

Article

Not peer-reviewed version

A Biological-Inspired Deep Learning Framework for Big Data Mining and Automatic Classification in Geosciences

[Paolo Dell'Aversana](#) *

Posted Date: 4 February 2025

doi: 10.20944/preprints202502.0181.v1

Keywords: Mineralogical Big Data; Rock samples; Deep Learning; Geosciences; Mycelia network



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Article

A Biological-Inspired Deep Learning Framework for Big Data Mining and Automatic Classification in Geosciences

Paolo Dell'Aversana

Independent researcher; dellaversanapaolo@gmail.com

Abstract: MycelialNet is a novel deep neural network (DNN) architecture inspired by natural mycelial networks. Mycelia, the vegetative part of fungi, form extensive underground networks that connect in a very efficient way biological entities, transport nutrients and signals, and dynamically adapt to environmental conditions. Drawing inspiration from these properties, MycelialNet integrates dynamic connectivity, self-optimization, and resilience into its artificial structure. This paper explores how mycelial-inspired neural networks can enhance big data analysis, particularly in mineralogy, petrology, and other Earth disciplines, where exploration and exploitation must be efficiently balanced during the process of data mining. We validate our approach by applying MycelialNet to synthetic data first, and then to a large petrological database of volcanic rock samples, demonstrating its superior feature extraction, clustering, and classification capabilities with respect to other conventional machine learning methods.

Keywords: Mineralogical Big Data; Rock samples; Deep Learning; Geosciences; Mycelia network

1. Introduction

The field of geoscience has undergone a radical transformation with the advent of big data and a large variety of machine learning (ML) methods [1–4]. These technologies have enabled the development of advanced algorithms and models capable of processing and analyzing vast amounts of geoscientific data, often characterized by complexity, multi-scale dimensions, and diverse data types [5–7]. As geological and geophysical exploration continue to generate enormous volumes of data from various sources—ranging from seismic surveys to remote sensing and rock sample analysis—the need for more sophisticated tools to extract meaningful insights from these datasets has become paramount. Deep learning methods, particularly in their ability to recognize patterns and make predictions, have proven to be powerful tools in geoscientific research, providing innovative solutions to challenges such as model inversion, data classification, and the identification of subsurface features. However, while machine learning algorithms excel at processing big data, they often lack the adaptive and decentralized intelligence observed in many natural biological systems.

This is where interdisciplinary research, particularly insights from biology, offers a compelling opportunity for innovation. For instance, mycology, the branch of biology dedicated to the study of fungi, provides fascinating models of decentralized intelligence, particularly through mycelial networks [8]. These networks exemplify efficient exploration and resource distribution strategies in natural ecosystems, making them an ideal inspiration for advanced AI models, particularly deep neural networks. Indeed, mycelial networks play a fundamental role in natural ecosystems by interconnecting plant roots and other biological entities through vast underground structures (see also supplementary materials at the end of the paper). These networks function as information highways, transporting nutrients, signaling molecules, and even electrical impulses. Their ability to dynamically explore their environment, adapt to stress, and optimize resource distribution has profound implications for artificial intelligence. As anticipated earlier, a key aspect of the fungal life is the mycelium. This is an intricate network of fine fungal threads beneath the forest floor. It forms

a vast, interwoven structure that permeates the organic geological layer. The majority of fungal biomass exists underground in this form, extending through the soil and intertwining with plant roots, soil organisms, and microbial communities. While mushrooms are commonly associated with fungi, they are merely the visible fruiting bodies of an expansive mycelial network that remains hidden beneath the surface. Despite their vast size, these underground networks remain largely unnoticed until they produce mushrooms or other fruiting structures. However, their role in ecological balance is crucial, particularly through the Common Mycelial Network (CMN), also known as the Common Mycorrhizal Network. In a CMN, plant and tree roots interconnect via a mycelial web, forming symbiotic partnerships with fungi that enable the exchange of isotopic carbon, nitrogen, phosphorus, water, and biochemical signals across species, space, and time.

Drawing inspiration from this natural network system, in this paper we propose “MycelialNet”, a biologically inspired deep learning framework designed to enhance big data analysis in geosciences (as well as in other scientific fields). MycelialNet mimics the decentralized, adaptive intelligence of mycelial networks to create robust and dynamic machine learning systems. Unlike traditional deep learning models, which rely on rigid, predefined structures, MycelialNet embraces a self-organizing, exploratory approach that balances efficient data mining with targeted exploitation of valuable patterns. By integrating biological principles [9–15] —such as emergent intelligence, adaptive learning, and distributed information processing—into AI architectures, MycelialNet offers a novel way to tackle the complexity and variability inherent in large geoscientific datasets.

Motivated by these considerations, we introduce a comprehensive approach that combines deep learning, reinforcement learning, and biological principles to address key challenges in the classification and interpretation of extensive rock sample datasets. By leveraging the fundamental properties of mycelial networks, we demonstrate how biologically (mycologically) inspired AI can improve the efficiency and accuracy of geoscientific data analysis. Specifically, our approach integrates exploratory search mechanisms, resilience to uncertainty, self-aware capabilities, and multi-scale pattern recognition, ensuring robust performance in big data mining even in case of complex and heterogeneous datasets.

Ultimately, this interdisciplinary framework highlights the potential of blending biological sciences, artificial intelligence, and geoscience to develop next-generation computational tools for Earth sciences. By integrating these domains, we aim to open new avenues for more effective and computationally efficient big data analysis, offering innovative solutions to some of the most pressing challenges in geoscientific research.

