Comparison of supervised and self supervised approaches for micro-CT lithology classification of carbonate rock samples

Carlos E. M. dos Anjos[1], Julio C. V. Fernandes[1], Manuel R. V. Avila[1], Alexandre G. Evsukoff[1], Nelson F. F. Ebecken[1], Rodrigo Surmas[2], Thais F. de Matos[2]

[1]*COPPE, Universidade Federal do Rio de Janeiro*
carlos.menezes@poli.ufrj.br
[2]*Petróleo Brasileiro S.A, Petrobras*

Abstract. The characterization of pre-salt reservoirs is a challenging task in the oil industry due to the geological peculiarities and the heterogeneity of carbonate rocks. These challenges gave rise to new methods in order to better characterize these rocks such as computed tomography for inner structure image generation and new computational methods to analyse them. One of such methods is the use of artificial intelligence techniques, such as deep learning that is considered the state of the art in several tasks and specially on computer vision. This work employs deep learning techniques for the lithological classification of rock samples in micro-tomography images of cylindrical rock samples referred usually as plugs. Two training paradigms are tested and compared, namely, supervised and self-supervised training. The experiments employed densenet161 pretrained on ImageNet as base model, which is a notorious model for image classification on the Imagenet dataset. The contrastive learning method called supervised contrastive (supervised adaptation of SimCLR) was chosen for the self-supervised experiments. This method allows the use of the label information in the SimCLR loss function while also enabling the authors to incorporate this information in three different ways: the standard SimCLR framework (where each image is considered unique), label-based (where each sample belonging to the same lithological classification is considered equal) and sample-based (where all images generated from the same rock sample are equal). The dataset consists of 46,185 images from 623 rock samples which are distributed between seven different classes labeled by specialists. Experiments with different training sizes were performed and compared for the supervised and self-supervised cases. The results obtained over 70% accuracy and F1 score on all experiments with standard deviations equal or lower than 7 % in most cases. The supervised experiments achieved the best results but the self-supervised approaches also displayed comparable results.

Keywords: Deep Learning, Transfer Learning, Self Supervised Learning, Lithology Classification, Carbonate Rocks

1 Introduction

Characterization of carbonate reservoir help describe the composition, disposition and properties of rock layers by comprising chemical, physical, mathematical and geological modeling techniques [1]. Porosity, permeability and lithology are three of the main properties associated with reservoir quality and viability of reservoir development [2]. An arrangement of rocks might be a reservoir if it has an interconnected porous system that includes void space large enough to deposit commercial volumes of hydrocarbons. These characteristics are linked to the lithology of the rock and help decide the viability of reservoir exploitation [2].

Rock samples lithology definition is a challenging task and is usually done by analysing two-dimensional thin-section images [3]. Valentín et al. [2] notices the subjectiveness of this procedure given its dependency on the chosen attributes for the classification. The specialist could focus on biological or petrological attributes, which may lead to different definitions. Usually, the intution rather than quantitative measurements defines the information used on each case to analyze a particular reservoir [4].

According to Ketcham and Carlson [5], data obtained from a three-dimensional rock model are closely related to thin-sections. The high resolution X-ray computed tomography (micro-CT) is a popular tool to generate three-

dimensional models and has become a standard technique in reservoir characterization workflows [6–8], since it permits the volume reconstruction helping estimate mechanical properties [9, 10] and understand the physical phenomena of fluid flow [11, 12].

Artificial intelligence algorithms have been used to automate processes in many industrial applications. Deep learning models excel on tasks such as computer vision, natural language processing and audio processing. Computer vision tasks, especially, have had a rapid growth in the past few years. Most of this evolution was achieved using supervised methods with ever increasing models and datasets. One of the first cases of success was Krizhevsky's Imagenet [13]. From that point on deeper and larger models were implemented, for example, the densenet model used in this work and proposed by Huang [14]. Recently self-supervised methods have achieved even better results. Two of these methods are the SimCLR method proposed by Chen [15] and the supervised adaptation of SimCLR proposed by Khosla in [16]. In the petroleoum industry, these techniques are useful and have already been applied in some works [2, 17].

This work aims to compare supervised and self-supervised training approaches for lithological classification of carbonate rock micro-CT on different traning sizes. This is an evolution of a previously work done by the authors, where deep convolutional models were proposed to automate the workflow for lithology classification in a supervised manner [17]. The distinction here is the use of unsupervised models, transfer learning and the dataset available.

