



# Phase segmentation of X-ray computer tomography rock images using machine learning techniques: an accuracy and performance study

Swarup Chauhan<sup>1</sup>, Wolfram Rühaak<sup>1,2</sup>, Hauke Anbergen<sup>3</sup>, Alen Kabdenov<sup>3</sup>, Marcus Freise<sup>3</sup>, Thorsten Wille<sup>3</sup>, and Ingo Sass<sup>1,2</sup>

<sup>1</sup>Department of Geothermal Science and Technology, Institute of Applied Geosciences, Technische Universität Darmstadt, Darmstadt, Germany

<sup>2</sup>Darmstadt Graduate School of Excellence Energy Science and Engineering, Technische Universität Darmstadt, Darmstadt, Germany

<sup>3</sup>APS Antriebs-, Prüf- und Steuertechnik GmbH, Göttingen, Rosdorf, Germany

Correspondence to: Wolfram Rühaak (ruehaak@geo.tu-darmstadt.de, w.ruehaak@online.de)

Received: 29 February 2016 – Published in Solid Earth Discuss.: 1 April 2016

Revised: 14 June 2016 – Accepted: 24 June 2016 – Published: 19 July 2016

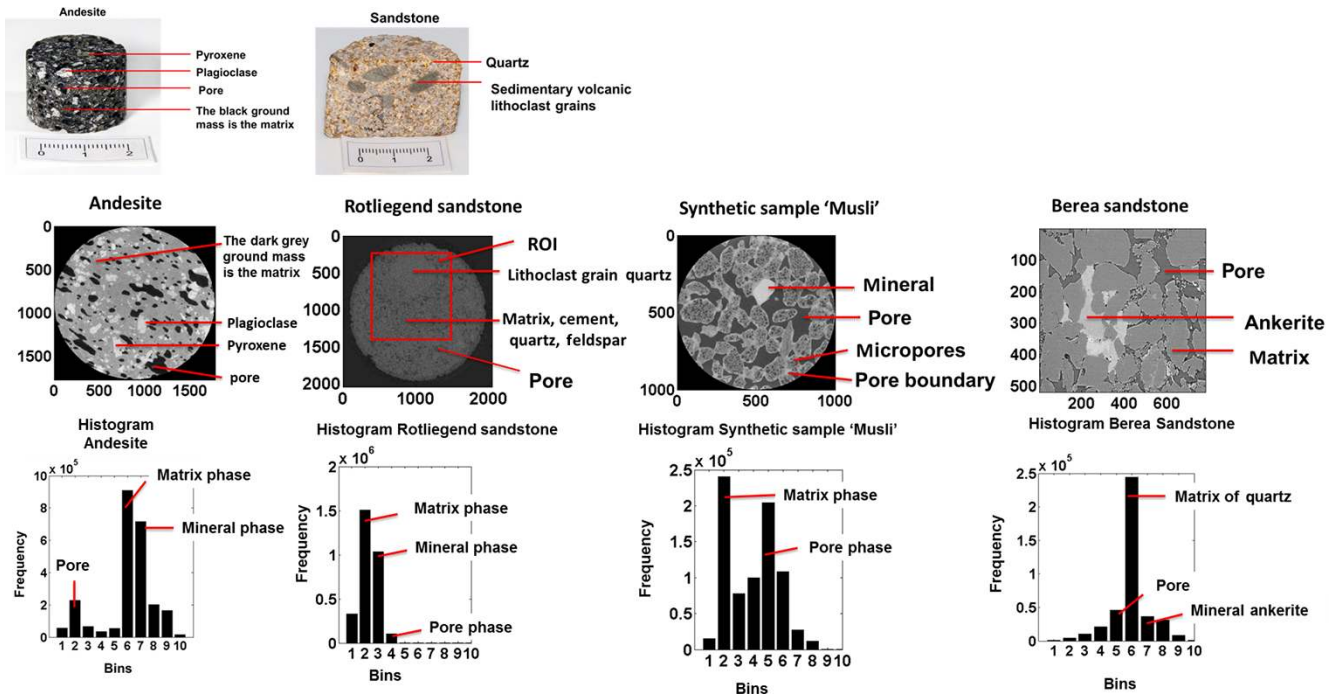
**Abstract.** Performance and accuracy of machine learning techniques to segment rock grains, matrix and pore voxels from a 3-D volume of X-ray tomographic (XCT) grayscale rock images was evaluated. The segmentation and classification capability of unsupervised (*k*-means, fuzzy *c*-means, self-organized maps), supervised (artificial neural networks, least-squares support vector machines) and ensemble classifiers (bragging and boosting) were tested using XCT images of andesite volcanic rock, Berea sandstone, Rotliegend sandstone and a synthetic sample. The averaged porosity obtained for andesite ( $15.8 \pm 2.5\%$ ), Berea sandstone ( $16.3 \pm 2.6\%$ ), Rotliegend sandstone ( $13.4 \pm 7.4\%$ ) and the synthetic sample ( $48.3 \pm 13.3\%$ ) is in very good agreement with the respective laboratory measurement data and varies by a factor of 0.2. The *k*-means algorithm is the fastest of all machine learning algorithms, whereas a least-squares support vector machine is the most computationally expensive. Metrics entropy, purity, mean square root error, receiver operational characteristic curve and 10 K-fold cross-validation were used to determine the accuracy of unsupervised, supervised and ensemble classifier techniques. In general, the accuracy was found to be largely affected by the feature vector selection scheme. As it is always a trade-off between performance and accuracy, it is difficult to isolate one particular machine learning algorithm which is best suited for the complex phase segmentation problem. Therefore, our investigation provides pa-

rameters that can help in selecting the appropriate machine learning techniques for phase segmentation.

## 1 Introduction

Micro X-ray computer tomography (XCT) images of a rock sample help in classification of pore space and assist in modeling of pore-network geometries. Pore-network geometries give an insight into the evolution of permeability and porosity of a rock sample. Image segmentation is the first step toward pore-network modeling. While developing this pore-network model discrimination between porous space and throat has to be resolved to the best possible extent. Currently this discrimination is still subjective (Piller, et al., 2009 and De Boever et al., 2012). A well-segmented 2-D or 3-D image of porous geometry provides a good foundation to obtain effective permeability and porosity trends.

Accurate segmentation of different phases from XCT rock images is a well-know and complex problem in the digital rock physics community (DRP). In general, tomography is a technique that generates a dataset (images), called a tomogram, which is a three-dimensional representation of the structure and variation of composition within a rock specimen. Each three-dimensional data point in the tomogram is called a voxel and contains a coefficient value associated with



**Figure 1.** The top panel shows the andesite and Rotliegend sandstone rocks used for XCT measurements. Middle panel shows the raw images of andesite (16 bit), Rotliegend sandstone (16 bit), synthetic sample (16 bit) and Berea sandstone (16 bit). Mineral composition of andesite and Rotliegend sandstone was determined from thin sections using a polarized microscope. Bottom panel shows histogram plot of the respective samples. Mineral composition of Berea sandstone is based on Madonna et al. (2012) and Andrä et al. (2013).

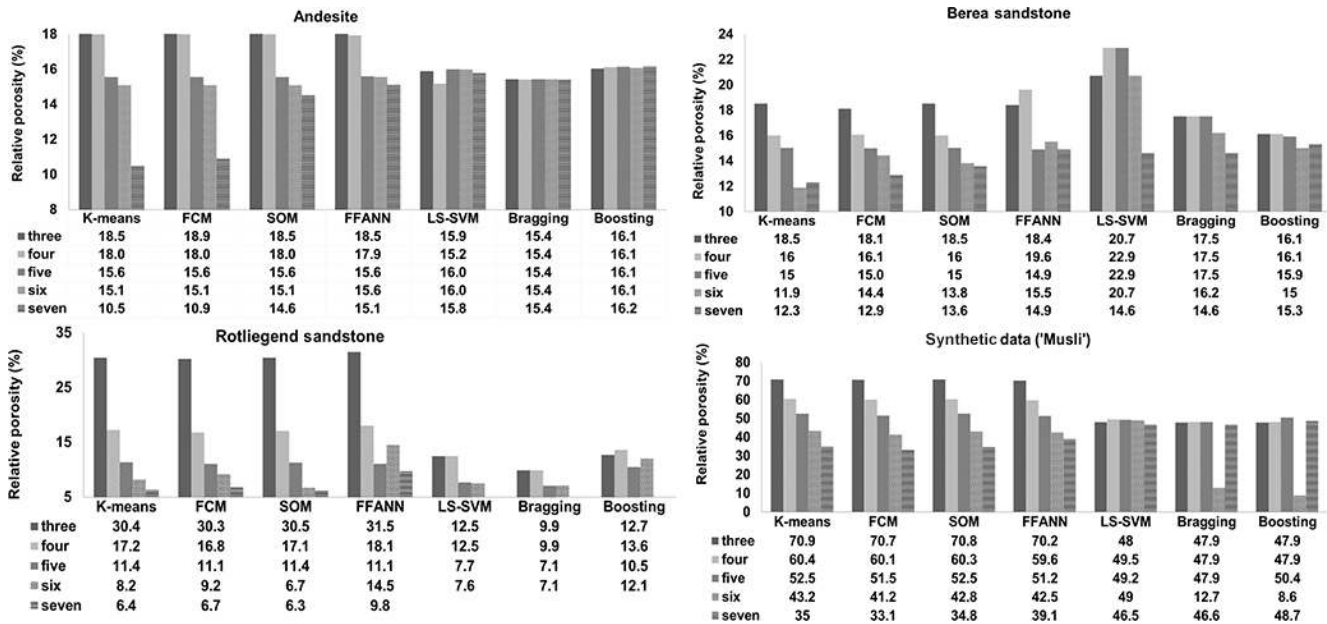
the density of the specimen. X-ray micro computed tomography involves collecting a tomogram using high-energy X-rays to achieve very high voxel resolution.

