Artificial Intelligence Assisted Quantitative Petrophysical Properties Analysis using Core Images and Well Logs

Tao Lin[1,*], *Mokhles* Mezghani[2], *Chicheng* Xu[1], and *Weichang* Li[1]

[1]Aramco Americas: Aramco Research Center - Houston, 16300 Park Row Dr., Houston, TX 77084, USA
[2]Saudi Aramco, Dhahran Heights, Building 9172, 31311 Dhahran, Saudi Arabia

Abstract. Reservoir characterization requires petrophysical properties, such as porosity and permeability, to be populated in a 3D reservoir model. It is critical to quantitatively estimate the vertical profiles of petrophysical properties in each well. Traditional approaches are time consuming and labor intensive, and sometimes invasive to the core. To overcome these issues, we propose to apply image analysis and automated artificial intelligence (AI) workflow to improve high-resolution quantitative petrophysical characterization. The workflow estimates the continuous vertical profiles of petrophysical properties in each well, by integrating core plug measurements, core images and well logs with computer vision (CV) and machine learning (ML) techniques. We achieved reasonably good correlation and accuracy from the models. The approach enables core digitization, core preservation and reduces exploration risk, as well as providing petrophysical insights to assist core description.

1. Introduction

Rock properties including petrophysical properties (porosity and permeability), geomechanical properties (Poisson's ratio and Young's modulus), and geochemical properties (Total organic carbon and kerogen volume) are critical for subsurface reservoir modeling. A critical step is to estimate the vertical profiles of petrophysical properties in each well based on core measurements and well logs [1, 2]. Traditionally, qualitative visual observations of depositional and diagenetic features are routinely recorded by geoscientists for geologic interpretation and petrophysical characterization. Quantitative core measurements are typically acquired in laboratory from core plugs at discrete depth levels. Both approaches are time consuming and labor intensive, and sometimes invasive to the core. Those petrophysical properties are then used to calibrate petrophysical models together with well logs. In very heterogeneous reservoirs, petrophysical properties can exhibit large variability on very small scale to be resolved by well logs. In addition, core analysis and other relevant information generate a large number of heterogeneous formats data, which can be difficult to integrate effectively into a single modeling framework. Therefore, integrating higher resolution measurements such as core scans or core images in an automated quantitative analysis workflow can improve petrophysical characterization at the cored interval resolution [3].

Core images are often used by geologists only on a descriptive basis for sedimentary (bedding, gain size, lithology), structural (dipping, deformation), and diagenetic features (fractures, vugs). It also has rich information about the subsurface rocks that can be extracted for quantitative characterization purposes. For example, different lithologies are often present in different brightness or colors. Shale or clay-rich rocks often appear darker in color than sandstones or limestones. For the same lithology, coarse-grained and porous intervals often exhibit different textures from the tight and fine-grained intervals, which is sometimes observable by human eyes. However, core-based visual information has not been routinely integrated into petrophysical models. Recent advances in computer vision and machine learning enable more quantitative use of core images in petrophysical characterization workflows with improved performances [3].

In this paper, we introduce an AI assisted quantitative workflow that estimates the continuous vertical profiles of petrophysical properties in each well, by integrating core plug measurements, core images and well logs with computer vision (CV) and machine learning (ML) techniques. We demonstrate that core images and well logs can be quantitatively integrated into petrophysical workflows to enhance estimation of petrophysical properties in cored intervals.

* Corresponding author: tao.lin@aramcoamericas.com

2. Data Preparation

While drilling exploration or development wells, cores are extracted from the subsurface. These cores are an important source of information for subsurface characterization. After cores are extracted, they are placed in cylinders, usually 3-ft long, and then sent to the core laboratory for further analysis and visual examination. Some of these analyses are performed before slabbing the core and others are done after the slabbing.

