A Review of Orebody Knowledge Enhancement Using Machine Learning on Open-Pit Mine Measure-While-Drilling Data

Daniel M. Goldstein *, Chris Aldrich and Louisa O’Connor

Western Australian School of Mines: Minerals, Energy and Chemical Engineering, Curtin University, GPO Box U1987, Perth, WA 6845, Australia; chris.aldrich@curtin.edu.au (C.A.); louisa.oconnor@riotinto.com (L.O.)

* Correspondence: 20878119@student.curtin.edu.au

Abstract: Measure while drilling (MWD) refers to the acquisition of real-time data associated with the drilling process, including information related to the geological characteristics encountered in hard-rock mining. The availability of large quantities of low-cost MWD data from blast holes compared to expensive and sparsely collected orebody knowledge (OBK) data from exploration drill holes make the former more desirable for characterizing pre-excavation subsurface conditions. Machine learning (ML) plays a critical role in the real-time or near-real-time analysis of MWD data to enable timely enhancement of OBK for operational purposes. Applications can be categorized into three areas, focused on the mechanical properties of the rock mass, the lithology of the rock, as well as, related to that, the estimation of the geochemical species in the rock mass. From a review of the open literature, the following can be concluded: (i) The most important MWD metrics are the rate of penetration (rop), torque (tor), weight on bit (wob), bit air pressure (bap), and drill rotation speed (rpm). (ii) Multilayer perceptron analysis has mostly been used, followed by Gaussian processes and other methods, mainly to identify rock types. (iii) Recent advances in deep learning methods designed to deal with unstructured data, such as borehole images and vibrational signals, have not yet been fully exploited, although this is an emerging trend. (iv) Significant recent developments in explainable artificial intelligence could also be used to better advantage in understanding the association between MWD metrics and the mechanical and geochemical structure and properties of drilled rock.

Keywords: measure-while-drilling (MWD); logging-while-drilling (LWD); open-pit mining; subsurface characterization; machine learning (ML); data acquisition; rock properties; geochemical analysis; artificial intelligence (AI); predictive modelling; feature importance

1. Introduction

Measurement while drilling (MWD) originated from Schlumberger’s downhole electrical logging system in 1911, which was exclusively successful within the oil industry [1]. They described MWD as a system of sensors collecting performance data during rock drilling of open-pit mining production blast holes that can be correlated with pre-excavation subsurface conditions. Using MWD as a characterization system is desirable due to the low cost and high resolution of blast holes. On the other hand, exploration drill holes are expensive and are rarely collected in comparison.

While MWD was introduced to the open-pit mining environment in the 1970s, its use to characterize the subsurface was limited due to the high volume of analogue data to manually analyse. To address this limitation, computerized data acquisition of MWD variables to broadly determine lithological boundaries in open-pit mining was introduced around the 1980s [2–7]. Evolving from this initial use of computerized MWD rock recognition, this literature review focuses on the progression of advancing analytics, including machine learning (ML) algorithms, in MWD data to determine the geotechnical, geological, and geochemical subsurface characteristics before open-pit mining.
To perform the above correlations, multiple MWD metrics are collected by researchers. Depending on the manufacturer system setup, MWD datapoints are generated as time series or depth series in blast holes ranging from 10 m deep in surface iron ore mines to over 60 m in open-pit coal mines [8]. The most common variables acquired in open-pit mining MWD systems are time, depth, rate of penetration (rop), torque or rotational pressure (tor), bit air pressure (bap; also called flushing air pressure), weight on bit (wob; also called weight on rods, thrust, or feed pressure), and rotary speed or revolutions per minute (rpm) [9]. Other less commonly collected variables are the drilling acoustic changes and vibrations of the drill rods [10].

Due to the continuous nature of drill and blast cycles for excavation in open-pit operations, enormous quantities of MWD datapoints are collected. Due to this sheer volume of data to analyse, researchers have recently begun applying artificial intelligence to MWD datasets to understand nonlinear trends between drill responses and subsurface composition. For instance, a single productive rig at an open-pit iron ore mine may drill around one hundred 10–12 m blast holes per day, generating a minimum of 10,000 MWD observations with several responses per observation, including rop, tor, wob, bap, and rpm [11]. Many major iron ore mines employ around a dozen simultaneously operating blast drills. Despite the abundance of data, as shown in Section 2 most studies have focused on broad classification of rock types or hardness using rop to denote lithological boundaries rather than detailed characterization of each rock type. In industry, mine technical services professionals are not necessarily appropriately skilled to analyse large volumes of data in the brief time period between drilling blast holes and loading of explosives.

The objective of this study is to present a systematic and thorough examination of the existing body of research pertaining to the utilization of ML algorithms in the analysis of MWD data for the purpose of characterizing rock in mining operations. Furthermore, this study incorporates a comprehensive analysis of algorithms employed for assessing the significance of each metric related to the MWD technique, in relation to variations in the subsurface, such as rock strength, fracture frequency, elemental composition, and density. Additionally, it explores the potential of ML algorithms and their prospective utilization in forthcoming applications. Figure 1 presents a flowchart that describes the research framework of this study. Section 2 presents a concise summary of the dissemination of scholarly literature through journals and the prevailing patterns in publications. Section 3 details the development of ML models from MWD data. In Section 4, this study presents the pragmatic implementations of data analysis derived from MWD techniques in the domains of geology identification using density; gamma; magnetic susceptibility and resistivity responses; assessment of rock mass characteristics, including rock strength, fracture frequency and rock mass classification scores; and geochemical composition, such as iron percentage as well as primary and secondary contaminants. These applications are expounded upon in Sections 4.1–4.3, respectively. Section 5 presents an exposition on the challenges and potential associated with the application of ML techniques for rock characterization utilizing MWD data. The concluding remarks are presented in Section 6.

