A framework for an AI pipeline for borehole data

John M. Aiken[1,2,3], Elliot Dufornet[1], Hamed Amiri[3], Lotta Ternieten[3,4], Oliver Plümper[3]

[1]Njord Centre, Departments of Physics and Geosciences, University of Oslo, PO BOX 1048, Blindern, Oslo, 0316, Oslo, Norway
[2]Expert Analytics, Oslo, Norway
[3]Department of Earth Sciences, Utrecht University, Utrecht, The Netherlands
[4]Department 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

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

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 CO_2. 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

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 & Lipson, 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).

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

At site BA1, borehole BA1B was cored and images including complete wrap-around scans, physical, chemical, and biological measurements (e.g., mean dry electrical resistivity, cell abundance), and lithographic information were recorded. Initial results of the OmanDP demonstrate that in borehole BA1B, between 65 and 100% of the peridotite has been hydrated to form serpentinite and related rock types. The decrease in the extent of peridotite alteration with depth may suggest that significant peridotite alteration in the region has been relatively young, within the last 50,000 years (Kelemen et al., 2021). H_2 and CH_4 outgassing have previously been detected in the Oman boreholes and are possible products of ongoing peridotite alteration (Ellison et al., 2021; Aiken et al., 2022).

The chemical reactions associated with peridotite alteration are well understood (e.g., (Kelemen & Matter, 2008; Plümper & Matter, 2023)). Olivine and pyroxenes react with water and carbon dioxide to form mainly serpentine minerals, brucite, iron oxides, and carbonates. Low-temperature alteration ($< 150°$ C) is possible and has been observed in Oman and other on-land environments (de Obeso & Kelemen, 2020; Corre et al., 2023). Redox reactions further produce H_2 and CH_4, which can be observed bubbling up continuously in alkaline springs found in peridotite-rich areas. The complete conversion of peridotite to serpentinite is not fully understood because the associated swelling should "armor" the reactive surfaces of the peridotite, thus preventing water from continuing to interact with unaltered rock (Hövelmann et al., 2012; Malvoisin et al., 2020, 2021). It is assumed that the volumetric expansion of the rock as a consequence of the hydration would induce stress on the surrounding host rock, thus opening new pathways to unaltered rock penetrating the "armor". This process, known as "reaction-driven cracking" (Kelemen & Matter, 2008; Jamtveit et al., 2009; Plümper et al., 2012), is expected to create hierarchical fracture networks within the host rock (Jamtveit et al., 2009). Thus, in addition to the geological attributes of the Oman peridotite, the density and complexity of the fracture networks should be indicative of recent and/or ongoing peridotite alteration (Iyer et al., 2008). Reaction driven cracking should develop a characteristic hierarchical network pattern dominated by four-sided domains (Aydin & DeGraff, 1988; Iyer et al., 2008). These fractures should grow from older fractures, linking different generations of fractures together. Thus, in an altered peridotite environment strongly influenced by reaction driven cracking we expect to see a fracture network made up of polygons with four or more sides and few single, linear fractures in regions of high alteration. Fractures in the OmanDP MBO cores have been qualitatively described through visual core descriptions using classification rubrics developed for ocean drilling expeditions (Blackman et al., 2006; MacLeod et al., 2017). These descriptions are insufficient to describe the complexity of fracture networks which would be necessary to identify potential regions of ongoing reaction driven cracking. To overcome the limitation of this qualitative description, in this study, we use a machine learning-based image segmentation model to identify fractures in the wrap-around core images. We then use statistical microstructure descriptors (SMDs) to describe the fracture network complexity (Lu & Torquato, 1992; P.-E. Chen et al., 2019; Amiri et al., 2023).

In this paper, we present a machine learning-oriented approach for treating multimodal data produced during the coring and subsequent investigations of OmanDP borehole BA1B. This framework is designed to normalize these multi-modal data (in our case, wrap around core images, physical measurements, and visual core descriptions) quickly. Much of the work presented here typically would take many months of work to complete compared to the computational workflow presented here. Specifically, we present two separate methods to ingest data from the OmanDP borehole BA1B: 1) we produced a machine learning image classifier for wrap-around core images that segmented fractures which is then used to calculate fracture network characteristics, and 2) we utilized the large language model ChatGPT to summarize handwritten visual core descriptions (VCDs) from the coring expedition. The VCDs represent on-site expert knowledge about the geology of the cores and also, observations that could help explain the presence of highly altered peridotites in the absence of complex fracture networks. They describe different morphometric features such as the presence of veins, alteration, and oxidation, as well as structural features and minerology. They are open-ended, semi-structured text documents written per core segment and thus make a depth-dependent, expert description of the BA1B core. The fracture network statistics and VCD keyword data are then compiled into a single dataset along with physical measurements (e.g., mean dry electrical resistivity) which is then used to train a gradient boosted trees (catboost) classification model predicting alteration in the peridotite core (Prokhorenkova et al., 2018). This model is then used to find a geological explanation from the machine learning classification model for the alteration of the core segments. A central objective of this paper is to establish a repeatable pattern for processing this type of data, enabling even individuals without earth science knowledge to exploit it. Additionally, it is to explore the impact of non-tectonic fracturing of rock on peridotite alteration using machine learning methods.