Before the slabbing, one of the core analyses is provided by the Gamma Ray measured with relatively small spacing ranging from 0.05 ft to 0.1 ft. Some of the analyses require the extraction of several samples (called plugs) from each core of 3 ft long. Plugs are generally used to measure different properties such as porosity, permeability, and grain density with a spacing ranging from 0.5 ft to 1.0 ft. An example of slabbed and plugged cores is shown in **Figure 1**.

Figure 1. Slabbed and plugged geological cores.

For visual examination, the core is slabbed in two parts, for example, 1/4 and 3/4 for 4 inches core diameter. On the slabbed core, a handwritten text is added to the slabbed face of the core to mark the extracted plug identifier and the core depth at every foot. Finally, a high-resolution photo (approximately 100 pixels per centimeter) of the slabbed core is taken, as shown in **Figure 2**.

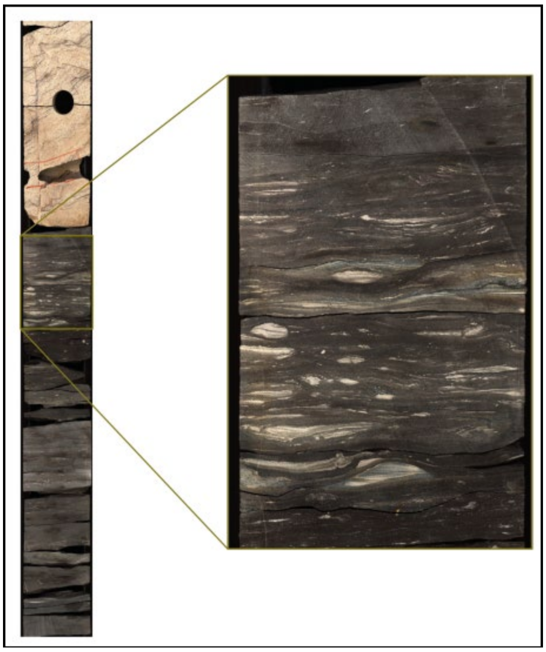

Figure 2. High resolution core image.

3. Core image features extraction

Firstly, we want to prepare some image features from the core images, to model the petrophysical properties quantitatively. Hence, we pre-process and analyze the core images using computer vision (CV) techniques.

Feature extraction is a core component of CV, both in traditional image processing, as well as the more recent deep learning frameworks. The traditional CV techniques have developed several approaches to extract image features from digital images or videos, so that a set of salient representations of the object of interest can be quantitatively captured. The features are chosen in a way sensitive to the target variables, to gain high-level understanding and perform model building and subsequent prediction. For instance, the color models provide simple yet useful characteristics about a rock description from image, represented in terms of the red/green/blue (RGB) color channels, as shown in **Figure 3**. In addition, the color model can be converted to the hue/saturation/brightness (HSV) model, to emphasize different characteristics of the same image, as shown in **Figure 4**. In addition, image entropy such as the Shannon entropy can be computed as a statistical measure of randomness that can be used to characterize the texture of the input core image.

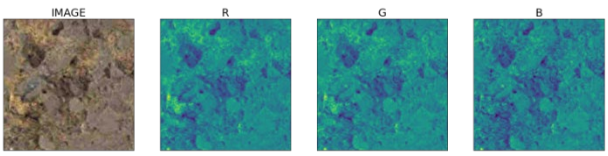

Figure 3. RGB color representation of a core image. (left) original core image in white color; (R/G/B) separated color channels of core image, R=red, G=green, B=blue.

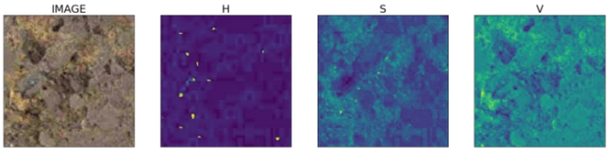

Figure 4. HSV color representation of a core image. (left) original core image in white color; (H/S/V) separated color channels of core image, H=hue, S=saturation, V=value/brightness.

