Machine Learning for Lithology Analysis using a Multi-Modal Approach of Integrating XRF and XCT data

Suraj Neelakantan[1], Alexander Hansson[2], Jesper Norell[2], Johan Schött[2], Martin Längkvist[1], Amy Loutfi[1]

***Abstract*—We explore the use of various machine learning (ML) models for classifying lithologies utilizing data from X-ray fluorescence (XRF) and X-ray computed tomography (XCT). Typically, lithologies are identified over several meters, which restricts the use of ML models due to limited training data. To address this issue, we augment the original interval dataset, where lithologies are marked over extensive sections, into finer segments of 10cm, to produce a high resolution dataset with vastly increased sample size. Additionally, we examine the impact of adjacent lithologies on building a more generalized ML model. We also demonstrate that combining XRF and XCT data leads to an improved classification accuracy compared to using only XRF data, which is the common practice in current studies, or solely relying on XCT data.**

I. INTRODUCTION

Drill cores are cylindrical rock samples drilled from the earth, of which an example is shown in Figure 1. Identification and classification of e.g. different rock types and lithofacies in drill cores is an important stage of mineral exploration.

The process of identifying and classifying distinguishable drill core depth ranges is often denoted as core logging [10]. Manual core logging can be inconsistent, leading to variability that complicates the development of reliable geological models [6], [10]. With the advent of new technologies for data collection from the drill cores using X-rays [1], [2], geochemical analysis can be done at a greater flexibility compared to traditional lab assays, also allowing for extraction of additional information such as 3D rock structures. X-ray fluorescence (XRF) scans are used to obtain elemental concentrations from drill cores [3] and X-ray computed tomography (XCT) scanning can give non-invasive access to the entire 3D volume of the drill core at high spatial resolution [4]. XCT data enables precise characterization of mineral grains based on density. This approach is especially useful for identifying high-density minerals such as gold, clearly differentiating them from other minerals and metals [18]. Additionally, by using 2D slices of attenuation values from 3D XCT data in machine learning (ML) algorithms, we can identify and segment euhedral minerals in the drill cores, improving our insights into geological structures [19]. Even though technology aids in collecting a vast amount of data from drill cores more efficiently, the analysis of this data still necessitates manual work and time. ML could therefore be an ideal tool for better leveraging the data at reduced manual effort.

[1]Center for Applied Autonomous Sensor Systems, Örebro University, Örebro, Sweden firstname.lastname@oru.se

[2]Orexplore AB, Torshamngatan 30B, 164 40 Kista, Stockholm, Sweden firstname.lastname@orexplore.com

Fig. 1: Illustration of drill core samples from a geological application arranged in a drill core tray.

For example, in a study of Zn-Pb-Ag deposits in a Swedish mine [6], researchers utilized chemical compositional data obtained from XRF analysis to assess the capabilities of various ML algorithms, including Self-Organizing Maps (SOM) and Classification and Regression Trees, in the classification of rock types. Here the ML algorithms were evaluated based on ground truth given by geologists, showcasing the utility of such algorithms in geological studies. In another study on the classification of rock types [7], SOMs were again used to classify rock types based on elemental compositions. Beyond the reliance on XRF data, the integration of digital images has also been utilized, allowing for the classification of rocks through their texture and color. This approach leverages ML and convolutional neural networks (CNN), marking an advancement in the field by combining traditional methods with different data types to achieve a more detailed geological analysis [8]–[10].

This study examines the integration of multi-modal data, specifically XRF and XCT measurements for classification of lithology in drill cores using ML. We tested our approach with both traditional ML models, like Random Forest (RF) and XGBoost, and deep learning (DL) models, like Bayesian Neural Networks (BNNs) and FT-Transformers.

II. DATA

Drill cores from three distinct mines were utilized. Specifically, drill core samples from three holes (LOV19001, LOV19002, and LOV19003) at the Lovisagruvan mine in Bergslagen, Sweden; six holes (MP0777, MP0779, MP0794, MP0802, MP0816, and MP0826) from the Mavres Petres mine in Greece; and one hole from an undisclosed mine in Sweden were used. These samples were scanned using Orexplore's GeoCore X10^{TM}, which provides XCT, XRF, and density data. The study utilized 662 meters of scanned drill core data, specifically 421 meters from the Lovisagruvan mine, 42 meters from the Mavres Petres mine, and 199 meters from the undisclosed mine in Sweden. When evaluating model

performance in later sections, we rely primarily on Lovisangruvan as the most complete and insightful dataset using the lithologies provided by geologists as ground truth. This as domain experts have confirmed the lithologies in Mavres Petres to be easily determinable even by visual inspection, whereas privacy concerns prevents the disclosure of actual lithologies and scan results details for the undisclosed dataset.

