Computational Geosciences (2022) 26:1151–1165
https://doi.org/10.1007/s10596-022-10151-9

ORIGINAL PAPER

Application of machine learning to characterize gas hydrate reservoirs in Mackenzie Delta (Canada) and on the Alaska north slope (USA)

Leebyn Chong[1,2] · Harpreet Singh[3,4] · C. Gabriel Creason[5,6] · Yongkoo Seol[3] · Evgeniy M. Myshakin[1,2]

Received: 11 July 2021 / Accepted: 2 May 2022 / Published online: 14 May 2022

Abstract
Artificial neural network-trained models were used to predict gas hydrate saturation distributions in permafrost-associated deposits in the Eileen Gas Hydrate Trend on the Alaska North Slope (ANS), USA and at the Mallik research site in the Beaufort-Mackenzie Basin, Northwest Territories, Canada. The database of Logging-While-Drilling (LWD) and wireline logs collected at five wells (Mount Elbert, Iġnik Sikumi, and Kuparuk 7–11–12 wells at ANS, plus 2L-38 and 5L-38 wells at the Mallik research site) includes more than 10,000 depth points, which were used for training, validation, and testing the machine learning (ML) models. Data used in training the ML models include the well logs of density, porosity, electrical resistivity, gamma radiation, and acoustic wave velocity measurements. Combinations of two or three out of these five well logs were found to reliably predict the gas hydrate saturation with accuracy varying between 80 and 90% when compared to the gas hydrate saturations derived from Nuclear Magnetic Resonance (NMR)-based technique. The ML models trained on data from three ANS wells achieved high fidelity predictions of gas hydrate saturation at the Mallik site. The results obtained in this study indicate that ML models trained on data from one geological basin can successfully predict key reservoir parameters for permafrost-associated gas hydrate accumulations within another basin. A generalized approach for selecting a well log combination that can improve model accuracy is discussed. Overall, the study outcome supports earlier work demonstrating that ML models trained on non-NMR well logs are a viable alternative to physics-driven methods for predicting gas hydrate saturations.

Keywords Machine learning · Gas hydrates · Well-logs · Neural network · NMR

1 Introduction

Machine learning (ML) is an effective data-driven approach for both regression and classification analysis of nonlinear systems. The systems can contain thousands of variables constituting a massive dataset for training with a random subset of data dedicated for independent validation. ML is particularly useful to assess problems and phenomena to which theoretical understanding is not complete and where empirical correlations and approximations are required. In the recent years, the application of ML to geoscience problems has become a rapidly emerging field [1–3]. An important category of geoscience problems where ML can contribute is spatio-temporal estimation of physical parameters that are difficult to monitor directly and/or require sophisticated techniques for interpretation of measured data. Deep learning methods that use several hidden layers in artificial neural network (ANN) architectures have been utilized to extract complex and highly

✉ Evgeniy M. Myshakin
evgeniy.myshakin@netl.doe.gov

[1] National Energy Technology Laboratory, 626 Cochrans Mill Road, P.O. Box 10940, Pittsburgh, PA 15236, USA

[2] NETL Support Contractor, 626 Cochrans Mill Road, P.O. Box 10940, Pittsburgh, PA 15236, USA

[3] National Energy Technology Laboratory, 3610 Collins Ferry Road, Morgantown, WV 26507, USA

[4] NETL Support Contractor, 3610 Collins Ferry Road, Morgantown, WV 26507, USA

[5] National Energy Technology Laboratory, 1450 Queen Avenue SW, Albany, OR 97321, USA

[6] NETL Support Contractor, 1450 Queen Avenue SW, Albany, OR 97321, USA

non-linear features from geoscience data [4, 5]. Supervised ML applies statistical learning to identify patterns in data and then makes predictions based on inferred patterns [6]. The term "supervised" refers to a set of samples where both the target output parameter (signal) and the predictive variables (e.g., well logs and derivatives of those logs) are known, such that the ML model is initially "supervised" by the known input and output before it is used to predict unknown targets. The data are split into a training set that serves to train a ML model against the known target parameter, and a testing set that represents an unbiased set of samples to assess the predictive performance of the trained model. In this study, supervised ML is applied to predict key gas hydrate reservoir parameters in permafrost-associated accumulations on the Alaska North Slope (ANS) and Northern Canada.

Gas hydrate is a crystalline ice-like compound where the gas molecules are included within a hydrate lattice comprising H-bonded water cages. Methane hydrate is widely distributed in permafrost areas and marine sediments. Recent U.S. Department of Energy and its partners research suggests that methane hydrate is a promising future source of energy [7]. In recent years, a number of field-scale drilling and testing programs were conducted at the Mallik research site in Northwest Canada [8], at the Mount Elbert [9], the Iġnik Sikumi [10], and the Kuparuk 7–1-12 [11] sites in Northern Alaska, in the eastern Nankai Trough offshore Japan [12], in the Bay of Bengal offshore India [13, 14], in the South China Sea offshore China [15], and in the Gulf of Mexico offshore the United States [16]. These programs were successful in confirming the technical viability of gas production from gas hydrate reservoirs through depressurization, understanding site-specific reservoir petrophysical parameters, and details of the geological settings necessary to develop geological models for reservoir simulations. The petrophysical properties of gas hydrate reservoirs were inferred through seismic surveys, Logging-While-Drilling (LWD) and wireline well logs, as well as depressurized and preserved core measurements. The production potential of a gas hydrate reservoir is mainly determined by reservoir porosity, permeability, and gas hydrate saturation (S_{gh}) in pore space in addition to initial pressure and temperature. S_{gh} is the parameter characterizing the amount of gas trapped in the hydrate lattice and that can be potentially released into a reservoir after the crystallographic structure of gas hydrate is decomposed. While most of those properties can be estimated directly in the field or in a laboratory using core samples, the evaluation of S_{gh} is a relatively complex process. S_{gh} can be estimated using three physics-based methods that are based on the following: (1) electrical resistivity logs and the empirical Archie's law [17, 18], (2) processing sonic logs of compressional and shear velocities in the acoustic velocity method, and (3) Nuclear Magnetic Resonance (NMR) and density well-logs in the NMR-density porosity method. Each of these three methods are briefly discussed below; a detailed overview of these methods including their approximations and limitations can be found elsewhere [19].

In the electrical resistivity method, gas hydrate acts as an electrical insulator and the gas hydrate-bearing sediments increases the resistivity of rock, which enables estimating S_{gh} using Archie's law [17, 18] that is used to assess mobile phase saturation in the pore space as follows:

$$S_w = \left(\frac{F^* R_w}{R_t}\right)^{\frac{1}{n}} = \left(\frac{a}{\phi^m}\right)^{\frac{1}{n}} \cdot \left(\frac{R_w}{R_t}\right)^{\frac{1}{n}} \quad (1)$$

where, R_t and R_w are log-measured resistivity of reservoir-rock saturated with all in situ fluids, and resistivity of connate water (without rock matrix), respectively. F is a formation factor, ϕ is porosity typically estimated from density logs, while a, m, n are empirical parameters: m is a function of rock cementation, n is a function of hydrate morphology, a is typically set to 1. Consequently, S_{gh} is calculated as $1 - S_w$.

For the acoustic velocity method, acoustic velocities (compressional and shear wave velocities) are used to estimate S_{gh} using relationships that are either empirical [20] or based on rock-physics effective medium theory [21]. The acoustic velocity model requires a knowledge of mineralogy and bulk modulus that must be acquired through other means, and an assumption of gas hydrate morphology (the way how gas hydrate is precipitated in pore space). The NMR-density porosity method is based on the analysis of the transverse magnetization relaxation time ($T2$) of the hydrogen atoms that can differentiate between the water in the aqueous phase and the water belonging to the gas hydrate crystal lattice. The NMR logging tool measures porosity filled with aqueous phase (ϕ_{nmr}), whereas the density porosity tool (ϕ_{den}) measures total porosity, such that the difference between ϕ_{den} and ϕ_{nmr} is used to estimate S_{gh} [22–24] as follows:

$$S_{gh} = \frac{\phi_{den} - \phi_{nmr}}{\phi_{den}} \quad (2)$$

These three physics-driven methods discussed above are based on certain assumptions and empirical parameters that are not always well constrained and such constraints associated with these methods can be overlooked [19]. On the other hand, S_{gh} can be estimated through a supervised ML method by using a combination of well logs that bear a footprint of gas hydrate presence in a reservoir without making assumptions about the pore morphology and without any laboratory-estimated empirical parameters. That implies that the estimation of S_{gh} through ML is akin to an indirect measurement whose accuracy is dependent on the accuracy of the logs measured directly in the field and the S_{gh} dataset assigned during the training of the supervised learning method. For instance, intervals of the wellbore, where its diameter (measured using

the caliper log) either exceeds or shrinks beyond a certain threshold, can affect well log readings depending on the investigation radius of the logging tool. In such a case the affected intervals are either assigned with corrections or removed from the original dataset.

Previous studies [19] presented a generalized approach to predict S_{gh} through ML using well log data. In that study, LWD well log data from two ANS wells (Mount Elbert [9] and Iġnik Sikumi [10]) were used to develop the ML models, and the NMR-derived S_{gh} along each depth was used as the target variable ("ground truth") in the supervised learning process. In that study, twelve different supervised ML algorithms belonging to five different classes were carefully examined for their ability to accurately predict the target variable and handle overfitting. Those classes were Ridge Regression [25] and its variants, Decision Tree [26] and its variants, k-Nearest Neighbor [27], ANN [28], and variants of Reduced Order Models [29, 30]. The results from that study revealed that Stochastic Gradient Descent Regression (SGDR) [31], the variant of Ridge Regression [25], and ANN [28] were the best algorithms with SGDR performance being more sensitive to the size of the training dataset. The accuracy of the best algorithms in predicting S_{gh} was ~84% that measured the accuracy of the fitted model in terms of the variation of predicted data set from its target values [19]; this methodology was adopted in another gas hydrate classification study [32] using unsupervised and supervised learning, where the accuracy increased up to 90% with supervised learning. However, the ML models presented in these two studies [19, 32] were investigated using data from sites within the same sedimentary basin, moreover, the same gas hydrate accumulation. Thus, the applicability of the ML models investigated in the previous studies [19, 32] to permafrost-associated gas hydrate accumulations in other basins remains unconfirmed.