2. Methodology

2.1. Key Components of MycelialNets

A core aspect of our biological-inspired approach to big data analysis in geosciences is the implementation of simulated Mycelial deep neural networks, here named “MycelialNets”. As anticipated in the introduction, these use a biologically (mycologically) inspired neural architecture that emulates the extraordinary adaptability of fungal networks. The core components include the following basic aspects:

- 1) “MycelialLayer”: this is a dynamic layer that adjusts its connectivity during training, pruning weak connections while regenerating new ones to optimize learning pathways.
- 2) Dynamic Connectivity: this functionality is inspired by mycelial exploration strategies. The network restructures itself iteratively, mirroring the adaptability of fungal networks.
- 3) Self-Monitoring Mechanism: the MycelialNet model incorporates self-reflection mechanisms. This aspect is inspired to the self-awareness of biological brain, as discussed in previous works [16]). Adjusting its connectivity ratio based on performance metrics such as accuracy, the

MycelialNet model can continuously monitor itself, adapting its own architecture to dynamic environmental conditions on time-varying data sets.

- 4) Exploration Factor: this is an additional component that encourages the model to explore diverse configurations and hyperparameters. It provides a dynamic balance between exploration/exploitation ratio, when the hyper-parameter space is explored by the network model, with the final goal to set an optimal MycelialNet architecture.

In the next sub-section, we define in a more quantitative way the concepts here just mentioned, using a detailed mathematical formulation of the MycelialNet model.

2.2. Mathematical Formulation

Let $X \in R^{m \times n}$ be the input data matrix, where m is the number of samples (for instance, rock samples) and n is the number of features (for instance, major oxides).

Let $Y \in R^{m \times k}$ the corresponding labels for classification with k output classes (for instance, Basalt, Andesite, Diorite, etc.).

Each input passes through multiple layers of dynamically changing artificial neurons.

Unlike standard artificial neural networks with fixed connections, MycelialNet introduces a time-dependent weight matrix, W_t , that dynamically adapts:

$$W_t = M_t \odot W_{t-1} \quad (1)$$

where:

- W_t is the weight matrix at time t ,
- W_{t-1} is the weight matrix at the previous time step $t-1$,
- $M_t \in \{0,1\}^{n \times d}$ is a binary mask matrix controlling active connections at time t ,
- n is the number of features, m is the number of samples (as anticipate earlier),
- d is the number of neurons in the layer.
- $\odot$ represents the Hadamard (elementwise) product,

The mask M_t is updated dynamically based on a connectivity ratio c_t :

$$M_t = 1(U_t < c_t) \quad (2)$$

where U_t is a uniform random matrix. The formula means that the mask M_t is updated by comparing each element of a random matrix U_t with the connectivity ratio c_t . Those elements that are smaller than c_t are “activated” (set to 1), and others are “deactivated” (set to 0). This dynamic update of the mask can be used to model how elements of a system are connected or disconnected in response to a changing connectivity threshold. The connectivity ratio c_t evolves over training:

$$c_{t+1} = c_t + \eta \nabla L_t \quad (3)$$

where:

- η is the learning rate,
- ∇L_t is the gradient of the loss function L at time t .
- The total loss function is

$$L_{total} = - \sum_{i=1}^m \sum_{j=1}^k y_{ij} \log(\hat{y}_{ij}) \quad (4)$$

where:

- m is the number of samples (e.g., rock samples), as state earlier,
- k is the number of output classes (e.g., types of rock), as state earlier,
- y_{ij} is the true label (one-hot encoded, where $y_{ij} = 1$ for the true class and $y_{ij} = 0$ otherwise),
- $\hat{y}_{ij}$ is the predicted probability for the j -th class for the i -th sample, computed using the Softmax function.

We briefly remind that the Softmax function is a mathematical function commonly used in machine learning, particularly in multi-class classification problems. It takes a vector of real numbers

as input and converts it into a probability distribution, where each element is in the range (0, 1) and the sum of all elements equals 1. In this case, we compute the Softmax for each sample and class. The final loss is the sum over all samples and classes.

Coming back to the computation of the gradient of the loss function, in our case, higher gradients increase MycelialNet connectivity, mimicking the mycelial network's expansion in response to environmental stimuli.

For a given neuronal layer l , the activation H_t^l at time t is computed as:

$$H_t^l = \sigma(H_t^{l-1} W_t^l + \mathbf{b}^l) \quad (5)$$

where:

- H_t^{l-1} is the activation from the previous neuronal layer,
- W_t^l is the dynamically adjusted weight matrix,
- $\mathbf{b}^l$ is the bias vector,
- σ is an activation function (e.g., Rectified Linear Unit, briefly ReLU, or sigmoid, as well as other activation functions settable by the user).

The output of the final layer is computed as:

$$\hat{\mathbf{Y}} = \text{Softmax} (H_t^{(P)} W_t^{(P)} + \mathbf{b}^{(P)}) \quad (6)$$

where P is the total number of layers.

This equation intuitively means that the final output of the neural network is computed in a multi-class classification problem. The model's final layer uses weights and biases to compute raw scores (logits), and the Softmax function is then applied to transform these raw scores into a probability distribution, making it suitable for classification tasks. The network minimizes a standard **cross-entropy loss** for classification. Additionally, we introduce a regularization term to encourage network sparsity. The gradient update rule for weights is given by:

$$W_{t+1} = W_t - \alpha \frac{\partial L_{total}}{\partial W_t} \quad (7)$$

This equation allows optimizing the weights in the MycelialNet model by moving them in the direction that reduces the total loss function. By iteratively applying this rule, the model learns to make better predictions. The learning rate α controls how quickly or slowly the weights are updated in each iteration.

Finally, to balance exploration and exploitation of the parameters and hyper-parameters space, we introduce an **entropy-based connectivity adjustment**:

$$c_{t+1} = c_t + \gamma H(X) \quad (8)$$

where $H(X)$ is the entropy of the activations:

$$H(X) = -\sum_{i=1}^N p_i \log p_i \quad (9)$$

This is the standard formula for the Shannon entropy, which measures the uncertainty in the system's state. It is used here to measure the "spread" or uncertainty in the activations, guiding the network's adaptability. Higher entropy leads to increased connectivity attempting to reduce uncertainties, mimicking mycelial expansion in high-information regions.