Segmentation is the partitioning of a tomogram (grayscale image) into disjoint regions that are homogeneous with respect to some characteristic. Porous materials like sedimentary and volcanic rocks contain areas of void, called the pore space, as well as a number of distinct mineral components, each with a comparatively uniform density. These different components are referred to as phases. Segmentation of a porous rock means deciding to which phase each voxel belongs. Tomographic images of such materials consist of a cubic array of reconstructed linear X-ray attenuation coefficient values each corresponding to a voxel of the sample. Ideally, one would wish to have a multi-modal distribution giving unambiguous phase separation of the pore and various mineral phase peaks. For flow properties, in particular, one would like to obtain a clear distribution separating the pore phase from mineral phase peaks. Unfortunately, the presence of low-density pore inclusions (e.g., microporosity, clays) below the image resolution can lead to a spread in the low-density signal making it difficult to unambiguously differentiate the pore from the microporous and solid mineral. As a consequence, significant features can be lost, and macroscopic properties of the segmented image can vary greatly with small changes in the segmentation parameters.

There have been extensive studies in various international groups to improve segmentation methods for better quantitative characterization of pore space feature. Iassonov et al. (2009) in their survey broadly classified segmentation algorithms into two types: (i) global thresholding segmentation scheme and (ii) local adaptive segmentation schemes.

The fundamental concepts behind global thresholding schemes is the histogram representation of the intensity and variation of all the gray pixels in a scene. There are many subcategories in the scheme, and the most commonly used are the histogram shape (triangulation) (Zack et al., 1977; Rosin et al., 2001 and Sund and Eilertsen, 2003) or the signal entropy consideration (Pal and Pal, 1989 and Pal, 1996).

The local adaptive segmentation scheme is governed by the fact that segmentation decision is made for each pixel in the scene. Utilization of local information generally provides better segmentation quality and account for some image artifacts, but it requires higher computation demand and memory. The most commonly used are locally adaptive (LA)-Kriging (Oh et al., 1999) and probabilistic fuzzy *c*-means (PMC)-Pham, which uses indicator Kriging, somewhat similar to LA-Kriging, except that the final result is obtained from fuzzy cluster membership (Pham, 2001). PMC-Pham belongs to the unsupervised segmentation category but due to the iteration scheme needs more computational power. Edge detection (ED)-Yanowitz is a technique based on edge



**Figure 2.** Relative porosity values obtained using unsupervised, supervised and ensemble classifier techniques for respective samples.

detection and surface procedure proposed by Yanowitz and Bruckstein (1989). Convergence active contours (CAC)-Sheppard is a hybrid method developed by Sheppard et al. (2005) which uses a combination of image enhancement, thresholding and convergence active contours. Markov random fields (MRF)-Berthod is an algorithm for supervised Bayesian image classification using Markov random fields developed by Berthod et al. (1996). The general drawback of CAC-Sheppard and MRF-Berthod methods can be attributed to long processing time caused by insufficient ground truth initialization and long processing time due to the simulated annealing technique. Jovanovic et al. (2013) proposed a segmentation scheme which can be performed already at the stage of sinograms. Cortina-Januchs et al. (2011) used a segmentation/classification technique based on a combination of clustering and artificial neural network (ANN) to segment binary soil images, whereas Khan et al. (2016) used the supervised technique least-squares support vector machine (LS-SVM) for segmentation of XCT rock images. Therefore, with the continuously improving CT technologies and computational resources, machine learning (ML) techniques can be an effective tool for segment and classify for phase segmentation of XCT rock images. Based on the heterogeneity of the sample the user can employ different ML techniques to obtain the best segmented image(s), which can be further used for simulating physical processes.

In Chauhan et al. (2016), we developed a workflow to segment XCT images using unsupervised, supervised and ensemble classifier ML techniques. The focus of this study is to assess the performance and accuracy of the above mentioned ML techniques to segment rock grain, matrix and pore phases

in heterogeneous rock samples such as andesite, Berea sandstone, Rotliegend sandstone and a synthetic sample containing microporosities.

## 2 Experimental approach

For this study andesite (Tongariro National Park, New Zealand); Berea sandstone (Andrä et al., 2013); Rotliegend sandstone (Rotliegend, Germany); and the synthetic sample “Musli” (henceforth referred to simply as Musli; provided by APS Antriebs, Prüf und Steuertechnik GmbH, Rosdorf, Germany), were used. Figure 2 shows the rock samples and respective histogram plots obtained from the XCT raw files. Effective porosity of andesite ( $17 \pm 2\%$ ) and Rotliegend sandstone ( $14 \pm 2\%$ ) was measured using a GeoPyc pycnometer (Micromeritics Instrument Corporation, Norcross, GA, USA). Thin-section analysis using a polarized microscope revealed andesite has a porphyritic texture with large plagioclase crystals (up to 3 mm in diameter), pyroxene in a cryptocrystalline matrix and isolated vesicles up to 6 mm in diameter (Chauhan et al., 2016). Rotliegend sandstone had different grain size (between 0.5 and 5 mm) of fine sand and gravel, with 26% monocrystalline quartz, up to 35% polycrystalline quartz, 8% feldspar, 9% sedimentary volcanic lithoclast grains and 13% cement (Aretz et al., 2016). Andrä et al. (2013) confirm that the porosity of the Berea sandstone (total porosity 19.97%; TM Petroleum Cores, Ohio, USA) was performed using an Helium Pycnometer 1330 (Micromeritics Instrument Corp., Belgium) and a mercury porosimetry using a Pascal 140+440 Mercury Porosimeter (Thermo Electron Corporation, Germany). Madonna et

al. (2012), using a scanning electron microscope, revealed Berea sandstone has ankerite, quartz, zircon, K feldspar and clay. The synthetic sample contained large pores, micropores and mineral grain.

Andesite volcanic rock and Rotliegend sandstone were imaged using a custom-built XCT scanner based on the CT-ALPHA system (ProCon, Sarstedt, Germany) at the Institute for Geosciences laboratory in Mainz, Germany. The samples were scanned by applying X-ray energy of 110 keV and using a prefilter of 0.3 copper. During the reconstruction of the projections, a noise filter was not used. The projections were radon-transformed in sinograms and thereafter converted through back projection into tomograms. These stacked tomograms resulted in 16 bit 3-D imagery, with a resulting voxel resolution of 13 and 21  $\mu\text{m}$  for andesite and sandstone, respectively. Andesite required no beam hardening correction (BHC), whereas BHC for sandstone was done based on regression analysis using 2-D paraboloid fitting. Finally, the tomograms are saved in raw format.

The Berea sandstone dataset was obtained from the GitHub FTP server (<https://github.com/cageo/Krzikalla-2012>). Andr a et al. (2013) performed XCT scans at the Tomographic Microscopy and Coherent Radiology Experiments (TOMCAT) (Stampanoni et al., 2006) beamline at the Swiss Light Source (Paul Scherrer Institute, Villigen, Switzerland). The beam energy was tuned for best contrast at 26 keV with an exposure time of 500 ms to retrieve a magnification of factor 10 (Andr a et al., 2013). The projections were magnified by microscope optics and digitized by a high-resolution CCD camera (PCO.2000) to obtain images of dimension  $1024 \times 1024 \times 1024$  with voxel resolution of 0.74  $\mu\text{m}$ . Tomographic images were reconstructed from the sinograms by applying Fourier transform spectroscopy (Marone et al., 2009) and saved in the desired file formats (Andr a et al., 2013).

## 2.1 Image pre-processing

Each of the 16 and 8 bit 3-D reconstructed raw images resulted in  $2048^3$  and  $1024^3$  voxels. The image filtering techniques such as blur, background intensity variation and contrast were tested on all the raw images before the segmentation and classification algorithms were initialized. In the case of Rotliegend sandstone (21  $\mu\text{m}$ ), as the XCT images were noisy, a contrast filter was used to enhance the image; for other XCT images (Berea, andesite and Musli), as the resolution and contrast were sufficiently high (7.5 to 13  $\mu\text{m}$ ), using filters did not show any noticeable change. The following sections describe the post-processing algorithm and how these were implemented in our image processing schemes.

## 3 Machine learning

The main focus of this study is to demonstrate the computational performance and accuracy of the different ML algorithms to segment/classify different phases in XCT rock samples – i.e., to map pixels of similar values into respective classes. ML algorithms rely on features; features are sets of instances which contain descriptive information based on which the ML algorithm trains its classification model and further identifies these features in an unknown dataset and groups them into respective classes, which in our case are the associated feature values of noise, rock grain, matrix and pore voxels. ML algorithms in general fall into categories of unsupervised, supervised and ensemble classifiers.

### 3.1 Unsupervised techniques

In the unsupervised technique  $k$ -means (MacQueen, 1967), fuzzy  $c$ -means (FCMs) (Dunn, 1973) and self-organized maps (SOMs) (Kohonen, 1990) were used for segmentation of pore, mineral and matrix phases. The  $k$ -means clustering algorithm is one of the simplest unsupervised ML algorithms commonly used to address the clustering problem. The  $k$ -means algorithm through an iterative scheme calculates the Euclidean distance between the data point (pixel value) and its nearest centroid (cluster). The algorithm converges when the mean square root error of Euclidean distance reaches minimum; that is, when no further pixel is left to be assigned to the nearest centroid (cluster). The performance of the  $k$ -means algorithm is strongly governed by the initial choice of the cluster centres. The  $k$ -means has the tendency to terminate without identifying the global minimum of the objective function (Chauhan et al., 2016). Therefore, it is recommended to run the algorithm several times to increase the likelihood that the global minimum of the objective function will be identified.