The GeoCore X10^{TM} drill core scanner is capable of measuring XRF signals for elements with atomic numbers ranging from 13 (Aluminum) to 92 (Uranium) [5], expressed as a function of drill core depth at $\sim$ cm resolution. In contrast, the XCT produces a full 3D reconstruction of the entire drill core volume, expressed as X-ray attenuation values in a voxel resolution of 0.2 mm [2]. To incorporate the XCT results in tabular format for the current application, they are summarized as statistical measures of the voxel-attenuation distribution, accumulated over the same depth intervals as the XRF results.

There are two types of such attenuation-derived features: percentile-based and volume fraction features. The percentile-based features represent specific percentiles of the attenuation values, covering all percentiles in steps of 5, from the 0th to the 100th. Volume fraction features, on the other hand, utilize sum of voxel counts within fixed ranges of attenuation values, expressed as volumetric fractions of the material classified as rock. In summary, the dataset has 68 distinct XRF features corresponding to individual chemical elements, 21 percentile feature columns plus 32 volume fraction feature columns (both of which are referred to as XCT features). In this study, the term 'multi-modal data' refers to XRF and XCT datasets, each representing a distinct modality; XRF features corresponding to individual chemical elements and XCT attenuation characteristics, respectively. The XRF and XCT data are used to train and test ML models. Data cleaning, in our context refers to removing columns containing only zeros from the XRF+XCT dataset.

III. METHODOLOGY

In problems involving tabular datasets, much of the existing literature leans towards traditional ML model such as gradient boosted decision trees [23]. Since this work is also based on tabular datasets, we naturally go by this trend. To complement this, we also explore the potential of some DL models on our dataset, providing a comparison to identify the most effective approach for our dataset. The ML models that we use in this work are described below.

A. Random Forest Classifier

The first choice is a traditional ML classifier, the random forest (RF) classifier. This ensemble technique operates by constructing numerous decision trees, with the predicted class determined by a majority vote across these trees. Compared to individual decision trees, this method is less prone to overfitting, making it a more dependable option [11]. Moreover, the ensemble approach of the RF model helps mitigate the impact of data point outliers [11], establishing it as the preferred method for predicting lithologies. Its robustness and efficiency in handling complex datasets mark the RF classifier as a standout choice in the field of geological analysis [11].

B. XG-Boost

XG-Boost, standing for *Extreme Gradient Boosting* [15], which is the second choice of ML model in our study. Unlike RF that operate on a majority voting principle from numerous decision trees, gradient boosting combines predictions from multiple decision trees sequentially. This approach aims to enhance the overall prediction accuracy by optimizing the model's weights based on errors identified in previous iterations. What sets XG-Boost apart is its incorporation of L1 and L2 regularization, which aids in constructing a more generalized machine learning model, making it particularly effective on sparse data [15]. This regularization approach supports the efficiency and reliability of the XG-Boost ML model.

C. Bayesian Neural Networks

The third choice of model in our study is Bayesian Neural Networks (BNNs) because as demonstrated in [14] to predict lithofacies boundaries, BNNs prevent over-fitting and the uncertainty estimation from BNNs can be vital for meaningful interpretation. BNNs are a type of AI models that are grounded in probability theory, enabling reasoning about data variables and incorporating prior beliefs about them [12]. They are believed to perform better with small datasets and in situations of data uncertainties by treating network weights as a distribution [13]. This probabilistic approach allows BNNs to offer not just predictions but also measures of uncertainty, making them especially valuable in applications where data may be sparse or noisy, such as geological core logging. This capability to quantify uncertainty in predictions sets BNNs apart as a crucial ML model, providing an insight into the reliability of the predictions made.

