

Machine learning in petrophysics: Advantages and limitations

Chicheng Xu^{a,*}, Lei Fu^a, Tao Lin^a, Weichang Li^a, Shouxiang Ma^b

^a *Aramco Americas: Aramco Research Center, Houston, United States*

^b *Saudi Aramco, Saudi Arabia*

ABSTRACT

Machine learning provides a powerful alternative data-driven approach to accomplish many petrophysical tasks from subsurface data. It can assimilate information from large and rich data bases and infer relations, rules, and knowledge hidden in the data. When the physics behind data becomes extremely complex, inexplicit, or even unclear/unknown, machine learning approaches have the advantage of being more flexible with wider applicability over conventional physics-based interpretation models. Moreover, machine learning can be utilized to assist many labor-intensive human interpretation tasks such as bad data identification, facies classification, and geo-features segmentation out of imagery data.

However, the validity of the outcome from machine learning largely depends on the quantity, quality, representativeness, and relevance of the feeding data including accurate labels. To achieve the best performance, it requires significant effort in data preparation, feature engineering, algorithm selection, architecture design hyperparameter tuning, and regularization. In addition, it needs to overcome technical issues such as imbalanced population, overfitting, and underfitting.

In this paper, advantages, limitations, and conditions of using machine learning to solve petrophysics challenges are discussed. The capability of machine learning algorithms in accomplishing different challenging tasks can only be achieved by overcoming its own limitations. Machine learning, if properly utilized, can become a powerful disruptive tool for assisting a series of critical petrophysics tasks.

1. Introduction

Machine learning (ML) is a powerful tool that has made significant business impacts in the oil and gas industry. Thus, its applications in petrophysics have gained popularity in recent years. One can easily find many ML related publications in conferences and journals (Xu et al., 2019). Some petrophysical problems are intrinsically routed with ML solutions such as rock typing, facies classification (Hall, 2016; Zhang and Zhan, 2017), and missing log prediction (Akinnikawe et al., 2018; Singh et al., 2020; Tokpanov et al., 2020). Most papers present successful cases of using one or more ML algorithms to solve some challenging petrophysical problems based on various types of data. However, few papers discuss the limitations of using ML and conditions required to ensure its successful applications. As it says that every sword has two edges, ML should not be touted as a perfect tool that solves all challenges. Instead, ML as a new tool practiced in the petrophysical technical community, if not done properly, may risk of being misused and generate seemingly good but erroneous results (Cuddy 2021).

In this paper, we first discuss the advantages of ML in performing some petrophysical task scenarios, then address its limitations and application conditions. In addition, we will emphasize what caution we need to exercise when using ML to maximize its benefits in petrophysical interpretation.

2. Advantages of machine learning in petrophysics

2.1. Well structured petrophysical data

The biggest advantage originates from the petrophysical data itself being well structured and understood. Petrophysical data acquired either from laboratory core analysis or downhole logging tools has well defined physical models. As such, many well-developed machine learning algorithms can be applied on petrophysical data. The results from machine learning models can also be verified. Xu et al. (2019) categorized petrophysical data by its source, type, index, and dimension. The most common datasets handled by machine learning are depth indexed core measurements and well logs. The data types and dimension may be different depending on data acquisition objectives, as exemplified below.

- Categorical data: e.g. petrophysical rock types, facies, fluid types
- Irregularly sampled numerical data: e.g. routine and special core analysis on core plugs
- Regularly sampled numerical (1D curves): most well logs are recorded in 1D numerical format such as gamma ray, neutron porosity, bulk density, and resistivity.

* Corresponding author. Aramco Americas, Houston, Texas, United States.

E-mail address: Chicheng.xu@aramcoamericas.com (C. Xu).

<https://doi.org/10.1016/j.aiig.2022.11.004>

Received 15 November 2022; Received in revised form 30 November 2022; Accepted 30 November 2022

Available online 16 December 2022

2666-5441/© 2022 The Authors. Publishing services by Elsevier B.V. on behalf of KeAi Communications Co. Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

- Arrays or waveforms: data from some advanced logging sensors are recorded in arrays or waveforms such as nuclear magnetic resonance, nuclear spectroscopy, and dipole sonic waveforms.
- Images or maps: petrography such as thin sections and scanning electronic microscope (SEM) images, borehole image logs. 2D maps can be generated from nuclear magnetic resonance (NMR) log and sonic log (Anemangely et al., 2019).