Figure 1. Research framework for review of machine learning (ML) of mining MWD data.
2. Literature Sources and Dissemination

As indicated by the literature review process, the publication of papers on this topic spans around thirty years. A total of 537 research articles mentioning “Measure-While-Drilling” or “Measurement-While-Drilling” have been indexed in Google Scholar (GS) from the late 1980s to the present, as shown in Figure 2’s yellow bars. Nonetheless, when the search results are filtered to studies employing MWD to strictly characterize the rock being drilled, there are just 129 publications. The 129 articles are organized according to the year of publication, as shown in Figure 2’s green bars. Indeed, the use of MWD data for rock characterization is limited in the first decade of the twenty-first century, whereas the number of papers has expanded dramatically in the last decade, with twenty-six publications expected in 2022. When the search parameters are restricted to using MWD to strictly characterize the rock being drilled using ML approaches, there are just 32 publications with a rapid employment of ML algorithms since 2016, shown in blue in Figure 2.

![Figure 2. Annual distribution of published journal papers mentioning MWD for rock characterization in all excavation types using ML from 1997 to 2022.](image)

To address readers’ academic interests in characterizing rock types using ML techniques on MWD data in open-pit mining, more journal articles than conference papers have been published. While master’s and doctorate dissertations were not featured on GS, a few of them have been published [12–17]. The leading journals are Mathematical Geosciences and Minerals, which have published over 22% of the discoveries concentrating on the use of ML on MWD data for open-pit mining rock characterization. Each of the remaining journals has published one paper.

Several ML techniques were applied to characterize rock conditions using MWD data in the reports identified with the GS search parameters. There are generally two types of ML techniques used in those findings: neural networks (NNs) [12,16,18–20] and Gaussian process (GP) [8,11,21]. Other approaches, including support vector machines (SVMs), random forests (RFs), boosting, self-organizing maps (SOMs), and fuzzy logic, have been compared to NNs in various studies [19,20]. Generally, the selection of any particular ML model would depend on its predictive power, its interpretability, and, to some extent, also the complexity of the data to be processed. Relevant studies to ML analysis of open-pit mining MWD data is presented in Table 1.
### Table 1. Summary of ML applications on MWD data for open-pit mining rock characterization.

<table>
<thead>
<tr>
<th>Method</th>
<th>Commodity</th>
<th>Holes</th>
<th>Target</th>
<th>Comment **</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>HOP</td>
<td>Coal</td>
<td>35</td>
<td>Rock Type</td>
<td>Input of full-waveform sonic velocity</td>
<td>[22]</td>
</tr>
<tr>
<td>LB</td>
<td>Iron Ore</td>
<td>28</td>
<td>Rock Type</td>
<td>2D analysis; drill data averaged per hole</td>
<td>[20]</td>
</tr>
<tr>
<td>FIS</td>
<td>Iron Ore</td>
<td>28</td>
<td>Rock Type</td>
<td>2D analysis; drill data averaged per hole</td>
<td>[20]</td>
</tr>
<tr>
<td>GP</td>
<td>Iron Ore</td>
<td>31</td>
<td>Rock Type</td>
<td>Labels exclusive of 1 m from contact</td>
<td>[8]</td>
</tr>
<tr>
<td>GP</td>
<td>Iron Ore</td>
<td>N/A</td>
<td>Rock Type</td>
<td>BIF and shale in Brockem Formation</td>
<td>[11]</td>
</tr>
<tr>
<td>GP</td>
<td>Iron Ore</td>
<td>N/A</td>
<td>Rock Type</td>
<td>CWT preprocessing of MWD data</td>
<td>[8]</td>
</tr>
<tr>
<td>LoR</td>
<td>Marble</td>
<td>204</td>
<td>Grade</td>
<td>6 classes of marble quality tested</td>
<td>[23]</td>
</tr>
<tr>
<td>NN</td>
<td>Coal</td>
<td>17</td>
<td>Rock Type</td>
<td>Backpropagation</td>
<td>[16]</td>
</tr>
<tr>
<td>NN</td>
<td>Iron Ore</td>
<td>33</td>
<td>Rock Type</td>
<td>Feedforward and Backpropagation</td>
<td>[12]</td>
</tr>
<tr>
<td>NN</td>
<td>Iron Ore</td>
<td>28</td>
<td>Rock Type</td>
<td>2D analysis; drill data averaged per hole</td>
<td>[20]</td>
</tr>
<tr>
<td>RF</td>
<td>Coal</td>
<td>35</td>
<td>Rock Type</td>
<td>ANN compares thresholding of SEM</td>
<td>[18]</td>
</tr>
<tr>
<td>RF</td>
<td>Marble</td>
<td>204</td>
<td>Grade</td>
<td>6 classes of marble quality tested</td>
<td>[23]</td>
</tr>
<tr>
<td>GP</td>
<td>Iron Ore</td>
<td>7000</td>
<td>Grade</td>
<td>Iron percentage, Phosphorous, Sulphur, Aluminum Oxide, Silicon Dioxide</td>
<td>[24]</td>
</tr>
<tr>
<td>RF</td>
<td>Iron Ore</td>
<td>33</td>
<td>Rock Type</td>
<td>To investigate specific fracture energy</td>
<td>[12]</td>
</tr>
<tr>
<td>RF</td>
<td>Iron Ore</td>
<td>302</td>
<td>Rock Type</td>
<td>Used for individual index calculation</td>
<td>[25]</td>
</tr>
<tr>
<td>PCA</td>
<td>Coal</td>
<td>17</td>
<td>Rock Type</td>
<td>Used as validation of 3 input features</td>
<td>[16]</td>
</tr>
<tr>
<td>PCA</td>
<td>Iron Ore</td>
<td>33</td>
<td>Rock Type</td>
<td>To investigate specific fracture energy</td>
<td>[12]</td>
</tr>
<tr>
<td>PCA</td>
<td>Iron ore</td>
<td>302</td>
<td>Rock Type</td>
<td>Used for individual index calculation</td>
<td>[25]</td>
</tr>
<tr>
<td>PCA</td>
<td>Marble</td>
<td>4</td>
<td>Rock Type</td>
<td>Intrusion detection of marble</td>
<td>[26]</td>
</tr>
<tr>
<td>AS</td>
<td>Coal</td>
<td>35</td>
<td>Rock Type</td>
<td>Seam roof detection to cease drilling</td>
<td>[27]</td>
</tr>
<tr>
<td>KMC</td>
<td>Coal</td>
<td>17</td>
<td>Rock Type</td>
<td>4 class clusters identified</td>
<td>[16]</td>
</tr>
<tr>
<td>GP</td>
<td>Iron Ore</td>
<td>138</td>
<td>Rock Type</td>
<td>2D analysis; drill data averaged per hole</td>
<td>[22]</td>
</tr>
<tr>
<td>GP</td>
<td>Iron Ore</td>
<td>14,340</td>
<td>Grade</td>
<td>Gaussian kernel</td>
<td>[28]</td>
</tr>
<tr>
<td>HMM</td>
<td>Marble</td>
<td>3</td>
<td>Rock Type</td>
<td>Sudden material type transition</td>
<td>[29]</td>
</tr>
<tr>
<td>NN</td>
<td>Iron Ore</td>
<td>33</td>
<td>Rock Type</td>
<td>Simple competitive learning</td>
<td>[12]</td>
</tr>
<tr>
<td>SOM</td>
<td>Iron Ore</td>
<td>33</td>
<td>Rock Type</td>
<td>3 class maps for 3 major rock types</td>
<td>[12]</td>
</tr>
<tr>
<td>SVM</td>
<td>Iron Ore</td>
<td>14,340</td>
<td>Grade</td>
<td>2D analysis; drill data averaged per hole</td>
<td>[28]</td>
</tr>
</tbody>
</table>