The goal of this study was to investigate the capability of ML models to predict S_{gh} for permafrost-associated gas hydrate deposits using well log training datasets for five wells drilled in two sedimentary basins. Locations of the five wells, i.e. the Mallik 2L-38 and 5L-38 wells in Northwest Territories, Canada, and Mt. Elbert, Iġnik Sikumi, Hydrate-01 (the Kuparuk 7–11–12 Site) on ANS are depicted in Fig. 1. The Mallik gas hydrate field is located in Mackenzie Delta on the coast of Beaufort Sea in Northern Canada. The Mallik 2L-38 and 5L-38 wells penetrate unconsolidated and lithified gas hydrate-bearing sediments from Mackenzie Bay and Kugmallit Sequences. The gas hydrate accumulation occurs within pebbles, coarse-grained sand, and silt layers [33, 34]. ANS gas hydrate occurs in laterally continuous Tertiary sandstone and conglomeratic units within the Sagavanirktok Formation of the Eileen Gas Hydrate Trend. The Trend overlies the Kuparuk River, Prudhoe Bay, and Milne Point oil fields. At the Mount Elbert Site gas hydrate occurs in pore space of thinly interbedded sandy-rich to silt-rich sediments within two reservoirs (Units D and C) [35]. At the Iġnik Sikumi site, gas hydrate is present in pores of four sandstone units (the "C-2 sand", the "C-1 sand", the "D sand", and the "E sand") within the Tertiary Sagavanirktok Formation [10]. The Hydrate-01 stratigraphic test well drilled from the Kuparuk 7–11–12 Pad confirmed the occurrence of two high-quality reservoirs (lithological Unit B and Unit D) highly saturated with gas hydrate. The sidewall core analysis identified gas hydrate presence in sandy slit and silty sand sediments of those units [36].

The ML models were trained using the gas hydrate saturation values inferred by the NMR-density porosity method, which was considered as the "ground truth" and helped in assessing the accuracy of ML-based predictions. The electrical resistivity method was also utilized to qualitatively confirm the presences of gas hydrate at corresponding depth points. The ML models are intended to complement the existing physics-driven methods discussed above and serve as a tool to probe S_{gh} distributions at drilling sites with either a limited number of logs or compromised logs that preclude data from being interpretated using the conventional methods. One potential application of ML is using well logs available at hundreds of legacy wells located on ANS and in Northern Canada to predict gas hydrate saturation in those locations. The majority of those wells were drilled without the intention to assess S_{gh}, hence the ML method could utilize the available logs to indicate the presence of gas hydrate in those locations. Thus, S_{gh} estimated using ML at locations where no exclusive measurements of S_{gh} are available can extend our knowledge about gas hydrate occurrence in geological formations and provide data on a larger scale that can be a suitable resource for geological and reservoir studies.

2 Methodology

The Keras (a Python library) [37] in TensorFlow was utilized in this study to engage ANN to train ML models with a dataset comprising various well log readings at depth intervals with and without gas hydrate presence. In other words, the dataset includes non-gas hydrate-bearing units mainly comprised by shale as well as gas hydrate-bearing units mainly occurring in sandy-rich sediments. The workflow for ML model preparation is outlined in Supplementary Materials Fig. S1 and involves the following steps: (1) data pre-processing, (2) hyperparameter tuning, (3) well log combination optimization, and (4) model validation. Well-log montages displaying downhole log data collected in those wells can be found elsewhere [38–42]. The Mallik 2L-38 and 5L-38, and Iġnik Sikumi well log data are publicly available [33, 34, 43]. The Mt. Elbert data are available upon request [44]. Obtaining Hydrate-01 data [42] currently requires approval from both NETL and JOGMEC (Japan). The following features (or well

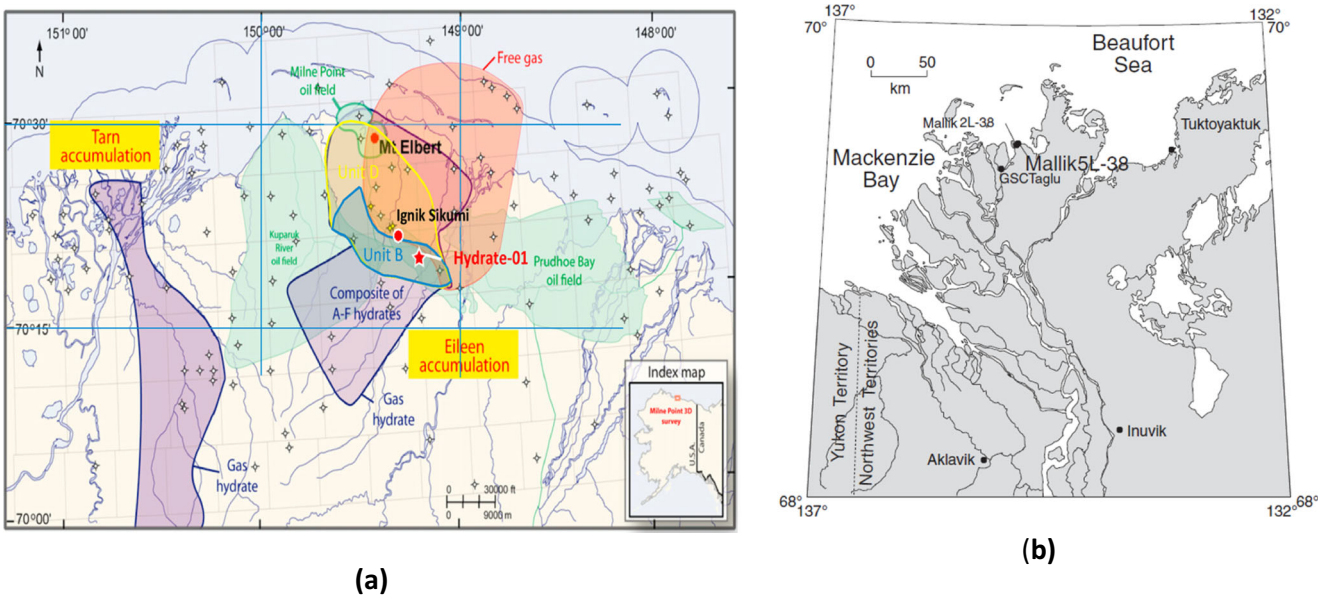

Fig. 1 **a** Location map of ANS regions showing Mt. Elbert, Iġnik Sikumi, and Hydrate-01 well locations in Eileen gas hydrate accumulation [35]. **b** Location map of the Mackenzie Delta region showing the Mallik 2L-38 and 5L-38 drilling sites [64]

logs) relevant to this study are selected: bulk density (ρ), density porosity (ϕ), gamma radiation (GR), resistivity (R_t), compressional acoustic wave velocity (V_p), and shear acoustic wave velocity (V_s). For each depth point, these six features correspond to a target variable (S_{gh}) known through the ground truth (NMR- derived gas hydrate saturations) and also desired as output of the ML-trained model. The well logs and NMR-derived S_{gh} for ANS sites (Mt. Elbert, Iġnik Sikumi, and Hydrate-01) are depicted in Fig. S2. Additionally, borehole diameters at different depths were available from caliper logs and used as part of the outlier removal process. If missing data was found in one well log feature, the values from the remaining log features were excluded from the analysis at that corresponding depth point.

Outlier removal was performed in a two-step manner: removal of data showing evidence of strong washouts and then removal of outliers based on the global-local outliers in subspaces (GLOSS) algorithm [45] analyzing all features in the dataset. Washouts indicate enlarged sections of a wellbore where the hole size is larger than the drill bit, which results in unreliable data readings from the borehole tools and thus compromised log data for that depth [46]. To mitigate washout effects, the data for each well was screened for large caliper values in the upper five percentile. The trimmed data was then scanned for outliers using the GLOSS algorithm that detects local subspace outliers using a global neighborhood search [45] by returning probability scores across features and depths instead of attempting to detect outliers within each feature separately. Table 1 lists the initial numbers of depth data points (each includes six features and a target value) and the numbers remaining after each step in the outlier removal procedure. A guiding principle for ANN implementation suggests that the training data size should be at least 30 times the number of features [47]; in this study, a training data size was defined by the number of well log features. Thus, for the six features the minimum size of the dataset should be 180. This is well below the sample size of the dataset used in this study, which is 5392 and 3915 samples for the wells in at the Mallik site and on ANS, respectively.

To assess the accuracy of prediction in the validation process, the metric consists of the coefficient of determination (R^2) or the accuracy score, which is calculated using Eq. 3. This accuracy score was used in our previous works [19], and has been used in other studies (e.g. assessing porosity distributions with ANN [48]).

$$R^2 = 1 - \frac{\sum_i \left(y_i - \hat{y}_i\right)^2}{\sum_i \left(y_i - \overline{y}\right)^2} \tag{3}$$

where y_i, $\hat{y}_i$, and $\overline{y}$ are the expected value of S_{gh} ("ground truth"), a predicted one, and average expected S_{gh}, respectively. The ideal R^2 value is 1 while any deviations of predicted values for expected ones result in a R^2 value below 1. The "ground truth" are taken as values of the NMR-derived S_{gh} which are confirmed by the electrical resistivity method. In other words, the electrical resistivity method should confirm the presence of gas hydrate at a particular depth if an NMR-derived S_{gh} value is accepted as "ground truth" at that depth. Using NMR-derived S_{gh} as the "ground truth" is a common practice used by other researchers to assess the accuracy of other methods, such as electrical resistivity and acoustic velocity methods [22, 49, 50].

