A Web-Based Information System for Ornamental Rock Classification and Similarity Search Using Siamese Networks

Carlos Henrique Costa Matos[1], **Hilario Seibel Junior**[1], **Karin Komati**[1]

[1]Programa de Pós-graduação em Computação Aplicada (PPComp)
Instituto Federal do Espirito Santo (IFES), Campus Serra
Av. dos Sabiás, 330 - Morada de Laranjeiras, Serra - ES, Brazil, 29166-630

carlos.1997matos@gmail.com, {hsjunior, kkomati}@ifes.edu.br

Abstract. ***Research Context****: The time-consuming, subjective, and error-prone process for identification of ornamental rocks, which is traditionally based on visual analysis and expert knowledge. This task is vital for competitiveness in construction, mining, and geology. Advances in computer vision create opportunities to automate such tasks and provide aid for professionals.* ***Scientific/Practical Problem****: The ornamental rock industry faces challenges in accurately and efficiently classifying rocks and retrieving similar units for high-demand users. The current manual effort is imprecise and slow, highlighting the need for AI solutions to improve the process.* ***Proposed Solution and/or Analysis****: This work proposes a web-based information system that uses Siamese Networks to aid analysts and general users in finding the class of a given ornamental rock, and its most similar pair of images in a dataset.* ***Related IS Theory****: Task-Technology Fit (TTF) guides our work. The proposed system is designed to allow smoother and quicker task completion, reducing time and effort. In addition, users can more easily and effectively achieve their desired outcomes. The smoother task execution can lead to lower costs associated with task performance.* ***Research Method****: A system is proposed using Siamese Networks to perform classification and similarity recognition of ornamental rocks.* ***Summary of Results****: The system successfully finds the class of the given image, as well as the image pair that most resembles it. The Siamese network comparison technique also allows for the correct identification of classes not used in training, but existing in the user's database.* ***Contributions and Impact to IS area****: This work bridges advanced vision models and decision support in ornamental rocks identification, transforming a neural network model into a reliable, evidence-based system and task-aligned evaluation.*

1. Introduction

Brazil stands out as one of the main global producers and exporters of ornamental rocks [Apex 2024], and the state of Espírito Santo has become the largest national exporter and a worldwide reference in the sector [FINDES 2024]. The identification of ornamental rocks is a step for various industrial sectors, such as civil construction, mining, and geology. The accurate classification of those materials is fundamental for the industry's competitiveness and sustainability. However, the traditional identification method, based on visual analysis and expert knowledge, is a time-consuming, subjective, and error-prone process [Zheng et al. 2024]. It is worth mentioning that the technical characterization of

a rock involves a series of complex laboratory tests, such as petrographic analysis and porosity tests, which are expensive and time-consuming, as evidenced by organizations such as the Mineral Technology Center [CETEM 2020]. Hence, with the exponential advancement of Artificial Intelligence (AI) and, in particular, of computer vision, an opportunity arises to optimize this process.

In this work, we propose a practical application for rock classification using Siamese neural networks, creating a support tool for both specialists and general users. The proposed solution does not aim to replace the specialist's role. Instead, it provides an additional tool that can be used during the validation process, increasing the efficiency and reliability of classifications, while also offering a new way to visualize the rocks without needing to handle them.

Figure 1 provides a visual representation of the model's separability and feature clustering. In this two-dimensional t-distributed Stochastic Neighbor Embedding (t-SNE) projection [Hinton and Roweis 2002], the distinct class regions are represented by colors, while the individual data points represent the model's predicted classifications. Effective performance is confirmed when the points reside within their respective color spaces. Furthermore, points positioned toward the center of a class region indicate a high level of confidence and a clear class separability achieved by the embedding space.