2 Materials and Methods

The common procedure employed for lithological classification is the analysis of individual bi-dimensional thin-section images which is a time-consuming and biased process since it depends on the subjective interpretation of the rock done by the interpreter. In Figure 1(A), the usual workflow for lithological classification is shown going through sample extraction untill thin-section extraction and analysis.

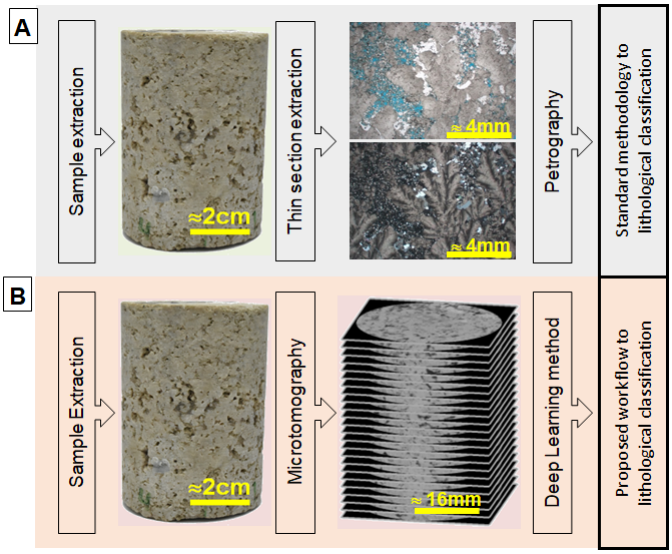

Figure 1. **a** Standard and **b** proposed workflows for lithological classification. Extracted and adapted from Anjos et al. [17].

Micro-CT is a technique incorporated into the reservoir characterization workflow and it produces images that closely correspond to a series of thin-sections enabling a better analysis of the volume as a whole when compared to just one thin-section. More details on micro-CT reconstruction and image acquisition of earth materials can be found in Hanna and Ketcham [18] and Cnnudde and Boone [19]. This work proposes a deep-learning approach for automatic lithological classification using micro-CT images as can be seen in Fig 1(B). The authors proposed a similar workflow in the past where deep convolutional models were used for lithology classification in a supervised manner [17]. The distinction here is the use of self-supervised models, transfer learning and the dataset available.

2.1 Dataset

The dataset available to train and test the models consisted of 623 micro-CT images of rock samples. Each image was cropped to select the region of interest, as seen in Figure 2, removing, thus, the void area around the sample. Each sample was annotated by specialists using the micro-CT image and thin-sections extracted from the same depth of the sample. The samples were grouped into seven categories (Fig 3) - coquinas (COQ), spherulites (ESF), stromatolites (ETR), laminites (LMT), others (OUTROS), reworked (RET), and massive reworked (RTM). These classes were selected and conceptualized by the specialists based on the description done by Terra et al. [20].

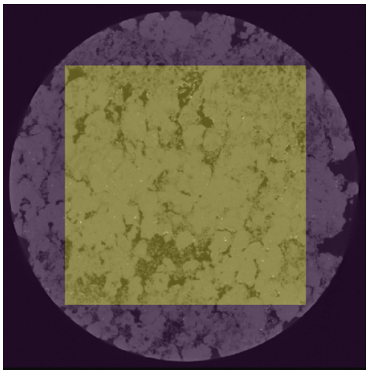

Figure 2. Cropped micro-CT slice, region of interest in yellow.

