Machine Learning Assisted Identification of Hydrocarbon Storage and Transport Pathways and Kerogen in Wolfcamp versus Barnett Shales

Eliza Ganguly, Dr. Siddharth Misra

Texas A&M University

For queries, please contact Dr Sid Misra (misra@tamu.edu)

Abstract

A robust machine learning (ML) workflow was developed to identify kerogen and hydrocarbon storage and transport pathways in the scanning electron microscopy (SEM) maps of organic-rich shale samples from two different formations, namely Wolfcamp Shale and Barnett Shale. Kerogen is most often wrongly identified as pores/cracks are poor when the machine learning model is developed on one shale formation and applied on the other shale formation. ML workflow developed only on Barnett shale cannot detect cracks in the Wolfcamp samples. The best performing machine learning approach learnt from both the formations and exhibits an average F1 scores of 0.99 and 0.91 on the inner-region and outer-region pixels, respectively. The machine learning workflow performs better on Barnett as compared to

Wolfcamp. Barnett shale, in comparison to Wolfcamp shale, provides better generalizable features and more complex microstructural aspects that are harder to identify. Overall, it is easier to identify kerogen as compared to pores and cracks due to their distribution, availability, connectivity, and pixel intensity. By learning from merely 50 thousand pixels with corresponding labels, the proposed machine learning workflow can successfully identify the hydrocarbon storage and transport pathways and kerogen in a large 20-GB dataset containing approximately 10 billion pixels.

1. Introduction

Microstructure determines the transport, chemical and mechanical properties of a material. High-resolution microscopy images capture the microstructural topology and morphology of various material constituents [1]. Analysis of microscopy image provides better understanding and quantification of the microstructural aspects of material constituents. For identifying, detecting, and locating a certain material constituent, it is crucial to perform image segmentation. Image segmentation partitions an image into segments by assigning each pixel a label. All the pixels having same label share certain common characteristics. Popular methods used for image segmentation are thresholding, region growing, clustering, edge detection, watershed, artificial neural network and partial-differential-equation based

techniques [2]. In this study, a robust machine learning (ML) workflow is developed to identify hydrocarbon storage and transport pathways (i.e. pores and cracks) and organic/kerogen material, which requires simultaneous identification of matrix and pyrite constituents, present in the scanning electron microscopy (SEM) images of organic-rich shales.

1.1 Image Segmentation Using Machine Learning

Compared to physics-driven or human-led efforts, data-driven models developed using machine learning can incorporate higher-order non-linear trends to accomplish iterative, manual, time-consuming, complex tasks in a much shorter duration while delivering more accurate results [3]. Andrew (2018) found the machine learning-based image segmentation to have a relatively high tolerance to noise and a considerably low misclassification rate, as compared to two traditional segmentation methods [4]. Machine learning assisted image segmentation can be achieved by implementing either supervised or unsupervised learning techniques. Supervised learning-based segmentation methods are categorized into pixel-wise and object-wise classification. Several researchers have implemented pixel-wise models on remotely sensed images [5, 6]. In terms of unsupervised machine learning methods, Shen et al. (2005) proposed a robust segmentation technique for magnetic resonance images of brain tissue with the help of a neural-network model [7].

Processing and analysis of high-resolution microscopic images has found ample applications in the oil and gas industry. Clelland et al. (1991) used scanning electron microscopy and X-ray to conduct automated petrographic analysis of sandstone samples. The method helped quantify mineralogy, pore and grain structure of the polished samples and proved to be a better alternative for the tedious manual petrographic methods [8]. Image processing techniques like filtering, segmentation and skeletonization were used to extract pore networks from microcomputed tomography images of rocks that are, in turn, used as input for network models to predict transport properties [9]. Budennyy et al. (2017) presented a method for petrographic thin section analysis combining statistical learning and image processing. The method makes use of watershed segmentation for rock classification and evaluation of properties of grains, cement, voids, and cleavage [10].

Many different segmentation methods have been developed to acquire nanoscale information from Scanning Electron Microscopy (SEM) images. A neural network was used by Tang and Spikes [1] to segment elemental SEM images of shales into five constituents, namely, quartz, calcite, feldspar, TOC and clay/pore. Wu and Misra (2019) and Misra et al. (2019) used random forest to process Hessian Matrix, Gaussian Filter, and Wavelet Transform based features extracted from the SEM

images of shale to identify four rock constituents [11,14]. A fast Random Forest algorithm, combined with other mathematical methods of image analysis, was used for automated and effective recognition of steel microstructures from optical microscopic images [12]. In 2018, Xiao et al. successfully demonstrated the implementation of a 3D fully residual convolutional network based on supervised deep learning to perform segmentation and reconstruction of mitochondrial structures from electron microscopy images [13].

1.2 Objectives of this Study

The objectives of this study are as follows:

- Locate and detect kerogen and hydrocarbon storage and transport pathways represented by pores/cracks in the scanning electron microscopy (SEM) maps from Wolfcamp and Barnett Shales by developing robust (highly generalizable) machine-learning-assisted image-segmentation method.
- Compare the robustness, reliability, and accuracy in identifying kerogen, pores, and cracks present in the SEM maps from Wolfcamp and Barnett shale formations, which differ in topology, morphology and distribution of these constituents. For example, Wolfcamp samples have predominantly inorganic pores and cracks, whereas Barnett samples have predominantly organic pores.

- A robust, highly generalizable model does not overfit the training data and exhibits reliable performance on new, unseen dataset. There are extremely limited prior studies on the generalization capability of machine-learning assisted identification of kerogen and pores in different types of organic-rich shales.
- For the outer regions of the pores, cracks, kerogen and other constituents, also referred as the transition zones, the pixel intensities transition from one constituent to another constituent. The proposed machine-learning assisted identification will generally have poor performance for transition zones. We investigate the robustness of the proposed ML-assisted characterization method for the transition zones.
- Identify the most informative yet independent features/attributes that improve the robustness of the proposed machine-learning assisted identification of constituents. For each pixel, features are computed by performing certain mathematical/statistical transformations on the intensities of the pixel of interest and the neighboring pixels.

2. Methodology

The micro-scale attributes of shale samples are best captured in a scanning electron microscopy (SEM) image. This paper investigates the robustness of machine learning assisted identification of pores, cracks and kerogen in the high-resolution SEM images of shale samples. The machine learning workflow aims to segment the SEM image into 4 constituents, namely pore/crack, organic/kerogen, matrix and pyrite. The segmentation workflow is illustrated in Figure 1.

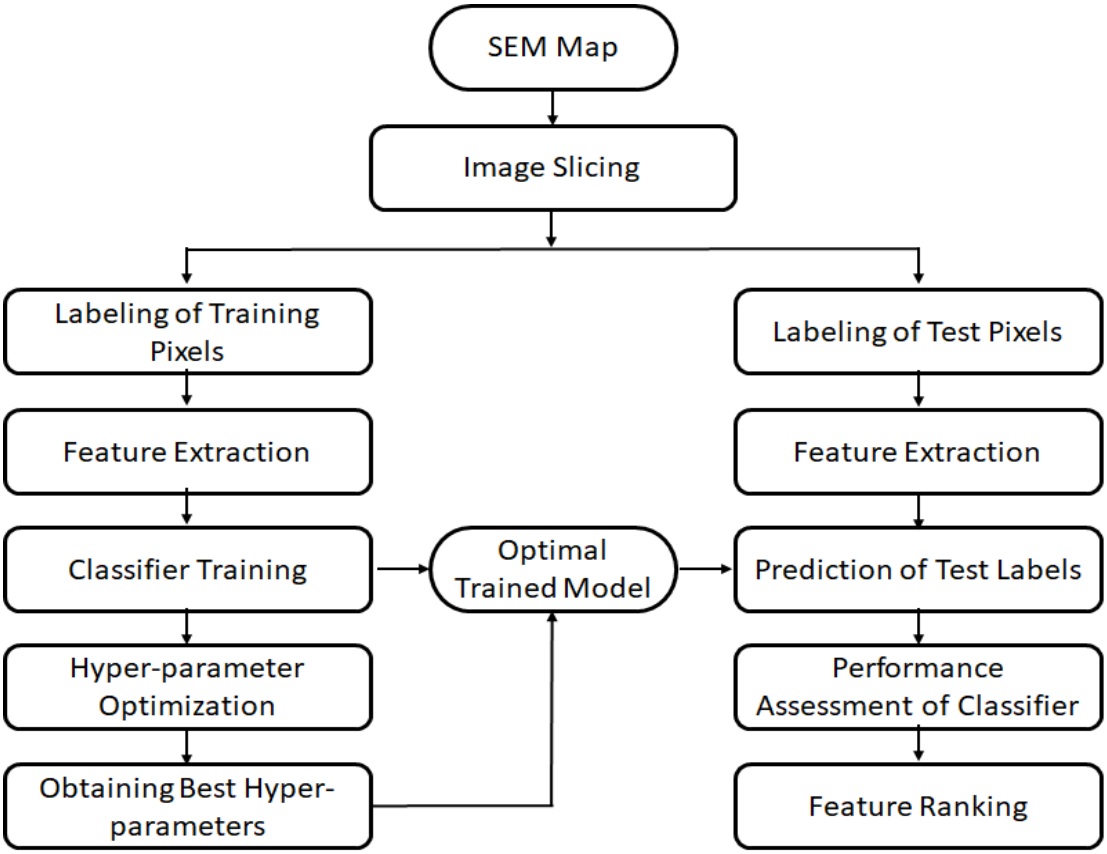

Fig. 1: Workflow for identifying kerogen, pores, and cracks in the SEM maps from Wolfcamp and Barnett shale formations using machine learning.

The study makes use of two SEM maps: Map-1 is from Wolfcamp Shale and Map-2 is from Barnett Shale, which are continuous sequence of multiple SEM images of organic-rich shale samples. The two SEM maps were captured using the FEI Helios Nanolab™ 650 DualBeam™ FIB/SEM machine and FEI SEM MAPS™ software in the Integrated Core Characterization (IC3) lab at the University of Oklahoma. We want to emphasize that the Map-1 contains 26060×205800 pixels representing an area of 260.6 μm by 2058 μm and the Map-2 contains 16464×18242 pixels representing an area of 164.64 μm by 182.42 μm. Map-1 and Map-2 together are 20 GB in size. Map-1 contains 1000 SEM images and Map-2 contains 54 SEM images, such that each image consists of 2058×2606 pixels that represent an area of 20.58 μm by 26.06 μm. Out of this large dataset of approximately 10 billion pixels, we select around 50 thousand pixels to train the machine learning workflow.

