided the models with a slightly broader range for accurate pick classification.

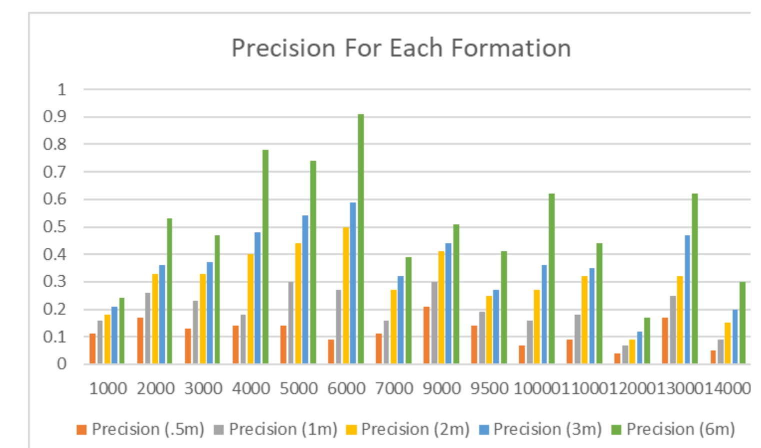
Additionally, we introduced a new variable named 'delta-T' to serve as a grace zone for model evaluation. 'Delta-T' represents a distance in meters, and if the model classifies a pick within this specified range, the classification is considered a true positive. This strategic inclusion allows the models to operate with a degree of flexibility, akin to the latitude typically afforded to experts in the field. These measures were implemented to enhance the models' performance and address the inherent challenges associated with the data imbalance in our dataset.

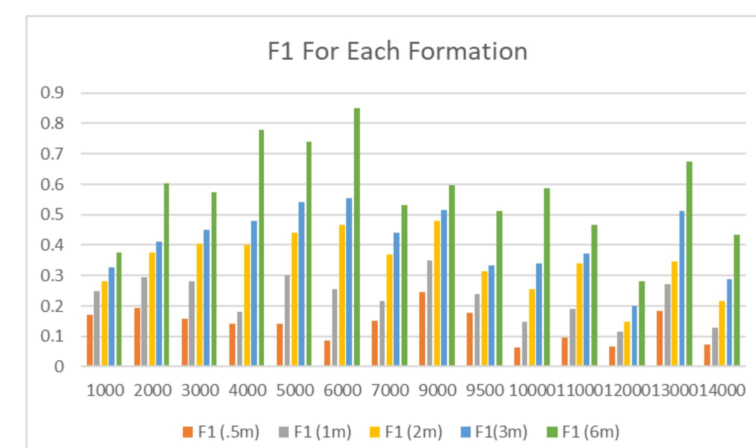
Method 1: Random Forest

Employing the **Random Forest** algorithm, we classify geological picks as either '1' (pick) or '0' (not a pick). The model's predictions are compared to geologists' picks, yielding accuracy scores and recalls. Systematic tuning, including adjustments in maximum depths and minimum data points per node, was key to optimizing performance.

Random Forest Results







Method 2: CNN with Soft-Attention

The model employs a **U-Net architecture** for a global view, facilitating an understanding of broad geological trends and structures across varied depths. A series of specialized inception modules with dilated convolutions constitutes the local view, focusing on finer details crucial for accurate formation top prediction. The **soft attention mechanism** employs element-wise multiplication to assign varying importance to elements based on attention scores from both global and local views, influencing the final result.