D. Feature Tokenizer Transformer

The fourth and final choice of ML model for comparison in our study is a Transformer, a revolutionary concept introduced in the seminal paper "Attention is All You Need" [16]. The main principle behind this powerful ML model is the *self-attention mechanism*, a process that assesses the relevance of each part of the input data relative to the others, enabling the model to concentrate on important features. Originally developed for text data analysis and forming the backbone of applications like ChatGPT, transformers have demonstrated their versatility across various domains, including handling tabular data [24]. The Feature Tokenizer Transformer (FT-Transformer) represents a specific adaptation of the Transformer architecture to the tabular domain [17]. Similar to how the original Transformer model converts words in text to embeddings, the FT-Transformer transforms all features in a tabular dataset, both categorical and numerical, into embeddings. These embeddings are then processed using the self-attention mechanism. This capability makes it particularly advantageous for complex tasks involving large and intricate

datasets, such as the classification of lithologies using XRF and XCT features.

IV. RESULTS

A. Model Performance: Original Intervals vs. High Resolution Intervals

Dataset	Original Intervals	High Resolution Intervals
Lovisagruvan	29	4340
Mavres Petres	576	4650
Undisclosed	37	1797

TABLE I: The number of samples in each dataset, original and high resolution intervals, excluding depth intervals scanned to less than 80% completeness from both the datasets.

Before comparing different ML models, We first introduce a data augmentation that enables the results presented in later sections. Similarly to image cropping, we utilize the inherent $\sim$ cm resolution of the scan data, to split each original log interval (typically on $\sim$ meter scale) into a number of $\sim$ 10 cm intervals, which greatly increases the total number of samples as summarized in Table I. This assumes that the label applied to an original log interval also applies for each $\sim$ 10 cm sub-interval within it, which is not necessarily valid for all labels and intervals. In fact, we find the assumption to be particularly questionable for "boundary samples", i.e. the outermost sub-intervals of each original log interval, as they often represent a transitional region of uncertainty between more distinguishable intervals, and therefore pay particular attention to this effect in section IV-B.)

A RF classifier is used to demonstrate the effect of augmentation, and given the limited size of the original interval dataset an extensive validation method was used. Therefore, a validation method was implemented where all unique combinations were generated to serve as individual test sets. Out of the 29 samples in the original interval dataset, 26 samples were used for training, and the remaining 3 samples were used for testing in each combination. The original interval dataset reflects the natural form in which data is typically available for core logging and is thus established as the baseline for this study. This baseline is essential as it mirrors the standard conditions under which geological data is collected, providing a reference point against which the efficacy of data augmentation can be assessed.

Lovisagruvan			
Dataset	**Test Acc**	**Precision**	**Recall**
XRF+XCT (Original intervals)	0.72	0.59	0.62
XRF+XCT (High resolution intervals)	**0.93**	**0.93**	**0.93**

TABLE II: Comparison of RF classifier on the XRF and XCT combined original interval and high resolution interval datasets.

Utilizing the high resolution dataset, RF classifier shows an improved performance (see Table II) compared to a RF classifier on original intervals. This improvement in the accuracy highlights the effectiveness of the high resolution dataset to train ML models. Consequently, this study will proceed with the high resolution dataset to explore further enhancements and applications. The high resolution dataset is first split into training and test sets, with a standalone test set used to evaluate all the models in this study. While training the RF classifier and XG-Boost models 5-fold stratified validation was used to ensure a balanced representation of all the classes within each fold. However, while training BNN and FT-transformer a hold-out validation technique was used.

B. Assessment of Split Strategies on Model Performance

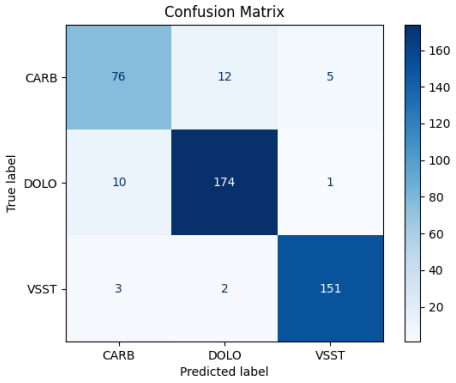
Boundary samples are those located at the edges of each depth interval, representing the transition between different rock types or lithologies. On the other hand, non-boundary samples, are found away from these edges, typically within the central portions of the depth intervals. They represent more stable, homogeneous conditions and are crucial for analyzing the attributes of a lithology without the variability introduced by transitional zones. Together, boundary and non-boundary samples can offer a holistic view of a geological study. ML models are evaluated across distinct dataset split scenarios: Random Split Evaluation, Testing with Only Non-Boundary Interval Samples, Testing on Only Boundary Samples, and Training and Testing without Boundary Samples. This allowed us to understand model performance across all depths of a drill core. In the study by Negin Houshmand et al. [10], a dataset was divided using an approach where continuous segments of each rock type was allocated across training, validation and test sets. By doing so, only boundary samples were included in the test set and this can hinder the performance of ML models based on findings.