2.1 Feature extraction

Performance of a data-driven model improves when informative and independent features are used to build the model (Wu and Misra, 2019). For high-dimensional dataset, feature extraction facilitates the learning and improves the robustness of data-driven models. Pixel intensity or brightness is the primary feature. Additionally, mathematical/statistical transformations are performed on the pixel intensities to generate 15 additional features belonging to seven categories. These 15 features

define each pixel in relationship to its neighboring pixels (similar to Wu et al., 2019) [15]. The features used in our study are explained in detail by Misra et al. (2019). In addition to pixel intensity, we extract the following 15 features belonging to seven categories:

- Gaussian blur (1 feature)
- Difference of Gaussian (1 feature)
- Two levels of Wavelet decompositions (6 features)
- Hessian affine region detector (3 features)
- Sobel edge detector (1 feature)
- Local statistical information (3 features: minimum, maximum, mean)

Gaussian blur is a low-pass filter that attenuates the high frequency signals and reduces the noise level and unnecessary details in the image that results in a blurred version of the image. The Difference of Gaussian (DoG) is a band-pass filter that delineates local structures (e.g. blobs) in image. The local information for each pixel is calculated as the maximum, minimum and mean intensities of all pixels inside a 3-pixel by 3-pixel region centered at the target pixel to be identified. Delineation of edges improves the segmentation of transition zones. Sobel operator is a discrete differentiation operator that measures the gradient of the image intensity function using a 3×3 kernel to emphasize the odd edges. Hessian affine region detector

describes the second-order partial derivatives of the local intensity around a pixel. Hessian detector is suited for detecting even edges, such as corners and stars. It can differentiate between tubular, sheet-like and blob-like structures present in the image. Multilevel wavelet decomposition aids in multi-resolution scale-space analysis of signals. Unlike Fourier transform, wavelet transform preserves local features and are suitable for non-stationary signals. More details about the feature extraction are provided in the work done by Wu and Misra (2019) and Misra et al. (2019) [14,19].

2.2 Machine Learning Model

To identify the pores/cracks and kerogen, we use supervised learning to train classifiers. In this study, Random Forest is the best performing (most generalizable) classifier in terms of learning to identify the pores, cracks, and kerogen along with two other constituents (i.e. pyrite and matrix) present in 8-bit grayscale SEM images of shale samples. To achieve this, Random Forest classifier relates the pixel intensity and 15 additional features corresponding to a specific pixel in SEM image to a specific constituent type. Random Forest is an ensemble technique that combines multiple decision trees in parallel into a single predictive model to achieve low bias and low variance. To that end, Random Forest uses bootstrapping followed by

aggregation, also referred to as bagging method. The final prediction in a Random Forest is made by averaging the predictions of all the decision trees in the ensemble.

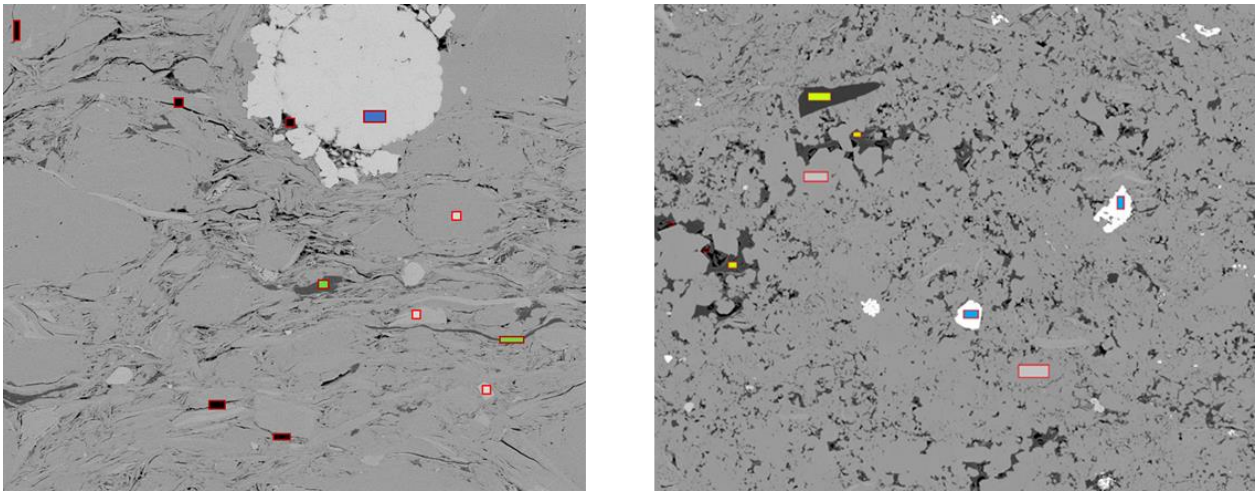

Fig. 2: Image Slice 90 of Map-1 from Wolfcamp shale showing the regions selected as training pixels, which constitute less than 0.7% of the total pixels in the image slice (left) and Image Slice 35 of Map-2 from Barnett shale showing the regions of training pixels, which constitute less than 0.5% of the total pixels in the image slice. Pixels corresponding to pore/crack, organic/kerogen, matrix and pyrite constituents are represented by black, green (light grey in the print version), grey and blue rectangles (dark grey in the print version), respectively.

2.3 Training Pixels and Training Dataset

Out of the large 20-GB SEM Maps dataset made of approximately 10 billion pixels, we select around 50 thousand pixels to train the machine learning workflow. The

proposed supervised learning task requires 15 features, pixel intensity, and the label representing the constituent type of the pixel. Machine learning model learns to identify the constituent of interest by learning to relate the pixel intensity and the 15 features extracted for a training pixel with the pre-assigned label, indicating the constituent type, i.e. pore/crack, organic/kerogen, pyrite, and matrix. Assigning labels to a pixel is a time-consuming step. Consequently, we assign labels to around 50 thousand pixels.

Training pixels should be carefully selected such that the pixels corresponding to each constituent type are representative of the most likely statistical distribution. Training pixels should not add adverse bias and noise. Number of training pixels for each constituent type should be relatively the same to avoid issues due to class imbalance. In our study, the number of pixels of the 4 constituents present in the SEM maps vary significantly. We choose the support pixels for the training dataset such that the volume fraction of each label in the original map is preserved in the training dataset. We use slice 90 of Map-1 and slice 35 of Map-2 for the purpose of selecting training pixels. For the Map-1, 705, 2074, 17373 and 15000 pixels belonging to the pore/crack, organic/kerogen, matrix and pyrite constituents, respectively (Figure 2: left), are selected as training pixels. For the Map-2, we selected 912, 8435, 5806 and 5387 pixels belonging to the pore/crack,

organic/kerogen, matrix, and pyrite constituents, respectively, as the training pixels (Figure 2: right). 15 features were then extracted for each pixel with respect to the neighboring pixels. These 15 features along with the pixel intensity and the available labels (information about the constituent type corresponding to a pixel) are used to train the Random forest classifier.

2.4 Pixels and Dataset for the Testing and Deployment

The robustness and generalization capability of the Random Forest classifier is tested on both the inner-region and outer-region pixels of Map-1 and Map-2. We select image slices 13, 649 and 860 from Map-1 and slices 15, 26 and 28 from Map-2 to build the inner-region test set. The outer region test set was formed by selecting pixels from slices 13, 203, 334 and 500 from the Map-1 and slices 13, 17 and 26 from the Map-2.

2.5 Evaluation Metric

As the random forest learns to identity the constituent type, the performance of the data-driven model needs to be evaluated using suitable metrics. We implement F1 score, precision, recall, and confusion matrix as the evaluation metrics for this study. F1 is the harmonic mean of precision and recall. F1 score, precision, and recall range from 0 to 1, such that 0 indicates poor identification. Precision of the classifier specific to a constituent type is a measure of the reliability of the constituent type

assigned by the classifier. Recall of the classifier specific to a constituent type is a measure of the classifier's ability to correctly assign the constituent type. In this study, we have an unequal distribution of the different classes/constituents, resulting in an imbalanced test dataset. Therefore, we use the weighted F1 score. Confusion matrix lists the total numbers of true positives, false positives, true negatives and false negatives for the classifier.

2.6 Hyper-parameter Optimization

Hyper-parameter optimization refers to the adjustment of the hyper-parameters of a machine learning model to optimize its performance. While parameters are learned by a model during training, hyper-parameters are set by the user prior to training. A reliable set of hyper-parameters ensures low memorization error (model performance on training set) and the lowest generalization error (model performance on testing set). We perform hyper-parameter optimization using a grid search method. Hyperparameter govern the learning process and the parameters learnt by the model as the model processes the available training data. Weighted F1 score is chosen as the basis of comparison to decide the best combination of hyper-parameters. A total of 1512 combinations of hyper-parameters were tried; followed by, a 3-fold stratified cross validation. Out of the 4536 models developed based on the different set of hyperparameters, only one model that generates low

memorization error on the training dataset and the lowest generalization error on the testing dataset was chosen as the optimum model.

3. Results and Discussion

In this section, we compare the performance of the machine-learning-assisted workflow on the SEM maps from Wolfcamp shale against that on Barnett shale. The SEM maps from the two shale formations exhibit significant difference in topology and distribution of constituents. The two SEM maps were also acquired with slightly different measurement settings; consequently, the two maps exhibit slightly different ranges of pixel intensity. To test robustness and high generalizability in identifying pores, cracks and kerogen, we develop the following three models:

- Model 1: Learning from Wolfcamp for deployment on Barnett
- Model 2: Learning from Barnett for deployment on Wolfcamp
- Model 3: Learning from both Wolfcamp and Barnet for deployment on both Wolfcamp and Barnet

In the following sections of the paper, the Random Forest classifier trained on Map-1 from Wolfcamp shale is referred to as Model-1, whereas that trained on Map-2 from Barnett shale is referred to as Model-2. Model-3 is trained on both of the maps. We present a comparative study of the performances of the three models on the two

maps and investigate the robustness of the proposed workflow in terms of Precision, Recall, F1 score, and confusion matrix for the 3 above-mentioned models.

