

Geochemical Biodegraded Oil Classification Using a Machine Learning Approach

Sizenando Bispo-Silva ^{1,*}, Cleverson J. Ferreira de Oliveira ¹ and Gabriel de Alemar Barberes ²

¹ Centro de Pesquisa e Desenvolvimento Leopoldo Américo Miguêz de Mello-CENPES, PDIEP-Gerência Geral de Pesquisa Desenvolvimento e Inovação, Departamento de Geoquímica do petróleo. Av. Horácio Macedo, 950-Cidade Universitária, CEP 21941-915 Rio de Janeiro, RJ, Brazil; cleveroli@petrobras.com.br

² Geosciences Center, Department of earth Sciences, University of Coimbra, Rua Silvio Lima S/n, 3030-790 Coimbra, Portugal; gabrielbarberes@gmail.com

* Correspondence: sizenando@petrobras.com.br

Abstract: Chromatographic oil analysis is an important step for the identification of biodegraded petroleum via peak visualization and interpretation of phenomena that explain the oil geochemistry. However, analyses of chromatogram components by geochemists are comparative, visual, and consequently slow. This article aims to improve the chromatogram analysis process performed during geochemical interpretation by proposing the use of Convolutional Neural Networks (CNN), which are deep learning techniques widely used by big tech companies. Two hundred and twenty-one chromatographic oil images from different worldwide basins (Brazil, the USA, Portugal, Angola, and Venezuela) were used. The open-source software Orange Data Mining was used to process images by CNN. The CNN algorithm extracts, pixel by pixel, recurring features from the images through convolutional operations. Subsequently, the recurring features are grouped into common feature groups. The training result obtained an accuracy (CA) of 96.7% and an area under the ROC (Receiver Operating Characteristic) curve (AUC) of 99.7%. In turn, the test result obtained a 97.6% CA and a 99.7% AUC. This work suggests that the processing of petroleum chromatographic images through CNN can become a new tool for the study of petroleum geochemistry since the chromatograms can be loaded, read, grouped, and classified more efficiently and quickly than the evaluations applied in classical methods.



Citation: Bispo-Silva, S.; de Oliveira, C.J.F.; de Alemar Barberes, G. Geochemical Biodegraded Oil Classification Using a Machine Learning Approach. *Geosciences* **2023**, *13*, 321. <https://doi.org/10.3390/geosciences13110321>

Academic Editors: Tomislav Malvić and Jesus Martinez-Frias

Received: 11 September 2023

Revised: 29 September 2023

Accepted: 30 September 2023

Published: 24 October 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: convolutional neural networks; biodegradation; organic geochemistry; orange data mining; chromatogram image

1. Introduction

The Gas Chromatography (GC) technique is widely used by the oil industry and can answer questions related to the origin of the oil and the physical-chemical conditions of production, refining, and storage [1]. Recently, the emergence of artificial intelligence (AI) techniques has opened up the data processing market, grouping, and classification of complex imaged data that could be used to classify chromatogram components [2].

Image data are part of the analytical routine practiced by petroleum geochemists, who use the proportion among chromatographic peaks to define the precursor geological environment and identify contamination by drilling fluid, light exhaust, mixing of oils, and even biodegradation [3–5].

It is important to create a routine for labeling geochemical data in a way that facilitates its extraction and transformation into information to support companies' decision-making. The most modern way to reach this level of management is through machine learning techniques controlled by experts in the field. In the case study of this paper, the users will quickly decide whether, in their analysis, they need to extract biodegraded oils from the data. Hence, the users will be able to download data efficiently with a low risk of noise, which will enable them to obtain more accurate information.

Biodegradation is a phenomenon caused by bacterial activity under 80 °C, often found in shallow reservoir conditions close to water/oil contact [6,7]. These bacteria tend to consume oil's light compounds in the saturate fraction (preferably n-alkanes and then isoalkanes) and then consume aromatics. Further, there are resistant compounds that form complex chemical structures. They are located at the chromatographic baseline hump called the unresolved complex mixture (UCM) [1]. As the biodegradation process is initiated, UCM tends to climb, whereas the concentration of n-alkanes decreases. These observations allowed Wenger [6] to build a biodegradation scale to rank the extent of biodegradation at five biodegradation levels: very slight, slight, moderate, heavy, and severe biodegradation (Figure 1). The biodegrading bacteria begin to consume the C₈–C₁₅ alkanes, accompanied by a very slight UCMs climb. Following, at a moderate level, bacteria consume the most part of n-alkanes (nC₁₅₊); however, UCM presents a tenuous hump.

