Apr 14th, 12:00 AM

Automated Feature Classification and Knowledge Extraction from Wireline Geophysical Observations: Big Data Potential for Offshore Resources Assessment

Galina Veres

Michelle Harris

Damon Teagle

Zoheir Sabeur

iEMSs International Congress on Environmental Modelling and Software

9th International Congress on Environmental Modelling and Software
Fort Collins, Colorado, USA, Mazdak Arabi, Olaf David, Jack Carlson, Daniel P. Ames (Eds.)
https://scholarsarchive.byu.edu/iemssconference/2018/

Automated Feature Classification and Knowledge Extraction from Wireline Geophysical Observations: Big data Potential for Offshore Resources Assessment

Galina Veres[1], Michelle Harris[2], Damon A.H. Teagle[3] and Zoheir A. Sabeur[1]
[1]IT Innovation Centre, School of Electronics and Computer Science, University of Southampton, UK
(gvv@it-innovation.soton.ac.uk; zas@it-innovation.soton.ac.uk)
[2]School of Geography, Earth and Environmental Sciences, Plymouth University
(michelle.harris@plymouth.ac.uk)
[3]Ocean & Earth Science,National Oceanography Centre Southampton, University of Southampton
(Damon.Teagle@southampton.ac.uk)

Abstract: Scientific drilling of the volcanic ocean crust recovers cores and undertakes downhole wireline logging. However, because core recovery rates are typically low (<30%), interpreting the wireline data is essential to gain a complete understanding of the stratigraphy. Ocean Drilling Program Hole 1256D samples 1500 m of in situ upper oceanic crust and has both core-derived lithostratigraphy and electrofacies classification based on geological interpretations of continuous downhole Formation MicroScanner imagery. We propose an automatic quantitative identification of electrofacies using Decision Trees. The cores and existing electrofacies classification provide training and verification of the automated classification. The identification of various classes is a challenging problem due to missing data, vertical shifts, horizontal misalignments, and multiclass unbalanced problem with 2 classes representing 50% of the data. Additionally, the structure of the same class changes with depth leading to large intra-class variations. Distinctive features for each class were identified by observation of images based on texture/shapes, and Decision Tree classifier was trained. Classification accuracy above 90% was achieved for the 3-classes for electrofacies with high recovery rates. In case of 9-classes, accuracy above 60% was achieved for some classes, though some challenges are remained due to strongly overlapped classes. A detailed analysis of the big data used for training the classifier and its performance is described. Combined analysis of drill cores and wireline geophysical data from scientific boreholes into volcanic rocks provides excellent training opportunities to develop automated rock classification methods for complex geological terranes that are of increasing interest to the hydrocarbons industry.

***Keywords**:* Machine learning; automated rock classification; big data analytics.

1 INTRODUCTION

A detailed understanding of the architecture and morphology of volcanic rocks that form the upper oceanic crust is essential to advance knowledge of the geodynamics of spreading ridges, hydrothermal exchanges between lavas and seawater, and geochemical recycling in subduction zones (Harris et al., 2015, 2017). Recovered rock types can range from thick coherent lava flows to highly fractured and brecciated flows. The distribution and abundance of the different flow types reflects the igneous processes operating at the mid ocean ridge and also has important consequences for the porosity and permeability of the crust. Undertaking sub-meter scale observations of oceanic lithosphere is challenging, primary because of the difficulty in direct continuous sampling by scientific ocean drilling, and the limited resolution of geophysical remote sensing methods (Tominaga, 2013). Ocean Drilling Program (ODP) Hole 1256D was the first hole to recover a complete section of intact fast spreading

rate oceanic crust and took place over 4 ODP/IODP Expeditions Leg 206, Expeditions 309, 312 and 335 (Teagle et al., 2006, 2012; Wilson et al, 2006). Core recovery was variable (average = 37%) but was biased towards the more coherent lithologies. Wireline logs provide quasi-continuous in situ physical property data with excellent quality in terms of borehole coverage compared to the variable core recovery. In particular, multidimensional wireline tool strings, such as the Formation Micro Scanner (FMS) and the ultrasonic borehole imager (UBI), provided high-resolution resistivity and sonic images in the situ borehole wall. These images enabled a qualitative reconstruction of the upper crustal architecture that comprises nine electrofacies identified from combined passes of formation microscanner images (Tominaga et al. 2009).