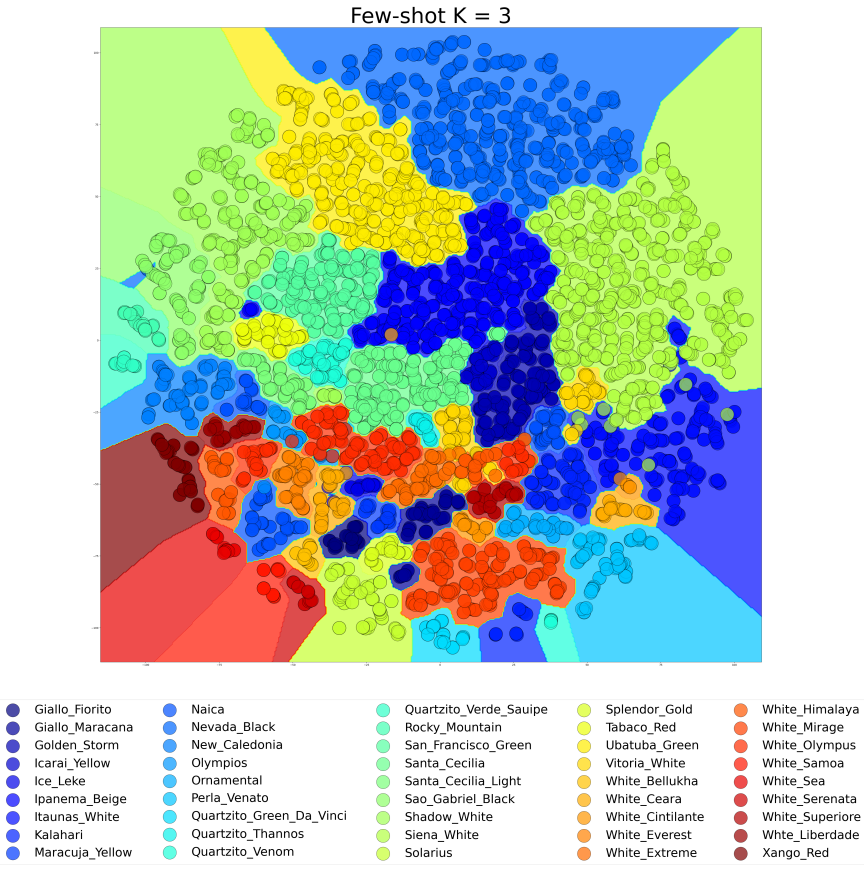

Figure 1. Visual representation of separability and feature clustering. Colors represent distinct class regions, and individual data points represent the predicted classifications ([Araujo 2022] dataset).

As a key advantage, our approach allows new rock classes to be added to the system without requiring model retraining. The system only requires a single reference image of the new rock, which makes the solution scalable and adaptable. The transfer learning technique, in turn, allows us to leverage the power of deep learning models pre-trained on large datasets to extract relevant visual features from the rocks. This accelerates the training process and improves the model's performance, even with a few samples [Chiodo Filho 2020]. In addition to the theoretical part and model development, this work includes the creation of a practical and accessible web application. The user can upload two or more images for comparison and receive a similarity percentage, which will indicate if the rocks have any resemblance.

The use of Siamese networks [Serrano and Bellogín 2023] for the identification of ornamental rocks is a key distinguishing feature of this research. In addition, integrating transfer learning optimizes the use of available data, while the implementation of a web tool makes the model usable in real-world scenarios. For example, we highlight the scenario wherein a person needs to find rocks that are so similar they can complement each other, as illustrated in Figure 2 with a pair of similar rock images.

Figure 2. Ornamental Rock Similarity for Complementarity

This work introduces an information system that uses artificial intelligence and Siamese neural networks to assist specialists in two main tasks: identifying the class of a given rock and finding its most similar pair in a dataset of rocks. The scientific contribution of this study lies in the transition from traditional Softmax-based classification to a metric-learning paradigm. By leveraging Siamese architectures, the proposed method overcomes the limitations of data scarcity and surpasses current state-of-the-art benchmarks in few-shot scenarios.

The system offers benefits to different stakeholders, e.g., enterprises can sell value-added products by ensuring a precise match between client demand and available materials. It also benefits the public sector, as it can attract tourism to geological sites and standardize the presentation and classification of materials, promoting fair trade. Additionally, it helps resellers by facilitating efficient inventory management through on-demand purchasing, which reduces waste and storage costs. The proposed system is designed to allow smoother and quicker task completion, reducing time and effort. In addition, users can more easily and effectively achieve their desired outcomes. The smoother task execution can lead to lower costs associated with task performance.

The remainder of this paper is organized as follows: Section 2 presents the related

work. Section 3 details the materials and methods, including data preparation, model architecture, and the training process. Section 4 discusses the results and prediction analysis. Finally, Section 5 concludes the study, followed by Section 6, Future Work, and Section 7, Acknowledgments.

2. Related Work

A systematic literature review was conducted in January 2026 via the Scopus database, restricted to journal and conference papers published in English. The search focused on three research axes: (1) neural networks for ornamental rock classification (2020–2026), (2) metric learning using Siamese Neural Networks (SNN) and K-Nearest Neighbors (KNN) (2020–2026), and (3) zero-shot learning in the mineral sector (2023–2026). From a total of 22 identified publications, the nine most relevant articles, three from each axis, were selected for analysis.

1. Query: (“ornamental rocks” OR “ornamental stones”) AND (“neural network”) in titles, abstracts, and keywords.
2. Query: (“image classification” AND “siamese network” AND “knn”) in titles, abstracts, and keywords.
3. Query: (“zero-shot” AND “mineral”) in titles, abstracts, and keywords.