2.2. Machine learning answers

As its adoption in petrophysics quickly grows, ML has proven to be a powerful tool for a wide range of problems. Despite an incredible amount of excitement generated by in various fields, ML is not a silver bullet that can solve all problems. In many cases, ML solutions may not be optimal solutions due to the nature of the problem or the quantity and quality of the available data. Before applying ML on any petrophysical problem, we may want to ask ourselves whether ML is necessary or cost effective?

Regardless, whatever model or method used, the aim of using ML is to find an answer to assist in business decision making. Sometimes, the answer does not need to be perfect to work, especially when considering the uncertainties of acquired data and data representativeness (Ma and Amabeoku, 2015). However, when a physical model is too complex or unavailable, problems may not be solvable from physics point of view. In this case, ML, as a data-driven tool, can usually provide an answer with quantified uncertainty (Khan et al., 2018; Basu et al., 2020; Chen et al., 2022). As it says, all models are wrong, but some are very useful. If the ML answer is proved to be good to assist in business decision making, it is useful.

2.3. Machine learning automation for productivity and consistency

Another big advantage of using ML is the enhanced efficiency performed by machine instead of human, such as in data quality check and correction, data labeling, grouping (Liu et al., 2021) and well log correlation (Brazell et al., 2019; Bakdi et al., 2020). Machine works relying on mathematics-based models while human interpretation is based on knowledge, experience, and skills. Thus, results from ML are more consistent, while products delivered by human may be subjective and biased. Consequently, for some tedious works, machine can finish much faster than human, with more consistent results. Akkurt et al. (2018) presented a case study of 59 wells using ML for accelerating and enhancing petrophysical analysis. The reported ML answers match the experts' work in 87% of wells for bulk density logs, 90% for compressional slowness logs and 88% for shear slowness logs. It was also claimed that automation of the workflow significantly improved project time efficiencies by providing a structured means to evaluate large, varying and complex datasets, which could potentially liberate the Petrophysicist from spending the bulk of the time on repetitive work.

2.4. Deciphering high-dimensional data

High-dimensional data such as arrays, images, waveforms, and 3D volumes are challenging to the capability of human recognition. Even with the assistance of modern 3D visualization, it is still hard for human to accurately label and interpret high-dimensional data. For example, it is nearly impossible to pin-point different minerals on a high-resolution thin section image or trace every pore on a 3D volume CT scan of a rock. In these cases, we must resort to machine to find a solution.

3. Limitations of data/model when applying machine learning

3.1. Data correctness and representativeness

Data acquired may be accurate, but could still be unrepresentative (Ma and Amabeoku, 2015). Reasons causing data to be unrepresentative

are many, including tool calibration, tool resolution, samples acquisition, and physical sample alteration before testing. For both ML and human working, if data is not representative to the targeted problem, the produced answers would be unreliable. ML techniques can be used to identify outlier data (Akkurt et al., 2018; Misra et al., 2019) and make corrections to the identified data with inputs from subject matter experts. However, if uncorrected non-representative data is used in ML modeling, it is not expected that machine would recognize this data quality issue, which will affect the modeled results. It is suggested that erroneous outliers be removed and all necessary corrections should be applied to the data before using them in the machine learning model.

3.2. Data consistency

Data consistency is another issue faced by both physical modeling and ML approach. Data acquisition settings may vary significantly from well-to-well and from field-to-field. Technology also evolves rapidly over time, thus tools used for data acquisition may be different from the same service provider or from different service providers. Borehole environments may change with advancement of drilling and drilling fluids. Geology always varies vertically and laterally. With field development, reservoirs are also undergoing fluid transitions (gas-oil-water) and changes in geo-mechanics with pore pressure changes. Correction and normalization of all these factors are difficult, but necessary to ensure data quality and consistency from different wells or locations.