Table 1 Numbers of depth data points for each well before and after outlier removal

Well	5L-38	2L-38	Mt Elbert	Iġnik Sikumi	Hydrate-01
Location	Mallik		ANS		
Initial Data	3120	2639	822	1127	2236
After Caliper Screening	2964	2507	778	1070	2124
After GLOSS algorithm	2920	2472	771	1053	2091

To account for the variability in performance of ML algorithms due to sample sizes used for training, cross-validation was performed using the *k*-fold method where *k*-1 sets are randomly chosen as training sets while a single set is used for validation. The *k*-fold cross-validation allows *k* different combinations of training and validation data and the average of accuracy scores from *k* different rankings are used to assess cross-validation. *k* between 2 and 5 was found to be sufficient to maintain model consistency in predictions. Generally, it is considered that 80/20 or 70/30 splits, where at least 80% or 70% of data are used for training, are sufficient to ensure the training of the ML models and provide highest accuracy in predictions. For this study, it was determined that 80/20 was an appropriate split based on similar studies that also employed well-logs with ML (for e.g. Song et al. [51]). The 20% of the data to be used for validation purposes was not fixed and could belong to different wells each time a ML model is trained. Furthermore, the effect of spatial location of the sample data was avoided by randomizing the data prior to using it in the ML model.

Each well log feature was normalized based on the minimum-maximum values of that feature, which is done to ensure the loss function is not adversely impacted by any large magnitude in the input data; specifically, normalization of the data helps make the computations relatively more efficient through faster convergence, but it may not necessarily improve the accuracy of the trained model. Various combinations of features were taken to train a ML model to predict the target variable. Before a ML model can be applied to a selected dataset, the hyperparameters of the algorithm must be tuned for optimal performance because, unlike the weights of the ANN, the hyperparameters are not optimized during the supervised learning process. More specifically, hyperparameters are non-trainable static parameters that affect the training procedure and a ML model quality, which means the tuning of hyperparameters must be done prior to training the ML model. To find an optimal set of hyperparameters, a global grid search method was used that entailed testing the accuracy of each discrete value combination of hyperparameters. The grid search method has been used for hyperparameter tuning for neural networks and compared to other search methods [52, 53]. For each set of hyperparameters, a ML model was trained and validated using all six features available from all wells. The Adam optimizer for the stochastic gradient descent algorithm and the mean squared error as the loss function were utilized [54]. Such choice is typical in the recently reported ML work in classifying wireline log shapes [51] and real-time well log predictions [55]. The neural network topology involves a number of hidden layers, where each layer is composed of several nodes (also called neurons), such that the nodes on each layer connect the layers per the activation function. Among different mathematical forms of activation function, it has been found that rectified linear unit (ReLU) works well for the vast majority of regression problems; therefore, ReLU is used for the hidden layers, while final output layer uses the linear activation function. After each hidden layer, a dropout layer was included in order to prevent overfitting. Dropout is a regularization method where nodes are randomly excluded during the training process to prevent certain nodes from dominating the prediction pattern. The frequency of node exclusion is called the dropout rate and it was set at 0.5 [56] in this study. The remaining hyperparameters are also tunable to ascertain the optimal training procedure: number of hidden layers, number of nodes per layer, learning rate of the optimizer, batch size, and epochs. Batch size refers to the interval of data points processed before the trainable parameters are updated, and epochs are the number of times the entire training set is used. We perform a broad tuning first with learning rates ranging from 0.0001 to 0.01, two to three layers, 10–50 nodes per layer, 100–500 batch size, and 100–500 epochs. After observing the hyperparameters that consistently lead to high R^2 on the testing data, fine tuning was performed by fixing those hyperparameters and grid searching smaller increments in other hyperparameters. A typical grid search for four hyperparameter involving their respective ranges is depicted in Fig. 2. The optimal values of hyperparameters for ML models trained using ANS wells include two hidden layers with 40 nodes per layer, 0.001 learning rate, 100 batch size, and 500 epochs. Similar hyperparameters tuning process for ML models trained using Mallik wells resulted in the same optimal values of hyperparameters.

3 Results and discussion

3.1 ML models utilizing various WLC

Features of the ML model were selected from the above-mentioned six well log inputs, and the selected features were

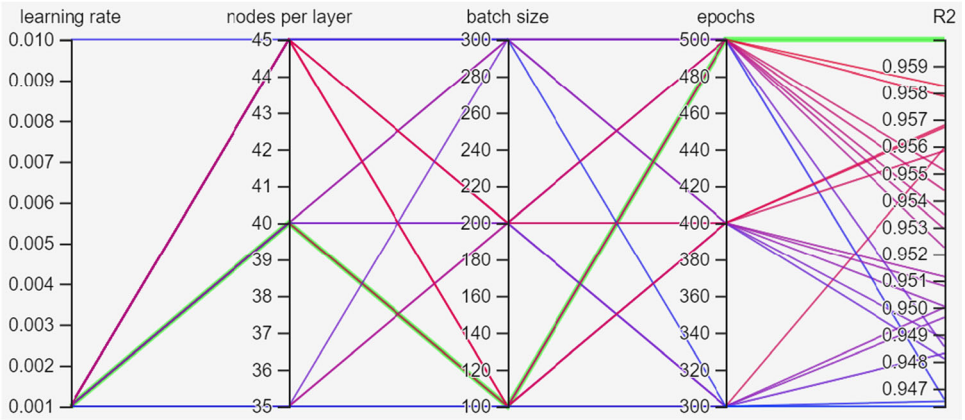

Fig. 2 Tensorboard parallel coordinate view of hyperparameter fine tuning two-layer neural networks showing R^2 and their corresponding hyperparameters. The green line represents a ML validated model with the highest R^2

referred to as well log combinations (WLC). To ascertain variance in predictions of the target variable, each ML model was trained and validated one hundred times, which produces 100 *Sgh* predictions (referred as realizations) and R^2 values for each WLC. Those *Sgh* predictions are not the same because of randomly initialized *weights* for every layer in ANN for a ML model [57]. The R^2 values were then used to calculate an average R^2 that enabled ranking the ML model performance per each WLC. The ML models based on a single well log and WLC with two and three well logs were used for comparative performance analysis. The number of well logs was intentionally limited to three in various WLC to achieve acceptable prediction accuracy with a fewer number of well logs, especially because a ML model utilizing a fewer number of well logs can be applied to a larger number of geological formations, where only a limited number of well logs may have been acquired.

To evaluate performance, first the models with only a single feature were considered. This allowed narrowing down the well log variables per their accuracy in predicting S_{gh} as a single independent feature. Table 2 displays the average R^2 values for the models utilizing only a single feature that were separately trained and validated for the Mallik and ANS wells. Following our previous works, the model providing accuracy above 80% was considered successful [19]. The average R^2 values indicate that models using acoustic (V_p, V_s) or resistivity (*Rt*) well logs provide distinctly high scores compared to the remaining well logs. This is expected given that *Rt* and (V_p, V_s) are sensitive to the presence of gas hydrate and these logs serve as data source to predict S_{gh} in the electrical resistivity and acoustic velocity methods, respectively. In Table 2, the R^2 values were calculated for the ML models validated on the same wells, within the same basin. Attempts to predict S_{gh} at the Mallik or ANS wells using ANS-trained or Mallik-trained ML models, respectively, with acoustic or resistivity logs resulted in poor R^2 (not shown). In other words, ML trained models with a single well log does not perform reliably when used to predict S_{gh} using "unseen" or "blind" data at wells in the other basin.

Next, all WLC pairs were sampled to train the respective ML models, validate them using the well log data from the same basin, and test them on the "blind" wells belonging to the other basin. Table 3 shows results for pair WLC with R^2 above 0.20. It is evident from R^2 values that using two features improves the accuracy of the ML models within the same basin where R^2 is between 0.67 and 0.92. Application of the ANS-trained and Mallik-trained models to predicting S_{gh} at Mallik and ANS wells, respectively, as "blind" wells (not participating in training and validation processes) led to declined performance. Among the two models trained using Mallik and ANS wells, respectively, the ANS-trained ML model provided better accuracy in predicting the target variable at a blind (Mallik) well compared to using Mallik-trained ML model in predicting the target variable at a blind (ANS) well. This variability in performance between the two models trained using datasets from two different basins can be

Table 2 Average R^2 for ML models validated using single logs

Feature	Mallik training	Mallik validation	ANS training	ANS validation
Vp	0.851	0.8487	0.764	0.7608
Rt	0.794	0.7937	0.773	0.7712
Vs	0.777	0.7755	0.529	0.5249
GR	0.132	0.1289	0.417	0.4139
ϕ	0.100	0.0974	0.238	0.2372
ρ	0.085	0.0829	0.340	0.3367

Table 3 Average R^2 for ML models validated and tested using pairs of well logs

WLC	Mallik trained				ANS trained		
	Mallik validated	Mt Elbert tested	Igṅik Sikumi tested	Hydrate-01 tested	ANS validated	5L-38 tested	2L-38 tested
ϕ Vp	0.9239	0.3604	0.5523	0.8264	0.8666	0.9330	0.5438
GR Vp	0.9116	0.4743	0.6166	0.8274	0.8504	0.8559	0.8918
Vp Vs	0.8969	0.2903	0.4153	0.7778	0.8491	0.7157	0.6505
Rt Vp	0.8964	0.3518	0.5111	0.6842	0.8588	0.8676	0.8370
GR Vs	0.8418	0.2020	0.4401	0.7233	0.6684	0.7247	0.7719

attributed to the fact that Mallik-trained ML models are based on the data available at one location, while ANS-trained models include data available from three different sites (Fig. 1). Using the dataset that includes three locations with variability in their rock properties, the ANS-trained ML models can capture the variability in gas hydrate hosting sediments (mineralogy, grain size, etc.) that show broad range of saturations. In other words, the variability exhibited by the ANS dataset is learned by the ML model that enables it to predict gas hydrate parameters in a natural environment more robustly than the Mallik-trained model. Table 3 indicates that ANS-trained ML models based on (*GR, V_p*) and (*Rt, V_p*) pairs provide excellent predictions (average $R^2 > 0.84$) for S_{gh} at the Mallik site. The (*GR, V_p*) pair is important from practical standpoint since *GR* is the one of the most primary logs that is acquired to characterize a formation in terms of its sand-shale geology, which means that *GR* log would be most likely available for for a majority of legacy wells. The *GR* log can be used to differentiate sandy silt or silty sand intervals, potential candidates to host pore-filling concentrated gas hydrate accumulations, from more finely sorted clay-rich intervals. In combination with the sonic tool readings, the *GR* log enables predicting S_{gh} by trained ML models that is similar to NMR-quality S_{gh} prediction. Notably, Saputro et al. [58] also found that *GR* and V_p are critical features in applying ANN to predict porosity log data with R^2 reaching 0.937.