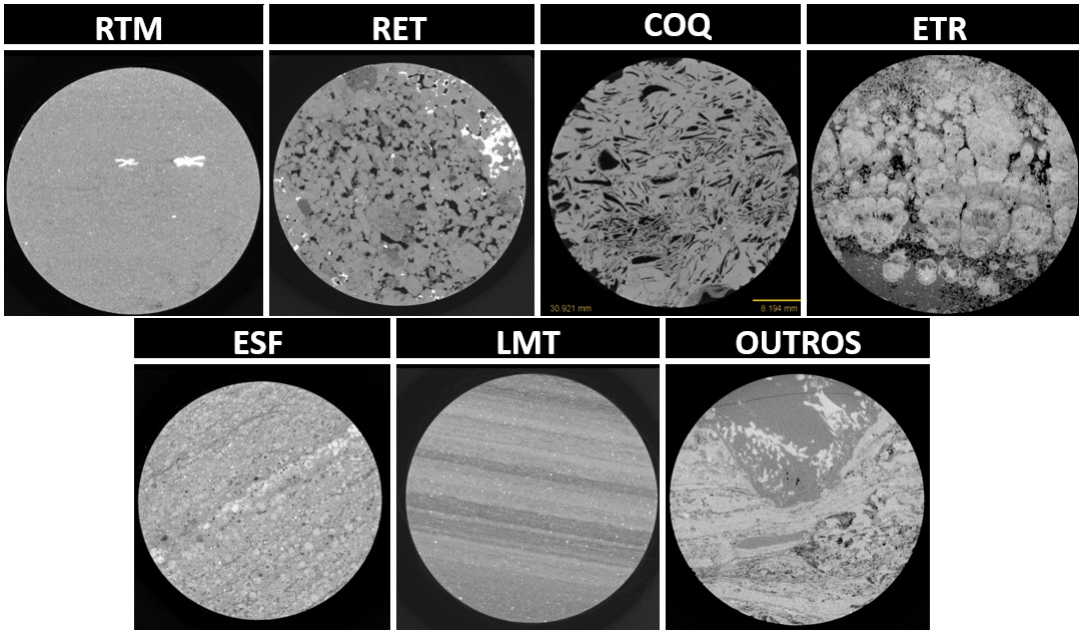

Figure 3. Lithology classes examples.

Different slices were taken from each plug sample, and assuming each plug sample is homogeneous, all slices from the same sample received the same lithological label. The dataset class distribution is shown on Table 1. A total of 46,185 slices from the 623 micro-CT images were used. Even though the plugs are not evenly distributed the slices used for each class are. This means that for classes with more plug samples, fewer slices were used from each plug. In our experience this type of balancing strategy shows better results than the gradient weighting method [21].

The dataset presents some intrinsic features that might cause mis-classification problems, such as ambiguity

Table 1. Dataset class distribution

	Total	COQ	ESF	ETR	LMT	OUTROS	RET	RTM
Plugs	623	56	61	194	48	94	104	66
Slices	46,185	6,397	6,671	6,467	6,421	6,648	6,845	6,736

caused by fragments of different classes in the same sample as is common in Reworked (RET) samples. Additionally, the samples have different luminosity conditions, produced by the polychromatic nature of the X-ray beam, which can also hinders the training. Lastly, artifacts caused by high-density materials may also affect the model's training.

The pre-processing steps applied in each samples consisted of resizing each image to the same dimensions (namely, 256x256) and scaling the image according to each slice's minimum and maximum values. Since the model's applied here are pre-trained we also need to re-scale to the model's expected scale (using pre-defined average and standard deviation values). The image slices used did not undergo any filtering or processing to remove high density artifacts.

2.2 Model

The model selected for this work was Densenet-161 proposed by Huang [14] chosen for its high accuracy on the ImageNet dataset and is easily accessible using the PyTorch framework. The framework allows you to download an Imagenet pre-trained version of the model, this pre-trained version of the densenet model was used in all experiments of this work. Since this model has many layers, as it is common in deep learning models, problems such as vanishing gradients may arise. To work around this problem direct connections between layers are used, thus, increasing the flow of information between layers during training; more details on the architecture can be found in the original paper [14].

2.3 Training methods

This work consists of the comparison between supervised and self-supervised approaches for micro-CT lithology classification of carbonate rock samples using deep learning. In the supervised paradigm the labels are used to directly train the model [21]. In our approach we used the cross-entropy loss function coupled with the SGD optimization method [21].

In the self-supervised approach two methods were chosen namely the SimCLR and the Supervised SimCLR. The SimCLR method was proposed by Chen [15] and uses the constrative loss to learn data representations by maximizing agreement between versions of the same data that have undergone different augmentations. The Supervised Contrastive Learning was proposed by Khosla in [16] and is a modification of the SimCLR loss function which enables the use of labels. This modified loss function allowed us not only to add labels to our self-supervised approach but also enabled us to test a different scenarios. The usual SimCLR approach considers each image, in our case plug slice, as unique, so every other slice, even from the same plug, is considered different during training. With the supervised contrastive learning loss function, we were able to consider slices from the same plug as equal during training without using labels.

The self-supervised approaches are presented as a method to pre-train your model without using labels. These methods aims to create a model that produces good representations features of the image, and with those features, they allow the usage of fewer data to train simpler models. Chen [15] managed to achieve accuracies even better than the common supervised approach on ImageNet using fewer data and a linear model. The author trained a supervised linear model using the features extracted from the self-supervised model trained using the SimCLR framework. The author tested different dataset sizes to check the impact of the sample size in the accuracy. Therefore, for a fair comparison of the models and methods, we trained all approaches using 100% (36,948), 50% (18,474), and 10% (3,694) of the training data available.