3.3. Data size, quality, and relevance

Most ML algorithms need the so called "big data" to train the model. When data is relevant and representative, it might be true to say "the bigger the better". However, in petrophysics domain, the data acquired are often small; it is typical to have measurements on a few hundreds of core plugs for certain rock properties for example. When data gets bigger by accumulating more data from different wells or even different fields, the data quality deteriorate significantly as those data may not be measured by following the same protocol, as discussed above in data consistency. Geological conditions may also have been changed across wells and fields, therefore the relevance becomes low for the data acquired far from the studied area. Sometimes, better results can be achieved by using a small-sized but a highly relevant training dataset to build a simple model. Fig. 1 shows a field example of cross-plotting neutron porosity vs. bulk density. As the number of wells increases from 1 well to 6 wells, more outlier data emerge and the correlation between the x and y parameters becomes poorer.

3.4. Accuracy of data labeling

The ML model trained by labeled data is strongly impacted by the accuracy of the labeled data. If the labeled data is inaccurate, the ML model will be misled. For example, core data are often considered as ground truth to calibrate well log interpretation or to train well log based ML model. If the core measurements are mishandled and give inaccurate results (e.g., permeability measurements without applying correct reservoir stresses), this inaccuracy will be transferred to the well log based model as well. Machine cannot tell whether the labeled data from core is accurate or not. Geologists label facies based on their own experience and the labeling process can be biased. Therefore, the ML model trained by the human labeled data will also show the same bias.

3.5. Model applicability and generalization

In petrophysics analysis, most empirical models have their own very specific applicability, such as in a field or a zone or for some rock types, such as using Archie equation for formation evaluation in clean and shaly formations. Similar limitations also apply to ML model. If the training data is from a specific field or zone or rock type, most likely the