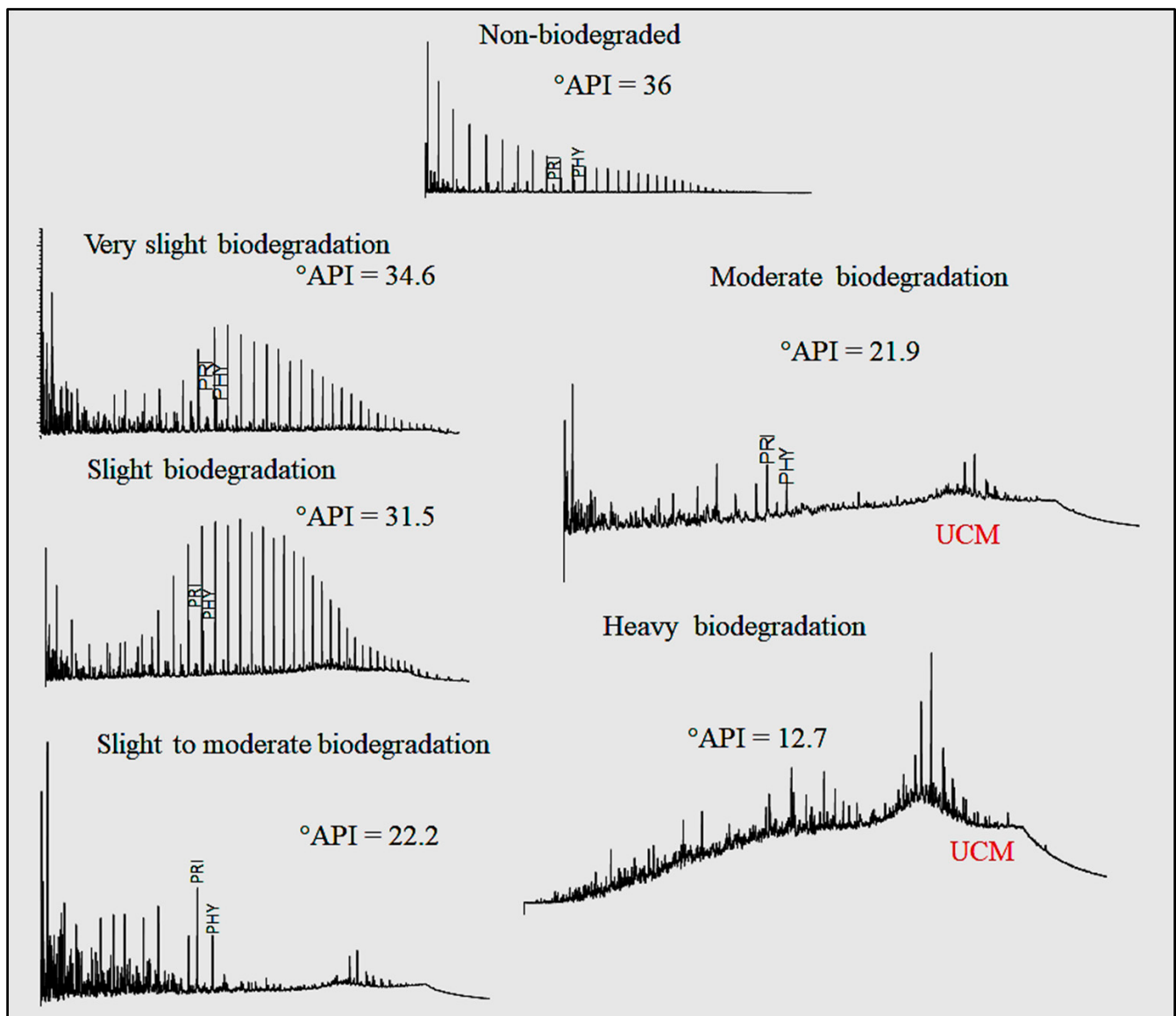


Figure 1. Biodegradation Stages. Based on Wenger [8].

The petroleum density is vital to the oil and gas industry because it implies reservoir recovery's cost reduction together with refined products' quality, which can reduce production costs for companies [5,8]. The °API gravity decreases with the light compounds' loss as well as petroleum quality [3,6,9]. This phenomenon is more sensitive at a slight to moderate biodegradation level than at a moderate to severe biodegradation level [9].

Pristane and phytane are two iso-alkanes commonly found in petroleum and represented in petroleum chromatograms next to nC17 and nC18, respectively. The ratio between the chromatographic peaks of these compounds indicates the probable degree of biodegradation. At the level of moderate biodegradation, the pristane/phytane ratio is little changed. At a heavy level, the UCM hump is very prominent, and n-alkanes become rare [3,6,10]. When the biodegradation reaches the severe stages, biomarkers begin to be consumed, and the demethylated hopanes (25-norhopane) are formed as a result of the ring-opening process by bacteria [10]. If the reservoir underwent more than one oil's charge and there is 25-norhopane together with n-alkanes, it suggests the oil's pulse mixture [3,6,8,11].

In geochemical studies of petroleum, it is common to analyze many samples or compare a few samples with previous analyses to group them, classify the characteristics of the oil, and propose a diagnosis of the studied area (well, reservoir, basin, etc.). So, in essence, the accurate evaluation of each chromatogram image can take a very long time for the geochemist due to the large number of analyses or the complexity of the samples. The use of AI in geochemical analysis brings cost and time savings and reduces the possibility of interpretation errors. However, topics related specifically to the organic geochemistry of petroleum involving the use of AI in image processing are still rare.

The use of statistics in petroleum geochemistry began around the 1960s, with simpler regression techniques and bivariate data. Subsequently, multivariate techniques with chemometrics and Machine Learning (ML) began to be used more widely because of the spread of computers and the increase in computational capacity [12].

Chemometrics aims to explain chemical phenomena through statistical methods, which, in turn, can be processed in a computer quickly by AI algorithms (Machine Learning (ML) and Deep Learning (DL)). A milestone in the use of AI in petroleum geochemistry is the work of McCammon [13], who used the separation of clusters (dendrograms) in oil constituents in order to unravel which of the three horizons producers (in fields in California) would preferentially drain. Wang et al. [14] did an extensive review of the use of chemometric and ML methods in petroleum geochemistry, introducing the possibility of using concentration data in certain situations.

One of the main Deep Learning (DL) algorithms for image classification is the Convolutional Neural Network (CNN), through which a mapping is made from images, finding recurring features and classifying them through neural networks. CNN is an algorithm used to process and classify files of the type of images that have been developed since the 1980s but gained popularity in 2012 [15,16] when it aroused the interest of big tech. CNN is a DL method that caught the attention of the scientific community at the International Skin Imaging Collaboration (2017) when the technique was used to classify images of melanomas with precision similar to experienced dermatologists, bringing speed to the diagnosis of this disease [17,18]. CNN uses a large amount of categorized image data (e.g., topographies such as hill, valley, and mountain) that are read pixel by pixel and transformed into a vector of scores, one for each category. The goal of the algorithm is that each category has the highest score, reducing the error between the output vector and the standard vector. To reduce error, the algorithm uses "weights" (millions of adjustable parameters) that control the input and output of the network and compute the vector that indicates how much a slight change in the weight could increase or decrease the mistake. This is possible because of the Stochastic Gradient Descent (SGD), a technique responsible for presenting the input vector, calculating the output ones and their respective errors repeatedly, and readjusting the weight with each new measurement. The sum of the vector

weights is computed, and when it is above a certain range, it is classified as a feature in a category [15,19,20].

Surveys involving the use of the same or similar algorithms began to be published with topics related to other areas of knowledge. In Geology, de Lima et al. [21] used images of fossils, rock samples, cores, and petrographic samples to classify and group them, and satisfactory results were obtained. Other authors were also able to classify rock images in order to improve petrographic analysis time through ternary diagrams [22,23]. CNN has been used to classify explosive volcanic plumes [24], fossil identification [25], and unstructured geological text data clustering [26]. Koeshidayatullah et al. [27] used transfer learning [28] to classify 4000 carbonate petrographic images in six classes as well as nine object detection classes. Pires de Lima et al. [28] also used transfer learning to make lithofacies classifications with approximately 7000 images split into 17 classes. These authors also compared different pre-trained models to accurately classify petrographic thin-section images [29]. CNN was successfully used to identify rock fractures from outcrop pictures and drills [30,31]. Kim et al. [32] applied CNN to identify saturation changes in core images caused by gas hydrate dissociation. With regard to source rock, the CNN coupled with an unsupervised algorithm was used in well logging data to predict total organic carbon (TOC), S_2 , and S_1 values [33,34] and was used in seismic images to identify petroleum system elements and consequently hydrocarbon leads [35]. In addition, some papers used semantic segmentation to identify coal macerals and determine their rank [36,37]. According to some authors, CNN can be used to predict rock porosity through data logging, seismic images [38,39], and permeability [40]. Zeng and Wang [41] were able to use CNN to classify SAR images from oil spills with greater accuracy than conventional ML methods. Moreover, some authors have used CNN to classify Remote-Sensing Scene [42–44].

In the forensic area, Bogdal et al. [45] used chromatogram image data to classify flammable waste and determine the presence of traces of gasoline. Furthermore, in the field of organic chemistry, some works used the CNN to qualify affected peaks by elution on GC-MS chromatograms in order to discriminate the noise from the true peak [46].