2.1. Neural Networks for Ornamental Rock Classification

The work by [Ouzounis et al. 2021] addresses the challenge of automated classification of dolomitic marble tiles on the production line, seeking to overcome the limitations of manual quality control, which is inherently subjective, slow, and prone to human error. The literature indicates that while Deep Learning models offer high efficacy in classifying ornamental rocks, their “black-box” nature generates distrust and limits acceptance in industrial environments that demand transparency. The distinctive contribution of this research lies in the evaluation of 15 Convolutional Neural Network (CNN) architectures, correlating technical performance with model interpretability through the Grad-CAM technique. By providing heatmaps that visually justify which stone features (e.g., veins and color) influenced decisions, the study shows that models like DenseNet and InceptionResNet achieve high precision, offering an explainable solution for stone sector automation.

The study by [Sidiropoulos et al. 2022] addresses the automated screening of marble tiles by exploring the synergy between hand-crafted descriptors and deep features extracted via Transfer Learning. The authors identify that, while CNNs are powerful, they often fail to capture fine-grained textural details that traditional descriptors can represent more efficiently. To overcome this, the work proposes a feature aggregation framework that combines these two distinct types of information. By integrating learned and hand-crafted features, the research achieves superior classification accuracy compared to using either method in isolation, providing a more comprehensive solution for visual quality control in the ornamental stone industry.

A recent study [Dias et al. 2024a] that its an evolution of [Dias et al. 2024b] addressed the identification of Brazilian ornamental rocks using a unified dataset, where the Vision Transformer (ViT) achieved a high F1-score of 98.42%. However, a persistent limitation of this traditional approach is the necessity of a Softmax output layer, which binds the model to a fixed number of classes and requires full retraining whenever a new rock

type is introduced. Our work overcomes this by replacing the final classification layer with a Siamese Network architecture. This modification not only eliminates the need for retraining for new classes but also optimizes the embedding space to the point of achieving 100% accuracy on the same domain, surpassing the performance of the traditional ViT model while offering superior scalability.

2.2. Metric learning with Siamese Neural Networks and KNN

The article by [Pal et al. 2021] addressed the automated visual evaluation of cervical images for precancer detection, using an algorithm based on Faster R-CNN that proved effective in initial tests. However, a persistent limitation of this traditional approach is the requirement for manual annotations of the cervical boundaries and the use of complex data augmentation techniques to handle the scarcity of positive samples. This work introduces a framework based on Deep Metric Learning (DML) that operates on the complete cervical image, eliminating the need for boundary marking by specialists. This modification handles class imbalance and data scarcity while optimizing the representation space to improve specificity in disease detection without compromising sensitivity. The proposed NasNet model with Batch-hard loss outperforms previous versions, offering a scalable solution and eliminating intensive manual annotation.

According to [Rajpoot and K.R. 2023], the reliance of fundus image classification on CNNs is limited by the requirement for large datasets, which hinders the identification of rare retinal diseases. To address this, the authors introduce a few-shot meta-learning (FSML) framework combining a Triplet Neural Network with a KNN classifier. By optimizing the embedding space through Triplet loss, the model classifies new conditions using only 24 to 28 samples. The approach reported an AUC of 0.9858, exceeding various state-of-the-art models and demonstrating scalability for data-scarce medical contexts. [Bhende et al. 2025] addresses intra-class variation and data scarcity in acne classification using a hybrid Siamese Neural Network (ResNet-50) and KNN architecture. By employing PCA and Optuna for hyperparameter tuning, the model achieved 81.32% accuracy. The approach provides a framework for dermatological characterization that balances classification performance with computational efficiency.

2.3. Zero-shot Learning in the Mineral Context

A study of [Nesteruk et al. 2023] addressed mineral identification using polarized light microscopy, where a Multi-Layer Perceptron architecture achieved 94.0% F1-Score. However, a persistent limitation of this traditional approach is the necessity of a Softmax output layer, which binds the model to a fixed and reduced number of classes, requiring full retraining for any new species. To address this, recent benchmarks such as MineralImage5k have expanded the experimental scope, conducting studies that range from smaller subsets of 10 and 98 classes to a significantly more complex dataset of 360 mineral classes for few-shot classification. This evolution highlights the shift toward foundation models like CLIP, which replace fixed classification layers with a shared semantic-visual embedding space. This transition not only enables the recognition of thousands of mineral species in a Zero-Shot manner but also addresses the extreme intra-class variability of raw samples, surpassing the taxonomic constraints of traditional supervised learning.

