

A framework for an AI pipeline for borehole data

John M. Aiken^{1,2,3}, Elliot Dufornet¹, Hamed Amiri³, Lotta Ternieten^{3,4}, Oliver Plümper³

¹Njord Centre, Departments of Physics and Geosciences, University of Oslo, PO BOX 1048, Blindern, Oslo, 0316, Oslo, Norway

²Expert Analytics, Oslo, Norway

³Department of Earth Sciences, Utrecht University, Utrecht, The Netherlands

⁴Department of Ocean Systems, Royal Netherlands Institute for Sea Research, The Netherlands

Key Points:

- We demonstrate an AI pipeline for ingesting data from the Oman Drilling Project Multi-borehole Observatory to predict peridotite alteration
- A large language model (ChatGPT) is able to summarize visual core descriptions, providing keywords that can be used in regression models
- Fractures are less predictive than other features for classifying highly altered (> 90%) peridotites.

Corresponding author: John M. Aiken, john@xal.no

Abstract

16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
Researchers analyzing data collected from borehole drilling projects can face dozens of terabytes of seismic, hydrologic, geologic, and rock mechanics data, including complex imagery, physical measurements, and expert-written reports. These diverse data sets play a pivotal role in understanding solid Earth processes. Ingesting and analyzing such data presents a colossal challenge that typically demands a team of experts and large amounts of time. The utilization of Artificial Intelligence (AI) and machine learning emerges as a compelling approach to help tackle the volume and complexity of drilling data. This paper presents an AI-based pipeline for ingesting data from the Oman Drilling Project’s Multi-borehole Observatory. The study focuses on the alteration of peridotite core segments taken from Borehole BA1B, utilizing a catboost classification model trained on an integrated data set of machine learning segmented core images, physical measurements, geological, lithographic data, and AI-summarized expert texts and feature selection. This paper’s central objective is to establish a repeatable, efficient pattern for processing such multifaceted borehole data through connecting fracture networks detected in the borehole BA1B imagery to the host rock alteration.

Plain Language Summary

32
33
34
35
36
37
38
39
40
41
42
43
Scientists studying the Earth using data from drilling into the ground often deal with huge amounts of information. This can include everything from seismic waves, water measurements, rock types, and complex images to detailed expert reports. Understanding this data is crucial for learning about the Earth’s processes. However, sorting through and making sense of it takes much work and requires a team of experts. This is where Artificial Intelligence (AI) and machine learning come in handy. They can help manage and understand these large and varied sets of data. This research focuses on data from the Oman Drilling Project, where scientists wanted to know how rocks in Oman change so they can be used to store CO₂. To answer this question, we trained several different AI models to analyze different kinds of data, including pictures and reports written by other scientists.

1 Introduction

44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
Ocean and continental drilling projects typically produce dozens of terabytes of data, including seismic, hydrological, geological, and rock mechanics data. These data are multi-modal and multi-source including imagery such as core photos or X-ray computed tomography scans, physical measurements such as resistivity, porosity, and permeability measurements, and expert data such as written visual core descriptions. The collection of these data is driven by scientific knowledge and theory. Given the volume and interdisciplinary scope of these data, analyzing them is a monumental task requiring many years of continuous work for a team of individual experts. Thus, there is a current strong need in the solid-Earth sciences for computational models and frameworks that ingest and interact with multi-modal, multi-source data and aid researchers in hypothesis testing (Goss, 2020; H. Wang et al., 2023). Artificial Intelligence (AI) and machine learning offer an attractive solution to this complex problem. New AI tools can produce more accurate simulations of multi-phasic fluid flow (Y. D. Wang et al., 2021) and Large Language Models (e.g., ChatGPT) can be used to summarize expert written drilling reports (Zhao et al., 2023). AI can aid scientists in going beyond simply ingesting and manipulating data and help generate scientific hypotheses from complex data (Schmidt & Lipsen, 2009; Guimerà et al., 2020; Li et al., 2022; H. Wang et al., 2023; Cornelio et al., 2023). This paper presents a framework for an AI pipeline that ingests multi-modal data (images and expert-written text) taken from the Oman Drilling Project (OmanDP).

64
65
The OmanDP multi-borehole observatory (MBO) is an example of a large-scale, interdisciplinary continental drilling project that has produced a multi-modal dataset.

66 At site BA1, borehole BA1B was cored and images including complete wrap-around scans,
67 physical, chemical, and biological measurements (e.g., mean dry electrical resistivity, cell
68 abundance), and lithographic information were recorded. Initial results of the OmanDP
69 demonstrate that in borehole BA1B, between 65 and 100% of the peridotite has been
70 hydrated to form serpentinite and related rock types. The decrease in the extent of peri-
71 dotite alteration with depth may suggest that significant peridotite alteration in the re-
72 gion has been relatively young, within the last 50,000 years (Kelemen et al., 2021). H₂
73 and CH₄ outgassing have previously been detected in the Oman boreholes and are possi-
74 ble products of ongoing peridotite alteration (Ellison et al., 2021; Aiken et al., 2022).

75 The chemical reactions associated with peridotite alteration are well understood
76 (e.g., (Kelemen & Matter, 2008; Plümpner & Matter, 2023)). Olivine and pyroxenes react
77 with water and carbon dioxide to form mainly serpentine minerals, brucite, iron ox-
78 ides, and carbonates. Low-temperature alteration (< 150° C) is possible and has been
79 observed in Oman and other on-land environments (de Obeso & Kelemen, 2020; Corre
80 et al., 2023). Redox reactions further produce H₂ and CH₄, which can be observed bub-
81 bling up continuously in alkaline springs found in peridotite-rich areas. The complete
82 conversion of peridotite to serpentinite is not fully understood because the associated
83 swelling should “armor” the reactive surfaces of the peridotite, thus preventing water
84 from continuing to interact with unaltered rock (Hövelmann et al., 2012; Malvoisin et
85 al., 2020, 2021). It is assumed that the volumetric expansion of the rock as a consequence
86 of the hydration would induce stress on the surrounding host rock, thus opening new path-
87 ways to unaltered rock penetrating the “armor”. This process, known as “reaction-driven
88 cracking” (Kelemen & Matter, 2008; Jamtveit et al., 2009; Plümpner et al., 2012), is ex-
89 pected to create hierarchical fracture networks within the host rock (Jamtveit et al., 2009).
90 Thus, in addition to the geological attributes of the Oman peridotite, the density and
91 complexity of the fracture networks should be indicative of recent and/or ongoing peri-
92 dotite alteration (Iyer et al., 2008). Reaction driven cracking should develop a charac-
93 teristic hierarchical network pattern dominated by four-sided domains (Aydin & DeGraft,
94 1988; Iyer et al., 2008). These fractures should grow from older fractures, linking differ-
95 ent generations of fractures together. Thus, in an altered peridotite environment strongly
96 influenced by reaction driven cracking we expect to see a fracture network made up of
97 polygons with four or more sides and few single, linear fractures in regions of high al-
98 teration. Fractures in the OmanDP MBO cores have been qualitatively described through
99 visual core descriptions using classification rubrics developed for ocean drilling expedi-
100 tions (Blackman et al., 2006; MacLeod et al., 2017). These descriptions are insufficient
101 to describe the complexity of fracture networks which would be necessary to identify po-
102 tential regions of ongoing reaction driven cracking. To overcome the limitation of this
103 qualitative description, in this study, we use a machine learning-based image segmen-
104 tation model to identify fractures in the wrap-around core images. We then use statis-
105 tical microstructure descriptors (SMDs) to describe the fracture network complexity (Lu
106 & Torquato, 1992; P.-E. Chen et al., 2019; Amiri et al., 2023).

107 In this paper, we present a machine learning-oriented approach for treating multi-
108 modal data produced during the coring and subsequent investigations of OmanDP bore-
109 hole BA1B. This framework is designed to normalize these multi-modal data (in our case,
110 wrap around core images, physical measurements, and visual core descriptions) quickly.
111 Much of the work presented here typically would take many months of work to complete
112 compared to the computational workflow presented here. Specifically, we present two sep-
113 arate methods to ingest data from the OmanDP borehole BA1B: 1) we produced a ma-
114 chine learning image classifier for wrap-around core images that segmented fractures which
115 is then used to calculate fracture network characteristics, and 2) we utilized the large
116 language model ChatGPT to summarize handwritten visual core descriptions (VCDs)
117 from the coring expedition. The VCDs represent on-site expert knowledge about the ge-
118 ology of the cores and also, observations that could help explain the presence of highly
119 altered peridotites in the absence of complex fracture networks. They describe different

120 morphometric features such as the presence of veins, alteration, and oxidation, as well
 121 as structural features and mineralogy. They are open-ended, semi-structured text doc-
 122 uments written per core segment and thus make a depth-dependent, expert description
 123 of the BA1B core. The fracture network statistics and VCD keyword data are then com-
 124 piled into a single dataset along with physical measurements (e.g., mean dry electrical
 125 resistivity) which is then used to train a gradient boosted trees (catboost) classification
 126 model predicting alteration in the peridotite core (Prokhorenkova et al., 2018). This model
 127 is then used to find a geological explanation from the machine learning classification model
 128 for the alteration of the core segments. A central objective of this paper is to establish
 129 a repeatable pattern for processing this type of data, enabling even individuals without
 130 earth science knowledge to exploit it. Additionally, it is to explore the impact of non-
 131 tectonic fracturing of rock on peridotite alteration using machine learning methods.

