Research Article

Digitalization of Legacy Datasets and Machine Learning Regression Yields Insights for Reservoir Property Prediction and Submarine-Fan Evolution: A Subsurface Example From the Lewis Shale, Wyoming

Thomas Martin[1] , Jared Tadla[1], Zane Jobe[1]

[1] Department of Geology and Geological Engineering, Colorado School of Mines

The Sedimentary Record

Machine-learning algorithms have long aided in geologic property prediction from well-log data, but are primarily used to classify lithology, facies, formation, and rock types. However, more detailed properties (e.g., porosity, grain size) that are important for evaluating hydrocarbon exploration and development activities, as well as subsurface geothermal, CO_2 sequestration, and hydrological studies have not been a focus of machine-learning predictions. This study focuses on improving machine-learning regression-based workflows for quantitative geological property prediction (porosity, grain size, XRF geochemistry), using a robust dataset from the Dad Sandstone Member of the Lewis Shale in the Green River Basin, Wyoming.

Twelve slabbed cores collected from wells targeting turbiditic sandstones and mudstones of the Dad Sandstone member provide 1212.2 ft. of well-log and core data to test the efficacy of five machine-learning models, ranging in complexity from multivariate linear regression to deep neural networks. Our results demonstrate that gradient-boosted decision-tree models (*e.g.*, CatBoost, XGBoost) are flexible in terms of input data completeness, do not require scaled data, and are reliably accurate, with the lowest or second lowest root mean squared error (RMSE) for every test. Deep neural networks, while used commonly for these applications, never achieved lowest error for any of the testing. We also utilize newly collected XRF geochemistry and grain-size data to constrain spatiotemporal sediment routing, sand-mud partitioning, and paleo-oceanographic redox conditions in the Green River Basin.

Test-train dataset splitting traditionally uses randomized inter-well data, but a blind well testing strategy is more applicable to most geoscience applications that aim to predict properties of new, unseen well locations. We find that using inter-well training datasets are more optimistic when applied to blind wells, with a median difference of 0.58 RMSE when predicting grain size in phi units. Using these data and results, we establish a baseline workflow for applying machine-learning regression algorithms to core-based reservoir properties from well-log and core-image data. We hope that our findings and open-source code and datasets released with this paper will serve as a baseline for further research to improve geological property prediction for sustainable earth-resource modeling.

INTRODUCTION

Well-log analysis, formation evaluation, and subsurface property modeling are an integral part of subsurface characterization for oil and gas exploration, carbon sequestration, geothermal development, mineral exploration, and water-resource characterization (Maries et al., 2017; Stumm & Como, 2017; Wallis et al., 2009; Williams & Lane, 1998). These workflows are typically performed on pro-

prietary datasets using closed-source commercial software (*e.g.*, Eriavbe & Okene, 2019). However, machine learning (ML) models and open-source datasets have led to major advances in the geosciences (Dramsch, 2020) due to their ease of use in high-level programming languages (*e.g.*, Python) that have thorough documentation and community support. Developing reproducible workflows (*e.g.*, reservoir property prediction) using open-source ML tools will allow researchers to investigate relationships and pre-

dictive performance on their own datasets. However, the large number of choices of ML models and methodologies (*e.g.*, classification, regression) makes it difficult to select a particular model that will be effective for a specific use case (Raschka, 2018).

This study compares five open-source ML regression models for subsurface formation property prediction, utilizing well-log data and derived core-image statistics. We chose a regression methodology (predicting a value) instead of classification (predicting a label) because (1) numerous studies have explored classification (Bormann et al., 2020; Hall & Hall, 2017) and (2) regression tends to be a better choice for predicting continuous reservoir property values. We demonstrate this workflow using an open-source dataset focusing on the Dad Sandstone Member within the Lewis Shale in south central Wyoming, which is analogous to many submarine channel-fan deposits that host significant oil and gas reserves (Pyles & Slatt, 2007). While these data and models are from a specific sedimentary depositional environment, the findings and workflows are transferable to other environments (*e.g.*, fluvio-deltaic systems, carbonate platforms) and datasets (*e.g.*, borehole image-log, hyperspectral core images) as long as there are trusted training and testing datasets from those environments.

GEOLOGIC SETTING

This study examines Upper Cretaceous submarine fan deposits within the Greater Green River Basin (GGRB) in south central Wyoming (Fig. 1). The GGRB is an active hydrocarbon-producing basin with exploration activity dating back to the early 1950's (Hettinger & Roberts, 2005). The cores in this study were retrieved from legacy vertical wells in several producing gas fields (Fig. 1; Asquith, 1975; Hettinger & Roberts, 2005); currently, operators utilize horizontal drilling techniques in the basin for hydrocarbon extraction activities (Levon & Mazza, 2020). Digitalizing these legacy vertical cores is imperative to provide better parameterization of reservoir models for future horizontal hydrocarbon development and/or carbon sequestration activities. The north, south, and east of the GGRB is bound by Precambrian basement thrust faults and Sevier fold and thrust belt structures to the west (Johnson & Andersen, 2009). The subsidence of the Mesozoic foreland basin in the northern part of the basin was driven by changes in the subduction angle of the Farallon plate (Johnson & Andersen, 2009; Yonkee & Weil, 2015), while subsidence in the southern part of the basin was mainly controlled by the nearby Uinta uplift (Johnson & Andersen, 2009; Liu & Nummedal, 2004). Today, the Great-Divide and Washakie sub-basins are separated by the Wamsutter Arch, but during the time of deposition, they formed a continuous deepwater basin (Olariu et al., 2012; Yonkee & Weil, 2015). The core presented in this study was collected from wells located east of the Rock Springs Uplift and south of the Wind River Range within the Great-Divide and Washakie sub-basins of the GGRB (Fig. 1).

The Dad Sandstone Member of the Lewis Shale was deposited during the Late Cretaceous (Fig. 1B), and consists of deep-water siliciclastic deposits, interpreted to be slope deposits and submarine fan deposits, with both channelized and lobate architectures (Fig. 1A, Asquith, 1970; Cain, 1986; Carvajal & Steel, 2012; Koo et al., 2016; Pyles & Slatt, 2007; van Horn & Shannon, 1989; Winn et al., 1987). Pyles (2000), Sapardina (2012), and Koo (2015) provide detailed core descriptions and interpretations of depositional processes of the Dad Sandstone. Common event-bed types in the Dad Sandstone are turbidites (Bouma, 1962; Lowe, 1982) and hybrid event beds (Haughton et al., 2009; Talling et al., 2012). The deep-marine Dad Sandstone member is coeval to the Fox Hills shallow-marine shoreface (Olariu et al., 2012), and both have rapid progradation rates of ~50km/My (Carvajal & Steel, 2012) that progressively fill the basin from north to south (Fig. 2). The deposition of the linked Fox Hills-Lewis/Dad depositional system occurred over ~2.2 My (Pyles et al., 2011).

DATASET

Overview

The dataset used in this study was collected from cored intervals from twelve wells in the GGRB (Fig. 1, Table 1) that have targeted the Dad Sandstone. Ten of the wells are available at the USGS Core Research Center (Hicks & Adrian, 2009) and two of them are available at the Colorado School of Mines core repository (PDRMT & CEPO). These wells were chosen for their accurate depth markings, spatial and stratigraphic location, presence of associated measurements (*e.g.*, porosity), and general core quality and condition; many other Dad/Lewis cores are available in the public domain, but do not satisfy these criteria. This dataset and associated code are open-source and available on the supplementary material and GitHub (Martin, 2022).

Well-log data

All wells in the study area have well-log data in the cored interval, ranging from having a full suite of logs (Fig. 3) to one well only having gamma-ray (S821). In cases of missing well-log data, the curve was imputed by using the *scikit-learn* Python package (Pedregosa et al., 2011) with a decision tree ML model using other wells that had all the logs desired (Fig. 3). This is not a replacement to collecting the full suite of well logs but allows for the ML models to be compared using the same datasets, as some models require matching and complete data. After the data was imputed, each well had the following curves: Caliper, Gamma Ray, Sonic, Spontaneous Potential, Density, Photoelectric Factor, Deep Resistivity, Density Porosity (Fig. 3). Each individual well-log curve was pre-processed on a per-well basis to normalize some of the differences between wells arising from different tools, vendors, vintages, and subsurface conditions. The well-log data were interpolated to a 0.1 Ft depth step basis to match the grain-size data we collected from the cores as part of this study. Further description of the original well-log data and the specific implementation of the log imputation and pre-processing is described in the GitHub repository (Martin, 2022).
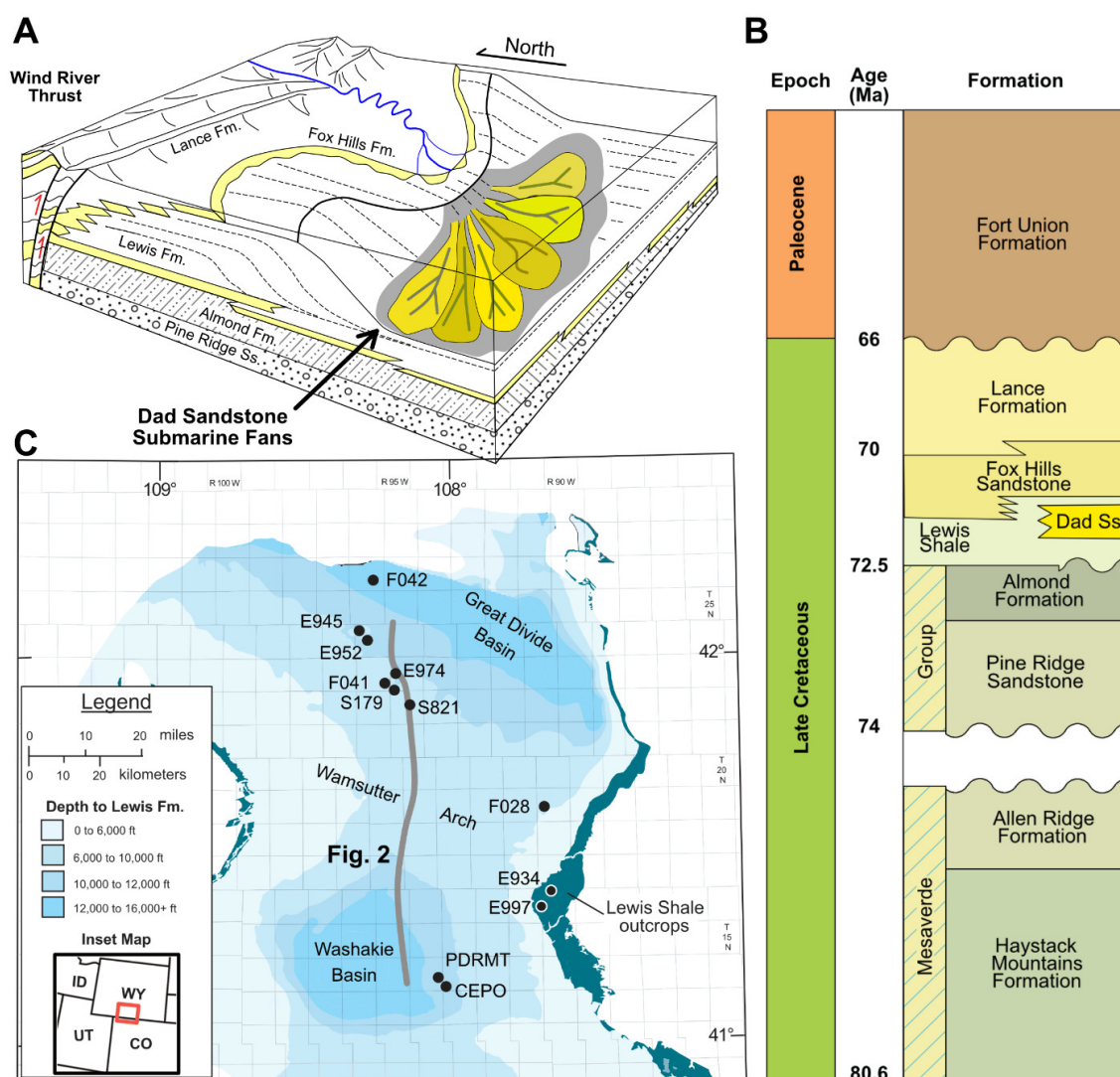
Figure 1. Dad Sandstone submarine fan deposits, Wyoming.