In addition to the manual electrofacies classification, Tominaga et al. (2009) also attempted to determine a quantitative lithostratigraphy using an artificial neural network (ANN). However, the attempt of automatically finding subsets of natural classes using unsupervised ANN clustering methods produced poor results, since unsupervised clustering aims to find natural clusters in the data which have strongly overlapping classes. At the same time, the automated interpretation of wireline data into an electrofacies stratigraphy would significantly reduce the current time required to generate it. It also adds value to the extensive database of wireline data generated by scientific ocean drilling and commercial hydrocarbon exploration. Previously some attempts were made towards automatic classification of different rock types. Yin et al. (2011) showed that using colour and shape features of rock images with reasonable correct recognition rates (>75%) of some rock types such as Lava, Tuff, Tuff Breccia, Volcanic Breccia and Ablation Breccia. Thomas et al. (2011) proposed automated lithology classification from core photographs with overall accuracy of 94.29% based on colour and texture of images and using representative sample for each class during training stage for shale, sandstone, carbonate cemented sandstone and 'no-core' (missing data). Recently, Lobos et al. (2016) developed a chain sequential classifier for 6 classes of rocks obtained during copper mining, where different features were identified and detected for each class. However, all the attempts mentioned above dealt with well-separated classes with intra-class variation significantly less than inter-class variations.

In this paper, the automatic quantitative identification of 9 strongly overlapped electofacies is proposed using selected features and Decision Trees classification (supervised learning). Labelling even a small subset of data for each class allows classification algorithms to identify automatically complex relationships and connections to separate strongly intertwined classes. To overcome the problem of overlapping classes, the analysis of FMS and UBI images was undertaken for each electrofacies (class), and discriminative features for each class were identified. This is then combined with additional wireline logging data for the electrofacies classification using Decision Trees (DT) algorithm.

2 MATERIALS AND METHODS

2.1 Data Acquisition and Pre-processing

Geophysical wireline data for Hole 1256D was measured over multiple logging phases during Leg 206, Expedition 309 and 312, and extends between 275– 1425 metres below the seafloor (mbsf) (Teagle et al. 2006, Tominaga et al. 2009). Data is available from the Borehole Research Group of the Lamont-Doherty Earth Observatory (http://brg.ldeo.columbia.edu/data/iodp-usio/). Standard data for wireline logging includes deep and shallow penetration resistivity, natural gamma ray, compressional velocity, density, photoelectric effect factor and porosity together with dynamic FMS (the Formation Microscanner) and UBI (the Ultrasonic Borehole imager) images with 50m intervals. However, both FMS and UBI images provided on the website are low resolution and do not have the same dynamic intervals for all three legs. Therefore, higher resolution processed FMS and UBI images were acquired from Dr. Masako Tominaga (Tominaga et al. 2009).

FMS and UBI images cover depths from 275 mbsf to 1425 mbsf and have a sampling rate of ~0.0017m. The FMS and UBI images were converted from RGB to grayscale format for further processing. Both images and wireline log data were brought to the same sampling rate of 0.002m and the same depth from 300 mbsf to 1412 mbsf. Logging runs from Leg 206, Expeditions 309 and 312 provide different quality of data for various parts of the hole. Generally, Leg 206 runs are better for the top of the hole, while Expedition 309 and 312 runs are better for the middle and base of the hole respectively. This is due to the degradation of the hole and increased caliper due to multiple passes of the drill string during

coring. Therefore, wireline log data corresponding to depths from 300 to 660 mbsf were extracted from Leg 206, while those at depths from 660 to 1084 mbsf from Expedition 309. The data at depths from 1084 to 1414 mbsf were extracted from Expedition 312.

The 1D wireline logging data used include porosity, natural gamma ray, deep and shallow penetration resistivity, and compressional velocity.