2 Data and Methods

We utilize three types of data extracted from the OmanDP borehole BA1B: wrap-around images of the borehole core, physical, chemical, and biological measurements made after the coring, and textual data comprising geologists' remarks regarding the drilled sections known as the "Visual Core Descriptions" (VCD). This data is processed (Figure 1) through fracture labeling via image segmentation of the wrap-around core images, fracture density and network connectivity estimation from the labeled fracture images, and summarization into keywords of the VCD text using ChatGPT. These data are combined with physical measurement data to create a depth-dependent database of borehole BA1B. This database is then used to predict the detected alteration within the core, as reported from the expedition (Kelemen et al., 2021).

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

2.1 Site Description

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

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

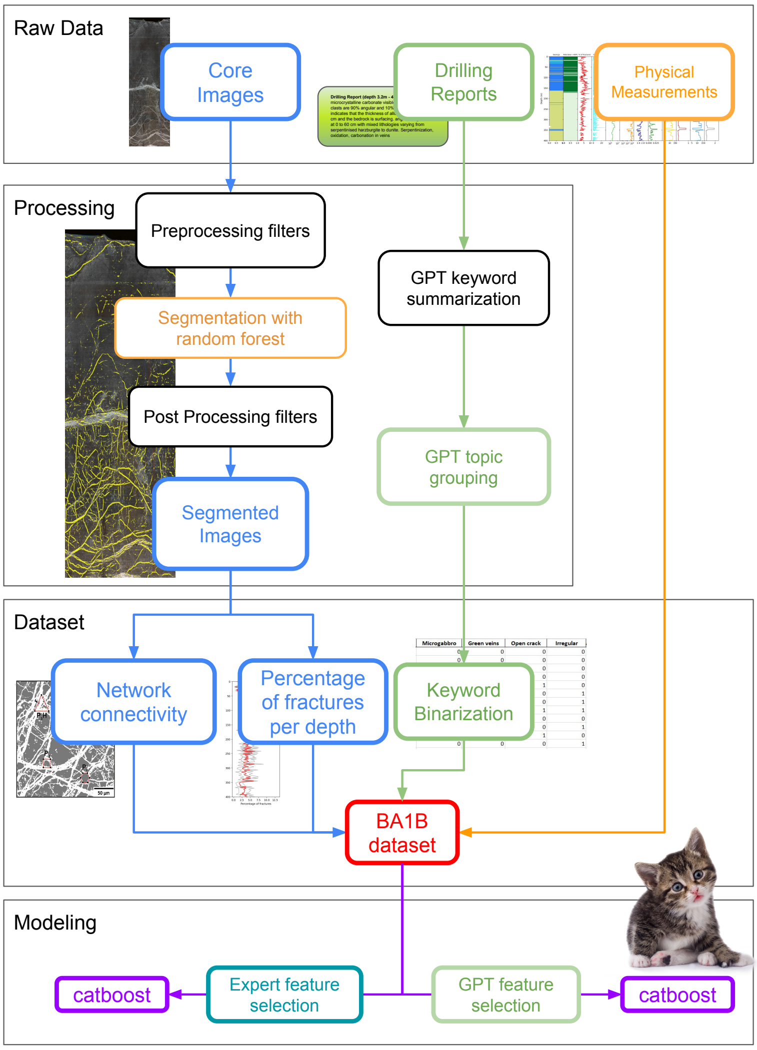

Figure 1. Pipeline utilizing AI and machine learning to ingest data taken from Oman Drilling Project Multi-borehole Observatory borehole BA1B. Ultimately this processes 505 wrap-around core images, 505 drilling reports per core segment, and 30 physical measurements into a data set of 96 columns ranging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

2.2 Core image analysis

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

2.2.1 Image segmentation

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

2.2.2 Estimating fracture density

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

$$F\% = \frac{I_{label}}{I_{area}} \tag{1}$$

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

2.2.3 Estimating Connectivity

Fracture network connectivity is another property of the observed fracture network in the core images that can have an impact on the alteration process. Thus, it is necessary to quantify such connectivity so we can use it as an additional feature to our machine learning model. Our approach involves utilizing n-point spatial correlation functions, i.e. SMDs (Lu & Torquato, 1992; P.-E. Chen et al., 2019; Amiri et al., 2023). These functions represent the probability of n random points separated by a distance r to lie in the same phase such as fractures. However, for $n \geq 3$ this probability calculation becomes computationally challenging. To address this, we focus on a subset of these functions: n-point polytope functions. These functions are defined by the probability that the n vertices of a random regular n-point polytope with an edge length r will fall within the same phase (P.-E. Chen et al., 2019). Given that reaction driven fractures should produce network patterns that are most likely to have four-sided polygons (Iyer et al., 2008), the detection of these prevelance of such polygons will indicate the complexity of the existing fracture network. That is, if the fracture network is made up of mostly longer linear segments and fewer polygons, it is less likely to present a hierarchical network generated from reaction driven cracking.

