Combining high-resolution core data and machine learning schemes to develop sustainable core analysis practices

Christophe Germay[1,*], *Tanguy* Lhomme[1], *Luc* Perneder[1], and *John* Cummings[2]

[1]EPSLOG, Liege, Belgium
[2]Stratum Reservoir, UK

Abstract. Traditional core analysis methods rely on extensive rock testing programs and require large numbers of plug samples. Numerous stakeholders compete for intact core material and often do not take the time to base their selection of samples on objective and reliable information. As a consequence, samples dedicated to core analysis programs consume a significant fraction of the material available and their selection is based on very little *a-priori* information. In an attempt to change this paradigm, we promote the CoreDNA workflow: a more sustainable approach of core analysis with the objective to optimize the value of cores by increasing the quantity of high-quality data extracted while keeping the analyst footprint to a minimum. The first step of this approach consists of combining several non-destructive tests to produce a multi-disciplinary set of continuous profiles of rock properties, as early as possible in the core analysis workflow, without causing any irreparable damage to the cores. Using fully mobile technology, continuous high-resolution and high-fidelity rock property data can be generated, immediately upon barrel-opening, prior to modification by slabbing and plugging. The technique can also be successfully mobilised on legacy core datasets. Multi-disciplinary data tracks ranging from textural and colour features of the rock, to grain size distribution statistics, elemental concentrations, elastic wave velocities and rock strength are generated. These tracks are integrated under a unique format and used as a road-map to guide core analysts during more complex and expensive discipline-specific tasks such as SCAL or RMT. These cost-efficient, non-destructive tests are designed to produce high-density multidisciplinary databases, providing data for targeted follow-on studies by not only core analysts but also geologists and petrophysicists. This data acquisition philosophy is completed by high-end data analytics using the latest developments in Machine Learning, with the objective to build and properly train predictive models for geological and petrophysical descriptions. Predictive AI models need to be trained with labelled data. CoreDNA provides multidisciplinary input data and output data constrained by rock/lithofacies, physical rock properties, elemental geochemistry data and possible diagenetic features, all concisely labelled and linked to high-resolution depth-referenced ultra-high resolution photographs. We describe cases in which supervised machine learning has been applied to CoreDNA analysis, and how this has improved the application of multidisciplinary core data in reservoir characterisation studies. The compatibility of CoreDNA data to support the classification of rocks with a convolutional neural network has been established. This represents a first step towards an Artificial Intelligence framework for the identification and classification of rock facies. The deployment of the iterative; multistage approach described above to real-world examples from extractive industries will further refine this novel approach. Such a coordinated effort will require extensive resources to create unambiguous labels for training sets. In turn, this will clear the path and lay the foundation for the building of robust AI models dedicated to the task of predicting rock properties from high quality, non-destructive core data.

1 Introduction

In a recent publication [1] we presented an integrated core analysis solution combining transdisciplinary, high resolution, non-destructive measurements on whole cores, for an early yet objective description of cores and the rapid estimation of formation properties.

Results of these fast tests are analysed real-time and turned into high resolution, continuous profiles of properties (petrophysical, geomechanical and geochemistry). This data is fed into (unsupervised) machine learning algorithms for the automated identification of lithofacies, the design of fit-for-purpose plug selections and the programing of subsequent steps in core analysis programs.

This unsupervised machine learning scheme enables the identification of a number of facies utilising a cluster analysis which establishes 'groupings' of data points with similar physical, visual and elemental characteristics. This approach does not presently enable sorting into pre-existing classes. To achieve this requires the development of an iterative, multi-stage approach, involving supervision, which is the focus of this paper.

* Corresponding author: tanguy.lhomme@epslog.com

2 Methodology

2.1 Workflow

Acquire CoreDNA data: CoreDNA data acquisition and analysis methods produce the input data to be fed into AI models. These cost-efficient, non-destructive tests are designed to produce high-density multidisciplinary databases, forming the input for predictive expert geologist and petrophysicist systems.

Construct a predictive model using a neural network: Predictive AI models need to be trained with labelled data. This implies the creation of training data sets for which CoreDNA input data is available together with output data, *i.e.* highly detailed geological descriptions with facies, diagenetic features and other geological markers already identified and labelled in ultra-high resolution photographs, together with accurate petrophysical properties.