This article aims to report a process automation of image analysis with the purpose of discriminating biodegraded oils from non-biodegraded oils. The success of this test, in addition to speeding up the analysis process, brings a new look at the geochemical characterization of oils.

2. Materials and Methods

2.1. Convolutional Neural Network (CNN)

The first step in using CNN was to group the image bank according to categories (continuing the example given above, hill, valley, or mountain) and load it into the algorithm. Subsequently, the data goes through a set of convolutional layers that work as an extractor of recurring features from the images, rearranging them in a feature map (Figure 2). Each neuron in the feature map of a given layer is connected with all neurons of the previous layer via weights (filter banks). Lecun et al. [15] state that all units in the feature maps share the same filter bank, mathematically corresponding to a convolution. To obtain more robust and less general features that can recognize patterns at any position in the image, a nonlinear (Kernel) calculation method is used. This step is called pooling layers and is responsible for reducing the variance in feature maps with distortions or translations (Figure 2). According to Lecun et al. [15], “although the role of the convolutional layer is to detect local conjunctions of features from the previous layer, the role of the pooling layer is to merge semantically similar features into one”.

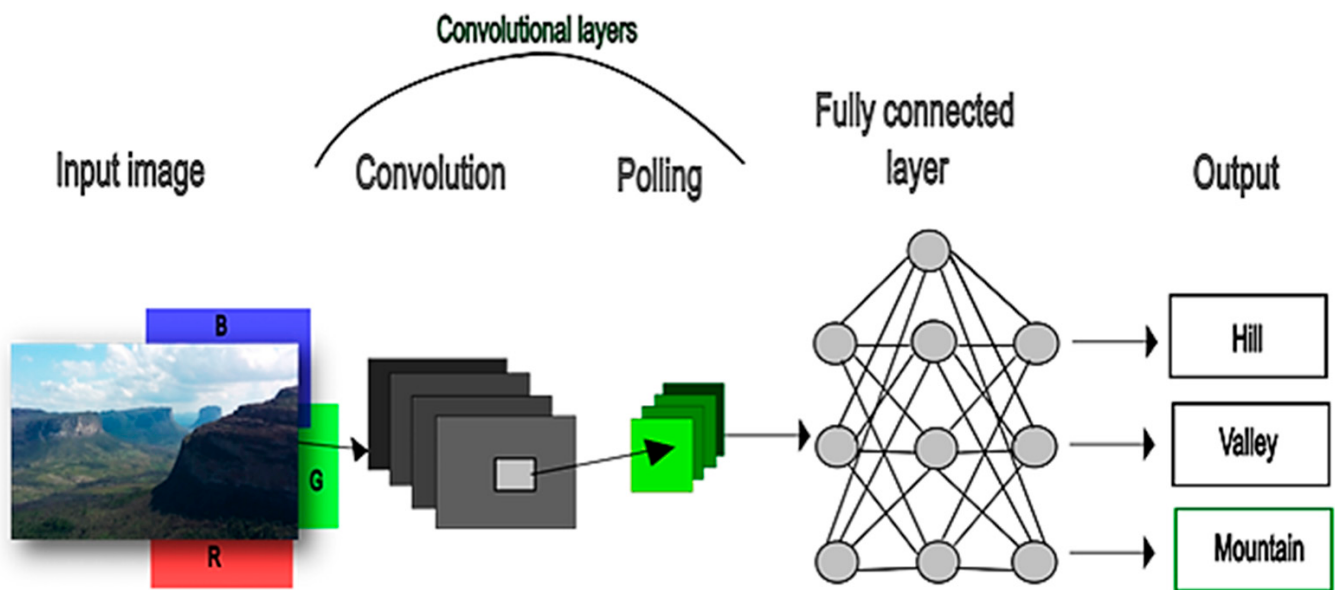


Figure 2. Materialization of a convolutional network and its analytical flow. Adapted from Lecun et al. [15] and Rawat et al. [16].

Soon after, each layer is stacked on top of the previous one to extract more features (fully connected layers) being extensively trained through the backpropagation mechanism and, as a result, comes out with a predicted value (category or class).

2.2. Convolutional Neural Network using Orange[®]

The chromatogram images were loaded into the Orange[®] software, where the InceptionV3 CNN algorithm was used for dedication (dimension reduction or embedded) and image processing by Deep Learning [47,48]. InceptionV3 is a CNN model that was trained on more than 1 million images. However, Orange[®] can import the inceptionV3 knowledge for training new image types (Transfer Learning). InceptionV3's transfer learning is important for data with a few samples since CNN works better with larger [2,28,49,50]. The DL processing via CNN determines the weights and feature maps of the images by finding patterns and creating filters from the training images (81% of the images). Next, Machine Learning algorithms (standard neural networks, logistic regression, decision tree, naive bayes, and random forest) were employed to classify the embedded images and compare them with each other. The algorithm with the best accuracy was utilized to generate a prediction model for the test samples (19% of the images). In the test, the model was effectively tested with untrained samples and revealed the actual efficiency of the technique for image classification. The complete flowchart of the Deep Learning analysis through CNN of gas chromatography imaged data can be seen in Figure 3.