3. Simulations

Previous works have demonstrated the efficacy of both chemical composition and thin section images for automatic rock (and facies) classification using various machine learning techniques [17–20]. These are particularly valuable in the context of big data in geosciences, where large datasets with complex features require sophisticated methods to extract meaningful insights. To verify the effectiveness of our MycelialNet method we apply it to both synthetic and real-world mineralogical datasets, comparing performance against traditional classifiers. Key performance evaluation metric is based on classification accuracy.

3.1. First Synthetic Test

The first classification test here discussed simulates a three-category classification task, where the goal is to classify a simulated large dataset of rock samples into three distinct hypothetical rock types. The features used for classification are typical oxides commonly found in geological studies, such as TiO_2 , SiO_2 , and others. These features are crucial in identifying different types of rocks based on their mineral composition, as we will see in the real-data test discussed ahead. The synthetic dataset consists of 1000 simulated rock samples, each with features which represent the concentrations or ratios of these oxides in each sample. The neural network model was trained on this data using the MycelialNet approach. This proved to be highly effective in handling the complexity of the data. In fact, the accuracy achieved by the model is relatively high (89%), demonstrating its ability to make reliable classifications despite the potential variability in the input features (Figure 1, left panel). The convergence during training was both fast and effective. Looking at the trend of the loss function (Figure 1, right panel), it is probable that increasing the number of epochs, accuracy can reach values between 90 and 95% for this test. Thanks to the entropy-driven adjustments in connectivity, the model was able to learn efficiently and avoid overfitting, leading to quick stabilization of the cost function (loss) as training progressed. This was reflected in the cost function trend, where a rapid and regular reduction in loss was observed over epochs. The entropy-based updates help the model adapt to the underlying data distribution, improving both speed and stability during the optimization process. Overall, in this test, the MycelialNet approach demonstrated significant advantages in both speed and accuracy, offering a robust solution for the classification of rock samples based on their oxide compositions.

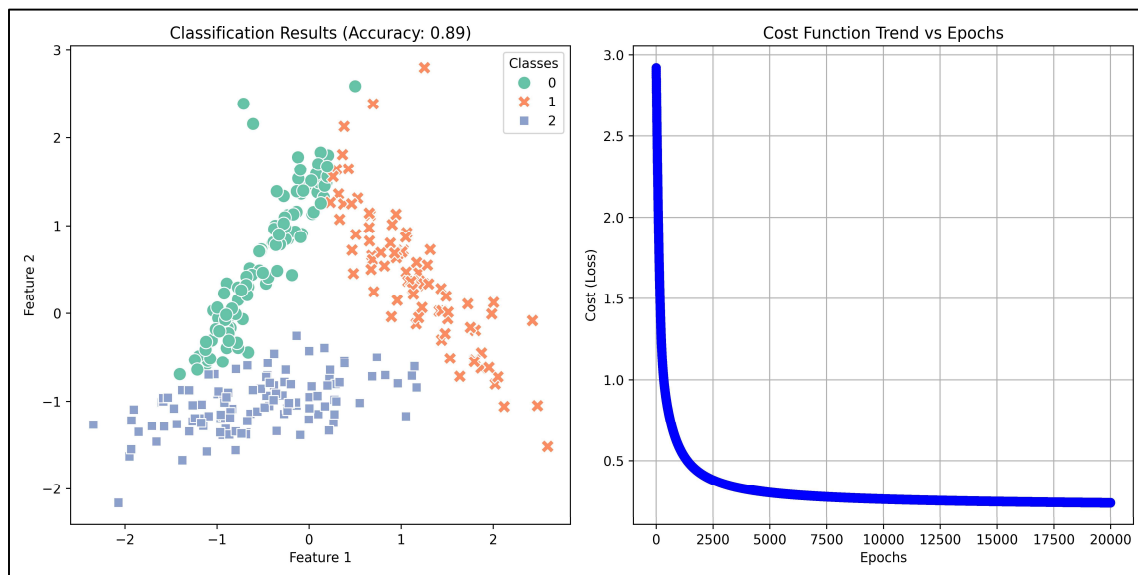


Figure 1. Scatter plot of classified rock samples (left panel) here represented with different symbols and colors. Loss function trend vs. epochs (right panel).

3.2. Addressing Non-Linear Classification Challenges with MycelialNet

A fundamental challenge in machine learning classification tasks arises when class boundaries in the feature space are non-linear. Many traditional classification models struggle with such problems because they rely on linear decision boundaries, making them ineffective for complex datasets where classes are intertwined in intricate ways. The MycelialNet model, however, overcomes this limitation thanks to its adaptive and dynamic connectivity. Inspired by the resilient and self-optimizing nature of fungal mycelial networks, the model continuously adjusts its internal structure, selectively pruning and regenerating connections to improve performance on challenging

classification tasks. This adaptability allows it to handle large, complex datasets with greater efficiency compared to conventional neural networks.

The second test discussed in this paper highlights the power of MycelialNet in addressing non-linearly separable classification problems. To demonstrate this, we use a synthetic dataset based on the “make_moons” function, which generates a dataset where the two classes are intertwined in a “crescent-moon” shape. Traditional linear classifiers often fail in this scenario, whereas MycelialNet successfully captures the underlying patterns and effectively distinguishes between the classes. In fact, this model shows strong adaptive pruning and regeneration capabilities. In simple words, the model masks and reconfigures connections during training, ensuring that it adapts to complex patterns dynamically.

Also for this test, as in the previous one, the network consists of multiple MycelialLayers, each capable of adjusting its structure dynamically. The final dense layer produces the output classification. Figure 2 shows the scatter plot of the synthetic data, displaying the distribution of data points with their respective classes. The “Decision Boundary Plot” illustrates how MycelialNet effectively separates the two intertwined classes, showcasing its superior ability to handle non-linear classification tasks.