Unlike  $k$ -means, in the FCM iterative scheme each data point can be a member of multiple clusters by varying the membership function (Jain, 2010 and Jain et al., 1999). The FCM clustering procedure involves minimizing the objective function

$$J_{\text{fcm}}(Z; U; V) = \sum_{j=1}^n \sum_{i=1}^k (\mu_{ij})^m \|x_i^{(j)} - c_k\|^2, \quad (1)$$

where  $c_k = \sum_{j=1}^n \mu_{ij} x_i$ ;  $c_k$  is the  $k$ th fuzzy cluster center,  $m$  is the fuzziness parameter (for  $m = 1$  FCM simplifies to  $k$ -means) and  $\mu_{ij}$  is the membership function. In our context, if we consider the entire raw image as a fuzzy set of data points (pixel values) which lie very close to each other, FCM uses membership criteria to “loosely” or “tightly” isolate subsets of rock grains, matrix and pore phase. The membership function influences the segregation of intersection subsets of values that lie in between rock grains/matrix phases for densely packed pixels (Rotliegend sandstone) and pore throat/matrix phases for the micropore dataset (Musli).

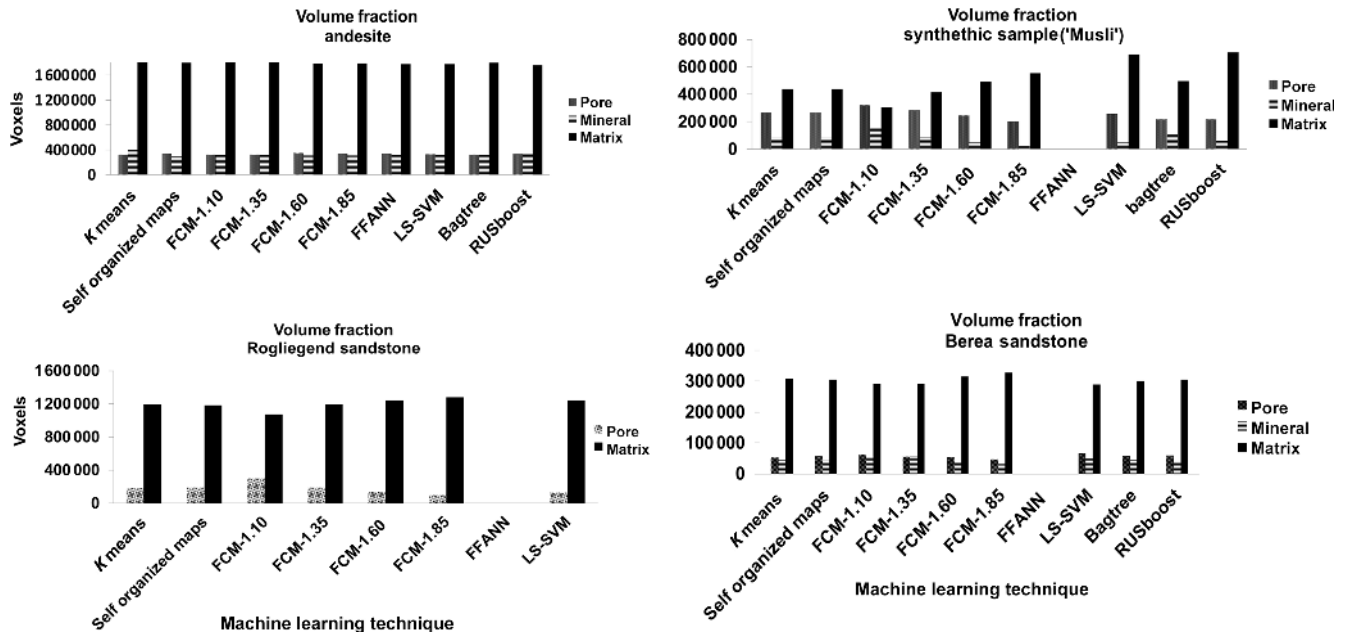


Figure 3. Total volume fraction plotted for respective samples.

FCMs can be a better choice in comparison to *k*-means, but it has a tendency to converge to the local minima of the objective function. Therefore, it is vital to test a range of membership values in combination with several centroids (classes) for accurate analysis (Cannon et al., 1986).

For a detailed description of SOMs the reader is referred to Kohonen (1990) and Chauhan et al. (2016). The SOM procedure uses a competitive learning process based on an ANN framework. In our context, a raw CT image is considered as an input pattern, which has to be classified. SOMs first arrange nodes (called neurons) in one of the desired topologies (grid, hexagon or random topology, as specified by the user) and assigns random weight (values). These nodes are trained using the pixel value of the CT image(s), iteratively using the Kohonen rule (Kohonen, 1990). During this competitive learning process the difference between the nodal weight and the neighboring pixel(s) is calculated. The iterative process stops when the difference reaches a minimum. The amount of adaptation of the nodal weight to its neighboring values can be influenced and monitored using learning rate parameter  $\alpha$ . The nodes that do not change to its surrounding value are classed as winner nodes. These winner nodes are nothing but different classes in the segmented image.

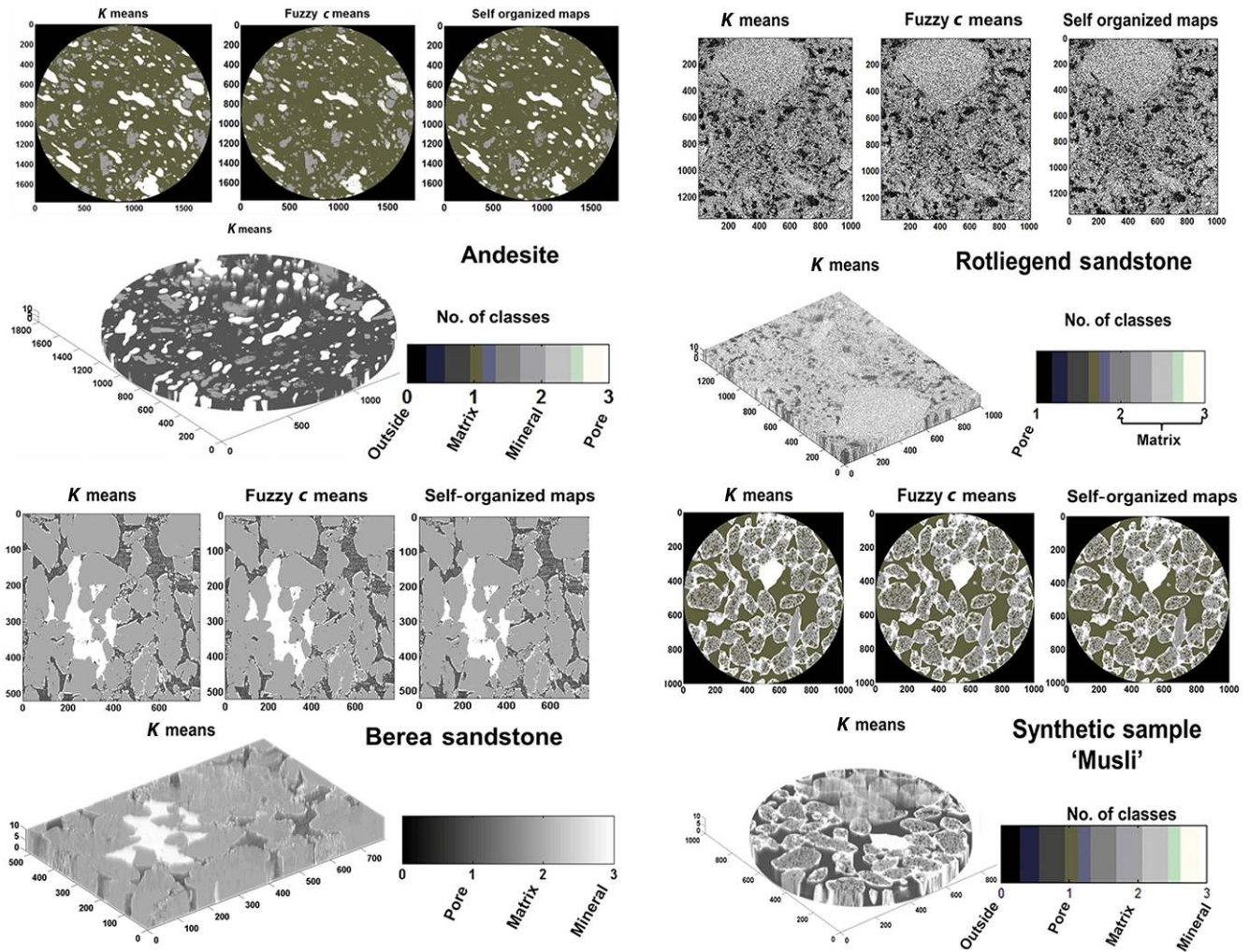
The unsupervised algorithms were configured to perform segmentation of three to seven classes. These classes in one-dimensional feature space are the non-overlapping segments of pixel bins in a histogram. Filter-based feature vector (FV) selection (Euclidian and Manhattan distance function) were used to initialize centroids for *k*-means, FCMs and SOMs. In the case of FCMs different degrees of membership values [1.10 to 1.85] were tested to loosely or tightly segregate pixel

values between rock grains and matrix phase. Grid topology was chosen in the case of SOMs.