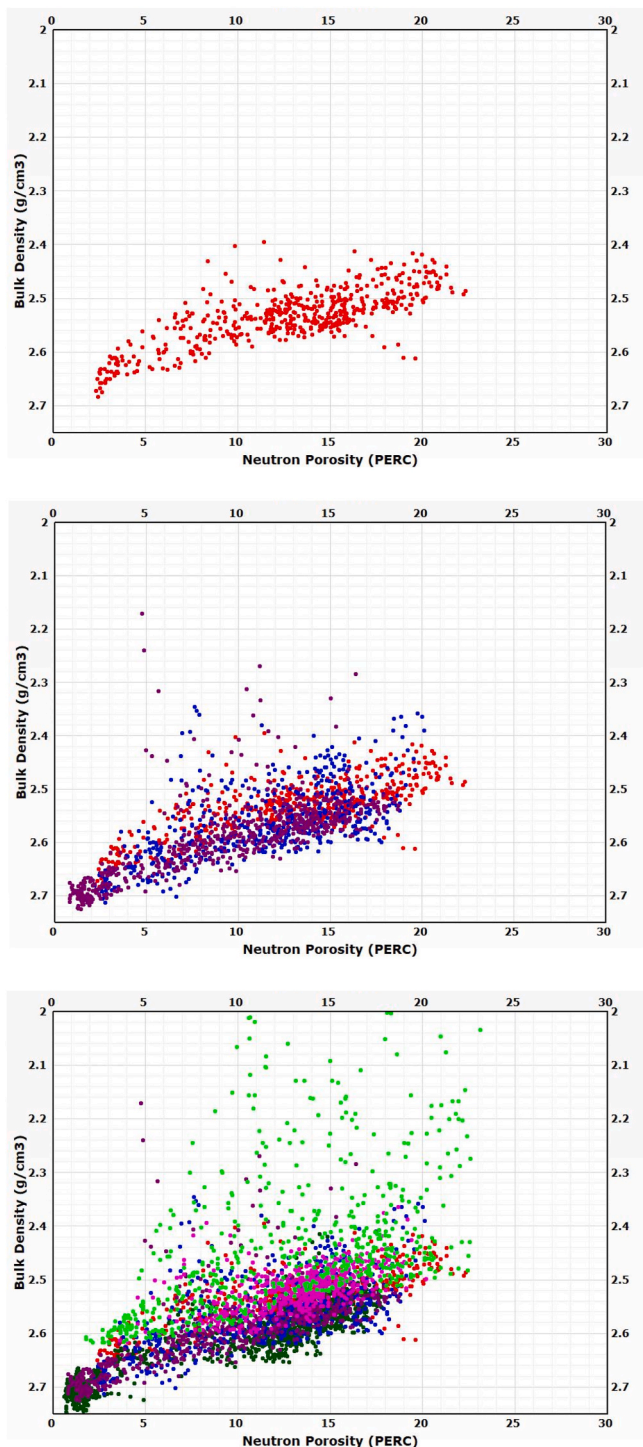


Fig. 1. An example of quality of correlating neutron porosity with bulk density deteriorates with increasing number of wells in the same field. Top panel: one well; Mid panel: three wells; Bottom panel: six wells.

ML model only applies to the same field/zone/rock type. It is hard for most ML models to be generalized out of the boundaries of the training data.

3.6. Machine learning still statistics based

Most ML algorithms are still based on statistics which shows averaging effects in regression. In certain regression tasks, ML models may fail to capture extreme variabilities. For example, in the 2021 ML contest

organized by the Society of Petrophysicist and Well Log Analysts (SPWLA) (Yu et al., 2021), all five teams predicted shear slowness (DTS) to be much higher than the log measurement in the fastest zone as highlighted in Fig. 2.

4. Suggestions for applying ML in petrophysics

4.1. Role of physics

The importance of physics cannot be underestimated even with machine learning approach. This is particularly true when the data size is small. Physics informed or guided ML potentially performs better, thus has gained much more attention recently (Liu et al., 2021). Understanding the physics, if possible, behind the data should always be helpful in correcting and/or normalizing data or extracting representative rock attributes (Xu et al., 2012). In many cases, forward physical modeling can be used to generate large volume of synthetic data for training or verifying the machine learning models.

4.2. Keep it simple

When one claims that a deep neural network with millions of parameters can solve a petrophysical problem, it may be against the basic interpretation rule; i.e., rule of simplicity or Occam's razor. When the data size is not big enough, overfitting becomes a common issue, if too many unknowns are intended to be solved. Therefore, it is also suggested that the machine learning model be as simple as possible. Applying sophisticated deep learning model on a small petrophysical dataset may lead to seemingly good but erroneous results.

5. Incorporation of human expertise and knowledge

Petrophysical interpretation heavily relies on human expertise and knowledge. An experienced petrophysics interpreter often sees a lot of features behind the noisy and complex dataset such as in borehole image interpretation. Machine learning assisted petrophysical interpretation workflows can benefit from human expert inputs and guidance at multiple stages of the project, including data quality control and improvement, data labeling, model selection and verification. More importantly, advanced machine learning model should be able to assimilate human expertise and accumulate a knowledge database from multiple projects.

6. Machine learning - a powerful complementary technology to assist petrophysicists

In many cases we reviewed from the literature, ML approves to be an alternative technology powerful enough to change business practices, but remains complementary and needs to be calibrated to physics-based model. Being complementary and alternative means that ML can solve what physical model can solve, and more efficient in cases of big data. However, it may not show huge advantage over conventional physical models in term of accuracy.

For example, in the 2022 SPWLA ML contest (Fu et al., 2021), ML is used as an alternative to solve multiple mineral composition model. The training data is from 10 wells that have results calculated with conventional multiminer solver. Therefore, we cannot verify the model performance beyond the conventional multiminer solver. The best case is that ML can achieve the same accuracy level as the conventional multiminer solver. For a practical field study, one can apply the same multiminer model to newly drilled wells for formation evaluation. In this case, ML may work as an alternative approach, without expecting extra advantage.