1) Random Split Evaluation: The results shown in Table III show that RF and XG-Boost models show high accuracy, precision, and recall across the dataset from Mavres Petres and Lovisagruvan mine, achieving scores well above our set baseline of 0.72. The performance of all models on the dataset Mavres Petres mine has consistently been on the higher side, because the lithologies present within this dataset are quite straightforward for classification as mentioned in subsection II. In contrast, the BNN showed slightly lower performance compared to the RF and XG-Boost models, except in the undisclosed dataset where it performed better. The FT-Transformer exhibited variable performance, with notably high precision on the dataset from Mavres Petres mine but low accuracy on the dataset from undisclosed and Lovisagruvan mine. Generally, the combination of XRF and XCT data contributed to better model performance than datasets featuring either XRF or XCT features alone, highlighting the advantage of using multi-modal data for an improved classification of lithologies.

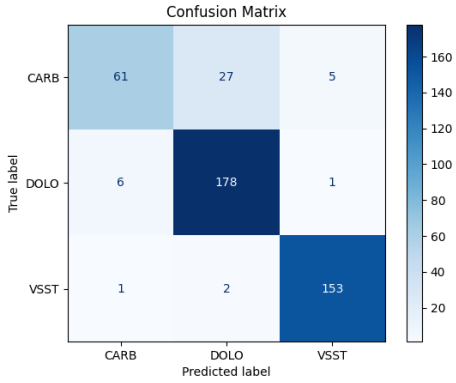
Figure 2 shows the comparison of confusion matrices for various models on the dataset from Lovisagruvan. The RF classifier exhibits strong performance in classifying Dolomite (DOLO) with 174 true positives and also achieves high accuracy for Volcanic Sand Siltstone (VSST) with 151 true posi-

TABLE III: Performance metrics on the test set across all the models and datasets. In the table, yellow highlights indicate the highest performance metrics for each dataset.

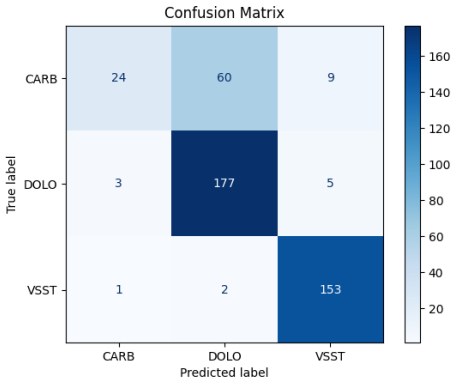
Model / Data	Lovisagruvan			Mavres Petres			Anonymous		
	Acc	Prec	Rec	Acc	Prec	Rec	Acc	Prec	Rec
Random Forest									
XRF+XCT	0.93	0.93	0.93	0.95	0.95	0.96	0.85	0.85	0.85
XRF+XCT Cleaned	0.92	0.92	0.92	0.95	0.95	0.96	0.82	0.82	0.83
XRF	0.84	0.84	0.83	0.95	0.95	0.96	0.81	0.81	0.81
XCT	0.84	0.84	0.83	0.95	0.95	0.96	0.74	0.74	0.75
XG-Boost									
XRF+XCT	0.90	0.90	0.90	0.95	0.95	0.96	0.86	0.86	0.86
XRF+XCT Cleaned	0.90	0.90	0.90	0.95	0.95	0.96	0.85	0.85	0.85
XRF	0.80	0.80	0.79	0.95	0.95	0.96	0.84	0.84	0.85
XCT	0.80	0.80	0.79	0.91	0.91	0.91	0.74	0.74	0.74
BNN									
XRF+XCT	0.82	0.83	0.82	0.72	0.61	0.72	0.82	0.82	0.82
XRF+XCT Cleaned	0.84	0.84	0.84	0.72	0.62	0.72	0.82	0.83	0.82
XRF	0.80	0.81	0.80	0.93	0.93	0.92	0.82	0.83	0.82
XCT	0.81	0.81	0.81	0.93	0.91	0.91	0.65	0.60	0.65
FT-Transformer									
XRF+XCT	0.82	0.84	0.74	0.95	0.97	0.92	0.70	0.70	0.68
XRF+XCT Cleaned	0.83	0.80	0.79	0.86	0.86	0.80	0.78	0.81	0.79
XRF	0.83	0.82	0.83	0.76	0.73	0.71	0.61	0.62	0.62
XCT	0.85	0.82	0.82	0.92	0.96	0.88	0.60	0.57	0.58