### 3.2 Supervised techniques

In the supervised category feed forward artificial neural network (FFANN) (Jain et al., 1999) and LS-SVM (Suykens and Vandewalle, 1999) were used to classify rock grains, matrix and pore phases (Chauhan et al., 2016). In general, the supervised algorithms rely on a classification model which has to be trained using an example set of data that represent each class.

ANN is an information-processing paradigm that mimics the behavior of the human brain (Haykin, 1994). FFANN is based on the ANN framework and uses a so-called error back-propagation algorithm (Hopfield, 1982). FFANN can be used for any input–output mapping problem but is best suited for modeling linear and nonlinear problems. In our case the XCT dataset was partitioned into a training and testing dataset. Thereafter, FFANN was set up with input layer, one hidden layer and output layer. The hidden layer was assigned 10 nodes, and the nodes of input and output layer varied depending on training and testing slices. The *k*-means and FCM segmented datasets were used as a feature vector to train the classification model using the Levenberg–Marquardt back-propagation method (Levenberg, 1944; Marquardt, 1963). The classification model was tuned using the 10 K-fold cross-validation function (repeated trained and testing), and the misclassification rate was determined using mean square root error (MSE). Once the classification model reached optimal accuracy, it was tested on the rest of the XCT raw slices.



**Figure 4.** The top, middle and last panel show the 2-D segmented images and volume rendered plots of respective samples using unsupervised networks (andesite figure has been modified after Chauhan et al., 2016).

For LS-SVM a training dataset was created which contained a range of pixel values which best represented pore, mineral, matrix, cracks, trapped pores and noise regions; these pixel ranges were further labeled into different classes, which ranged from one to seven. For FFANN and LS-SVM the models were tuned using the 10 K-fold cross-validation function (repeated training and testing), and the misclassification rate was determined using mean MSE in the case of FFANN. Once the classification model reached an optimal performance threshold, it was tested on the rest of the XCT slices.

### 3.3 Ensemble classifier techniques

In the ensemble classifier technique RUSBoost and Bragtree algorithms are used (Seiffert et al., 2008; Breiman, 1996) to classify pore, rock grains and matrix phases (Chauhan et al., 2016). In general ensemble classifiers are a “bootstrap aggregation” of different weak classifiers. In general, weak

classifiers are algorithms which perform classification with a substantially high error rate ( $< 0.5$ ) – but slightly better than random guessing. The main advantage of bootstrapping such weak learner is to gain speed. The main difference between bragging and RUSBoost is the way they train their weak classifiers. Bragtree is an iterative scheme in which classifiers are trained with randomly chosen samples from the training dataset; in the second step the misclassified instances are collected and its classifiers are retrained until the misclassification error is minimized. However, RUSBoost sequentially trains its classifiers using the whole training set, essentially focusing on retraining inaccurate classifiers with the large dataset until its misclassification error is minimized. The ensemble classifiers were trained using the same FV which was used for LS-SVM. During the training process the ensemble models (of type RUSBoot and Bragtree) were parameterized using a (weak) classifier of type “Decision Tree” with a leaf size of five and trained up to 1000 training cycles. The learn-

ing rate, which is a parameter from a [0.0, 1.0] control overfitting range, was set to 0.1. Smaller values of learning rate require large numbers of weak learners to maintain a constant training error. Empirical evidence suggests that small values of learning rate favor better test error, as the constraint on the given number of weak learners maintains a constant training error.

### 3.4 Feature selection

In a practical rock CT segmentation/classification task, a priori information representing different phases (pore, matrix, rock, cracks, trapped pores etc.) in the XCT image is given to ML algorithms for segmentation or training the classification model. The dataset used as a priori information contains pixel values representing different phases in the XCT image, termed feature vectors. For unsupervised  $k$ -means, FCMs and SOMs, 10 slices from a XCT images were used to develop the FVs. For FFANN 5 images out of 10 were used to train the network; for LS-SVM and ensemble based classifiers different subset of pixels representing the pore, mineral, matrix, cracks, trapped pore and noise regions were used as feature vectors. The total number of pixels used to train and test each ML algorithm is shown in Table 1.

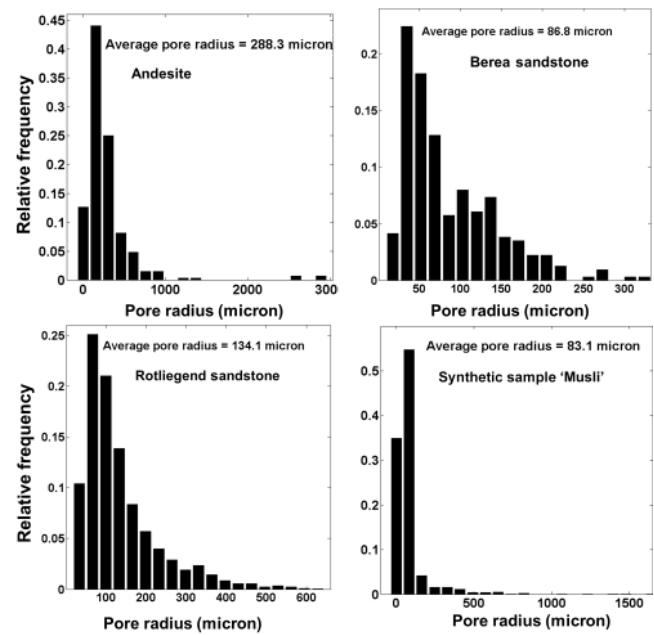
### 3.5 Performance

Computational performance was measured in terms of the segmentation and classification speed of the ML algorithms. Tests were performed on a Windows Server 2008 R2 Standard 64 bit operating system, with two six-core Intel Xeon processors, CPU (E645, 2.40 GHz) and installed memory (RAM) of 48.0 GB.

### 3.6 Accuracy

There is a wide set of evaluation metrics available to compare the quality of clustering and classification algorithms. For unsupervised clustering techniques accuracy can be evaluated intrinsically; i.e how close are the elements to each other within a cluster and how far apart from elements of other clusters (Amigó et al., 2009). Extrinsic metrics, on the other hand, are a comparison between the output of the clustering system and the *gold standard* usually built using human assessors (Amigó et al., 2009). Stehl (2002), Meilă (2003) and Amigó et al. (2009) proposed several types of cluster evaluation metrics tested on different mathematical constraints. However, the appropriate metrics for cluster evaluation is nontrivial and is still a subject of discussion. In this work, we use extrinsic evaluation metrics “purity” and “entropy”, which are most commonly used for clustering problems. The idea is to identify ideal class(es), representing the “best” porosity values, and to compare the clustering algorithm.

Any supervised classification is incomplete until the assessment of its accuracy has been performed. The supervised classification models are trained with a priori informa-



**Figure 5.** The pore size distribution of different rock samples using a watershed technique.

tion which is almost a subset of classes under investigation. Stehman et al. (2003) pointed out the accuracy assessment of a supervised classification problem can be assessed in three different steps: (1) the design of the model, (2) the response of the designed model to obtain the true classification rate (minimum error rate) and (3) the analysis of the classified data. The most common methods used for the analysis of the classified data are confusion matrix or  $\kappa$  – error statistics introduced by Fleiss et al. (1969). For our accuracy assessment study we have used step 2, the response of the designed model, to obtain the true classification rate: namely, metrics such as MSE, ROC and 10 K-fold cross-validation for AANN, LS-SVM and ensemble classifiers, respectively.

Subsections below illustrate all the metrics used for evaluating unsupervised, supervised and ensemble classifiers.

#### 3.6.1 Entropy and purity

The entropy of a class reflects how the members of the  $k$  pixels are distributed within each class; the global quality measure is by averaging the entropy of all classes:

$$\text{entropy} = -\sum_j \frac{n_j}{n} \sum_i P(i, j) \times \log_2 P(i, j), \quad (2)$$

where  $P(i, j)$  is the probability of finding an item from the category  $i$  in the class  $j$ , where  $n_j$  is the number of items in class  $j$  and  $n$  the total number of items in the distribution.

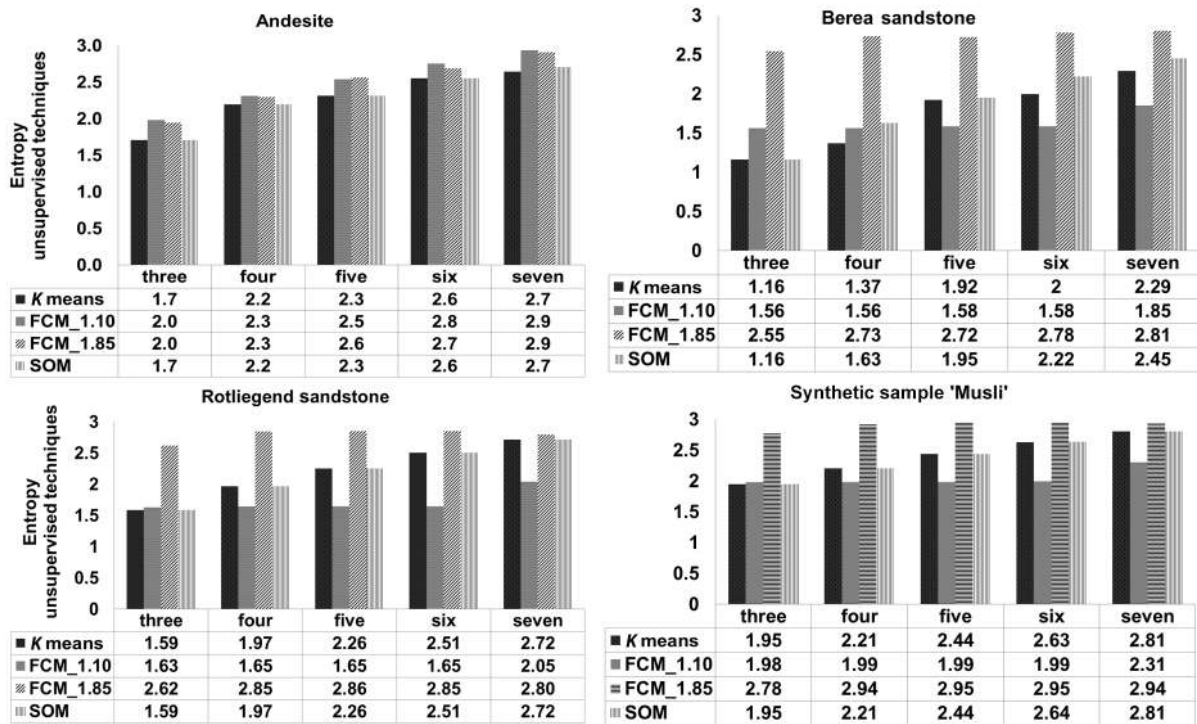
Purity focuses on the frequency of the most common category in each class. If  $C$  is the set of pixels to be evaluated