For example, while linear models or simple decision trees may not have the same predictive power as multilayer perceptrons or support vector machines, it may be easier to interpret the results generated by these models. Likewise, deep learning models, such as convolutional neural networks or vision transformers may be better able to analyze more complex signals and images than traditional multilayer perceptrons or random forests and can also be coupled with explanatory models, where required. Further investigations of these trade-offs could yield significant advances in the field.

### 3. Development of Machine Learning Models from MWD Data

The development of machine learning models comprises multiple discrete stages, depicted in Figure 3. The process begins with data in its unrefined state, a phase commonly known as data preprocessing (Figure 3A). Here, the focus is on improving the consistency and integrity of the data. In the case of sequential data, such as time-series data, it is often necessary to adjust or eliminate identifiable patterns and trends. Furthermore, to provide a cohesive basis for the following stages, any discrepancies resulting from different intervals of data recording are reconciled.

Feature engineering is the segment that follows preprocessing. In this phase, data is transformed into a form that computer algorithms can easily understand (Figure 3B). When confronted with complex data types, such as images or sound signals, it is necessary to perform transformations to convert them into a structured format. Depending on the ML model, a crucial aspect of this stage could entail reducing redundancy by replacing...
groups of interconnected variables with a more succinct set that encompasses most of the initial information. After suitably structuring the data, it is segregated into distinct subsets known as training, validation, and test datasets (Figure 3C). The training data functions as the fundamental dataset upon which the model is initially constructed. The utilisation of a validation set facilitates the process of refining the model, whilst the test set is specifically put aside to act as a benchmark for assessing the model’s performance in real-world situations.

![Figure 3. A general methodology for the building of ML models for the interpretation of data, including data preprocessing (A), feature engineering (B), structuring of training, validation and test sets (C) model validation (D), model optimisation (E), construction of a complementary explanatory model (F) validation of new data (G) prior to implementation of the model.]

After data partitioning, the focus shifts to the training of the model itself (Figure 3D). There is a wide array of models available for this purpose, encompassing neural networks, decision trees, and support vector machines. The selection of a model is primarily contingent upon the characteristics of the data and the goals of the research. If the proposed model’s performance is not acceptable, then the model’s hyperparameters can be tuned (Figure 3E), then trained, validated, and tested (Figure 3C) before acceptable predictions can be made. In situations where it is necessary to explain the reasoning behind a model’s decisions, particularly in diagnostic contexts, an additional explanatory model can be developed (Figure 3F). Nevertheless, if the decisions made by the primary model possess intrinsic transparency, this step may be deemed unnecessary.

Ultimately, when novel datasets are introduced, they undergo identical preliminary phases of preprocessing and feature engineering as the initial training data (Figure 3G). An essential component of this phase involves determining the statistical similarity between the newly acquired data and the first training dataset, although this is not necessarily always done in practice. Several analytical techniques, such as principal component analysis and t-stochastic neighbour approaches, offer valuable insights into this congruence. If a noticeable discrepancy exists between the datasets, it may be necessary to recalibrate the model.

**4. Practical Applications**

4.1. Geology Recognition from MWD

This part pertains to the utilization of ML techniques on MWD data to understand the intricate relationships between drilling metrics and forecast subsurface geological
conditions before commencing mining operations. For example, a banded iron formation (BIF) deposit might have classification zones of shale and BIF. These broad categorization of geological zones from MWD data are useful for identifying the contacts between highly contrasting material types to update a deposit-scale model and increase its accuracy within localized areas [8]. This increased accuracy leads to improved blasting outcomes, in terms of achieving fragmentation around geological boundaries. However, these nongranular assortments do not necessarily provide resolution within each lithological unit that are required to optimize mining operations downstream of the drill and blast process, including excavation, haulage, and beneficiation.