On the other hand, ML can potentially bring disruptive changes to some petrophysical interpretation work scenarios. For example, an intelligent outlier detection algorithm can screen out the major portion of bad data from thousands of wells in a short moment (Akkurt et al.,

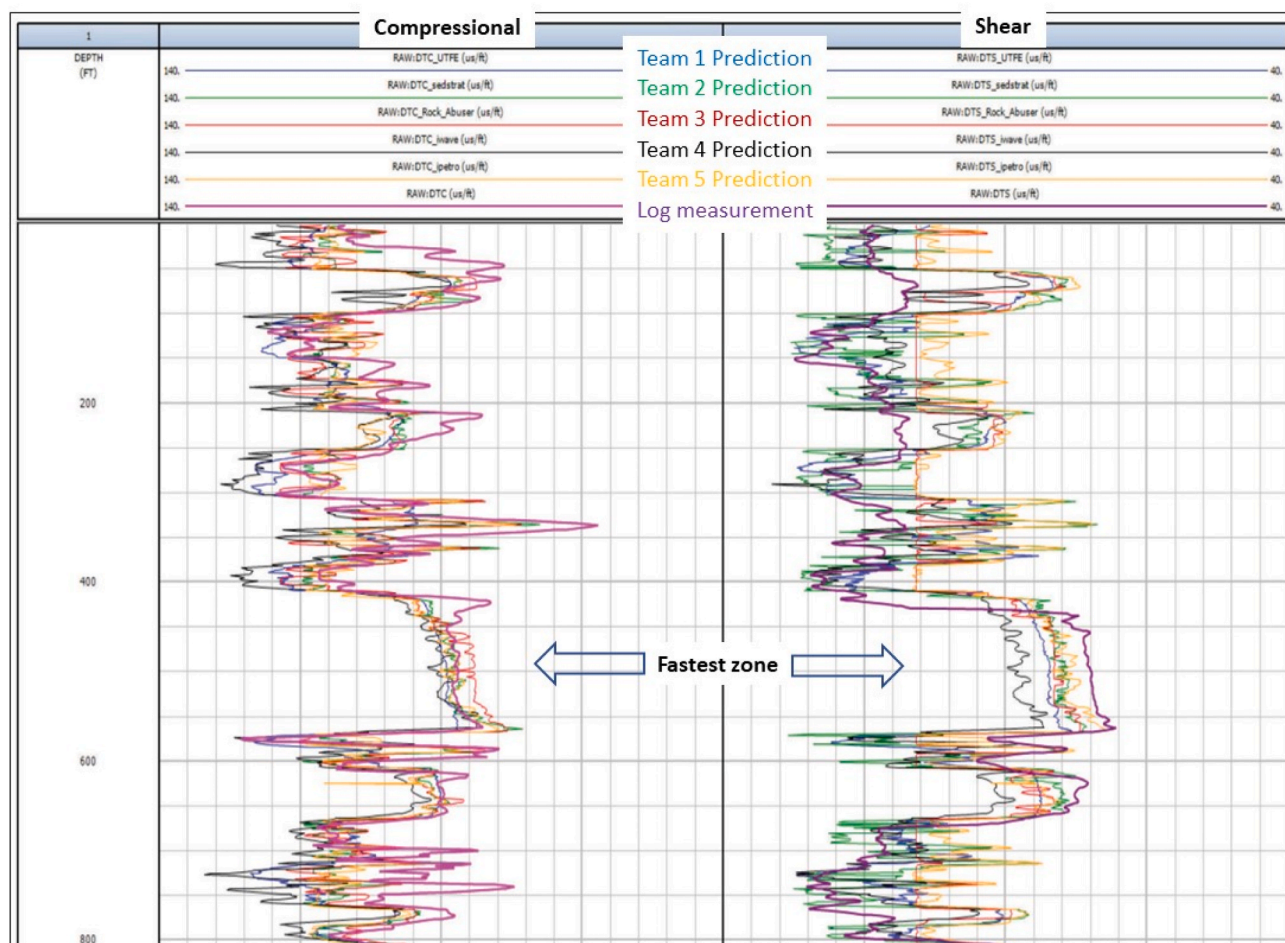


Fig. 2. Machine learning prediction performance of DTC (compressional) and DTS (shear) by five teams in the 2021 SPWLA Machine Learning contest (Yu et al., 2021).

2018). Another example may be that a well-trained convolutional neural network can be used to automatically segment geological features including vugs and fractures out of huge volume of imagery data quickly (Liu et al., 2022). We believe more disruptive applications will emerge by leveraging deep learning to achieve automation in interpreting high-dimensional data, such as borehole image analysis, evaluation of cement quality and pipe corrosion (Viggen et al., 2020; Xu et al., 2022).