The features that appear in the highest performing pair WLC (Table 3) were selectively augmented with an additional feature to produce triplet WLC in attempt to further improve the ML model accuracy. Since V_p appears in all top pair WLC performers (Table 3), it is therefore preserved in the sampled triplet WLC. Table 4 lists the average R^2 of triplet WLC selectively sampled based on pair WLC results from Table 3. Compared to pair WLC, the addition of the third feature does not significantly improve the R^2 when the trained models are validated against the data within the same basin. However, there is an improvement in the R^2 values when the models are tested to predict S_{gh} at the sites in the other basin. For example, there are only two pair WLC performers yielding R^2 above 0.80 for the models tested at the Mallik wells, while there are six such triplet WLC performers. Significantly, two WLCs including (*ϕ, Rt, V_p*) and (*ϕ, GR, V_p*) lead to ANS-trained ML models yielding the higher R^2 values for the "blind" Mallik wells compared to the R^2 values deduced from the validated datasets (Table 4). The general trend of better performance of the ANS-trained ML models in predicting the target variable at the "blind" Mallik wells is preserved for the triplet WLC. The reason additional logs improve the predictions from a different basin is intuitive in the sense that adding additional logs enable the ML model to better capture the signatures of the formation that control S_{gh}, which means that complete signatures controlling S_{gh} are unlikely to be present in a single log; this hypothesis is also supported by the physical knowledge of gas hydrate deposits that are known to be affected by various characteristics of the formation that is measurable using well logs.

Table 4 Average R^2 for ML models validated and tested using triplets as WLCs

WLC	Mallik trained				ANS trained		
	Mallik validated	Mt Elbert tested	Igṅik Sikumi tested	Hydrate-01 tested	ANS validated	5L-38 tested	2L-38 tested
ϕ Rt Vp	0.9505	0.7436	0.7569	0.7205	0.8833	0.9234	0.9038
GR Rt Vp	0.9417	0.7589	0.6768	0.5447	0.8794	0.8196	0.8243
ϕ GR Vp	0.9352	0.3912	0.5785	0.8321	0.8780	0.9115	0.8179
ϕ Vp Vs	0.9289	0.4199	0.5849	0.8288	0.8919	0.8845	0.5839
Rt Vp Vs	0.9249	0.7109	0.6319	0.6075	0.8854	0.8482	0.8402
GR Vp Vs	0.9128	0.5083	0.6412	0.8262	0.8814	0.8235	0.8887

3.2 R^2 distributions

The performance of an ML model was measured using an averaged R^2 value, which is a metric of accuracy, and evaluated using R^2 distributions, which provide precision or consistency in predictions. Tables 2, 3 and 4 collect the average R^2 values illustrating the accuracy of ML models in predicting the target variable. Figure 3 depicts the R^2 distributions for ANS-trained ML models for the three WLC including *Vp*, (*Vp* and ϕ), and (V_p, ϕ, and *GR*). Figure 4 shows the R^2 distributions for Mallik-trained ML models using the same WLC sequence. The figures demonstrate that consequent addition of a feature into a WLC improve average R^2, although it does not always lead to increase in prediction accuracy, or in other words there is no clear consistency between improvement in R^2 and improvement in prediction. As an example, the distribution of R^2 for Mallik-trained ML models using the triplet WLC appears to be broader than those utilizing the pair

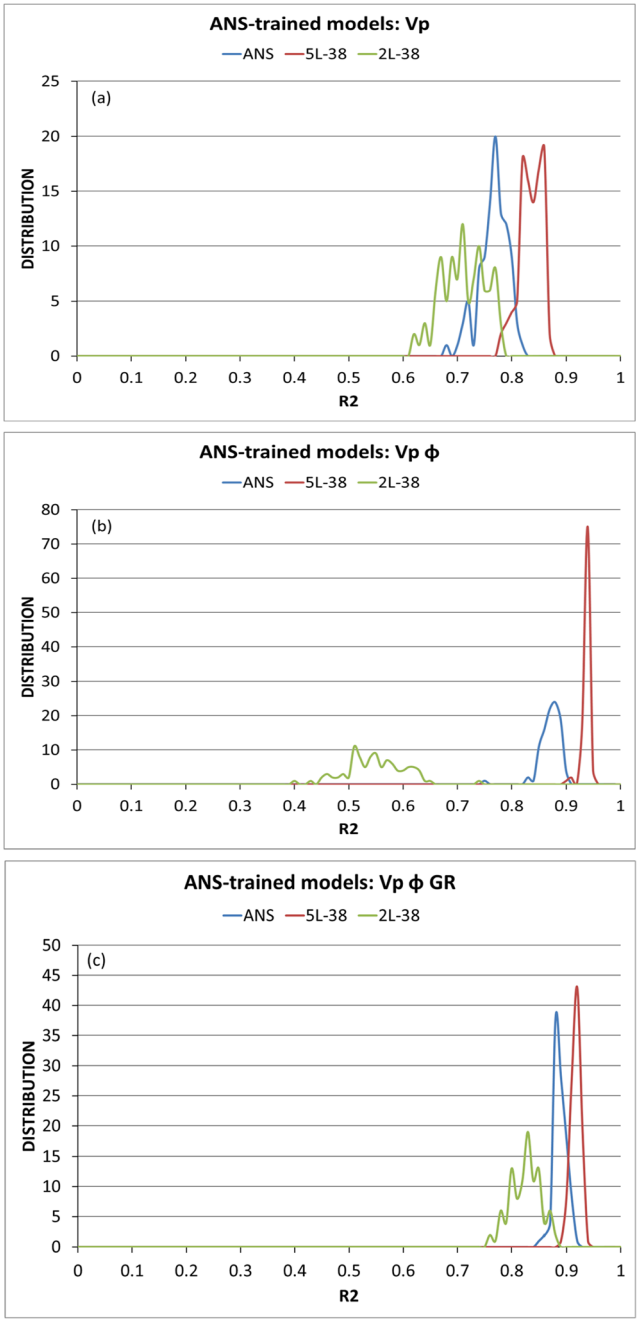

Fig. 3 R^2 distributions of ANS-trained models for (**a**) Vp, (**b**) Vp ϕ, and (**c**) Vp ϕ GR WLC

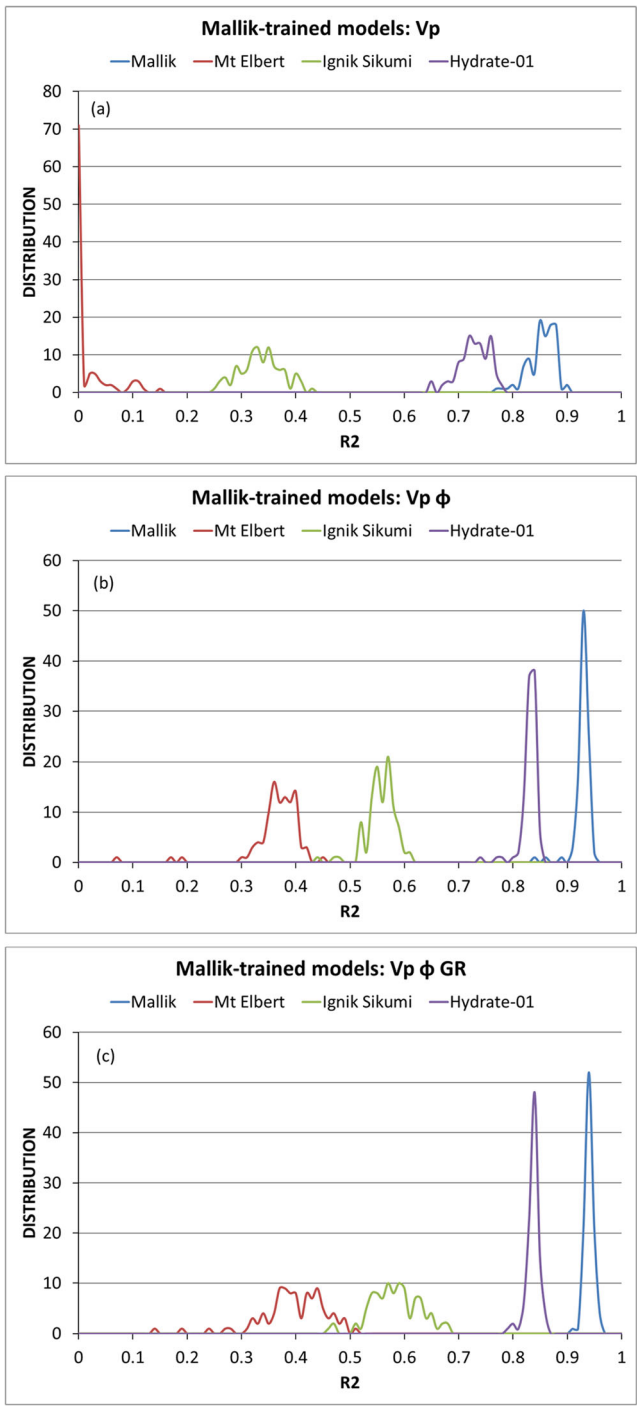

Fig. 4 R^2 distributions of Mallik-trained models for (**a**) Vp, (**b**) Vp ϕ, and (**c**) Vp ϕ GR WLC

WLC. The addition of ϕ to V_p lead to decrease in accuracy for ANS-trained ML models when applied to the 2L-38 well, however, further augmentation of the ϕ and V_p pair with *Rt* increased the R^2 value over 0.90, similar to 5L-38 (Tables 3 and 4). The corresponding R^2 distributions were narrowed down and yielded higher R^2 values exceeding the value obtained using the validation sets (Fig. 3). A similar trend was obtained when *GR* was added to the pair instead of *Rt* (Fig. S3a). However, using V_s instead of *Rt* did not improve the R^2 for 2L-38 (Fig. S3b).