2.4 Validation statistics and cross-validation

There are many ways to evaluate a machine learning algorithm [22]. Selecting the appropriate evaluation functions depends on the task executed by the model and the characteristics of the problem. In this work, accuracy and F_1 score were used as figures of merit.

Cross-validation is a common method for prediction error estimation and is widely used in machine learning problems [22]. It consists of splitting the dataset into k subsets of equal or nearly equal sizes, where each subset is known as a fold and retains the same class distribution as the dataset. In our case, the dataset was divided into ten folds, and eight folds were used for training, one for validation and one for testing, as shown in Figure 4. That way, ten different models were trained for each experiment. It is important to note that the test fold was only used after all training steps finished, and on self-supervised approaches, it was only used for inference after the linear model was trained.

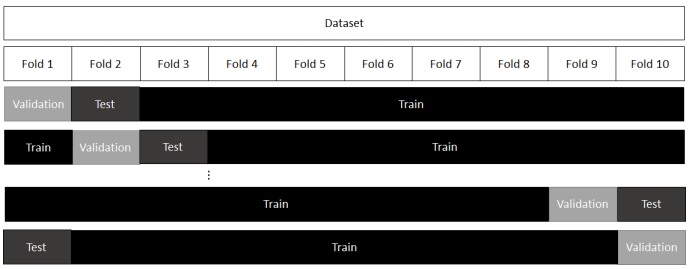

Figure 4. Ten fold cross-validation with train, validation and test sets.

3 Results

Five different experiments were conducted in this work varying the training paradigm from supervised to self-supervised. The first test is the training of a linear model (ImageNetFeatures) with the features extracted using a pre-trained Densenet161, it is important to notice that the Densenet161 model is not trained on the data. The second approach is the common supervised approach training of the Densenet161 model. The third approach uses the SimCLR framework to train the Densetnet161 model to function as a feature generator and then train a supervised linear model using the generated features (SimCLR). The fourth method differs from the third in the sense that it uses the labels of the plugs by utilizing the supervised contrastive loss function during the training and then, just as the third, uses the learned features to train a linear model (SupConLabels). This approach is considered supervised since it uses labels. The last approach uses the supervised contrastive learning loss once more but this time the test differs because instead of using the labels directly to determine which images are equal or not we consider all images coming from the same plug as equal and then like the method mentioned before a linear model is trained on the learned features (SupConPlugs). This is considered self-supervised since it doesn't use the plug labels. This test could have been done without this loss function but it was easier to implement this way.

There are two ways of evaluating the models' predictions. The first is using the figures of merit per slice prediction (each slice is considered a sample), and the second is considering the most frequent prediction among all slices from any given plug. In this work, we use the second type of evaluation since all plugs were considered homogenous during training, and it is, usually, how specialists use the model predictions on their workflow.

The accuracy results on the test set for each experiment and for each training dataset size used are shown in Table 2. It's notable that all models are within each other's standard deviation when evaluating the accuracy. This behavior repeats itself for the F1 score, seen in Table 3. All models achieved over 70% score on both metrics, and the standard deviation is lower than 7% in most cases. These results have a higher average and lower standard deviation than previously achieved [17]. However, the dataset used in this work is ten times larger than before.

Table 2. Test set acurracy for each experiment on different training dataset sizes

Experiment	100%	50%	10%
Supervised	79.79 ± 5.36	78.66 ± 5.67	77.40 ± 5.99
SimCLR	76.42 ± 6.60	75.94 ± 6.72	75.94 ± 6.39
SupConLabels	80.11 ± 5.76	79.63 ± 5.96	78.19 ± 5.80
SupConPlugs	74.98 ± 5.76	73.54 ± 5.99	73.86 ± 4.91
ImageNetFeatures	77.55 ± 6.54	77.71 ± 6.13	76.43 ± 6.44

Table 3. Test set F1 score for each experiment on different training dataset sizes

Experiment	100%	50%	10%
Supervised	79.81 ± 7.26	78.65 ± 6.79	77.82 ± 6.74
SimCLR	76.28 ± 8.11	75.90 ± 8.05	75.17 ± 8.43
SupConLabels	80.16 ± 6.70	79.63 ± 7.03	78.12 ± 6.96
SupConPlugs	74.03 ± 7.10	72.37 ± 7.47	72.73 ± 4.70
ImageNetFeatures	77.41 ± 7.41	77.38 ± 6.63	76.40 ± 7.12