To specifically assess fracture connectivity, we compute the lineal-path L function(Lu & Torquato, 1992). This function measures the probability of a whole segment of a random line to lie within the fractures, providing an efficient means to evaluate the linear connectivity in complex fracture networks such as those found in serpentinites (Amiri et al., 2023). In our study, six correlation functions are calculated: S_2 for two-point correlation, P_{3H} for horizontal triangles, P_{3V} for vertical triangles, P_4 for squares, P_6 for hexagons, and L for the lineal-path function. Alongside these, we also compute normalized versions of these SMDs, termed "scaled autocovariance functions" (Jiao et al., 2007), altogether introducing 12 features representing geometrical patterns and linear connectivity, and ultimately the complexity, of the fracture network within the BA1B core segments.

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

2.3 Hand Written Expert Visual Core Descriptions

After successful, drilling the recovered cores are processed and described during core description campaign following a protocol created by reviewing and adapting procedures of previous scientific ocean drilling expeditions (Blackman et al., 2006; Teagle et al., 2006; D. Teagle et al., 2012; Gillis et al., 2014; MacLeod et al., 2017). The protocol contains the optical description of the cores and various scientific analyses. Multiple teams perform the core characterization, each focusing on specific aspects. The teams are as follows: igneous petrology, alteration/metamorphic petrology, structural geology, geochemistry, paleomagnetism, physical properties, near-visible infrared scanning, microbiology, and wireline geophysical logging and hydrogeological testing. At the end of the campaign, visual core descriptions (VCDs) are produced, which are section-by-section summaries of the core description observables and most pertinent instrumental measurement parameters of the recovered cores.

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

2.3.1 ChatGPT for Drilling Reports

Recent efforts in the development of large language models (LLMs) have caused a paradigm shift in the availability of easy-to-use text summarizing tools (Zhao et al., 2023). We choose to use ChatGPT due to its ease of use Application Programming Interface (API). Other LLMs likely offer similar utility. When compared to the effort and lack of utility of traditional text analysis methods (traditional natural language processing methods such as Lemmatization (Miller, 1995), Frequency distributions and collocations (Gledhill, 2000), and TextRank (Barrios et al., 2016) did not produce valid keywords) new LLMs provide a new way forward for accessing high density, hard to quantify text data.

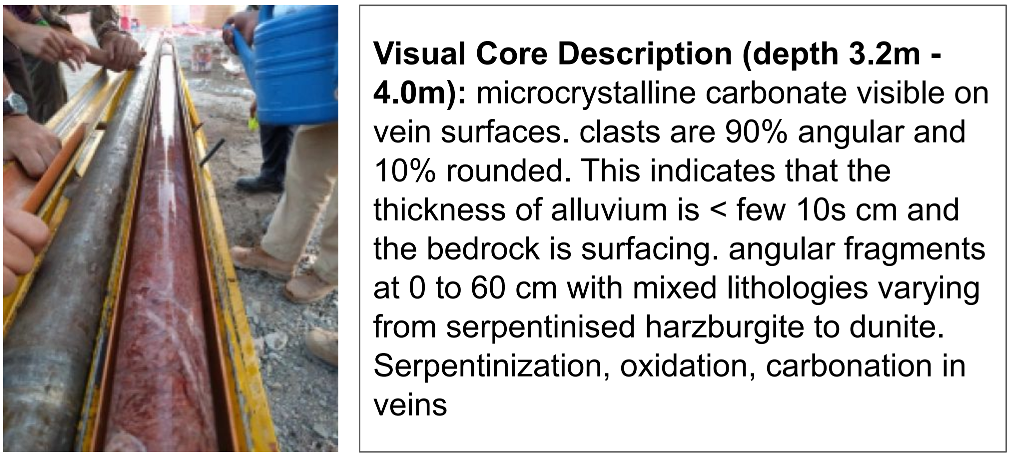

Figure 2. Field researchers cataloging the visual core description (left). An example of the visual core description text that is summarized into keywords using ChatGPT (right).

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

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

with **{text}** being replaced by the geologist's remarks. This produced hundreds of different keywords that emerged from the process, many of which were close duplicates or similar keywords. These keywords were then condensed for duplicates and/or similarities (e.g., "vein" versus "veins"). Keywords that were reported by ChatGPT less than 50 times (representing less than 10% of the BA1B total cored depth) were removed. Ultimately, 52 keywords remained. Those keywords were integrated in the dataset as binary variables for each core segment. Then, we asked ChatGPT to group the different keywords into topics based on the type of information they convey. We plotted the graph of keywords depending on depth to have preliminary information about the keyword features of the core (Figure 5).

2.4 Data set

Ultimately this process produced a depth-dependent data set composed of 690 rows corresponding to 690 different sections of the core and involved image analysis fracture-related data, textual data from geologists' reports, geological data, and physical measurements. In total, the dataset is comprised of 96 features: 13 of them derived from core image segmentation of fractures, 51 from the extracted keywords from VCDs, 30 from direct physical measurements, plus the depth and the alteration (we wish to predict the alteration). For each of the 690 core segments that include: depth range (in meters), percentage of alteration ($\geq 90\%$), keywords picked by ChatGPT, fracture density estimate, fracture network connectivity estimates, and physical measurements including precise mineral composition, electrical resistivity, magnetic susceptibility, cell abundance, and trace of volatile elements. Not all rows are assigned to an image because the image report provides only 505 images of the borehole. The physical measurements are only available for a limited number of sections throughout the borehole (typically only every 20m of core sections, please see the dataset for more details). We extrapolated missing data by imputing the sample without physical data using the values from the samples above. This data set was then used in a catboost model to predict per core segment if the percentage 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 ≈25m 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 ≈ 2-3mm. 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 ($\text{AUC}_{fractures}$=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