The research by [Eppel et al. 2024] addresses the challenge of zero-shot material state segmentation by tackling the inherent limitations of both manual annotation and

purely synthetic datasets. The authors point out that manually labeling real-world images for material states is prohibitively expensive and often imprecise due to the gradual nature of these states, while traditional synthetic data fails to capture the diversity and complexity of the real world. The work introduces an unsupervised framework that infuses patterns automatically extracted from real-world images into synthetic scenes. By establishing a comprehensive benchmark across multiple domains, the study demonstrates that models trained on these infused synthetic datasets can generalize to unseen materials in a zero-shot manner, overcoming the scalability and precision bottlenecks of conventional supervised learning.

[Dong et al. 2025] addresses the multi-scale complexity and overlapping optical features of minerals in tight sandstone through a hybrid architecture for segmentation and classification. By integrating automated feature extraction with quantitative analysis, the method provides greater objectivity and speed than manual petrographic examination. This framework balances high-resolution geological characterization with computational efficiency.

Table 1. Comparative analysis of related works

Reference	Learning Paradigm	Decision Logic	Context	Dataset	Best Results
[Ouzounis et al. 2021]	CNN	Softmax	Ornamental Rocks	Marble Tiles	82.84% Acc
[Sidiropoulos et al. 2022]	CNN	Softmax	Lithology	Marble dolomitics	99.04% AUC
[Dias et al. 2024a]	CNN	Softmax	Ornamental Rocks	Marble Tiles	98.36% Acc
[Pal et al. 2021]	SNN	Softmax	General Rocks	Marble Tiles	98.36% Precision
[Rajpoot and K.R. 2023]	SNN	Metric Learning	General Rocks	Retinal disease	90% AUC
[Bhende et al. 2025]	SNN	KNN	Skin Disease	Acne classification	81.32% Acc
[Nesteruk et al. 2023]	SNN	Metric Learning	General Rocks	Raw Mineral	94% F1
[Eppel et al. 2024]	CNN	Zero-shot	Image Segmentation	Multi-sources	N/A
[Dong et al. 2025]	CNN	Softmax	Mineralogy	Sandstones	92.01% mIoU

3. Material and methods

This section shows the methodology for training, validating, and testing Siamese neural networks for few and zero-shot multi-class classification of ornamental rock images. In addition, the section addresses the pre-trained backbones for the network and the web-based user interface of the system.

3.1. Datasets

Our experiments have been conducted on two ornamental stone slab image datasets: one from [Dias et al. 2024a] (Figure 3) and the other from [Araujo 2022] (Figure 4). Both datasets include images with large and non-standardized dimensions, from a minimum of 921×667 pixels to a maximum of 1228×1632 pixels. The following table illustrates the image distribution among the training, validation, and test sets:

Table 2. Distribution of data into training, validation, and test sets.

Dataset	Train	Valid	Test	Classes
Ornamental Stone Slabs [Dias et al. 2024a]	1268	263	263	12
Ornamental Stone Slabs [Araujo 2022]	24308	5259	5198	45

Figure 3. Dataset 1: 12 Classes of Ornamental Stone Slabs [Dias et al. 2024a]

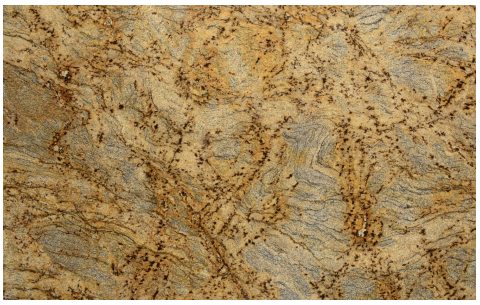

Figure 4. Dataset 2: 45 Classes of Ornamental Stone Slabs [Araujo 2022]

3.2. Implementation Details

The training experiments have been performed in a workstation with Intel i5 8600k processor and NVIDIA RTX 3060 GPU. The model has been entirely implemented using Python 3.10 within the PyTorch framework and compiled with GCC 11.2.0. All dependencies have been managed within a dedicated Anaconda virtual environment.

The datasets used in this study were selected because they consist of polished slab images, which directly align with our research objectives. Conversely, other repositories—such as the Rock Classification Dataset[1], Rock Images[2], and Gemstones Images[3]—were excluded. These sources feature lithologies that diverge from the ornamental stone focus, including raw minerals and uncut gemstones, as exemplified in Figures 5(a, b, and c).

(a) Rock Classification

(b) Rock Images

(c) Gemstones Images

Figure 5. Examples of excluded samples which do not meet the criteria for ornamental slabs.

The web-based user interface has been built using standard HTML and CSS, leveraging the utility-first classes provided by the Tailwind CSS framework for responsive and efficient design. Finally, communication between this front-end and the Flask API has been managed via standard HTTP requests, using Base64 encoding for image transfer.