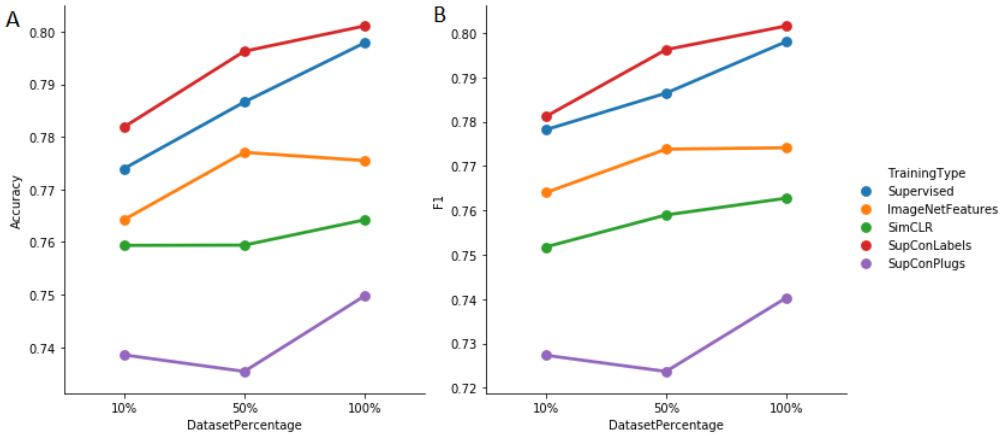

Figure 5. Mean accuracy and F1 score on the test set for all experiment on different training sizes.

In Figure 5, the comparison between the model's average accuracy and average F1 score are shown for increasing dataset sizes. Most models had higher averages with bigger datasets. Even though the models average accuracy and F1 score are close, two experiments had a higher average than the rest, these are the Supervised and SupConLabels experiments. Both experiments are considered supervised since they use labels to train the models. Those two also had a lower standard deviation than most models. The ImageNetFeatures experiment performed almost as well as the supervised approaches and better than the self-supervised approaches. The self-supervised approaches had the lowest averages on accuracy and F1 score, but it's important to note that our dataset is way smaller than ImageNet which could be the reason that these approaches didn't outperform the supervised ones. This dataset size difference could also be the reason that the models performed almost the same on the different dataset sizes tested. It's also notable that the SupConLabels managed to increase the average accuracy of the supervised approach by 0.32% on 100% of the dataset, 0.97% using 50% and 0,79% using only 10%.

4 Conclusions

In this paper, we compared different supervised and self-supervised approaches for lithological classification of carbonate rock micro-CT. The results shown were satisfactory and improved previous works results also increasing the amount of classes, it's important to note that this work had more images available than before. Although the results from the self-supervised approaches weren't better than the supervised approaches, they weren't that far behind and as stated before this could have happened due to the amount of data available. Also the self-supervised approaches with a large enough dataset could create a model capable of generating good features for different kind of problems, for example the same features could be used to train a linear model for lithology classification, porosity estimation and permeability estimation, unlike that the supervised approaches need to train a deep learning model from scratch.

Acknowledgements. The authors would like to thank Petrobras for providing the data and financial support and NVIDIA Corporation for the GPU provided by the NVIDIA Grant Program. The authors also thank the Brazilian Research Council (CNPq) for the scholarships for students and researchers.

Authorship statement. The authors hereby confirm that they are the sole liable persons responsible for the authorship of this work, and that all material that has been herein included as part of the present paper is either the property (and authorship) of the authors, or has the permission of the owners to be included here.