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

4 Discussion

This paper presents an AI-based pipeline for ingesting high-density, high-complexity disparate data sets into a single data set that can be used for analysis. This included core imagery, expert remarks (VCDs), and various physical, chemical, and biological measurements. These data were condensed into a single data set cataloging various features per depth increment of the ODP Multi-borehole Observatory borehole BA1B. A random forest classifier was used to label fractures in the data which were then quantified using fracture network connectivity statistics. An LLM (ChatGPT) was employed to summarize VCD text and describe and group the dataset features into topical groups to compare during modeling. This analysis produced the following result: complex fracture networks do not appear in high-density arrangements that correlate with a high degree of peridotite alteration. Moreover, this process used free and open source tools to automate much of the workflow reducing the time it would take to identify and label fractures in pictures, identify relevant text in thousands of comments, and combine this information together to visualize and then apply statistical analysis such as the catboost model presented in this paper.

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

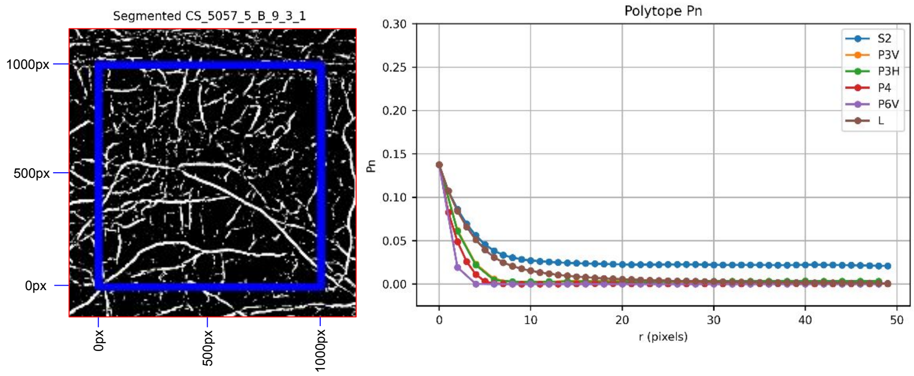

Figure 3. An example of fracture labeling and fracture network polygon identification. In the left image, segmented fractures are presented with white as a labeled fracture and black as a labeled host rock. The blue square represents the 1000px by 1000px selection taken from the larger core image. On 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.

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

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

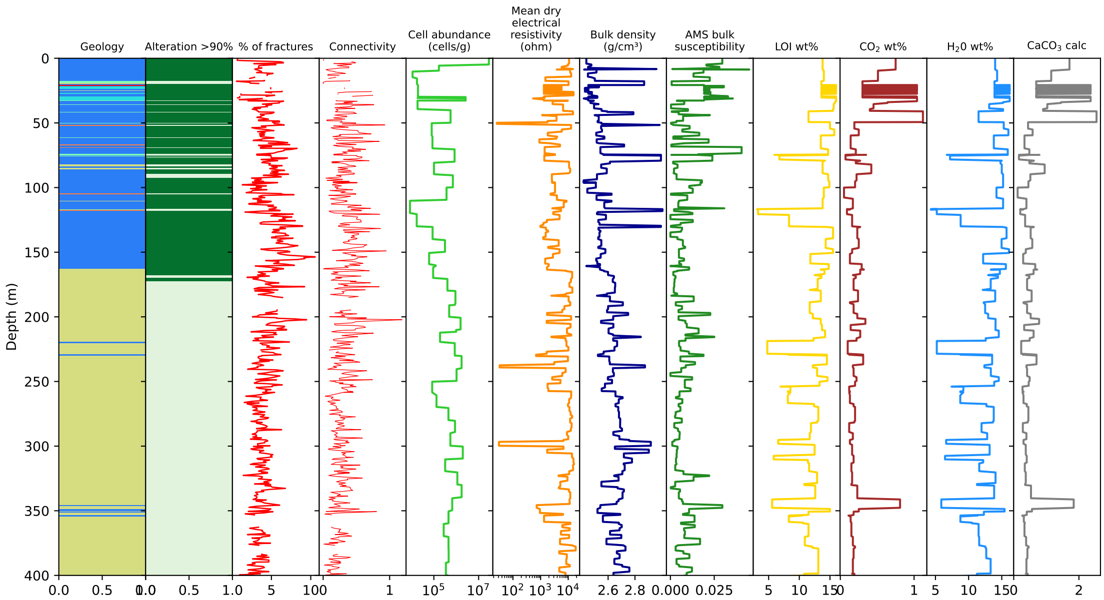

Figure 4. A subset of measurements taken in borehole BA1B. The % of fractures and connectivity are calculated from the random forest segmentation. Other quantities were measured on-site during or after the coring of borehole BA1B. These represent a cross-section of dataset features related to fracturing, 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 catboost model.