2.2 Qualitative and Quantitative Electrofacies Analyses

During a qualitative electrofacies analysis of in situ wireline logs nine classes were identified. These were based on texture, presence/absence/distribution of fractures, and ranges of conductivity. Specifically, the classes include *Massive flows (MAS), Massive off-axis Ponded Lava (MOPL), Fractured Massive Flows (FMAS), Fragmented Flows (FF), Pillow lavas (P), Thin flows or Thick Pillows (TP), Breccias (BR), Isolated dikes (ID) and Dikes (VD)* (Tominaga et al. 2009) which are identified as:

- *MAS:* flows more than 2m thick, homogeneous texture on FMS and UBI images, fracture network with narrow fractures**,** sub-horizontal contacts**,** variable density
- *MOPL:* rare fractures that are narrow and very discrete, many fractures are sub-horizontal, fewer fractures than MAS/FMAS, low porosity structure and homogenous texture
- *FMAS:* flows more than 2m thick with less homogeneous texture than MAS, rare shallow-to-moderate contacts**,** very common rounded fractures with no preferred orientation
- *FF:* contacts of variable dip**,** dense fractures, 'mesh' texture**,** patches of high conductivity
- *P:* curved margins/contacts, radial fractures, downward drooping tear drop shapes, irregular patches within more massive interiors, highly conductive interstitial material
- *TP:* ~1m thick flow, curved margins**,** shallow angle/contacts**,** small scale/thin fracture networks
- BR: highly irregular and high contrast**,** no discernible fractures**,** overall has spotty appearance**,** highly variable**,** no morphology preservation, highly fragmented textures
- *ID:* vertical fractures – step deep fractures**,** more uniform texture than MAS/FMAS/FF.
- *VD:* sub-parallel sub-vertical fractures, high conductivity contacts

This qualitative assessment of electrofacies is used as a basic foundation for testing our automated machine classification method.

The quantitative analysis of these classes is a challenging machine classification problem. Firstly, the FMS images had significant missing data with some vertical shifts, while UBI images had horizontal misalignments, as show in Figure 1. Data gaps in FMS images are quite large due to partial coverages of the borehole wall by each pass. Combined with the vertical shifts, it adds on more challenges for machine classifiers to recognize the representative features and the relevant distinction of the respective classes. Horizontal shifts in UBI images cause images to become blurred, and potentially smudge important features such as lines, curves and blobs.

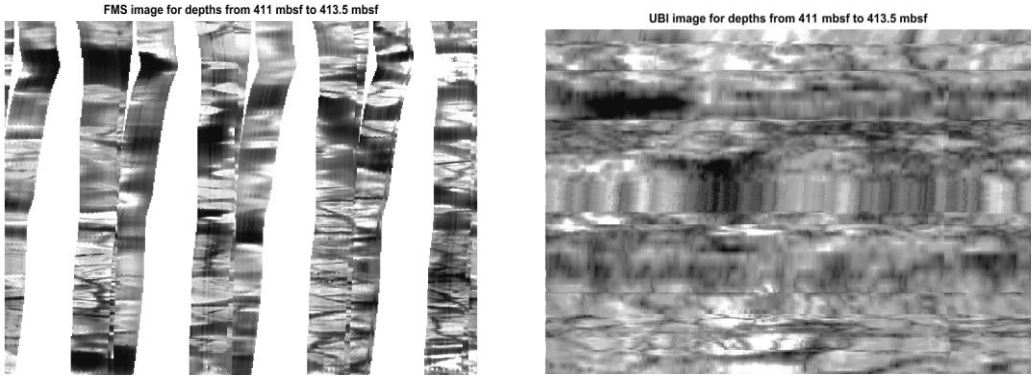

Figure 1. FMS image (left) with missing data and vertical shifts; UBI image (right) with horizontal data. Both images from 411 – 413.5 mbsf.

Secondly, we are being challenged with an unbalanced multiclass problem (Veres and Sabeur, 2013, 2015). The relative proportion of the different classes range from 1.4% up to 31.3% (Table 1). VD and FF classes are the most abundant and P and TP are the least abundant.

Table 1. Classes' distribution for a whole dataset, % (Tominaga et al. 2009).

BR	FF	FMAS	ID	MAS	MOPL	P	TP	VD
12.9	23.1	5.2	8.6	9.8	4.3	1.4	3.4	31.3

Finally, the structure captured by FMS and UBI images for the same class can change with depth, and the very large intra-class variations lead to poor class separation.