A somewhat different pattern emerges after examining the trend for Mallik-trained ML models (Fig. 4). The models are noticeably robust and accurate with single feature (V_p) and (V_p, ϕ) pair when Hydrate-01 was used as a "blind" well, however, a triplet WLC was required to reliably predict S_{gh} when Mt. Elbert and Iġnik Sikumi were used as "blind" wells. The addition of *Rt* to (V_p, ϕ) WLC created distinguishable distributions for each ANS well (Fig. S4a) with Hydrate-01 achieving closest accuracies with Mallik-trained models. Although the average R^2 increased when *Rt* was used instead of *GR*, the R^2 distribution becomes wide, indicating deterioration in predictions from the Mallik-trained models for those wells. Fig. S4b shows that replacing *Rt* with V_s in the triplet WLC improved the consistency but sacrificed the accuracy. Thus, the ANS-trained ML models achieved narrower distributions and higher R^2 than the Mallik-trained models, likely due to larger variability captured by the training data from three wells at different locations in the case of ANS-trained model compared to the training data from two wells at one site for the Mallik-trained model. Tradeoff between precision and accuracy due to removal or addition of features in the WLC impacted the Mt Elbert and 2L-38 wells the most; this was likely due to the quality or quantity of the key feature impacting gas hydrate saturation for the respective wells.

3.3 Gas hydrate saturation prediction

The average R^2 values collected in Tables 3 and 4, and R^2 distributions shown in Figs. 3 and 4 characterize the performance of the ML models. Those values and distributions of R^2 were created using 100 S_{gh} predictions compared to the "ground truth" of S_{gh}. The predicted S_{gh} values with the highest R^2 were used to select the ML models. Figure 5 shows predicted vs expected ("ground truth") S_{gh} scatter plot for the triplet WLC (ϕ, *GR*, *Vp*), where the straight line with unit-slope represents the perfect match. The figures demonstrate the formation of two distinct clusters of data (highlighted by ovals), where the first cluster indicates accurate predictions at high gas hydrate saturations, and the second cluster designates the predictions underestimating the expected values at low S_{gh} (<0.5). A similar pattern is found in predictions using other WLC (Fig. S5). Here, the main source of deviation of R^2 from a perfect value of 1.0 was likely due to poor prediction of S_{gh}

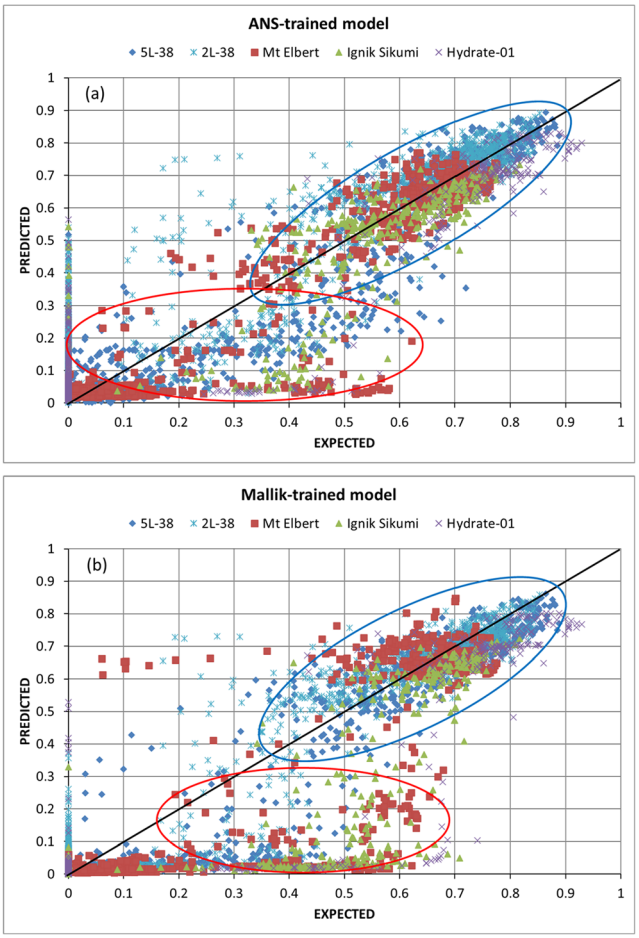

Fig. 5 Prediction vs expectation of a (**a**) ANS-trained and a (**b**) Mallik-trained model using Vp ϕ GR WLC

at lower range ($S_{gh} < 0.5$). This is an interesting result that may be exhibiting the shortcoming in the NMR method in reliably predicting S_{gh} at lower range ($S_{gh} < 0.5$). Consistent under-prediction for $S_{gh} < 0.5$ by ML models utilizing different WLCs is an indication of systematic lack in quality of the original "ground truth" data in those lower range of gas hydrate saturations. One could argue that under-prediction for $S_{gh} < 0.5$ by ML models could also mean shortcoming of ML to match the data at lower S_{gh}, however, there is no rationale that explains this shortcoming in ML. The hypothesis about the shortcoming in the NMR mentioned above is discussed in detail below.

In a recent development of the NMR-based method, the longitudinal relaxation time (T_1) and transverse relaxation time T_2 distributions are jointly inverted, compared to conventional processing which inverts a (T_2) distribution from the echo signal using a constant T_1/T_2 ratio [24]. According to [24], using the constant T_1/T_2 ratio causes the NMR porosity in gas hydrate bearing zones to be underestimated by about 3–6 porosity units, and the derived gas hydrate saturations to be overestimated by ~8–10%. This indicates that the predicted vs expected S_{gh} values in Fig. 5 should experience a bias towards

higher expected S_{gh} values. This is confirmed in Figs. 5 and Fig. S5 that show the data clusters corresponding to high S_{gh} values lying below the unit-slope line, which is a line depicting perfect match; this means that the expected S_{gh} values obtained by the conventional NMR processing are systematically higher than the predicted values. Figure 6 and Fig. S6 show the comparison of predicted and expected S_{gh} values (vs depth) for select wells and ML models, where many intervals corresponding to high S_{gh} clearly show that the expected values are higher the predicted S_{gh}.

The overestimation of conventional NMR-derived S_{gh} by 8–10% cannot fully explain large underprediction of S_{gh} at the lower range (S_{gh} < 50%) (Fig. 5 and Fig. S5). To elaborate further, it should be recalled that the degree of gas hydrate saturation is strongly controlled by reservoir quality, such that a low-quality reservoir can be maximally-saturated up to 50%. In contrast, a high-quality reservoir can achieve saturations above 80% [56]. A low-quality reservoir implies more finely sorted particles, more clay, silt and mud content, and is typically characterized as silty clay or clayey silt facies. Thus, the sediment-bearing low gas hydrate saturations are enriched with clay such as illite, kaolinite, chloride, smectite-family,

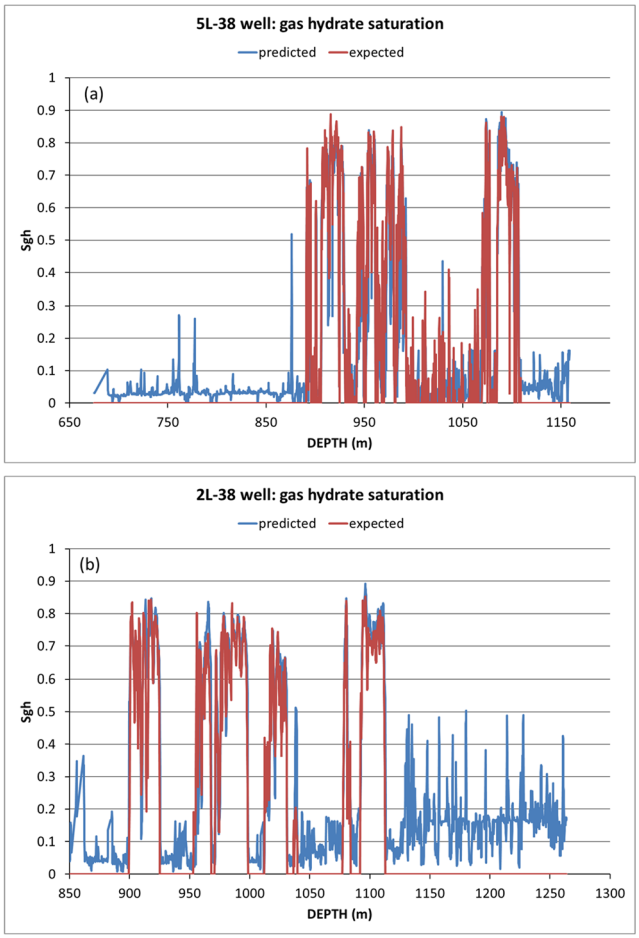

Fig. 6 ANS-trained model predicting gas hydrate saturations using Vp ϕ GR WLC in Mallik wells (a) 5 L-38 and (b) 2 L-38

and others. Recently, Elsayed et al. [59] reported the effect of clay content on the spin−spin NMR relaxation time measured in porous media, such that the increase in oscillating magnetic pulses (TE spacing) leads to strong reduction in T_2 distribution (tails <0.1 ms) due to clay-induced internal field gradients. Consequently, if using a conventional technique, interpretation of T_2 distribution causes underestimation of NMR-based porosity, implying that the NMR-derived gas hydrate saturation will be overestimated in case of gas hydrate in clay-rich sediments. This can be a leading factor responsible for discrepancy between NMR-derived S_{gh} and ML-derived S_{gh} at many depths where S_{gh} values fall lower than 50%. In such locations, the low S_{gh} in pore space generally occur in sediments with elevated clay content that would cause the excessive assignment to S_{gh}.

It must be noted that for systems with moderate complexity, such as pure fluid dynamics without the complexity of geological formations, first principles-based models are sufficiently accurate to depict their behavior. However, for complex systems, such as gas hydrate deposits or multiphase flow in porous media, first principles-based models may not be accurate enough to depict behavior of such complex systems. In such systems, directly-measured features (e.g. well logs) are better representation of the underlying correlations with the target variable (S_{gh}) than the first principles-based models (e.g. S_{gh} based on NMR-derived model), therefore, combining feature engineering (e.g. selective WLC) with data-driven ML appears to be a more accurate representation of S_{gh} than NMR-derived model that is considered the "ground truth". The improved performance exhibited by ML combined with feature engineering over first principles-based models is also exhibited by other studies with relatively complex systems, such as fluid flow characterization in pipes/wells [47, 60] and PVT modeling of complex fluids [61]. The improved performance of ML combined with feature engineering over first principles-based models for relatively complex systems as observed in this study and also confirmed by others [47, 60, 61] suggest that variability represented in first principle-based features (e.g. well logs measured based on fundamental properties) and their appropriate combination when used with ML enables more accurate prediction of target variable than possible through first principles-based models.