CNN Results

	Delta T (Accepted Distance from Ground Truth)				
	0.5 m	1 m	2 m	3 m	6 m
Mannville	0.24	0.42	0.62	0.72	0.92
t61	0.36	0.56	0.68	0.78	0.8
t51	0.16	0.34	0.54	0.68	0.78
t41	0.1	0.16	0.2	0.22	0.3
t31	0.08	0.14	0.32	0.4	0.52
Clearwater Wabiskaw	0.38	0.12	0.72	0.76	0.82
t21	0.5	0.62	0.74	0.78	0.82
t15	0.04	0.12	0.2	0.28	0.44
e14	0.18	0.28	0.42	0.46	0.7
t11	0.26	0.44	0.62	0.74	0.78
t10.5	0.2	0.26	0.34	0.44	0.68
e10	0.1	0.14	0.26	0.42	0.64
McMurray	0.1	0.24	0.28	0.32	0.5
Paleozoic	0.22	0.34	0.56	0.68	0.84

The model exhibited a range of accuracies across different formations. Notably, it achieved higher accuracy in formations like Mannville, especially at larger Delta T values, reaching up to 0.92 at 6m. However, the model encountered challenges in formations with complex lithological characteristics, such as the t41 and McMurray formations, where lower accuracies were observed. These results suggest difficulties in accurately delineating boundaries in formations characterized by rapid lithological variations. The choice of data may have limited the model's effectiveness in distinguishing subtle formation changes, particularly in shaly sands, converting gamma ray logs into more interpretable variables such as Shaliness (VSH) could enhance the model.

Conclusion

This research explored the use of a Random Forest algorithm for classifying geological picks into '1' (pick) or '0' (not a pick), with a comparative analysis against geologists' picks revealing the algorithm's accuracy and recall. Systematic tuning, including adjustments in maximum depths and minimum data points per node, was key to optimizing performance.

Further, the integration of a U-Net architecture offered a comprehensive view of geological features, capturing both broad trends and detailed structures at various depths. The model's specialized inception modules with dilated convolutions were instrumental in focusing on finer details essential for precise formation top predictions.
A soft attention mechanism was also incorporated, enhancing model interpretability. This mechanism, using element-wise multiplication based on attention scores from global and local views, variably weighted elements to influence the final prediction.

In summary, this research sheds light on the effectiveness of algorithmic approaches and architectural choices in geological predictions, while also acknowledging the challenges encountered. Future work should focus on better handling lithologically complex formations, experimenting with alternative log data transformations, and adding more interpretable variables. Continued collaboration between data scientists and geologists is vital for improving the accuracy and practicality of geological predictions in real-world applications.

Future Direction

The intricate nature of the formation geology necessitated a corresponding increase in the complexity of the CNN model. This adaptation, while somewhat effective in addressing the specificities of the dataset, did not universally enhance performance across all geological formations. Simplifying and fine-tuning the model could lead to more consistent and accurate results, though this might reduce its efficacy in identifying the most complex formations in diverse oil fields.

Additionally, improving the model with better log data, such as VSH (Volume of Shale) data, could significantly enhance its accuracy. VSH data, providing a more precise measure of shaliness, could offer a more detailed and nuanced understanding of the geological formations, thereby improving the model's predictive capabilities in complex geological scenarios.

The complexity of the formations in our study, predominantly shale, sand, and shaly-sand, posed significant challenges for our models. These geologies, known for their difficulty in identification even by experts, impacted the models' performance. It is hypothesized that in less complex geological settings, the models could potentially achieve accuracy levels comparable to those of expert geologists and petrophysicists.

Looking ahead, with a larger team and extended project timeline, a more streamlined and effective model could be developed. This model would aim to meet or even surpass industry standards, applicable across various oil fields. Such an endeavor would require extensive testing and a multitude of datasets, but it holds the promise of a more efficient and cost-effective approach for energy companies embarking on drilling in new fields.

Acknowledgments

Abubakar, A., Kulkarni, M., & Kaul, A. (Year). Soft Attention Convolutional Neural Networks for Rare Event Detection in Sequences. Schlumberger, USA.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., & Rabinovich, A. (2015). Going deeper with convolutions. Presented at The IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

Chen, L.-C., Yang, J., Wang, W., Xu, W., & Yuille, A.L. (2016). Attention to Scale: Scale-aware Semantic Image Segmentation. Proceedings of the IEEE conference on computer vision and pattern recognition, 3640–3649.