A) Conceptualized block and depositional model, modified from Van Horn and Shannon (1989). B) Stratigraphic column modified from Wyoming State Geological Survey. C) Basin map with well locations. Color contours represent subsurface depth to the top of the Lewis Shale. Grey line denotes approximate location of Figure 2.

Grain size

Most ML studies utilize subjective geologic classifications determined by geoscientists (*e.g.*, lithofacies, systems tracts) (Hall & Hall, 2017). Our goal was to minimize the interpretive error in lithologic description, and thus grain size was determined by the authors using a grain size card and physical inspection of the core itself (Compton, 2016). While there is still human error and bias in determining grain size visually (see discussion in Jobe et al., 2021), it is more objective than interpreting stratigraphic facies or other more subjective/qualitative classifications. The grain-size data were collected using a categorical guide (*e.g.*, upper-fine, Wentworth, 1922) and digitized and converted to the phi scale to have a linear, numeric scale (Krumbein, 1938); finally, these data were interpolated onto a 0.1 Ft scale (Fig. 4).

XRF Geochemistry

X-Ray Fluorescence (XRF) is a non-destructive measurement technique commonly used on core material to obtain efficient and accurate elemental composition (Young et al., 2016). For this study, we collected data using a Bruker Handheld Tracer 5G portable XRF every 0.5Ft on the core, using a 50kV energy level and 90 second collection time. Comparing results of the standards used before and after every day of data collection, most differences were below measurement error and therefore no corrections were needed. We used a helium purge to further enhance data quality of lighter elements such as magnesium, aluminum, and silicon. The five elements we will focus on for this study are aluminum (Al), titanium (Ti), silicon (Si), calcium (Ca), and magnesium (Mg). These five elements were chosen due to relatively higher proportion of total weight percentage, and use as a detrital indicator (e.g. Al, Ti, Si ratios). The en-
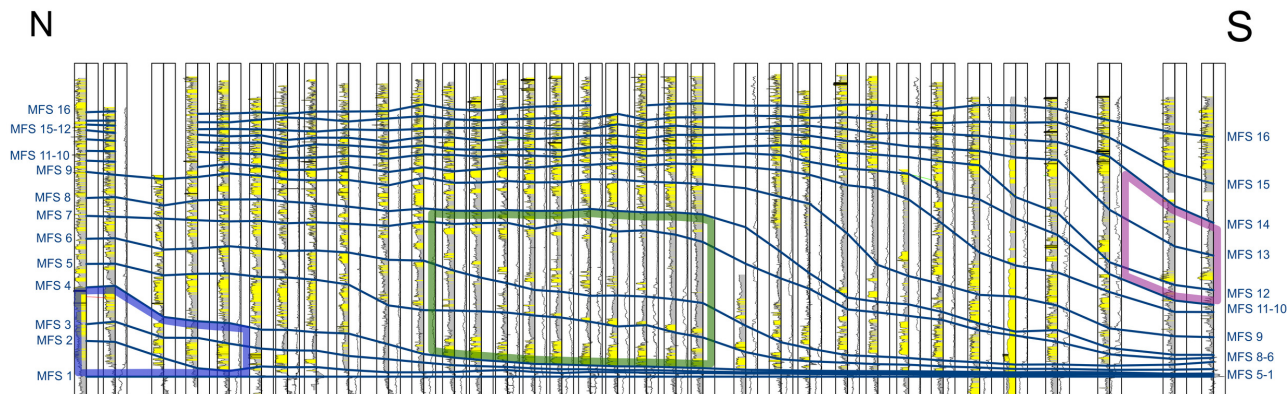
Figure 2. Stratigraphic framework developed by Carvjal (2006) based on maximum flooding surfaces (MFS).

Blue, green, and pink outlines represent the stratigraphic intervals of the 12 cores used in this study (Table 1). Blue outline denotes the MFS 2-4 interval for wells F042, E945, E952, E974, F041, S179, S821 (;eft). Green outline denotes the MFS 5-8 interval for wells F028, E934, E997 (central). Red outline denotes the MFS 12-14 interval for wells CEPO and PDRMT (right). Figure modified from cross section NS2 of Carvajal (2007); see location in Figure 1.

Table 1. Overview of the 12 cored wells that form the dataset for this study. Grain-size and XRF measurements collected every 0.1 and 0.5 feet, respectively, by this study. Porosity measurements obtained from published reports. All well names except PDRMT and CEPO are the identifiers used by the USGS Core Research Center.

	Well Name	Core Thickness (ft)	Grain Size Measurements (n)	Porosity Measurements (n)	XRF Points (n)	Carvajal MFS (Fig. 2)
1	F042	76.5	765	71	150	2 to 4
2	E945	61.2	612	34	122	2 to 4
3	E952	57.2	572	46	112	2 to 4
4	E974	61.3	613	41	123	2 to 4
5	F041	62.1	621	60	124	2 to 4
6	S179	52.2	522	52	105	2 to 4
7	S821	25.1	251	25	49	2 to 4
8	F028	116.1	1161	24	231	5 to 6
9	E934	41.1	411	0	83	5 to 8
10	E997	576.8	5768	153	1133	5 to 8
11	PDRMT	48.5	485	27	97	12 to 14
12	CEPO	34.1	341	26	67	12 to 14
	Total	1212.2	12122	559	2396	

tire elemental suite is available in the online supplementary material.

Porosity

Porosity measurements from core plugs were taken from legacy scanned PDF reports available on Wyoming Oil and Gas Conservation Commission and USGS Core Research Center websites. These data were not processed in any way to attempt to normalize for differing methodologies and standards between various laboratories, The largest potential differences are different laboratories, improved methodology with time, and the specific calculation methodology of porosity. The depths of porosity measurements were interpolated to the most reasonable decifoot (*e.g.*, 11-12 Ft. became 11.5 feet).

Image Data

Image data were collected by the authors and by USGS core research center staff on the cores in this study (table 1). Images have various core-tray sizes, amount of core, lighting conditions, and resolution. Due to this variability, we did not employ an automated ML model to trim, edit, and stack core images into a depth registered core column (*e.g.*, Meyer et al., 2020), but rather manually performed this task. To reduce errors from shadows and edge effects, we cropped the middle 60% of the core image for the analysis (Fig. 5). After the cropped image was manually depth registered, the standard (RGB format) images were converted to Hue-Saturation-Value values (HSV, Joblove & Greenberg, 1978; Smith, 1978) using the scikit-image package (Van Der Walt et al., 2014). This transformation reduces the differences caused due to lighting, shadows, and cam-
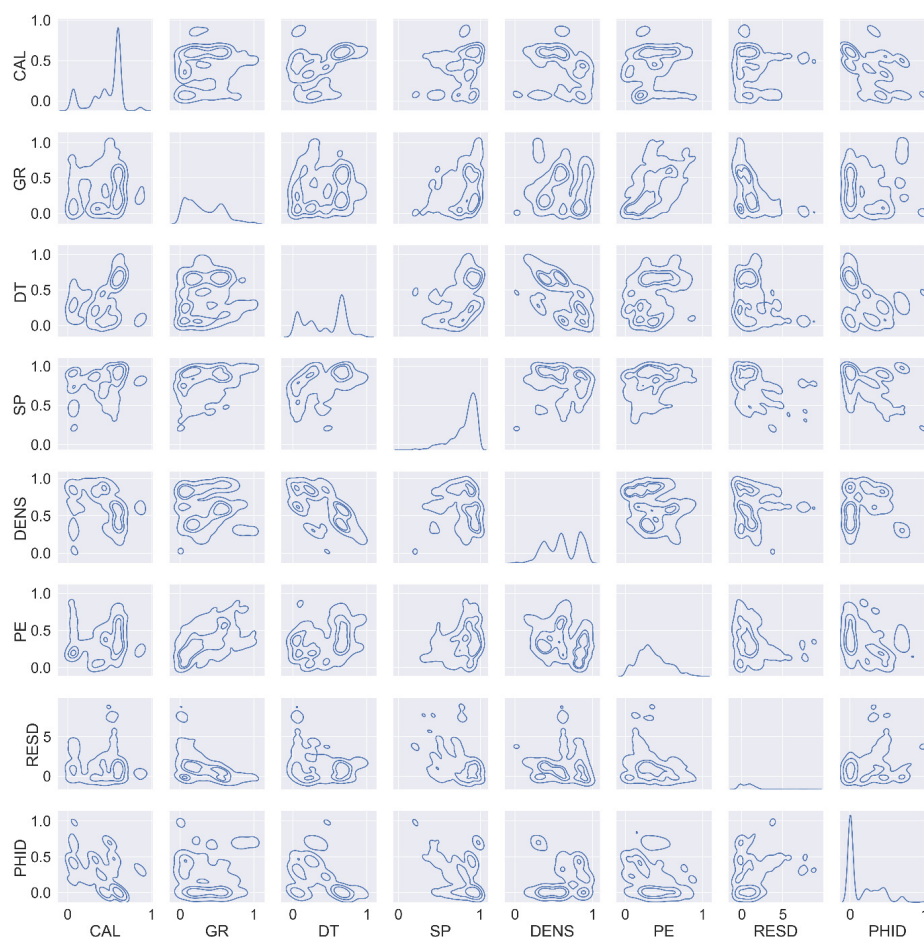
Figure 3. Pairplot of kernel density estimates of the complete well-log dataset (including imputed data) used in the study with three contours separating the data into quartiles.

Abbreviations are: CAL = Caliper, GR = Gamma Ray, DT= Sonic, SP = Spontaneous Potential, DENS = Density, PE = Photoelectric Factor, RESD = Deep Resistivity, PHID = Density Porosity. This plot is included to explore linearity between different well log data types, as additional available well-log data is available but was not used due to strong linear correlation.. Strongly correlated input variables can have downstream effects on machine learning because of duplicate information.

era type/setup. We tested using Red-Green-Blue (RGB) channels from the original image as input features rather than HSV, and RGB did not improve results in our testing because the RGB channels are highly correlated to the value channel. The HSV values for each discretized depth step (0.1 ft, Fig. 5) are then further reduced from the entire image to median and mode of H, S, and V values, along with the inter-quartile range of saturation and value. This compresses the amount of data for each depth step by 2-3 orders of magnitude compared to full image data; similar transformations were used by Martin et al. (2021) for core image data, and in both cases allow for efficient model runtime and use of computational infrastructure.