The petrophysical properties can also correlate to the texture of the core, thus it is desirable to further include the core image texture representation in the analytic. We represent the core image texture through the Haralick texture features [4]. These features are computed from the Gray Level Co-occurrence Matrix (GLCM), which counts the co-occurrence of neighboring gray levels in an image [5]. Haralick then introduced 14 statistics calculated from the GLCM to describe the texture of an image. Typically, the maximum correlation coefficient is not calculated due to computational instability, so this feature is skipped in our analysis. The core images are partitioned into patches before calculating the GLCM and extracting the Haralick features from each image patch. **Figure 5** shows an example of Haralick texture features extracted from the core image. Although the Haralick textures are not straightforward to human, they are very informative to the machine learning data analytics.

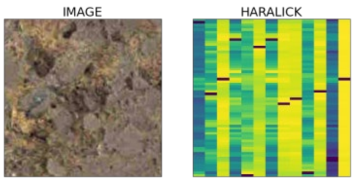

Figure 5. Haralick texture features a core image. (left) original core image; (right) Haralick texture features at each depth strips, each column represents one calculated coefficient in sequence of depth.

Each one of the image features described above are then aggregated horizontally into a profile in depth, resampled and aligned with well logs and core analysis data. To eliminate potential local artifacts in the core image, such as plug holes or markers, each feature depth profile is smoothed with box-car windows that can cover those artifacts, as shown in **Figure 6**. The smoothed profiles are then used subsequently in the model building. An example of core images and extracted image features in depth profile is shown in **Figure 6**.

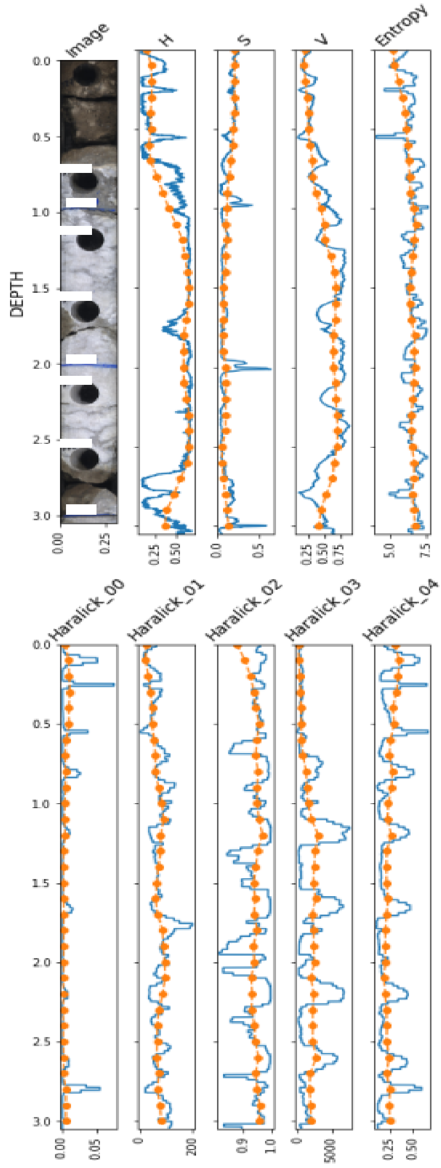

Figure 6. Example of core image (left) and extracted image features (right & bottom). Only a few of image features are shown here, from left to right: hue (H), saturation (S), value (V), Shannon entropy (Entropy), angular second moment (Haralick_00), contrast (Haralick_01), correlation (Haralick_02), sum of squares variance (Haralick_03), inverse difference moment (Haralick_04). The blue curves are the depth-wise aggregation of raw computation, and the orange curves are smoothed features profile for model building.

4. Machine learning modeling

We formulate the task of petrophysical property prediction as a supervised regression problem. The core image features described in the previous section and the well logs, such as core gamma log, are used as the input features for model building. The target properties from core analysis are used as the ground truth labels. In this paper, we investigate the prediction of three properties including grain density, porosity, and permeability.