3.1. Model 1: Learning from Wolfcamp shale for deployment on Barnett shale

Model-1 was trained on pixels from Slice 90 of Map-1 from the Wolfcamp formation and then tested on both the inner and outer region pixels from several slices of Map-1. Model-1 achieved average F1 scores of 1.00 and 0.89 on the inner region pixels and the outer region pixels, respectively. A lower F1 score for outer region pixels is primarily due to low precision and recall of matrix and pyrites constituents. A low precision of a certain constituent indicates that one or more of the other constituents are being identified as that particular constituent whereas a low recall indicates misclassification of a certain constituent as a different one. In case of the outer region pixels, we also observe a low precision but high recall for the pore/crack and organic/kerogen constituents. This leads to the explanation that in the boundary region of two constituents, the pyrites are being identified as matrix and in turn, matrix pixels are being labelled as pore/cracks and organic/kerogen constituents. This observation is justified because in the Map-1, the matrix shares considerable boundary area with all other constituents, including pore/cracks.

After the training and testing, in order to evaluate the robustness, Model-1 is deployed on pixels from the inner-region and outer-region of Map-2. Map-1 and Map-2 significantly vary in the pore/crack constituents. While Map-1 is characterized by thin strips of black pixels present as cracks (Figure 2: left), Map-2 is dominated by cluster of pixels denoting pores embedded in organic matter (Figure 2: right). We compare the performances of Model-1 on the inner region of the two maps in Figure 3 (left) and those on the outer region pixels in Figure 3 (right). For the inner region pixels, we observe a low precision of 0.41 for the pore/crack constituent coupled with a low recall of 0.49 for the organic/kerogen constituent. This denotes that pixels belonging to kerogen constituent are being labelled as pore/cracks by Model-1. This is also evident from the confusion matrix (Figure 4). In Figure 4, for the matrix and pyrite constituents, the number of support pixels are equal to the number of diagonal elements; thereby, proving that they have been correctly identified. But a significant number of support pixels (1615 out of 5263 pixels) in kerogen constituent has been identified as pores or cracks, resulting in a low value of the F1 score for these two constituents. The model performs well for the matrix and pyrite constituents of inner region of both the maps (Figure 3: left). This is because the variation in pixel intensities between these constituents and pore/cracks and organic/kerogen are more detectable than that between pore/cracks and organic/kerogen.

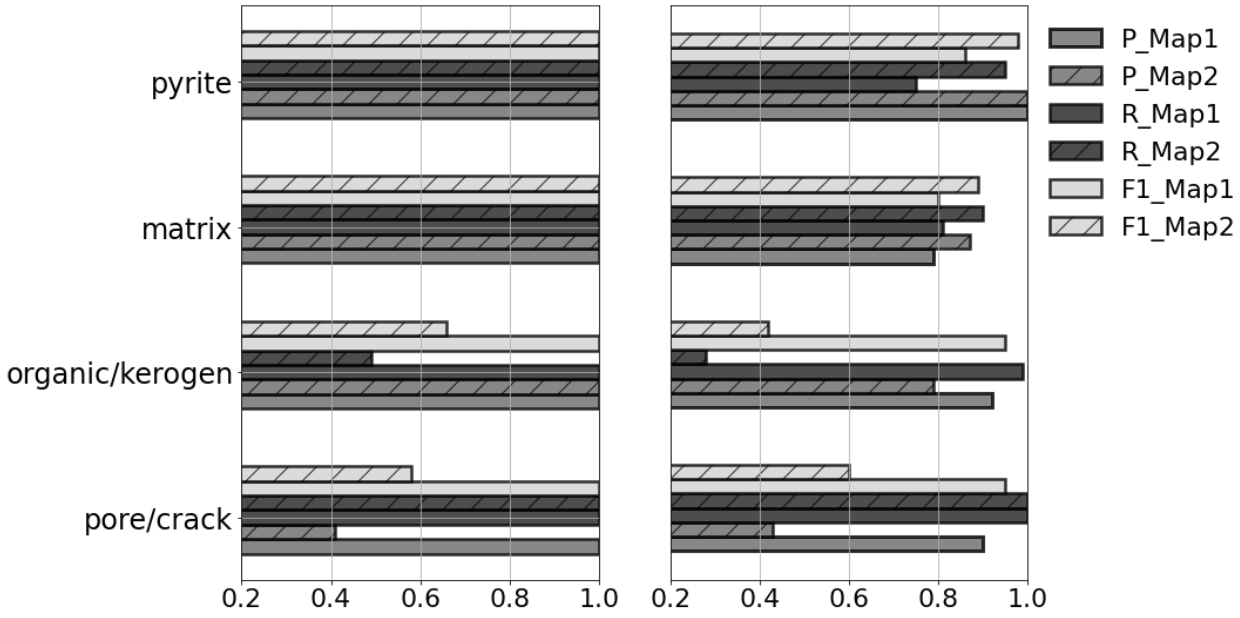

Fig. 3: Comparison of segmentation performance (P: precision; R: recall; F1: F1 Score) of Model-1 on inner-region test pixels from Wolfcamp against those on inner-region test/deployment pixels from Barnett (left) and comparison of segmentation performance of Model-1 on outer-region test pixels from Wolfcamp against those on outer-region test/deployment pixels from Barnett (right). Model-1 was trained on training pixels from Wolfcamp. Model-1 generalizes well for the inner and outer regions of matrix and pyrite constituents.

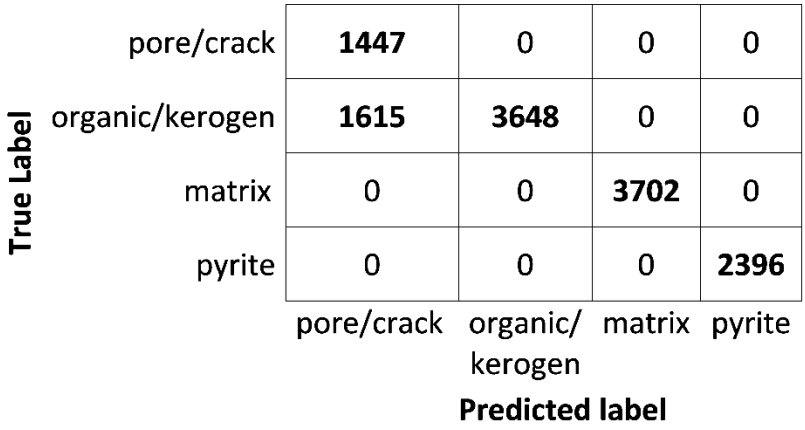

Fig. 4: Confusion matrix related to the segmentation performance of Model-1 trained on Wolfcamp when applied on the inner-region pixels from Barnett.

The F1 scores are 0.89 and 0.81 for the outer-region pixels from Map-1 and Map-2, respectively. Compared to that, the F1 scores are 1.00 and 0.82 for the inner-region pixels from Map-1 and Map-2, respectively. The greyscale intensity ranges of the pore/crack and organic/kerogen constituents shows a greater overlap in Map-2 than Map-1 leading to a lower recall and precision values for these constituents (Figure 3: right). This proves that Model-1 is not suitable for accurate segmentation of pore/crack and organic/kerogen constituents of Barnett shale. In case of Map-1 (Figure 3: right), a lower recall for pyrites coupled with a lower precision for matrix tells us that the model classifies pyrites as matrix at the boundary of these two constituents. Interestingly, Model-1 exhibits better performance on the matrix and pyrites constituents in Barnett shale as compared to that in Wolfcamp shale. A

sharper intensity contrast at the interface of these two constituents in Map-2 may be responsible for an improved segmentation.

3.2. Model 2: Learning from Barnett Shale for deployment on Wolfcamp Shale

In the previous section, we trained and tested Model-1 on Map-1 from the Wolfcamp formation and deployed on Map-2 from the Barnett formation. In this section, we will investigate the robustness of the Model-2 when it is trained and tested on Map-2 from Barnett shale formation and deployed on Map-1 from Wolfcamp shale formation. First, Model-2 was validated on the inner as well as outer regions of Map-2. Model-2 showed an efficient segmentation performance with an average F1 score of 1.00 for inner region and 0.98 for outer region validation pixels. In both cases, pyrite constituents were most accurately identified.

Figure 5 plots the confusion matrices for the performance of Model-1 on inner region pixels of Map-1 and that of Model-2 on inner region pixels of Map-2. The diagonal values denote the number of pixels that have been correctly identified by the two models. We observe that no more than 11 pixels in Map-1 (shown by the off-diagonal values) and 18 pixels in Map-2 has been incorrectly labeled. Despite using

limited training pixels, the performance of the two models prove the robustness of machine learning-assisted segmentation workflow.

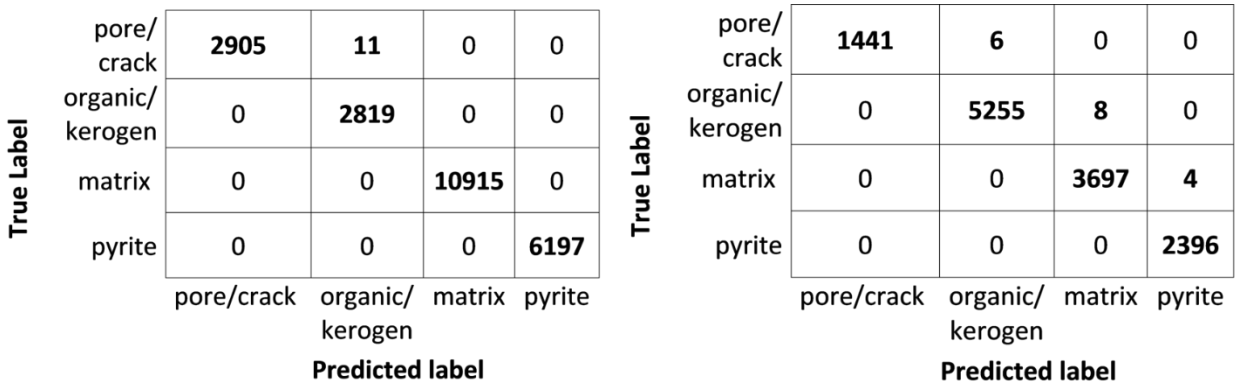

Fig. 5: Confusion matrices related to the segmentation performance on the inner-region test pixels. Model-1's segmentation performance on Wolfcamp is in Left and Model-2's segmentation performance on Barnett is in Right.