132 2 Data and Methods

133 We utilize three types of data extracted from the OmanDP borehole BA1B: wrap-
 134 around images of the borehole core, physical, chemical, and biological measurements made
 135 after the coring, and textual data comprising geologists’ remarks regarding the drilled
 136 sections known as the “Visual Core Descriptions” (VCD). This data is processed (Fig-
 137 ure 1) through fracture labeling via image segmentation of the wrap-around core images,
 138 fracture density and network connectivity estimation from the labeled fracture images,
 139 and summarization into keywords of the VCD text using ChatGPT. These data are com-
 140 bined with physical measurement data to create a depth-dependent database of bore-
 141 hole BA1B. This database is then used to predict the detected alteration within the core,
 142 as reported from the expedition (Kelemen et al., 2021).

143 Below we provide a full pipeline description (Figure 1) including a site description
 144 for the OmanDP MBO, describing the wrap-around core image processing, the VCD text
 145 processing, and the regression models that are built from this analysis.

146 2.1 Site Description

147 The OmanDP borehole BA1B is part of a multi-borehole observatory (MBO) that
 148 was established during the second drilling phase of the OmanDP in the Wadi Tayin Mas-
 149 sif to address a spectrum of questions that connect the deep mantle and the ancient ocean
 150 floor with modern hydrology and ongoing biogeochemical reactions in the mountains and
 151 wadis of the Samail Ophiolite (Kelemen et al., 2021). The Wadi Tayin Massif is one of
 152 the southern massifs of the Oman ophiolite complex, which was formed primarily via a
 153 mid ocean ridge basalt like, single-stage process at a submarine spreading ridge (Godard
 154 et al., 2003). The Massif is characterized by an extensive mantle sequence consisting al-
 155 most entirely of harzburgite and minor lherzolite that host 5%–15% dunites and mul-
 156 tiple mafic intrusions and is overlain by a 5–7 km thick gabbroic crustal section, sheeted
 157 dikes, and pillow lavas (Boudier & Coleman, 1981; Pallister & Knight, 1981). Gravity
 158 anomalies (Ravaut et al., 1997) suggest that the Massif composed of 30%–60% (Falk &
 159 Kelemen, 2015; de Obeso & Kelemen, 2018) serpentinized mantle peridotite, extending
 160 up to 5 km below the present-day surface.

161 BA1B is one of three boreholes from the active alteration zone (BA) site, which
 162 targets alteration at temperatures < 50 °C. It is of specific interest because it is one of
 163 the boreholes instrumented with hydrophones (Aiken et al., 2022) which could provide
 164 direct evidence of ongoing seismic activity due to reaction driven cracking. The cores re-
 165 covered from BA1B consist of $\sim 55\%$ harzburgite, $\sim 35\%$ dunite, and $\sim 10\%$ mafic dykes
 166 and alluvium. Contacts between ultramafic and mafic domains are marked by chlorite,
 167 prehnite, talc, and hydrogrossular, indicating metasomatism on a millimeter scale. Carbonate-
 168 rich zones occur in the upper 150 m and are characterized by a distinct decrease in vein
 169 abundance with depth (Kelemen et al., 2021).

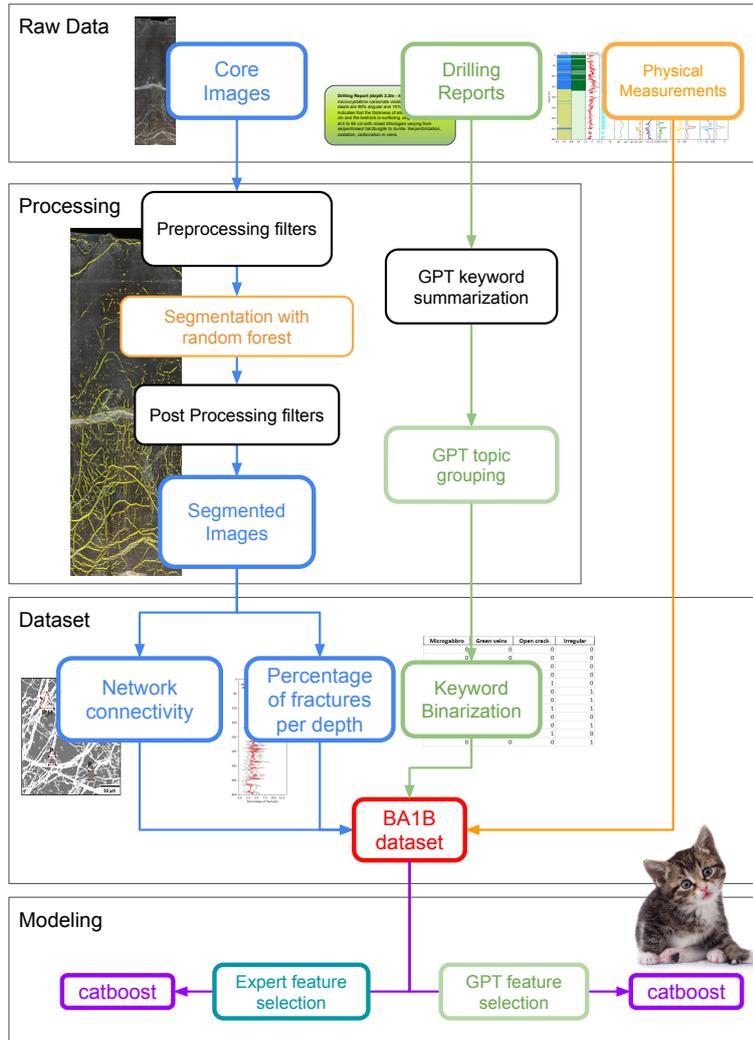


Figure 1. Pipeline utilizing AI and machine learning to ingest data taken from Oman Drilling Project Multi-borehole Observatory borehole BA1B. Ultimately this processes 505 wrap-around core images, 505 drilling reports per core segment, and 30 physical measurements into a data set of 96 columns ranging from 0m at the top of the borehole to the cored depth of 400m.

170

2.2 Core image analysis

171

172

173

174

175

176

177

178

179

180

181

The BA1B wrap around core images provide the primary images to identify fractures. Following the drilling process, the borehole is segmented into 505 equivalent sections, and were photographed. Each of the 505 core cut images typically measures 60 to 100 cm in length (75 cm average length) and 10 to 20 cm in width. The complete image set is then made up of 505 core segments of approximately 1 m in depth, spanning from the uppermost layer of the core to a depth of 400 m. Additionally, there are sections of core that were taken immediately for microbiological analysis, and not photographed, such that the entire data set includes 690 core sections. After applying pre-processing filters to ensure proper treatment, we use the Ilastik software (Berg et al., 2019) for segmentation and extraction of fractures (Section 2.2.1). Post-processing filters were subsequently implemented to enhance the accuracy of our segmentation. These segmented

182 images could then be used to calculate the percentage of fractures at various depths for
 183 each core image and estimate fracture network connectivity.

184 **2.2.1 Image segmentation**

185 We employ a multi-step process for successfully segmenting fractures/alteration prod-
 186 uct veins. Raw wrap-around core images are first pre-processed using Gaussian, Hessian,
 187 Roberts, and Sobel edge-enhancing filters. This flattens differences in color content of
 188 the image, and highlights abrupt changes in edges, making it ultimately easier to pick
 189 out fracture veins. Twenty images taken from 20 m segments distributed depth-wise along
 190 the borehole were then labeled using the Ilastik software (Berg et al., 2019). We then
 191 used the built-in random forest algorithm within Ilastik to label the remaining 485 im-
 192 ages. We drop all labeled pixel groupings with ≤ 50 pixels. We then apply a post-processing
 193 eccentricity filter to remove small round erroneously labeled pixel groupings as they are
 194 not physically representative of a fracture or vein network. This is then considered the
 195 final labeled fracture/vein network data set. In this study, we do not differentiate frac-
 196 tures closed by mineral precipitation (veins) from open fractures. This is because if we
 197 differentiated between these two, we would not capture the full network of fractures and
 198 would likely underestimate the network connectivity and complexity.