6.1. Summary

From the above description and discussions, the following are summarized;

- With demonstrated benefits of using machine learning in petrophysical operations, there is a need for practitioners to have an in-depth understanding of both advantages and limitations of the technology so that we can better harness machine learning as an effective tool to solve petrophysics challenges.
- As a rule of thumb for best practices, we need to carefully query the data size, quality, representativeness, and relevance and use the optimal training data set to build machine learning models.
- Best efforts need to be committed to understand the physics behind the data even though operational issues can be very complex.
- Once machine learning is determined to be used, significant effort in data preparation, feature engineering, algorithm selection, architecture design and hyperparameter tuning need to be committed to achieve the best performance.

- In addition, technical issues, such as imbalanced population and overfitting, need to be handled carefully to avoid erroneous answers.
- It is mutually understood that know-hows and best practices on achieving the best petrophysical interpretation using machine learning should be shared and documented with field applications.

Declaration of competing interest

We submit the manuscript titled as *Machine Learning in Petrophysics: Advantages and Conditions* to the AI in Geosciences journal. We declare the manuscript will not be considered publishing on other journals and all authors have no conflict of interest.

References

- Akinikawe, Oyewande, Lyne, Stacey, Roberts, Jon, 2018. Synthetic well log generation using machine learning techniques. In: SPE/AAPG/SEG Unconventional Resources Technology Conference. OnePetro.
- Akkurt, Ridvan, Conroy, Tim T., Psaila, David, Paxton, Andrea, 2018. Jacob Low, and Paul Spaans. "Accelerating and enhancing petrophysical analysis with machine learning: a case study of an automated system for well log outlier detection and reconstruction. In: SPWLA 59th Annual Logging Symposium. OnePetro.
- Anemangely, Mohammad, Ahmad, Ramezanzadeh, Amiri, Hamed, Hoseinpour, Seyed-Ahmad, 2019. Machine learning technique for the prediction of shear wave velocity using petrophysical logs. J. Petrol. Sci. Eng. 174, 306–327.
- Bakdi, Sara, Kannan, Nitish, Reddy Masini, Shashipal, Chennakrishnan, Balaji, 2020. Automated well correlation using machine learning and facial recognition techniques. In: Abu Dhabi International Petroleum Exhibition & Conference. OnePetro.
- Basu, Saptaswa, Johansen, Kyrre, Sergey Skvortsov, 2020. Petrophysical workflows vs machine learning—A comparative case study in the Dakota group, Williston basin. In: SPWLA 61st Annual Logging Symposium. OnePetro.