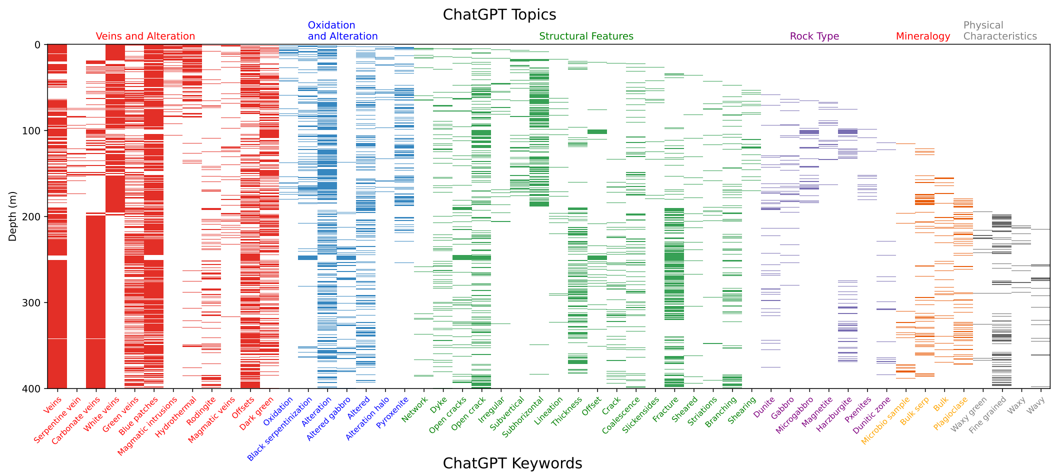

Figure 5. Presence of keywords per depth, grouped by type of information they convey.

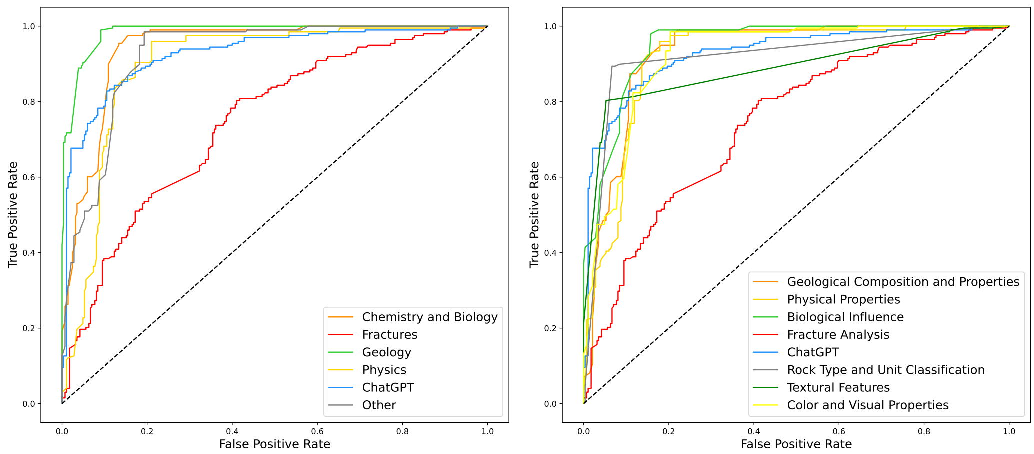

Figure 6. Plotting of the ROC curve for a different subset of features. The left panel has features categorized into groups by an expert. The right panel has features categorized into groups by ChatGPT. Some curves (e.g., Fractures (left) and Fracture Analysis (right)) are identical because they contain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

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

In addition to summarizing the drilling reports we also used ChatGPT to group dataset features that were put into the BA1B dataset (see Section Supporting information). ChatGPT created feature groupings very similar to the expert groupings (Figure 6). ChatGPT correctly classified the columns related to the connectivity code which was not expected due to these columns being shortened acronyms with very little information provided otherwise and b) coming from a paper publishing a new method for detecting network connectivity (Amiri et al., 2023). ChatGPT's reasoning behind each grouping is sensible. It recognizes chemical composition relationships to the mineralogical composition and how that affects alteration. However, in some cases, its reasoning is not very deep. For example, it gives the rationale for separating Cell Abundance into its category as "Microbial activity can significantly impact mineral alteration processes". While this is true, it does not describe how it may occur. Although we did not prompt ChatGPT for this effort. Ultimately, using ChatGPT for feature selection allowed ChatGPT to arrive at the same results as the expert reasoned groupings do. Namely, fracture location and complexity have much lower predictive power than other features with regard to where high peridotite alteration occurs (Figure 6).