Core to Log offset

For studies that use both core and well-log derived properties for analysis, typically a core-to-log depth shift is performed to correct wireline stretch and core loss/breakage during drilling operations (Fontana et al., 2010; Jeong et al., 2020). In this study, we compared 3 different strategies: no offset, a qualitative offset, and a quantitative offset. The no offset is a base case, where no depth shifting was

done to either dataset. The qualitative depth shift was done by visually inspecting log patterns, and matching patterns. The quantitative offset utilizes a simple linear regression method to compare gamma-ray log and the grain size to further fine tune the qualitative offset to the nearest 0.1Ft. This linear regression methodology assumes that mudstones generally have a higher gamma-ray value compared to sandstones and testing shows that this method matches these well-log peaks to core-derived grain size quite well.

MACHINE LEARNING REGRESSION MODELS, EVALUATION AND PARAMETERS
Overview

This study focuses on prediction of core-derived properties using standard well-log and core-image summary data (Fig. 4, 5). We chose to explore five machine learning models, ranging in complexity from multiple linear regression to deep neural networks. The models were selected to show a range of algorithm complexity and structure. There are many ML models for supervised machine learning regression problems, but we chose models that have a Python in-
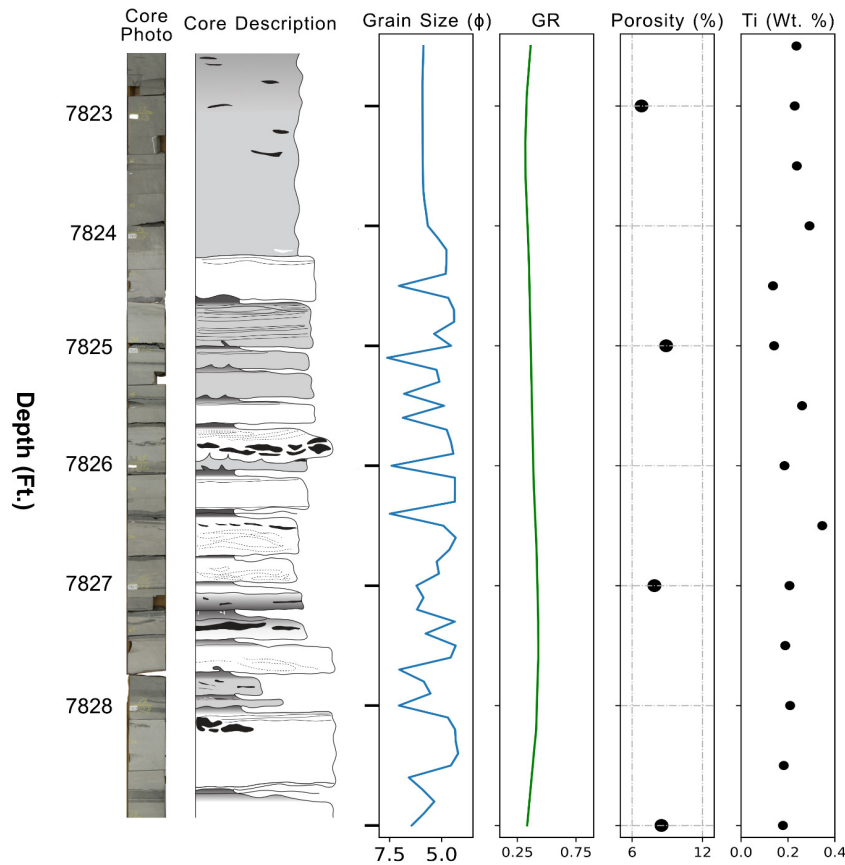
Figure 4. Subset of Data from F028, depth in feet.

From right to left; RGB (Red-Green-Blue) Core photo; core description modified from Koo (2015); grain size (in Phi); Gamma ray (In API); porosity (percentage), Titanium (Ti) from XRF.

terface, are actively maintained (i.e., public releases within the last twelve months), and have detailed documentation. There are many other models we could have chosen, and we hope other researchers will explore those models with the same dataset published with this study. A few examples of untested models are AdaBoost, LightGBM, and k-nearest neighbors, and we expect these models to perform similarly to models tested in this study. To replicate this research, the models and code are available at https://github.com/ThomasMGeo/LewisML (Martin, 2022). To install and use all code provided in the github repository, prior installation of some common Python packages is required, including Python 3.6+, Jupyter Notebooks (Kluyver et al., 2016), Lasio, Welly, Pandas (McKinney, 2010), Numpy (Harris et al., 2020), Scikit-Learn (Pedregosa et al., 2011), Scikit-Image (Van Der Walt et al., 2014), XGboost (Chen & Guestrin, 2016), Catboost (Prokhorenkova et al., 2018), and Tensorflow (Abadi et al., 2016).

Hyperparameter Tuning

Many ML models have hyperparameters that can be tuned to improve the learning process of the model (Mantovani et al., 2019, 2017). One example of a hyperparameter for XGBoost is the maximum depth of each decision tree; tuning these values often results in a lower model error compared to an untuned model (Table 2). We used a 'grid search' method to automatically perform hyperparameter tuning (Liashchynskyi & Liashchynskyi, 2019). Grid search is a parameter sweep method where each set of specific parameters are evaluated. While there are other methods for hyperparameter tuning, the 'grid search' method has built-in functionality for the five models used in this study. The bounds and hyperparameters that were tested are available in the GitHub repository (Martin, 2022).

Evaluating Regression Models

Evaluating machine learning models requires a scoring metric to evaluate model performance. To evaluate regression models, we use two main metrics: Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE, equations 1 & 2 respectively). Minimizing these two error types gives insight into model ranking and performance for skewed and non-gaussian datasets common to geology. This non-linearity is also why R^2 was not used, as we are comparing linear and non-linear models and datasets.

$$\text{RMSE} = \sqrt{(f-o)^2} \tag{1}$$

$$\text{MAE} = \frac{\sum_{i=1}^{n}|f-o|}{n} \tag{2}$$

Where f is predictions, o is observations, and n is number of errors (Chai & Draxler, 2014).

Utilizing both metrics gives a more complete insight into the error structure than using one alone. There are advan-
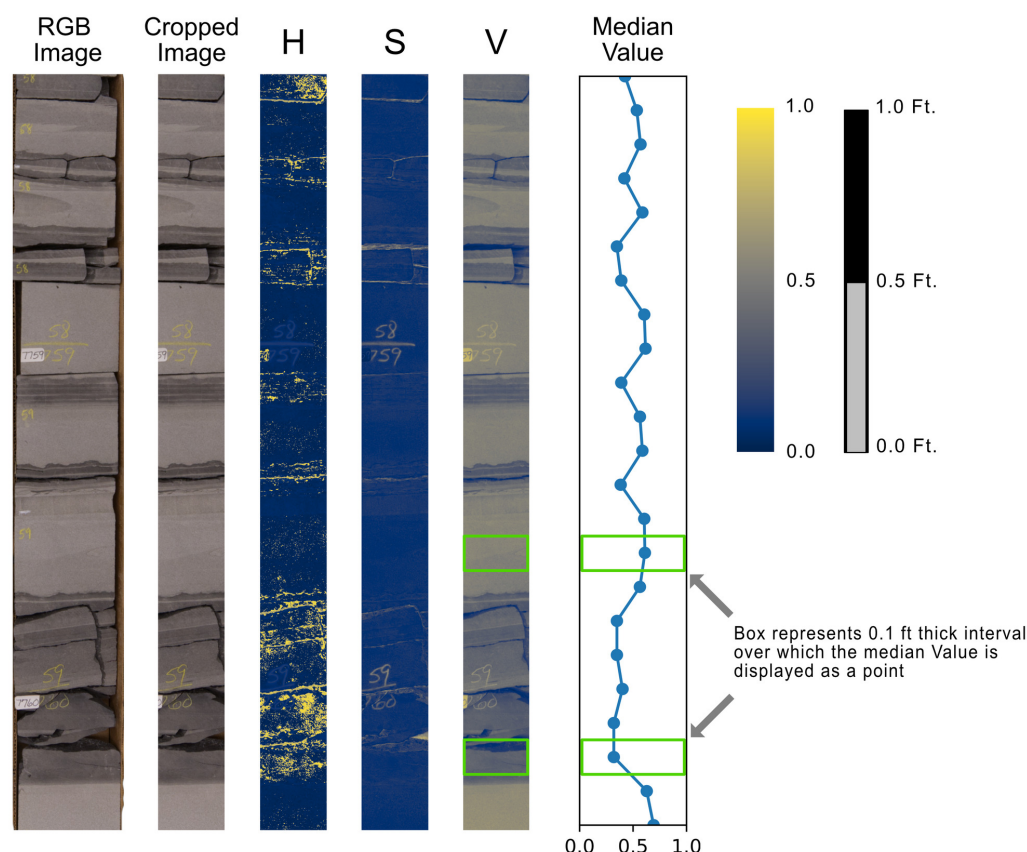
Figure 5. Core image data methodology.

The RGB image was cropped to 60% of its original pixel width and converted to HSV channels, with a plot shown for each channel. Core image from F028, 7758 to 7760 Ft. depth. Colorbar is equal for H (hue), S (saturation), and V (value) channels. Line plot (column 6) shows the median Value (V) for every 0.1 Ft depth interval.

tages and disadvantages to both metrics (Chai & Draxler, 2014; Willmott et al., 2009; Willmott & Matsuura, 2005). RMSE tends to have better assessment of performance when errors are gaussian and is also more flexible as a distance metric compared to MAE (Chai & Draxler, 2014). This manuscript will focus on RMSE but will use MAE as a differentiator if the RMSE is very similar.

MACHINE LEARNING MODELS
Multi-variate Linear Regression

Linear regression is a data analysis method that constructs a linear relationship between n variables (Montgomery et al., 2021). There are various ways to calculate error; for this study we used least-squares (Myers, 1990). Linear regression for this study utilized multiple variables (*e.g.*, the well-log curves in Fig. 3) and is thus termed multivariate regression (MVR). Major advantages of MVR are its computational efficiency, model interpretability (using R^2 or other metrics), and it has no need for hyperparameter tuning. One downside to MVR is the input variables need to be linear, and many geoscience relationships are non-linear; the non-linear to linear transformation creates information loss, which is sub-optimal.

Support Vector Machine

Support Vector Machine models (SVM) (Chang & Lin, 2011) have been commonly used since the 1990's and have been used with success to model physical and earth science processes (Okwuashi and Ndehehe, 1996; Balabin & Lomakina, 2011; Bermúdez et al., 2019; Pisel & Pyrcz, 2021; and others). While SVM's are common for classification tasks, we will be using a specific type of SVM built for regression, specifically the Support Vector Regression Machine (SVRM, Drucker et al., 1997). An advantage of SVRM is its robustness to outliers (Drucker et al., 1997), while a downside is the difficulty of scaling it to very large datasets.