Optimize the performance of the predictive model: Creating such a training set is a highly resource-intensive task and should be anticipated as an iterative improvement cycle. The data set used to train the AI model is meant to grow over time, as more and more CoreDNA input data become available along with corresponding standardized geological labels. This continuous optimization ensures that the AI model prediction performances increase with the number of projects.

2.2 Data acquisition and integration

As a core data acquisition principle, CoreDNA combines transdisciplinary, high-resolution, non-destructive measurements for a rapid and objective description of cores and the quantitative estimation of formation properties. Core samples, which may be whole cores still in their half-open liners, or fractions of legacy cores, are scanned with multiple sensors interfaced to one single table-top equipment. These smart sensors share the same depth reference and compatible resolution ranges. The complete CoreDNA test series includes the following tests:

- Portable XRF measurements for elemental composition,
- High-resolution photo for panoramic viewing of the cores, under white and UV lights,
- Laser scan for surface rugosity and grain properties
- Ultra-high-resolution (UHR) photos (1.8µm/pixel) for the rock texture and grain properties,
- Probe permeability,
- P- and S-wave ultrasonic velocity logging,
- Strength log from the scratch test,
- Grain size and grain fabric data.

CoreDNA data sets are designed to be used as a single source, unified format platform to unravel the complexity of tested cores. The basic analysis concept for such multi-dimensional data sets involves data checking and augmentation routines that prepare for unsupervised and supervised machine learning schemes.

2.3 Supervised Machine Learning

The relevance of CoreDNA data to train neural networks was established with a dedicated study. The findings are summarized in this section.

Data set: This study uses ultra-high-resolution photos together with physical data such as elemental compositions from pXRF, grain size, and strength, acquired on a series of 27 outcrop samples (17 sandstones and 10 carbonates), each of 20cm in length, with known labels (Figure 1 and Table 1).

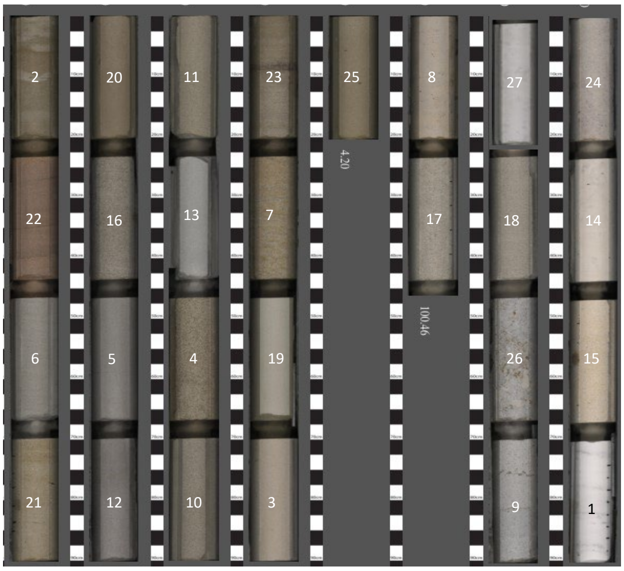

Figure 1: Panoramic photos of the 27 outcrop samples used to train a supervised machine learning scheme for rock classification purposes.

Table 1: Class labels

Class	Label
1	Alabama Marble
2	Bandera Brown
3	Bentheimer
4	Berea Buff
5	Berea Gray
6	Berea Sister Gray
7	Boise
8	BonneTerre Dolomite
9	Burlington (Carthage Marble)
10	Carbon Tan
11	Castlegate
12	Colton

13	Crab Orchard
14	Edwards White
15	Edwards Yellow
16	Idaho Gray
17	Indiana High Perm
18	Indiana Low Perm
19	Kentucky
20	Kirby
21	Leapord
22	Nugget
23	Parker
24	Rocheron
25	Sans Saba
26	Silurian Dolomite
27	Wisconsin

Using standard augmentation techniques [2] the number of 600x600 pixel photos of samples was brought to 479. The multidimensional data set was completed with 16 CoreDNA vectors including strength, mean grain size derived from laser topographical maps, and elemental compositions from XRF measurements (Figure 2, Figure 3, Table 1, Table 2).