The fracture density and network complexity measurements were less predictive than other features in the catboost model (Figure 6). This suggests that peridotite alteration within borehole BA1B could be driven by multiple factors. It is possible that some of the primary fractures in the network were created tectonically. These tectonic fractures could provide pathways for meteoric fluids to access unaltered peridotites from the surface. This corresponds to the in-situ oxygen isotope study of serpentinites from the Oman MBO (Scicchitano et al., 2023). Namely, the microscale oxygen isotope composition in two serpentinite samples from the BA1B core confirms varying stages of hydrothermal alteration. This process likely began in an oceanic environment and progressed within a continental context, influenced by low-temperature (T $<$ 50°C) interactions with groundwater possessing distinct δ^{18}O values. There does seem to be some hierarchical networks in the Oman peridotite as we can see primary fractures with branching, connecting fractures in the example segmented image (Figure 3). Why would these networks occur yet not correlate with alteration? First, it is possible that the predicted value of $> 90\%$ alteration (or not) is too course grained a measurement of alteration to be of value in this setting. Given that bulk density of the rock can be a proxy for peridotite alteration as it scales linearly with alteration, we would expect that the fracture density would strongly correlate with bulk density (Figure 4). However this is not the case ($\rho = -0.0225$). These fracture networks must act as fluid pathways given adequate fluid pressures, and this fluid motion is detected acoustically in BA1B (Aiken et al., 2022). And there is evidence that there is ongoing, low-temperature alteration in the Oman peridotites (Kelemen et al., 2021). Thus, we are left with the conclusion that the relationship between peridotite alteration and fracture networks is more complex. As such, fracture network development is likely a result of various processes including reaction driven and tectonic fracturing. Future studies may use our classification as a ground basis for in depth investigations.

It is important to note that throughout these results depth is likely the single variable that dictates how much of the peridotites are altered. This high correlation with alteration (ρ=-0.70) is why depth is removed from the training data sets. Many of the physical measurements and VCD-based keywords are depth-dependent as well (Figures 4, 5). This is likely because these measurements correlate with the already altered peridotite. However, they do not necessarily correlate with the process that drove this alteration.

4.1 Conclusion

This paper has presented an AI-based pipeline to ingest and analyse multi-modal data from the Oman Drilling Project's Multi-borehole observatory. In this pipeline, a random forest algorithm was used for image segmentation of core images. Additionally, ChatGPT was utilised to summarize the expert knowledge from the drilling reports. These were coupled with physical, chemical, and biological measurements and used to predict the presence of highly altered peridotites via a catboost model. The catboost model provided valuable outlooks of the main factors influencing peridotite alteration. It indicates textual and physical data such as depth and mineral composition are of primary importance in the classification, but the network analysis data taken from segmentation represent a suitable alternative and provide acceptable results. Moreover, it shows an AI-based treatment of geological data can equal a physical measurements-oriented method, and is a viable substitute for this classification problem. While this pipeline is particular to the research questions related to the Oman Drilling Project's Borehole BA1B, much of the AI-based framework presented in this paper applies to a great many drilling-related data sets.

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

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

5 Open Research

All codes can be found at `https://zenodo.org/doi/10.5281/zenodo.10226092`, data is available on the International Continental Drilling Project Webpage `https://www.icdp-online.org/projects/by-continent/asia/oodp-oman/`. A detailed tutorial about the use of Ilastik in this paper is available here. A detailed tutorial of the use of the connectivity estimation software can be found here.

Acknowledgments

Drilling in the Oman Drilling Project was supported by the Alfred P. Sloan Foundation (in association with the Deep Carbon Observatory, DCO), the International Continental Scientific Drilling Program (ICDP), US National Science Foundation (NSF) Research (Grants NSF-EAR-1516300, the Japanese Marine Science and Technology Center (JAMSTEC), and the Japanese Society for the Promotion of Science (JSPS) grant number 16H06347, and contributions from the Sultanate of Oman Ministry of Regional Municipalities and Water Resources, the Oman Public Authority for Mining, and Sultan Qaboos University. The project also received funding from the European Research Council (ERC DIME, grant no. 669972 and ERC nanoEARTH, grant no. 852069), the US National Science Foundation (grant no. EAR-1516313), and the Norwegian Research Council (SerpRateAI, grant no. 334395).

References

Aiken, J. M., Sohn, R. A., Renard, F., Matter, J., Kelemen, P., & Jamtveit, B. (2022). Gas migration episodes observed during peridotite alteration in the samail ophiolite, oman. *Geophysical Research Letters*, *49*(21), e2022GL100395. doi: https://doi.org/10.1029/2022GL100395

Amiri, H., Vasconcelos, I., Jiao, Y., Chen, P.-E., & Plümper, O. (2023). Quantifying microstructures of earth materials using higher-order spatial correlations and deep generative adversarial networks. *Scientific reports*, *13*(1), 1805. doi: 10.1038/s41598-023-28970-w

Aydin, A., & DeGraff, J. M. (1988). Evolution of polygonal fracture patterns in lava flows. *Science*, *239*(4839), 471-476. doi: 10.1126/science.239.4839.471

Barrios, F., López, F., Argerich, L., & Wachenchauzer, R. (2016). Variations of the similarity function of textrank for automated summarization. *CoRR*, *abs/1602.03606*. Retrieved from http://arxiv.org/abs/1602.03606

Berg, S., Kutra, D., Kroeger, T., Straehle, C. N., Kausler, B. X., Haubold, C., ... Kreshuk, A. (2019, September). ilastik: interactive machine learning for (bio)image analysis. *Nature Methods*. doi: 10.1038/s41592-019-0582-9

Blackman, D., Ildefonse, B., John, B., Ohara, Y., Miller, D., MacLeod, C., & Scientists, E. . (2006). *Proceedings IODP Expedition 304/305* (Tech. Rep.). Integrated Ocean Drilling Program Management International, Inc. Retrieved from https://doi.org/10.2204/iodp.proc.304305.2006