First, an overview of the various types of geological deposits in which MWD data have been used to identify geology in open-pit mining operations is presented. The next section critically reviews the types of analysis performed by researchers to understand which MWD metrics are important to describe changing subsurface conditions from a data science perspective, including lithology, density, rock strength, and weathering or fracture intensity. Several authors have used principal component analysis (PCA) to interpret feature importance [12,16,25]. However, it is worth noting that PCA is generally not an effective approach for rating the significance of input variables. The final section details the application of ML to categorize rock types from MWD data. The classification methods employed by previous studies can be improved upon gaining a granular comprehension of the changing subsurface geological conditions.

4.1.1. Mined Commodities

Most researchers delineating lithologies from MWD data in open-pit mining have focused on sedimentary commodities, predominantly iron ore and coal. In these deposits, the resource geology has been reasonably defined by exploration drilling at hole spacings of 50 m grids in iron ore and 100–200 m grids in coal. The use of MWD data attempts to add local accuracies in between the exploration drillholes through blast holes located at approximately 5 m burden and spacing. Most of the research to recognize subsurface geology from drilling metrics has taken place in Canada [7,12,16] and Australia [8,11,18,20–22,27]. These countries have had strong research collaboration between the mining industry and universities.

The initial findings to link six open-pit blast hole drilling responses with the subsurface geology via the use of downhole geophysical responses was the basis for later work [7]. Nearly 20 years later, postgraduate research applied NN for open-pit geological classification from MWD datasets in coal and iron ore mining blast holes [12,16]. Since these earlier Canadian-based studies, research on open-pit classification of geology has been entirely within Australian coal and iron ore deposits. The research in coal geology recognition aimed to accurately predict coal roof locations to prevent blast damage in open-pit mining with several reports presenting the results of various ML classification methods on the same MWD dataset from 35 blast holes in the Australian Hunter Valley coal region [18,22,27]. A few years later, coal seams in six gas wells were identified with 96% accuracy from the Surat Basin using five categorization algorithms [19]. In the iron ore industry, various classification methods were used on datasets of 28 holes and approximately 120 holes were utilized [20,21]. Both studies successfully distinguished BIF rock from shale in the Pilbara region to improve fragmentation by tailoring explosive loading for each rock unit. Nearly a decade later, multiple findings from the University of Sydney presented successful classification of BIF and shale units from a significantly larger dataset of several thousand blast holes [8,11].

While the majority of rock type recognition has occurred in open-pit mining in Canada and Australia, examples from civil engineering, tunnelling, oil and gas, and other mining industries have also been investigated. MWD data were used to distinguish geological zones in an urban construction project from Hong Kong. In the tunnelling industry, MWD data were also used to predict geology ahead of blasting in the Norwegian Loren, Swedish Stockholm Bypass, and the Chinese Jiuding Shan projects [30–32]. The use of MWD data within the oil and gas sector has demonstrated their efficacy in the classification of lithology
within narrow targets [33–36]. This classification aids to attain precise directional drilling, hence optimizing the selection of well locations. In the underground coal mining industry, linear correlation and ML algorithms were applied to MWD data from roof support drills to determine geological zones and adjust locations and spacings of rock bolts and cable bolts to improve ground support [37–40]. In the open-pit quarry industry, successful use of MWD metrics to observe sill and marble boundaries was reported in four blast holes and altered explosive loading practices to achieve optimal fragmentation, which prevented unnecessary rock breaking off boulders from under-blasting [29]. More recently, six classes of marble qualities were predicted using Logistical Regression and RF [23]. All the above examples establish the wider applicability of determining geology from MWD data outside of open-pit mining of sedimentary deposits.

4.1.2. MWD Metrics

MWD metrics (rop, tor, fob, wob, rpm, etc.) have shown varying levels of feature importance when correlated with subsurface geological conditions [41]. Feature importance refers to methods that calculate a score for each of the MWD metrics in a particular model. The resulting scores represent the contribution of each feature in the prediction of the target variable.

A high score indicates that the characteristic will have a greater influence on the model used to forecast a particular subsurface geological variable. Initial manual attempts to interpret the feature importance of open-pit mining MWD variables with wireline gamma responses identified local minimum and maximum variations; the changes in rop, tor, wob, and SED parameters were associated with different lithological factors [7]. They also reported that rpm did not significantly correlate with any rock type, due to being controlled by the operator at a relatively constant rate. However, these univariate experiments each focused on an isolated, individual drilling parameter. Indeed, all other drilling metrics were required to be maintained near constant in these studies, which is unrealistic for common production blast hole drilling. An example of a surface blast hole drill rig is presented in Figure 4 (left), together with the typical MWD variables collected (Figure 4, right).

![Diagram](image-url)

**Figure 4.** Typical layout of a surface mining blast drill rig (Type: Epiroc D65) (left) and MWD variables (right).
Early attempts to manually interpret the feature importance for each MWD input metric on rock type recognition models have given way to more advanced analytical methods, primarily principal component analysis (PCA). PCA aims to show patterns in multivariate data by reducing the dimensionality of these datasets and creating new variables, called principal components [42]. The principal components are a set of orthogonal vectors that are linear combinations of the input variables and best describe which of these inputs represent the most variation to the data. The rop and tor were identified as the MWD metrics that were most closely linked with subsurface geological conditions, while rpm was the least significant [12,16]. Both researchers identified feature importance by using the loading plots that detailed the most variation (the first and second principal components), although this may not necessarily account for most of the variation in the target variable. PCA has also been applied to determine coal vs. non-coal rock types by including gamma response and hole diameter as inputs along with the MWD metrics of rop, wob, and tor [19]. This work also applied a fit-for-purpose feature importance algorithm based on random forests and determined that the rate of penetration was the single most important MWD metric to classify coal vs. non-coal rock types in the investigated holes. The wob, tor, and rpm were presented as relatively insignificant compared to rop [19].