3.3. Dataset Manipulation

To ensure consistency between development and production, the same preprocessing pipeline is applied to all images during both training and in system inference. First, pixel values are rescaled to the range [0.0, 1.0]. Subsequently, we apply a normalization step

[1]https://www.kaggle.com/datasets/salmaneunus/rock-classification
[2]https://www.kaggle.com/datasets/neelgajare/rocks-dataset
[3]https://www.kaggle.com/datasets/lsind18/gemstones-images

with a mean of [0.485, 0.456, 0.406] and a standard deviation of [0.229, 0.224, 0.225], following the ImageNet distribution to maintain compatibility with the pre-trained backbones. Finally, all images are resized to 224x224 pixels. This unified pipeline guarantees that the model receives data in a standardized format, regardless of whether it originates from the training set or a user-uploaded image in the deployed system.

3.4. Convolutional Neural Network

Convolutional Neural Networks play a fundamental role in our work, serving as the primary feature extractor for ornamental rock images. Leveraging their intrinsic ability to learn hierarchical representations directly from visual data, CNNs are utilized as the backbones in our Siamese network architecture. This approach is enhanced by the concept of transfer learning, where we use CNNs pre-trained on the ImageNet dataset. By freezing the convolutional layers of these backbones, we are able to leverage the vast pre-existing knowledge of visual patterns, allowing our model to focus on learning similarities and differences between rock types, even in few-shot scenarios. This optimizes the training process and improves the quality of the generated embeddings, which are essential for the classification task and for success in a few-shot learning environment.

We have conducted an exhaustive comparison among 10 neural network architectures for feature extraction in the task of rock classification using Siamese neural networks. The 10 models evaluated were the same as in [Dias et al. 2024a]. As in the original work, the ViT model has been the best-performing model in terms of accuracy in the task of class discrimination, and will be used in our system.

3.5. Siamese Neural Network

Figure 6 illustrates the architecture of the Siamese Neural Network, designed to process features extracted from the backbone. This network is composed of two sequential dense layers, designed to refine the embeddings and learn the similarity between images of the same class and the dissimilarity between images of different classes. To ensure the architecture's flexibility, the model was configured to dynamically adapt its classification head based on the output dimensions of the pre-trained backbone.

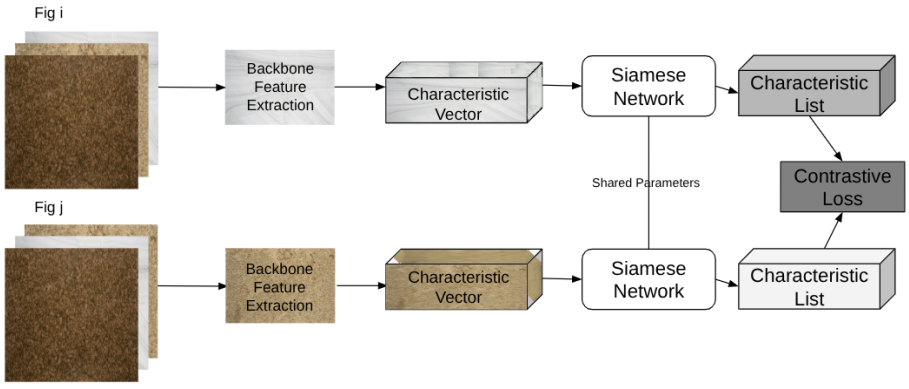

Figure 6. Siamese Neural Network Architecture for Similarity Learning

Using an automated method, the input dimension for the first fully connected layer is determined by passing a dummy input tensor through the feature extractor. This approach eliminates the need for manual adjustments, allowing for the seamless and efficient integration of a variety of backbones. The classification head then uses dense layers

to process the extracted features, with the goal of learning more discriminative representations for the classification task.

To mitigate overfitting and promote generalization, a Dropout layer with a 50% rate was inserted between each of the dense layers. This technique was used to reduce the dependency on certain neurons, deactivating them during the training process. This forces other neurons to discover a set of features to achieve a better result [Hinton et al. 2012, Srivastava et al. 2014, Vignesh Baalaji et al. 2023].

The output layer of the Siamese network consists of a feature list, which represents the compact and discriminative representation of the input image. To determine the similarity between pairs of images, cosine similarity is calculated between their respective embedding vectors. This metric is chosen for its effectiveness in measuring the angular orientation of the vectors, indicating how semantically similar the images are, regardless of the magnitude of their embeddings.

Figure 7 illustrates the image evaluation process within the test sets. Initially, two images are input into a feature extraction module, which generates a feature vector for each. These vectors are then compared using cosine similarity, a metric that returns a value between -1 and 1, indicating how aligned the vectors are. The closer the cosine similarity value is to 1, the higher the similarity between the images. After comparing the new image with all available labeled images in the test set, the model assigns the new image the class of the one that exhibited the highest cosine similarity.