199 **2.2.2 Estimating fracture density**

200 The first essential piece of data to acquire is the degree of fracturing in the core
 201 at any given depth, enabling the establishment of a correlation between depth and the
 202 number of fractures. We calculate the percentage of fractures using the following rela-
 203 tionship:

$$F\% = \frac{I_{label}}{I_{area}} \quad (1)$$

204 Where I_{label} is the number of pixels labeled as a fracture in a wrap around core segment
 205 and I_{area} is the total number of pixels of a wrap around core segment. We calculated
 206 the fracture percentage using three distinct approaches to segmentation: raw segmen-
 207 tation, segmentation with an area filter, and segmentation incorporating an eccentric-
 208 ity filter. In the end, the variations in filters used have negligible impact on the results,
 209 as the curves share similar trends with a translation shift thus we choose to apply only
 210 the eccentricity filter to the data sets as it is most relevant to identifying small artifacts
 211 that are not fractures.

212 **2.2.3 Estimating Connectivity**

213 Fracture network connectivity is another property of the observed fracture network
 214 in the core images that can have an impact on the alteration process. Thus, it is nec-
 215 essary to quantify such connectivity so we can use it as an additional feature to our ma-
 216 chine learning model. Our approach involves utilizing n-point spatial correlation func-
 217 tions, i.e. SMDs (Lu & Torquato, 1992; P.-E. Chen et al., 2019; Amiri et al., 2023). These
 218 functions represent the probability of n random points separated by a distance r to lie
 219 in the same phase such as fractures. However, for $n \geq 3$ this probability calculation be-
 220 comes computationally challenging. To address this, we focus on a subset of these func-
 221 tions: n-point polytope functions. These functions are defined by the probability that
 222 the n vertices of a random regular n -point polytope with an edge length r will fall within
 223 the same phase (P.-E. Chen et al., 2019). Given that reaction driven fractures should
 224 produce network patterns that are most likely to have four-sided polygons (Iyer et al.,
 225 2008), the detection of these prevalence of such polygons will indicate the complexity
 226 of the existing fracture network. That is, if the fracture network is made up of mostly

227 longer linear segments and fewer polygons, it is less likely to present a hierarchical net-
 228 work generated from reaction driven cracking.

229 To specifically assess fracture connectivity, we compute the lineal-path L function (Lu
 230 & Torquato, 1992). This function measures the probability of a whole segment of a ran-
 231 dom line to lie within the fractures, providing an efficient means to evaluate the linear
 232 connectivity in complex fracture networks such as those found in serpentinites (Amiri
 233 et al., 2023). In our study, six correlation functions are calculated: S_2 for two-point cor-
 234 relation, P_{3H} for horizontal triangles, P_{3V} for vertical triangles, P_4 for squares, P_6 for
 235 hexagons, and L for the lineal-path function. Alongside these, we also compute normal-
 236 ized versions of these SMDs, termed “scaled autocovariance functions” (Jiao et al., 2007),
 237 altogether introducing 12 features representing geometrical patterns and linear connec-
 238 tivity, and ultimately the complexity, of the fracture network within the BA1B core seg-
 239 ments.

240 In our analysis, the SMDs were computed within 1000x1000 pixel windows (one
 241 pixel is 0.2mm x 0.2mm) extracted from all core images. In each core image segment,
 242 a calculated SMD presents a probability curve (Figure 3) for that particular type of poly-
 243 gon to be present for the specified distance r ($r=1$ is a single pixel). To reduce these curves
 244 to data that can be utilized in the catboost model, we utilize the sum of the values for
 245 each SMD at edge length $r < 50$ pixels (<10 mm) as input data for our model.

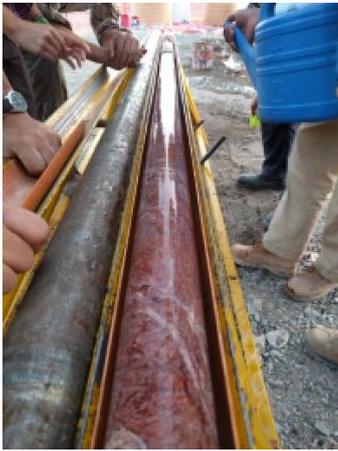
246 **2.3 Hand Written Expert Visual Core Descriptions**

247 After successful, drilling the recovered cores are processed and described during core
 248 description campaign following a protocol created by reviewing and adapting procedures
 249 of previous scientific ocean drilling expeditions (Blackman et al., 2006; Teagle et al., 2006;
 250 D. Teagle et al., 2012; Gillis et al., 2014; MacLeod et al., 2017). The protocol contains
 251 the optical description of the cores and various scientific analyses. Multiple teams per-
 252 form the core characterization, each focusing on specific aspects. The teams are as fol-
 253 lows: igneous petrology, alteration/metamorphic petrology, structural geology, geochem-
 254 istry, paleomagnetism, physical properties, near-visible infrared scanning, microbiology,
 255 and wireline geophysical logging and hydrogeological testing. At the end of the campaign,
 256 visual core descriptions (VCDs) are produced, which are section-by-section summaries
 257 of the core description observables and most pertinent instrumental measurement pa-
 258 rameters of the recovered cores.

259 To ensure consistency throughout the cores, especially during the optical core de-
 260 scription, each team member was responsible for observing a specific set of character-
 261 istics; however, an entire team would work together for initial descriptions (e.g., units
 262 and lithologies, critical features) to guarantee continuity. The terminology and abbrev-
 263 viations during description and classification were adapted from previous expeditions (Blackman
 264 et al., 2006; Whitney & Evans, 2010; Früh-Green et al., n.d.; MacLeod et al., 2017).

265 **2.3.1 ChatGPT for Drilling Reports**

266 Recent efforts in the development of large language models (LLMs) have caused
 267 a paradigm shift in the availability of easy-to-use text summarizing tools (Zhao et al.,
 268 2023). We choose to use ChatGPT due to its ease of use Application Programming In-
 269 terface (API). Other LLMs likely offer similar utility. When compared to the effort and
 270 lack of utility of traditional text analysis methods (traditional natural language process-
 271 ing methods such as Lemmatization (Miller, 1995), Frequency distributions and collo-
 272 cations (Gledhill, 2000), and TextRank (Barrios et al., 2016) did not produce valid key-
 273 words) new LLMs provide a new way forward for accessing high density, hard to quan-
 274 tify text data.



Visual Core Description (depth 3.2m - 4.0m): microcrystalline carbonate visible on vein surfaces. clasts are 90% angular and 10% rounded. This indicates that the thickness of alluvium is < few 10s cm and the bedrock is surfacing. angular fragments at 0 to 60 cm with mixed lithologies varying from serpentinised harzburgite to dunite. Serpentinization, oxidation, carbonation in veins

Figure 2. Field researchers cataloging the visual core description (left). An example of the visual core description text that is summarized into keywords using ChatGPT (right).

275 The handwritten VCDs (Section 2.3) were given to ChatGPT to summarize. Each
276 set of remarks per depth unit (505 in total) was given to ChatGPT (gpt-turbo-3.5) with
277 the prompt:

278 *“Please summarize the following text into ten keywords and explain why you picked*
279 *each keyword. The text to summarize is: {text}”*,

280 with {text} being replaced by the geologist’s remarks. This produced hundreds of dif-
281 ferent keywords that emerged from the process, many of which were close duplicates or
282 similar keywords. These keywords were then condensed for duplicates and/or similar-
283 ities (e.g., “vein” versus “veins”). Keywords that were reported by ChatGPT less than
284 50 times (representing less than 10% of the BA1B total cored depth) were removed. Ul-
285 timately, 52 keywords remained. Those keywords were integrated in the dataset as bi-
286 nary variables for each core segment. Then, we asked ChatGPT to group the different
287 keywords into topics based on the type of information they convey. We plotted the graph
288 of keywords depending on depth to have preliminary information about the keyword fea-
289 tures of the core (Figure 5).

290 2.4 Data set

291 Ultimately this process produced a depth-dependent data set composed of 690 rows
292 corresponding to 690 different sections of the core and involved image analysis fracture-
293 related data, textual data from geologists’ reports, geological data, and physical mea-
294 surements. In total, the dataset is comprised of 96 features: 13 of them derived from core
295 image segmentation of fractures, 51 from the extracted keywords from VCDs, 30 from
296 direct physical measurements, plus the depth and the alteration (we wish to predict the
297 alteration). For each of the 690 core segments that include: depth range (in meters), per-
298 centage of alteration ($\geq 90\%$), keywords picked by ChatGPT, fracture density estimate,
299 fracture network connectivity estimates, and physical measurements including precise
300 mineral composition, electrical resistivity, magnetic susceptibility, cell abundance, and
301 trace of volatile elements. Not all rows are assigned to an image because the image re-
302 port provides only 505 images of the borehole. The physical measurements are only avail-
303 able for a limited number of sections throughout the borehole (typically only every 20m
304 of core sections, please see the dataset for more details). We extrapolated missing data
305 by imputing the sample without physical data using the values from the samples above.
306 This data set was then used in a catboost model to predict per core segment if the per-
307 centage of alteration was above or below 90%.