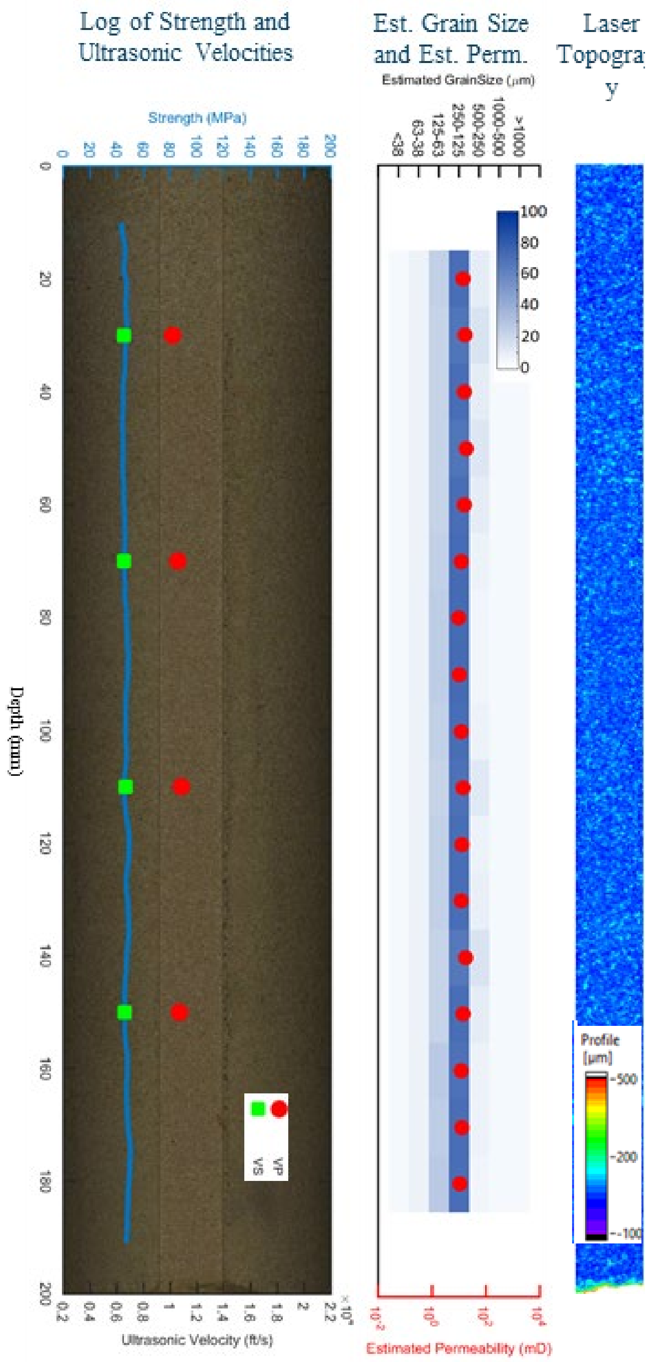

Figure 2: CoreDNA data for the San Saba outcrop sample: Strength and ultrasonic velocity profiles, high resolution

panoramic core photo, rugosity map and permeability index estimated from grain size distributions.