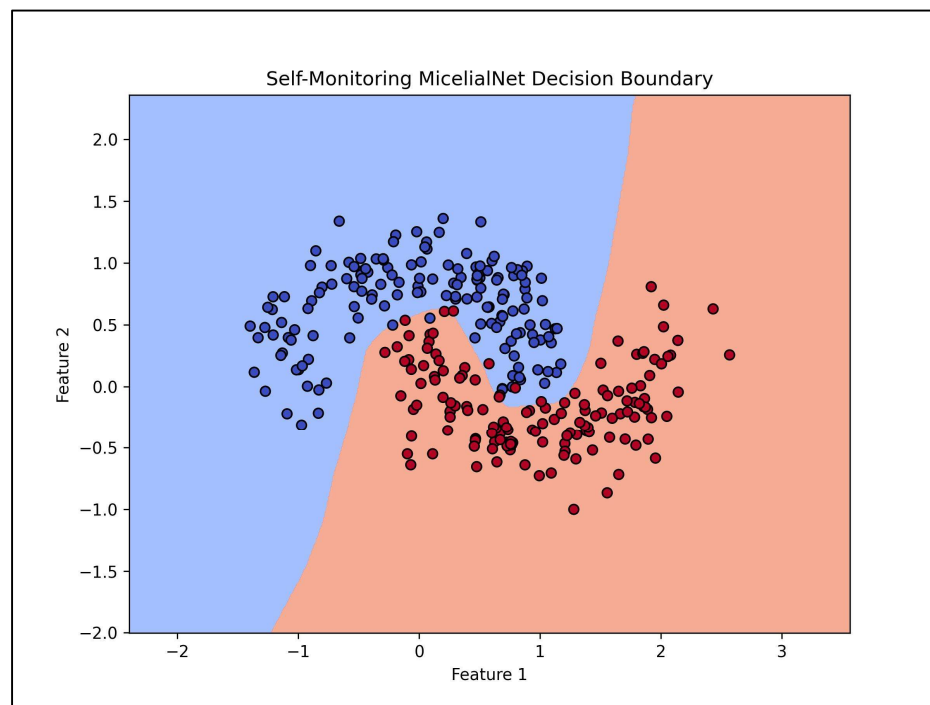


Figure 2. Scatter plot of the two classes (here represented with different colors) and final decision boundary obtained through the application of the MycelialNet classification model.

4. Test on a Real Data Set of Rock Samples

4.1. Introducing the Test

After discussing synthetic test, we apply the MycelialNet to a real data set. We remark that this type of deep neural network model is designed to perform both supervised and unsupervised analysis of big datasets, offering powerful capabilities for data mining and statistical analysis. It is particularly well-suited for tasks where large amounts of data must be processed, analyzed, and categorized, such as the case of rock sample datasets in mineralogical and petrological disciplines or geophysics. In the case of rock sample datasets, MycelialNet can be trained to classify different types of rocks (e.g., Andesite, Basalt, Granite) based on their chemical composition, specifically oxide

percentages (SiO_2 , Al_2O_3 , Fe_2O_3 , etc.), by learning patterns in the data that map these features to rock types. However, it is well known that every automatic classification approach is improved if it is anticipated by adequate analysis of the entire data set, allowing clear identification of key features, correlations, and hidden relationships between the data. For that reason, MycelialNet is designed also for performing effective statistical analysis, unsupervised learning and clustering tasks, where there are no predefined labels. MycelialNet mining techniques includes identifying hidden correlations between different oxides and rock types or identifying anomalies in the dataset (such as outliers in chemical composition). This type of deep learning method allows assessing how different oxide percentages correlate with each other, identifying highly correlated features (which can be important for feature selection and dimensionality reduction). Moreover, it is very effective in determining which features (oxides, in this case) have the most influence on the classification of rock types, as well as in identifying groups of unlabeled rock samples that share similar chemical compositions. MycelialNet also integrates self-reflection mechanisms [16], which allow the model to evaluate and modify its architecture dynamically based on feedback from the data. This helps improve the model's performance by optimizing its internal parameters during data mining, and adjusting its structure (e.g., layer sizes, neuron connections) for better representation of the data. This feature is particularly useful in unsupervised learning scenarios, where the model may adapt its structure to uncover patterns not originally anticipated.

In the test discussed in this section, we applied MycelialNet to a real rock-sample dataset. We started with unsupervised clustering of rock types based on their oxide content, helping geologists identify natural groupings or categories of rocks based on their chemical signatures. Next, we performed correlation analysis between oxides, to determine which oxides tend to vary together, which could provide insights into how geological processes influence rock formation. After performing the unsupervised analysis, we trained MycelialNet in a supervised learning process to classify rock samples into predefined categories (e.g., Andesite, Basalt, etc.). For that classification task, the model uses labeled training data (with known rock types, based on analysis by human experts) and learn from these samples to predict the rock type of new, unseen samples. We evaluated the performances of the MycelialNet model using standard metrics for classification (accuracy, precision, recall, AUC) By incorporating cross-validation, the model can provide robust performance metrics.

4.2. The Data Set

The dataset used in this test is sourced from the publicly available GEOROC (Geochemistry of Rocks of the Oceans and Continents) database (GEOROC website, accessed on 15/01/2025). It consists of a compilation of more than 1500 major- and trace-element data points, and 570 Pb-isotopic analyses of Mesozoic-Cenozoic (190–0 Ma) magmatic rocks in southern Peru, northern Chile, and Bolivia (Central Andean orocline) [21]. The chemical oxides used for this classification test include: SiO_2 , TiO_2 , Al_2O_3 , Fe_2O_3 , MnO , MgO , CaO , Na_2O , K_2O , and P_2O_5 (in weight %). The rock samples encompass both the dominant rock types, such as Andesite, Basaltic Andesite, Rhyolite, and Dacite, as well as less common classes with far fewer samples. The chemical and mineralogical compositions of these rock types exhibit considerable overlap, adding complexity to the classification process. As highlighted in the previous synthetic test, one of the key challenges in this classification task is the presence of non-linear decision boundaries, due to the overlapping chemical compositions of different rock classes. Traditional machine learning models, such as Support Vector Machines (SVMs), Random Forest, and conventional Neural Networks, often struggle with this high-dimensional, non-linearly separable dataset. For that reason, we attempted to perform this classification task using MycelialNet model, comparing the results with those obtained by other “traditional” machine learning methods.