2.5 catboost Classifier

Using the data set created via the above pipeline, we analyze this data using catboost to predict, per core segment, if the percentage of alteration is above or below 90%. Catboost is an open-source library designed to implement machine learning model based on the Gradient Boosting technique (Friedman, 2002; Prokhorenkova et al., 2018) that has been used across Earth sciences to solve problems such as fracture development (McBeck et al., 2020), climatic and metamorphic effects on glacier instabilities (Bouchayer et al., 2022), and understanding stick-slip motion (Hulbert et al., 2019). Catboost builds sequential decision tree models where each tree is trained on the residuals of the previous model using data that is out of the sample of the previous model, effectively improving the model’s accuracy with each step. We use 1000 boosting iterations with a learning rate of 0.1 with trees having a maximum depth of 3.

To evaluate this model, we first analyze its accuracy using the area under (AUC) the receiver operator characteristic curve (ROC) (Hastie et al., 2009). The ROC is the ratio of the True positive rate to the false positive rate for different decision thresholds for the classification model. The AUC is the area under this curve. An AUC value of 0.5 indicates that the model is no better than random chance because it is no more likely that the a true positive will occur than a false positive, while an AUC of 1.0 indicates a perfect model. By modulating the different features given to the model (by topic of feature), we can estimate which class of features are predictive of alteration: fracture network estimates, physical estimates, geologists remarks, and fracture network data and geologist remarks together.

2.5.1 ChatGPT for Automated Feature Selection

Automated feature selection has long been a staple of machine learning and is integrated in a variety of methods (Zou & Hastie, 2005; X.-w. Chen & Jeong, 2007; Hastie et al., 2009; Sharma et al., 2021). These methods often used a combination of model variance and complexity to determine which features to eliminate from a training data set. For example, recursive feature elimination (X.-w. Chen & Jeong, 2007) removes features one by one from the most important feature to the least, re-ordering features after each removal, to determine the subset of features that can be kept in the final model. The best model then chosen through recursive feature elimination typically minimizes the model variance, i.e., maintains the highest amount of fittedness through fit statistics such as R^2 , and minimizes complexity by removing training variables that have little to no impact on the model output. These methods do not consider the conceptual constraints a model may have placed on it by scientists. For example, in this study, we are interested in the impact of fracture networks on alteration. Thus, instead of using these methods, we rely on ChatGPT to provide expert model feature groupings as an automated feature selection tool.

We asked ChatGPT to classify all model features in the BA1B data set into groups that we could use to separate for model comparison analysis (see Section 2.5). We excluded the text summarization features already classified since ChatGPT had already seen these. We gave ChatGPT the prompt:

You are an expert physicist, chemist, biologist, and computational scientist ai helper bot. I will give you a list of columns for a catboost classifier. These columns are the features in the model. The catboost classifier is designed to determine whether a section of a borehole core has greater than 90% peridotite alteration or not. We are attempting to measure the impact of fractures in the sample against other features that impact the total alteration assuming this is related to reaction driven cracking.

You are to first:

1. *define each column*
2. *provide an overarching category for the column*
3. *describe why you picked this category for the column*

Features provided to you should be grouped into categories. Please reply saying you understand the task and then I will give you the column names.

ChatGPT replied categorizing all of the features into groups. These categorizations are then used for model feature selection. We compare these AI-selected feature groupings to the expert-selected feature groupings (Figure 6).

3 Results

We first report the results of the fracture network detection and ChatGPT’s summarization of expert geologist remarks (examination of the physical results can be found in (Kelemen et al., 2021)). Then we will describe the results of the catboost model.

Throughout the top section mainly composed of dunite (0-160 m), a moderate amount of fractures is detected (Figure 4), from 3% at the very top to 6% near the transition area. There appears to be a slight linear trend increasing in fractures between ≈ 25 m and 100m. This trend decreases in the location at approximately 90m where there is a 5-m section of less altered rock. A peak in fractures is detected near the transition zone between dunite and harzburgite rocks (160-180 m), with up to 10% of the image fractured. Fewer fractures are observed in the bottom harzburgite section (180-400 m).

Figure 3 shows an example of quantifying geometrical patterns and linear connectivity using the SMDs. All the SMDs start from the same probability, approximately 0.4, at $r=0$. This probability indicates the phase fraction (aka fracture fraction) as it measures the probability of only one point occurring in the same phase. The r at which S_2 stabilizes ($\approx 10 - 15$ pixels) gives a rough estimation of average fracture width to be $\approx 2-3$ mm. Moreover, the lineal-path curve consistently shows higher values compared to other polytope functions, suggesting that linear connectivity is a predominant pattern in these images. That is, there is low network complexity across most regions of the borehole.

The ChatGPT summarization analysis had two steps, first was the keyword analysis, and second was the topic analysis of the selected keywords (see Figure 5). Some keywords appear prolifically across the entire depth cross-section (e.g., Serpentine veins, Black Serpentinization, Gabbro). Others have a clear depth dependence either occurring in the upper Dunite sequence (e.g., Irregular, Lineation, Open cracks, Alteration halo) or in the lower Harzburgite sequence (e.g., Hydrothermal, Shearing, Magmatic veins). These keywords generally appear where we would expect them to when referencing the full-text reports. ChatGPT was also able to group keywords into meaningful topics of “veins and alteration”, “oxidation and alteration”, “structural features”, “rock type”, “mineralogy”, and “physical characteristics”.

Additionally, we had ChatGPT categorize the other features that were available to the catboost model using the prompt given in Section 2.3.1. ChatGPT replied with features similar to the expert-chosen features (Figure 6) producing the same result, that fracture features are much less predictive than other features collected about the peridotite alteration. This creates two classes of catboost model (expert-guided and GPT-guided).

The expert-guided catboost model is quite performant when using all data ($AUC=0.99$). When split by expert-guided topic the models (Figure 6, Table 3) perform well for all groups of features except for fracture-related features ($AUC_{fractures}=0.74$, $AUC_{mean}=0.93$). Similarly, the GPT-guided CatBoost models, using analogous feature groupings, demon-

Table 1. Area under the receiver operator characteristic curves (AUC) for different feature groups used in catboost model.

Feature Group	AUC
Expert Selected Feature Groups	
Chemistry and Biology	0.94
Fractures	0.74
Geology	0.98
Physics	0.90
ChatGPT Keywords	0.93
ChatGPT Selected Feature Groups	
Geological Composition	0.93
Physical Properties	0.92
Biological Influence	0.95
Fractures	0.74
Rock Type	0.92
Textural Features	0.89
Color and Visual Properties	0.92

405 strate comparable performance. ($AUC_{fractures}=0.74$, $AUC_{mean}=0.92$). An AUC above
 406 0.7 is considered to provide some discrimination while an AUC above 0.9 is considered
 407 to provide excellent discrimination (Hosmer Jr et al., 2013).

408 4 Discussion

409 This paper presents an AI-based pipeline for ingesting high-density, high-complexity
 410 disparate data sets into a single data set that can be used for analysis. This included core
 411 imagery, expert remarks (VCDs), and various physical, chemical, and biological measure-
 412 ments. These data were condensed into a single data set cataloging various features per
 413 depth increment of the ODP Multi-borehole Observatory borehole BA1B. A random for-
 414 est classifier was used to label fractures in the data which were then quantified using frac-
 415 ture network connectivity statistics. An LLM (ChatGPT) was employed to summarize
 416 VCD text and describe and group the dataset features into topical groups to compare
 417 during modeling. This analysis produced the following result: complex fracture networks
 418 do not appear in high-density arrangements that correlate with a high degree of peri-
 419 dotite alteration. Moreover, this process used free and open source tools to automate much
 420 of the workflow reducing the time it would take to identify and label fractures in pic-
 421 tures, identify relevant text in thousands of comments, and combine this information to-
 422 gether to visualize and then apply statistical analysis such as the catboost model pre-
 423 sented in this paper.