Bouchayer, C., Aiken, J. M., Thøgersen, K., Renard, F., & Schuler, T. V. (2022). A machine learning framework to automate the classification of surge-type glaciers in svalbard. *Journal of Geophysical Research: Earth Surface*, *127*(7), e2022JF006597. doi: https://doi.org/10.1029/2022JF006597

Boudier, F., & Coleman, R. G. (1981). Cross section through the peridotite in the samail ophiolite, southeastern oman mountains. *Journal of Geophysical Research: Solid Earth*, *86*(B4), 2573-2592. doi: https://doi.org/10.1029/JB086iB04p02573

Chen, P.-E., Xu, W., Chawla, N., Ren, Y., & Jiao, Y. (2019). Hierarchical n-point polytope functions for quantitative representation of complex heterogeneous materials and microstructural evolution. *Acta Materialia*, *179*, 317–327.

Chen, X.-w., & Jeong, J. C. (2007). Enhanced recursive feature elimination. In *Sixth international conference on machine learning and applications (icmla 2007)* (p. 429-435). doi: 10.1109/ICMLA.2007.35

Cohen, J. (2013). *Statistical power analysis for the behavioral sciences.* Academic press.

Cornelio, C., Dash, S., Austel, V., Josephson, T. R., Goncalves, J., Clarkson, K. L., ... Horesh, L. (2023). Combining data and theory for derivable scientific discovery with ai-descartes. *Nature Communications*, *14*(1), 1777. doi: 10.1038/s41467-023-37236-y

Corre, M., Brunet, F., Schwartz, S., Gautheron, C., Agranier, A., & Lesimple, S. (2023). Quaternary low-temperature serpentinization and carbonation in the new caledonia ophiolite. *Scientific Reports*, *13*(1), 19413. doi: 10.1038/s41598-023-46691-y

de Obeso, J. C., & Kelemen, P. B. (2018). Fluid rock interactions on residual mantle peridotites overlain by shallow oceanic limestones: Insights from wadi fins, sultanate of oman. *Chemical Geology*, *498*, 139-149. doi: https://doi.org/10.1016/j.chemgeo.2018.09.022

de Obeso, J. C., & Kelemen, P. B. (2020). Major element mobility during serpentinization, oxidation and weathering of mantle peridotite at low temperatures. *Philosophical Transactions of the Royal Society A*, *378*(2165), 20180433. doi: 10.1098/rsta.2018.0433

Ellison, E. T., Templeton, A. S., Zeigler, S. D., Mayhew, L. E., Kelemen, P. B., Matter, J. M., & Party, T. O. D. P. S. (2021). Low-temperature hydrogen formation during aqueous alteration of serpentinized peridotite in the samail ophiolite. *Journal of Geophysical Research: Solid Earth*, *126*(6), e2021JB021981. doi: 10.1029/2021JB021981

Falk, E. S., & Kelemen, P. B. (2015). Geochemistry and petrology of listvenite in the samail ophiolite, sultanate of oman: Complete carbonation of peridotite during ophiolite emplacement. *Geochimica et Cosmochimica Acta*, *160*, 70-90. doi: https://doi.org/10.1016/j.gca.2015.03.014

Friedman, J. H. (2002). Stochastic gradient boosting. *Computational statistics & data analysis*, *38*(4), 367–378. doi: 10.1016/S0167-9473(01)00065-2

Früh-Green, G., Orcutt, B., Green, S., Cotterill, C., & Scientists, E. . (n.d.). *Proceeding IODP Expedition 357* (Tech. Rep.). International Ocean Discovery Program. Retrieved from https://doi.org/10.14379/iodp.proc.357.102

.2017

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

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

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

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

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

Guimerà, R., Reichardt, I., Aguilar-Mogas, A., Massucci, F. A., Miranda, M., Pallarès, 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

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

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

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

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

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

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

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

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

Kelemen, P. B., & Matter, J. (2008). In situ carbonation of peridotite for co¡sub¿2¡/sub¿ storage. *Proceedings of the National Academy of Sciences*, *105*(45), 17295-17300. doi: 10.1073/pnas.0805794105

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

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

H. Larochelle, M. Ranzato, R. Hadsell, M. Balcan, & H. Lin (Eds.), *Advances in neural information processing systems* (Vol. 33, pp. 9459–9474). Curran Associates, Inc. Retrieved from https://proceedings.neurips.cc/paper_files/paper/2020/file/6b493230205f780e1bc26945df7481e5-Paper.pdf

Li, Z., Ji, J., & Zhang, Y. (2022). From kepler to newton: Explainable AI for science discovery. In *Icml 2022 2nd ai for science workshop.* Retrieved from https://openreview.net/forum?id=vA9hti-Fi7H

Lu, B., & Torquato, S. (1992). Lineal-path function for random heterogeneous materials. *Physical Review A*, *45*(2), 922.