References

[1] L. Buryakovsky, G. V. Chilingar, S. Shin, and H. H. Rieke. *Fundamentals of the Petrophysics of Oil and Gas Reservoirs*. John Wiley & Sons, 2012.
[2] M. B. Valentín, C. R. Bom, J. M. Coelho, M. D. Correia, P. Márcio, P. Marcelo, and E. L. Faria. A deep residual convolutional neural network for automatic lithological facies identification in brazilian pre-salt oilfield wellbore image logs. *Journal of Petroleum Science and Engineering*, vol. 179, pp. 474–503, 2019.
[3] K. Remeysen and R. Swennen. Application of microfocus computed tomography in carbonate reservoir characterization: possibilities and limitations. *Marine and Petroleum Geology*, vol. 25, n. 6, pp. 486–499, 2008.
[4] V. Tschannen, M. Delescluse, M. Rodriguez, and J. Keuper. Facies classification from well logs using an inception convolutional network. *arXiv preprint*, vol. , 2017.
[5] R. A. Ketcham and W. D. Carlson. Acquisition, optimization and interpretation of x-ray computed tomographic imagery: applications to the geosciences. *Computers & Geosciences*, vol. 27, n. 4, pp. 381–400, 2001.
[6] M. A. Knackstedt, C. H. Arns, A. Limaye, A. Sakellariou, T. J. Senden, A. P. Sheppard, R. M. Sok, W. V. Pinczewski, and G. F. Bunn. Digital core laboratory: Properties of reservoir core derived from 3d images. In *SPE Asia Pacific Conference on Integrated Modelling for Asset Management*, pp. 14. Society of Petroleum Engineers, 2004.
[7] C. H. Arns, F. Bauget, A. Limaye, A. Sakellariou, T. Senden, A. Sheppard, R. M. Sok, V. Pinczewski, S. Bakke, L. I. Berge, P. E. Øren, and M. A. Knackstedt. Pore scale characterization of carbonates using x-ray microtomography. *Spe Journal*, vol. 10, n. 04, pp. 475–484, 2005.
[8] S. Claes, J. Soete, V. Cnudde, and R. Swennen. A three-dimensional classification for mathematical pore shape description in complex carbonate reservoir rocks. *Mathematical Geosciences*, vol. 48, n. 6, pp. 619–639, 2016.
[9] C. H. Arns, M. A. Knackstedt, W. V. Pinczewski, and E. J. Garboczi. Computation of linear elastic properties from microtomographic images: Methodology and agreement between theory and experiment. *Geophysics*, vol. 67, n. 5, pp. 1396–1405, 2002.
[10] M. A. Knackstedt, S. Latham, M. Madadi, A. Sheppard, T. Varslot, and C. Arns. Digital rock physics: 3d imaging of core material and correlations to acoustic and flow properties. *The Leading Edge*, vol. 28, n. 1, pp. 28–33, 2009.
[11] J. Dvorkin, N. Derzhi, Q. Fang, A. Nur, B. Nur, A. Grader, C. Baldwin, H. Tono, and E. Diaz. From micro to reservoir scale: Permeability from digital experiments. *The Leading Edge*, vol. 28, n. 12, pp. 1446–1452, 2009.
[12] B. P. Muljadi, M. J. Blunt, A. Q. Raeini, and B. Bijeljic. The impact of porous media heterogeneity on non-darcy flow behaviour from pore-scale simulation. *Advances in Water Resources*, vol. 95, pp. 329–340, 2016.
[13] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pp. 1097–1105, 2012.
[14] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4700–4708, 2017.
[15] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton. A simple framework for contrastive learning of visual representations. In *International conference on machine learning*, pp. 1597–1607. PMLR, 2020.
[16] P. Khosla, P. Teterwak, C. Wang, A. Sarna, Y. Tian, P. Isola, A. Maschinot, C. Liu, and D. Krishnan. Supervised contrastive learning. *arXiv preprint arXiv:2004.11362*, vol. , 2020.
[17] dos C. E. Anjos, M. R. Avila, A. G. Vasconcelos, A. M. P. Neta, L. C. Medeiros, A. G. Evsukoff, R. Surmas, and L. Landau. Deep learning for lithological classification of carbonate rock micro-ct images. *Computational Geosciences*, vol. 25, n. 3, pp. 971–983, 2021.
[18] R. D. Hanna and R. A. Ketcham. X-ray computed tomography of planetary materials: A primer and review of recent studies. *Chemie der Erde-Geochemistry*, vol. 77, n. 4, pp. 547–572, 2017.
[19] V. Cnudde and M. N. Boone. High-resolution x-ray computed tomography in geosciences: A review of the current technology and applications. *Earth-Science Reviews*, vol. 123, pp. 1–17, 2013.
[20] G. J. S. Terra, A. Spadini, A. França, C. Sombra, E. Zambonato, L. d. S. Juschaks, L. Arienti, M. Erthal, M. Blauth, M. Franco, and others. Classificação de rochas carbonáticas aplicável às bacias sedimentares brasileiras. *Boletim de Geociencias da PETROBRAS*, vol. 18, n. 1, pp. 9–29, 2010.
[21] I. Goodfellow, Y. Bengio, and A. Courville. *Deep Learning*. MIT Press, 2016.
[22] J. Friedman, T. Hastie, and R. Tibshirani. *The elements of statistical learning*. Springer series in statistics New York, 2001.