424 The effort by ChatGPT to categorize the VCDs into summarization via keywords
 425 represents an enormous amount of person hours worth of work. The summarization is
 426 not simply the segmentation of individual words, it is the conceptualization of the vi-
 427 sual core description into a summary that can be represented by descriptive keywords.
 428 In order to convert this data into keywords, at least two humans would need to first ran-
 429 domly select a subset of the data (i.e., the training data), agree on the keyword summa-
 430 rization database (i.e., a list of words that researchers agree should appear in the VCDs),
 431 then read a subset of VCDs until each researcher agrees with each other (typically de-
 432 termined by a statistic such as Cohen’s Kappa (Cohen, 2013)). This process would oc-

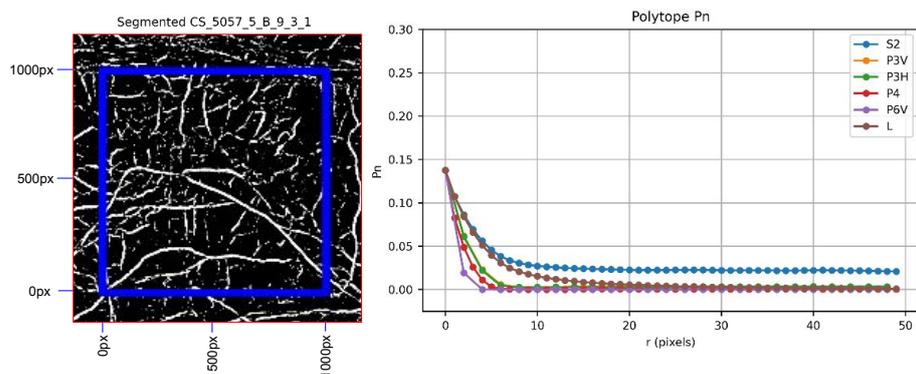


Figure 3. An example of fracture labeling and fracture network polygon identification. In the left image, segmented fractures are presented with white as a labeled fracture and black as a labeled host rock. The blue square represents the 1000px by 1000px selection taken from the larger core image. On the right, the polytope functions are calculated. Each curve represents the probability of finding a polygon of different shapes and number of sides in the entire core section (S2 for two-point correlation, P3H for horizontal triangles, P3V for vertical triangles, P4 for squares, P6 for hexagons, and L for the lineal-path function). As we can see in the left image, there are typically linear fractures that do not segment into hierarchical regions of larger fractures connected with smaller fracture spaces. This is typical for the entire borehole. This is confirmed in the curves to the right, it is far more likely to find linear fractures or two points than any other polygon shape.

433 cur iteratively until researchers felt there was no need for new keywords and little dis-
 434 agreement when searching the subset of training data. Then this keyword search would
 435 be applied to the full data set. This process would take at least days of fulltime work
 436 for a single borehole. In comparison, approximately ten lines of python codes were writ-
 437 ten to use ChatGPT to perform the keyword search (in addition to the prompt, Section
 438 2.3.1) and then the code run time took only a couple hours. Thus, this process allows
 439 for the processing and quantification of dense, expert derived data that would otherwise
 440 be time consuming to use.

441 For the exploitation of the textual data, relevant keywords were extracted from the
 442 geologists' visual core descriptions (VCD) using ChatGPT, and made into binary vari-
 443 ables for an easy use in our data set. Plotting the graphs of keywords per depth reveals
 444 internal correlation between keywords, and thus links subsets of properties with a given
 445 depth, showing a regressive model could effectively predict sample properties. Overall,
 446 ChatGPT was able to competently summarize expert knowledge. When comparing the
 447 summarized keywords to the original intent of the language written, we found that these
 448 were relevant to the original meaning within the text written in the drilling report. More-
 449 over, by utilizing ChatGPT, we were able to leverage expert knowledge by summariz-
 450 ing the drilling report within the statistical model (catboost) without needing an expert
 451 to convert said knowledge to a format that could be used by the statistical model. This
 452 has profound implications across much field-based science which includes a large collec-
 453 tion of written notes and remarks by experts. ChatGPT, or other LLMs, are not replace-
 454 ments for these experts, but they do provide a profound tool for converting long-form
 455 text data written by experts into data sets that can be ingested and compared using re-
 456 gression and classification techniques such as catboost with other physical, chemical, and/or
 457 biological data captured. It is likely that better prompt engineering and a more tightly
 458 coupled pipeline, general purpose LLMs would greatly extend the ability of the AI pipeline
 459 presented here (Ge et al., 2023; Lewis et al., 2020; Nori et al., 2023). This was not the

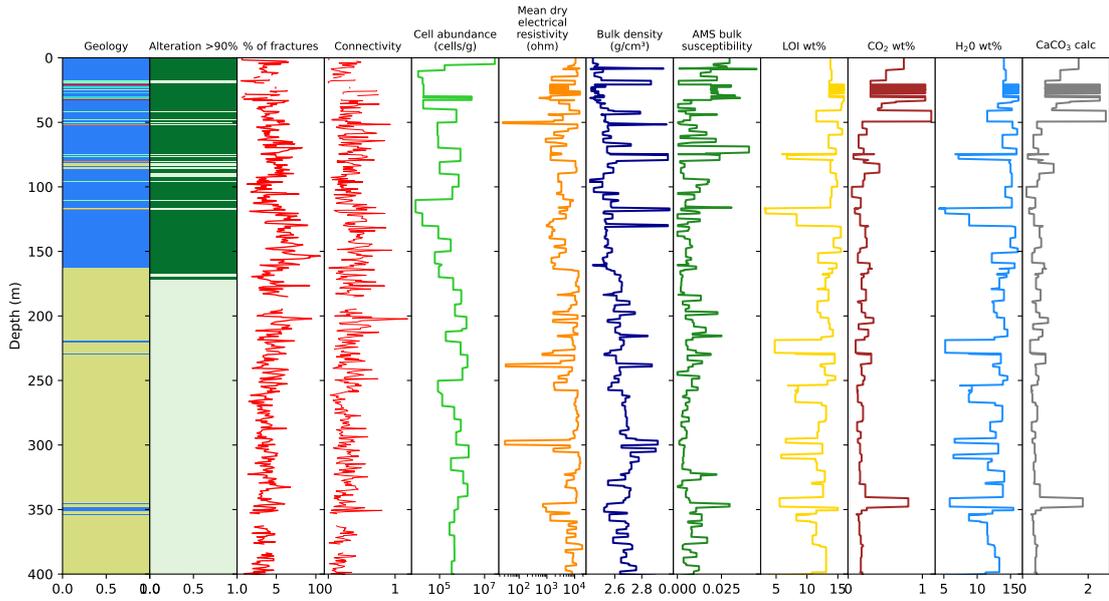


Figure 4. A subset of measurements taken in borehole BA1B. The % of fractures and connectivity are calculated from the random forest segmentation. Other quantities were measured on-site during or after the coring of borehole BA1B. These represent a cross-section of dataset features related to fracturing, physical and chemical attributes, and biology found in borehole BA1B. The geology column (furthest left) the colors represent the lithology with blue being dunite and yellow-green being harzburgite. For a complete description of the lithology please see (Kelemen et al., 2021). In the connectivity column, the red represents the lineal feature, the greater the value, the more likely there are to be connected fractures in this region. These data, along with other connectivity measures, and text keywords (Figure 5) are used to train the cat-boost model.

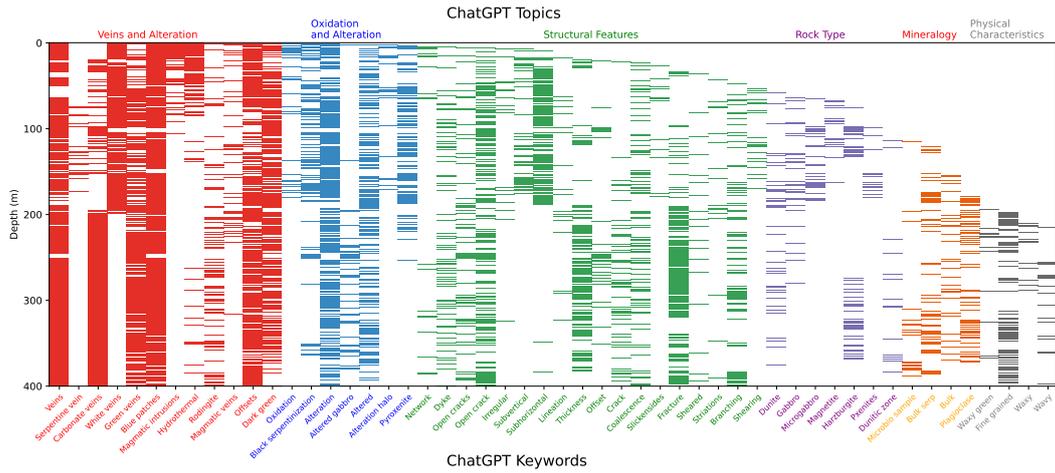


Figure 5. Presence of keywords per depth, grouped by type of information they convey.

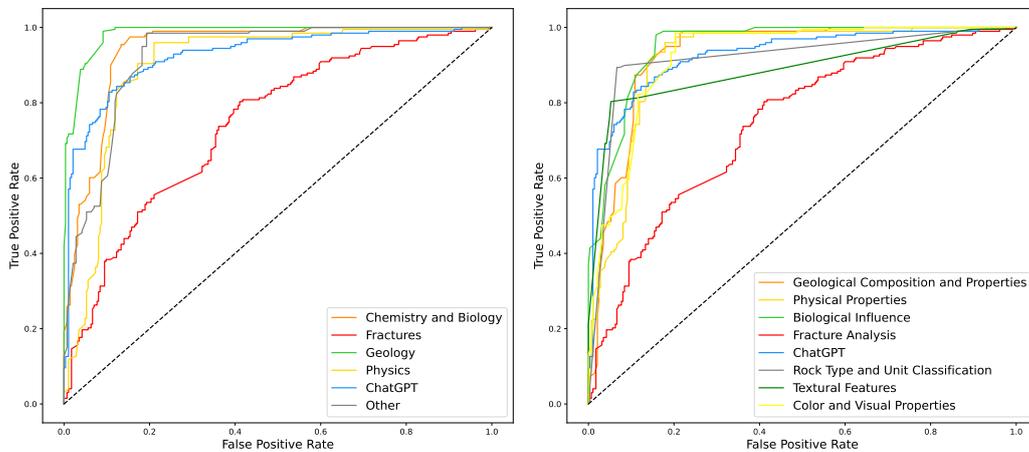


Figure 6. Plotting of the ROC curve for a different subset of features. The left panel has features categorized into groups by an expert. The right panel has features categorized into groups by ChatGPT. Some curves (e.g., Fractures (left) and Fracture Analysis (right)) are identical because they contain identical feature groups.