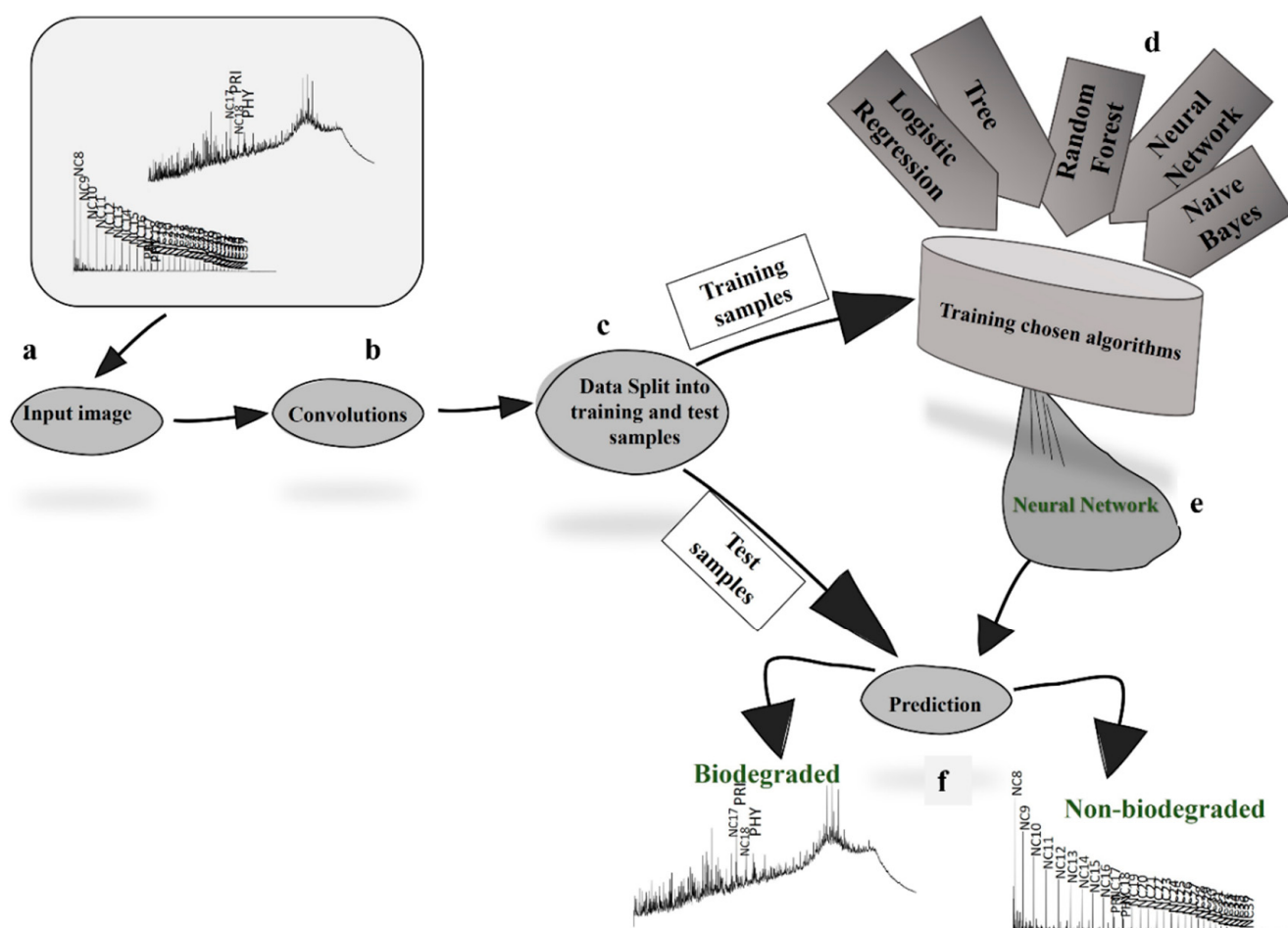


Figure 3. Complete flowchart of image analysis in the Orange® software. (a) input image; (b) convolutional calculations; (c) separation of test samples and training samples; (d) sample training with five algorithms; (e) the best model's testing; and (f) output class.

A total of 221 whole oil images (chromatograms) in JPEG format from gas chromatography analysis were used and tested. These data show oils from foreign basins (East Venezuela, Lusitanian, and Lower Congo, among others); however, the vast majority belong to Brazilian basins (Campos, Santos, Recôncavo, and Potiguar, among others). The samples were previously classified as both biodegraded and non-biodegraded (Table 1 and Figure 4). However, some samples were purposely misclassified as biodegraded (they are not currently biodegraded) in order to evaluate the efficiency of the classification model with mistakes still in the training stage.

Table 1. Number of images used and original classification.

Biodegraded	Non-Biodegraded
92	129

The data were processed by CNN, which measured the images (180 images) and created specific filters for each category. Next, the image classifier was trained using the results calculated by the CNN to create a robust image prediction model of the chromatograms from biodegraded oils. There is a moderate difference in the number of images for each class. Nevertheless, in the test stage, the samples were stratified to avoid any bias in the model. For that, it was necessary to find the algorithm that would present the best result (accuracy).

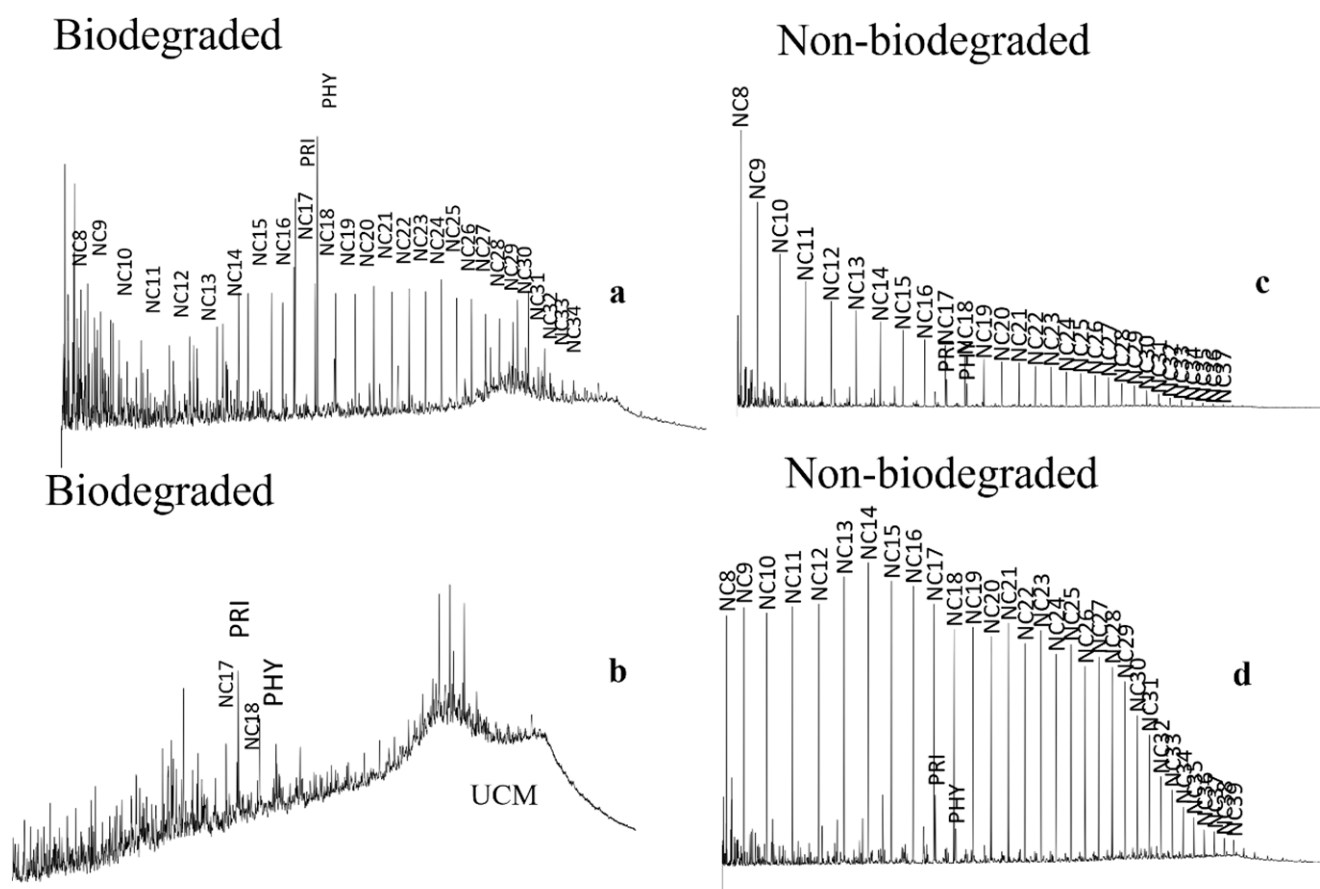


Figure 4. Chromatogram images used in the analysis and their pre-training classification. Figures (a,b) are chromatograms of biodegraded oil samples. Figure (a) presents the loss of light compounds (the peaks have a smaller carbon number than nC₁₆). Figure (b) shows the total loss of light compounds in addition to the rise of UCM. Figures (c,d) are chromatograms of non-biodegraded oil samples.