**Table 1.** The number of pixels used for training and testing the classification model.

Type of classifiers	Andesite		Rotliegend sandstone	
	No. of training pixels	No. of testing pixels	No. of training pixels	No. of testing pixels
<i>K</i> -means		31 577 290		13 681 600
Fuzzy <i>c</i> -means		31 577 290		13 681 600
Self-organized maps		31 577 290		13 681 600
Artificial neural networks	15 788 645	31 577 290	6 840 800	13 681 600
Least-squares support vector machine	2077	31 577 290	1511	41 943 040
Brugging and boosting	2077	31 577 290	1511	41 943 040

Type of classifiers	Synthetic sample (Musli)		Berea sandstone	
	No. of training pixels	No. of testing pixels	No. of training pixels	No. of testing pixels
<i>K</i> -means		10 000 000		4 056 000
Fuzzy <i>c</i> -means		10 000 000		4 056 000
Self-organized maps		10 000 000		4 056 000
Artificial neural networks	5 000 000	10 000 000	20 28 000	4 056 000
Least-squares support vector machine	1655	10 000 000	1366	4 056 000
Brugging and boosting	1655	10 000 000	1366	4 056 000



**Figure 6.** Entropy values of unsupervised techniques plotted for respective samples.

and  $L$  the set of classes,

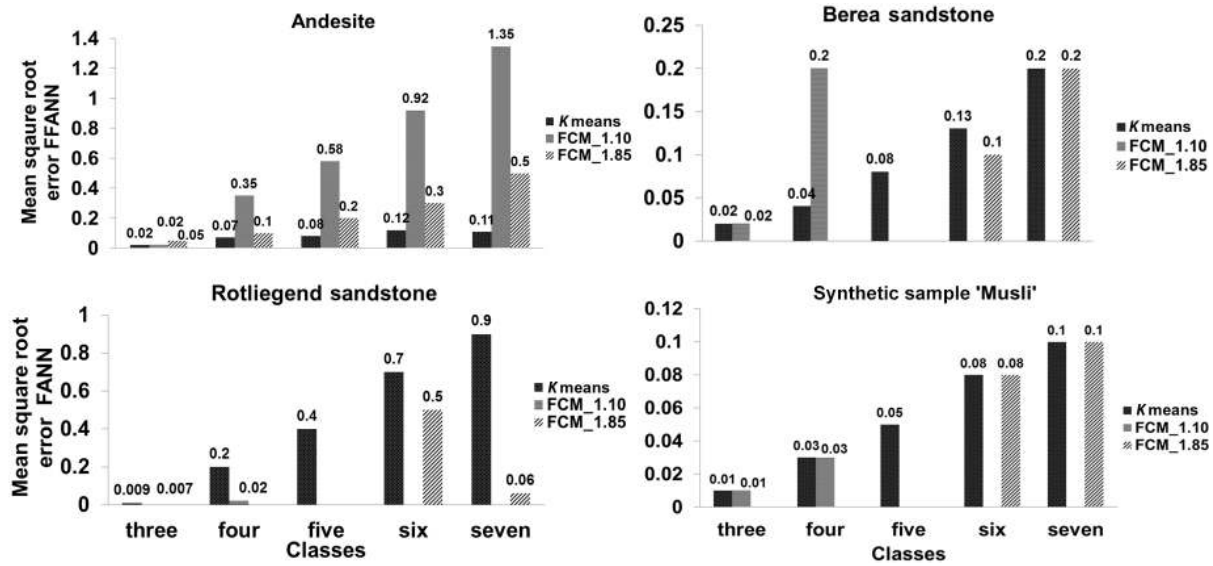
$$\text{purity} = \sum_i \frac{|C_i|}{N} \max_j \text{precision}(C_i L_j), \quad (3)$$

where the precision of a pixels  $C_i$  for a given classes  $L_j$  is defined as

$$\text{precision}(C_i L_j) = \frac{|C_i \cap L_j|}{|C_i|}. \quad (4)$$

Purity is a real number within  $[0, 1]$ ; the larger the purity values, the better is the clustering method. Conversely, the lower the entropy value, the better is the clustering performance.





**Figure 7.** Mean square root error values of feed forward artificial neural network (FFANN) obtained for respective samples. FFANN was trained using segmented datasets of *k*-means, fuzzy *c*-means with a membership function of 1.10 and 1.85

### 3.6.2 Mean square root error

The most commonly used error metrics to assess the accuracy of the FFANN are the MSE, the mean square relative error (MSRE), the coefficient of efficiency (CE) and the coefficient of determination ( $R^2$ ), shown in the equations below:

$$MSE = \sum_{i=1}^n \frac{(Q_i - \tilde{Q}_i)^2}{Q_i^2}, \tag{5}$$

$$MSRE = \frac{\sum_{i=1}^n \frac{(Q_i - \tilde{Q}_i)^2}{Q_i^2}}{n}, \tag{6}$$

$$CE = 1 - \frac{\sum_{i=1}^n (Q_i - \tilde{Q}_i)^2}{\sum_{i=1}^n (Q_i - \bar{Q}_i)^2}, \tag{7}$$

$$R^2 = \left[ \frac{\sum_{i=1}^n (Q_i - \bar{Q}_i) (\tilde{Q}_i - \bar{Q}_i)}{\sqrt{\sum_{i=1}^n (Q_i - \bar{Q}_i)^2 \sum_{i=1}^n (\tilde{Q}_i - \bar{Q}_i)^2}} \right]^2, \tag{8}$$

where  $\tilde{Q}_i$  are the classified images by FFANN.  $Q_i$  are the images used for training the FFANN (*k*-means and FCM images),  $\bar{Q}_i$  is the mean of the images used for training FFANN and  $\bar{Q}_i$  is the mean of the classified images To evaluate accuracy of our FFANN model, we looked at the MSRE values. The lower the MSRE value, the higher is the accuracy of the prediction.

### 3.6.3 Receiver operational characteristics

Receiver operational characteristic (ROC) curves have long been used in the signal detection theory (Bradley, 1997). It is

a good way of cross-validation of classifiers' accuracy (probability of classifiers correct response  $P(C)$ ).

$$\text{accuracy } (1 - \text{error}) = \frac{T_p + T_n}{C_p + C_n} = P(C), \tag{9}$$

$$\text{sensitivity } (1 - \beta) = \frac{T_p}{C_p} = P(T_p), \tag{10}$$

$$\text{specificity } (1 - \alpha) = \frac{T_n}{C_n} = P(n), \tag{11}$$

where  $T_p$  and  $T_n$  are the true positive and true negative examples and  $C_p$  and  $C_n$  are total number of true positive and true negative examples.

Probability of false positive is  $P(F_p) = \alpha$   
 Probability of true positive is  $P(T_p) = (1 - \beta)$

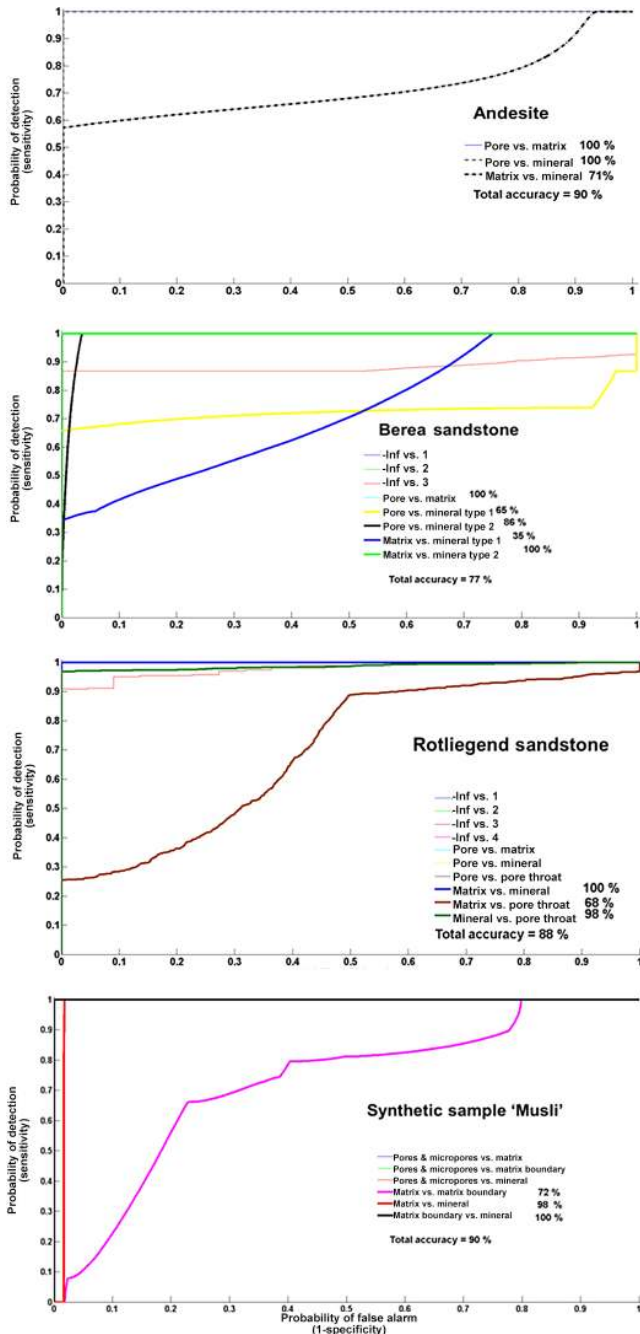
The accuracy is determined by calculating the area under the curve (AUC), and the simplest was to do that is by using trapezoidal approximation.