3.4 Selection of WLC for a "blind" well

In the previous section, it was determined that triplet WLC should provide reliable prediction of the target variable using ANS-trained ML models (Table 4). Depending on the number of LWD or wireline well logs available at a site of interest, there could be several combinations of logs constituting a triplet WLC. A trained ML model applied to a "blind" well with a triplet WLC as input would predict a realization of S_{gh}. One hundred such applications of the model provides 100

realizations or S_{gh} predictions. Without "ground truth" knowledge to assess the accuracy of prediction through R^2 values, precision is the remaining statistical metrics to characterize those realizations. Precision indicates how close S_{gh} predictions are to each other and for ML applications where "ground truth" is known it visually manifests itself through the spread of R^2 values. For cases where "ground truth" (and R^2 values) is not known, another metric should be utilized.

To quantify the precision of realizations obtained from an ANS-trained model using a triplet WLC, Pearson's correlation coefficients were calculated for each of the 100 realizations against the remaining 99. The Pearson's correlation coefficient (r_{xy}) between two realizations x and y is estimated using Eq. (4) below:

$$r_{xy} = \frac{\sum_{i=1}^{n}\left(x_i - \overline{x}\right)\left(y_i - \overline{y}\right)}{\sqrt{\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2}\sqrt{\sum_{i=1}^{n}\left(y_i - \overline{y}\right)^2}} \tag{4}$$

where, n is a realization size (2472 depth points at the Mallik 2L-38 well); x_i and y_i are individual S_{gh} predictions at a depth point; $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ (the realization mean) and $\overline{y}$ is analogous to $\overline{x}$. r_{xy} is a measure of correlation between 0.0 (no correlation), and 1.0 (identical realizations). Each realization received 99 correlation coefficients, which were processed to obtain mean and median values. The sums of all averaged coefficients were defined as "scores" that constitute the metric to assess the precision of those realizations.

To identify a triplet WLC leading to better reliability in predicting S_{gh}, the triplet WLC reporting best precision (best "scores") is recommended to be selected. That recommendation is based on the analysis of R^2 distributions depicted in Figs. 3-4 and S3–4. As a general trend, triplet WLC providing high averaged R^2 value (accuracy) also deliver high precision showing narrow spreads of individual one hundred R^2 values. Therefore, selecting a WLC with high precision most likely results in reliable estimates of the target variable. To present an instructive example, the ANS-trained ML models using the following triplet WLC; *(ϕ, GR, Vp)*; *(ϕ, Rt, Vp)*; *(ϕ, Vp,* Vs*)* were selected to predict S_{gh} at the Mallik 2 L-38 well. The R^2 distributions are depicted in Figs. 3c, S3a, and S3b, respectively. Their corresponding averaged R^2 values are 0.8179, 0.9038, and 0.5839 (Table 4). Those figures and numbers clearly show that better averaged R^2 is accompanied with more narrow distribution of R^2 values.

Figures 7 and S7 show the averaged Pearson's correlation coefficients for select triplet WLC together with standard deviations for the mean values. In Figs. 7 and S7, the median lines consistently appear above the mean, thus indicating correlation coefficients are grouped towards high values. The scores are depicted in Figs. 7 and S7 and Table 5 together with other descriptive statistics. Comparison of scores with averaged R^2 values reveals the triplet WLC that provides the best performance demonstrated by the highest "scores" (Table 5). In other words, for a "blind" well, which provides well logs "unseen" in the training process by the ANS-trained ML models, the "scores" based on Pearson's correlation coefficients can be used to assess a triplet WLC performance with respect to the target variable.

Table 5 also collects information about the Pearson's coefficients for realizations providing R^2 values closest to averaged R^2 values and R^2 values for realizations providing maximum Pearson's coefficients. Comparison of those numbers allows to deduce that selecting a realization with the highest Pearson's coefficient most likely leads to a prediction with a R^2 value on the right tail of a R^2 distribution. Thus, the suggested screening approach implies selecting first a top triplet WLC performer and then choosing a suitable realization based on Pearson's statistics. It should be emphasized that this approach would not guarantee selection of the best WLC among performers delivering close averaged R^2 (like 0.90 and 0.91), but it does filter out WLC performers with poor averaged R^2 (like 0.60). Similarly, choosing a realization with the maximum Pearson's coefficient would not guarantee selection with maximum R^2 value, but most likely the realization would have R^2 higher than average R^2.

3.5 Further development

The ML models considered above were trained using either ANS or Mallik well logs to predict the target variable at the Mallik site or ANS wells, respectively. Training a ML model using well logs from all five wells on ANS and at the Mallik site can bring even higher accuracy for a "blind" well compared to that computed for the ANS-trained models. Such models were trained following the approach described in the Methodology section of this report, and the validation step has shown consistent accuracy with R^2 above 0.9. This confirms that including data from additional wells (as they become available) in a training dataset further increases the accuracy of predictions. The ML models utilizing all five wells are available for further study to analyze logs available at legacy wells located on ANS and in Northern Canada. Data provided by numerous industry wells have confirmed that gas hydrate exists widely on the North Slope both within and below the permafrost section and almost exclusively occurs within the sand-rich units [23, 62]. These legacy wells were drilled without intention to assess S_{gh} and were mainly associated with oil exploration activity at formations deeper than the gas hydrate stability zone. Access to the routine well logs such as ϕ, *GR*, and readings from sonic and/or electrical resistivity tools would open up a possibility to screen for gas hydrate presence at those locations and help refine our knowledge about gas hydrate resources in North America.

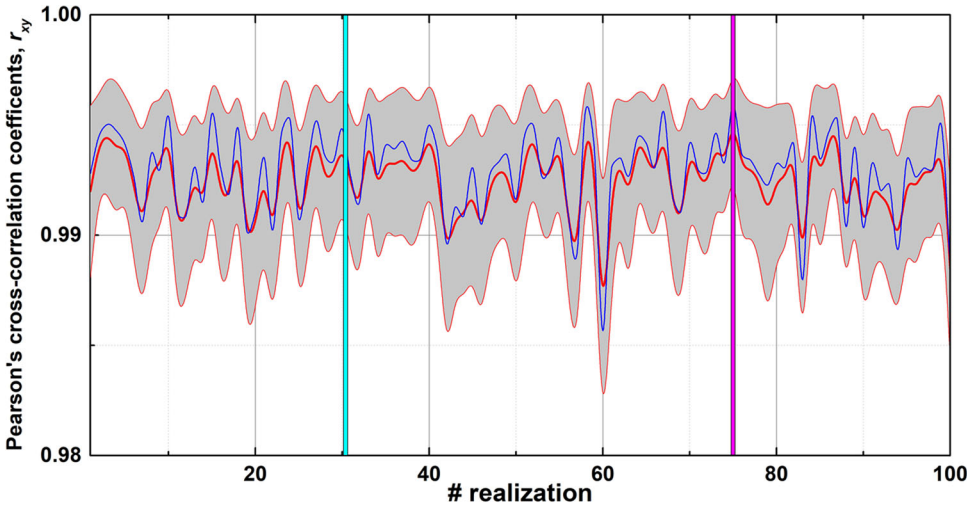

Fig. 7 Averaged Pearson's cross-correlation coefficients for 100 realizations to predict S_{gh} at the Mallik 2 L-38 well using the ANS-trained ML model utilizing ϕ *GR Vp* WLC triplets. The gray areas highlight the standard deviation for the coefficient's means (red curves). The median values are shown in blue curves. The cyan and magenta vertical lines designate the realizations corresponding to the averaged R^2 coefficient and the maximum r_{xy}, respectively

The presence of acoustic logs in the top WLC performers (Table 3) is a promising sign that these measurements can be used towards 3D gas hydrate reservoir characterization using ML in the future. Such models can utilize surface seismic and vertical seismic profile (VSP) data to connect the attributes extracted from those surveys with acoustic well logs through ML training. The ML models can provide a means to evaluate high-resolution acoustic data in the lateral direction from the vertical wellbore and use them to acquire detailed S_{gh} spatial distributions.

The ML models presented in this study are applicable to permafrost-associated continuous sand-rich accumulations typically characterized with high porosity and absolute permeability. To provide the ML model's application to marine gas hydrate-bearing sediments and/or gas hydrate sitting in fractures and veins, a training database should be extended with corresponding suites of well logs at locations of interest. However, the potential of ML in characterizing marine gas hydrate is unclear since gas hydrate in marine sediments often occurs in interbedded stratigraphic units where thin mud layers with no gas hydrate alternate with gas hydrate sandy sections. This variability in mud/sand facies at a resolution that can be finer than the resolution of the well logging tool could lead to averaging of the facies properties by well logs, thereby leading to a poorly characterized target property of closely spaced layers with different lithology [63]. Furthermore, the "ground truth" data used to train the ML models should be scrutinized and, if possible, compared with core laboratory measurements. The physics-driven methods like the NMR-based one that proved to provide reliable estimates for high *Sgh* in high-quality reservoirs might not deliver correct saturation values for fracture-filled gas hydrate and/or gas hydrate in thinly interbedded sedimentary sections.

A brief summary of the results obtained in this study is given below:

- R^2 distributions show that addition of a third feature into a WLC improves the accuracy of ML models;
- ML models trained using three well-logs from ANS (ϕ, *Rt*, V_p) predict S_{gh} with excellent accuracy ($R^2 > 0.90$) for the Mallik site;
- ML predictions of S_{gh} are excellent in the range where S_{gh} values above 50%, while predictions in the lower range of S_{gh} are substantially underestimated, and;
- Pearson's descriptive statistics was utilized to identify a candidate WLC and realization for a "blind well" where no "ground truth" is available.