The image features are first extracted from quality-checked and cleaned core images. Then the extracted features are combined with well log data to create the ensemble feature profile. The core analysis data is used as the ground truth label set. The features are fed into selected machine learning regression algorithms to train the corresponding models, test on new image data, and for performance evaluation. The overall workflow is shown in **Figure 7**.

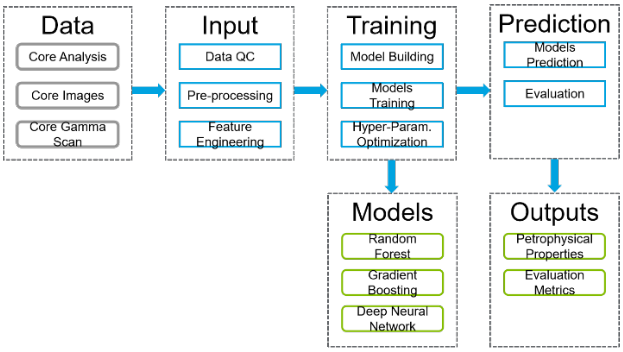

Figure 7. Workflow of AI-assisted petrophysical properties regression based on core images and well logs

4.1. Classification and Regression Tree

Classification and Regression Trees (CART) are among the most popular machine learning algorithms given their intelligibility and simplicity [6]. CART is a decision tree-based prediction model. Each decision tree is a set of internal (non-leaf) nodes and leaves. Each internal nodes are associated with an input feature, and the selected feature is used to make decision on how to divide the data set into two separate sets with certain criteria. Each leaf of the tree is labeled with a class or regression function, and the algorithm walks through the tree by making decisions at each node from input features and arrives at the corresponding leaf to compute the classification or regression. The procedure is shown as the sub-trees in **Figure 8**(a).

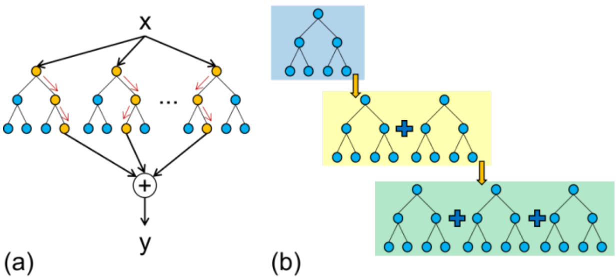

Figure 8. Diagram of classification and regression trees (CART) based regression models: (a) Random Forest; (b) Gradient Boosting.

In this work, two ensemble techniques are used to build the regression models, which effectively construct more than one CART. The random forest (RF) technique is one type of bootstrap aggregating or bagging ensemble, which builds multiple weak CART with repeatedly resampled training data and averages the trees for a consensus prediction [7]. The gradient boosting (GB) technique incrementally builds an ensemble by training new weak CART to minimize the residual of previous iteration [8]. **Figure 8**(a)&(b) shows the diagrams of both RF and GB algorithms.

4.2. Deep Neural Network

Among the recent advances of machine learning, deep learning (DL) has attracted broad attention [9]. DL technology is based on deep neural networks (DNN), which are networks composed of layers of artificial neurons [10]. The neuron receives input signals from other neurons or external sources, sums them up with weights in transfer function, and produces output with non-linear activation function. The model is trained by learning optimal weights from training data.

In this work, the feed-forward network topology is adopted to model the petrophysical properties. The network consists of multiple layers of neurons, and connections between neighboring layers. The neurons in each layer are shown in **Figure 9**, and the last layer has a linear activation to output the target variable. Three individual networks are trained to model each petrophysical properties, grain density, porosity, and permeability respectively.