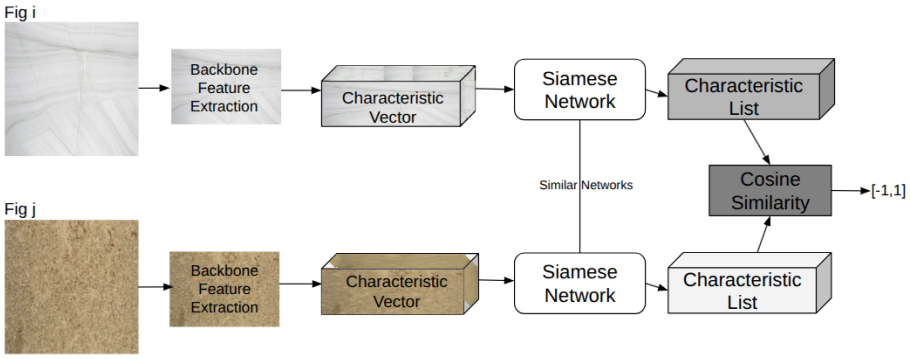

Figure 7. Image Classification Using Cosine Similarity and Feature Vectors

The Siamese network's training was conducted with a specific set of parameters to ensure stability and optimize performance. For the optimization process, the AdamW optimizer was utilized, with an initial learning rate of 0.01 and a weight_decay of 1e-2. Additionally, a ReduceLROnPlateau scheduler was implemented to dynamically adjust the learning rate, reducing it by 75% if the model's performance did not show improvement after three consecutive epochs on the validation set. Training was run for 40 epochs, with a patience of 10 for early stopping, which automatically interrupted the process when the loss on the validation set stopped improving, and 10 model versions were created to ensure the veracity of the results.

The head of the Siamese network was constructed with linear layers, normalization, and Dropout, projecting the output from the feature extraction backbone into the final low-dimensional embedding space. This structure, detailed in the Projection Head, begins with a Linear transformation to 1024 dimensions, followed by BatchNorm1d for

stabilization. The Tanh activation function was specifically chosen to ensure the embedding values were normalized between -1 and 1. This normalization prevents the creation of outlier values and, consequently, improves the stability of feature comparison. The Dropout was applied with a rate of 0.5 for regularization, and finally, a second Linear layer projects the vector to 256 dimensions, defining the feature space where similarity will be calculated.

To shape the embedding space and make it discriminative for subsequent kNN classification, the Contrastive Loss function was employed during training [Wang and Liu 2021]. This loss function operates with a margin parameter of 1.0, which establishes a distance threshold to penalize the network when the distance between the embedding vectors of dissimilar image pairs is less than this limit. In contrast, the model is trained to ensure high similarity for images of the same class (positive pairs), for which a cosine similarity threshold of 0.9 was defined. A value greater than or equal to 0.9 is interpreted as a positive match. This configuration forces the network to maximize the inter-class distance while minimizing the intra-class distance, resulting in highly cohesive and well-separated clusters for the final kNN classification stage.

During training, validation was performed by comparing images from the validation set with reference images from the training set, in an approach similar to kNN classification. The final model evaluation in the testing phase was conducted using two distinct approaches. The first approach, simulating a multi-neighbor proximity classification, utilized a kNN strategy with k=3. In this method, each test image was compared against all images in the validation set, and the final class was determined by the majority vote of the three nearest neighbors.

The second approach consisted of a global 1-NN classification, where the model's performance was measured by comparing each test image with all images in the training set. The predicted class was initially determined by the label of the closest neighbor. However, the final classification was additionally conditioned on a cosine similarity threshold of 0.85, the prediction of the nearest neighbor's class was only accepted if the similarity between the samples was equal to or greater than this threshold. This strategy allowed for the evaluation of the global discrimination capability of the embedding space molded by the Contrastive Loss.

Figure 8 illustrates how Contrastive Loss shapes the embedding space for kNN classification by enforcing separation between dissimilar classes and cohesion within the same category. The loss function maintains a minimum inter-class distance of 1.0 while grouping similar vectors to meet a 0.9 similarity threshold. This optimization maximizes inter-class distance and minimizes intra-class distance, establishing the spatial distinction required for proximity-based classification. In the provided diagram, a new query (represented by "?") is classified by identifying its nearest neighbors in the structured space. Because the query vector is positioned closest to three samples from Class A, the algorithm assigns it to that category. This process demonstrates how the embedding space enables the identification of new samples based on their proximity to established clusters.