2.3 Feature Identification and Extraction

Taking into consideration qualitative and quantitative electrofacies analyses and manually inspecting FMS and UBI images, it was decided that FMS images can be used to extract distinctive shapes, while UBI images are suitable for extracting distinctive textures for each class. Shapes extraction were selected for FMS images since they have very sharp fractures where data is available, and at the same time missing pixels can be ignored in this case. UBI images are more suitable for texture extractions due to smudging of shape features.

Shapes were detected using 8-connected objects (blobs) and labelled for each FMS binary image (Haralic and Shapiro, 1992). Then the properties of these objects were measured such as 'Orientation', 'MinorAxisLength', 'MajorAxisLength', 'EulerNumber', 'EquivDiameter' and 'Eccentricity'. The shape feature set for every FMS image consisted of number of blobs detected in the image, and [5, 25, 50, 75, 95] percentiles for each property mentioned above.

Texture features extracted from UBI images are Entropy, Contrast, Correlation, Energy and Homogeneity, since they can help to identified the important distinctive properties of different classes such as presence and number of fractures/contacts, uniformity, variance/inertia and homogeneity. Contrast, Correlation, Energy and Homogeneity are calculated using gray-level co-occurrence matrix

$$Homogeneity = \sum_{i,j} \frac{p(i,j)}{1+|i-j|}, \quad Energy = \sum_{i,j} p(i,j)^2, \quad Contrast = \sum_{i,j} |i-j|^2 \, p(i,j) \qquad (1)$$

2.4 Rock Classification Framework

The overlap of electrofacies (classes) required a two-stage approach for automatic rock classification. It involves: 1) Identification of the most distinct electrofacies with high core/wireline coverage recovery rate (used for validation) and application of the classification algorithms and; 2) Application of the same classification algorithms to all nine classes. The latter will show how performance of the classification algorithms changes with decreasing recovery rate and overlapping classes.

Three classes with high core recovery rates were identified: 1) Massive off-axis ponded lava flow (MOPL) with core recovery 88-100%, homogenous texture, few fractures and sharp contacts of the lava flows; 2) Fragmented flows (FF) with highly fractured rock displaying a variety of orientations and intersections; 3) Dikes (VD) with core recoveries of 57-110% (more than 100% recovery not uncommon in hard rock coring due to incomplete recovery of the previously drilled interval), regularly spaced steeply dipping features and homogenous textures away from contacts. Contacts may be fragmented, with present fracturing. These are sub-vertical to steeply dipping and sub-parallel.

Additionally, we proposed two automatic rock classification methods: histogram-based and feature-based. The histogram-based method takes into consideration the distribution of parameters within each class. Histograms are calculated for both FMS and UBI images as follows: each image is summed up along depth dimension and then the distribution of summations is estimated by fitting histogram with n number of bins. In case of FMS images, the summation is done only over non-missing pixels. The features described above are used in feature-based method. Furthermore, the 1D wireline logging data is added to either distributions or features of the FMS and UBI images. Any generic geological constraints applicable to Hole 1256D were used to update the initial class predictions.

Different classifiers were applied to achieve automatic identification of electrofacies such as Random Forest (RF), Support Vector Machines (SVM), Linear Discriminant Analysis (LDA), Echo State Networks (ESN) and Decision Trees (DT). They have been purposely selected because of their robustness in

coping with overlapping classes, missing data and unbalanced data such as in the case of this study. The Decision Trees algorithm proved to be faster and produce the best classification results.

3 RESULTS

Both feature extraction and classification were done using Matlab on PC with Intel(R) Core™ i7-3770CPU@3.40GHz and 16 GB RAM. For classification purposes, both FMS and UBI images were divided into 50cm sliding windows with 30cm step, i.e. each image size is 100 (2 samples per 1 cm) by 750 (circumference of the Hole) pixels. Histograms with 20 bins were calculated for both types of images, and centre of each bin together with number of pixels inside of each bin are fed to classifiers (80 all together) for histogram-based approach. 31 shape (blob) and 49 texture features were extracted from FMS and UBI images respectively, and fed to classifiers for feature-based approach. 1D wireline logging data were also fed to classifiers for both histogram- and feature-based approaches adding extra 500 features, i.e. 580 features were used both for histogram-based and feature-based approaches.