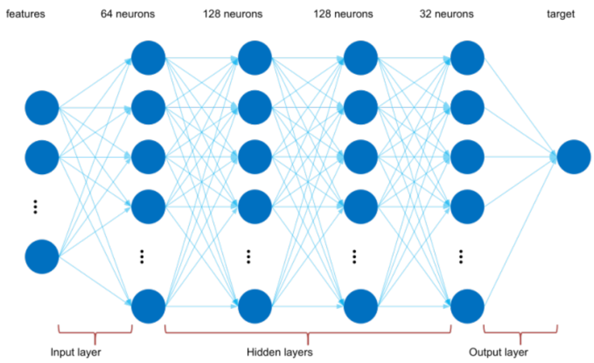

Figure 9. Diagram of deep neural networks. The image features and well logs data are fed as input to the network. The network consists of 4 hidden layers and output one target petrophysical property. Individual networks are trained to model each petrophysical property.

5. Field example

We tested the workflow of AI-assisted petrophysical properties prediction in a field from Saudi Arabia. The workflow is described in the previous section. The machine learning models used in this study are Random Forest, Gradient Boosting (GB) and Deep Fully Connected Neural Network (NN).

We plot the cross plots of the predicted against the actual values for each petrophysical properties **Figure 10**, to visualize the performance of the model prediction. All values are normalized to [0, 1] for the visualization. A guideline of y=x is plotted in a red line to represent a perfect match. The prediction results from each model (NN, RF, GB) are reasonably close to the ground truth.

The grain density has a small dynamic range and almost discrete values related to rock types. The porosity has a good distribution, and all models can capture the property very well. The permeability is more challenging to predict due to weaker correlation between the physics of the rock and the image. Despite that, the models are still able to capture the overall trend with slightly higher level of variances than those of density and porosity prediction.

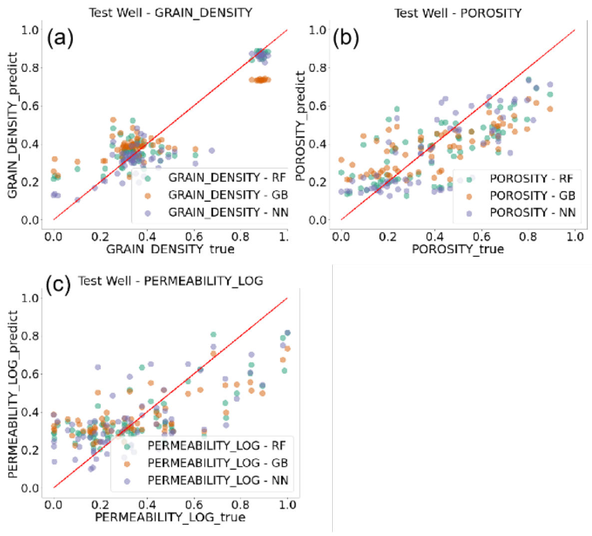

Figure 10. The cross plot of predicted against actual petrophysical properties. The target properties are: (a) grain density; (b) porosity; and (c) permeability (logarithmic). The regression algorithms are color-coded: (1) random forest as green dots; (2) gradient boosting as orange dots; and (3) deep neural network as purple dots. The red line represents the guide of y=x. All values are normalized to [0, 1].

The depth tracks of predicted and actual petrophysical properties are shown in **Figure 11**. All values are normalized to [0, 1] for the visualization, and the depth is represented in sample index. In grain density estimation, NN and RF models' predictions have very good performance and similar accuracy, while GB model prediction is also reasonable. In porosity estimation, all three models can track the trend well, while NN performs slightly better on capturing low porosity and RF performs slightly better on capturing high porosity. In permeability estimation, the prediction error is also reasonable with more challenging on lower and higher ends, while RF is slightly better at the average trend, and NN and GB are slightly better at providing the overall correlation between the prediction and ground truth.