3. Results

The algorithms Naive Bayes, Neural Networks, Random Forest, Decision Tree, and Logistic Regression were chosen to test the classification of images (Table 2). Neural Networks presented the best classification result because, despite having an area under the curve (AUC) as high as Logistic Regression (both with 99.7%), it presented the highest accuracy (CA) among all algorithms with 96.7%, followed by Logistic Regression and its 96.1%. Among the 6 samples that were misclassified, 4 show mild biodegradation with the loss of light compounds (<nC₁₆) or a slight rise in UCM (Figure 5).

Table 2. Classification training results for the five ML algorithms. Note that the Neural Networks algorithm presented the highest accuracy (CA) of the group, followed by Logistic Regression.

Model	AUC	CA
Decision Tree	0.889	0.928
Random Forest	0.973	0.939
Neural Network	0.997	0.967
Naive Bayes	0.94	0.939
Logistic Regression	0.997	0.961

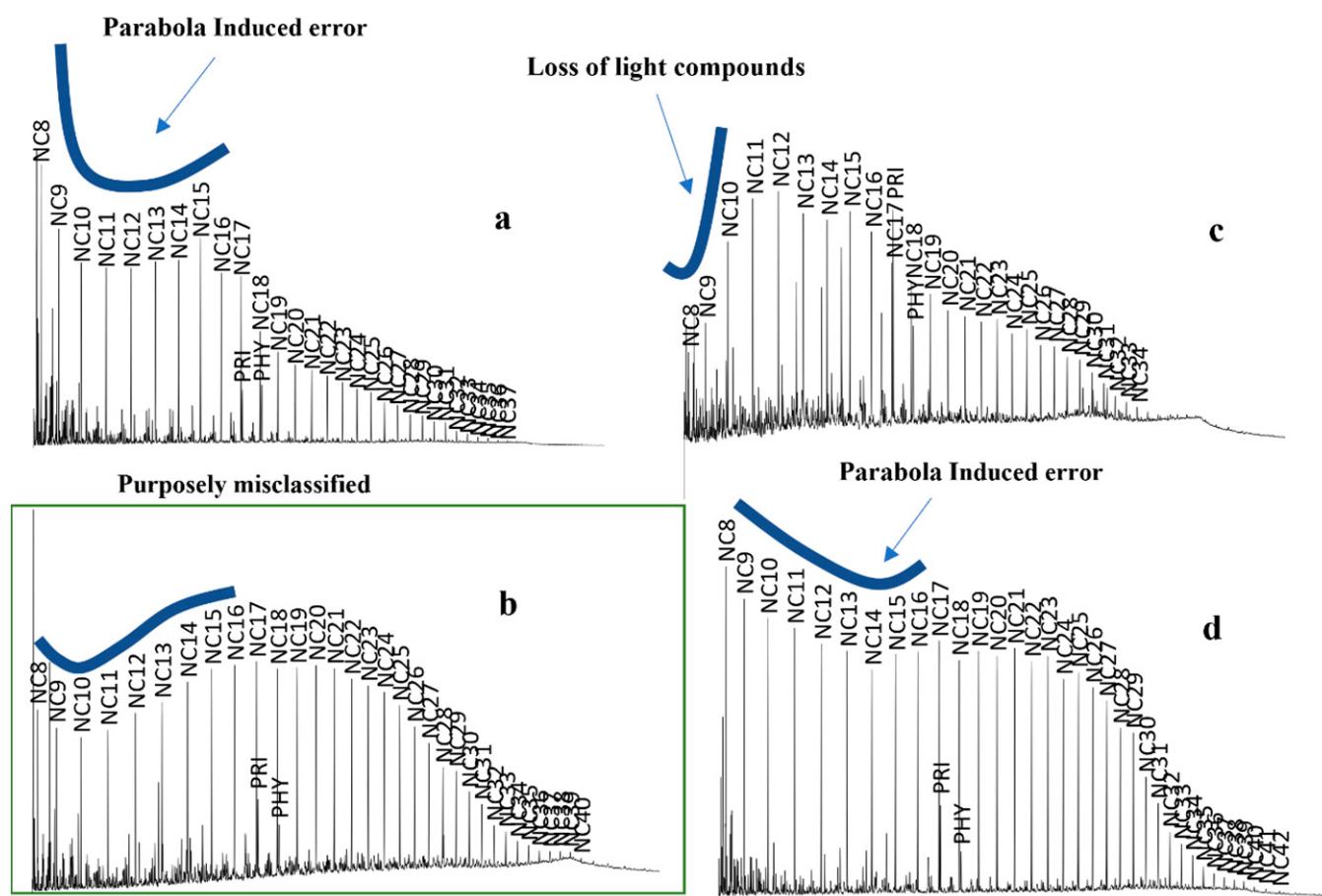


Figure 5. Results of misclassified samples in the training step. Pictures (a,d) represent non-biodegraded oils; however, CNN classified them as biodegraded. Note the small parabola in the region of the lighter compounds, which is related to the original composition of the organic matter and may have misled CNN analysis for pictures (a,d). Pictures (b,c) represent biodegraded oils; however, CNN classified (c) as non-biodegraded. Observing the lighter compounds' loss means there was a slight biodegradation, which may have misled CNN analysis for picture (c). Picture (b) was purposely misclassified as non-biodegraded in the training step; however, CNN classified it as biodegraded.

Once the prediction model was established, the next step was intended to test the model through the processing and classification of 41 images not yet classified. The test result (Table 3) shows that the AUC achieved was 99.7%, with an accuracy of 97.6%, which is even better than the training result. The confusion matrix of the test samples indicates that only one sample was misclassified; however, this sample shows characteristic elements of contamination by drilling fluid, like a prominent pike at nC₁₃ to nC₁₇ compounds (Figure 6a) [5]. The result of the mixture of severe biodegraded oil (note the 25-norhopane peak in Figure 6b) and drilling fluid will be an oil-derived chromatogram with no distinguishable elements of biodegradation. Therefore, the test's prediction error is actually a hit (Figure 6).

Table 3. Test result and model classification.