4.3. Workflow

We started performing accurate data mining by creating a complete correlation matrix (Figure 3). Creating a correlation matrix in data mining is essential for understanding the quantitative relationships between oxides in a large rock sample dataset. That is crucial for multiple reasons. First, correlations can reveal geochemical associations. In fact, certain oxides co-vary due to mineralogical and petrogenetic processes (e.g., Al_2O_3 and K_2O in feldspars). Second, strongly correlated oxides indicate redundancy, allowing for dimensionality reduction and feature selection to improve computational efficiency. Third, the correlation matrix helps cluster rock types based on oxide interdependencies. Next, this type of data representation allows identifying which oxides best separate rock types, improving decision boundary accuracy. Moreover, it can reveal latent geochemical trends useful for both classification and exploratory analysis. Finally, correlation matrix help highlight economic mineralization trends. For instance, correlations between oxides (e.g., Fe_2O_3 with TiO_2) can indicate ore deposit formation.

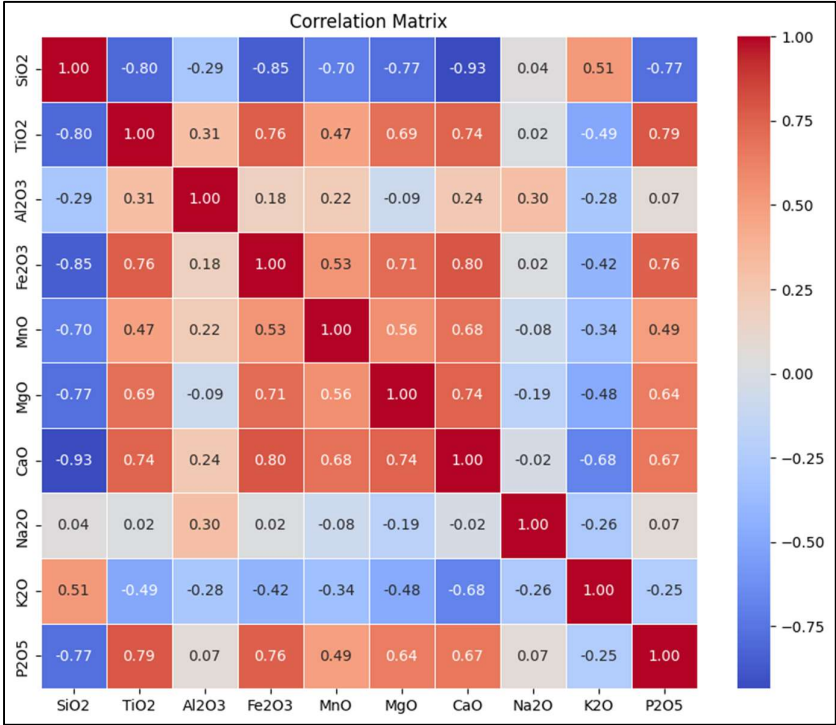


Figure 3. Oxides correlation matrix.

Next, we created histograms of all oxides (Figure 4). The x-axis of each histogram corresponds to the concentration of a specific oxide in the samples, measured in percentages or weight ratios (e.g., SiO_2 , Fe_2O_3 , Al_2O_3). Each bin in the histogram represents a range of values for that oxide, showing how often certain concentration levels appear in the dataset. The y-axis represents the frequency (or count) of samples that fall into each concentration range for a given oxide. Higher bars indicate a higher concentration of samples in that range, while shorter bars represent fewer samples in that range. These histograms help identify asymmetrical distributions, guiding MycelialNet to apply adaptive normalization techniques instead of conventional scaling, which might misrepresent feature importance. MycelialNet uses histogram-based variance analysis to automatically drop or reduce weight for such features, making the model more computationally efficient. Furthermore, histograms help visualize how distributions overlap, guiding feature engineering for MycelialNet’s dynamic connectivity. In addition, histograms allow MycelialNet to detect natural groupings of rock types before labels are applied, leading to a more efficient classification process. Finally, if histograms show that an oxide has a multimodal distribution, MycelialNet can apply different learning rates for different sub-populations in the data, improving convergence speed. Another important aspect

strictly linked with the mining industry is that histograms can reveal geochemical signatures of ore deposits. In fact, many ore deposits are characterized by specific oxide enrichments (e.g., Fe_2O_3 for iron ore, TiO_2 for titanium deposits). Histograms help identify whether certain oxide levels correlate with economic mineralization, guiding prospecting models. In conclusion, histograms of oxides are not just visual tools. Instead, they are a fundamental step in intelligent data mining, feature selection, and classification. That is true in general, but it is important especially when using MycelialNet. They enhance the model’s ability to dynamically adapt, extract meaningful patterns, and make accurate predictions in large-scale rock sample datasets.