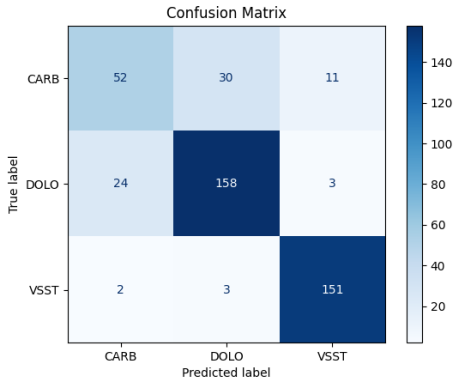
(a) Random Forest



(b) XG-Boost



(c) Bayesian Neural Network



(d) FT-Transformer

Fig. 2: Confusion matrices of various models on the the combined XRF and XCT dataset from Lovisagruvan.

tives. However, it has confusion between Carbonate (CARB) and DOLO, misclassifying 12 instances of CARBs as DOLOs. The XG-Boost model has an improved DOLO classification with 178 true positives and mirrors this strength in VSST

classification with 153 true positives. Yet, it displays a slightly higher rate of confusion between CARB and DOLO, with 27 instances being misclassified, highlighting a challenge in differentiating these two lithologies. In classification of DOLO, BNN records 177 true positives, ranking second only to XG-Boost. Additionally, it achieves 153 true positives in VSST classification, equaling XG-Boost, the best model in classification of VSST. The FT-Transformer model, while presenting a lower true positive count for DOLO at 158, maintains consistent performance for VSST with 151 true positives, aligning with the other models. However, the FT-Transformer model faces difficulty in distinguishing between CARB and DOLO compared to its counterparts, with a higher misclassification count of 30. Across all models, the consistent challenge lies in the misclassification of CARB, albeit to varying degrees. Despite this, all models demonstrate a shared strength in accurately classifying VSST, indicating a common proficiency across the different machine learning approaches.

TABLE IV: RF classifier results for the combined XCT and XRF dataset tested on only non-boundary interval samples.

Lovisagruvan			
Model	**Accuracy**	**Precision**	**Recall**
XRF+XCT	0.88	0.88	0.87
XRF+XCT Cleaned	**0.88**	**0.88**	**0.88**
XRF	0.78	0.78	0.78
XCT	0.78	0.78	0.78

2) Testing with only Non-Boundary Interval Samples: In the analysis of non-boundary samples, the RF model, on the combined XRF and XCT dataset, the metrics in Table IV demonstrates robust performance with accuracy and precision both at 0.88, and recall of 0.87. Although the performance surpasses our baseline metrics, it falls short when compared to results from a randomly split dataset. When the XRF+XCT dataset is cleaned, there is no significant changes in the performance. In contrast, performance declines when the model is trained solely on XRF or XCT data, with accuracy, precision, and recall all dropping to 0.78. This highlights the advantage of using multi-modal data.

3) Testing on only Boundary Samples: When focusing on performance on boundary samples (see Table V), the RF classifier, using the combined XRF and XCT dataset, has an accuracy and a precision of 0.65, alongside a higher recall of 0.76. However, after cleaning the XRF and XCT combined dataset, there's a slight improvement in the model's accuracy and precision to 0.69, with recall of 0.75. There is decline when the model is restricted to using only XRF or XCT data, with accuracy and precision dropping further to 0.63 and recall to 0.67. Compared to the performance on the non-boundary sample test set, there's a decrease in the performance, underscoring the challenges faced by the ML model in classifying boundary samples. Additionally, it's important to highlight that the accuracy on boundary samples falls slightly below our baseline, again, emphasizing the increased difficulty in predicting outcomes accurately in these edge cases.

TABLE V: RF classifier results for Lovisagruvan dataset tested on only boundary interval samples.