Model	AUC	CA
Neural Network	0.997	0.976

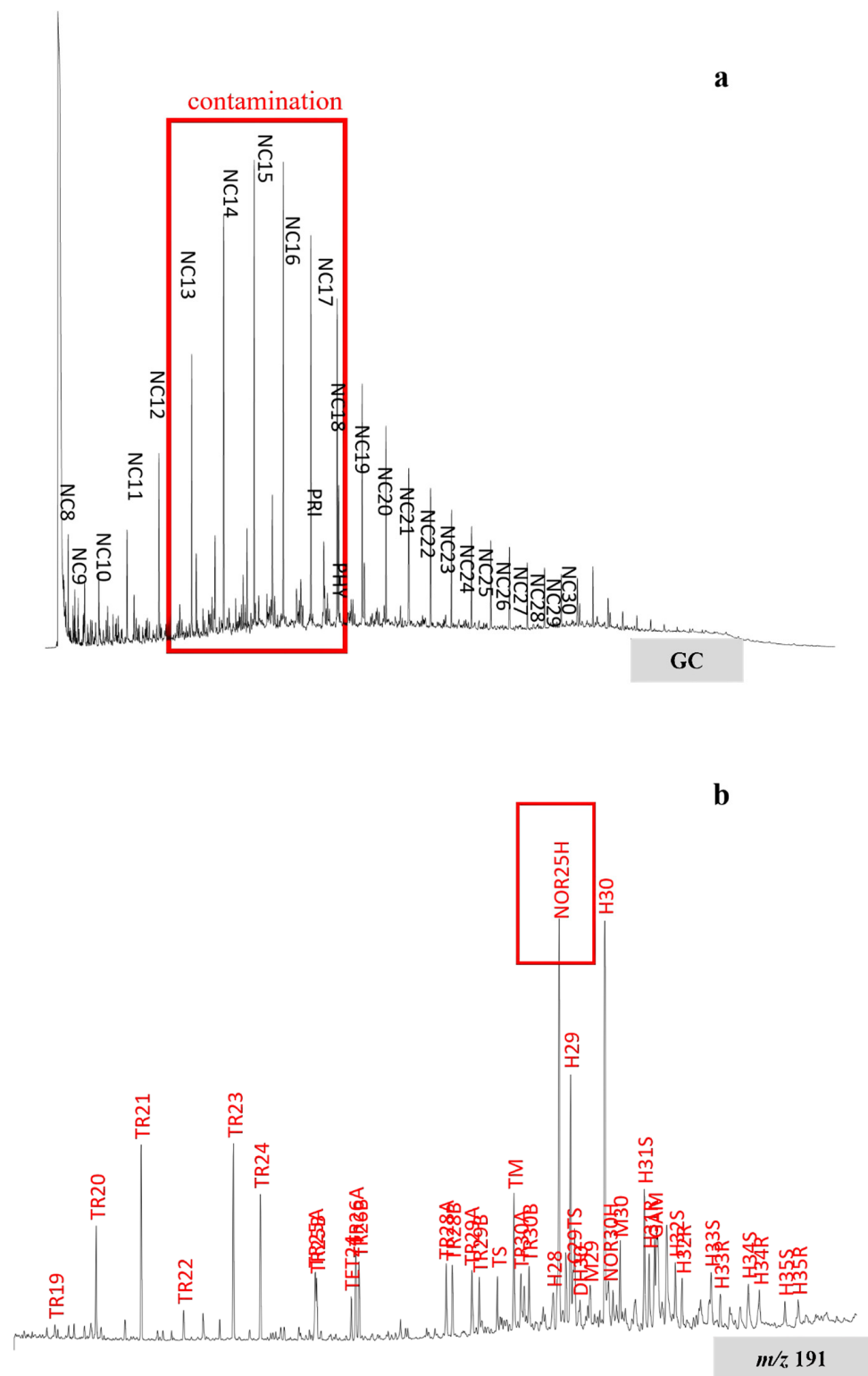


Figure 6. This sample was misclassified by the algorithm, as there was a mixture of biodegraded oil and non-biodegraded fluid in this well. (a) A gas chromatography (GC) sample previously classified for training as biodegraded was predicted to be non-biodegradable by the model, which is correct as the chromatography sample results in an oil with non-biodegradable characteristics. (b) Note that the terpene fragmentogram highlights the high peak of 25-norhopane, a diagnostic biomarker of severe biodegradation.

4. Discussion

Samples that show only mild biodegradation features or mixtures of fluids from different sources may induce the algorithm's prediction error (Figures 5 and 6 and Table 4). It is important to note that the CNN algorithm is highly dependent on the number of samples used for training. In fact, with more images, the algorithm tends to have more accurate and complex responses. Samples purposely misclassified serve as a screen for simulating cases in which the previous manual classification presents some misclassified samples. The critical point in this case study is that even small errors in the pre-classification can generate a useful and adjusted model.

Table 4. Confusion matrix shows the number of samples classified correctly (in blue) and incorrectly (in green).

ACTUAL	PREDICTED		Σ
	Biodegraded	Non-biodegraded	
Biodegraded	14	1	15
Non-biodegraded	0	26	26
Σ	14	27	41

Some authors pointed out that it is possible to mix biodegraded oil with younger oils from fresh charges into reservoirs [6,11]. However, the better way to identify biodegraded and fresh's mixed oils is through a m/z 177 or 191 mass chromatogram, because mass chromatograms display 25-norhopanes peaks. Nevertheless, studying mass chromatograms is beyond the scope of the present paper.

Despite the increasing use of CNN in images from rock, paleontological, and petrographic materials, the use of CNN for the improvement of organic geochemical analysis is still quite rare. Geochemists typically take between 8 and 16 h to interpret 221 chromatogram images. A deep learning model can reduce this time to almost 10 min. Notwithstanding the success, unfortunately, CNN does not give the main parameters and details used for your interpretative mechanism [15,21]. Nevertheless, the use of CNN can open a new horizon for geochemistry when it comes to analysis by GC-FID, GC-MS, and GC-MS/MS (total and selected ion chromatogram), identification of contaminants (as well as environmental pollutants), identification of analysis defects, and, finally, identification and characterization of origin and oil maturation.

5. Conclusions

Each well drilled for the petroleum industry increases the amount of generated oil data (isotopes, biomarkers, composition, etc.). Therefore, it is vital for these companies' managers to manage their databases in order to simplify the download by users, who can use these geochemical data to obtain information and provide rapid support for geological modeling, well locations, and drilling resolutions.

This research proposes a new way to interpret petroleum by using a deep learning approach. The experiments were feasible to achieve high accuracy by modeling with low computational cost. This approach is sufficient to reduce the time of geochemist interpretation, and it allows companies to manage their geochemical data bank adroitly.

It is worth noting that the CNN model may also be applied to other oil classification problems such as clustering analysis, drill contamination, or even the environmental origin of parental source rock. There are possibilities for using CNN in bitumen, oil shows, or even gas samples.

Author Contributions: S.B.-S.: Conceptualization, Methodology, Writing—Original Draft, Software manipulation. G.d.A.B.: Writing—Reviewing, and Editing; Resources. C.J.F.d.O.: Writing—Reviewing, and Editing e Data Curation. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Data will be made available on request.