3.6. The web-based System

The workflow of the proposed system, as detailed in Figure 9, involves three main roles: the Ornamental Rock Seller, the Specialist, and the Buyer. The core use case involves

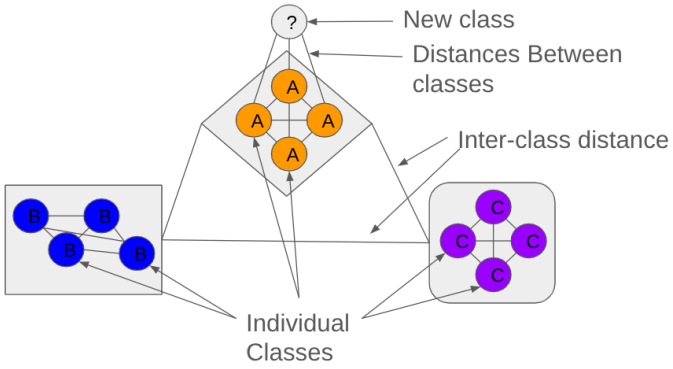

Figure 8. kNN for K = 3

a Buyer presenting an image of a rock. Although a Seller's experience ensures they can identify the rock's basic type, determining the best possible alternative or substitute for the client's specific needs is complex. The primary goal of this work is not just rock identification, but facilitating a similarity match. The system provides a solution when the client requires a rock highly similar to the one pictured, for instance, when substituting a broken piece or creating matching material for a crafted object. The process is completed when the Seller uploads the image.The system identifies and presents the most similar rock classes and individual images to the Buyer. This allows the Buyer to make an informed choice based on the closest visual match that specifically satisfies their requirement.

As illustrated in Figure 10, the system follows a Model-View-Controller (MVC) architectural pattern to ensure modularity and maintainability. The View layer provides a web interface where sellers upload images and view similarity results. The Controller layer, implemented as a RESTful API, handles image processing requests and orchestrates the similarity matching workflow. The Model layer contains the core image analysis and matching logic: it extracts visual features from uploaded images, computes similarity scores via cosine distance, and queries a pre-indexed database of ornamental rock images to retrieve the most similar matches. This three-tier separation enables efficient processing, scalable search, and a clean user experience tailored to the needs of sellers, specialists, and buyers in the ornamental stone trade.

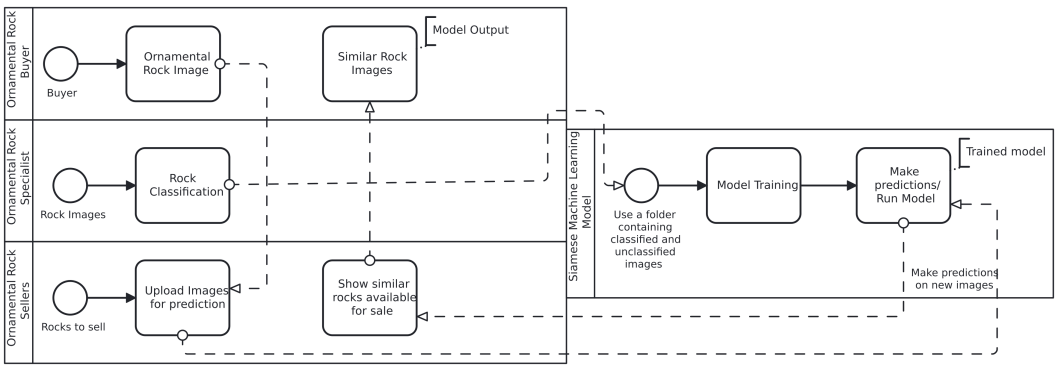

Figure 9. End-to-End Flowchart

The system interface presents a screen wherein the user starts by choosing a folder

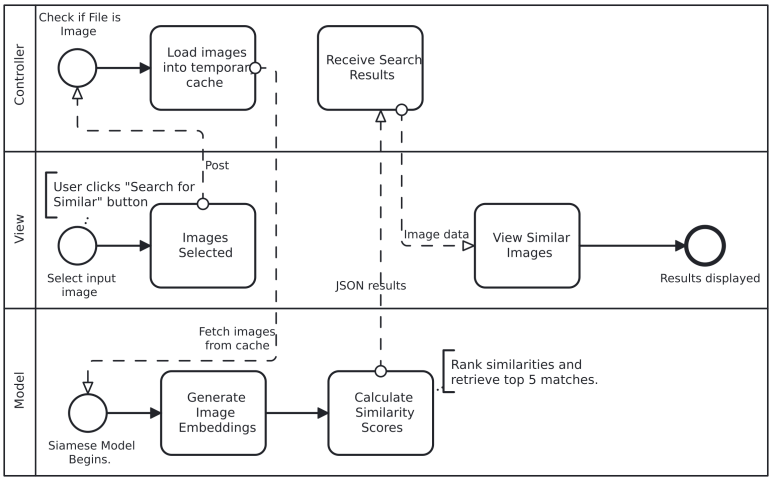

Figure 10. MVC System Flowchart

containing the images to search (Figures 11 and 12). Then, after successfully selecting the directory (Figure 13), the system starts the prediction steps. By clicking the "Search For Similar Images" button (Figure 14), the process starts. This folder must contain only the rock images to be recognized. Subsequently, the system performs a visual analysis of the results, as illustrated in Figure 15. This figure displays all image predictions and the respective reference images selected for the class, along with their calculated similarity scores.