Lovisagruvan dataset			
Random Forest	**Test Acc**	**Precision**	**Recall**
XRF+XCT	0.65	0.65	0.76
XRF+XCT Cleaned	**0.69**	**0.69**	**0.75**
XRF	0.63	0.63	0.67
XCT	0.63	0.63	0.67

4) Training and Testing without Boundary Samples: When excluding boundary samples from both the training and testing sets, and using the combined XRF and XCT data, the RF classifier achieves notable accuracy, precision, and recall of 0.91. When the dataset is cleaned, it further enhances the model's performance, with accuracy, precision, and recall slightly increasing to 0.92. This improvement suggests that cleaning the data of non-informative values might lead to more accurate predictions. Additionally, when the RF model is trained on data using only XRF or only XCT, it still exhibits commendable performance, with accuracy and precision at 0.85 and a marginally higher recall of 0.86. These results collectively underscore the efficacy of the RF model in handling varied datasets, particularly when the ML model is trained without boundary samples, and the accuracy achieved is above our established baseline.

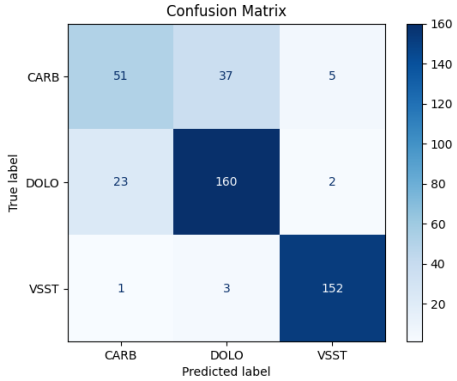
TABLE VI: RF results for the dataset trained and tested on a dataset excluding all the boundary samples.

Lovisagruvan			
Random Forest	**Test Acc**	**Precision**	**Recall**
XRF+XCT	0.91	0.91	0.91
XRF+XCT Cleaned	**0.92**	**0.92**	**0.92**
XRF	0.85	0.85	0.86
XCT	0.85	0.85	0.86

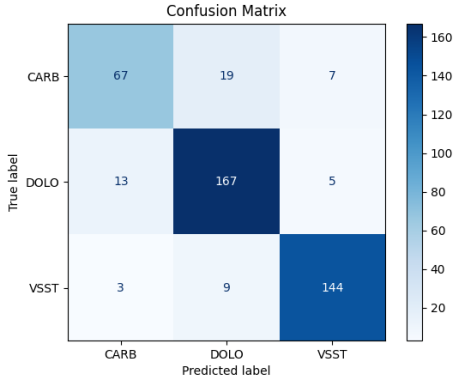
V. DISCUSSIONS

A. Effect of Integrating XRF and XCT Data

The combination of XRF and XCT enhances the classification of lithologies, as evidenced by the comparative analysis of model performances across different test set splits. While the XRF features offer individual chemical elements details crucial for identifying specific lithologies, XCT data offers insights into the distribution of XCT attenuation values, which may vary among different lithologies due to differences in density, mineral composition, and porosity. For instance, a highly porous rock type or a lithology can typically exhibit lower attenuation values at lower percentiles compared to a denser one. These are insights that are not apparent in chemical data alone. This combination is particularly effective in lithologies with similar XRF features but differing in structures. CARB and DOLO are two such lithologies that have similar XRF features [20] and we can notice in Figure 3a this confusion in classifying them when only using XRF features. However this confusion reduces when the model used only XCT features to classify these lithologies, as seen in Figure 3b. Thus, merging these two diverse sets of features allows us to take advantage of each, resulting in an improved classification performance.



(a) Confusion matrix of the XRF based RF model on the XRF Lovisagruvan dataset.



(b) Confusion matrix of the XCT based RF model on the XCT Lovisagruvan dataset.

Fig. 3: Confusion matrices of RF model on Lovisagruvan dataset using scanned XRF and XCT features respectively.

B. Effect of Boundary and Non-Boundary Samples

Here we discuss the impact of various dataset splitting methods on model's outcomes having recognized the advantages of using the combined XRF and XCT dataset. The strategy of random splitting consistently achieved the highest performance across all models when applied to the combined XRF and XCT dataset with accuracy, precision, and recall all above 0.90 for the dataset from Lovisagruvan. This superior performance suggests that training on randomly selected samples, which contains a broad range of geological characteristics and boundaries, more accurately captures the complexity of natural environment. Despite the success of the random split method in providing a comprehensive training through a diverse representation of lithologies, other splitting strategies were also explored. Comparing random split with tests containing only non-boundary samples, since these two splits included more boundary samples while training the models than other splits, the latter showed a slight reduction in accuracy, precision, and recall (approximately 0.88 for XRF+XCT) for the Lovisagruvan dataset. Conversely, tests solely on boundary samples saw the lowest performance, underscoring the challenges models face in predicting lithologies from transitional zones where features may blend with neighboring lithologies or appear less distinct.