4 Conclusions

The ML models trained using the well logs from ANS predicted gas hydrate saturation distributions at the Mallik 5L-38

Table 5 Descriptive statistics collected across 100 realizations to predict gas hydrate saturation distributions at the Mallik 2L-38 well using the ANS-trained ML models

WLC	Score (mean values)	Score (median values)	Pearson's coefficient, r_{xy} at aver. R^2	Averaged R^2	Max Pearson's coefficient, max r_{xy}	R^2 value at max r_{xy}
ϕ, Rt, Vp	99.275	99.335	0.9927	0.9038	0.9950	0.9235
ϕ, GR, Vp	99.245	99.295	0.9923	0.8179	0.9954	0.8267
ϕ, Vp, Vs	98.937	98.983	0.9897	0.5839	0.9932	0.6214

and 2L-38 wells with high accuracy, thus demonstrating the applicability of the models for sites in a different geological basin. The pair of well logs that combine compressional velocity either with gamma ray or electric resistivity readings provide high prediction accuracy (R^2 above 0.82) relative to NMR-based gas hydrate saturations, which can be further improved (R^2 up to a 0.92) by including density porosity to the two pairs. The ML models trained using the Mallik wells were also tested to predict gas hydrate saturations at the three ANS wells (Mount Elbert, Iġnik Sikumi, and Kuparuk 7–11–12), such that the accuracy of the best ML model was within 0.72–0.76 (R^2) with density porosity, electrical resistivity, and compressional velocity well logs as the input data. The better performance of the ANS-trained ML models over the Mallik-trained models was attributed to the nature of geological variability within the training datasets. Specifically, in the former case, the training data collected from three locations on ANS contained more variability in geological settings compared to the training data with only one site in Mackenzie Delta. The analysis of ML-predicted values against the "ground truth" indicates an excellent match when gas hydrate saturations are above 50%, whereas the ML-predicted gas hydrate saturations are underestimated when gas hydrate saturation in the "ground truth" are below 50%. It is speculated that the systematic underestimated predictions of S_{gh} by the ML models in the case of poor quality gas hydrate intervals (below 50% S_{gh}) could be a limitation of the NMR-log based technique in reliably predicting S_{gh} values.

ML models using all five wells were trained for further application at sites without prior knowledge of gas hydrate distribution in reservoir units. The models were validated showing consistent accuracy with R^2 above 0.9. These trained ML models utilized various triplet WLC comprising φ, GR, Vp, Vs, and Rt logs; and without prior knowledge of the "ground truth" a generalized approach was developed to select triplet WLCs that provided best predictions for a "blind" well. The approach was based on the descriptive statistics collected using one hundred realizations to predict gas hydrate saturations at every ML model.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10596-022-10151-9.

Acknowledgements This work was performed in support of the U.S. Department of Energy's Fossil Energy Crosscutting Technology Research Program. The research was executed through the NETL Research and Innovation Center's Hydrate Research Field Work Proposal. This research was supported in part by an appointment to the National Energy Technology Laboratory Research Participation Program, sponsored by the U.S. Department of Energy. The authors would like to express their sincere appreciation to the U.S. DOE-NETL, and the Japanese Ministry of Economy, Trade and Industry for providing the permission to disclose this research. Any use of trade, product, or firm names is for descriptive purposes only and does not imply endorsement by the U.S. or Japanese Government.

References

1. Karpatne, A., Ebert-Uphoff, I., Ravela, S., Babaie, H.A., Kumar, V.: Machine learning for the geosciences: challenges and opportunities. IEEE Trans. Knowl. Data Eng. **31**, 1544–1554 (2019). https://doi.org/10.1109/TKDE.2018.2861006
2. Lary, D.J., Alavi, A.H., Gandomi, A.H., Walker, A.L.: Machine learning in geosciences and remote sensing. Geosci. Front. **7**, 3–10 (2016). https://doi.org/10.1016/j.gsf.2015.07.003
3. Caté, A., Perozzi, L., Gloaguen, E., Blouin, M.: Machine learning as a tool for geologists. Lead. Edge. **36**, 215–219 (2017). https://doi.org/10.1190/tle36030215.1
4. Karpatne, A., Watkins, W., Read, J., Kumar, V.: Physics-Guided Neural Networks (PGNN): an Application in Lake Temperature Modeling. (2017)
5. Racah, E., Beckham, C., Maharaj, T., Kahou, S.E., Prabhat, Pal C : ExtremeWeather: A large-scale climate dataset for semi-supervised detection, localization, and understanding of extreme weather events. (2016)
6. Raschka, S., Mirjalili, V.: Python machine learning: machine learning and deep learning with Python, scikit-learn, and TensorFlow 2, 3rd edn. Packt Publishing (2019)
7. Collett, T., Johnson, A., Knapp, C., Boswell, R.: Natural gas hydrates—energy resource potential and associated geologic hazards. American Association of Petroleum Geologists (2009)
8. Uchida, S., Soga, K., Klar, A., Yamamoto, K.: Geomechanical study of the Mallik gas hydrate production field trials. Bull. Geol. Surv. Canada. **601**, 191–204 (2012). https://doi.org/10.4095/291751
9. Anderson, B.J., Kurihara, M., White, M.D., Moridis, G.J., Wilson, S.J., Pooladi-Darvish, M., Gaddipati, M., Masuda, Y., Collett, T.S., Hunter, R.B., Narita, H., Rose, K., Boswell, R.: Regional long-term production modeling from a single well test, mount Elbert gas hydrate stratigraphic test well. Alaska North Slope. Mar. Pet. Geol. **28**, 493–501 (2011). https://doi.org/10.1016/j.marpetgeo.2010.01.015
10. Boswell, R., Schoderbek, D., Collett, T.S., Ohtsuki, S., White, M., Anderson, B.J.: The Iġnik Sikumi field experiment, Alaska north slope: design, operations, and implications for CO 2 –CH 4 exchange in gas hydrate reservoirs. Energy Fuel. **31**, 140–153 (2017). https://doi.org/10.1021/acs.energyfuels.6b01909
11. Collett, T.S., Zyrianova, M. V, Okinaka, N., Wakatsuki, M., Boswell, R., Marsteller, S., Minge, D., Crumley, S., Itter, D., Hunter, R.D.: Design and operations of the Hydrate 01 Stratigraphic test well, Alaska North Slope, http://pubs.er.usgs.gov/publication/70213245, (2020)
12. Yamamoto, K., Terao, Y., Fujii, T., Ikawa, T., Seki, M., Matsuzawa, M., Kanno, T.: Operational overview of the first offshore production test of methane hydrates in the Eastern Nankai Trough. In: Day 3 Wed, May 07, 2014. OTC (2014)