$$AUC = \sum_i \left\{ (1 - \beta_i \cdot \Delta\alpha) + \frac{1}{2} (\Delta(1 - \beta) \cdot \Delta\alpha) \right\} \tag{12}$$

In our case the AUC was determined using the trapezoidal approximation for each exponential curve, and the values were the fraction multiplied by 100 to obtain the value in percent.

### 3.6.4 10 K-fold cross-validation

The idea for cross-validation was first proposed by Larson (1931). Cross-validation is a statistical method of evaluating and comparing learning algorithms by dividing data



**Figure 8.** Receiver operational characteristic curves depicting the accuracy of the least-squares support vector machine multi-classification scheme for class four. A few curves which appear in the legend have close proximity to the x axis and lie behind other curves and therefore are invisible.

into two segments: one used to learn or train a model and the other used to validate the model. The problem with evaluating such a model is that it may demonstrate adequate prediction capability on the training data, but it might fail to predict future unseen data. Cross-validation is a procedure for estimating the generalization performance in this context.

Later, Kohavi (1995) and Dietterich (1998) investigated several approaches to estimate the accuracy of classifiers using different combinations of 10 K-fold cross-validation techniques; they recommended 10 K-fold cross-validation as one of the best cross-validation techniques, as it mitigates biases despite variances in the size of training and testing datasets.

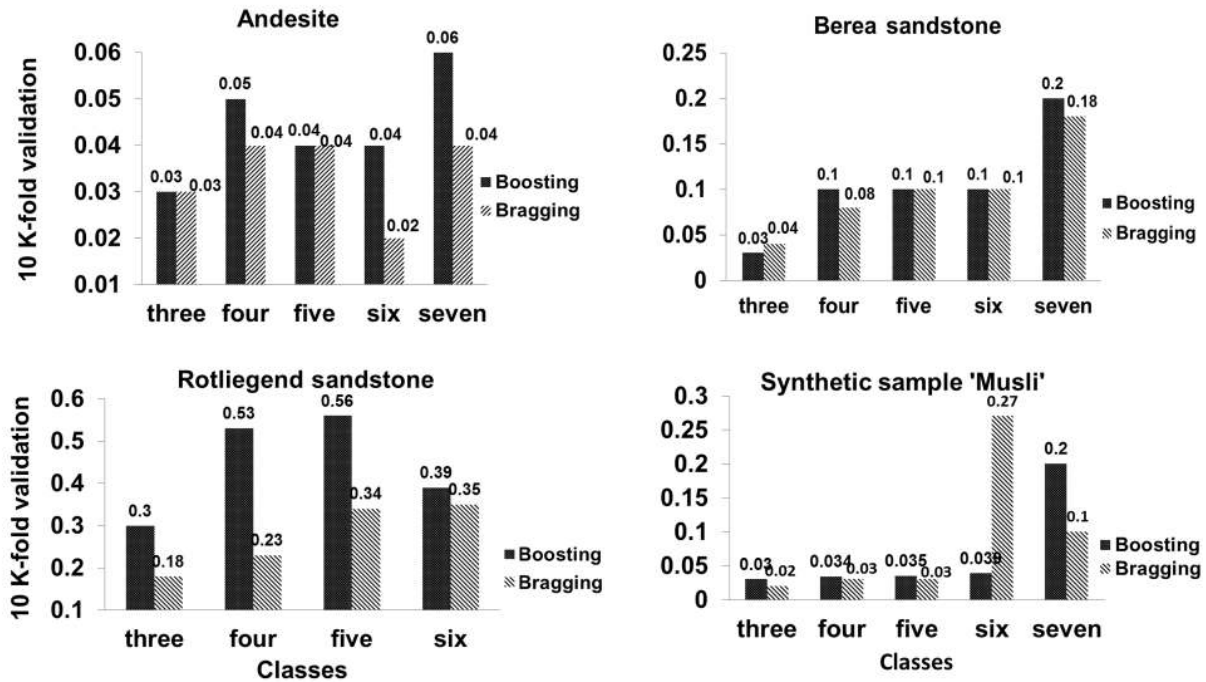
At the onset of 10 K-fold cross-validation, the dataset is initially stratified and partitioned into 10 equal (or nearly equal) segments or folds. Subsequently 10 iterations of training and validation are performed such that within each iteration a different fold of the data is held out for validation, while the rest of the folds are used for learning.

## 4 Results and discussions

### 4.1 Porosity and pore size distribution

The porosities which were determined from the stack of 10 XCT slices for three to seven classes using different ML techniques are shown in the Fig. 2. The estimated porosity is the ratio between the pore phase voxels and entire sample volume multiplied by 100. In general, the porosity using unsupervised ML techniques agrees well for all the four samples within  $\pm 1.2\%$  for each class. For andesite, Berea, sandstone, Rotliegend sandstone and Musli, the average estimated porosity sum over all classes is  $15.8 \pm 2.5$ ,  $16.3 \pm 2.6$ ,  $13.4 \pm 7.4$  and  $48.3 \pm 13.3\%$ , respectively. This is in good agreement with the experimental porosity values obtained for andesite and Rotliegend sandstone using a GeoPyc pycnometer and Berea sandstone as reported in Andr a et al. (2013). The large standard deviation in the case of sandstone and Musli is caused by the FCM segmentation scheme. When the membership function is tightly constrained [1.10, 1.35], the segregation between pore phase voxels and pore throat voxels is underestimated, contributing to the increase in porosity. Conversely, when the membership function is loosely constrained [1.60, 1.85], pore throat and micropores are segmented as matrix phases, resulting in a decrease in porosity and increase in matrix phase, which is clearly visible in the volume fraction plot of sandstone and Musli in Fig. 3. The low standard deviation in the estimated porosity values of andesite is due to the absence of microporosity and interconnected pores. The pore, mineral and matrix phases are distinct from each other; therefore the ML techniques have less difficulty in segmentation and classification. Figure 4 shows the segmented images using unsupervised technique and respective volume rendered images.

Pore size distribution (PSD) of andesite, sandstones and Musli was computed using the method suggested by Rabani et al. (2014). The segmented grayscale images were first converted to binary images using thresholding technique. Morphological and filtering operations were performed based on the complexity of the segmented images.



**Figure 9.** Accuracy of ensemble classifiers boosting and bragging calculated using 10 K-fold validation for respective samples.

Distance transform to convert the bright area into catchment basin and later watershed transformation was performed to segment the pore boundaries. Figure 5 shows the PSD and average pore radius of andesite, Berea sandstone, Rotliegend sandstone and Musli from *k*-means segmented images.

#### 4.2 Performance and accuracy analysis

Performance in the form of computational time is tabulated in Table 2. The *k*-means algorithm is the fastest among all the ML techniques because segmentation of phases into different classes is based on nearest-neighborhood distance measurements, unlike other ML techniques (exception: FCM), where the classification is governed by classification models.

In the case of supervised techniques the computational speed is correlated with the size of the feature vector used for training the classification model and post-processing of the unknown dataset. One reason is that supervised techniques are based on a “single” classification model; training and cross-validation of the model with a large amount of feature vectors consumes time. This can be related to the high computational time of the andesite sample using FFANN, where five slices were used to train the classification model compared to other samples where the classification model was trained using only one slice. For LS-SVM – as feature vector pixels are less than 1 % of the total pixel values for the all the samples – the training of the classification model took 1 to 10 min. The high computational time was consumed in post-processing, where a large unknown

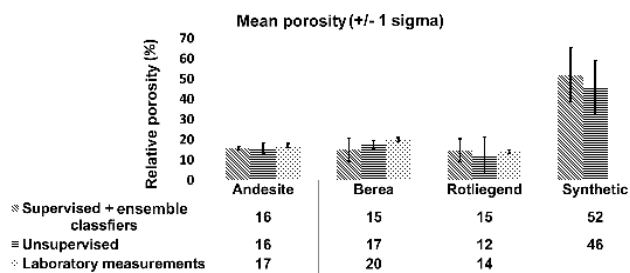
dataset was subjected to the trained model. In the case of ensemble classifiers the post-processing of an unknown dataset took longer compared to the training of the respective (bootstrapped weak) classification schemes. As the Rotliegend sandstone is densely packed with very low porosity, it resulted in low contrast and a badly resolved XCT dataset. As a consequence, the individual (weak) classification models required more computational time to achieve a consolidated, nearly accurate, well-classified result. Therefore, the processing time of Rotliegend sandstone images by ensemble classifiers was higher compared to other XCT samples.