Excluding boundary samples from both the training and testing phases has improved the performance of ML models. However, the results still lags behind those obtained from a random split that includes both boundary and non-boundary samples. This indicates that models classify distinct, homogeneous lithologies with relative ease.

C. Evaluation of ML Models

Across all models, the recurring misclassification between CARB and DOLO points to a potential intrinsic similarity in how these lithologies are represented in the dataset. The variability in geological features across different depth ranges within a drill core, although slight, does exist. However, this variability is often overlooked in ML model training, as samples are randomly divided into train and test sets without considering depth ranges. Despite this, the high recall observed across models suggests their effectiveness in broadly classifying lithologies, even when characteristics vary by depth. RF classifiers, in particular, demonstrate a high recall, highlighting their ability to classify lithologies amidst these variations.

Geological datasets are susceptible to noise, incompleteness, measurement errors, and limited sample availability, making a model's resilience to such imperfections can be valuable. Unlike traditional CNNs, which primarily focus on point estimates, BNNs provide a probabilistic approach to predictions. As shown in an application of BNNs in lithology [14], BNNs prevent overfitting and provide uncertainty estimates, key factors in developing reliable geological models. Transformer based models have the ability to emphasize key features because of their self-attention mechanism. E.g. in [25] Vision Transformer (ViT) is used in image-based lithology classification. Traditional neural networks are only capable of collecting local information, which makes it difficult to identify complex patterns while ViT's self-attention technique enables it to identify complex patterns and offers insights through attention rollout visualizations, as detailed in [26]. These visualizations in [25] elucidate the decision-making process, reflecting geological expertise. Given the demonstrated effectiveness of BNN and transformer models in lithology, as shown in the cited works, we compare the accuracy of BNNs and FT-Transformers to traditional ML models. Analyzing the results further from these models is a scope for future work.

VI. CONCLUSION

Our study of lithology classification, using the combined capabilities of XRF and XCT data, highlights the promise of ML models in the field of drill core logging. The integration of XRF and XCT data notably enhances the performance of these models, offering an improvement over approaches that rely solely on either XRF or XCT for training. The combination of XRF and XCT features not only increase the

accuracy of the classification outcomes but also shows the value of multidimensional data analysis in geological studies. We tested four different ML techniques on scanned drill core data collected from three distinct mine sites. RF classifier, XG-Boost, and FT-Transformer showed strong performance on the dataset from the Mavres Petres mine. XG-Boost excelled with the dataset from an undisclosed mine and RF classifier stood out for its effectiveness on the dataset from Lovisagruvan.

Augmenting the scanned data, where lithologies are identified over larger depth intervals, by segmenting them into finer slices of 10cm, can significantly boost the performance of ML models. Another advantage of high resolution data is the ability to be re-composited into different depth intervals like e.g. those corresponding to the intervals selected for geochemical lab assays. Our findings indicate that classifying samples near lithological boundaries presents a challenge.

This work lays a foundation for future exploration into hybrid models that merge the strengths of RF classifier, BNNs, and FT Transformers, potentially leading to more comprehensive lithology classification methods. A direction for future research is to include additional types of data, like digital images of the drill cores or 2D slices from XCT scans, together with XRF and XCT data used in this study. Further work on how to use the uncertainty estimates from Bayesian Neural Networks in real-world geological decisions could be important, especially given the requirements in exploration and resource estimation. By following these paths, future research can make lithology classification models not just more accurate and reliable, but also more useful in practical situations.

ACKNOWLEDGMENT

The authors thank Lovisagruvan AB for permission to include the data displayed in Figure 1. The authors would also like to thank to Stefan Luth from the Department of Mineral Resources, Geological Survey of Sweden, Uppsala, Sweden, and to Tim Baker from Eldorado Gold Corporation for their insights on Lovisagruvan and Mavres Petres datasets respectively.

This work has been supported by the Industrial Graduate School Collaborative AI and Robotics funded by the Swedish Knowledge Foundation Dnr:20190128 and in collaboration with the industrial partner Orexplore Technologies.

REFERENCES

[1] Ross, P-S., Alexandre Bourke, and Bastien Fresia. "A multi-sensor logger for rock cores: Methodology and preliminary results from the Matagami mining camp, Canada." Ore Geology Reviews 53 (2013): 93-111.