Next, Model-2 trained on Map-2 is deployed on the inner and outer region pixels of Map-1 (Figure 6). Model-2 achieves an F1 score of 0.89 when deployed on the inner region pixels and 0.91 when deployed on the outer region pixels of Map-1; this is very encouraging performance. These scores are found to be higher than the F1 scores obtained when Model-1 was deployed on Map-2 inner region (F1 score of 0.82) and outer region (F1 score of 0.81) pixels. This suggests that Map-2 has more generalizable statistical features that can serve to better train the segmentation model and Map-1 has simpler microstructural features that are easier to identify. Figure 6

shows the generalization capability of Model-2 when deployed on a different map with respect to its performance when tested on the map on which it is trained.

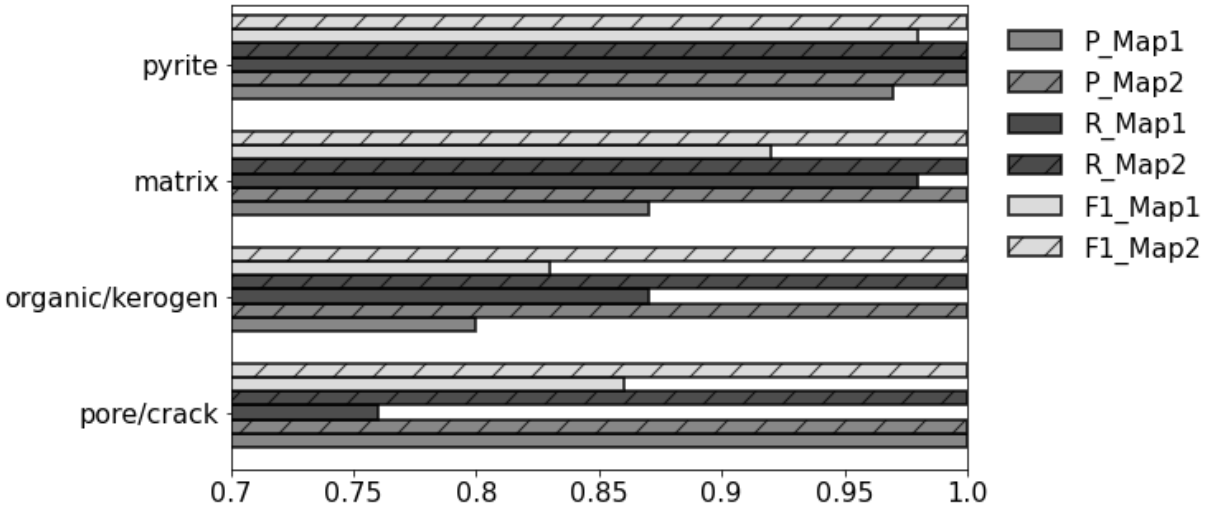

Fig. 6: Comparison of segmentation performance (P: precision; R: recall; F1: F1 Score) of Model-2 on inner-region test/deployment pixels of Wolfcamp against that on inner-region test pixels from Barnett. Model-2 was trained on training pixels from Barnett.

Model-2 learns from assorted clusters of pixels labeled as pores and therefore, fails to accurately identify cracks that are present in Map-1 as elongated strips of pixels. This observation can be verified from Figure 7 which shows the segmentation of an image slice of Map-1 using Model-1 (left) and Model-2 (right). We see pixels belonging to cracks being identified as organic/kerogen in the right image.

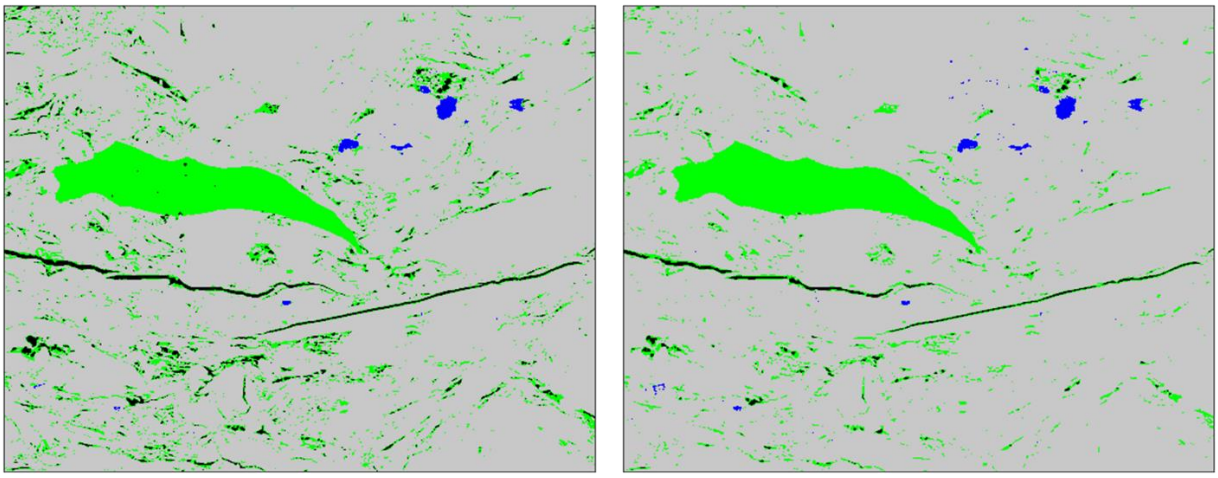

Fig. 7: Images showing the difference in segmentation of Slice 334 from Wolfcamp shale using Model-1 (left) and Model-2 (right). The difference is most prominent in pore/crack and organic/kerogen components.

3.3. Model 3: Learning from both Wolfcamp and Barnett shales for deployment on Wolfcamp and Barnett shales

The results obtained using Model 1 and Model 2 show the machine learning workflow performs reliably and accurately when the model is trained and tested on the same map. The performance deteriorates for certain constituents when deployed on a SEM map from a different formation. In this section, Model-3 is trained on a combined dataset with training pixels from both Map-1 and Map-2. For evaluating the robustness of the Model-3, we deployed the Model-3 on the inner-region pixels and then on the outer region pixels of Map-1 and Map-2, respectively. In Figure 8, the performance of Model-3 does not fall below 0.95 in terms of F1 score for the

inner region pixels of the two maps. The average F1 scores are 1.00 and 0.99 for Map-1 and Map-2, respectively. For the outer region pixels, the model exhibits incorrect classification (Precision/Recall of less than 0.8) for pore/cracks and pyrites constituents (Figure 8). Model-3 performs the best in terms of average F1 score for all the constituents, namely Map-1 inner region, Map-1 outer region, Map-2 inner region and Map-2 outer region (Figure 8). For the outer-region pixels, F1 score is larger than 0.8. For the kerogen, pores and cracks, the machine learning workflow performs better on Barnett as compared to Wolfcamp.

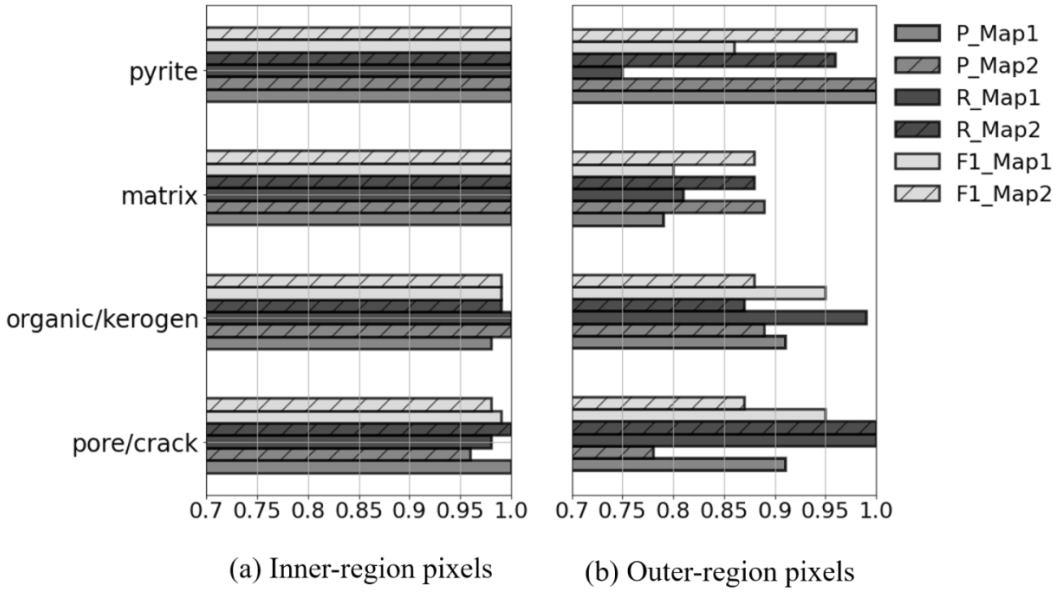

Fig. 8: Comparison of segmentation performance (P: precision; R: recall; F1: F1 Score) of Model-3 on inner-region test/deployment pixels from Barnett and Wolfcamp (left) against that on outer-region test/deployment pixels from Barnett and Wolfcamp (right). Model-3 was trained on training pixels from Slice 90 from Wolfcamp shale and Slice 35 from Barnett shale.