XGBoost

XGBoost (Chen & Guestrin, 2016) is a gradient-boosting algorithm typically applied to decision trees (Quinlan, 1986). Gradient Boosted decision trees (GBDTs) are generally flexible for use with missing, noisy, and unscaled data. This makes them particularly useful for practical predictions on real geoscience data, although this algorithm can struggle on sparse data. XGboost has been used with great success for subsurface well-log classification tasks (Bormann et al., 2020; Dev & Eden, 2019; Hall & Hall, 2017; Martin et al., 2021).

Table 2. Results for predicting grain size, porosity, and XRF geochemistry using a 75:25 test:train random split.

Grainsize (phi units)	No Offset		Qualitative Core-log Offset		Quantitative Core-log Offset	
	RMSE	MAE	RMSE	MAE	RMSE	MAE
MVR	1.41	0.90	1.19	0.74	1.12	0.70
SVRM	1.14	0.41	0.97	0.24	0.89	0.23
SVRM w/HPT	1.29	1.17	1.23	1.07	1.23	1.06
XGBoost	0.76	0.21	0.70	0.18	0.68	0.15
XGBoost w/HPT	0.74	0.19	0.67	0.14	0.67	0.14
Catboost	0.92	0.42	0.80	0.32	0.76	0.26
Catboost w/HPT	0.75	0.23	0.68	0.17	0.68	0.17
DNN	1.08	0.38	0.92	0.26	0.89	0.28
Median	1.00	0.40	0.86	0.25	0.82	0.24

Porosity (porosity units, %)	No Offset		Qualitative Core-log Offset		Quantitative Core-log Offset	
	RMSE	MAE	RMSE	MAE	RMSE	MAE
MVR	2.99	1.60	2.78	1.45	2.46	1.43
SVRM	2.15	0.60	1.80	0.43	1.79	0.51
SVRM w/HPT	1.93	1.07	1.84	0.97	1.87	0.87
XGBoost	1.64	0.76	2.12	0.65	1.55	0.64
XGBoost w/HPT	1.60	0.84	1.88	0.64	1.61	0.52
Catboost	1.69	0.76	1.77	0.67	1.33	0.54
Catboost w/HPT	1.80	0.79	1.83	0.63	1.41	0.52
DNN	N/A	N/A	N/A	N/A	N/A	N/A
Median	1.80	0.79	1.84	0.65	1.61	0.54

Multi-XRF (combined wt. % Error)	No Offset		Qualitative Core-log Offset		Quantitative Core-log Offset	
	RMSE	MAE	RMSE	MAE	RMSE	MAE
MVR	1.33	0.67	1.27	0.61	1.25	0.59
SVRM	1.26	0.46	1.16	0.40	1.12	0.41
SVRM w/HPT	N/A	N/A	N/A	N/A	N/A	N/A
XGBoost	1.11	0.52	1.06	0.44	1.09	0.45
XGBoost w/HPT	N/A	N/A	N/A	N/A	N/A	N/A
Catboost	1.11	0.52	1.06	0.43	1.00	0.40
Catboost w/HPT	1.14	0.49	1.07	0.42	1.02	0.39
DNN	1.35	0.49	1.24	0.44	1.21	0.40
Median	1.20	0.50	1.11	0.44	1.11	0.41

The green cell denotes lowest error, and the yellow denotes second lowest error for each type of error calculation and each offset type. Multi-XRF is a combined score of the five predicted elements (Al, Ti, Si, Ca, Mg). Relative instability in results between the different data types could be from relative differences in dataset size, grainsize has the largest dataset while porosity is the smallest. The results are the median from three experiments for each model/dataset combination.

Catboost

Catboost (Prokhorenkova et al., 2018)) is another gradient boosting decision tree (GBDT) algorithm, but with some important differences as compared to XGBoost. The ability to use Catboost with noisy, unscaled data is unchanged from XGBoost, but Catboost reformulates the specific boosting algorithm to better honor the data distributions throughout the training cycle (Prokhorenkova et al., 2018). On a range of test datasets, Catboost has shown up to 12% improvement over XGBoost (Prokhorenkova et al., 2018). Catboost has recently been used for subsurface in-vestigations with success, particularly when compared to other methods (Tang et al., 2021; Zhong et al., 2021).

Deep Neural Networks

Deep Neural Networks (DNN) are a specific type of neural network, with many (2+) layers between the input and output layers. DNN can be used for either classification or regression on nonlinear data types and perform best with large (n > 100,000) datasets. One downside to DNNs is the lack of interpretability (i.e., feature importance, hyperparameters) compared to other ML methods like decision

trees. DNNs have been used successfully for geologic image recognition and classification problems (Jeong 2020; Chawshin et al., 2021; Lauper et al., 2021; Martin et al., 2021; Meyer et al., 2020; Falivene et al., 2022). DNN regression has also been used to estimate river sediment yield (Cigizoglu & Alp, 2005). The DNN in our study is two layers, with further model parameters available on the GitHub (Martin, 2022).

K Means Clustering

While this study is generally focused on regression-based prediction, the XRF data collected from the cores is quite useful for constraining basin evolution as well. Because multiple clinothems were sampled in varying spatial locations (Fig. 2), we chose k-means clustering to find similarities in the XRF data, which include ratios and amounts of specific elements. K-means clustering (Arthur & Vassilvitskii, 2007) is an unsupervised ML method that organizes data into clusters of similar data around a centroid. K-means algorithm minimizes the average squared distance between an observation and a cluster, while continually testing out new centroid locations until the optimum locations are discovered. One advantage and disadvantage to k-means is that the number of clusters must be chosen by the user. The advantage of a user defined number of clusters is you can match the same number to number of facies or other geologic classification scheme, but the disadvantage is that this selection can promote bias. K-means is also generally computationally efficient compared with supervised ML methods and does not require hyperparameter tuning.

RESULTS

Grain size

Table 2 shows the results for the five ML models predicting grainsize using well-log data (Fig 3.) reporting both RMSE and MAE, for each core-to-log offset. The two models that had the lowest RMSE were both gradient-boosted decision trees (GBDT), XGBoost and Catboost respectively. These two models had very similar results, but XGBoost with hyperparameter tuning had the lowest error for both RMSE and MAE for all core-log offsets (Table 2). The overall median error for all models was the lowest for quantitative offset compared to the qualitative and no offset cases. XGBoost had the lowest RMSE on the quantitative offset with an RMSE of 0.67. Deep neural networks performed similarly to SVRM. All models improved on multi-variate linear regression results.

Using core image data alone did not improve the ability to predict grain size demonstrably compared to the well log data with Catboost, as shown in Table 3. Combining both the well-log and core-image data also did not result in a significant decrease in RMSE or MAE compared to well-log data only. Issues with lighting, crack detection, and similar coloring of the different lithologies (specifically in the E997 core) were insurmountable for the ML model. We tested various pre-processing pipelines to help eliminate these issues (*e.g.*, variable lighting and legacy photo types) without success. Future core-image studies should prioritize even

Table 3. Results comparing image, well log, and combined results, from random test-train dataset separation.

Image Only	RMSE	MAE
Catboost	1.13	0.51
Catboost w/HPT	1.14	0.48
Well Log Only		
Catboost	0.71	0.19
Catboost w/HPT	0.69	0.15
Combined		
Catboost	0.70	0.20
Catboost w/HPT	0.68	0.19

The error here is the median value from three experiments. HPT= Hyperparameter tuning. Note the image only dataset preformed worse than models that used well log datasets.

and normalized lighting during data acquisition (cf. Martin et al., 2021).

Porosity

Table 2 shows the results for the four ML models, reporting both RMSE and MAE, for each offset on predicting plug-derived porosity measurements from well-log data (Fig. 3). The error did not reduce between the no offset to qualitative offset, but the quantitative offset had the best performance out of the three offsets. Along with GBDT, the SVRM performed very well for the qualitative offset, but not on the other two offsets. Deep neural networks were not utilized for porosity due to the small dataset size. E997 is a 'behind-the-outcrop' core, and the shallowest core in the dataset with porosity data. Due to the non-stationarity of these porosity data points for deeper and more relevant subsurface conditions, E997 was not used for testing or training datasets. Testing also revealed that including E997 resulted in a higher RMSE.

XRF elemental data

Table 2 shows the results for the five ML models predicting XRF measurements, reporting both RMSE and MAE, for each offset from well-log data (Fig. 3). XGBoost had the lowest (best) RMSE for each offset, with SVRM and Catboost with the second-best results, respectively. The DNN results are similar with the multi-variate linear regression results. The overall median was relatively similar between all offsets, but the qualitative offset had the lowest RMSE. Catboost had the lowest RMSE overall, with a combined RMSE of 1.00 of the five element panel (Al, Ti, Si, Ca, Mg) using the quantitative offset.

Test-Train Splits: Random Split vs. Blind Well

Figure 6C highlights the results for predicting grain size using Catboost, comparing the error from randomly test-train splitting of the dataset vs. using a single blind well as the test split (Fig. 6A-B). On the randomly split dataset,
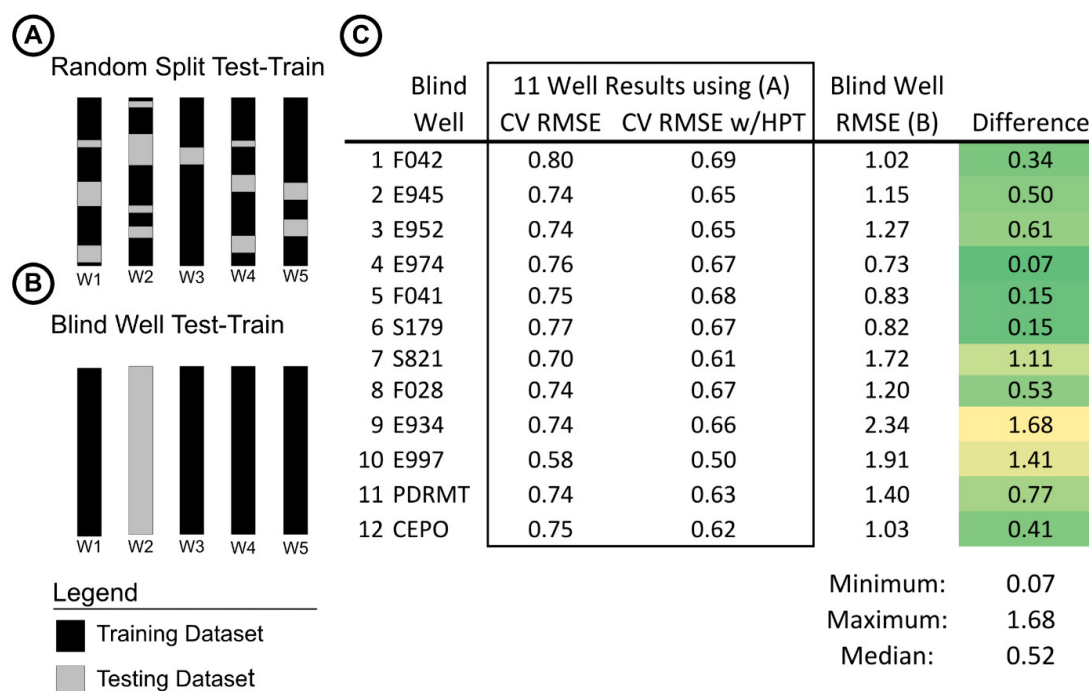
Figure 6. Comparison between random and blind-well test-train splits (median values from three model iterations).