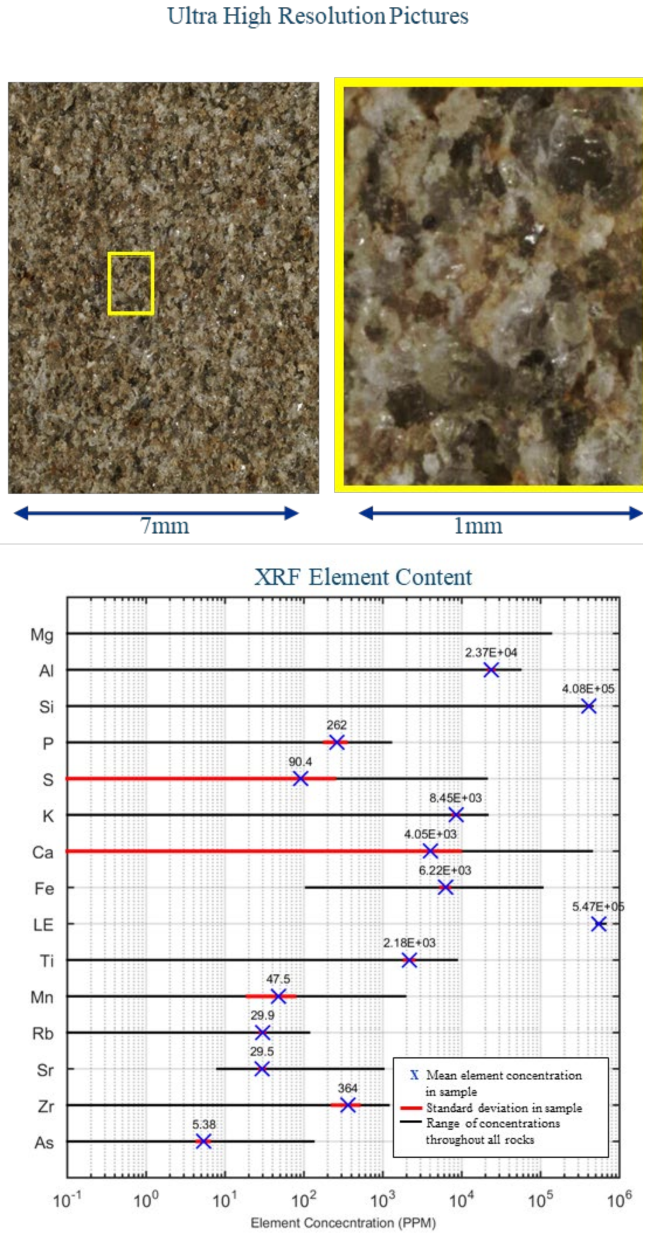

Figure 3 : CoreDNA data for the San Saba outcrop sample: high and ultra-high definition photographs; elemental composition from XRF measurements.

Table 2 : CoreDNA data for the Sans Saba outcrop sample: geomechanical properties.

Test Results		Rock Strength [MPa]	Ultrasonic Wave Velocity [ft/s]	
			Vp	Vs
Sample A	Mean	45	11,053	7,044
	STD	3	582	87
	Min	37	10,572	6,944
	Max	53	11,581	7,118
Sample B	Mean	47	10,572	6,568
	STD	2	280	72
	Min	44	10,192	6,518
	Max	50	10,842	6,671

Table 3: CoreDNA data for the Sans Saba outcrop sample: petrophysical properties.

Test Results		Est. Grain Size [mm]	Est. Perm. [mD]
Sample A	Mean	144.2	12.5
	STD	5.8	3.3
	Min	135.3	8.0
	Max	154.5	18.9
Sample B	Mean	146.4	13.7
	STD	4.2	2.5
	Min	139.6	9.9
	Max	153.9	18.5

This data set was split in training, validation and test sets according to the proportion shown in Figure 4.

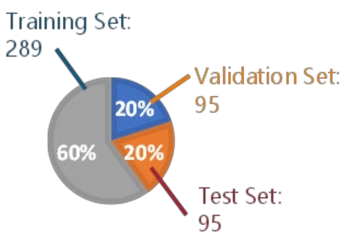

Figure 4: Data set proportions for neural network training and validation.

Convolutional Neural Network: A version of the convolutional neural network EfficientNet pre-trained with 14 million images from the ImageNet database corresponding to 20.000 categories, was used for transfer learning [3].

Results. After training the neural network with 60% of the multi-dimensional input data set (Figure 5), we tested its ability to match tested rock with existing labels.

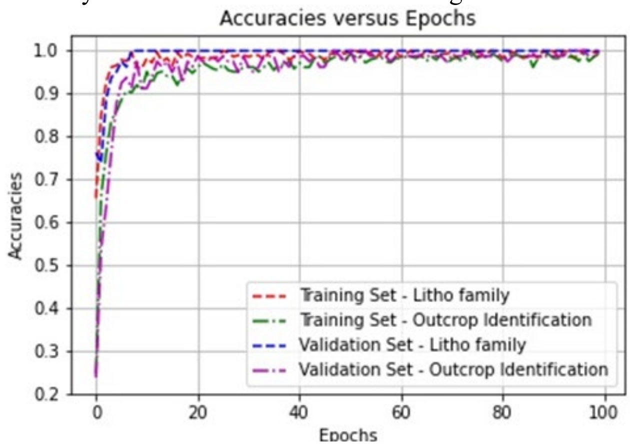

Figure 5: Neural network training phase.