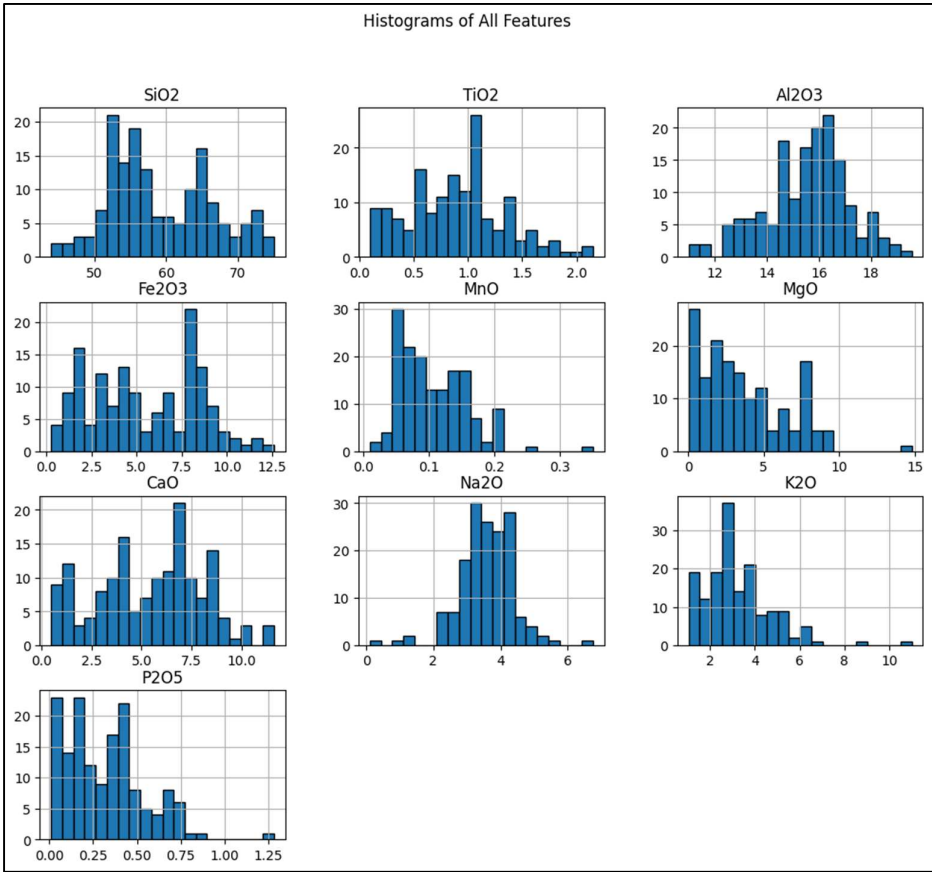


Figure 4. Oxides histograms in the data set (see text for explanations).

An additional step of the unsupervised data mining workflow is to visualize the oxides content for each rock type (Figure 5). When using MycelialNet, visualizing the oxides content for each rock type is essential in the unsupervised data mining workflow because it helps uncover geochemical patterns before classification. This step identifies the specific geochemical signatures of different rock types, enhancing the ability to cluster similar samples and refine feature selection. It also improves data interpretation by validating results against geological knowledge while detecting potential outliers or rare rock types. Additionally, it supports dimensionality reduction by guiding techniques like Principal Component Analysis (PCA), and helps in designing more effective supervised learning strategies. By first analyzing oxide distributions, MycelialNet structures the data more effectively, leading to more accurate and meaningful classifications.

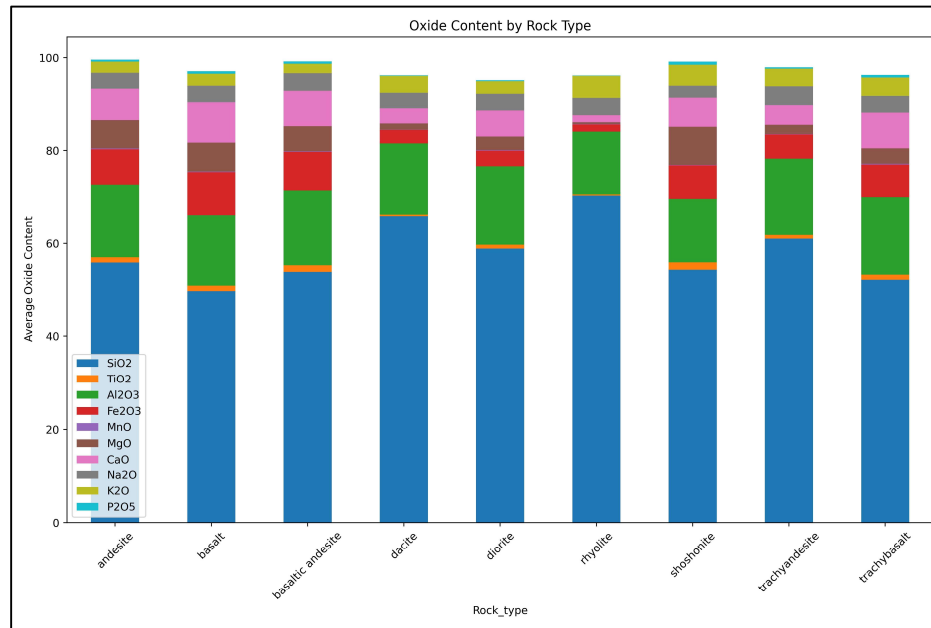


Figure 5. Oxides content for each rock type.

4.4. Supervised Learning and Classification Results

After accurate analysis of the data set, we applied the MycelialNet model to perform the supervised classification of the rock samples. Classification accuracy was almost 90% on the test-data set. (This result is particularly significant given the complexity and variability of the dataset). The network was trained on a percentage of the data set ranging between 80 and 90%. The remaining unlabeled data was used as test data set. The classification results were visualized using cross-plots that display the predicted rock-class labels in a two-feature space. Figure 6 presents an example of a classification cross-plot where the test data are color-coded according to their assigned rock-class labels in the SiO_2 – TiO_2 feature space. This visualization highlights the model's ability to distinguish rock types, with some uncertainties, based on their oxide composition. The trend is generally correct: SiO_2 tends to increase from basalt, to andesite, dacite, trachyandesite, rhyolite. Instead TiO_2 decreases with increasing SiO_2 , being highest in basalts-basaltic andesite, and lowest in rhyolites. There is some minor misclassification case. For instance, some diorite samples (yellow dots) appear at lower SiO_2 (around 50-52%), where basalts and basaltic andesites should be. A possible explanation is that some diorite compositions can be transitional or can contain mafic inclusions. More in general, rock classifications based on oxides are not always rigid: some transitional types exist. Furthermore, geochemical variations in different samples or minor alteration effects can influence SiO_2 and TiO_2 values. In addition, we remark that this plot (as well as the plot in Figure 7) considers the projection in a two-feature space only (SiO_2 and TiO_2), but our classification is based on multiple oxides (e.g., K_2O , Na_2O , Al_2O_3).

Similarly, Figure 7 provides an additional classification display, showing the results in the Fe_2O_3 – MgO feature space. General Trends: Fe_2O_3 and MgO decrease as rocks evolve from basalt, andesite, and basaltic classes to dacite, trachyandesite and rhyolite classes. Mafic rocks (basalt, andesite) show high Fe_2O_3 and MgO , while felsic rocks (dacite, rhyolite) show low levels of these oxides. Some minor misclassification cases can be justified as in Figure 6.