Acknowledgments: The authors would like to thank especially Petrobras for granting the geochemistry data used in this paper and Jarbas V. P. Guzzo for their contribution to data curation and the incentive provided.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Peters, K.E.; Walters, C.C.; Moldowan, J.M. *The Biomarker Guide: Biomarkers and Isotopes in the Environment and Human History*; Cambridge University Press: Cambridge, UK, 2004; Volume 1, ISBN 9780521781589.
- Alzubaidi, L.; Zhang, J.; Humaidi, A.J.; Al-Dujaili, A.; Duan, Y.; Al-Shamma, O.; Santamaría, J.; Fadhel, M.A.; Al-Amidie, M.; Farhan, L. Review of Deep Learning: Concepts, CNN Architectures, Challenges, Applications, Future Directions. *J. Big Data* **2021**, *8*, 53. [[CrossRef](#)] [[PubMed](#)]
- Peters, K.E.; Walters, C.C.; Moldowan, J.M. *The Biomarker Guide: Biomarkers and Isotopes in Petroleum Systems and Earth History*; Cambridge University Press: Cambridge, UK, 2004; Volume 2, ISBN 9780521781589.
- Kotarba, M.J.; Bilkiewicz, E.; Jurek, K.; Więclaw, D.; Machowski, G. Origin, Migration and Secondary Processes of Oil and Natural Gas in the Western Part of the Polish Outer Carpathians: Geochemical and Geological Approach. *Int. J. Earth Sci.* **2021**, *110*, 1653–1679. [[CrossRef](#)]
- Wenger, L.M.; Davis, C.L.; Evensen, J.M.; Gormly, J.R.; Mankiewicz, P.J. Impact of Modern Deepwater Drilling and Testing Fluids on Geochemical Evaluations. *Org. Geochem.* **2004**, *35*, 1527–1536. [[CrossRef](#)]
- Wenger, L.M.; Davis, C.L.; Isaksen, G.H. Multiple Controls on Petroleum Biodegradation and Impact on Oil Quality. *SPE Reserv. Eval. Eng.* **2001**, *5*, 375–383.
- Röling, W.F.M.; Head, I.M.; Larter, S.R. The Microbiology of Hydrocarbon Degradation in Subsurface Petroleum Reservoirs: Perspectives and Prospects. *Res. Microbiol.* **2003**, *154*, 321–328. [[CrossRef](#)]
- Wenger, L.M.; Isaksen, G.H. Control of Hydrocarbon Seepage Intensity on Level of Biodegradation in Sea Bottom Sediments. *Org. Geochem.* **2002**, *33*, 1277–1292. [[CrossRef](#)]
- Elias, R.; Vieth, A.; Riva, A.; Horsfield, B.; Wilkes, H. Improved Assessment of Biodegradation Extent and Prediction of Petroleum Quality. *Org. Geochem.* **2007**, *38*, 2111–2130. [[CrossRef](#)]
- Connan, J. *Biodegradation of Crude Oils in Reservoirs*; Academic Press Inc. (London) Ltd.: London, UK, 1984; Volume 1, ISBN 0120320010.
- Nascimento, L.R.; Rebouças, L.M.C.; Koike, L.; de A.M Reis, F.; Soldan, A.L.; Cerqueira, J.R.; Marsaioli, A.J. Acidic Biomarkers from Albacora Oils, Campos Basin, Brazil. *Org. Geochem.* **1999**, *30*, 1175–1191. [[CrossRef](#)]
- Hempkins, W.B. Multivariate Statistical Analysis In Formation Evaluation. In Proceedings of the SPE: All Days, San Francisco, CA, USA, 12 April 1978.
- McCammon, R.B. The Dendrograph—A New Tool for Correlation. *Bull. Geol. Soc. Am.* **1968**, *79*, 1663–1670.
- Wang, Y.-P.; Zou, Y.-R.; Shi, J.-T.; Shi, J. Review of the Chemometrics Application in Oil-Oil and Oil-Source Rock Correlations. *J. Nat. Gas Geosci.* **2018**, *3*, 217–232. [[CrossRef](#)]
- Lecun, Y.; Bengio, Y.; Hinton, G. Deep Learning. *Nature* **2015**, *521*, 436–444.
- Rawat, W.; Wang, Z. Deep Convolutional Neural Networks for Image Classification: A Comprehensive Review. *Neural Comput.* **2017**, *29*, 2352–2449. [[PubMed](#)]
- Esteva, A.; Kuprel, B.; Novoa, R.A.; Ko, J.; Swetter, S.M.; Blau, H.M.; Thrun, S. Dermatologist-Level Classification of Skin Cancer with Deep Neural Networks. *Nature* **2017**, *542*, 115–118. [[CrossRef](#)] [[PubMed](#)]
- Marchetti, M.A.; Liopyris, K.; Dusza, S.W.; Codella, N.C.F.; Gutman, D.A.; Helba, B.; Kalloo, A.; Halpern, A.C.; Soyer, H.P.; Curiel-Lewandrowski, C.; et al. Computer Algorithms Show Potential for Improving Dermatologists' Accuracy to Diagnose Cutaneous Melanoma: Results of the International Skin Imaging Collaboration 2017. *J. Am. Acad. Dermatol.* **2020**, *82*, 622–627. [[CrossRef](#)]
- Liang, X. Theoretical Basis. In *Ascend AI Processor Architecture and Programming*; Elsevier: Amsterdam, The Netherlands, 2020; pp. 1–40, ISBN 9780128234884.
- Albawi, S.; Mohammed, T.A.; Al-Zawi, S. Understanding of a Convolutional Neural Network. In Proceedings of the 2017 International Conference on Engineering and Technology (ICET), Antalya, Turkey, 21–23 August 2017; pp. 1–6. [[CrossRef](#)]
- de Lima, R.P.; Bonar, A.; Duarte Coronado, D.; Marfurt, K.; Nicholson, C. Deep Convolutional Neural Networks as a Geological Image Classification Tool. *Sediment. Rec.* **2019**, *17*, 4–9. [[CrossRef](#)]
- Xu, Z.; Ma, W.; Lin, P.; Shi, H.; Pan, D.; Liu, T. Deep Learning of Rock Images for Intelligent Lithology Identification. *Comput. Geosci.* **2021**, *154*, 104799. [[CrossRef](#)]
- Alfárez, G.H.; Vázquez, E.L.; Martínez Ardila, A.M.; Clausen, B.L. Automatic Classification of Plutonic Rocks with Deep Learning. *Appl. Comput. Geosci.* **2021**, *10*, 100061. [[CrossRef](#)]
- Wilkes, T.C.; Pering, T.D.; McGonigle, A.J.S. Semantic Segmentation of Explosive Volcanic Plumes through Deep Learning. *Comput. Geosci.* **2022**, *168*, 105216. [[CrossRef](#)]