- Brazell, Seth, Bayeh, Alex, Ashby, Michael, Burton, Darrin, 2019. A machine-learning-based approach to assistive well-log correlation. *Petrophysics-The SPWLA Journal of Formation Evaluation and Reservoir Description* 60, 469–479, 04.
- Chen, S., Shao, W., Sheng, H., Kwak, H., 2022. Use of Symbolic Regression for Developing Petrophysics Models. paper SPWLA 2022 0113, Stavanger, Norway, 15 June.
- Cuddy, Steve, 2021. The benefits and dangers of using artificial intelligence in petrophysics. *Artific. Intelligence Geosci.* 2, 1–10. <https://doi.org/10.1016/j.aiig.2021.04.001>. ISSN 2666-5441.
- Fu, Lei, Xu, Chicheng, Yu, Yanxiang, Ashby, Michael, McDonald, Andy, Dai, Bin, 2021. Well-Log-Based Reservoir Property Estimation with Machine Learning: A Tutorial for the PDDA 2021 ML Contest. *SPWLA Today Newsletter*. November.
- Hall, Brendon, 2016. Facies classification using machine learning. *Lead. Edge* 35 (10), 906–909.
- Khan, Mohammad Rasheed, Tariq, Zeeshan, Abdurraheem, Abdulazeez, 2018. Machine learning derived correlation to determine water saturation in complex lithologies. In: *SPE Kingdom of Saudi Arabia Annual Technical Symposium and Exhibition*. OnePetro.
- Liu, Hui-Hai, Zhang, Jilin, Liang, Feng, Temizel, Cen, Basri, Mustafa A., Mesdour, Rabah, 2021. "Incorporation of physics into machine learning for production prediction from unconventional reservoirs: a brief review of the gray-box approach". *SPE Reservoir Eval. Eng.* 24, 847–858. <https://doi.org/10.2118/205520-PA>.
- Liu, Yubo, Liao, Guangzhi, Xiao, Lizhi, Liang, Zhen, Zhang, Jiawei, Zhang, Xinyu, Zhang, Zhe, Zhou, Jun, Guojun, Li, 2022. Automatic Fracture Segmentation and Detection from Image Logging Using Mask R-CNN" Paper Presented at the SPWLA 63rd Annual Logging Symposium, Stavanger, Norway. <https://doi.org/10.30632/SPWLA-2022-0115>. June.
- Viggen, Erlend Magnus, Merciu, Ioan Alexandru, Løvstakken, Lasse, Måsøy, Svein-Erik, 2020. Automatic interpretation of cement evaluation logs from cased boreholes using supervised deep neural networks. *J. Petrol. Sci. Eng.* 195, 107539 <https://doi.org/10.1016/j.petrol.2020.107539>. ISSN 0920-4105.
- Ma, Shouxiang, Amabeoku, Maclean, 2015. Core analysis with emphasis on carbonate rocks- quality assurance and control for accuracy and representativeness. *SEG Interpretation Journal*, V. 3 (1) <https://doi.org/10.1190/INT-2014-0072.1>.
- Misra, Siddharth, Osogba, Oghenekaro, Powers, Mark, 2019. Unsupervised outlier detection techniques for well logs and geophysical data. *Mach. Learn. Subsurface Charact* 1.
- Singh, Maniesh, Makarychev, Gennady, Mustapha, Hussein, Voleti, Deepak, Akkurt, Ridvan, Daghar, Khadija Al, Ahmed Mawlod, Arwa, et al., 2020. Machine learning assisted petrophysical logs quality control, editing and reconstruction. In: *Abu Dhabi International Petroleum Exhibition & Conference*. OnePetro.
- Tokpanov, Yury, Smith, Johanna, Ma, Zheren, Deng, Lichi, Benhallam, Wassim, Salehi, Amir, Xiang, Zhai, Darabi, Hamed, Castineira, David, 2020. Deep-learning-based automated stratigraphic correlation. In: *SPE Annual Technical Conference and Exhibition*. OnePetro.
- Xu, Chicheng, Torres-Verdin, Carlos, 2012. Saturation-Height and Invasion Consistent Hydraulic Rock Typing Using Multi-Well Conventional Logs. Cartagena, Colombia. Paper presented at the SPWLA 53rd Annual Logging Symposium, June.
- Xu, Chicheng, Misra, Siddharth, Srinivasan, Poorna, Ma, Shouxiang, 2019. When petrophysics meets big data: what can machine do?. In: *Paper Presented at the SPE Middle East Oil and Gas Show and Conference*. Manama, Bahrain. <https://doi.org/10.2118/195068-MS>. March.
- Xu, Chicheng, Li, Weichang, Abudu, Olanrewaju, Ma, Shouxiang, 2022. System and method using machine learning for automated casing or tubing inspection log interpretation related application. *US Patent Appl.* 16, 63/175738, April.
- Yu, Yanxiang, Xu, Chicheng, Misra, Siddharth, Li, Weichang, Ashby, Michael, Pan, Wen, Deng, Tianqi, Jo, Honggeun, Santos, Javier E., Fu, Lei, Wang Chengran, Kalbekov, Arkhat, Suarez, Valeria, Kusumah, Prasetya, Epo, Avilandito, Mohammad, Pamadya, Yogi, Izadi, Hossein, 2021. Synthetic sonic log generation with machine learning: a contest summary from five methods. *Petrophysics* 62, 393–406. <https://doi.org/10.30632/PJV62N4-2021a4>.
- Zhang, Licheng, Zhan, Cheng, 2017. Machine learning in rock facies classification: an application of XGBoost. In: *International Geophysical Conference, Qingdao, China, 17-20 April 2017*. Society of Exploration Geophysicists and Chinese Petroleum Society, pp. 1371–1374.