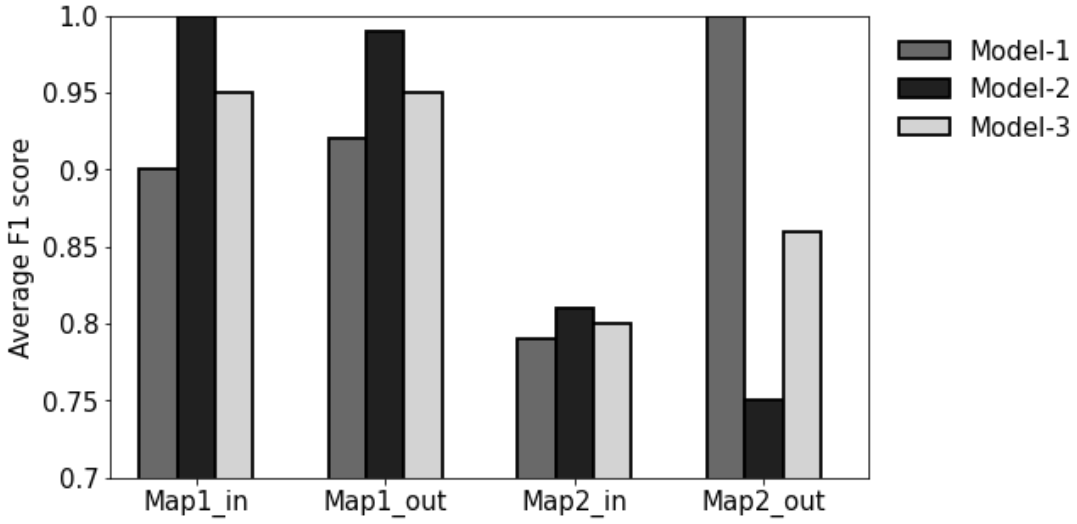

Fig. 9: Comparison of segmentation performances (in terms of average F1 score) of the three models on the inner (subscript in) and outer (subscript out) regions pixels of the two maps.

It is evident from Figure 9, Model-3 is better than Model-1 for inner-region and outer-region pixels of Map-1 and significantly better than Model-2 for the outer region pixels of Map-2. Overall, among the three models, Model-3 performs consistently well in all the four cases shown in Figure 9. Therefore, it can be said that, the model trained on the combined dataset is more robust than a model trained on one map and deployed on a map from a different formation. This suggests that training the workflow with pixels from different maps results in a more robust model, capable of segmenting SEM maps from different formations. Figure 10 aids

a visual comparison of segmented images obtained using the three previously discussed models. A large fraction of organic constituent in Slice 55 of Map-2 is incorrectly identified as pores by Model-1 (top right). On the other hand, Model-2 (bottom left) underpredicts the total pore space as it fails to identify the pores present in the form of cracks. Model-3 generates the best generalized segmentation performance (bottom right) among the three approaches.

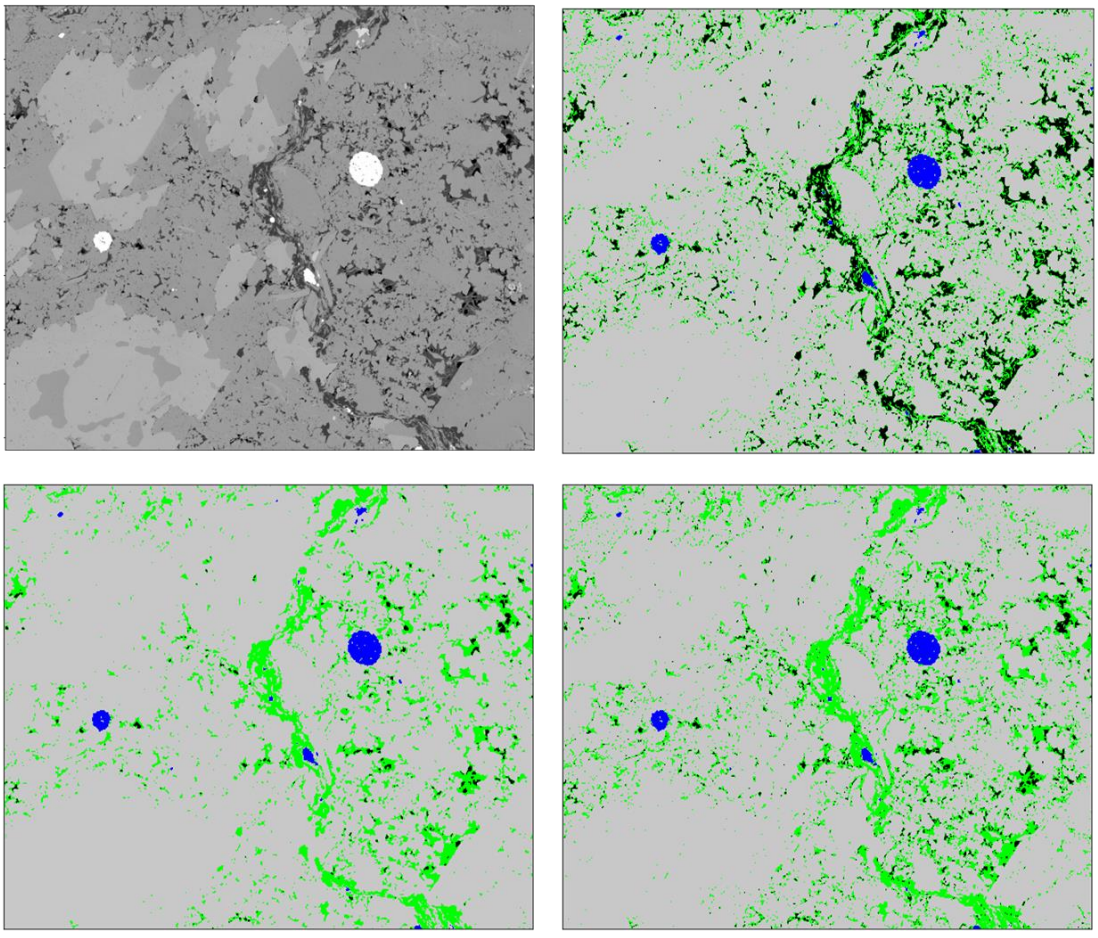

Fig. 10: Slice 55 from the Barnett shale (top left), after segmentation using Model-1 trained only on Wolfcamp (top right), Model-2 trained only on Barnett (bottom left) and Model-3 trained on both Wolfcamp and Barnett (bottom right).

4. Estimation of Petrophysical Parameters Based on the Identification of Pores, Cracks and Kerogen

Above sections, describe the development and evaluation of a robust method to identify hydrocarbon storage and transport pathways along with the kerogen distribution in the SEM maps from shale formations. After the identification of pores, kerogen, and cracks, one can quantify various petrophysical and petrological characteristics. On those lines, we quantify porosity distribution, pore size distribution, and kerogen size distribution in the Map-2 from the Barnett shale formation. Map-2 covers an area of 164.64 μm by 182.42 μm containing 16464×18242 pixels. Map-2 was sliced into 56 image slices, respectively, with no overlapping pixels. Image slices are numbered 1 to 56 for Map-2, each consisting 2058×2606 pixels and covering an area of 20.58 μm by 26.06 μm. The total porosity from the Barnett map was found to be 0.45%. Figure 11 shows a grayscale heatmap with cells representing the 56 (8 rows by 7 columns) slices of Map-2. The plot is annotated with the porosity values associated with each slice of the Barnett Map. The porosity ranges from 0.1 % to 0.8% in the entire sample representing an area of 164.64 μm by 182.42 μm. The right side of the map is dominated by low-porosity regions. Both high porosity and low porosity regions are localized and predominantly surrounded by regions having 0.4% porosity. Extremely low and high

porosity regions are isolated, whereas 0.5% porosity regions are the most well connected.

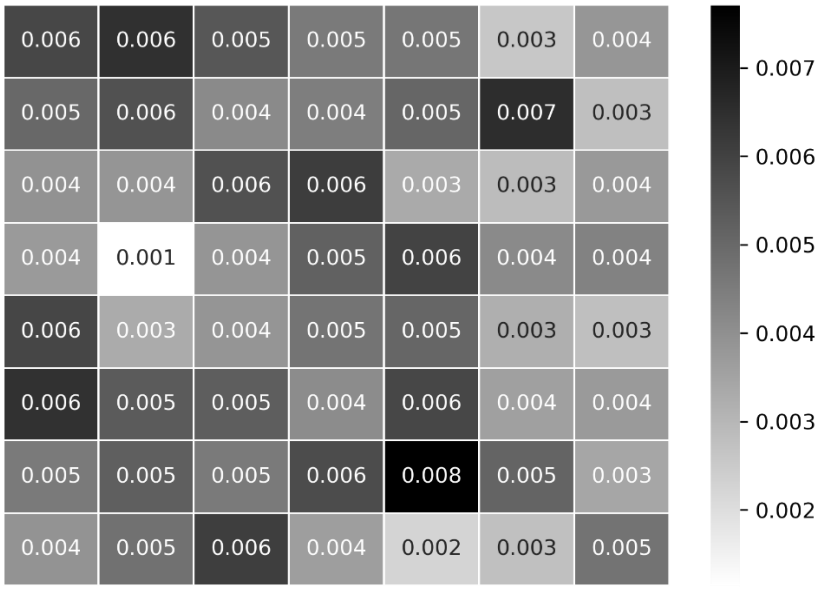

Fig. 11: Slice-wise porosity distribution in the Barnett shale sample (Map-2) representing variations over an area of 164.64 µm by 182.42 µm.

The pore-size and kerogen-size distributions are quantified with the help of a cluster function that calculates the number of pixels present in each cluster of a specified component. A multiscale pore size distribution was observed with 17% of the pore volume consisting of pore sizes less than 20 pixels, 68% of the pore volume made of pixel size between 20 and 500 pixels and the remaining 15% pore volume consisting of pore sizes greater than 200 pixels. The micro-scale (<20 pixels) and macro-scale (>500 pixels) pore sizes contribute equally to the total pore volume of the Barnett sample. The macro-scale porosity of the system is primarily made of the

organic pores. Figure 12 shows the Barnett map (a) and the distribution of the three different pore size scales (b: micro-scale pore size < 20 pixels, c: meso-scale pore size between 20 and 500 pixels, d: macro-scale pore size > 500 pixels). Number of macroscale pores is two order of magnitude lower than the number of microscale pores, while the number of mesoscale pores is half of the number of microscale pores. Figure 13 plots the distribution of the kerogen cluster sizes present in Map-2. Only the large clusters consisting of more than 10,000 pixels are considered for this plot. We observe presence of very large kerogen clusters with sizes more than 100,000 pixels. The map predominantly contains kerogen clusters of size ranging from 10,000 pixels to 20,000 pixels. In the cluster-size range from 100,000 to 200,000 pixels, there are on an average 1 to 2 clusters in the map. These clusters host the organic pores and are important in the calculation of TOC of the shale sample.