The current classification performances of the network reach an accuracy of 99%, as seen in Figure 6 showing the numbers of actual samples in the test set which were properly identified for each class.

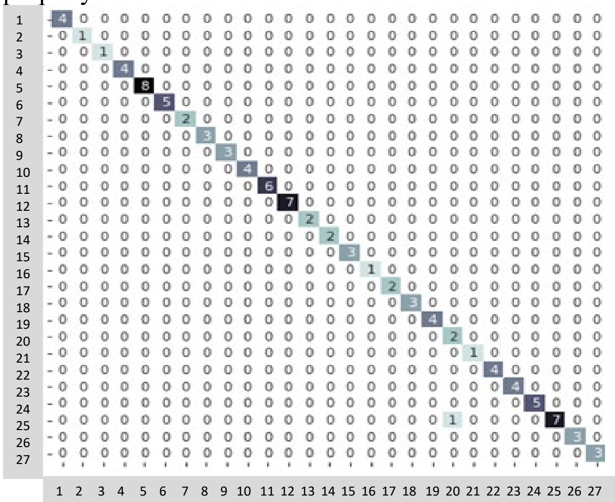

Figure 6: Actual vs predicted labels by the trained EfficientNet convolutional neural network.

Only one data point (corresponding to the Sans Saba rock sample) out of 95 was mislabelled and confused with another rock (Kirby) with very similar physical properties and visual aspect. This represents an exciting first step towards executing CoreDNA analysis, and in particular, ultra-high resolution image analysis, with the aim of automated identification and classification of rock types.

2.4 Optimizing the performance of the predictive model

Creating such a training set is a highly resource-intensive task and should be anticipated as an iterative improvement cycle described in Figure 7.

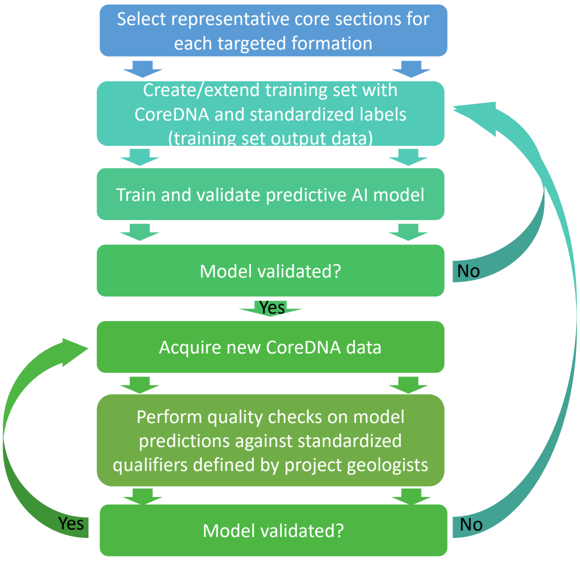

Figure 7: Iterative AI model improvement loop.

The value of the multistage approach described above comes from the significant reduction of the effort required by the quality checks to validate the AI model predictions compared to the requirements for the extension of the training set. The deployment of such an approach to real-world examples from extractive industries will require significant resources to create extensive sets of digital core data and strictly repeatable standards, if this data is to be transformed into unambiguous digital labels used in training sets. In an attempt to enhance the coherence and reduce the subjectivity of man-made observations and interpretations, which can often be applied in conjunction with objective physical data, we designed a software tool dedicated to the registering of such information into shared multidisciplinary databases.

2.5 LithoLog

LithoLog software is a product developed with input from expert sedimentologists, with the goal of digitalising sedimentological information using standardized qualifiers to minimise subjectivity and inconsistency in the description process, while also generating robust, fully quantitative sedimentological data at a scale normally reserved for detailed petrographic analysis/optical microscopy. It is designed to simplify navigation through large multiscale datasets composed of (i) core data logs with centimetre resolution, (ii) panoramic high-resolution pictures covering 1m (3 feet) of core sample, and (iii) ultra-high resolution pictures of each centimetre of core. The software ergonomics and accessibility have been prioritized to create a user-friendly platform, suitable for a diverse range of disciplines/contributors, thereby accelerating the creation of large databases for AI model training.