Cross validation using random Split test-train **(A)** is common in ML research. However, with well-based geoscience applications, using a blind well **(B)** is more applicable for geoscientists wanting accurate error estimates on unseen wells. **(C)** Table Comparing prediction of grain size using test-train splits defined in (A) and (B). Each row is a single well that was held out from the twelve well dataset, with the other eleven wells being the randomly sampled testing-training dataset (using method A), with 25% of the data being held out for testing. For every blind well that was tested, the random split with inter-well data RMSE (column 3, Fig 6C) was always lower than the individual blind well prediction (ie the blind well RMSE is always higher than the random inter-well testing and training, column 5, Fig. 6C). While this lower RMSE using random split (Fig 6A) may seem better, it does not honor the true error structure for geoscience-based predictions on unseen wells. Thus, we advocate for using blind well testing.

25% of the data is held out for testing. Figure 6A illustrates visually how this is implemented. Using the RMSE from the randomly split testing dataset is overly optimistic (ie lower, column 3, Fig 6C) for each of the tests compared to the blind well, with the minimum RMSE increase of 0.05, the maximum of 2.02 and the median difference of 0.58 (Fig. 6C). This is only for one core-log offset and one ML model but highlights the possible variability in prediction results using different test-train split methodologies.

DISCUSSION

1. Machine Learning Model Selection

Many studies have focused on the best approaches and models for well-log classification and regression tasks, with gradient-boosted decision trees (GBDTs) commonly being the most effective models (Bormann et al., 2020; Hall & Hall, 2017). Our study generally confirms this, with Catboost and XGBoost occupying the first- or second-best model type <80% of the time and scoring as the winners of both the SEG 2016 and FORCE 2020 well-log classification contests (Bormann et al., 2020; Hall & Hall, 2017). GBDTs are generally flexible to missing and unscaled data and can be trained on a relatively small amount of data compared to a DNN due to model architecture. While each data type and offset had a ML model with the lowest error, the difference between the first-place and the second-place model is generally very small (1-5%, see Table 2). These differences have

a minor impact on final interpretation and utilization, so a ML model that is easier to implement and interpret may be preferred even if it is slightly less accurate. For example, while SVRMs have low error with smaller datasets, they are slower to train, and require more specific pre-processing compared to gradient boosted decision trees due to model architecture. Due to these reasons, a GBDT (Catboost, XGBoost) would generally be preferable. For this dataset, a second- or third-place decision-tree model has significant benefits over MVR due to the need to linearize MVR inputs. For subsurface investigations using well-logs, MVR is a useful baseline comparison due to computational efficiency and interpretability, but using MVR alone will generally underperform compared to other ML methods because well log data often have non-linear distributions.

While DNNs are popular for a variety of deep learning tasks, including image recognition, none of the DNN models had the lowest error in this study (Table 2). This poor performance, paired with a lack of clear interpretability suggests that DNNs are not recommended for datasets of this size, as larger, higher dimensional datasets might have increased performance using a DNN where interpretability is not a main factor for evaluation or use. For well-log analysis and prediction, DNN class of models would need a specialized dataset or research question to take advantage of the inherent model complexity and infrastructure. While further research will continue to refine which models and class of models have the best performance, future needs
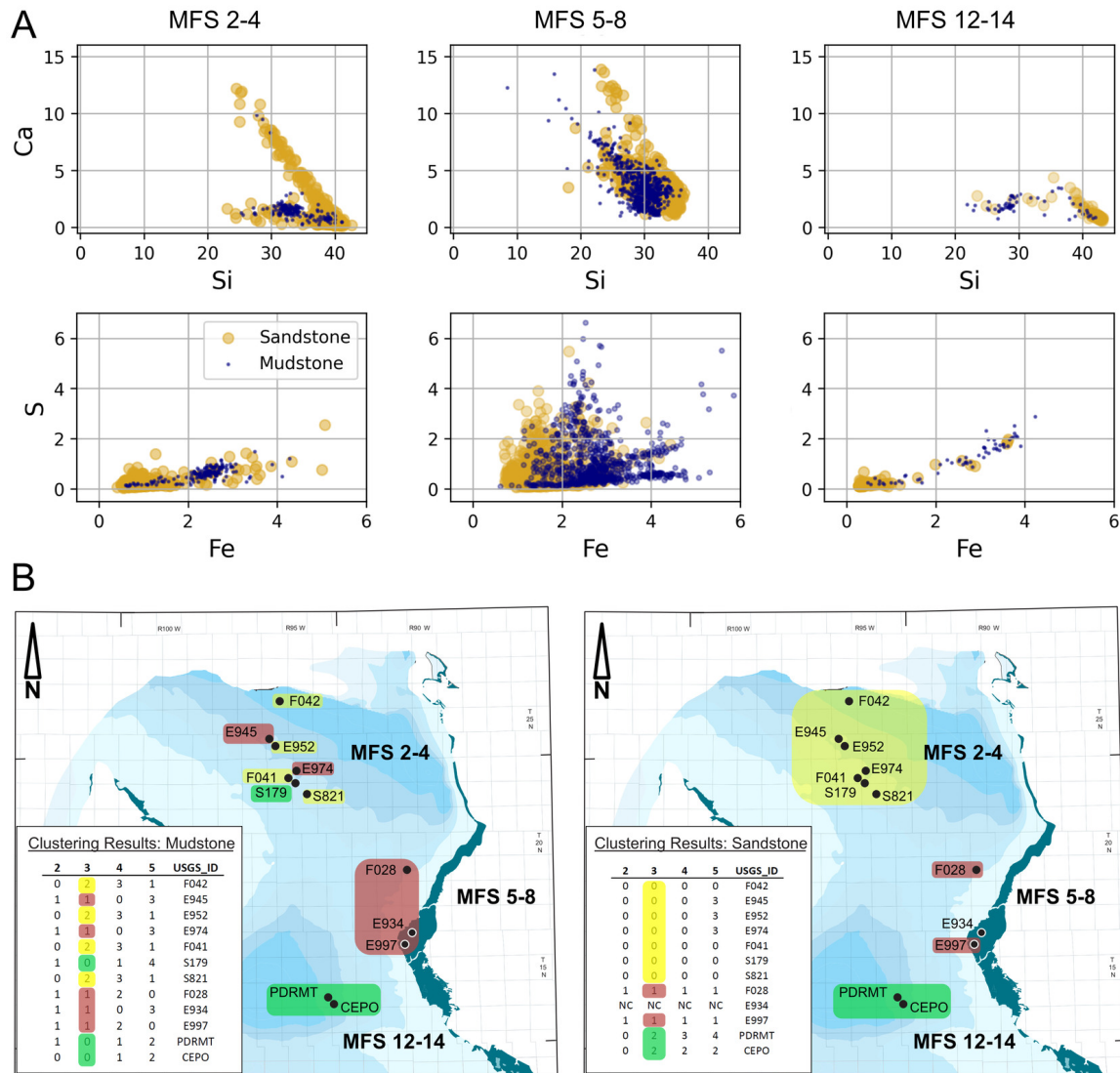
Figure 7. A) XRF Cross plots comparing calcium (Ca) to silicon (Si), and sulfur (S) to iron (Fe), grouped into Maximum Flooding Surface roups defined by Carvajal (2006) (Fig. 2).

The elevated calcium in the sandstone lithology is interpreted to be calcite cement visible in some of the cores. The elevated sulfur relative to iron in the MFS 5-8 group is interpreted to have relative anoxic conditions at time of deposition compared to the MFS 2-4 and 12-14 group (Dijkstra et al., 2016). B) K-means clustering results, using XRF geochemical ratios as input data. Mudstone (left) did not cluster into distinct spatiotemporal patterns, perhaps due to the ease of long-distance mud transport. However, sandstone (right) had spatiotemporal segmentation for both two and three clusters, suggesting a potential change in sediment routing or provenance through time. Three clusters are highlighted on the map for both mudstone and sandstone. E934 is completely mudstone and thus was not used for the sandstone clustering workflow.

for most geoscience professionals will include easier model deployment, interpretation, and dataset augmentation (Burkov, 2020; Curtis et al., 2020; Jahic et al., 2019; Molnar, 2022). Future researchers can build on the results this study to improve ML model parameterization and deployment practices.

2. Core Image Data

While core-images have been used for classification of facies and lithology primarily using DNNs, image-based ML methods require substantial computing resources (Chawshin et al., 2021; Jeong et al., 2020; Lauper et al., 2021; Martin et al., 2021; Pires de Lima et al., 2019; Solum et al., 2022; Zhang et al., 2021). Generally, these DNN work-flows are used on well curated datasets with standardized core imagery (*e.g.*, Martin et al., 2021). However, the core photographs in this study are more variable (*e.g.*, various cameras, formats, resolutions, and lighting conditions). For example, the core images in this study had visible dust, debris, labels, and other surface imperfections. Lighting is also generally inconsistent, causing variations in the HSV value from the top to the middle of the core boxes. The core-image data underperformed compared to other data types in this study, perhaps due to this image-quality variability. Furthermore, while some studies have had success using color to predict lithology (*e.g.*, Martin et al., 2021), color alone is not diagnostic of grain size. For example, the E997 core has dark grey mudstones but also mudstones that are light grey in color, similar to the color of sandstone in

this specific core (Fig. 4). Physical inspection of the core makes differentiating between grain sizes straightforward rather than relying on photographs alone. These issues emphasize the need for high resolution, depth-registered, fit-for-purpose core photography (*e.g.*, using methods of Meyer et al., 2020), preferably paired with another sensor type (hyperspectral, XRF, LiDAR) to differentiate between differences in lithology, grain size, and other useful geological indicators.

3. Test-Train Splits: Blind-Well Splits Honor Traditional Geoscience Workflows

Randomly splitting datasets into testing and training sets is common in ML research, especially for data consisting of independent measurements (*e.g.*, image datasets, web traffic). However, in the geological realm and particularly with borehole-based datasets, some of the measurements are correlated to each other and thus dependent. However, the identification of geologic parameters from well-logs is non-unique, thus calling into question the concept of random test-train splits.

Blind predictions (*e.g.*, predicting facies in a new well location) is needed for most geoscience applications. Using random test-train splits with multi-well datasets (*e.g.*, Fig. 6A) is still a valid way to test and verify ML modeling approaches and can be useful for specific well-based workflows such as predicting missing core data, confirming data quality, verifying previous interpretations, or other specific needs. However, the results from random, inter-well testing and training datasets will be overly optimistic on blind wells (Fig. 6) because every well (and corresponding well log) has different data conditions (age, wireline contractor, pressure, equipment, size, etc.) We expect slight changes within the Dad Sandstone facies during basin evolution, but it is the choice of the modeler to decide if the dataset has stationarity, and if the test-training splitting methodology is appropriate. For this study, the core dataset encompasses a wide range of Dad Sandstone facies due to the spatial and age variability of the wells we chose (Fig. 1C), and therefore we assume stationarity. But with these model types, extrapolating outside the range of test data is not recommended.