[2] Luth, Stefan, et al. "Combined X-Ray Computed Tomography and X-Ray Fluorescence Drill Core Scanning for 3-D Rock and Ore Characterization: Implications for the Lovisa Stratiform Zn-Pb Deposit and Its Structural Setting, Bergslagen, Sweden." Economic Geology 117.6 (2022): 1255-1273.

[3] Croudace, Ian W., Anders Rindby, and R. Guy Rothwell. "ITRAX: description and evaluation of a new multi-function X-ray core scanner." Geological Society, London, Special Publications 267.1 (2006): 51-63.

[4] Williams, Jack N., et al. "Controls on fault zone structure and brittle fracturing in the foliated hanging wall of the Alpine Fault." Solid Earth 9.2 (2018): 469-489.

[5] Element Concentrations - Orexplore Insight User Manual,insight.orexplore.se/manual/latest/concentrations.html. Accessed 16 May 2024.

[6] Simán, Filip, et al. "Rock classification with machine learning: a case study from the Zinkgruvan Zn-Pb-Ag deposit, Bergslagen, Sweden." 2021 Swedish Artificial Intelligence Society Workshop (SAIS). IEEE, 2021.

[7] Klawitter, Mathias, and Rick Valenta. "Automated geological drill core logging based on XRF data using unsupervised machine learning methods." (2019).

[8] Tiu, Glacialle. "Classification of drill core textures for process simulation in geometallurgy: Aitik mine, Sweden." (2017).

[9] Ran, Xiangjin, et al. "Rock classification from field image patches analyzed using a deep convolutional neural network." Mathematics 7.8 (2019): 755.

[10] Houshmand, Negin, et al. "Rock type classification based on petrophysical, geochemical, and core imaging data using machine and deep learning techniques." Applied Computing and Geosciences 16 (2022): 100104.

[11] Breiman, L. (2001). Random forests. Machine learning, 45, 5-32.

[12] Mihaljević, Bojan, Concha Bielza, and Pedro Larrañaga. "Bayesian networks for interpretable machine learning and optimization." Neurocomputing 456 (2021): 648-665.

[13] Jospin, Laurent Valentin, et al. "Hands-on Bayesian neural networks—A tutorial for deep learning users." IEEE Computational Intelligence Magazine 17.2 (2022): 29-48.

[14] Maiti, Saumen, and Ram Krishna Tiwari. "Neural network modeling and an uncertainty analysis in Bayesian framework: a case study from the KTB borehole site." Journal of Geophysical Research: Solid Earth 115.B10 (2010).

[15] Chen, Tianqi, and Carlos Guestrin. "Xgboost: A scalable tree boosting system." Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining. 2016.

[16] Vaswani, Ashish, et al. "Attention is all you need." Advances in neural information processing systems 30 (2017).

[17] Gorishniy, Yury, et al. "Revisiting deep learning models for tabular data." Advances in Neural Information Processing Systems 34 (2021): 18932-18943.

[18] Kyle, J. Richard, and Richard A. Ketcham. "Application of high resolution X-ray computed tomography to mineral deposit origin, evaluation, and processing." Ore Geology Reviews 65 (2015): 821-839.

[19] Neelakantan, Suraj, et al. "Neural network approach for shape-based euhedral pyrite identification in X-ray CT data with adversarial unsupervised domain adaptation." Applied Computing and Geosciences 21 (2024): 100153.

[20] Sun, Junmin, et al. "A Raman spectroscopic comparison of calcite and dolomite." Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy 117 (2014): 158-162.

[21] Levin, Roman, et al. "Transfer learning with deep tabular models." arXiv preprint arXiv:2206.15306 (2022).

[22] Fernández, Alberto, et al. "SMOTE for learning from imbalanced data: progress and challenges, marking the 15-year anniversary." Journal of artificial intelligence research 61 (2018): 863-905.

[23] Shwartz-Ziv, Ravid, and Amitai Armon. "Tabular data: Deep learning is not all you need." Information Fusion 81 (2022): 84-90.

[24] Hegselmann, Stefan, et al. "Tabllm: Few-shot classification of tabular data with large language models." International Conference on Artificial Intelligence and Statistics. PMLR, 2023.

[25] Koeshidayatullah, Ardiansyah, et al. "Faciesvit: Vision transformer for an improved core lithofacies prediction." Frontiers in Earth Science 10 (2022): 992442.

[26] Dosovitskiy, Alexey, et al. "An image is worth 16x16 words: Transformers for image recognition at scale." arXiv preprint arXiv:2010.11929 (2020).