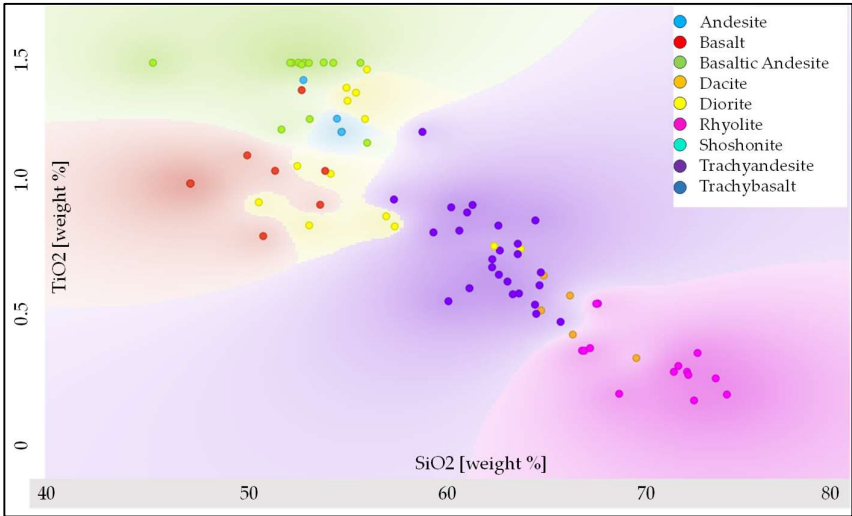


Figure 6. Example of test-data classification cross-plot, showing with different colors in the 2-feature display of SiO₂ and TiO₂.

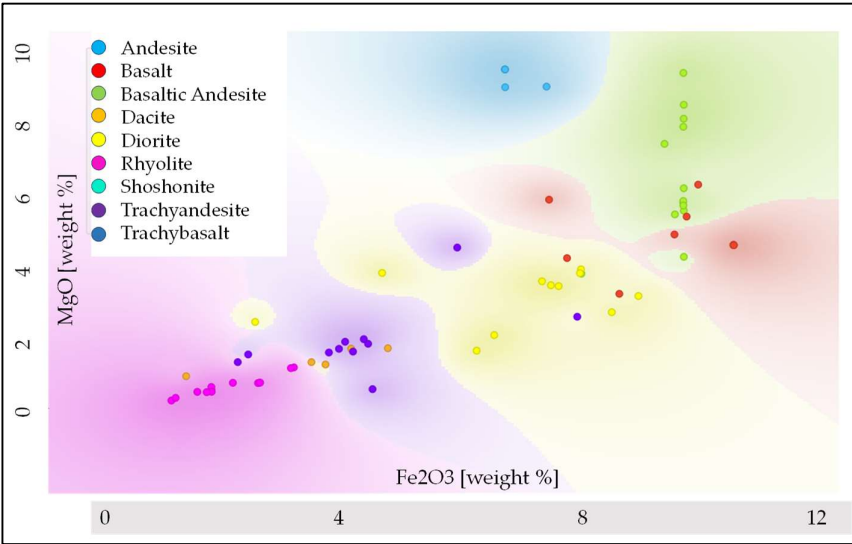


Figure 7. Example of test-data classification cross-plot in the 2-feature display of Fe₂O₃ and MgO.

In summary, both figures illustrate the effectiveness of MycelialNet in correctly assigning rock classes while preserving the geochemical relationships within the dataset. For comparison purposes, we applied other types of classifiers to the same data, and MycelialNet model demonstrated the highest classification accuracy, as summarized in Table 1.

Table 1. Accuracy comparison.

Method	Accuracy
Random Forest	0.625
Logistic Regression	0.65
Standard Neural Network	0.69
MycelialNet model	0.875

5. Discussion

All the tests discussed in the previous sections (on both simulated and real data) show that the MycelialNet model offers several advantages over conventional deep learning models when it comes to data mining, unsupervised learning, and supervised learning. These advantages stem from its unique architecture and learning mechanisms, which make it more adaptable, self-organizing, and efficient in handling complex datasets. Unlike conventional machine learning models with static architectures, MycelialNet incorporates self-organizing connectivity, inspired by the behavior of mycelial networks in nature. This means that the model dynamically adjusts its internal neuron connections based on data patterns, instead of relying on pre-defined layers and connections. This allows it to reconfigure itself for both unsupervised data mining and supervised classification. As shown by both synthetic and real data tests, this approach enables a more organic analysis of information, avoiding bottlenecks and reducing unnecessary computational costs. This helps automatically uncover hidden relationships in large datasets without requiring predefined structures, making it superior to conventional models that often struggle with fixed architectures. As clearly shown in the test on real data, MycelialNet first analyzes the dataset without labels, using clustering, density estimation, and correlation studies to understand the structure of the data. It can self-discover relationships between features and classes before formal classification begins. Once patterns are identified, the model can transition to supervised learning, using labeled data to train a classifier more efficiently. The prior unsupervised step enhances generalization because the network already understands the data distribution. In other words, instead of blindly training on labeled rock types, MycelialNet first explores systematically the oxide compositions, identifies hidden clusters, and then fine-tunes a classification model. This makes the final model more accurate and generalizable compared to conventional neural networks or other machine learning methods. Furthermore, MycelialNet integrates self-aware deep learning techniques, allowing it to self-evaluate its own learning process and self-adjust parameters dynamically, without any external human intervention. Finally, MycelialNet model evolves by allowing multiple competing subnetworks to solve the same task and selecting the most optimal configuration dynamically. Instead of converging on a single fixed solution (as conventional models do), MycelialNet continuously evaluates and adapts, ensuring higher accuracy and more diverse representations of data. This capability has a significant impact on rock classification: in fact, different rock types can have similar oxide compositions, making classification challenging. MycelialNet tests multiple evolving decision pathways, leading to more robust and confident classifications. In summary, MycelialNet is not just a deep learning model; it is a dynamic, evolving system that integrates self-awareness, biological inspiration, and hybrid learning to offer a fundamentally new approach to big data analysis.