One potential issue with applying this methodology to other formations is increased complexity and heterogeneity within one formation compared to the Dad Sandstone/Lewis Shale. Mixed carbonate and siliciclastic deepwater systems have centimeter-scale geochemical and facies changes, which may affect model performance as compared to this study. Some formations will need to be split into sub-regional areas or temporal zones to maintain stationarity depending on research questions and dataset quality/quantity. While using ML can be a data-driven approach, many of the decisions done before a ML modeling workflow are subjective, but important decisions that require geoscience domain knowledge. Another reason for inter-well underperformance is that each well generally has different vendors collecting the downhole data, vintages, reservoir conditions, depths, and other variables that make it more difficult for the data to be directly transferrable between wells. With a more modern, curated dataset, the differences

would potentially be smaller than results presented in this study.

Future work exploring multiple facies in a single well-log, or other stacked succession multi-classification research might require different model types (Long short-term memory, convolutional neural networks, etc.). These types of models use the context above and below the region of interest for classification (*e.g.*, Martin et al., 2021) and are well-developed due to research for audio and time-series analysis. These model types also require full sequences of data and would not perform well with randomly split datasets. Exploring these model types and relating them to geological processes and the completeness of the stratigraphic record (*e.g.*, Sadler, 1981) is an exciting area of future research.

Whichever method is used for splitting data into testing and training datasets, it is important to describe the specific workflow that was performed, ideally with the release of code (*e.g.*, Martin, 2022) so that the work is reproducible for future ML research. These details could have an outsized impact on how the results from ML workflows would be generalizable and transferrable to truly blind data (Fig. 6).

4. Implications for chemostratigraphic heterogeneity prediction

While ML models can predict elemental data from well-log data (with a degree of error and uncertainty), the results presented in this study do not replace the need for original data collection for research questions focused on chemostratigrapic patterns and processes in deep-water systems, which are highly variable both spatially and temporally (i.e., in depth), and can change within centimeters. A limitation of this dataset is that XRF measurements were only collected every 0.5 Ft (15.24 cm), which cannot resolve the true lithologic and geochemical variability that is seen in core photos (Fig. 4).

We assume stationarity in the XRF data from the entire twelve-well dataset, a reasonable assumption for most practitioners because the data is within the same formation within the same basin. Exploring the k-means clustering results (Fig. 7) shows a spatial pattern for the sandstone samples, organizing into three groups that are geographically consistent (Fig. 7B). These clusters correlate with time of deposition as well, as the basin was filled north to south (Fig. 2; Carvajal, 2007; Carvajal & Steel, 2006; Koo, 2015). We tested various combinations of XRF elemental data as inputs to the k-means clustering algorithm, and the trends in Figure 7 were consistent with multiple experiments. The change in sandstone bulk geochemistry could be caused by temporal changes in source provenance, sediment routing, or diagenetic differences (*e.g.* Hunt et al., 2015). The basin sediment was primarily sourced from the north, but also had input from the west (Rock Springs Uplift, Hettinger & Roberts, 2005). Foreland basins commonly show temporal variability in transverse and axial sediment sourcing that affect sandstone composition (Koshnaw et al., 2020; Malkowski et al., 2016; Primm et al., 2018; Sharman et al., 2018; Szwarc et al., 2015), and the Green River

Basin may have had temporal changes in the three dominant source areas (Wind River Range, Granite Mountains, Rock Springs Uplift, Hettinger & Roberts, 2005). However, untangling these relationships for our dataset and basin modeling would require additional datatypes (*e.g.*, detrital zircon data) and sample locations throughout the basin, compared to our single North-South transect (Fig. 1). Different present day burial depths (~500 ft. for E997 to 13,000 ft. in the south) could also impact the differences we see today due to differential diagenesis.

The mudstone XRF data do not show strong spatial and/ or temporal trends with the available geochemical data. Using k-means clustering with 2, 3, 4, and 5 clusters, none of the resulting clusters were geologically meaningful in either relative age or location within the basin (Fig. 7). Changing the element and element ratios used in the clustering experiments significantly affected the cluster results; the results shown in figure 7B is just from one set of feature inputs. We attribute the lack of strong spatiotemporal clustering in mudstone XRF data to be related to (1) the relative ease of basin-wide mud transport in marine depositional systems as compared to sand (Drake et al., 1978), (2) a more consistent mud provenance through time (cf. Fildani & Hessler, 2005), (3) sub-regional diagenetic and anoxia changes that affect mudstone geochemistry more than sandstone. However, there are minor differences in mudstone elemental data through time that were not revealed in the clustering results. For example, the relatively enriched sulfur and calcium for the MFS 5-8 group suggests a more anoxic environment during time of deposition (cf. Dijkstra et al., 2016) as compared to the older and younger intervals (Fig. 7). This difference might be driven by being on the edge of the basin compared to the more central sediment fairway (F028, E934, E997, eastern edge of basin). Additional data in the center of the basin may shed light on these hypotheses, helping to improve our understanding of basin circulation dynamics and geochemical evolution. To capture finer-scale detail and vertical heterogeneity, a core-scanning workflow that collects depth-continuous data (*e.g.*, Birdwell et al., 2020) would be useful, and future work could compare results from a core scanner to the hand-held XRF data collected in this study.

5. Implications for stratigraphic heterogeneity prediction

This study provides a workflow and best practices to predict properties, including grain size, from well-log data. But grain size alone is not indicative of depositional process, sub-environment (levee, channel, lobe), level of confinement, and other stratigraphic information needed to make a holistic interpretation. We propose that this work should be used in conjunction with other more conventional interpretation methods (*e.g.*, graphic logging, facies classification) and other ML workflows to further constrain reservoir and geologic properties. For example, the ML workflow described here could be used to generate initial results for a basin-wide study focused on sediment routing and chemostratigraphy, upon which more detailed, bespoke studies could provide more accurate information about de-

positional sub-environments, depositional processes, chemofacies, and diagenesis.

CONCLUSIONS

This study utilizes a newly digitized twelve-well dataset from the Dad Sandstone Member of the Lewis Shale in the Greater Green River Basin, Wyoming to evaluate performance of five machine learning models and workflows to predict common core-derived properties from well logs and core images. The overall best performing models with the lowest RMSE are gradient-boosted decision trees (XG-Boost, CatBoost). These types of ML models are especially useful for geoscience problems, as they can handle both regression and classification while being less sensitive to unscaled, missing, and noisy data. For example, utilizing CatBoost model to predict grainsize, using well-log data the RMSE is 0.75 which is below the 5-model median RMSE of 1.00. Using core image data performed worse, with a RMSE of 1.13. The core-image regression result emphasizes the requirement to use high quality, fit for purpose core image data with consistent lighting. Core imaging in the future should be treated as a potential digital data type, and not just as a storage bookkeeping activity. The XRF data also yielded insights into basin evolution with the sandstone showing a strong spatiotemporal organization into north, central, south groupings, while the mudstones show a more mixed depositional pattern. This might be due to diagenetic, mixing, or other geochemical differences during time of deposition.

While the models developed in this study can predict grain size, XRF elemental measurements, and porosity (with a certain amount of error) from well logs and core images, this does not negate the need to collect more data in the future for other Dad Sandstone datasets, and other, more general deep-water stratigraphy datasets. More data would allow for better quality control of the specific ML research question, and confirmation that the models perform on a satisfactory level. The detail provided by predicted grain size every 0.1Ft is not the same level of detail as physical core inspection and description done manually by a trained sedimentologist (Fig. 4), which will still be needed as ML models are improved and refined for geoscience applications.

This research demonstrates the possibility of predicting both inter-well and blind-well core-derived reservoir properties from well logs and core images using fully open-source Python packages (Martin, 2022). Utilizing a blind well testing strategy compared to a random split will have large differences on results depending on use case. We recommend using blind-well holdouts for the majority of geoscience uses, as even for inter-well training and testing it will still be a valid measure of error. The research presented here will also guide further ML model choice for geoscientists and researchers alike. This workflow can be scaled from twelve wells in south-central Wyoming to thousands of wells with digital well-log data worldwide.

ACKNOWLEDGMENTS

References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., & Zheng, X. (2016). TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems. *ArXiv*. http://arxiv.org/abs/1603.04467

Arthur, D., & Vassilvitskii, S. (2007). K-means++: The advantages of careful seeding. *Proceedings of the Annual ACM-SIAM Symposium on Discrete Algorithms*, 1027–1035.

Asquith, D. O. (1970). Depositional Topography and Major Marine Environments, Late Cretaceous, Wyoming. *AAPG Bulletin*, *54*(7), 1184–1224. https://doi.org/10.1306/5d25cbb3-16c1-11d7-8645000102c1865d

Asquith, D. O. (1975). *Petroleum Potential of Deeper Lewis Washakie and Sand Wash Basins, Wyoming and Colorado* (pp. 159–162). Rocky Mountain Association of Geologists.

Balabin, R. M., & Lomakina, E. I. (2011). Support vector machine regression (SVR/LS-SVM)—an alternative to neural networks (ANN) for analytical chemistry? Comparison of nonlinear methods on near infrared (NIR) spectroscopy data. *The Analyst*, *136*(8), 1703–1712. https://doi.org/10.1039/c0an00387e

Bermúdez, M., Cea, L., & Puertas, J. (2019). A rapid flood inundation model for hazard mapping based on least squares support vector machine regression. *Journal of Flood Risk Management*, *12*(S1), 1–14. https://doi.org/10.1111/jfr3.12522

Birdwell, J. E., Fonteneau, L. C., & Martini, B. A. (2020). Mineralogy and lithology of the Upper Cretaceous Niobrara Formation determined by hyperspectral core imaging. *The Mountain Geologist*, *57*(2), 121–143. https://doi.org/10.31582/rmag.mg.57.2.121

Bormann, P., Aursand, P., Dilib, F., Manral, S., & Dischington, P. (2020). *FORCE 2020 Well well log and lithofacies dataset for machine learning competition* [Data set]. Zenodo. https://doi.org/10.5281/ZENODO.4351156

Bouma, A. (1962). *Sedimentology of Some Flysch Deposits: A Graphic Approach to Facies Interpretation*. Elsevier.

Burkov, A. (2020). *Machine learning engineering* (Vol. 1). True Positive Incorporated.

Cain, M. R. (1986). Depositional environment of Upper Cretaceous Lewis sandstones of the Lewis Shale, Sand Wash Basin, Colorado. *New Interpretations of Northwest Colorado Geology*, *171–181*.