Our clustering problem is to determine the most appropriate class for each pixel. That is, we wish to identify which of the unsupervised ML techniques satisfies properties of “cluster homogeneity” (i.e., not mixing items belonging to different categories) and “cluster completeness” (i.e., how well items belonging to same categories are grouped together) defined by Amigó et al. (2009). Therefore, the metrics entropy and purity were chosen to evaluate the accuracy of unsupervised ML techniques. The entropy values were calculated using Eq. (2) on the 3-D stack of 10 slices for each class and are shown in Fig. 6. In general class three and four have the lowest entropy values compared to other classes. This shows that if cluster homogeneity and cluster completeness get violated this may lead to misclassification. Among the three unsupervised ML techniques, *k*-means has the lowest entropy values; therefore it can be assumed that *k*-means performs the best segmentation compared to SOMs and FCMs.

**Table 2.** The computational time for processing 10 slices.

Machine learning techniques	CPU: time (h:min:s)			
	Andesite	Rotliegend sandstone	Synthetic sample (Musli)	Berea sandstone
<i>K</i> -means	00:15:35	00:12:04	00:10:59	00:05:33
FCM	00:29:19	00:56:03	00:42:21	00:41:05
SOM	01:07:06	1:41:47	01:11:23	00:33:32
FFANN (training using <i>k</i> -means)	08:58:18	00:11:50	00:10:40	00:11:12
LS-SVM*	63:29:35	03:22:58	03:02:15	01:45:17
Bragging	05:57:05	07:32:22	12:19:40	03:51:13
Boosting	07:47:05	09:52:56	06:14:58	03:20:42

\* Open-source public library provided by the University of Leuven's Department of Electrical Engineering (ESAT) SCD-SISTA division was used: <http://www.esat.kuleuven.be/sista/lssvmlab/>.



**Figure 10.** Mean porosity value obtained using supervised and ensemble classifiers as well as unsupervised machine learning techniques.

For FFANN the accuracy was interpreted using Eqs. (6) and (8) and the MSE shown in Fig. 7. FFANN was trained using *k*-means and FCM and was tested on raw XCT images of the respective samples. The testing dataset (3-D stack of raw images) was scaled between three and seven class values before the start of the testing cycle. In the case of Beria, Rotliegend and the synthetic sample, when the membership function was tightly constrained to 1.10, FCM was able to segment pore, matrix and mineral grain phases into a maximum of three and four classes. Similarly, on a moderate (1.60) and loosely constrained (1.85) membership function FCMs yield a maximum of five, six and seven classes, respectively. This explains the variance in the number of datasets used for validation of FFANN. The lower the MSE value, the better is the accuracy; the accuracy decreases with over-classification (for class five to six). Different settings, such as increasing the number of training slices up to five and increasing the number of neurons from 10 to 30, did not show any significant improvement in the accuracy. Among all the XCT samples, the worst accuracy was found for Rotliegend sandstone. Based on our analysis, we suggest that FFANN may not be the best-suited ML technique for clustering analysis.

In the case of LS-SVM, the low variance seen in the porosity values up to class six, is the indication that LS-SVM is one

among the most suitable ML techniques for phase segmentation analysis of XCT images. As the hand-picked feature vector dataset of class four had an appropriate mix of all the phases and the desired amount of noise, it gave the best trade-off between quality and speed. Hence we show the accuracy of LS-SVM for classification of class four using the ROC curve (Metz, 1978) in Figure 8. The slope of the ROC curve gives the accuracy of classification which was computed using Eq. (12). The accuracy ranges from 77 % for Beria sandstone to 88 % for Rotliegend sandstone and 90 % for andesite and Musli. Up to 100 % accuracy is achieved in discriminating the pore phase with respect to mineral and matrix phases.

Ensemble classifiers also show low variance in the porosity values as LS-SVM because of the same feature vectors used. The accuracy of the ensemble classifiers tested using the 10 K-fold cross-validation technique (Quinlan, 1996) is shown in Fig. 10. Both Bragging and Boosting classifiers were trained using the training dataset. The training dataset comprises the pixel values representing pore, mineral, matrix, noise phases and feature vectors. The initial growth of the leaf size was started with 5, and the corresponding weak classifiers were trained up to 1000 iterations (Breiman et al., 1996). The accuracy was determined by 10 K-fold cross-validation techniques. The best accuracy was achieved for andesite and Musli XCT (with an exception for class six) images, and the worst for Rotliegend sandstone, going up to 0.56.

## 5 Conclusions

In this study the performance and accuracies of ML techniques were validated, and relative porosity and pore size distribution of andesite (altered minerals), Beria sandstone, Rotliegend sandstone (interconnected pores) and Musli (microporosity) rock samples were computed. The total averaged porosity values obtained using unsupervised, supervised and ensemble classifiers are shown in Fig. 10 and are in good agreement with the experimental values obtained using the GeoPyc pycnometer and data reported in Andrä et

al. (2013). The high standard deviations up to 13 % seen in the case of Musli can be attributed to the misclassification caused by ensemble classifiers at class six. This can be seen in the porosity value of Musli in Fig. 2. The feature vector set corresponding to class six introduces noise information in the form of 73 pixels. When the training/testing was performed using feature vector up to class six, the ensemble classifiers showed high misclassification. Thereafter, when additional information on cracks and specks represented as 300 and 97 pixels is introduced as class seven (feature vector), the ensemble classifier stabilizes. It is difficult to speculate why this happens.

Our analysis shows unsupervised ML techniques perform well with filter-based feature extraction techniques. In terms of computational time,  $k$ -means outperforms all the other ML techniques. Fuzzy  $c$ -means can distinguish well between pore and pore-throat boundaries, given that the membership function is loosely constrained between 1.60 and 1.85. It was found that different tuning parameters (such as different FCM membership criteria and different SOM topologies and distance functions) need to be tested for the unsupervised techniques. A SOM topology “grid top” layout (neurons arranged in a grid format) and a SOM Manhattan distance function (sum of the absolute difference) gave consistent results, and FCM membership function within [1.35–1.85] gave consistent results. Low entropy values of  $k$ -means indicate that  $k$ -means is more accurate compared to fuzzy  $c$ -means and self-organized maps.

In the case of supervised techniques the computational time was significantly improved by reducing the training dataset of FFANNs and by careful selection of feature vector dataset for LS-SVM. Based on our analysis we conclude that FFANNs may not be best suited for clustering analysis; due to difficulty in scaling the training dataset (XCT raw files), the interpretation of clustering labels and accuracy becomes extremely difficult. Additionally, the accuracy in terms of mean square root error of the validation cycle (training and repeated testing) is largely regularized by fine and coarse scaling of the testing dataset, which may not always correspond to the image classification. As a consequence, there were cases where despite low accuracy (high MSE) the classification performed by FFANN was good. LS-SVM, however, proved to be one of the best and accurate supervised ML techniques for phase segmentation problem. However, it strongly relies on the craft with which the feature vector dataset is constructed. The user has the flexibility to decide which phases or feature are most relevant for phase segmentation. The authors suggest using the histogram plot of the raw image or  $k$ -means (or any other unsupervised ML technique) as an orientation for feature vector selection. It is further recommended that the first and second class labels (e.g., class three and class four) should contain predominantly phases such as pore, matrix, mineral and noise pixels. Consequently, other interesting feature pixels can be included. A suitable balance has to be found, such that the

classifier is not excessively trained on one particular feature and does not get stuck in local minima. Thereafter, the ROC curve validation technique is best suited for accuracy assessment of LS-SVM.

Ensemble classifier can be the second-best alternative to tackle phase segmentation problems as it also relies on the feature vector dataset to train the classification model; therefore, the user has more control over the classification scheme. However, the weak learners involved in the ensemble classification scheme remain as a black box to a large extent; therefore, appropriate tuning of the individual weak learners to optimize computational speed and accuracy may be cumbersome. To have a better control over the ensemble classification scheme, and for future work, we suggest an ensemble classifier with  $k$ -means, FCMs and LS-SVM as weak learners.

*Acknowledgements.* We thank Michael Kersten, Frieder Enzmann and his group at the Institute for Geoscience, Johannes-Gutenberg-Universität Mainz, for the high-resolution X-ray tomography measurements of andesite and Rotliegend sandstone. We thank Phillip Mielke and Achim Aretz for the fieldwork and collection of rock samples, and Heiko Andrä and her team at Fraunhofer ITWM Germany for making the synchrotron dataset of Berea sandstone freely available online.

This work is funded within the framework of the SUGAR (Submarine Gashydrat Ressourcen) III project by the German Federal Ministry of Education and Research (BMBF grant no. 03SX381H) and also partly supported by the DFG in the framework of the Excellence Initiative, Darmstadt Graduate School of Excellence Energy Science and Engineering (GSC 1070). The sole responsibility for the content lies with the authors.

We thank two anonymous reviewers and the editor, Steven Henkel, for their valuable comments which improved the manuscript.