Several studies introduced derived drilling metrics calculated from the collected MWD variables to determine if derived features were more important than the raw drilling metrics. Modulated specific energy (MSE) takes advantage of the variations in the ratio of tor to wob to identify rock type in coal blast holes [18]. SEM is calculated by modifying the specific energy of drilling (SED) with a rotational work fraction, which is determined by dividing the tor by the sum of the tor and the wob. They reported SEM as an important feature from the drilling metrics to detect boundaries of multiple coal seams amongst predominantly sandstone units. A second novel drilling parameter called adjusted penetrate rate (APR) was developed to account for variations between drilling operators in manually operated rigs as well as between manual and autonomous drilling rigs in iron ore blast holes [21]. APR is calculated by dividing the rop by the product of the wob times the square root of tor. Their finding reported APR was a more important feature than SED for the categorization of iron ore rock types of BIF waste, mineralized BIF and shale. rule-based labelling of geology from multiple drilling metrics outperformed the univariate APR method of rock type recognition [8]. However, not all variables are consistently collected in open-pit iron ore production drilling conditions due to sensor faults or breakdowns [8]. As a result, the missing data have resulted in up to 90% of blast hole observations not being able to be used for APR and SED.

Determination of feature importance reveals which variables are essential to collect for accurate prediction of rock type. Conversely, feature importance should identify which features may be discarded during selection of variables to include for subsequent analysis using ML algorithms. As mentioned previously, the use of PCA on MWD data for the selection of features with the highest variation as the most important predictors may not be effective. In contrast, developments in interpretable and explainable AI have opened significantly more advanced approaches for this purpose. This could include Shapley value regression, permutation analysis, interpretation of model structures, kernel SHAP, LIME, etc., and in the case of deep learning models, also even direct interpretation of image or signal data, where applicable [43–45].

4.1.3. Machine Learning Classification

Classification of geology using ML methods on open-pit blast hole MWD data has been demonstrated on mostly small datasets of a few dozen holes. Feature engineering has consistently been applied to clean noisy MWD data, including filtering, smoothing, normalizing, and removing outliers to prepare the raw data for ML analysis. Although a range of different ML models have been used, multilayer perceptrons and Gaussian process regression have featured prominently in recent studies, as discussed in Section 2.
NN algorithms have demonstrated the probability of geological recognition in both iron ore and coal deposits. The classification accuracy for five rock types reached 95% using back propagation neural network (BPNN) algorithms on 17 blast holes from a Canadian coal mine [16]. The classification accuracy for three rock types in an iron ore deposit averaged 57% based on 33 blast holes in the United States [12]. One study used NN as an assessment tool to determine if the SEM metric was superior to APR to identify lithologies, while another study demonstrated a prediction accuracy of 96% using NN to classify coal and non-coal rock types [18,19]. A multilayer perceptron NN algorithm classified four BIF rock types at 78% accuracy [20].

GPs have been exclusively used at the Pilbara BIF iron ore mines to classify geology. GP clustering of hardness values (approximated by regression modelling of APR) was used to categorize three types of iron ore units (waste BIF, mineralized ore, and shale) [21]. The figures presented are two-dimensional, indicating the researchers used an average MWD value for each hole to classify geology. This two-dimensional method provides a singular rock type, and thus, a limited perspective for mine planning. In a novel three-dimensional method, MWD datapoints were labelled with categorized units from downhole gamma responses [11]. The GP approximations resulted in a thresholding strategy that classified two BIF units in the Dales Gorge Member in three blast patterns at 87%, 84%, and 81% accuracy, respectively. GP-derived results were reported as only 62% accurate to classify two rock types (unmineralized shale and mineralized iron ore) [8]. They attributed this poor performance to noisy data requiring further data cleaning.

The use of NN and GP algorithms on MWD metrics has been largely successful in classifying rock types. Several findings reported other ML algorithms, including fuzzy inference systems (FISs), SVM, random forest (RF), and boosting to compare results with NNs, but none demonstrated significantly better performance to NNs [19,20]. Despite ML classification accuracy being greater than 90%, the prediction of a handful of broad geological categories does not adequately capture subtle differences within each geological unit. An increased resolution of rock type (including subcategories, such as massive, fractured, or deformed rock) is required for informed and effective mine planning.

4.2. Rock Mass Properties from MWD

Rock mass quality has been correlated with drilling response data [41]. As with determining geological zones, the geotechnical properties are categorized rather than predicted from discrete values using ML analytical techniques. The main rock mass properties under investigation are rock strength, discontinuities or fractures, and rock mass categorization scores, such as, rock mass rating (RMR), the geological strength index (GSI), Q-system, and rock quality designation (RQD) [46–49]. These geotechnical conditions are traditionally determined by logging and laboratory testing of diamond drill cores from holes that are spaced hundreds of meters apart without any regular pattern. Rock mass characterization data collection from exploration drilling is sparse due to the expense of drilling. This low-resolution data capture requires interpolation and broad geotechnical domaining to classify a rock mass with significant uncertainty between holes. Surface mine ML applications on MWD data were generally focused on geotechnical characterization to improve fragmentation, or rock breakage from blasting [1,6,25,50]. The focus of understanding rock mass properties from MWD data in underground mining- and tunnelling-based research is to reduce strata failure by adjusting the spacing and locations of ground support equipment [13,51–59].

First, an overview of the various environments and mined commodities in which MWD data have been used to identify rock mass properties is presented. Next, the types of feature importance analysis conducted by researchers to understand which MWD metrics are useful to describe changing geotechnical conditions, such as strength and fracture state, are critically reviewed from a data science perspective. Finally, the application of ML in characterizing rock mass zones from MWD data is surveyed. As with the geology-based
findings in the previous section, these broad rock mass zone classifications are useful but do not accurately predict the complex fabric of a rock mass.