The *generic geological constraints*, applicable for Hole1256D, are 1) MOPL is only present above 350 mbsf; 2) VD are only present below 1050 mbsf. These geological constraints were applied to classification results as post-processing filter. For each electrofacies analysis, the data were divided into training and testing sets. The training set consisted of the most representative samples for a given class due to difficulties to obtain reliable labels for all samples by observation. Results for both the histogram- and feature-based approaches are presented.

3.1 Classification results for 3 Classes

The training set consisted of the following intervals:

1. MOPL: 300-325 mbsf (83 images of each type)
2. FF: 406.78-416 and 488-493.43 mbsf (48 images of each type)
3. VD: 1130-1147 and 1210-1255 mbsf (207 images of each type)

The remaining intervals sampled for each of these classes were used for testing, i.e. 77, 800 and 954 images respectively. Overall, training set had 338 records, while testing set had 1831 records.

At first performance of classification algorithms such as DT, RF, SVM, LDA and ESN were compared using histogram-based approach. Each algorithm was tuned using grid-based approach to obtain the best results and functions provided in Matlab except for ESN. DT was trained using function fitctree which fits a binary classification tree for multiclass classification with the following parameters: 'MinParentSize'=10, 'MinLeafSize'=6 and 'NumVariablesToSample'=17. TreeBagger function was used to fit RF with 10 trees, with 25% of observations to be randomly selected with replacement for each bootstrap replica and computing variable importance by permuting observations. SVM were trained using function fitcecoc which fits multiclass models using one-versus-one coding design and automatically finds optimal parameters. LDA were trained using fitcdiscr function with 'SaveMemory' option 'on'. ESN was a plane ESN with 300 internal states, spectral radius of 0.45, input and teacher scaling of 1 and 0 for the remaining parameters.

Figure 2 shows correct classification rates (CCR) for all algorithms and 3 classes during testing stage. Though all algorithm achieved 100% CCR for FF class, MOPL class was the most challenging for SVM, LDA and ESN, while VD class was for RF. The best results for each class and overall ('All') were achieved using DT. Additionally DT was the fastest algorithm requiring 4.86s to run, while it took 5.97s for RF, 14.31s for SVM, 6.53s for LDA and 6.98s for ESN to run. Note that DT produces better results than RF in this case for MOPL and VD classes. It could be due to several reasons such as small sample size, discontinuous nature of the data, overfitting of the data. Therefore, further results will be reported only for DT algorithm due to space restriction.

The comparison of histogram- and feature-based approaches using DT for 3 class problem is given in Figure 3. Both approaches classified 100% of FF samples correctly. Histogram-based approached performed better in this case classifying 100% of MOPL samples correctly in comparison to 93.5% by feature-based approach, and 99% of VD samples correctly in comparison to 97% by feature-based approach. Next we will investigate 9 class problem using DT.