Edited by: S. Henkel

Reviewed by: two anonymous referees

## References

- Amigó, E., Gonzalo, J., Artiles, J., and Verdejo, F.: A comparison of extrinsic clustering evaluation metrics based on formal constraints, *Inform. Retrieval*, 12, 461–486, 2009.
- Andrä, H., Combaret, N., Dvorkin, J., Glatt, E., Han, J., Kabel, M., Keehm, Y., Krzikalla, F., Lee, M., Madonna, C., Marsh, M., Mukerji, T., Saenger, E.H., Sain, R., Saxena, N., Ricker, S., Wiegmann, A., and Zhan, X.: Digital rock physics benchmarks – Part I: Imaging and segmentation, *Comput. Geosci.*, 50, 25–32, 2013.
- Aretz, A., Bär, K., Götz, A. E., and Sass, I.: Outcrop analogue study of Permocarboniferous geo-thermal sandstone reservoir formations (northern Upper Rhine Graben, Germany): impact of mineral content, depositional environment and diagenesis on petrophysical properties, *Int. J. Earth Sci.*, 105, 1431–1452, doi:10.1007/s00531-015-1263-2 2016.

- Bradley, A. P.: The use of the area under the ROC curve in the evaluation of machine learning algorithms, *Pattern Recogn.*, 30, 1145–1159, 1997.
- Breiman, L.: Bagging predictors, *Mach. Lear.*, 24, 123–140, 1996.
- Berthod, M., Kato, Z., Yu, S., and Zerubia, J.: Bayesian image classification using Markov random fields, *Image Vision Comput.*, 14, 285–295, 1996.
- Cannon, R. L., Dave, J. V., and Bezdek, J.: Efficient Implementation of the Fuzzy c-Means Clustering Algorithms, *IEEE T. Pattern Anal.*, 8, 248–255, 1986.
- Chauhan, S., Rühaak, W., Khan, F., Enzmann, F., Mielke, P., Kersten, M., and Sass, I.: Processing of rock core microtomography images: Using seven different machine learning algorithms, *Comput. Geosci.*, 86, 120–128, 2016.
- Cortina-Januchs, M. G., Quintanilla-Dominguez, J., Vega-Corona, A., Tarquis, A. M., and Andina, D.: Detection of pore space in CT soil images using artificial neural networks, *Biogeosciences*, 8, 279–288, doi:10.5194/bg-8-279-2011, 2011.
- De Boever, E., Varloteaux, C., Nader, F.H., Foubert, A., Békri, S., Youssef, S., and Rosenberg, E.: Quantification and prediction of the 3D pore network evolution in carbonate reservoir rocks, in: *Oil & Gas Science and Technology – Rev. IFP Energies nouvelles*, Vol. 67, No. 1., 161–178, 2012.
- Dietterich, T. G.: Approximate statistical tests for comparing supervised classification learning algorithms, *Neural Comput.*, 10, 1895–1923, 1998.
- Dunn, J. C.: A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters, *J. Cybernetics*, 3, 32–57, 1973.
- Fleiss, J. L., Cohen, J., and Everitt, B. S.: Large sample standard errors of kappa and weighted kappa, *Psychol. Bull.*, 72, 323–327, doi:10.1037/h0028106, 1969.
- Haykin, S.: *Neural Networks: A Comprehensive Foundation*, 1st Edn., Upper Saddle River, NJ, USA, Prentice Hall PTR, 1994.
- Hopfield, J. J.: Neural networks and physical systems with emergent collective computational abilities, *P. Natl. Acad. Sci. USA*, 79, 2554–2558, 1982.
- Iassonov, P., Gebrenegus, T., and Tuller, M.: Segmentation of X-ray computed tomography images of porous materials: A crucial step for characterization and quantitative analysis of pore structures, *Water Resour. Res.*, 45, W09415, doi:10.1029/2009WR008087, 2009.
- Jain, A. K.: Data clustering: 50 years beyond K-means, *Pattern Recogn. Lett.*, 31, 651–666, 2010.
- Jain, A. K., Murty, M. N., and Flynn, P. J.: Data clustering: a review, *ACM Comput. Surv.*, 31, 264–323, 1999.
- Jovanović, Z., Khan, F., Enzmann, F., and Kersten, M.: Simultaneous segmentation and beam-hardening correction in computed microtomography of rock cores, *Comput. Geosci.*, 56, 142–150, doi:10.1016/j.cageo.2013.03.015, 2013.
- Khan, F., Enzmann, F., and Kersten, M.: Multi-phase classification by a least-squares support vector machine approach in tomography images of geological samples, *Solid Earth*, 7, 481–492, doi:10.5194/se-7-481-2016, 2016.
- Kohavi, R. A.: study of cross-validation and bootstrap for accuracy estimation and model selection, in the *International Joint Conference on Artificial Intelligence (IJCAI)*, 1137–1145, 1995.
- Kohonen, T.: The self-organizing map, *P. IEEE*, 78, 1464–1480, 1990.
- Larson, S. C.: The shrinkage of the coefficient of multiple correlation, *J. Educ. Psychol.*, 22, 45–55, doi:10.1037/h0072400, 1931.
- Levenberg, K.: A method for the solution of certain problems in least squares, *Q. Appl. Math.*, 5, 164–168, 1944.
- MacQueen, J. (Ed.): *Some Methods for classification and Analysis of Multivariate Observations*, University of California Press, 1967.
- Madonna, C., Bjarne, S., Almqvist, G., and Saenger, E. H.: Digital rock physics: numerical prediction of pressure dependent ultrasonic velocities using micro-CT imaging, *Geophys. J. Int.*, 189, 1475–1482, 2012.
- Marone, F., Hintermüller, C., McDonald, S., Abela, R., Miluljan, G., Isenegger, A., and Stampanoni, M.: X-ray tomography microscope at TOMCAT, 9th International Conference on X-Ray Microscopy, *J. Phys. Conf. Ser.*, 186, 012042, doi:10.1088/1742-6596/186/1/012042, 2009.
- Marquardt, D.: An algorithm for least-squares estimation of nonlinear parameters, *SIAM J. Appl. Math.*, 11, 431–441, 1963.
- Meilä, M.: Comparing clusterings by the variation of information. *Learning theory and kernel machines*, Volume 2777 of the series *Lecture Notes*, in: *Computer Science*, Springer, Berlin, Heidelberg, 173–187, doi:10.1007/b12006, 2003.
- Metz, C. E.: *Basic principles of ROC analysis*, *Seminars in nuclear medicine*, 8, WB Saunders, 1978.
- Oh, W. and Lindquist, B.: Image thresholding by indicator Kriging, *IEEE T. Pattern Anal.*, 21, 590–602, 1999.
- Pal, N. R.: On minimum cross-entropy thresholding, *Pattern Recogn.*, 29, 575–580, 1996.
- Pal, N. R. and Pal, S. K.: Entropic thresholding, *Signal Processing*, 16, 97–108, doi:10.1016/0165-1684(89)90090-X, 1989.
- Pham, T. D.: Image segmentation using probabilistic fuzzy c-means clustering, in: *Image Processing, 2001. Proceedings. 2001 International Conference on*, vol. 1. IEEE, 722–725, doi:10.1109/ICIP.2001.959147, 2001.
- Piller, M., Schena, G., Nolich, M., Favretto, S., Raddelli, F., and Rossi, E.: Analysis of Hydraulic Permeability in Porous Media: From High Resolution X-ray Tomography to Direct Numerical Simulation, in: *Transp Porous Med* 80, 57–78, 2009.
- Quinlan, J. R.: Bagging, boosting, and c4.5, in: *Proceedings of the Thirteenth National Conference on Artificial Intelligence*, AAAI/MIT Press, 725–730, 1996.
- Rabbani, A., Jamshidi, S., and Salehi, S.: An automated simple algorithm for realistic pore network extraction from microtomography Images, *J. Petrol. Sci. Eng.*, 123, 164–171, 2014.
- Rosin, P. L.: Unimodal thresholding. *Pattern recognition*, 34, 2083–2096, doi:10.1016/S0031-3203(00)00136-9, 2001.
- Seiffert, C., Khoshgoftaar, T., Van Hulse, J., and Napolitano, A.: RUSBoost: Improving classification performance when training data is skewed, *ICPR 2008, 19th International Conference on Pattern Recognition*, 1–4, 2008.
- Sheppard, A. P., Sok, R. M., and Averdunk, H.: Improved pore network extraction methods, in: *International Symposium of the Society of Core Analysts*, 21–25, 2005.
- Stampanoni, M., Groso, A., Isenegger, A., Mikuljan, G., Chen, Q., Bertrand, A., Henein, S., Betemps, R., Frommherz, U., Böhler, P., Meister, D., Lange, M., and Abela, R.: Trends in synchrotron-based tomographic imaging: the SLS experience, in: *Proceedings of SPIE, The International Society for Optical Engineering*, 6318, M1–M14, 2006.

- Stehman, S. V. and Czaplewski, R. L. Introduction to special issue on map accuracy, *Environ. Ecol. Stat.*, 10, 301–308, 2003.
- Strehl, A.: Relationship-based Clustering and Cluster Ensembles for High-dimensional Data Mining, PhD thesis, The University of Texas at Austin, 2002.
- Sund, T. and Eilertsen, K.: An algorithm for fast adaptive image binarization with applications in radiotherapy imaging, *IEEE T. Med. Imaging*, 22, 22–28, 2003.
- Suykens, J. A. and Vandewalle, J.: Least Squares Support Vector Machine Classifiers, *Neural Process. Lett.*, 9, 293–300, 1999.
- Yanowitz, S. D. and Bruckstein, A. M.: A new method for image segmentation, *Comput. Vision Graph.*, 46, 82–95, 1989.
- Zack, G. W., Rogers, W. E., and Latt, S. A.: Automatic measurement of sister chromatid exchange frequency, *J. Histochem. Cytochem.*, 25, 741–753, 1977.