460 case when we attempted to use the general-purpose image segmentation tool Segment
 461 Anything (Kirillov et al., 2023) which was unable to accurately label fractures.

462 In addition to summarizing the drilling reports we also used ChatGPT to group
 463 dataset features that were put into the BA1B dataset (see Section Supporting informa-
 464 tion). ChatGPT created feature groupings very similar to the expert groupings (Figure
 465 6). ChatGPT correctly classified the columns related to the connectivity code which was
 466 not expected due to these columns being shortened acronyms with very little informa-
 467 tion provided otherwise and b) coming from a paper publishing a new method for de-
 468 tecting network connectivity (Amiri et al., 2023). ChatGPT’s reasoning behind each group-
 469 ing is sensible. It recognizes chemical composition relationships to the mineralogical com-
 470 position and how that affects alteration. However, in some cases, its reasoning is not very
 471 deep. For example, it gives the rationale for separating Cell Abundance into its category
 472 as “Microbial activity can significantly impact mineral alteration processes”. While this
 473 is true, it does not describe how it may occur. Although we did not prompt ChatGPT

474 for this effort. Ultimately, using ChatGPT for feature selection allowed ChatGPT to ar-
 475 rive at the same results as the expert reasoned groupings do. Namely, fracture location
 476 and complexity have much lower predictive power than other features with regard to where
 477 high peridotite alteration occurs (Figure 6).

478 The fracture density and network complexity measurements were less predictive
 479 than other features in the catboost model (Figure 6). This suggests that peridotite al-
 480 teration within borehole BA1B could be driven by multiple factors. It is possible that
 481 some of the primary fractures in the network were created tectonically. These tectonic
 482 fractures could provide pathways for meteoric fluids to access unaltered peridotites from
 483 the surface. This corresponds to the in-situ oxygen isotope study of serpentinites from
 484 the Oman MBO (Scicchitano et al., 2023). Namely, the microscale oxygen isotope com-
 485 position in two serpentinite samples from the BA1B core confirms varying stages of hy-
 486 drothermal alteration. This process likely began in an oceanic environment and progressed
 487 within a continental context, influenced by low-temperature ($T < 50^{\circ}\text{C}$) interactions
 488 with groundwater possessing distinct $\delta^{18}\text{O}$ values. There does seem to be some hierar-
 489 chical networks in the Oman peridotite as we can see primary fractures with branching,
 490 connecting fractures in the example segmented image (Figure 3). Why would these net-
 491 works occur yet not correlate with alteration? First, it is possible that the predicted value
 492 of $> 90\%$ alteration (or not) is too coarse grained a measurement of alteration to be of
 493 value in this setting. Given that bulk density of the rock can be a proxy for peridotite
 494 alteration as it scales linearly with alteration, we would expect that the fracture den-
 495 sity would strongly correlate with bulk density (Figure 4). However this is not the case
 496 ($\rho = -0.0225$). These fracture networks must act as fluid pathways given adequate fluid
 497 pressures, and this fluid motion is detected acoustically in BA1B (Aiken et al., 2022).
 498 And there is evidence that there is ongoing, low-temperature alteration in the Oman peri-
 499 dotites (Kelemen et al., 2021). Thus, we are left with the conclusion that the relation-
 500 ship between peridotite alteration and fracture networks is more complex. As such, frac-
 501 ture network development is likely a result of various processes including reaction driven
 502 and tectonic fracturing. Future studies may use our classification as a ground basis for
 503 in depth investigations.

504 It is important to note that throughout these results depth is likely the single vari-
 505 able that dictates how much of the peridotites are altered. This high correlation with
 506 alteration ($\rho = -0.70$) is why depth is removed from the training data sets. Many of the
 507 physical measurements and VCD-based keywords are depth-dependent as well (Figures
 508 4, 5). This is likely because these measurements correlate with the already altered peri-
 509 dotite. However, they do not necessarily correlate with the process that drove this al-
 510 teration.

511 4.1 Conclusion

512 This paper has presented an AI-based pipeline to ingest and analyse multi-modal
 513 data from the Oman Drilling Project’s Multi-borehole observatory. In this pipeline, a
 514 random forest algorithm was used for image segmentation of core images. Additionally,
 515 ChatGPT was utilised to summarize the expert knowledge from the drilling reports. These
 516 were coupled with physical, chemical, and biological measurements and used to predict
 517 the presence of highly altered peridotites via a catboost model. The catboost model pro-
 518 vided valuable outlooks of the main factors influencing peridotite alteration. It indicates
 519 textual and physical data such as depth and mineral composition are of primary impor-
 520 tance in the classification, but the network analysis data taken from segmentation rep-
 521 resent a suitable alternative and provide acceptable results. Moreover, it shows an AI-
 522 based treatment of geological data can equal a physical measurements-oriented method,
 523 and is a viable substitute for this classification problem. While this pipeline is particu-
 524 lar to the research questions related to the Oman Drilling Project’s Borehole BA1B,

525 much of the AI-based framework presented in this paper applies to a great many drilling-
526 related data sets.

527 A critical component of this project was also using openly available, easy to use
528 tools. Ilastik (Berg et al., 2019) is free and open source and can be used without any pro-
529 gramming expertise. OpenAI’s ChatGPT tool is also offered as a free option. Catboost
530 (Prokhorenkova et al., 2018) is designed to be used out of the box without a long and
531 expert driven hyperparameter search. The single expert driven, programming task in this
532 project was the fracture network complexity estimations. In this way, we present a frame-
533 work for using modern, sophisticated tools to address multi-modal and interdisciplinary
534 data.

535 We hope that future work can use AI-agents to bulk process the vast quantities of
536 data that have been collected by international continental and oceanic drilling opera-
537 tions. Using AI in this way can both automate the extensive work required to ingest such
538 datasets, but also it can leverage the massive resources that have been used across the
539 world to generate these data sets. As such, our workflow shows how we can utilize AI
540 and machine learning to streamline the analysis of large, disparate, and multi-modal datasets.
541 This provides the basis to utilize often largely unused data such as the visual core de-
542 scription to develop a systematic dataset for the further depth and correlative analyses.

543 5 Open Research

544 All codes can be found at <https://zenodo.org/doi/10.5281/zenodo.10226092>, data
545 is available on the International Continental Drilling Project Webpage [https://www.icdp-
546 -online.org/projects/by-continent/asia/odp-oman/](https://www.icdp-online.org/projects/by-continent/asia/odp-oman/). A detailed tutorial about
547 the use of Ilastik in this paper is available [here](#). A detailed tutorial of the use of the con-
548 nectivity estimation software can be found [here](#).

549 Acknowledgments

550 Drilling in the Oman Drilling Project was supported by the Alfred P. Sloan Foundation
551 (in association with the Deep Carbon Observatory, DCO), the International Continen-
552 tal Scientific Drilling Program (ICDP), US National Science Foundation (NSF) Research
553 (Grants NSF-EAR-1516300, the Japanese Marine Science and Technology Center (JAM-
554 STEC), and the Japanese Society for the Promotion of Science (JSPS) grant number 16H06347,
555 and contributions from the Sultanate of Oman Ministry of Regional Municipalities and
556 Water Resources, the Oman Public Authority for Mining, and Sultan Qaboos Univer-
557 sity. The project also received funding from the European Research Council (ERC DIME,
558 grant no. 669972 and ERC nanoEARTH, grant no. 852069), the US National Science
559 Foundation (grant no. EAR-1516313), and the Norwegian Research Council (SerpRateAI,
560 grant no. 334395).