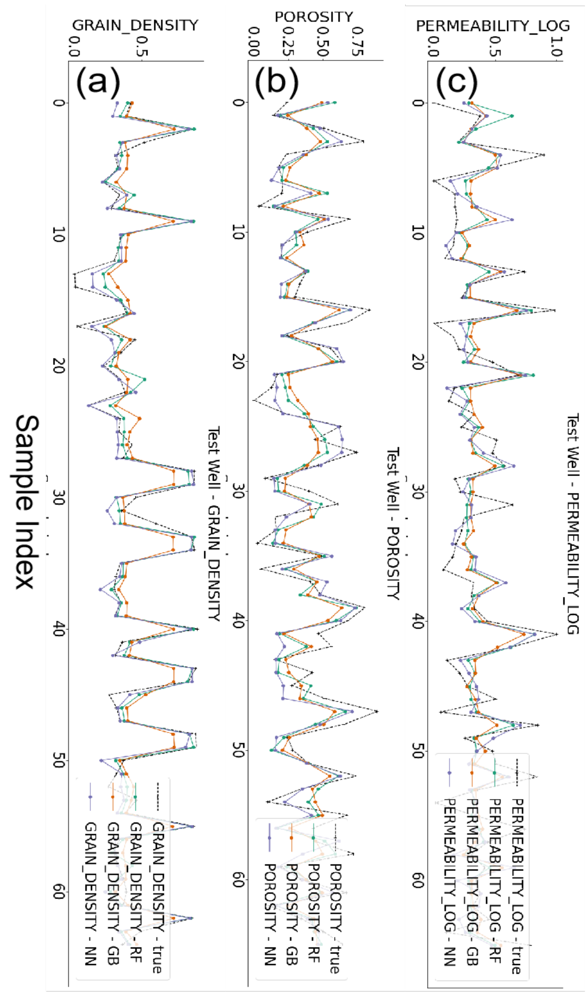

Figure 11. The depth track of predicted and actual petrophysical properties. The target properties are: (a) grain density; (b) porosity; and (c) permeability (logarithmic). The regression algorithms are color-coded: (1) random forest as green lines; (2) gradient boosting as orange lines; (3) deep neural network as purple lines; and (4) ground truth labels as black dash lines. All values are normalized to [0, 1].

We also derived the feature importance from RF models respecting to each petrophysical properties, as shown in **Figure 12**. The feature importance describes which features are relevant in the modeling, where the relative scores are effective. It can help domain experts with better understanding of the core characteristics, by providing insights into the data set and models. The relative scores can highlight which features may be more relevant to the target property and model building, and which features are less relevant. Such knowledge can be interpreted by petrophysicists and data scientists as the basis for gathering more or different data. We can also improve the predictive models by selecting more significant features using feature importance. Such feature selection can simplify the modeling problem, speed up the modeling process, and sometimes improve the model performance. In this case study, we note that grain density is more relevant to hue, which is correlated to the rock type. For example, shale rocks often appear darker in color than sandstones or limestones, where correlates to distinguishable grain density. Meanwhile, porosity and permeability are relevant to both image features and textures, especially some Haralick textures, due to the complex correlations between core images and physical properties.

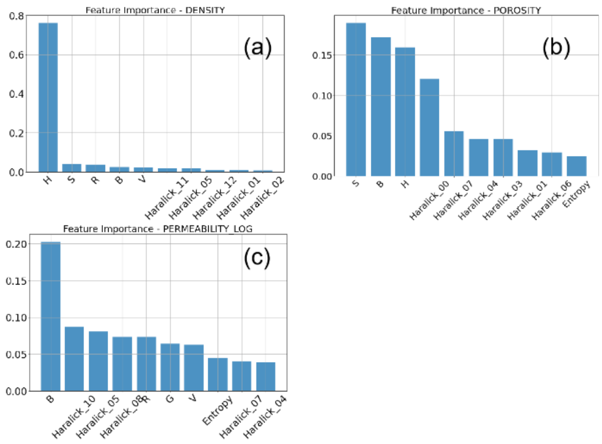

Figure 12. The feature importance of modeled petrophysical properties by Random Forest. The target properties are: (a) grain density; (b) porosity; and (c) permeability (logarithmic). The features are sorted by their importance to the prediction, and only the top 10 most important features are shown.