4.2.1. Commodities

As discussed in Section 2, there are considerably fewer studies on the ML classification of MWD data for rock mass characteristics in open-pit mining than for geological types. However, the use of ML algorithms on MWD data to characterize a rock mass has been demonstrated in other industries. These include tunnelling [51,53,54] and underground mining [52,57–60], which employ drilling prior to excavation. Exploration and laboratory-based activities [61–63] have also correlated rock mass properties with drilling metrics.

In a series of classic studies, Schunnesson’ (1990, 1996, 1997, 1998) drill monitoring was used in underground Scandinavian iron ore and copper mines to distinguish changing conditions in each investigated rock mass [57–60]. More than two decades later, potential blasting issues in the underground Swedish Malmberget iron ore mine were predicted by assessing trends in MWD data [52]. In the tunnelling space, rock mass classification zones, including RMR and rock mass quality, were predicted from intelligent MWD data analysis to improve ground support patterns [51,53,54].

Exploration drilling, and laboratory experiments also support the use of ML methods to predict geotechnical properties from drilling metrics. Rock strength could be predicted from an exploration drill rig’s MWD data in the seven holes of diamond core in Turkey [61]. In laboratory experiments involving drilling through prepared samples, it was demonstrated that rock properties, including density, porosity, P-wave velocity, Schmidt hardness, UCS, tensile strength, and elasticity modulus, could be successfully predicted from acoustics produced during the drilling of laboratory samples [62]. On the other hand, in the tunnelling space, simulated fractures were detected in the rock mass from MWD data trends during the excavation of a 20 m wide tunnel [53]. More recently, several Chinese tunnel excavations have used regression- and classification-based methods to determine rock properties from MWD data. NNs were used to predict rock strength values, including UCS, elasticity modulus, and Poisson’s ratio, from MWD data [64]. In terms of a classification approach, a bi-directional long short-term memory NN was used to predict trimodal rock mass classes [65].

However, few studies have attempted to characterize rock mass conditions from MWD data in the open-pit mining environment utilizing thresholding or simple correlation. It was observed that high and low rod correlated with weak and strong rock strength zones in 1267 blast holes from the Aitik copper mine in Sweden [66]. To determine fracture locations from MWD data, downhole geophysical televsioner logs were interpreted to correlate fracture locations in eight blast holes from the Canadian Highland Valley copper mine [50]. Also using televsioner logs to calibrate their findings, two novel rock description indexes were developed based exclusively on MWD data from 302 production blast holes at the Erzberg mine [25]. These systems categorized rock mass zones for structural condition and strength properties and were corroborated by highwall mapping via photogrammetrically generated models. However, no studies were identified that used ML analytical methods on MWD data to characterize subsurface geotechnical properties in open-pit operations.

4.2.2. Feature Importance

As with the determination of geological zones from MWD data, research has attempted to identify the feature importance that each MWD metric has in predicting subsurface geotechnical conditions. Determination of feature importance using advanced analytical methods has almost entirely taken place within underground mining and tunnelling operations [51,52,57]. Multiple studies have utilized manual correlation in open-pit mining to identify feature importance for drillability or blastability. These are outside the scope of this review due to their focus on other aspects that are not directly related to subsurface geotechnical conditions [1,6,14,55,66–68] (Ghosh et al., 2014; Khorzoughi, 2011; Liu and Yin, 2001; Manzoor et al., 2019; Navarro et al., 2018; Scoble and Peck 1987; Segui and
Higgins, 2001). PCA has only been used on open-pit mining MWD data to determine feature importance for a novel MWD-based rock mass classification system (drilling rock factor—DRF) [25].

PCA has been used to determine the feature importance of MWD metrics for geotechnical properties in underground mining and tunnelling projects. Specifically, the features ranking highly on the first principal component have generally been interpreted to be the most important. As discussed, the variables with the most variation are not necessarily the most important (Section 4.1). Schunnesson determined that the rop was most important from PCA loading plots showing rop as low in higher rock strength zones and, conversely, the rop was observed to be higher in weaker rock strength zones [57]. A later study focusing on the link between rock mass conditions and chargeability issues in underground blast holes employed PCA to identify ropS, torS, rop, and tor highly in terms of the first principal component, accounting for 62% of data variability [52]. Using a combination of PCA and ordered weighted average (OWA), Galende-Hernández et al. determined the top three features to select for predictive analysis were wob, damper pressure, and tor [51]. The inconsistent feature importance results between these two findings may be attributed to discrepancies with data collection, data cleaning equipment, or geological settings. It is also likely that the results do not support one another because PCA is not a statistically appropriate method for feature selection.

PCA has also been used on an open-pit mining MWD dataset from 286 blast holes at the Erzberg Mine in central Europe to interpret the feature importance of drilling metrics for input to their DRF [25]. The DRF rock mass classification system is composed of a structural factor and strength-grade factors. The structural factor consists of huge, fractured, and extensively fractured zones, while the strength-grade factor has soft-waste, hard-waste, transition zone, and hard-ore categories. Navarro et al. reported that the metrics rop, ropS, and torS ranked highly on PC1, the first principal component, for the structural factor evaluation, while tor dominated the positive side of PC2, the second principal component [25]. The first two main components of the structural factor explain a combined 66.02% of the entire variation, with PC1 contributing 47.47% and PC2 contributing 18.55%. In contrast, for the strength-grade factor, rop and wob dominated the positive sides of PC1 and PC2, respectively. The results of the study demonstrate that the strength-grade PC1 outcomes can differentiate between two distinct regions characterized by varying rock strength. Specifically, the schisted sandstone exhibits a uniaxial compressive strength (UCS) value of 30 MPa, while the limestone displays a UCS value of 125 MPa [25]. These results correlate well with digital photogrammetry reconstructions of the pit walls after blasting and excavation [25]. However, the DRF results are heavily dependent on calibration with manually interpreted televiewer logs. In addition, since the rock factor zones are also based on the local iron ore geology of the Erzberg Mine, adapting DRF to different mineral deposits would require intensive re-calibration. A deposit-agnostic approach without manual calibration would be more readily adopted to determine geotechnical properties from open-pit mining MWD data.