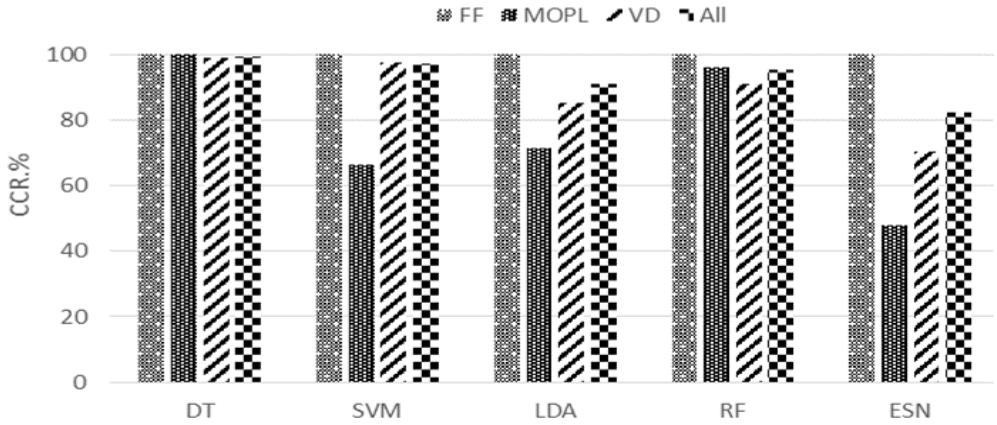

Figure 2. Comparison of different classification algorithms for 3 class problem

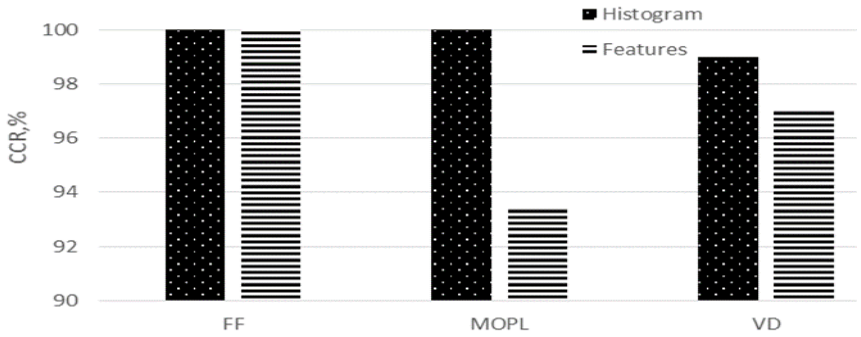

Figure 3. Comparison of histogram- and feature-based approaches for 3 classes

3.2 Classification results for 9 Classes

The intervals used for training the remaining 6 classes were as follows:

- BR: 348.02-351.98, 517-523.66 and 530.57-535.39 mbsf (51 from 478 images)
- FMAS: 652.21-655.32 and 659.84-664-57 mbsf (26 from 190 images)
- ID: 817-51-823.21 and 966.46-969.77 mbsf (30 from 318 images)
- MAS: 439-451.66 and 482.96-488 mbsf (59 from 360 images)
- P: 750.39-754.61 and 700.61-703.02 mbsf (22 from 51 images)
- TP: 706.66-708.28 and 737.9-739.17 mbsf (9 from 125 images)

Overall, training set consisted of 536 records, while testing set had 3165 records. DT was tuned using grid-based approach and trained using parameters: 'MinParentSize'=13, 'MinLeafSize'=9 and 'NumVariablesToSample'=96. The classification results did not achieve high accuracy as with the 3 classes case (Table 2). In this case the average CCR was poor (46.4%) after applying geological constraints to the results produced by the DT classifier. The best recognizable class is VD with ~75% of samples classified correctly. The DT classifier was able to achieve CCRs above 50% for BR and ID classes; and above 40% for MAS and MOPL classes. P and TP were scored poor classification accuracies. P was often misclassified either as BR (more than 40% samples) or MAS (~20% of samples). TP was misclassified as BR, FF and ID classes. 60% of samples belonging to FF class were misclassified as a BR class.

Classification results for 9 classes using the feature-based approach are shown in Table 3. Overall CCR was 49.9% and the VD class remained the best recognized class with CCR of 82.9%, following by MOPL (68.8%), BR (62.5%) and MAS (56.8%). The class with the lowest CCR is P at 3.4%. A comparison of the histogram and feature-based approaches indicate that the improvement of classification accuracy of one class usually occurs at the expense of reducing classification accuracy for other classes (Figure 4). For example, using the features approach improved CCR for such classes as BR, MAS, MOPL, TP and VD, but CCRs were reduced for such classes as FF, FMAS, ID and P. The VD class is the best recognised class irrespective of the approach used. The most challenging

classes are P and TP, since they have similar features and are not easy to distinguish manually. Although the overall degree of accuracy is reduced with a 9 class problem, the misclassifications in the confusion matrix are restricted to comparable electrofacies for the more fractured lithologies (BR, FF, FMAS). This could potentially allow a "coarse" estimate on the relative proportion of "highly fractured" lithologies which are in general, poorly recovered during coring operations.

Table 2. Confusion matrix for 9 classes for histogram-based approach.