Carvajal, C. R. (2007). *Sediment Volume Partitioning, Topset Processes and Clinoform Architecture. Understanding the Role of Sediment Supply, Sea Level and Delta Types in Shelf Margin Building and Deepwater Sand Bypass: The Lance-Fox Hills-Lewis System in S. Wyoming*. University of Texas. http://media.proquest.com/media/pq/classic/doc/1459905661/fmt/ai/rep/NPDF?hl=&cit:auth=Carvajal,+Cristian+Rene&cit:title=Sediment+volume+partitioning,+topset+processes+and+clinoform+…&cit:pub=ProQuest+Dissertations+and+Theses&cit:vol=&cit:

Carvajal, C. R., & Steel, R. (2006). Thick turbidite successions from supply-dominated shelves during sea-level highstand. *Geology*, *34*(8), 665–668. https://doi.org/10.1130/g22505.1

Carvajal, C. R., & Steel, R. (2012). Source-to-sink sediment volumes within a tectono-stratigraphic model for a Laramide shelf-to-deep-water basin: Methods and results. In C. Busby & A. Azor (Eds.), *Tectonics of Sedimentary Basins: Recent Advances* (pp. 131–151). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781444347166.ch7

Chai, T., & Draxler, R. R. (2014). Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature. *Geoscientific Model Development*, *7*(3), 1247–1250. https://doi.org/10.5194/gmd-7-1247-2014

Chang, C.-C., & Lin, C.-J. (2011). LIBSVM: A Library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, *2*(3), 1–39. https://doi.org/10.1145/1961189.1961199

Chawshin, K., Gonzalez, A., Berg, C. F., Varagnolo, D., Heidari, Z., & Lopez, O. (2021). Classifying lithofacies from textural features in whole core CT-scan images. *SPE Reservoir Evaluation and Engineering*, *24*(02), 341–357. https://doi.org/10.2118/205354-pa

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 13-17-,785-794. https://doi.org/10.1145/2939672.2939785

Cigizoglu, H. K., & Alp, M. (2005). Generalized regression neural network in modelling river sediment yield. *Advances in Engineering Software*, *37*(2), 63–68. https://doi.org/10.1016/j.advengsoft.2005.05.002

Compton, R. R. (2016). *Geology in the Field* (2nd ed.). Earthspun Books.

Curtis, A., O'Malley, D., Beroza, G. C., Johnson, P. A., Li, E., & Haber, E. (2020). Tackling 21st century geoscience problems with machine learning. *Eos*, *101*. https://doi.org/10.1029/2020eo150184

Dev, V. A., & Eden, M. R. (2019). Gradient Boosted Decision Trees for Lithology Classification. *Computer Aided Chemical Engineering*, 978–970. https://doi.org/10.1016/b978-0-12-818597-1.50019-9

Dijkstra, N., Slomp, C. P., & Behrends, T. (2016). Vivianite is a key sink for phosphorus in sediments of the Landsort Deep, an intermittently anoxic deep basin in the Baltic Sea. *Chemical Geology*, *438*, 58–72. https://doi.org/10.1016/j.chemgeo.2016.05.025

Drake, D. E., Hatcher, P. G., & Keller, G. H. (1978). In D. J. Stanley & G. Kelling (Eds.), *Sedimentation in Submarine Canyons, Fans and Trenches* (pp. 33–41). Dowden Hutchinson and Ross.

Dramsch, J. S. (2020). 70 Years of Machine Learning in Geoscience in Review. *Advances in Geophysics*, *61*, 1–55. https://doi.org/10.1016/bs.agph.2020.08.002

Drucker, H., Surges, C. J. C., Kaufman, L., Smola, A., & Vapnik, V. (1997). Support vector regression machines. *Advances in Neural Information Processing Systems*, *1*, 155–161.

Eriavbe, F. E., & Okene, U. O. (2019, August 5). *Machine learning application to permeability prediction using log & core measurements: A realistic workflow application for reservoir characterization*. Society of Petroleum Engineers - SPE Nigeria Annual International Conference and Exhibition 2019, NAIC 2019. https://doi.org/10.2118/198874-ms

Fildani, A., & Hessler, A. M. (2005). Stratigraphic record across a retroarc basin inversion: Rocas Verdes–Magallanes Basin, Patagonian Andes, Chile. *Geological Society of America Bulletin*, *117*(11), 1596. https://doi.org/10.1130/b25708.1

Fontana, E., Iturrino, G. J., & Tartarotti, P. (2010). Depth-shifting and orientation of core data using a core–log integration approach: A case study from ODP–IODP Hole 1256D. *Tectonophysics*, *494*(1–2), 85–100. https://doi.org/10.1016/j.tecto.2010.09.006

Hall, M., & Hall, B. (2017). Distributed collaborative prediction: Results of the machine learning contest. *The Leading Edge*, *36*(3), 267–269. https://doi.org/10.1190/tle36030267.1

Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., del Río, J. F., Wiebe, M., Peterson, P., … Oliphant, T. E. (2020). Array programming with NumPy. *Nature*, *585*(7825), 357–362. https://doi.org/10.1038/s41586-020-2649-2

Haughton, P., Davis, C., McCaffrey, W., & Barker, S. (2009). Hybrid sediment gravity flow deposits – Classification, origin and significance. *Marine and Petroleum Geology*, *26*(10), 1900–1918. https://doi.org/10.1016/j.marpetgeo.2009.02.012

Hettinger, R. D., & Roberts, L. N. R. (2005). Lewis total petroleum system of the southwestern Wyoming province, Wyoming, Colorado, and Utah. In *USGS Digital Data Series: Vol. DDS-69-D*.

Hicks, J., & Adrian, B. (2009). Core Research Center. *U.S. Geological Survey*. https://doi.org/10.3133/gip94

Hunt, J. E., Croudace, I. W., & MacLachlan, S. E. (2015). Use of Calibrated ITRAX XRF Data in Determining Turbidite Geochemistry and Provenance in Agadir Basin, Northwest African Passive Margin. *Micro-XRF Studies of Sediment Cores*, 127–146. https://doi.org/10.1007/978-94-017-9849-5_4

Jahic, B., Guelfi, N., & Ries, B. (2019). Software engineering for dataset augmentation using generative adversarial networks. *Proceedings of the IEEE International Conference on Software Engineering and Service Sciences, ICSESS*, 59–66. https://doi.org/10.1109/icsess47205.2019.9040806

Jeong, J., Park, E., Emelyanova, I., Pervukhina, M., Esteban, L., & Yun, S.-T. (2020). Interpreting the Subsurface Lithofacies at High Lithological Resolution by Integrating Information From Well-Log Data and Rock-Core Digital Images. *Journal of Geophysical Research: Solid Earth*, *125*(2), 1–17. https://doi.org/10.1029/2019jb018204

Jobe, Z., Howes, N., Meyer, R., Martin, J., Coutts, D., & Hou, P. (2021). Sedimentary Graphic Logs: A Toolkit for Digitalization and a Template for Standardized Description. *Geological Society of America Abstracts with Programs*, *19*(3), 15–29. https://doi.org/10.1130/abs/2020am-354696

Joblove, G. H., & Greenberg, D. (1978). Color Spaces for Computer Graphics. *ACM SIGGRAPH Computer Graphics*, *12*(3), 20–25. https://doi.org/10.1145/965139.807362

Johnson, P. L., & Andersen, D. W. (2009). Concurrent growth of uplifts with dissimilar orientations in the southern green river Basin, wyoming: Implications for paleocene- eocene patterns of foreland shortening. *Rocky Mountain Geology*, *44*(1), 1–16. https://doi.org/10.2113/gsrocky.44.1.1

Kluyver, T., Ragan-Kelley, B., Pérez, F., Granger, B., Bussonnier, M., Frederic, J., Kelley, K., Hamrick, J., Grout, J., Corlay, S., Ivanov, P., Avila, D., Abdalla, S., & Willing, C. (2016). Jupyter Notebooks—a publishing format for reproducible computational workflows. *Positioning and Power in Academic Publishing: Players, Agents and Agendas - Proceedings of the 20th International Conference on Electronic Publishing, ELPUB 2016*, 87–90. https://doi.org/10.3233/978-1-61499-649-1-87

Koo, W. M. (2015). *Architecture and Evolution of Submarine-Fans, and Coupling with Shelf-Edge Processes in Supply-Dominated Margins: Example from Maastrichtian Washakie Basin*. University of Texas at Austin.

Koo, W. M., Olariu, C., Steel, R. J., Olariu, M. I., Carvajal, C. R., & Kim, W. (2016). Coupling between shelf-edge architecture and submarine-fan growth style in a supply-dominated margin. *Journal of Sedimentary Research*, *86*(6), 613–628. https://doi.org/10.2110/jsr.2016.42

Koshnaw, R. I., Horton, B. K., Stockli, D. F., Barber, D. E., & Tamar-Agha, M. Y. (2020). Sediment routing in the Zagros foreland basin: Drainage reorganization and a shift from axial to transverse sediment dispersal in the Kurdistan region of Iraq. *Basin Research*, *32*(4), 688–715. https://doi.org/10.1111/bre.12391

Krumbein, W. C. (1938). Size Frequency Distributions of Sediments and the Normal Phi Curve. *SEPM Journal of Sedimentary Research*, *Vol. 8*(3), 84–90. https://doi.org/10.1306/d4269008-2b26-11d7-8648000102c1865d

Lauper, B., Zimmerli, G. N., Jaeggi, D., Deplazes, G., Wohlwend, S., Rempfer, J., & Foubert, A. (2021). Quantification of Lithological Heterogeneity Within Opalinus Clay: Toward a Uniform Subfacies Classification Scheme Using a Novel Automated Core Image Recognition Tool. *Frontiers in Earth Science*, *9*(May), 1–25. https://doi.org/10.3389/feart.2021.645596

Levon, T., & Mazza, J. (2020). A Multi-Disciplinary Modeling Approach to Determine Economic Viability of Infill Horizontal Well Development in the Green River Basin. *Proceedings of the 8th Unconventional Resources Technology Conference*, 1–20. https://doi.org/10.15530/urtec-2020-2968

Liashchynskyi, P., & Liashchynskyi, P. (2019). *Grid Search, Random Search, Genetic Algorithm: A Big Comparison for NAS*. 1–11. http://arxiv.org/abs/1912.06059

Liu, S., & Nummedal, D. (2004). Late Cretaceous subsidence in Wyoming: Quantifying the dynamic component. *Geology*, *32*(5), 397–400. https://doi.org/10.1130/g20318.1

Lowe, D. R. (1982). Sediment gravity flows: II. Depositional models with special reference to the deposits of high-density turbidity currents. *Journal of Sedimentary Petrology*, *Vol. 52*(1), 279–297. https://doi.org/10.1306/212f7f31-2b24-11d7-8648000102c1865d

Malkowski, M. A., Schwartz, T. M., Sharman, G. R., Sickmann, Z. T., & Graham, S. A. (2016). Stratigraphic and provenance variations in the early evolution of the Magallanes-Austral foreland basin: Implications for the role of longitudinal versus transverse sediment dispersal during arc-continent collision. *Geological Society of America Bulletin*, *129*(3–4), 1–23. https://doi.org/10.1130/b31549.1

Mantovani, R. G., Horvath, T., Cerri, R., Vanschoren, J., & De Carvalho, A. C. P. L. F. (2017). Hyper-Parameter Tuning of a Decision Tree Induction Algorithm. *Proceedings - 2016 5th Brazilian Conference on Intelligent Systems, BRACIS 2016*, 37–42. https://doi.org/10.1109/bracis.2016.018