4.2.3. Machine Learning Classification

Classification of rock mass properties using ML methods on open-pit blast hole MWD data has not been demonstrated, despite an exhaustive search. It was discovered that while many reports claim to characterize rock mass conditions using ML on MWD data, what was reported was often rock type, with an assumption of strong vs. weak UCS values for contrasting rock types. The results of these are discussed in Section 4. No studies were found that attempted to predict geotechnical properties using ML methods in the open-pit mining, underground mining, or construction industries.

4.3. Geochemical Properties from MWD

Very few studies attempt to predict geochemical properties from MWD data [24,25,57]. Commonly reported geochemical properties of ore included aluminium oxide (Al₂O₃),
calcium oxide (CaO), iron (Fe), total loss of ignition (LOI), magnesium oxide (MgO), manganese (Mn), phosphorus (P), sulfur (S), silica oxide (SiO$_2$), and titanium oxide (TiO$_2$). All findings reporting geochemical properties from mining MWD data were located within an iron ore deposit and listed grade or iron ore percentage.

Fe was reported as the only ore quality investigated [25,57]. Partial least squares (PLS) analysis was applied on MWD data from an underground mine and closely predicted iron content against assays of drill cuttings in areas of high and low phosphorus geology [57]. The ore grade prediction of Navarro et al. was split into two simple categories of (1) waste consisting of less than 20% iron and (2) ore consisting of greater than 20% iron [25]. These waste and ore class results appear equivalent to the DRF results. However, there is no explicit mention of the ore grade prediction accuracy, which suggests it is probable that no validation of results was conducted against laboratory assays.

Only one study reported the use of ML classification for geochemical properties from MWD and drill cutting assay data from two Australian iron ore mines [24]. Unlike previous studies that only reported the investigated ore properties were Al$_2$O$_3$, CaO, Fe, LOI, MgO, Mn, P, S, SiO$_2$, and TiO$_2$. Fe was predicted with acceptable R$^2$ values of 0.79, 0.79, and 0.78 in the first mine and 0.64, 0.64, and 0.63 in the second mine for the GP, SVM, and RF algorithms, respectively. The datasets were combined to predict P and S. Acceptable R$^2$ values for P were 0.79, 0.78, and 0.81, and S were 0.91, 0.90, and 0.92 for the GP, SVM, and RF algorithms, respectively. The report also identified that several of the geochemical properties correlated with each other. Based on this, cross-assay predictive models showed improved results for the remaining ore properties. For example, the R$^2$ value for Al$_2$O$_3$ of 0.65 on MWD data alone in an RF model was improved to 0.90 and 0.85 when adding the cross-assay results for Fe and SiO$_2$, respectively. The primary benefit of this result is that if one assay can be directly measured with a sensor or strongly predicted from a model, these measurements can be utilized to enhance estimates for other geochemical properties.

5. Discussion
5.1. Current Challenges

This summary makes it clear that the motivation behind these findings is to circumvent the limitations of conventional subsurface characterization prior to excavation. NNs and GPs are the two primary categories of ML algorithms utilized with MWD data for rock categorization. While some artificial intelligence algorithms have been applied to the same datasets, different ML architectures produced different outcomes [18,22,27]. Thus, no model has been proposed as optimal for characterizing geological, geotechnical, and geochemical features from MWD data, as model selection depends on underlying aims, scientific objectives, and model restrictions. Notably, some ML approaches are more attractive (NN and RF) than others (GP) due to their computational efficiency [19]. In addition to there being no accepted, optimal ML modelling algorithm to characterize subsurface conditions from MWD data, there is no uniformly consistent approach to MWD data collection and processing.

There are still some challenges in data collection and processing that must be overcome. First, most of the investigated projects featured limited amounts of data from blast holes, making it difficult to extensively analyse and interpret the subsurface. Second, the quality of the output is dependent not just on the model but also on the condition of the gathered data. Quality assurance and quality control processes are not commonplace with open-pit MWD data acquisition. During the development of predictive models, the robustness of the model may be affected due to the existence of missing values or outliers within the dataset. In addition, the feature importance among MWD variables may also affect the modelling robustness. Existing studies [16,25,52] disagree as to which metrics should be utilized to characterize rock properties under different environmental situations, as they vary based on deposit-specific conditions. To identify the feature importance of drilling variables, PCA has been used to choose input variables from the MWD data. A statistically suitable strategy for determining feature relevance should enhance feature
selection to improve computational precision and speed. In conclusion, the subsurface characterization applications of ML techniques applied to MWD data are in an escalating stage of development.

In summary, the relative frequencies of machine learning methods used in the analysis of MWD data reported are shown in Figure 5. The analysis is based on a keyword search of the Scopus database. Interestingly, convolutional neural networks (tied with SVM at 16%) are comparatively high up the list, despite being a relative new approach focused on image analysis.