%		***Predicted electrofacies***								
		BR	**FF**	**FMAS**	**ID**	**MAS**	**MOPL**	**P**	**TP**	**VD**
Actual electrofacies	**BR**	**57.6**	23.7	0.2	9.1	0	0	8.0	1.4	0
	FF	60.7	**17.3**	0.6	7.2	2.3	0	8.7	3.1	0
	FMAS	17.1	18.3	**20.1**	18.9	15.9	0	8.5	1.2	0
	ID	9.4	19.1	2.8	**52.8**	14.2	0	1.4	0	0
	MAS	17.9	14.6	1.0	19.7	**42.59**	0	4.6	0	0
	MOPL	3.9	11.7	0	3.9	37.7	**42.9**	0	0	0
	P	44.8	10.3	0	10.3	20.7	0	**13.8**	0	0
	TP	30.2	21.6	1.7	16.4	4.3	0	12.9	**12.9**	0
	VD	9.7	3.5	0.6	8.4	1.6	0	0.8	0	**75.4**

Table 3. Confusion matrix for 9 classes for feature-based approach.

%		***Predicted electrofacies***								
		BR	**FF**	**FMAS**	**ID**	**MAS**	**MOPL**	**P**	**TP**	**VD**
Actual electrofacies	**BR**	**62.5**	17.1	0.5	10.1	0	0	5.8	1.2	2.8
	FF	44.6	**22.7**	2.8	2.3	4.3	0	8.4	14.7	0
	FMAS	7.9	35.4	**17.7**	9.1	16.5	0	3.0	10.4	0
	ID	25.7	21.9	2.4	**20.1**	28.5	0	0.7	1.7	0
	MAS	3.7	21.9	5.3	9.6	**56.8**	0	1.7	1.0	0
	MOPL	0	3.9	0	0	27.3	**68.8**	0	0	0
	P	24.1	17.2	3.4	10.3	17.2	0	**3.4**	24.1	0
	TP	10.3	37.9	7.8	2.6	11.2	0	7.8	**22.4**	0
	VD	0.2	3.8	0.2	5.8	7.12	0	0	0	**82.9**

4 CONCLUSIONS

An automatic quantitative identification of electofacies based on Decision Trees is proposed for wireline geophysical logs from Ocean Drilling Program Hole 1256D. The big data are quite complex and challenging due to large gaps and vertical shifts in FMS images, horizontal misalignments in UBI images, and strongly overlapped classes. However, good classification accuracy (above 90%) for the 3 classes' problem was achieved. In the case of the 9-classes' problem, classification accuracy above 60% were obtained for some classes. The proposed approach requires only a small labelled representative sample for each class and does not need expert knowledge for tuning classification and feature selection algorithms. However, features suggested in this approach are not discriminative enough for some intertwined classes, and further work is required to identify such features. Also, deep learning algorithms such as Convolutional Neural Networks need to be tested. They are specifically designed for imagery data and exploit the strong spatially local correlation that could be embedded in images. Indeed, these will be investigated in our future studies within our big data science programme.

ACKNOWLEDGMENTS

ZAS and GV were supported by the Southampton Marine and Maritime Institute, University of Southampton, HEIF Research Collaboration Stimulus Fund (2014/15). DAHT and MH were supported by NER/T/S/2003/00048, NE/E001971/1 to DAHT, NERC studentships to MH (NER/S/A/2005/13475A), post-cruise funding to MH (NE/L00059/1; IODP Exp 335), DAHT acknowledges support of a Royal Society Wolfson Research Merit Award (WM130051). Also, the authors acknowledge that this research used data samples from the Ocean Drilling Program and also the Integrated Ocean Drilling Program.