561 References

- 562 Aiken, J. M., Sohn, R. A., Renard, F., Matter, J., Kelemen, P., & Jamtveit, B.
563 (2022). Gas migration episodes observed during peridotite alteration in the
564 samail ophiolite, oman. *Geophysical Research Letters*, *49*(21), e2022GL100395.
565 doi: <https://doi.org/10.1029/2022GL100395>
- 566 Amiri, H., Vasconcelos, I., Jiao, Y., Chen, P.-E., & Plümper, O. (2023). Quantifying
567 microstructures of earth materials using higher-order spatial correlations and
568 deep generative adversarial networks. *Scientific reports*, *13*(1), 1805. doi:
569 [10.1038/s41598-023-28970-w](https://doi.org/10.1038/s41598-023-28970-w)
- 570 Aydin, A., & DeGraff, J. M. (1988). Evolution of polygonal fracture patterns in lava
571 flows. *Science*, *239*(4839), 471-476. doi: [10.1126/science.239.4839.471](https://doi.org/10.1126/science.239.4839.471)
- 572 Barrios, F., López, F., Argerich, L., & Wachenchauser, R. (2016). Variations of

- 573 the similarity function of textrank for automated summarization. *CoRR*,
574 *abs/1602.03606*. Retrieved from <http://arxiv.org/abs/1602.03606>
- 575 Berg, S., Kutra, D., Kroeger, T., Straehle, C. N., Kausler, B. X., Haubold, C., ...
576 Kreshuk, A. (2019, September). ilastik: interactive machine learning for
577 (bio)image analysis. *Nature Methods*. doi: 10.1038/s41592-019-0582-9
- 578 Blackman, D., Ildfonse, B., John, B., Ohara, Y., Miller, D., MacLeod, C., & Sci-
579 entists, E. . (2006). *Proceedings IODP Expedition 304/305* (Tech. Rep.).
580 Integrated Ocean Drilling Program Management International, Inc. Retrieved
581 from <https://doi.org/10.2204/iodp.proc.304305.2006>
- 582 Bouchayer, C., Aiken, J. M., Thøgersen, K., Renard, F., & Schuler, T. V. (2022).
583 A machine learning framework to automate the classification of surge-type
584 glaciers in svalbard. *Journal of Geophysical Research: Earth Surface*, *127*(7),
585 e2022JF006597. doi: <https://doi.org/10.1029/2022JF006597>
- 586 Boudier, F., & Coleman, R. G. (1981). Cross section through the peridotite in
587 the samail ophiolite, southeastern oman mountains. *Journal of Geophysi-
588 cal Research: Solid Earth*, *86*(B4), 2573-2592. doi: [https://doi.org/10.1029/
589 JB086iB04p02573](https://doi.org/10.1029/JB086iB04p02573)
- 590 Chen, P.-E., Xu, W., Chawla, N., Ren, Y., & Jiao, Y. (2019). Hierarchical n-point
591 polytope functions for quantitative representation of complex heterogeneous
592 materials and microstructural evolution. *Acta Materialia*, *179*, 317–327.
- 593 Chen, X.-w., & Jeong, J. C. (2007). Enhanced recursive feature elimination. In *Sixth
594 international conference on machine learning and applications (icmla 2007)*
595 (p. 429-435). doi: 10.1109/ICMLA.2007.35
- 596 Cohen, J. (2013). *Statistical power analysis for the behavioral sciences*. Academic
597 press.
- 598 Cornelio, C., Dash, S., Austel, V., Josephson, T. R., Goncalves, J., Clarkson, K. L.,
599 ... Horesh, L. (2023). Combining data and theory for derivable scientific
600 discovery with ai-descartes. *Nature Communications*, *14*(1), 1777. doi:
601 10.1038/s41467-023-37236-y
- 602 Corre, M., Brunet, F., Schwartz, S., Gautheron, C., Agranier, A., & Lesimple,
603 S. (2023). Quaternary low-temperature serpentinization and carbona-
604 tion in the new caledonia ophiolite. *Scientific Reports*, *13*(1), 19413. doi:
605 10.1038/s41598-023-46691-y
- 606 de Obeso, J. C., & Kelemen, P. B. (2018). Fluid rock interactions on resid-
607 ual mantle peridotites overlain by shallow oceanic limestones: Insights
608 from wadi fins, sultanate of oman. *Chemical Geology*, *498*, 139-149. doi:
609 <https://doi.org/10.1016/j.chemgeo.2018.09.022>
- 610 de Obeso, J. C., & Kelemen, P. B. (2020). Major element mobility during serpen-
611 tinization, oxidation and weathering of mantle peridotite at low temperatures.
612 *Philosophical Transactions of the Royal Society A*, *378*(2165), 20180433. doi:
613 10.1098/rsta.2018.0433
- 614 Ellison, E. T., Templeton, A. S., Zeigler, S. D., Mayhew, L. E., Kelemen, P. B.,
615 Matter, J. M., & Party, T. O. D. P. S. (2021). Low-temperature hydro-
616 gen formation during aqueous alteration of serpentinized peridotite in the
617 samail ophiolite. *Journal of Geophysical Research: Solid Earth*, *126*(6),
618 e2021JB021981. doi: 10.1029/2021JB021981
- 619 Falk, E. S., & Kelemen, P. B. (2015). Geochemistry and petrology of listvenite in
620 the samail ophiolite, sultanate of oman: Complete carbonation of peridotite
621 during ophiolite emplacement. *Geochimica et Cosmochimica Acta*, *160*, 70-90.
622 doi: <https://doi.org/10.1016/j.gca.2015.03.014>
- 623 Friedman, J. H. (2002). Stochastic gradient boosting. *Computational statistics &
624 data analysis*, *38*(4), 367–378. doi: 10.1016/S0167-9473(01)00065-2
- 625 Früh-Green, G., Orcutt, B., Green, S., Cotterill, C., & Scientists, E. . (n.d.). *Pro-
626 ceeding IODP Expedition 357* (Tech. Rep.). International Ocean Discovery
627 Program. Retrieved from <https://doi.org/10.14379/iodp.proc.357.102>

628

.2017

629

Ge, Y., Hua, W., Mei, K., Ji, J., Tan, J., Xu, S., ... Zhang, Y. (2023). Openagi: When llm meets domain experts. *In Advances in Neural Information Processing Systems (NeurIPS)*. doi: 10.48550/arXiv.2304.04370

630

631

632

Gillis, K., Snow, J., Klaus, A., & Scientists, E. . (2014). *Proceedings IODP Expedition 345* (Tech. Rep.). Integrated Ocean Drilling Program. Retrieved from <https://doi.org/10.2204/iodp.proc.345.102.2014>

633

634

635

Gledhill, C. J. (2000). *Collocations in science writing* (Vol. 22). Gunter Narr Verlag.

636

637

Godard, M., Dautria, J.-M., & Perrin, M. (2003). Geochemical variability of the oman ophiolite lavas: Relationship with spatial distribution and paleomagnetic directions. *Geochemistry, Geophysics, Geosystems*, 4(6). doi: <https://doi.org/10.1029/2002GC000452>

638

639

640

641

Goss, H. (2020). The rise of machine learning. *Eos*, 101. Retrieved from <https://api.semanticscholar.org/CorpusID:225472155>

642

643

644

Guimerà, R., Reichardt, I., Aguilar-Mogas, A., Massucci, F. A., Miranda, M., Pallerès, J., & Sales-Pardo, M. (2020). A bayesian machine scientist to aid in the solution of challenging scientific problems. *Science Advances*, 6(5), eaav6971. doi: 10.1126/sciadv.aav6971

645

646

647

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: data mining, inference, and prediction*. Springer Science & Business Media.

648

649

650

Hosmer Jr, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression*. John Wiley & Sons.

651

652

Hulbert, C., Rouet-Leduc, B., Johnson, P. A., Ren, C. X., Rivière, J., Bolton, D. C., & Marone, C. (2019). Similarity of fast and slow earthquakes illuminated by machine learning. *Nature Geoscience*, 12(1), 69–74. doi: 10.1038/s41561-018-0272-8

653

654

655

656

Hövelmann, J., Putnis, C. V., Ruiz-Agudo, E., & Austrheim, H. (2012). Direct nanoscale observations of co2 sequestration during brucite [mg(oh)2] dissolution. *Environmental Science & Technology*, 46(9), 5253–5260. doi: 10.1021/es300403n

657

658

659

660

Iyer, K., rn Jamtveit, B., Mathiesen, J., Malthe-Sørenssen, A., & Feder, J. (2008). Reaction-assisted hierarchical fracturing during serpentization. *Earth and Planetary Science Letters*, 267(3), 503–516. doi: <https://doi.org/10.1016/j.epsl.2007.11.060>

661

662

663

664

Jamtveit, B., Putnis, C. V., & Malthe-Sørenssen, A. (2009). Reaction induced fracturing during replacement processes. *Contributions to Mineralogy and Petrology*, 157(1), 127–133. doi: 10.1007/s00410-008-0324-y

665

666

667

Jiao, Y., Stillinger, F., & Torquato, S. (2007). Modeling heterogeneous materials via two-point correlation functions: Basic principles. *Physical review E*, 76(3), 031110.