![Figure 5. Relative frequency of machine learning methods reported in the analysis of MWD data in the Scopus database (GP—Gaussian processes, SVM—support vector machines, CNN—convolutional neural networks, FUZ—fuzzy systems, PCA—principal component analysis, MLP—multilayer perceptron, LOG—logistic regression, BAY—Bayes modelling, RF—random forest).]](image)

5.2. Future Developments

Even though non-traditional approaches to comprehend mining subsurface conditions require further development, they are gaining popularity. ML methods can be utilized as a supplement to conventional theory [8,25]. They have the potential to extract insights about a vast array of spatial data, which can then be narrowed down to investigate areas of interest in greater detail. However, despite the widespread use of ML classification algorithms for categorizing geology in MWD data, ML analytical methods have seldom been applied. A common assumption has been that it is far more crucial to understand the boundaries between dissimilar rock types than the variation within each rock type [11,18,25]. As a result, the categorization of ‘strong’ and ‘weak’ rock types, such as BIF and shale, respectively, has dominated projects by investigating univariate relationships, mainly rop and tor [69]. In contrast, existing multivariate ML regression algorithms have significant prospects for solving the complicated challenge of variance within each rock type.

To interpret subsurface properties using MWD data, supervised [8,18,22,25] and unsupervised learning [12,16] techniques have been implemented. Supervised learning can use all available information to make predictions and classifications, while unsupervised learning may identify potential relationships and extract features from a substantial volume of unlabelled data. In the context of both supervised and unsupervised learning applications, it is essential to employ suitable data preprocessing approaches. These techniques encompass the removal of noise, outliers, and false attributes. This necessity arises from the considerable geographic variability observed in measurements while considering the data related to MWD and subsurface conditions [70].

Existing techniques of MWD data processing and analysis rely heavily on manual interpretation, such as determining the most critical rock-property-dependent metrics [31]. Consequently, it is difficult to repeatedly process data from new sources and places. A pertinent aim of applied research is therefore to design an automated workflow in which
ML methods and available input data, connected with specific conditions, can correlate, process, and select drilling variables for subsequent estimation of rock properties. In the interim, the gathering of a training dataset during algorithm execution could facilitate data collecting for future applications operating under comparable conditions.

The implementation of ML techniques is dependent on the training data set accurately representing the connection between the input datasets, MWD, and subsurface properties, respectively. It is crucial to select attributes for the training set that accurately represent the population. As a result, ML training progress is impeded by the challenges of overfitting, extended training time, and a notable inclination to become trapped in local minima [12]. As demonstrated in the tunnelling sector, these obstacles may be overcome with optimization algorithms, including genetic algorithms, particle swarm optimization and imperialist competition algorithms [54]. Due to their ability to rapidly analyse massive and complex MWD datasets, ML methods provide an effective solution to provide accurate and relevant predictions of subsurface properties in open cut mining operations.

Interpretation of MWD features can benefit considerably from recent advances in explainable artificial intelligence, as discussed among other by Saranya and Subshini [71] and Lundberg and Lee [72]. Moreover, although having emerged only recently, deep learning models have significant potential in the analysis and interpretation of measure-while-drilling (MWD) data in hard-rock mining [73,74]. These models are particularly well suited to deal with complex data, such as associated with panoramic borehole imaging, for example [73,75,76], or vibrational data [77]. In addition, they could also better support the processing of MWD metrics when considered as time series data [78].

6. Conclusions

The mining industry is increasingly using methods of artificial intelligence to overcome the uncertainty associated with data on geological objects. ML algorithms have been applied on mining MWD datasets to understand pre-excavation subsurface conditions. Numerous studies have been conducted on the collection, processing, analysis, and application of MWD variables for the characterization of rock type, rock mass, and ore properties.

From this review of the literature, the following conclusions can be made:

- The most commonly measured MWD variables measured together with the depth of drilling in open-pit mines are rate of penetration (rop), torque or rotational pressure (tor), bit air pressure (bap), weight on bit (wob), and rotary speed or revolutions per minute (rpm).

- Several studies have analysed the relative importance of each of these variables related to the identification of rock types and characteristics with mixed results. This is an emerging area that can benefit from recent significant advances in machine learning, where models are interpreted or explained. This would particularly be the case where other MWD variables, such as image, vibrational or acoustic signals are included in models.

- In most studies, ML models could successfully categorize geological zones or rock types from MWD data, but further work is required to capture more subtle differences within geological units.

- Classification of discrete rock mass properties, including rock strength and fracture count, using ML methods on open-pit blast hole MWD data does not appear to have been demonstrated yet. Instead, rock type is typically identified as a proxy for rock strength based on assumptions associated with each type.

- Only one study considered the prediction of geochemical properties using ML classification on MWD and drill cutting assay datasets. ML analytical methods applied to MWD data resulting in discrete values for geotechnical, geological, and geochemical will enable a low-cost, high-resolution comprehension of subsurface conditions beyond simple rock type classification.

- Overall, these studies have not yet taken full advantage of recent developments in deep learning and, as more data are collected, these ML approaches are likely to play
a more important role in MWD data analysis for enhanced comprehension of geology, rock mass characteristics, and ore attributes for enhanced sustainability, efficiency, and safety of mining operations.

**Author Contributions:** Conceptualization, D.M.G.; methodology, D.M.G.; software, D.M.G.; validation, D.M.G.; formal analysis, D.M.G.; investigation, D.M.G.; resources, D.M.G.; data curation, D.M.G.; writing—original draft preparation, D.M.G.; writing—review and editing, D.M.G., C.A., and L.O.; visualization, D.M.G. and C.A.; supervision, C.A. and L.O.; project administration, D.M.G.; funding acquisition, D.M.G. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the MRIWA Postgraduate Research Scholarship and the Australasian Institute of Mining and Metallurgy Education Endowment Fund Postgraduate Scholarship.

**Acknowledgments:** The research was supported by the advanced computing resources provided by the Pawsey Supercomputing Research Centre in Perth, Australia.

**Conflicts of Interest:** The authors declare no conflicts of interest.

**References**


**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.