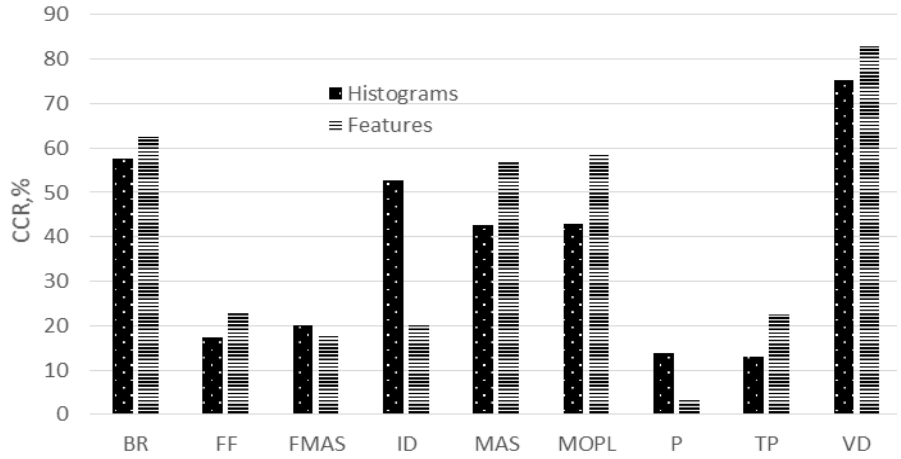

Figure 4. Comparison of DT performance for histogram-based and feature-based approaches.

REFERENCES

Haralick, R. M, Shapiro L.G., 1992, Computer and Robot Vision, Volume I, Addison-Wesley, pp. 28-48

Harris, M., Coggon, R.M., Smith-Duque, C.E., Cooper, M.J., Milton, J.A., Teagle, D.A.H., 2015. Sr isotopes as tracers for channeled hydrothermal fluid flow during the accretion and evolution of fast spread ocean crust: ODP Hole 1256D. Earth Planet. Sci. Lett. 416,56-66.

Harris, M., Coggon, R.M., Wood, M., Smith-Duque, C.E., Henstock, T.J., Teagle, D.A.H., 2017. Hydrothermal cooling of the ocean crust: Insights from ODP Hole 1256D. Earth Planet. Sci. Letts. 462,110–121. doi: 10.1016/j.epsl.2017.01.010

Lobos R., Silva J.F., Ortiz J.M., Diaz C., Egaba A., 2016. Analysis and Classification of Natural Rock Textures based on New Transform-based Features. Math Geosci., 48, 835-870

Niekum S., 2005. Reliable rock detection and classification for autonomous science. Master's thesis, Carnegie Mellon University, December 2005.

Teagle D., Alt J., Umino S., Miyashita S., Banerjee N., Wilson D. and the Expedition 309/312 Scientists, 2006. Expedition 309/3012 summary. Proceedings of the Integrated Ocean Drilling Program, 309/312, 1-27. doi: 10.2204/iodp.proc.309312.2006

Teagle, D.A.H., Ildefonse, B., Blum, P., and the Expedition 335 Scientists, 2012. *Proc. IODP,* 335: Tokyo (Integrated Ocean Drilling Program Management International, Inc.)

Thomas A, Rider M, Curtis A, MacArthur A., 2011. Automated lithology extraction from core photographs. First Break, 29 (6), 103–109

Tominaga M, Teagle. D.A.H., Alt J.C., Umino S., 2009. Determination of the volcanostratigraphy of oceanic crust formed at superfast spreading ridge: Electrofacies analysis of ODP/IODP Hole 1256D. Geochemistry, Geophysics, Geosystems, 10 (1), Q01003

Tominaga M., 2013. Imaging" the cross section of oceanic lithosphere: The development and future of electrical microresistivity logging through scientific ocean drilling. Tectonophysics, 608, 84-96.

G. V. Veres and Z. A. Sabeur (2013). Automated operational states detection for drilling systems control in critical conditions. 21st European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning, 24th-26th April 2013, Bruges, Belgium.

G. V. Veres and Z. A. Sabeur (2015). Data Analytics for Operational Drilling States Classifications. 23rd European Symposium on Artificial Neural Networks, In Computational Intelligence and Machine Learning, 22nd -24th April 2015, Bruges, Belgium. ISBN 978-287587014-8.

Wilson D. S., Teagle D. A. H., and 57 others (2006) Drilling to gabbro in intact ocean crust. Science 312, 1016-1020. DOI: 10.1126/science.1126090

Yin X.-C., Liu Q., Hao H.-W., Wang Z.-B., Huang K., 2011. FMI image based rock structure classification using classifier combination. Neural Comput. Appl., 20 (7), 955–963