Mantovani, R. G., Rossi, A. L. D., Alcobaça, E., Vanschoren, J., & de Carvalho, A. C. P. L. F. (2019). A meta-learning recommender system for hyperparameter tuning: Predicting when tuning improves SVM classifiers. *Information Sciences*, *501*, 193–221. https://doi.org/10.1016/j.ins.2019.06.005

Maries, G., Malehmir, A., Bäckström, E., Schön, M., & Marsden, P. (2017). Downhole physical property logging for iron-oxide exploration, rock quality, and mining: An example from central Sweden. *Ore Geology Reviews*, *90*, 1–13. https://doi.org/10.1016/j.oregeorev.2017.10.012

Martin, T. (2022). *LewisML*. Github: Https://Github.Com/ThomasMGeo/LewisML. https://doi.org/10.5281/zenodo.5847637

Martin, T., Meyer, R., & Jobe, Z. (2021). Centimeter-Scale Lithology and Facies Prediction in Cored Wells Using Machine Learning. *Frontiers in Earth Science*, *9*(June), 1–18. https://doi.org/10.3389/feart.2021.659611

McKinney, W. (2010). Data Structures for Statistical Computing in Python. *Proceedings of the 9th Python in Science Conference*, *1*(Scipy), 56–61. https://doi.org/10.25080/majora-92bf1922-00a

Meyer, R., Martin, T., & Jobe, Z. (2020). CoreBreakout: Subsurface Core Images to Depth-Registered Datasets. *Journal of Open Source Software*, *5*(50), 1969. https://doi.org/10.21105/joss.01969

Molnar, C. (2022). *Interpretable Machine Learning: A Guide for Making Black Box Models Explainable* (2nd ed.). https://christophm.github.io/interpretable-ml-book/

Montgomery, D. C., Peck, E., & Vining, G. G. (2021). *Introduction to Linear Regression Analysis* (6th ed.). Wiley. https://www.wiley.com/en-us/Introduction+to+Linear+Regression+Analysis%2C+6th+Edition-p-9781119578727

Myers, R. H. (1990). *Classical and modern regression with applications (Second)*. PWS-KENT Publishing Co.

Olariu, M. L., Carvajal, C. R., Olariu, C., & Steel, R. J. (2012). Deltaic process and architectural evolution during cross-shelf transits, Maastrichtian Fox Hills Formation, Washakie Basin, Wyoming. *AAPG Bulletin*, *96*(10), 1931–1956. https://doi.org/10.1306/03261211119

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Müller, A., Nothman, J., Louppe, G., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Cournapeau, D., Brucher, M., & Perrot, M. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830. http://scikit-learn.org

Pires de Lima, R., Suriamin, F., Marfurt, K. J., & Pranter, M. J. (2019). Convolutional neural networks as aid in core lithofacies classification. *Interpretation*, *7*(3), SF27–SF40. https://doi.org/10.1190/int-2018-0245.1

Pisel, J. R., & Pyrcz, M. J. (2021). Classifying basin-scale stratigraphic geometries from subsurface formation tops with machine learning. *The Depositional Record*, *7*(1), 64–76. https://doi.org/10.1002/dep2.129

Primm, J. W., Johnson, C. L., & Stearns, M. (2018). Basin-axial progradation of a sediment supply driven distributive fluvial system in the Late Cretaceous southern Utah foreland. *Basin Research*, *30*(2), 249–278. https://doi.org/10.1111/bre.12252

Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V., & Gulin, A. (2018). Catboost: Unbiased boosting with categorical features. In *Advances in Neural Information Processing Systems* (Vol. 4, pp. 6638–6648).

Pyles, D. R. (2000). *A high-frequency sequence stratigraphic framework for the Lewis shale and Fox Hills sandstone, Great Divide and Washakie basins, Wyoming*. Colorado School of Mines. https://mines.primo.exlibrisgroup.com/permalink/01COLSCHL_INST/jlttbh/alma993061323502341

Pyles, D. R., & Slatt, R. M. (2007). Stratigraphy of the Lewis Shale, Wyoming, USA: applications to understanding shelf-edge to base-of-slope changes in stratigraphic architecture of prograding basin margins. *Atlas of Deep-Water Outcrops*, 485–489. https://doi.org/10.1306/12401031St562471

Pyles, D. R., Syvitski, J. P. M., & Slatt, R. M. (2011). Defining the concept of stratigraphic grade and applying it to stratal (reservoir) architecture and evolution of the slope-to-basin profile: An outcrop perspective. *Marine and Petroleum Geology*, *28*(3), 675–697. https://doi.org/10.1016/j.marpetgeo.2010.07.006

Quinlan, J. R. (1986). Induction of decision trees. *Machine Learning*, *1*(1), 81–106. https://doi.org/10.1007/bf00116251

Raschka, S. (2018). *Model Evaluation, Model Selection, and Algorithm Selection in Machine Learning*. http://arxiv.org/abs/1811.12808

Sadler, P. M. (1981). Sediment accumulation rates and the completeness of stratigraphic sections. *The Journal of Geology*, *89*(5), 569–584. https://doi.org/10.1086/628623

Sapardina, D. (2012). *Contrasting Facies in Slope and Basin-floor Deposits that Correspond to Rising and Flat Shelf Edge Trajectories, Lewis Shale, Washakie Basin, Wyoming*. Colorado School of Mines. https://mines.primo.exlibrisgroup.com/permalink/01COLSCHL_INST/jlttbh/alma996307683502341

Sharman, G. R., Hubbard, S. M., Covault, J. A., Hinsch, R., Linzer, H.-G., & Graham, S. A. (2018). Sediment routing evolution in the North Alpine Foreland Basin, Austria: Interplay of transverse and longitudinal sediment dispersal. *Basin Research*, *30*(3), 426–447. https://doi.org/10.1111/bre.12259

Smith, A. R. (1978). Color gamut transform pairs. *Proceedings of the 5th Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH 1978*, *2*, 12–19. https://doi.org/10.1145/800248.807361

Solum, J. G., Auchter, N. C., Falivene, O., Cilona, A., Kleipool, L., & Zarian, P. (2022). Accelerating core characterization and interpretation through deep learning with an application to legacy data sets. *Interpretation*, *10*(3). https://doi.org/10.1190/int-2021-0189.1

Stumm, F., & Como, M. D. (2017). Delineation of salt water intrusion through use of electromagnetic-induction logging: A case study in Southern Manhattan Island, New York. *Water*, *9*(9), 1–17. https://doi.org/10.3390/w9090631

Szwarc, T. S., Johnson, C. L., Stright, L. E., & McFarlane, C. M. (2015). Interactions between axial and transverse drainage systems in the Late Cretaceous Cordilleran foreland basin: Evidence from detrital zircons in the Straight Cliffs Formation, southern Utah, USA. *Geological Society of America Bulletin*, *127*(3–4), 372–392. https://doi.org/10.1130/b31039.1

Talling, P. J., Masson, D. G., Sumner, E. J., & Malgesini, G. (2012). Subaqueous sediment density flows: Depositional processes and deposit types. *Sedimentology*, *59*(7), 1937–2003. https://doi.org/10.1111/j.1365-3091.2012.01353.x

Tang, J., Fan, B., Xiao, L., Tian, S., Zhang, F., Zhang, L., & Weitz, D. (2021). A new ensemble machine-learning framework for searching sweet spots in shale reservoirs. *SPE Journal*, *26*(01), 482–497. https://doi.org/10.2118/204224-pa

Van Der Walt, S., Schönberger, J. L., Nunez-Iglesias, J., Boulogne, F., Warner, J. D., Yager, N., Gouillart, E., & Yu, T. (2014). Scikit-image: Image processing in python. *PeerJ*, *2014*(1), 1–18. https://doi.org/10.7717/peerj.453

van Horn, M. D., & Shannon, L. T. (1989). Hay reservoir field; a submarine fan gas reservoir within the Lewis Shale, Sweetwater County, Wyoming. *United States Guidebook Wyoming Geological Association*, *40*, 155–180.

Wallis, I., McCormick, S., Sewell, S., & Boseley, C. (2009). Formatioon Assessment in Geothermal Ussing Wireline Tools – Application and Early Results from the Nggatamarikii Geothermmal Field, New Zealand. *New Zealand Geothermal Workshop 2009 Proceedings*, 1–8.

Wentworth, C. K. (1922). A Scale of Grade and Class Terms for Clastic Sediments. *The Journal of Geology*, *30*(5), 377–392. https://doi.org/10.1086/622910

Williams, J. H., & Lane, J. W., Jr. (1998). Advances in Borehole Geophysics for Ground-Water Investigations. *US Geological Survey Bulletin*. https://doi.org/10.3133/fs00298

Willmott, C. J., & Matsuura, K. (2005). Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate Research*, *30*(1), 79–82. https://doi.org/10.3354/cr030079

Willmott, C. J., Matsuura, K., & Robeson, S. M. (2009). Ambiguities inherent in sums-of-squares-based error statistics. *Atmospheric Environment*, *43*(3), 749–752. https://doi.org/10.1016/j.atmosenv.2008.10.005

Winn, R. D., Bishop, M. G., & Gardner, P. S. (1987). Shallow-water and sub- storm-base deposition of Lewis Shale in Cretaceous Western Interior seaway, south-central Wyoming. *American Association of Petroleum Geologists Bulletin*, *71*(7), 859–881. https://doi.org/10.1306/948878ca-1704-11d7-8645000102c1865d

Yonkee, W. A., & Weil, A. B. (2015). Tectonic evolution of the Sevier and Laramide belts within the North American Cordillera orogenic system. *Earth-Science Reviews*, *150*, 531–593. https://doi.org/10.1016/j.earscirev.2015.08.001

Young, K. E., Evans, C. A., Hodges, K. V., Bleacher, J. E., & Graff, T. G. (2016). A review of the handheld X-ray fluorescence spectrometer as a tool for field geologic investigations on Earth and in planetary surface exploration. *Applied Geochemistry*, *72*, 77–87. https://doi.org/10.1016/j.apgeochem.2016.07.003

Zhang, J., Ambrose, W., & Xie, W. (2021). Applying convolutional neural networks to identify lithofacies of large-n cores from the Permian Basin and Gulf of Mexico: The importance of the quantity and quality of training data. *Marine and Petroleum Geology*, *133*(June), 105307. https://doi.org/10.1016/j.marpetgeo.2021.105307

Zhong, C., Geng, F., Zhang, X., Zhang, Z., Wu, Z., & Jiang, Y. (2021). Shear Wave Velocity Prediction of Carbonate Reservoirs Based on CatBoost. *2021 4th International Conference on Artificial Intelligence and Big Data (ICAIBD)*, 622–626. https://doi.org/10.1109/icaibd51990.2021.9459061

Supplementary Materials

Supplementary datasets

Download: https://thesedimentaryrecord.scholasticahq.com/article/36638-digitalization-of-legacy-datasets-and-machine-learning-regression-yields-insights-for-reservoir-property-prediction-and-submarine-fan-evolution-a-sub/attachment/94158.zip