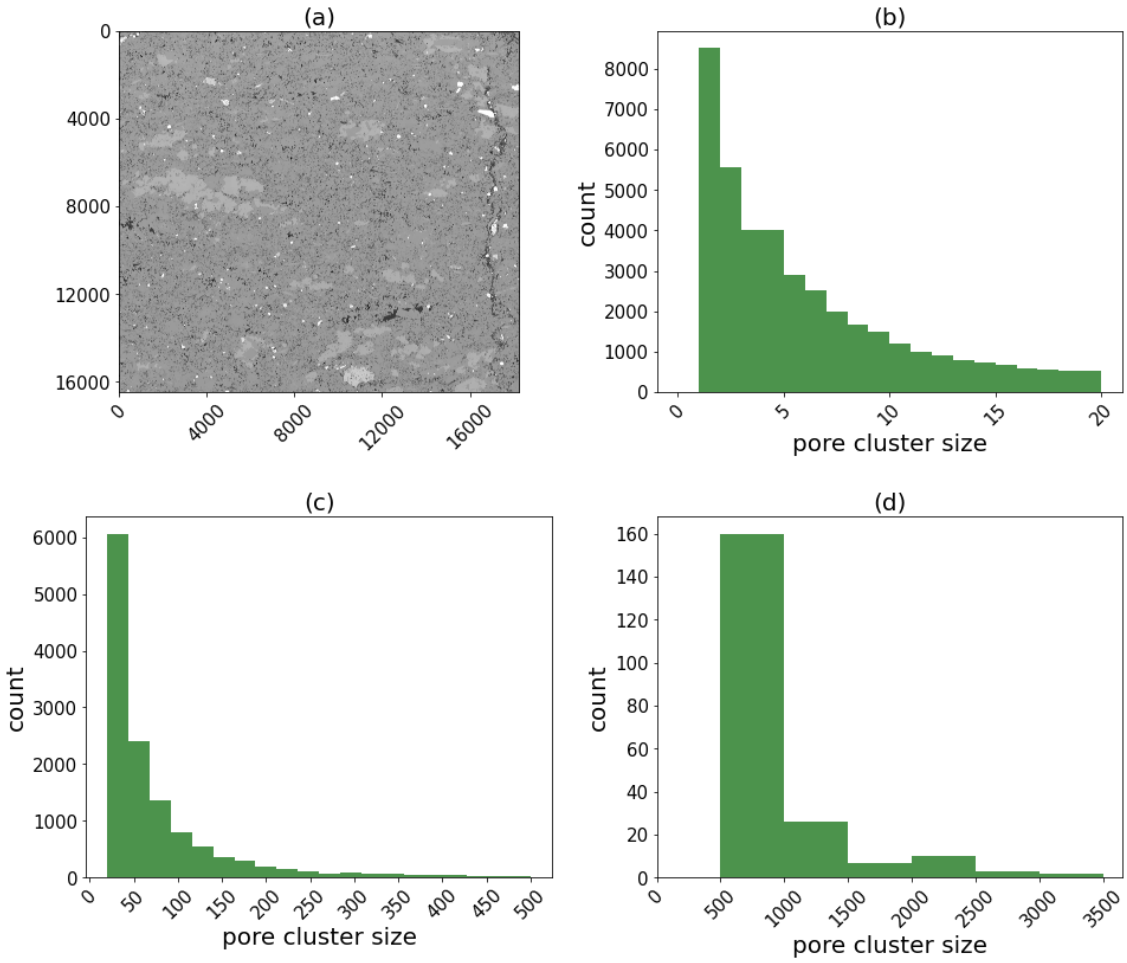

Fig. 12: Multiscale pore size distribution of the Barnett shale sample (Map-2).

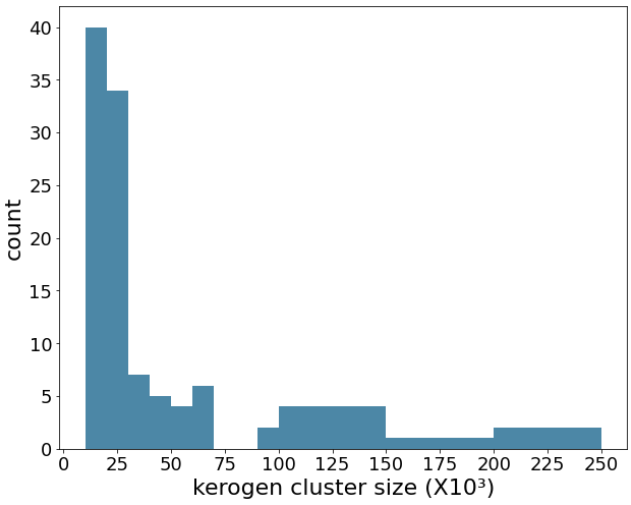

Fig. 13: Organic/kerogen size distribution of Barnett shale sample (Map-2).

5. Further Analysis of the Robustness of the ML-Assisted Identification of Kerogen, Cracks and Pores

5.1. Effect of feature scaling

Feature scaling is an essential step for learning algorithms based on density and distance; for example, K-nearest neighbors and K-means. However, Random forest is a tree-based method that does not require feature scaling. We trained and tested our model on both scaled and unscaled features. Min-Max Scaler was used to perform feature scaling. Min-max scaler is suitable when features do not have outliers and are specially suited when the features have a specific minimum and maximum value, similar to pixel intensity of SEM image that ranges between 0 to 255. However, min-max scaler compresses the feature distribution to range between 0 and 1 leading to a sparse representation that adversely affects feature values close to the minimum value. In Figure 14, feature scaling (the image on right) considerably deteriorates the segmentation of pore/crack and organic/kerogen. Most pixels belonging to pore/crack constituent are not detected and the pixels belonging to organic/kerogen constituent are labeled as matrix. Pixel intensities of matrix and organic matter and those of pores and organic matter exhibit overlap over certain range of values; consequently, min-max feature scaling makes it harder to differentiate between these constituents.

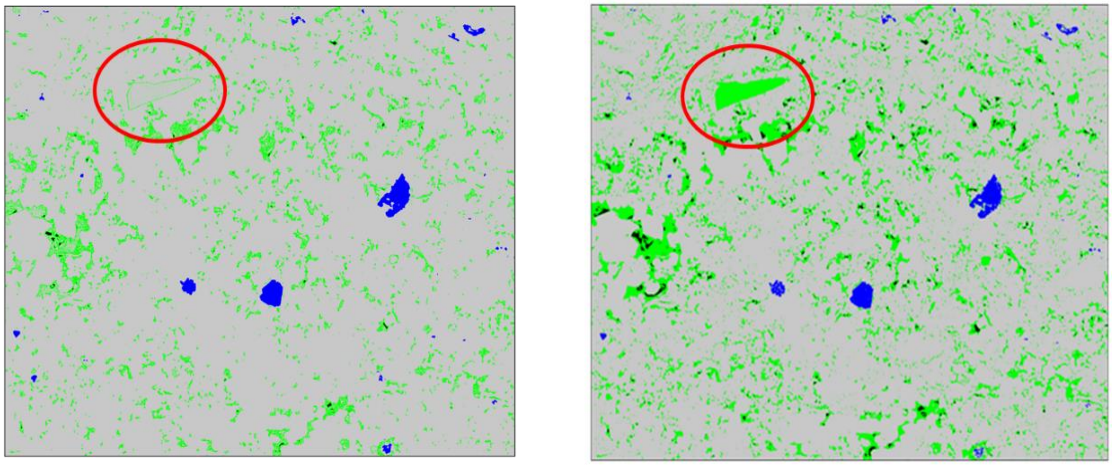

Fig 14: Effect of feature scaling on the segmentation of Slice 35 of Map-2. Images on left and right are segmented using Random Forest classifiers trained on unscaled features and features scaled using MinMax Scaler, respectively.

Table 1: Comparison of segmentation performance obtained by training the Random Forest on scaled features (I) and unscaled features (II) corresponding to the inner-region pixels of Map-2 and then testing the segmentation performance on the inner-region pixels from various slices of Map-2.

Constituent	**Precision**		**Recall**		**F1 score**		**Support**
Cases	**I**	**II**	**I**	**II**	**I**	**II**	**Total:12807**
Pore/crack	1.00	1.00	0.20	0.99	0.34	0.99	1447
Organic/kerogen	0.71	1.00	0.15	1.00	0.25	1.00	5263
Matrix	0.36	1.00	0.99	1.00	0.52	1.00	3701
Pyrite	0.95	1.00	0.46	1.00	0.62	1.00	2396

As shown in Table 1, a near perfect score is obtained for model trained using unscaled features as compared to that trained using scaled features. Extremely low precision for matrix indicates that certain constituents are being wrongly identified as the matrix when using scaled features. Extremely low recall of pore/crack and organic/kerogen constituents indicates that these constituents are very hard to detect when using scaled features. These observations align with Figure 14.

5.2. Effect of histogram equalization

Table 2 portrays the effect of pixel-intensity histogram equalization between the two maps. A histogram matching is carried out when images are believed to be acquired under different conditions. In our study, the two SEM maps were found to differ significantly from each other in certain regions in terms of exposure. Therefore, histogram equalization was carried out for certain slices as a pre-processing step. In this process, one image is considered as the reference or template and other images are modified such that the range of the pixel intensities are same as that of the reference image. We performed histogram equalization by considering slice 90 from Map-1, which has been used to create the training set of Model 1, as the template for matching the intensity histograms of test slices from Map-2. However, the process did not yield a better segmentation performance. The results show a significant reduction in the F1 scores for all the constituents due to the histogram equalization

as compared to using unscaled original images (Table 2). Detection of matrix is not affected by the equalization. Due to the equalization, the pixels that were detected as pore/crack are now identified as matrix or pyrite. The result is most evident for the matrix and pyrite constituents. A major portion of the matrix constituent is being identified as pyrite (Fig 15: bottom right). This is because the range of values of pyrite pixels in Map-2 is higher than that in Map-1. This range is narrowed as a result of histogram matching (Fig 15: top right) which consequently causes some of the pixels of the pyrite constituents of Map-2 to fall under the range of pixel intensities of matrix in Map-1 that leads to wrong identification of matrix and pyrite.

Table 2: Comparison of segmentation performance obtained by training the Random Forest on unscaled pixels from Map-1 and then testing the segmentation performance on the original unscaled inner-region pixels (I) and on the histogram-equalized inner-region pixels (II) from Map-2.