6. Conclusions

Inspired by fungal mycelial networks, MycelialNet introduces a biologically inspired approach to deep learning that enhances big data mining and automatic classification. The integration of dynamic connectivity creates an efficient, self-adaptive neural architecture. Applied to geosciences, this approach facilitates better feature ranking, correlations discovery and high performance in classification tasks, offering a powerful tool for data-intensive research fields. For rock sample datasets, it allows for both exploratory geochemical studies and accurate supervised classification, making it an invaluable tool for geology, geophysics, and beyond. Future work will explore further refinements and applications in other domains, such as geophysical data inversion, composite well-logs analysis and so forth, where adaptive data analysis is critical.

Supplementary Materials: Supporting information can be found at: <https://www.spun.earth/>, where there is an accurate description and high-resolution images of mycelium. Furthermore, a very informative video about mycelium and fungal life is “How Fungi Make our Worlds”, by Merlin Sheldrake at <https://www.youtube.com/watch?v=ZRFmCXBv5R4>.

Author Contributions: Paolo Dell’Aversana: Conceptualization, methodology, software; validation, formal analysis, investigation, writing, review and editing.

Funding: This research received no external funding”

Data Availability Statement: The data set used in this paper is sourced from the publicly available GEOROC (Geochemistry of Rocks of the Oceans and Continents) database (GEOROC website, accessed on 15/01/2025). Link to the web site: <https://georoc.eu/georoc/new-start.asp>.

Conflicts of Interest: The author declares no conflicts of interest.

References

1. Russell, S. and Norvig, P., Artificial Intelligence: A Modern approach, Global Edition, published by Pearson Education, Inc. 2016, publishing as Prentice Hall.
2. Raschka, S. and Mirjalili, V., Python Machine Learning: Machine Learning and Deep Learning with Python, scikit-learn, and TensorFlow, 2nd Edition, PACKT Books, 2017.
3. Ravichandiran, S., Deep Reinforcement learning with Python, Packt Publishing, 2020.
4. Ribeiro, C. and Szepesvári, C., Q-learning combined with spreading: Convergence and results. In Proceedings of the ISRF-IEE International Conference: Intelligent and Cognitive Systems (Neural Networks Symposium), 1996, pages 32–36.
5. Barnes, A.E.; Laughlin, K.J., Investigation of methods for unsupervised classification of seismic data. In Expanded Abstracts; SEG Technical Program: Salt Lake City, UT, USA 2002; pp. 2221–2224. <https://doi.org/10.1190/1.1817152>.
6. Bestagini, P.; Lipari, V.; Tubaro, S., A machine learning approach to facies classification using well logs. In Expanded Abstracts; SEG Technical Program: Houston, TX, USA 2017; pp. 2137–2142. <https://doi.org/10.1190/segam,2017-17729805.1>.
7. Dell’Aversana, P. Comparison of different Machine Learning algorithms for lithofacies classification from well logs, Bull. Geophys. Oceanogr. 2017, 60, 69–80. <https://doi.org/10.4430/bgta0256>.
8. Sheldrake, M., Entangled life: how fungi make our worlds, change our minds & shape our futures. First US edition 2020, New York, Random House.
9. Damasio, A., Self Comes to Mind: Constructing the Conscious Brain; Pantheon: New York, NY, USA 2010, [Google Scholar].
10. Edelman, G.M., Neural Darwinism: The Theory of Neuronal Group Selection; Basic Books: New York, NY, USA 1987; ISBN 0-19-286089-5. [Google Scholar].
11. Edelman, G.M, Bright Air, Brilliant Fire: On the Matter of the Mind; Reprint Edition 1993; Basic Books: New York, NY, USA 1992; ISBN 0-465-00764-3. [Google Scholar].
12. Tononi, G.; Boly, M.; Massimini, M.; Koch, C., Integrated information theory: From consciousness to its physical substrate. Nat. Rev. Neurosci. 2016, 17, 450–461. [Google Scholar] [CrossRef] [PubMed].
13. Tononi, G.; Edelman, G.M., Consciousness and complexity. Science 1998, 282, 1846–1851. [Google Scholar] [CrossRef] [PubMed].
14. Panksepp, J. and Biven, L., The Archaeology of Mind: Neuroevolutionary Origins of Human Emotions (Norton Series on Interpersonal Neurobiology), 2012.
15. Panksepp, J., & Moskal, J., Dopamine and SEEKING: Subcortical “reward” systems and appetitive urges. In A. J. Elliot (Ed.) 2008, Handbook of approach and avoidance motivation (pp. 67–87). Psychology Press.
16. Dell’Aversana, P., Enhancing Deep Learning and Computer Image Analysis in Petrography through Artificial Self-Awareness Mechanisms. Minerals 2024, 14, 247. <https://doi.org/10.3390/min14030247>.
17. Dell’Aversana, P., Deep Learning for automatic classification of mineralogical thin sections. Bull. Geophys. Oceanogr. 2021, 62, 455–466. <https://doi.org/10.4430/bgo00367>.
18. Hall, B. Facies classification using machine learning. Lead. Edge 2016, 35, 906–909.
19. She, Y.; Wang, H.; Zhang, X.; Qian, W., Mineral identification based on machine learning for mineral resources exploration. J. Appl. Geophys. 2019, 168, 68–77.
20. Liu, K.; Liu, J.; Wang, K.; Wang, Y.; Ma, Y., Deep learning-based mineral classification in thin sections using convolutional neural network. Minerals 2020, 10, 1096.

21. Mamani, M., Wörner, G., & Sempere, T., Geochemical variations in igneous rocks of the Central Andean orocline (13°S to 18°S): Tracing crustal thickening and magma generation through time and space. GSA Bulletin 2010 122(1–2), 162–182. <https://doi.org/10.1130/B26538.1>.

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.