25. Wang, H.; Li, C.; Zhang, Z.; Kershaw, S.; Holmer, L.E.; Zhang, Y.; Wei, K.; Liu, P. Fossil Brachiopod Identification Using a New Deep Convolutional Neural Network. *Gondwana Res.* **2022**, *105*, 290–298. [CrossRef]
26. Wang, B.; Wu, L.; Xie, Z.; Qiu, Q.; Zhou, Y.; Ma, K.; Tao, L. Understanding Geological Reports Based on Knowledge Graphs Using a Deep Learning Approach. *Comput. Geosci.* **2022**, *168*, 105229. [CrossRef]
27. Koeshidayatullah, A.; Morsilli, M.; Lehrmann, D.J.; Al-Ramadan, K.; Payne, J.L. Fully Automated Carbonate Petrography Using Deep Convolutional Neural Networks. *Mar. Pet. Geol.* **2020**, *122*, 104687. [CrossRef]
28. Bengio, Y. Deep Learning of Representations for Unsupervised and Transfer Learning. *Proc. ICML Work. Unsupervised Transf. Learn.* **2012**, *27*, 17–36.
29. Pires De Lima, R.; Suriamin, F.; Marfurt, K.J.; Pranter, M.J. Convolutional Neural Networks as Aid in Core Lithofacies Classification. *Interpretation* **2019**, *7*, SF27–SF40. [CrossRef]
30. Byun, H.; Kim, J.; Yoon, D.; Kang, I.S.; Song, J.J. A Deep Convolutional Neural Network for Rock Fracture Image Segmentation. *Earth Sci. Inform.* **2021**, *14*, 1937–1951. [CrossRef]
31. Alzubaidi, F.; Makuluni, P.; Clark, S.R.; Lie, J.E.; Mostaghimi, P.; Armstrong, R.T. Automatic Fracture Detection and Characterization from Unwrapped Drill-Core Images Using Mask R-CNN. *J. Pet. Sci. Eng.* **2022**, *208*, 109471. [CrossRef]
32. Kim, S.; Lee, K.; Lee, M.; Lee, J.; Ahn, T.; Lim, J.T. Evaluation of Saturation Changes during Gas Hydrate Dissociation Core Experiment Using Deep Learning with Data Augmentation. *J. Pet. Sci. Eng.* **2022**, *209*, 109820. [CrossRef]
33. Wang, H.; Wu, W.; Chen, T.; Dong, X.; Wang, G. An Improved Neural Network for TOC, S1 and S2 Estimation Based on Conventional Well Logs. *J. Pet. Sci. Eng.* **2019**, *176*, 664–678. [CrossRef]
34. Wang, H.; Lu, S.; Qiao, L.; Chen, F.; He, X.; Gao, Y.; Mei, J. Unsupervised Contrastive Learning for Few-Shot TOC Prediction and Application. *Int. J. Coal Geol.* **2022**, *259*, 104046. [CrossRef]
35. Souza, J.F.L.; Santos, M.D.; Magalhães, R.M.; Neto, E.M.; Oliveira, G.P.; Roque, W.L. Automatic Classification of Hydrocarbon “Leads” in Seismic Images through Artificial and Convolutional Neural Networks. *Comput. Geosci.* **2019**, *132*, 23–32. [CrossRef]
36. Lei, M.; Rao, Z.; Wang, H.; Chen, Y.; Zou, L.; Yu, H. Maceral Groups Analysis of Coal Based on Semantic Segmentation of Photomicrographs via the Improved U-Net. *Fuel* **2021**, *294*, 120475. [CrossRef]
37. Santos, R.B.M.; Augusto, K.S.; Iglesias, J.C.A.; Rodrigues, S.; Paciornik, S.; Esterle, J.S.; Domingues, A.L.A. A Deep Learning System for Collotelinite Segmentation and Coal Reflectance Determination. *Int. J. Coal Geol.* **2022**, *263*, 104111. [CrossRef]
38. Feng, R. Estimation of Reservoir Porosity Based on Seismic Inversion Results Using Deep Learning Methods. *J. Nat. Gas Sci. Eng.* **2020**, *77*, 103270. [CrossRef]
39. Wang, J.; Cao, J.; Yuan, S. Deep Learning Reservoir Porosity Prediction Method Based on a Spatiotemporal Convolution Bi-Directional Long Short-Term Memory Neural Network Model. *Geomech. Energy Environ.* **2021**, 100282. [CrossRef]
40. Wu, J.; Yin, X.; Xiao, H. Seeing Permeability from Images: Fast Prediction with Convolutional Neural Networks. *Sci. Bull.* **2018**, *63*, 1215–1222. [CrossRef]
41. Zeng, K.; Wang, Y. A Deep Convolutional Neural Network for Oil Spill Detection from Spaceborne SAR Images. *Remote Sens.* **2020**, *12*, 1015. [CrossRef]
42. de Lima, R.P.; Marfurt, K. Convolutional Neural Network for Remote-Sensing Scene Classification: Transfer Learning Analysis. *Remote Sens.* **2020**, *12*, 86. [CrossRef]
43. Hu, F.; Xia, G.S.; Hu, J.; Zhang, L. Transferring Deep Convolutional Neural Networks for the Scene Classification of High-Resolution Remote Sensing Imagery. *Remote Sens.* **2015**, *7*, 14680–14707. [CrossRef]
44. Yu, D.; Xu, Q.; Guo, H.; Zhao, C.; Lin, Y.; Li, D. An Efficient and Lightweight Convolutional Neural Network for Remote Sensing Image Scene Classification. *Sensors* **2020**, *20*, 1999. [CrossRef]
45. Bogdal, C.; Schellenberg, R.; Lory, M.; Bovens, M.; Höpli, O. Recognition of Gasoline in Fire Debris Using Machine Learning: Part II, Application of a Neural Network. *Forensic Sci. Int.* **2022**, *332*, 111177. [CrossRef]
46. Risum, A.B.; Bro, R. Using Deep Learning to Evaluate Peaks in Chromatographic Data. *Talanta* **2019**, *204*, 255–260. [CrossRef]
47. Demšar, J.; Erjavec, A.; Hočevár, T.; Milutinovič, M.; Možina, M.; Toplak, M.; Umek, L.; Zbontar, J.; Zupan, B. *Orange: Data Mining Toolbox in Python* Tomaž Curk Matija Polajnar Laň Zagar. 2013, Volume 14. Available online: <https://jmlr.csail.mit.edu/papers/v14/demsar13a.html> (accessed on 29 September 2023).
48. Godec, P.; Pančur, M.; Ilenič, N.; Čopar, A.; Stražar, M.; Erjavec, A.; Pretnar, A.; Demšar, J.; Starič, A.; Toplak, M.; et al. Democratized Image Analytics by Visual Programming through Integration of Deep Models and Small-Scale Machine Learning. *Nat. Commun.* **2019**, *10*, 4551. [CrossRef] [PubMed]
49. Pires de Lima, R.; Duarte, D. Pretraining Convolutional Neural Networks for Mudstone Petrographic Thin-Section Image Classification. *Geosciences* **2021**, *11*, 336. [CrossRef]
50. Ribani, R.; Marengoni, M. A Survey of Transfer Learning for Convolutional Neural Networks. In Proceedings of the 2019 32nd SIBGRAPI Conference on Graphics, Patterns and Images Tutorials (SIBGRAPI-T), Rio de Janeiro, Brazil, 28–31 October 2019; pp. 47–57. [CrossRef]

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.