Constituent	**Precision**		**Recall**		**F1 score**		**Support**
Cases	**I**	**II**	**I**	**II**	**I**	**II**	**Total: 12807**
Pore/crack	0.44	0.76	1.00	0.46	0.61	0.58	1447
Organic/kerogen	1.00	0.96	0.65	0.46	0.79	0.62	5263
Matrix	1.00	0.45	1.00	1.00	1.00	0.62	3701
Pyrite	1.00	0.81	1.00	0.40	1.00	0.54	2396

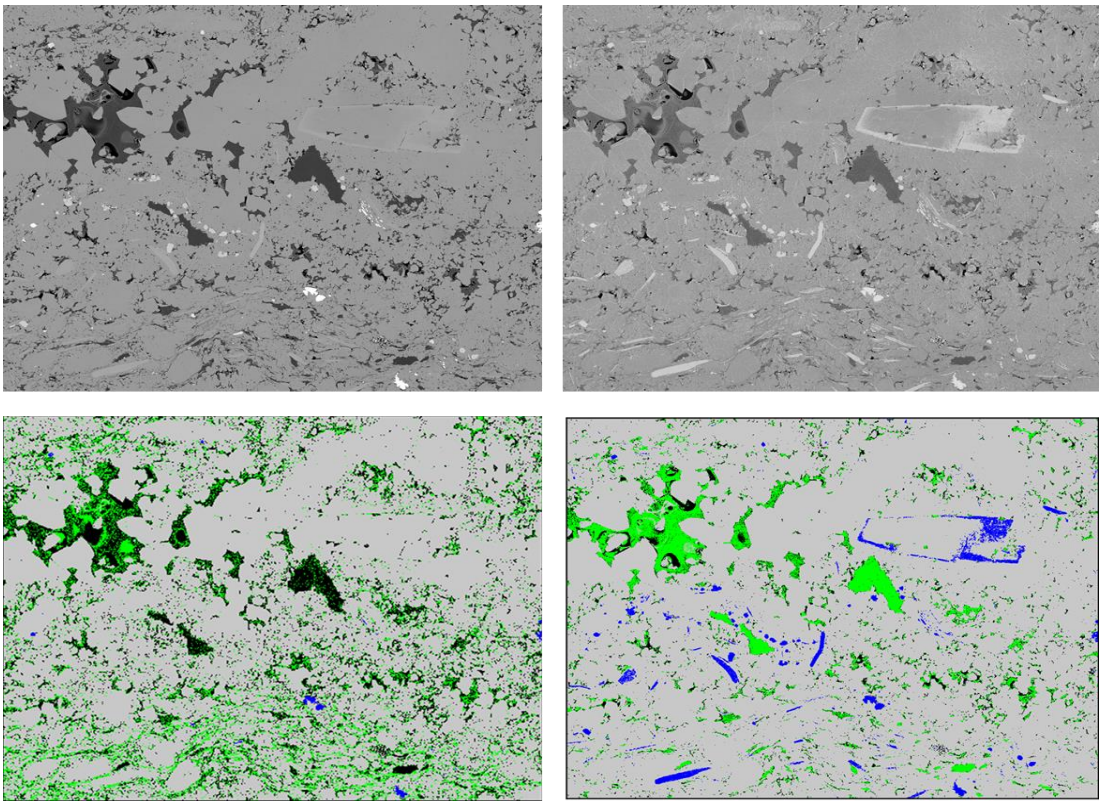

Fig 15: Effect of histogram equalization on Slice 28 of Map-2. Original image (top left), image after matched with histogram of Slice 90 of Map-1 (top right), original image segmented by Model-1 (bottom left), and matched image segmented by Model-1 (bottom right).

5.3. Effect of the size of testing data

There is always a trade-off associated with the number of support (number of samples) to be chosen to test a machine learning model. Selecting a large number of pixels for testing the ML-assisted segmentation method ensures its robustness because the increase in number of pixels ensures similarity between the probability

distribution of features in both training and testing datasets. When selecting the pixels to form the testing dataset, some constituents such as pores/cracks that have a lower volume fraction in the image, are hard to select resulting in a limited number of pixels for that constituent. Imbalanced testing dataset, i.e. when the number of samples per category (in this case, constituent type), penalizes the constituent having an order of magnitude lower number of samples as compared to other constituents. We present a comparison of the performance of Model 2 when tested on small vs. larger number of support pixels (Table 3).

In our image, we have an abundance of pixels belonging to organic/kerogen constituent and limited number of pixels for pore/crack. As we increase the support of pore/crack constituent from 1447 (Table 3: Case I) to 4656 pixels (Table 3: Case II), we observe an decrease in the recall of the pore/crack constituent from 1.00 to 0.86. Recall measures the number of times a constituent is being correctly identified by the model. A drop in the recall suggests that Case I support pixels were selected from easily identifiable regions of pore/crack constituent whereas the newly introduced support pixels were selected from regions where the model failed to correctly identify the pore/crack constituent.

Table 3: Comparison of segmentation performance of the proposed ML-assisted segmentation method trained on pixels of Map-2 when tested on limited number of support pixels (Case I) against that when tested on large number of support pixels (Case II) from slices 15, 26 and 28 of Map-2.

Constituent	**Precision**		**Recall**		**F1 score**		**Support**	
Cases	**I**	**II**	**I**	**II**	**I**	**II**	**I**	**II**
Pore/crack	0.44	1.00	1.00	0.88	0.61	0.93	1447	4656
Organic/kerogen	1.00	0.97	0.65	1.00	0.79	0.98	5263	23584
Matrix	1.00	1.00	1.00	1.00	1.00	1.00	3701	13835
Pyrite	1.00	1.00	1.00	1.00	1.00	1.00	2396	5154
							Tot:12807	**Tot:47229**

In the case of organic/kerogen constituent, we increase the support from 5263 (Table 3: Case I) to 23584 (Table 3: Case II) pixels. This causes the recall of this constituent to increase from 0.65 to 1.00 coupled with an increase in the precision of pore/crack constituent from 0.44 to 1.00. This suggests a considerable improvement in the segmentation accuracy of the organic/kerogen constituent. In Case I, a significant part of the pixels chosen for the test set were being wrongly identified as pore/crack by the model. An increase in the recall of organic/kerogen as well precision of pore/crack proves that the newly introduced support pixels of the organic/kerogen

constituent are being correctly identified by the classifier. Taking into account the above inferences, it can be said that the Random Forest classifier performs slightly better on the organic/kerogen constituent than on pore/crack. Therefore, increasing the number of support pixels is advantageous to understand the performance of a model on a constituent, but it is subjected to the availability and distribution of that constituent.

5.4. Effect of feature selection

Elimination of redundant features, also referred to as feature selection, helps enhance the efficiency of the model and reduce the curse of dimensionality. Feature ranking facilitates feature selection/elimination. Feature ranking is a crucial step required to understand the contribution of each feature to the overall performance of any supervised learning process. In this study, we use 16 features to train the classifier models. We investigated the effect of the features by a mean decrease accuracy (MDA) method. We used a Scikit-Learn compatible black-box estimator called Permutation Importance provided in the python library eli5. This method ranks a feature based on the decrease in the specified scoring metric when a particular feature is not available while testing a trained model. Applying this on the train set and re-training the model for each feature can be computationally intensive. Therefore, this method is implemented only on the test set where the feature values

are replaced with random noise and the mean decrease in model accuracies on the test dataset with randomized features are calculated.

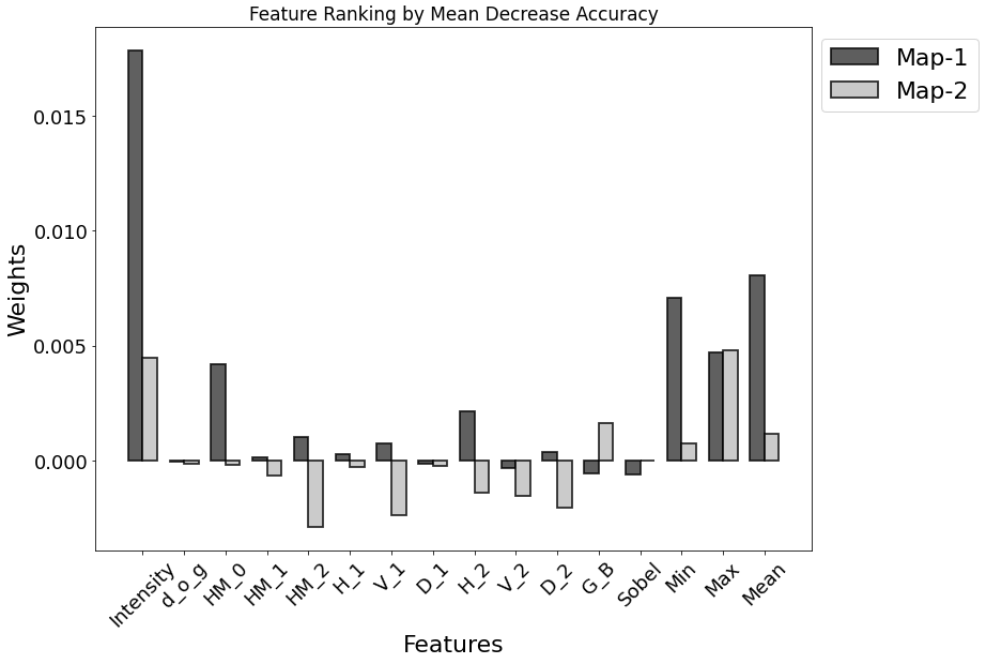

Fig. 16: Importances of the 16 features used in Model 3 to predict the pores, kerogen, and cracks.

In our study, we tested the Permutation Importance for the Model-3 trained on both Barnett and Wolfcamp:

- Case I: Testing of Model-3 on randomized features corresponding to pixels from Wolfcamp shale sample (Map-1)
- Case II: Testing of Model-3 on randomized features corresponding to pixels from Barnett shale sample (Map-2)

In Figure 16, the most important features in these 2 cases as calculated by the MDA method are: Pixel Intensity and Local statistical information (minimum, maximum, and mean). Difference of gaussian, Sobel edge detector and level-1 horizontal and diagonal Wavelet-transforms are the least important features. Use of Hessian matrix and wavelet transform as features adversely affects the performance on Map-2. Notably, even though Cases I and II are trained on the same training dataset (Model 3), the dominant features change depending on the testing dataset processed by the Model 3.