668

669

670

Kelemen, P. B., Leong, J. A., Carlos de Obeso, J., Matter, J. M., Ellison, E. T., Templeton, A., ... Team, T. O. D. P. S. (2021). Initial results from the oman drilling project multi-borehole observatory: Petrogenesis and ongoing alteration of mantle peridotite in the weathering horizon. *Journal of Geophysical Research: Solid Earth*, 126(12), e2021JB022729. doi: <https://doi.org/10.1029/2021JB022729>

671

672

673

674

675

676

Kelemen, P. B., & Matter, J. (2008). In situ carbonation of peridotite for co₂ storage. *Proceedings of the National Academy of Sciences*, 105(45), 17295–17300. doi: 10.1073/pnas.0805794105

677

678

679

Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., ... Girshick, R. (2023). Segment anything. *arXiv:2304.02643*.

680

681

682

Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., ... Kiela, D. (2020). Retrieval-augmented generation for knowledge-intensive nlp tasks. In

- 683 H. Larochelle, M. Ranzato, R. Hadsell, M. Balcan, & H. Lin (Eds.), *Advances*
684 *in neural information processing systems* (Vol. 33, pp. 9459–9474). Curran
685 Associates, Inc. Retrieved from [https://proceedings.neurips.cc/paper](https://proceedings.neurips.cc/paper_files/paper/2020/file/6b493230205f780e1bc26945df7481e5-Paper.pdf)
686 [_files/paper/2020/file/6b493230205f780e1bc26945df7481e5-Paper.pdf](https://proceedings.neurips.cc/paper_files/paper/2020/file/6b493230205f780e1bc26945df7481e5-Paper.pdf)
- 687 Li, Z., Ji, J., & Zhang, Y. (2022). From kepler to newton: Explainable AI for science
688 discovery. In *Icml 2022 2nd ai for science workshop*. Retrieved from [https://](https://openreview.net/forum?id=vA9hti-Fi7H)
689 openreview.net/forum?id=vA9hti-Fi7H
- 690 Lu, B., & Torquato, S. (1992). Lineal-path function for random heterogeneous mate-
691 rials. *Physical Review A*, *45*(2), 922.
- 692 MacLeod, C., Dick, H., Blum, P., & Scientists, E. . (2017). *Proceedings IODP Ex-*
693 *pedition 360* (Tech. Rep.). International Ocean Discovery Program. Retrieved
694 from <https://doi.org/10.14379/iodp.proc.360.102.2017>
- 695 Malvoisin, B., Podladchikov, Y. Y., & Myasnikov, A. V. (2021). Achieving com-
696 plete reaction while the solid volume increases: A numerical model applied
697 to serpentinisation. *Earth and Planetary Science Letters*, *563*, 116859. doi:
698 <https://doi.org/10.1016/j.epsl.2021.116859>
- 699 Malvoisin, B., Zhang, C., Müntener, O., Baumgartner, L. P., Kelemen, P. B., &
700 Party, O. D. P. S. (2020). Measurement of volume change and mass trans-
701 fer during serpentinization: Insights from the oman drilling project. *Jour-*
702 *nal of Geophysical Research: Solid Earth*, *125*(5), e2019JB018877. doi:
703 <https://doi.org/10.1029/2019JB018877>
- 704 McBeck, J. A., Aiken, J. M., Mathiesen, J., Ben-Zion, Y., & Renard, F. (2020). De-
705 formation precursors to catastrophic failure in rocks. *Geophysical Research Let-*
706 *ters*, *47*(24), e2020GL090255. doi: 10.1029/2020GL090255
- 707 Miller, G. A. (1995, nov). Wordnet: a lexical database for english. *Commun. ACM*,
708 *38*(11), 39–41. doi: 10.1145/219717.219748
- 709 Nori, H., Lee, Y. T., Zhang, S., Carignan, D., Edgar, R., Fusi, N., . . . Horvitz, E.
710 (2023). *Can generalist foundation models outcompete special-purpose tuning?*
711 *case study in medicine*.
- 712 Pallister, J. S., & Knight, R. J. (1981). Rare-earth element geochemistry of the
713 samail ophiolite near ibra, oman. *Journal of Geophysical Research: Solid*
714 *Earth*, *86*(B4), 2673-2697. doi: <https://doi.org/10.1029/JB086iB04p02673>
- 715 Plümper, O., & Matter, J. (2023, 06). Olivine—The Alteration Rock Star. *Ele-*
716 *ments*, *19*(3), 165-172. doi: 10.2138/gselements.19.3.165
- 717 Plümper, O., Røyne, A., Magrasó, A., & Jamtveit, B. (2012, 12). The interface-
718 scale mechanism of reaction-induced fracturing during serpentinization. *Geol-*
719 *ogy*, *40*(12), 1103-1106. doi: 10.1130/G33390.1
- 720 Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V., & Gulin, A. (2018).
721 Catboost: unbiased boosting with categorical features. In S. Bengio, H. Wal-
722 lach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, & R. Garnett (Eds.), *Ad-*
723 *vances in neural information processing systems* (Vol. 31). Curran Associates,
724 Inc. Retrieved from [https://proceedings.neurips.cc/paper_files/paper/](https://proceedings.neurips.cc/paper_files/paper/2018/file/14491b756b3a51daac41c24863285549-Paper.pdf)
725 [2018/file/14491b756b3a51daac41c24863285549-Paper.pdf](https://proceedings.neurips.cc/paper_files/paper/2018/file/14491b756b3a51daac41c24863285549-Paper.pdf)
- 726 Ravaut, P., Bayer, R., Hassani, R., Rousset, D., & Yahya’ey, A. (1997). Struc-
727 ture and evolution of the northern oman margin: gravity and seismic con-
728 straints over the zagros-makran-oman collision zone. *Tectonophysics*, *279*(1),
729 253-280. (The Adolphe Nicolas Volume) doi: [https://doi.org/10.1016/](https://doi.org/10.1016/S0040-1951(97)00125-X)
730 [S0040-1951\(97\)00125-X](https://doi.org/10.1016/S0040-1951(97)00125-X)
- 731 Schmidt, M., & Lipson, H. (2009). Distilling free-form natural laws from experimen-
732 tal data. *Science*, *324*(5923), 81-85. doi: 10.1126/science.1165893
- 733 Scicchitano, M. R., de Obeso, J. C., Blum, T. B., Valley, J. W., Kelemen, P. B.,
734 Nachlas, W. O., . . . Roddatis, V. (2023). An empirical calibration of the
735 serpentine-water oxygen isotope fractionation at t=25–100°C. *Geochim-*
736 *ica et Cosmochimica Acta*, *346*, 192-206. doi: [https://doi.org/10.1016/](https://doi.org/10.1016/j.gca.2023.02.015)
737 [j.gca.2023.02.015](https://doi.org/10.1016/j.gca.2023.02.015)

- 738 Sharma, A., Lysenko, A., Boroevich, K. A., Vans, E., & Tsunoda, T. (2021,
739 08). DeepFeature: feature selection in nonimage data using convolu-
740 tional neural network. *Briefings in Bioinformatics*, *22*(6), bbab297. doi:
741 10.1093/bib/bbab297
- 742 Teagle, Alt, Umino, Miyashita, Banerjee, Wilson, & Scientists, E. . (2006). *Pro-*
743 *ceedings IODP Expedition 309/312* (Tech. Rep.). Integrated Ocean Drilling
744 Program Management International, Inc. Retrieved from [https://doi.org/
745 10.2204/iodp.proc.309312.2006](https://doi.org/10.2204/iodp.proc.309312.2006)
- 746 Teagle, D., Ildefonse, B., Blum, P., & Scientists, E. . (2012). *Proceedings IODP Ex-*
747 *pedition 335* (Tech. Rep.). Integrated Ocean Drilling Program Management In-
748 ternational, Inc. Retrieved from [https://doi.org/10.2204/iodp.proc.335
749 .2012](https://doi.org/10.2204/iodp.proc.335.2012)
- 750 Wang, H., Fu, T., Du, Y., Gao, W., Huang, K., Liu, Z., . . . others (2023). Scientific
751 discovery in the age of artificial intelligence. *Nature*, *620*(7972), 47–60. doi: 10
752 .1038/s41586-023-06221-2
- 753 Wang, Y. D., Blunt, M. J., Armstrong, R. T., & Mostaghimi, P. (2021). Deep learn-
754 ing in pore scale imaging and modeling. *Earth-Science Reviews*, *215*, 103555.
755 doi: <https://doi.org/10.1016/j.earscirev.2021.103555>
- 756 Whitney, D. L., & Evans, B. W. (2010). Abbreviations for names of rock-forming
757 minerals. *American Mineralogist*, *95*(1), 185–187. doi: 10.2138/am.2010.3371
- 758 Zhao, W. X., Zhou, K., Li, J., Tang, T., Wang, X., Hou, Y., . . . others (2023). A
759 survey of large language models. *arXiv preprint arXiv:2303.18223*. doi: 10
760 .48550/arXiv.2303.18223
- 761 Zou, H., & Hastie, T. (2005, 03). Regularization and Variable Selection Via the
762 Elastic Net. *Journal of the Royal Statistical Society Series B: Statistical
763 Methodology*, *67*(2), 301-320. doi: 10.1111/j.1467-9868.2005.00503.x