The overall performance is summarized in Table 1. In the table, y denotes the ground truth measurements and $\tilde{y}$ denotes the model prediction. The R^2 denotes the coefficient of determination and is defined by:

$$R^2 = 1 - \frac{\sum(y - \tilde{y})^2}{\sum(y - \text{avg}(y))^2}$$

The full-scale error (FSE) is defined by:

$$\text{FSE} = (y - \tilde{y})/\max(y)$$

And the zero-normalized cross-correlation (CC) is defined by:

$$\text{CC} = \frac{1}{n}\frac{\sum(y - \text{avg}(y))(\tilde{y} - \text{avg}(\tilde{y}))}{\sqrt{\sum(y - \text{avg}(y))^2}\sqrt{\sum(\tilde{y} - \text{avg}(\tilde{y}))^2}}$$

Table 1. Evaluation metrics of all models to target petrophysical properties

Target	Model	R^2	FSE	CC
Grain Density	NN	70.4%	8.4%	86.0%
	RF	68.3%	8.7%	82.6%
	GB	56.6%	12.2%	77.5%
Porosity	NN	40.6%	14.3%	69.2%
	RF	40.7%	15.1%	68.1%
	GB	37.8%	16.0%	69.2%
Permeability (LOG10)	NN	33.2%	16.5%	62.9%
	RF	37.0%	16.4%	61.5%
	GB	34.4%	16.9%	62.2%

In this case study, NN has overall slightly better performance in terms of metrics of accuracy and correlation, while RF and GB also provide reasonable prediction. However, it should be noted that prediction accuracy depends on many factors including feature engineering, hyper-parameter tuning, and training vs. validation vs. testing data splitting.

6. Conclusion

We have demonstrated that it is feasible to integrate core images and well logs into quantitative petrophysical characterization workflow, with the assistance of computer vision and machine learning technology. We successfully applied the workflow on data from several heterogeneous reservoirs in Saudi Arabia. The core images are preprocessed and analyzed with CV techniques to extracted features as a depth profiles. The features are aligned with core gamma logs and plug analysis. Image features such as RGB and HSV provide more information related to lithology while entropy and Haralick provide more information regarding rock textures. The integrated features vector is imported in ML models building and training with difference ML regression algorithms, such as CART and DNN. The models produce high-resolution profiles of petrophysical properties that can be very useful for up-scaling. We evaluated the importance of each feature from CART model for individual target, to provide petrophysical insights of core characteristics and help in future model building. We also evaluated the model performances and achieved reasonably good correlation and accuracy from the models.

The AI-assisted workflow leverages machine learning techniques to automate the process of core data analysis, reduces human effort and cost, and provides fast core petrophysical properties prediction. The approach enables core digitization and core preservation, and reduces exploration risk, as well as provides petrophysical insights to assist core description. Machine leaning algorithms add value to business cases with workflow automation and improved efficiency.

References

1. F. J. Lucia, *Carbonate Reservoir Characterization* (Springer-Verlag Berlin, Heidelberg, 2007)
2. C. Xu, Z. Heidari, C. Torres-Verdín, SPE ATCE (2012)
3. A. Gonzalez, L. Kanyan, Z. Heidari, L. Olivier, Petrophysics, **61** 495–518 (2020)
4. R. M. Haralick, K. Shanmugam, I. Dinstein, IEEE Trans. on Systems, Man, and Cybernetics **SMC-3, No.6**, 610-621.(1973)
5. M. Partio, B. Cramariuc, M. Gabbouj, A. Visa, Norsig2002 (2002)
6. M. Krzywinski, N. Altman, Nat Methods **14**, 757-758 (2017)

7. L. Breiman, Machine Learning **45(1)**, 5–32 (2001)
8. J. H. Friedman, Computational Statistics and Data Analysis **28**, 367-378 (2002)
9. J. Jiang, R. Xu, S. C. James, C. Xu, SPE Res Eval & Eng **24**, 250–261 (2021)
10. Y. LeCun, Y. Bengio, G. Hinton, Nature **521**, 436–444 (2015)