13. Kumar, P., Collett, T.S., Shukla, K.M., Yadav, U.S., Lall, M.V., Vishwanath, K.: India National gas Hydrate Program Expedition-02: operational and technical summary. Mar. Pet. Geol. **108**, 3–38 (2019). https://doi.org/10.1016/j.marpetgeo.2018.11.021
14. Boswell, R., Myshakin, E., Moridis, G., Konno, Y., Collett, T.S., Reagan, M., Ajayi, T., Seol, Y.: India National gas Hydrate Program Expedition 02 summary of scientific results: numerical simulation of reservoir response to depressurization. Mar. Pet. Geol. **108**, 154–166 (2019). https://doi.org/10.1016/j.marpetgeo.2018.09.026
15. Yang, S., Liang, J., Lei, Y., Gong, Y., Xu, H., Wang, H., Lu, J., Holland, M., Schultheiss, P., Wei, J., others: GMGS4 gas hydrate drilling expedition in the South China Sea. Fire ice. 17, 7–11 (2017)
16. Flemings, P., Boswell, R., Collett, T., Cook, A., Divins, D., Frye, M., Guerine, G., Goldberg, D., Malinverno, A., Meazell, K., Morrison, J., Pettigrew, T., Phillips, S., Santra, M., Sawyer, D., Shedd, W., Thomas, C., You, K.: GOM2: Prospecting, Drilling and Sampling Coarse-Grained Hydrate Reservoirs in the Deepwater Gulf of Mexico. (2017)
17. Archie, G.E.: The electrical resistivity log as an aid in determining some reservoir characteristics. Trans. AIME. **146**, 54–62 (1942). https://doi.org/10.2118/942054-G
18. Lu, S., McMechan, G.A.: Estimation of gas hydrate and free gas saturation, concentration, and distribution from seismic data. GEOPHYSICS. **67**, 582–593 (2002). https://doi.org/10.1190/1.1468619
19. Singh, H., Seol, Y., Myshakin, E.M.: Prediction of gas hydrate saturation using machine learning and optimal set of well-logs. Comput. Geosci. **25**, 267–283 (2021). https://doi.org/10.1007/s10596-020-10004-3
20. Lee, M.W., Hutchinson, D.R., Collett, T.S., Dillon, W.P.: Seismic velocities for hydrate-bearing sediments using weighted equation. J. Geophys. Res. Solid Earth. **101**, 20347–20358 (1996). https://doi.org/10.1029/96JB01886
21. Helgerud, M.B., Dvorkin, J., Nur, A., Sakai, A., Collett, T.: Elastic-wave velocity in marine sediments with gas hydrates: effective medium modeling. Geophys. Res. Lett. **26**, 2021–2024 (1999). https://doi.org/10.1029/1999GL900421
22. Kumar, D., Dash, R., Dewangan, P.: Methods of gas hydrate concentration estimation with field examples. Goehorizons. 76–86 (2009)
23. Collett, T.S., Lee, M.W.: Well log characterization of natural gas-hydrates. Petrophysics - SPWLA J. Form. Eval. Reserv. Descr. **53**, 348–367 (2012)
24. Jain, V., Saumya, S., Vij, J., Singh, J., Singh, B., Pattnaik, S., Oli, A., Kumar, P., Collett, T.S.: New technique for accurate porosity estimation from logging-while-drilling nuclear magnetic resonance data, NGHP-02 expedition, offshore. India. Mar. Pet. Geol. **108**, 570–580 (2019). https://doi.org/10.1016/j.marpetgeo.2018.11.001
25. Hastie, T., Tibshirani, R., Friedman, J.: The elements of statistical learning: data mining, inference, and prediction. Springer New York (2013)
26. Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J.: Classification And Regression Trees. Routledge (2017)
27. Dudani, S.A.: The distance-weighted k-nearest-neighbor rule. IEEE Trans. Syst. Man. Cybern. **SMC-6**, 325–327 (1976). https://doi.org/10.1109/TSMC.1976.5408784
28. Pal, S.K., Mitra, S.: Multilayer perceptron, fuzzy sets, and classification. IEEE Trans. Neural Netw. **3**, 683–697 (1992). https://doi.org/10.1109/72.159058
29. Friedman, J.H.: Multivariate adaptive regression splines. Ann. Stat. **19**, (1991). https://doi.org/10.1214/aos/1176347963
30. Hastie, T., Tibshirani, R.: Generalized additive models. Stat. Sci. **1**, (1986). https://doi.org/10.1214/ss/1177013604
31. Robbins, H., Monro, S.: A stochastic approximation method. Ann. Math. Stat. **22**, 400–407 (1951). https://doi.org/10.1214/aoms/1177729586
32. Singh, H., Seol, Y., Myshakin, E.M.: Automated well-log processing and lithology classification by identifying optimal features through unsupervised and supervised machine-learning algorithms. SPE J. **25**, 2778–2800 (2020). https://doi.org/10.2118/202477-PA
33. Dallimore, S.R., Yamamoto, K., Wright, J.F., Bellefleur, G.: Scientific results from the JOGMEC/NRCan/Aurora Mallik 2007-2008 gas hydrate production research well program, Mackenzie Delta, Northwest Territories. Canada. Geol. Surv. Canada. **601**, (2012). https://doi.org/10.4095/291751
34. Dallimore, S.R., Collett, T.S., Taylor, A.E., Uchida, T., Weber, M., Chandra, A., Mroz, T.H., Caddel, E.M., Inoue, T.: Scientific results from the Mallik 2002 gas hydrate production research well program, Mackenzie Delta, northwest territories. Canada: Preface. Bull. Geol. Surv. Canada. **585**, (2005)
35. Rose, K., Boswell, R., Collett, T.: Mount Elbert gas hydrate stratigraphic test well, Alaska north slope: coring operations, core sedimentology, and lithostratigraphy. Mar. Pet. Geol. **28**, 311–331 (2011). https://doi.org/10.1016/j.marpetgeo.2010.02.001
36. Yoneda, J., Jin, Y., Muraoka, M., Oshima, M., Suzuki, K., Walker, M., Otsuki, S., Kumagai, K., Collett, T.S., Boswell, R., Okinaka, N.: Multiple physical properties of gas hydrate-bearing sediments recovered from Alaska North Slope 2018 Hydrate-01 stratigraphic test well. Mar. Pet. Geol. 123, 104748 (2021). https://doi.org/10.1016/j.marpetgeo.2020.104748
37. Géron, A.: Hands-on machine learning with Scikit-learn, Keras, and TensorFlow: concepts, tools, and techniques to build intelligent systems. O'Reilly Media. (2019)
38. Collett Denver, CO (United States)], T.S. [United S.G.S., Lewis Oklahoma City, OK (United States)], R.E. [Schlumberger W.L.S., Dallimore Pacific Geoscience Centre, Sidney, BC (Canada)], S.R. [Geological S. of C.: JAPEX/JNOC/GSC et al. Mallik 5L-38 gas hydrate production research well downhole well-log and core montages. Natural Resources Canada, Geological Survey of Canada, Vancouver, BC (Canada), Canada (2005)
39. Yang, X.W., Murray, D.R., Noguchi, S., Fujii, T., Fujii, K., Yamamoto, K., Dallimore, S., R: geophysical well-log montage for the Aurora/JOGMEC/NRCan Mallik 2L-38 Gas Hydrate Production Research Well. (2012)
40. Collett, T.S., Lewis, R.E., Winters, W.J., Lee, M.W., Rose, K.K., Boswell, R.M.: Downhole well log and core montages from the mount Elbert gas hydrate stratigraphic test well. Alaska North Slope. Mar. Pet. Geol. **28**, 561–577 (2011). https://doi.org/10.1016/j.marpetgeo.2010.03.016
41. Schoderbek, D., Farrell, H., Howard, J., Raterman, K., Silpngarmlert, S., Martin, K., Smith, B., Klein, P.: ConocoPhillips gas hydrate production test. , Pittsburgh, PA, and Morgantown, WV (United States) (2013)
42. Boswell, R., Collett, T.S., Suzuki, K., Yoneda, J., Haines, S.S., Okinaka, N., Tamaki, M., Crumley, S., Itter, D., Hunter, R.: Alaska north slope 2018 Hydrate-01 stratigraphic test well: technical results. Presented at the. (2020)
43. 2011 2012 IGNIK SIKUMI DATASETS, https://netl.doe.gov/oil-gas/methane-hydrates/2011-2012-IGNIK-SIKUMI-DATASETS
44. WELL LOG DATA FROM THE 2007 MT ELBERT TEST, https://netl.doe.gov/node/6865
45. van Stein, B., van Leeuwen, M., Back, T.: Local subspace-based outlier detection using global neighbourhoods. In: 2016 IEEE International Conference on Big Data (Big Data). pp. 1136–1142. IEEE (2016)
46. Ugborugbo, O., Rao, T.: Impact of Borehole Washout on Acoustic Logs and Well-to-Seismic Ties. In, Nigeria Annual International Conference and Exhibition. Society of Petroleum Engineers (2009)

47. Bikmukhametov, T., Jäschke, J.: Combining machine learning and process engineering physics towards enhanced accuracy and explainability of data-driven models. Comput. Chem. Eng. **138**, 106834 (2020). https://doi.org/10.1016/j.compchemeng.2020.106834
48. Singh, S., Kanli, A.I., Sevgen, S.: A general approach for porosity estimation using artificial neural network method: a case study from Kansas gas field. Stud. Geophys. Geod. **60**, 130–140 (2016). https://doi.org/10.1007/s11200-015-0820-2
49. Kleinberg, R.L., Flaum, C., Collett, T.S.: Magnetic resonance log of JAPEX/JNOC/GSC et al. Mallik 5L-38 gas hydrate production research well: gas hydrate saturation, growth habit, and relative permeability. Bull. Surv. Canada. 585, 114 (2005)
50. Haines, S.S., Collett, T., Boswell, R., Lim, T., Okinaka, N., Suzuki, K., Fujimoto, A.: Gas hydrate saturation estimation from acoustic log data in the 2018 Alaska north slope Hydrate-01 stratigraphic test well. In: Proceedings of the 10th International Conference on Gas Hydrates (ICGH10). US Department of Energy – NETL program. Singapore. (2020)
51. Song, S., Hou, J., Dou, L., Song, Z., Sun, S.: Geologist-level wireline log shape identification with recurrent neural networks. Comput. Geosci. **134**, 104313 (2020). https://doi.org/10.1016/j.cageo.2019.104313
52. Pontes, F.J., Amorim, G.F., Balestrassi, P.P., Paiva, A.P., Ferreira, J.R.: Design of experiments and focused grid search for neural network parameter optimization. Neurocomputing. **186**, 22–34 (2016). https://doi.org/10.1016/j.neucom.2015.12.061
53. Liashchynskyi, P., Liashchynskyi, P.: Grid Search. Random Search, Genetic Algorithm, A Big Comparison for NAS (2019)
54. Scalero, R.S., Tepedelenlioglu, N.: A fast new algorithm for training feedforward neural networks. IEEE Trans. Signal Process. **40**, 202–210 (1992). https://doi.org/10.1109/78.157194
55. Kanfar, R., Shaikh, O., Yousefzadeh, M., Mukerji, T.: Real-time well log prediction from drilling data using deep learning. Int. Pet. Technol. Conf. 2020, IPTC 2020. (2020). https://doi.org/10.2523/iptc-19693-ms
56. Baldi, P., Sadowski, P.: The dropout learning algorithm. Artif. Intell. **210**, 78–122 (2014). https://doi.org/10.1016/j.artint.2014.02.004
57. Glorot, X., Statistics, Y.B.B.T.-P. of the T.I.C. on A.I. and: Understanding the difficulty of training deep feedforward neural networks, http://proceedings.mlr.press/v9/glorot10a/glorot10a.pdf
58. Saputro, O.D., Maulana, Z.L., Latief, F.D.E.: Porosity Log Prediction Using Artificial Neural Network. J. Phys. Conf. Ser. 739, 012092 (2016). https://doi.org/10.1088/1742-6596/739/1/012092
59. Elsayed, M., Glatz, G., El-Husseiny, A., Alqubalee, A., Adebayo, A., Al-Garadi, K., Mahmoud, M.: The effect of clay content on the spin–spin NMR relaxation time measured in porous media. ACS Omega. **5**, 6545–6555 (2020). https://doi.org/10.1021/acsomega.9b04228
60. Arief, H.A., Wiktorski, T., Thomas, P.J.: A survey on distributed fibre optic sensor data modelling techniques and machine learning algorithms for multiphase fluid flow estimation. Sensors. **21**, 2801 (2021). https://doi.org/10.3390/s21082801
61. Truc, G., Rahmanian, N., Pishnamazi, M.: Assessment of cubic equations of state: machine learning for rich carbon-dioxide systems. Sustainability. **13**, 2527 (2021). https://doi.org/10.3390/su13052527
62. Collett, T.S., Lee, M.W., Agena, W.F., Miller, J.J., Lewis, K.A., Zyrianova, M.V., Boswell, R., Inks, T.L.: Permafrost-associated natural gas hydrate occurrences on the Alaska north slope. Mar. Pet. Geol. **28**, 279–294 (2011). https://doi.org/10.1016/j.marpetgeo.2009.12.001
63. Boswell, R., Collett, T.S., Myshakin, E., Ajayi, T., Seol, Y.: The increasingly complex challenge of gas hydrate reservoir simulation. In: proceedings of the 9th international conference on gas hydrates (ICGH9). Denver. (2017)
64. Medioli, B.E., Wilson, N., Dallimore, S.R., Paré, D., Brennan-Alpert, P., Oda, H.: Sedimentology of the cored interval, JAPEX/JNOC/GSC et al. Mallik 5L-38 gas hydrate production well, Mackenzie Delta, Northwest Territories. (2005)