5.5. Comparison of the robustness of the proposed machine learning workflow against conventional threshold-based and Fiji-based segmentation methods

Here we compare the performance of the proposed machine learning workflow against two popular conventional segmentation methods, namely threshold-based segmentation and ML-assisted segmentation using the Fiji software. Threshold-based segmentation is a simple but popular method that identifies different constituents present in an image based on the pixel intensity ranging from 0 to 255. On the inner-region pixels from Slice 35 of Map-2, we compare the performance of the threshold-based segmentation method against the performance of the proposed machine learning workflow trained on 16 features of pixels from Map-2. We observe a small difference in the F1 scores between the two cases (Table 4). However, this

difference is more in case of outer region pixels (Table 5). With exposure to more features in the image, the classification of the interfaces, especially between pore/crack and organic/kerogen constituents, is considerably improved when using the proposed machine learning workflow.

Table 4: Comparison of threshold-based (I) and ML-based (II) segmentation methods for inner region pixels

Constituent	**Precision**		**Recall**		**F1 score**		**Support**
Cases	**I**	**II**	**I**	**II**	**I**	**II**	**Total: 5171**
Pore/crack	0.96	1.00	1.00	0.97	0.98	0.99	104
Organic/kerogen	0.88	0.99	0.99	1.00	0.93	1.00	476

Table 5: Comparison of threshold-based (I) and ML-based (II) segmentation methods for outer region pixels

Constituent	**Precision**		**Recall**		**F1 score**		**Support**
Cases	**I**	**II**	**I**	**II**	**I**	**II**	**Total: 3825**
Pore/crack	0.29	1.00	0.31	0.88	0.30	0.93	395
Organic/kerogen	0.28	0.91	0.29	0.98	0.28	0.95	722

Table 6: Comparison of segmentation performance of the proposed ML-assisted segmentation (II) method against the Fiji-based segmentation method (I) trained on Slice 45 of Map-2 when tested on inner region pixels of slice 10 of Map-2.

Constituents	**Inner region**						**Outer region**					
	Precision		**Recall**		**F1 score**		**Precision**		**Recall**		**F1 score**	
	I	**II**	**I**	**II**	**I**	**II**	**I**	**II**	**I**	**II**	**I**	**II**
Pore/ crack	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Organic/ kerogen	1.00	1.00	1.00	1.00	1.00	1.00	0.992	1.00	1.00	1.00	0.996	1.00
Matrix	1.00	1.00	1.00	1.00	1.00	1.00	0.994	1.00	0.984	1.00	0.989	1.00
Pyrite	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.993	1.00	0.996	1.00

We also compared the robustness of the proposed workflow with the automated segmentation results from the plugin WEKA (Waikato Environment for Knowledge Analysis) of the open source image processing software Fiji (Table 6). Both the methods have been trained on the same training set from Slice 45 of Map-2 and tested on the same test set from Slice 10 of Map-2. Random Forest classifier have been used in both cases keeping the hyper-parameters and the extracted features same for both the models. Both models have been tested on the inner and outer

region pixels of Slice 10 of Map-2. From the inner region, around 1900, 10000, 19000 and 1700 support pixels were chosen for pore/crack, organic/kerogen, matrix and pyrite constituents, respectively. For those constituents, support of around 650, 1100, 450 and 400 pixels were selected from the outer region. It was found that both the methods result in a perfect F1 score of 1.00 with almost zero rate of misclassification for the inner region pixels. Both machine-learning-assisted segmentation methods are robust for inner-region and outer-region pixels.

6. Conclusions

Machine learning (ML) can identify pores and cracks representing the hydrocarbon storage and transport pathways as well as kerogen in the scanning electron microscopy (SEM) images of organic-rich shale samples from two shale formations, namely Wolfcamp (Map-1) and Barnett (Map-2) shales. The two shale formations and the corresponding SEM maps differ in topology and distribution of pores, cracks, and kerogen. In our study, pixel intensity, Gaussian Blur and local pixel information (minimum, maximum and mean) are the most important feature out of the 16 features for purposes of identifying pores, cracks and kerogen. Wavelet decompositions and Hessian affine edge detector are the lowest ranked feature for the desired task.

Machine learning assisted segmentation method when trained and tested on the same formation exhibits 99% accuracy on inner region pixels and more than 80% accuracy on outer regions. The proposed machine learning workflow do not accurately and reliably identify pores, cracks and kerogen when it is trained on SEM images from one formation and then applied on images from different formation. Machine learning workflow trained on Wolfcamp shale can robustly detect matrix and pyrite in Barnett shale, but performs poorly when identifying the pores, cracks and kerogen. This is primarily due to identification of kerogen in the transition zone as pores. Moreover, matrix in the transition zone generally gets identified as pores, cracks or kerogen. In contrast, the machine learning workflow trained on Barnett shale can robustly detect pyrite in Wolfcamp shale but exhibits poor performance for the remaining constituents; for example, fails to detect cracks in Barnett shale. This is due to pores and cracks getting wrongly identified as kerogen or matrix; in addition, kerogen is wrongly identified as matrix. Machine learning workflow trained on both the shale formations exhibits the best performance with an average F1 scores of 0.99 and 0.91 on the inner-region and outer-region pixels, respectively.

The machine learning assisted identification of pores shows that the porosity ranges from 0.1 % to 0.8% in the 164.64 μm by 182.42 μm region of the Barnett sample.

Extremely high porosity and extremely low porosity regions in the Barnett sample are localized. Further, the machine learning assisted identification of kerogen shows that the micro-scale (<20 pixels) and macro-scale (>500 pixels) pore sizes contribute equally to the total pore volume of the Barnett shale sample. The macro-scale porosity of the system is primarily made of the organic pores surrounded by kerogen. Barnett shale sample predominantly contains kerogen clusters of size ranging from 10,000 pixels to 20,000 pixels.

Following scenarios improve the robustness of machine learning assisted identification of pores, cracks and kerogen: (1) large sizes of training and testing datasets sampled from different formations, (2) hyperparameter tuning, (3) use of random forest classifier and unscaled features, (4) perform simple feature extraction by considering the variations in intensities of neighboring pixels, (5) perform feature ranking after feature extraction to reduce the dimensionality, and (6) create a balanced testing dataset and use weighted F1 score to better evaluate the performance.

Acknowledgements

We thank the Financial support from the American Chemical Society's Petroleum Research Foundation (ACS-PRF #59363-DNI Grant). We also thank the University of Oklahoma Research Council's Faculty Investment Program that supported a few aspects of this project. We thank the Integrated Core Characterization Center, led by Dr. Chandra Rai and Dr. Carl Sondergeld, at the University of Oklahoma for providing us the SEM images for this study.

Declarations

The authors declare that they have no competing interests. EG, YW and SM developed and tested the machine learning and feature extraction methods used in the study. EG and SM prepared the first complete draft of the paper.

For queries, please contact Dr Sid Misra (misra@tamu.edu)

References

1. Tang, D., Spikes, K., 2017, Segmentation of Shale SEM Images using Machine Learning, SEG International Exposition and Annual Meeting. https://doi.org/10.1190/segam2017-17738502.1
2. Kaur, D., Kaur, Y., 2014, Various Image segmentation Techniques: A Review, IJCSMC, Vol. 3, Issue. 5, pg. 809-814.
3. Misra, S., & Wu, Y. (2019). Machine learning assisted segmentation of scanning electron microscopy images of organic-rich shales with feature extraction and feature ranking. Machine Learning for Subsurface Characterization, 289.
4. Matthew, A., 2018, A quantified study of segmentation techniques on synthetic geological XRM and FIB-SEM images, Computational Geosciences 22:1503–1512. https://doi.org/10.1007/s10596-018-9768-y.
5. Anemone, R., Emerson, C., Conroy, G., Finding Fossils in New Ways: An Artificial Neural Network Approach to Predicting the Location of Productive Fossil Localities, 2011, Evolutionary Anthropology.
6. Bauer, T., Strauss, P, 2014, A rule-based image analysis approach for calculating residues and vegetation cover under field conditions, CATENA 113, pg. 363-369.

7. Shen, S., Sandham, W., Granat, M., Sterr., 2005, MRI fuzzy segmentation of brain tissue using neighborhood attraction with neural-network optimization, IEEE Transactions on Information Technology in Biomedicine, Vol. 9, Issue. 3.
8. Clelland, W.D., Fens, T.W.,1991, Automated Rock Characterization With SEM/Image-Analysis Techniques, SPE Formation Evaluation, Vol. 6, Issue. 4.
9. Idowu, N.A., Nardi, C., Long, H., Varslot, T., Oren, P., 2014, Effects of Segmentation and Skeletonization Algorithms on Pore Networks and Predicted Multiphase-Transport Properties of Reservoir-Rock Samples, SPE Reservoir Evaluation & Engineering, Vol. 17, Issue. 4.
10. Budennyy, S., Pachezhertsev, A., Bukharev, A., Erofeev, A., Mitrushkin, D., 2017, Image Processing and Machine Learning approaches for Petrographic Thin Section Analysis, SPE Russian Petroleum Technology Conference.
11. Wu, Y., & Misra, S. (2019). Intelligent Image Segmentation for Organic-Rich Shales Using Random Forest, Wavelet Transform, and Hessian Matrix. IEEE Geoscience and Remote Sensing Letters.
12. Bulgarevich, D.S., Tsukamoto, S., Kasuya, T., Demura1, M., Watanabe, M., 2018, Pattern recognition with machine learning on optical microscopy images of typical metallurgical microstructures, Scientific Reports.

13. Xiao C, Chen Xi, Li W, Li L, Wang L, Xie Q and Han H (2018) Automatic Mitochondria Segmentation for EM Data Using a 3D Supervised Convolutional Network, 2018, Frontiers in Neuroanatomy, Vol 12, Article 92.
14. Misra, S., Ganguly, E., & Wu, Y. (2019). Generalization of machine learning assisted segmentation of scanning electron microscopy images of organic-rich shales. Machine Learning for Subsurface Characterization, 315.
15. Wu, Y., Misra, S., Sondergeld, C., Curtis, M., & Jernigen, J. (2019). Machine learning for locating organic matter and pores in scanning electron microscopy images of organic-rich shales. Fuel, 253, 662-676.