MacLeod, C., Dick, H., Blum, P., & Scientists, E. . (2017). *Proceedings IODP Expedition 360* (Tech. Rep.). International Ocean Discovery Program. Retrieved from https://doi.org/10.14379/iodp.proc.360.102.2017

Malvoisin, B., Podladchikov, Y. Y., & Myasnikov, A. V. (2021). Achieving complete reaction while the solid volume increases: A numerical model applied to serpentinisation. *Earth and Planetary Science Letters*, *563*, 116859. doi: https://doi.org/10.1016/j.epsl.2021.116859

Malvoisin, B., Zhang, C., Müntener, O., Baumgartner, L. P., Kelemen, P. B., & Party, O. D. P. S. (2020). Measurement of volume change and mass transfer during serpentinization: Insights from the oman drilling project. *Journal of Geophysical Research: Solid Earth*, *125*(5), e2019JB018877. doi: https://doi.org/10.1029/2019JB018877

McBeck, J. A., Aiken, J. M., Mathiesen, J., Ben-Zion, Y., & Renard, F. (2020). Deformation precursors to catastrophic failure in rocks. *Geophysical Research Letters*, *47*(24), e2020GL090255. doi: 10.1029/2020GL090255

Miller, G. A. (1995, nov). Wordnet: a lexical database for english. *Commun. ACM*, *38*(11), 39–41. doi: 10.1145/219717.219748

Nori, H., Lee, Y. T., Zhang, S., Carignan, D., Edgar, R., Fusi, N., ... Horvitz, E. (2023). *Can generalist foundation models outcompete special-purpose tuning? case study in medicine.*

Pallister, J. S., & Knight, R. J. (1981). Rare-earth element geochemistry of the samail ophiolite near ibra, oman. *Journal of Geophysical Research: Solid Earth*, *86*(B4), 2673-2697. doi: https://doi.org/10.1029/JB086iB04p02673

Plümper, O., & Matter, J. (2023, 06). Olivine—The Alteration Rock Star. *Elements*, *19*(3), 165-172. doi: 10.2138/gselements.19.3.165

Plümper, O., Røyne, A., Magrasó, A., & Jamtveit, B. (2012, 12). The interface-scale mechanism of reaction-induced fracturing during serpentinization. *Geology*, *40*(12), 1103-1106. doi: 10.1130/G33390.1

Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V., & Gulin, A. (2018). Catboost: unbiased boosting with categorical features. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, & R. Garnett (Eds.), *Advances in neural information processing systems* (Vol. 31). Curran Associates, Inc. Retrieved from https://proceedings.neurips.cc/paper_files/paper/2018/file/14491b756b3a51daac41c24863285549-Paper.pdf

Ravaut, P., Bayer, R., Hassani, R., Rousset, D., & Yahya'ey, A. (1997). Structure and evolution of the northern oman margin: gravity and seismic constraints over the zagros-makran-oman collision zone. *Tectonophysics*, *279*(1), 253-280. (The Adolphe Nicolas Volume) doi: https://doi.org/10.1016/S0040-1951(97)00125-X

Schmidt, M., & Lipson, H. (2009). Distilling free-form natural laws from experimental data. *Science*, *324*(5923), 81-85. doi: 10.1126/science.1165893

Scicchitano, M. R., de Obeso, J. C., Blum, T. B., Valley, J. W., Kelemen, P. B., Nachlas, W. O., ... Roddatis, V. (2023). An empirical calibration of the serpentine-water oxygen isotope fractionation at t=25–100°c. *Geochimica et Cosmochimica Acta*, *346*, 192-206. doi: https://doi.org/10.1016/j.gca.2023.02.015

Sharma, A., Lysenko, A., Boroevich, K. A., Vans, E., & Tsunoda, T. (2021, 08). DeepFeature: feature selection in nonimage data using convolutional neural network. *Briefings in Bioinformatics*, *22*(6), bbab297. doi: 10.1093/bib/bbab297

Teagle, Alt, Umino, Miyashita, Banerjee, Wilson, & Scientists, E. . (2006). *Proceedings IODP Expedition 309/312* (Tech. Rep.). Integrated Ocean Drilling Program Management International, Inc. Retrieved from https://doi.org/10.2204/iodp.proc.309312.2006

Teagle, D., Ildefonse, B., Blum, P., & Scientists, E. . (2012). *Proceedings IODP Expedition 335* (Tech. Rep.). ntegrated Ocean Drilling Program Management International, Inc. Retrieved from https://doi.org/10.2204/iodp.proc.335.2012

Wang, H., Fu, T., Du, Y., Gao, W., Huang, K., Liu, Z., ... others (2023). Scientific discovery in the age of artificial intelligence. *Nature*, *620*(7972), 47–60. doi: 10.1038/s41586-023-06221-2

Wang, Y. D., Blunt, M. J., Armstrong, R. T., & Mostaghimi, P. (2021). Deep learning in pore scale imaging and modeling. *Earth-Science Reviews*, *215*, 103555. doi: https://doi.org/10.1016/j.earscirev.2021.103555

Whitney, D. L., & Evans, B. W. (2010). Abbreviations for names of rock-forming minerals. *American Mineralogist*, *95*(1), 185–187. doi: 10.2138/am.2010.3371

Zhao, W. X., Zhou, K., Li, J., Tang, T., Wang, X., Hou, Y., ... others (2023). A survey of large language models. *arXiv preprint arXiv:2303.18223*. doi: 10.48550/arXiv.2303.18223

Zou, H., & Hastie, T. (2005, 03). Regularization and Variable Selection Via the Elastic Net. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *67*(2), 301-320. doi: 10.1111/j.1467-9868.2005.00503.x