Digital core description: The main features of LithoLog software include visualisation panels for high (35µm/pixel) and ultra-high (1.8µm/pixel) resolution pictures. The resolution enables detailed sedimentological core description, which provides hugely important information and data on the fabric and texture of reservoir rocks, while providing input into the interpretive elements of core description such as depositional process *and environment assignments.*

Standardized qualifiers: Functionalities are built in the software so that users can edit primary lithology and qualifiers at a centimetre scale. The introduction of a single set of standardized qualifiers shared among multiple users is taking one step towards standardized workflows and methodologies for core descriptions, which are lacking compared to RCA/CCA/SCAL type studies.

Visualizing Grain size distributions: Graphic tools are included to visualize and perform quality checks on grain size distribution data from CoreDNA laser scans for entire cored sections, with a resolution of a centimetre. A UHR photo (1.8µ per pixel) is loaded with a digital grain size overlay, with 1 image per cm. The digital grain size overlay can be selected from pre-defined Wentworth-scale grainsize bins [4] until the circles match visible grains. Individual grain diameter can be measured with micron scale accuracy, meaning even poorly sorted sediments can be accurately quantified (Figure 8).

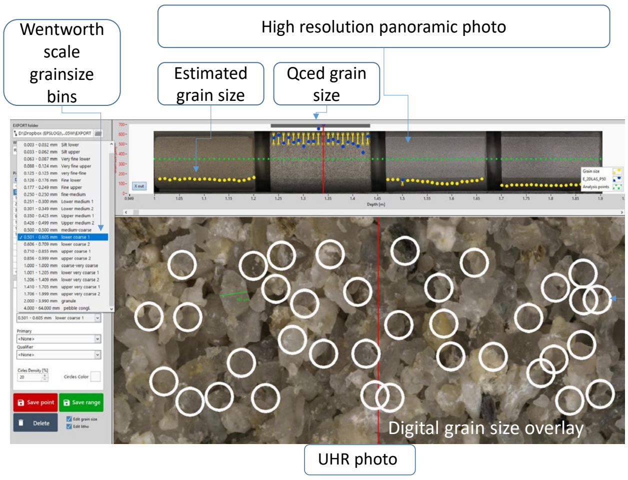

Figure 8: Grain size distributions visualisation and quality checking in LithoLog.

Third-party data visualization: Data can be imported from any *.las or *.csv file, ensuring applicability across industry. This provides further assistance to select appropriate qualifiers to label the formation under study.

3 Conclusions

CoreDNA is a technology that can be applied on fresh cores as soon as barrels are opened but also on legacy cores. Its main objective is to assist core analysists with continuous, high-resolution profiles of high fidelity properties measured directly on cores, before the core has been permanently modified by slabbing and plugging. For this, we create multi-disciplinary data tracks, which can be all integrated under one unique format and used as a road-map to guide core analysists during more complex and expensive discipline-specific tasks such as SCAL or RMT.

CoreDNA arrays of transdisciplinary, high-resolution, non-destructive tests were deployed as an integrated core analysis package in conjunction with unsupervised and supervised machine learning schemes.
The compatibility of CoreDNA data to support the classification of rocks with convolutional neural network was established. The development of an Artificial Intelligence framework for the identification and classification of rock facies is a piece of work in progress. Next steps will require the deployment of the iterative, multistage approach described above to real-world examples from extractive industries. Such a coordinated effort will require extensive resources to create unambiguous labels for training sets. In turn, this will clear the path and lay the foundation for the building of robust AI models dedicated to the task of predicting rock properties from high quality non-destructive core data.

References

[1] C. Germay, T. Lhomme and P. Bisset, “Combining high-resolution core data with unsupervised machine learning schemes for the identification of rock types and the prediction of reservoir quality” in The 34th International Symposium of the Society of Core Analysts, 2021.

[2] C. Shorten and T. Khoshgoftaar, “A survey on Image Data Augmentation for Deep Learning” Journal of Big Data, vol. 6, no. 60, 2019.

[3] M. Tan and Q. V. Le, “EfficientNet: Rethinking Model Scaling for Conventional Neural Networks” in International Conference on Machine Learning, Long Beach, 2019.

[4] C. K. Wentworth, “A Scale of Grade and Class Terms for Clastic Sediments” The Journal of Geology, 1922.