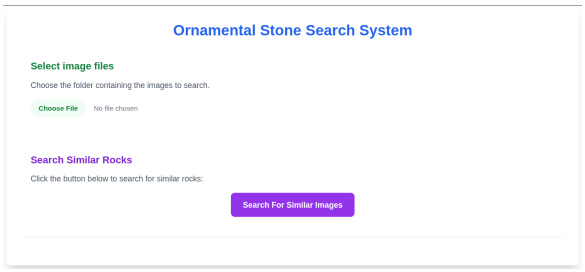

Figure 11. System Initial Screen.

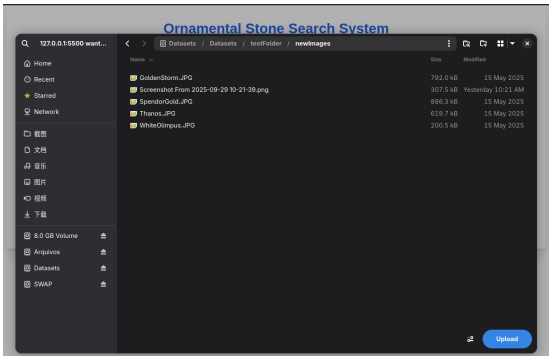

Figure 12. Select Folder Images.

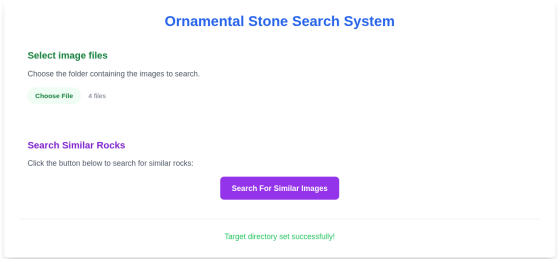

Figure 13. Folder successfully selected.

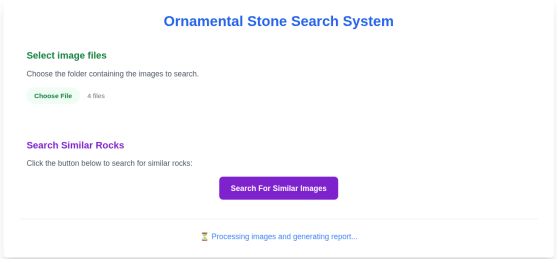

Figure 14. Start prediction processing.

The similarity between images was evaluated using cosine similarity, which ranges from -1 (maximum dissimilarity) to 1 (perfect similarity). For the end-user, how-

Figure 15. Color vs. Accuracy Mapping: Chromatic Variations in Inter-Class Similarity

ever, a negative value is not intuitively meaningful, as these values indicate strong dissimilarity or an opposing relationship between the feature vectors. Therefore, to simplify interpretation and focus on similarity, negative values were normalized to 0, indicating only that the images are not considered similar. The visual feedback is communicated via a color scale defined by specific similarity thresholds. Green (left image in Figure 16) indicates high similarity (above 80%). Yellow (right image in Figure 15) is used for medium similarity (between 60% and 80%). Finally, Red (also in the right image in Figure 16) represents low similarity (below 60%, including the normalized 0 value).

Finally, Figure 16 displays the most similar images, along with their respective classes and similarity scores. These samples allow us to confirm the model's prediction by visually inspecting the most analogous examples that led to the final class assignment.

4. Results

The baseline approach (Softmax Activation) relies on a conventional ViT classification method where the final layer outputs the probability distribution over a set of predefined classes. In contrast, our work utilizes a Siamese Network, which inherently returns a similarity measure between image pairs, rather than a probabilistic class output like the Softmax function. For a quantitative comparison, we implemented a Similarity-Based Classification mechanism: the query image is compared with all images in the reference database through a global nearest-neighbor search within the learned embedding space. Tables 3 and 4 present the performance comparison between the traditional classification results from the cited literature and our proposed work.

On the other hand, to make the Figure 17 and 1, predictions were made using the kNN algorithm applied to the embeddings generated by the Siamese Network, to generate the predictions shown in the figures. This difference in methodology justifies the resulting variance in accuracy. For instance, in Figure 17, the kNN model made 3 wrong predictions, resulting in a 98.85% classification accuracy, while the corresponding accuracy in Table 3 is 99.24%. This discrepancy merely serves to validate the model's robustness and consistency across different evaluation metrics.

The classification result is determined by similar images: the query image is as-

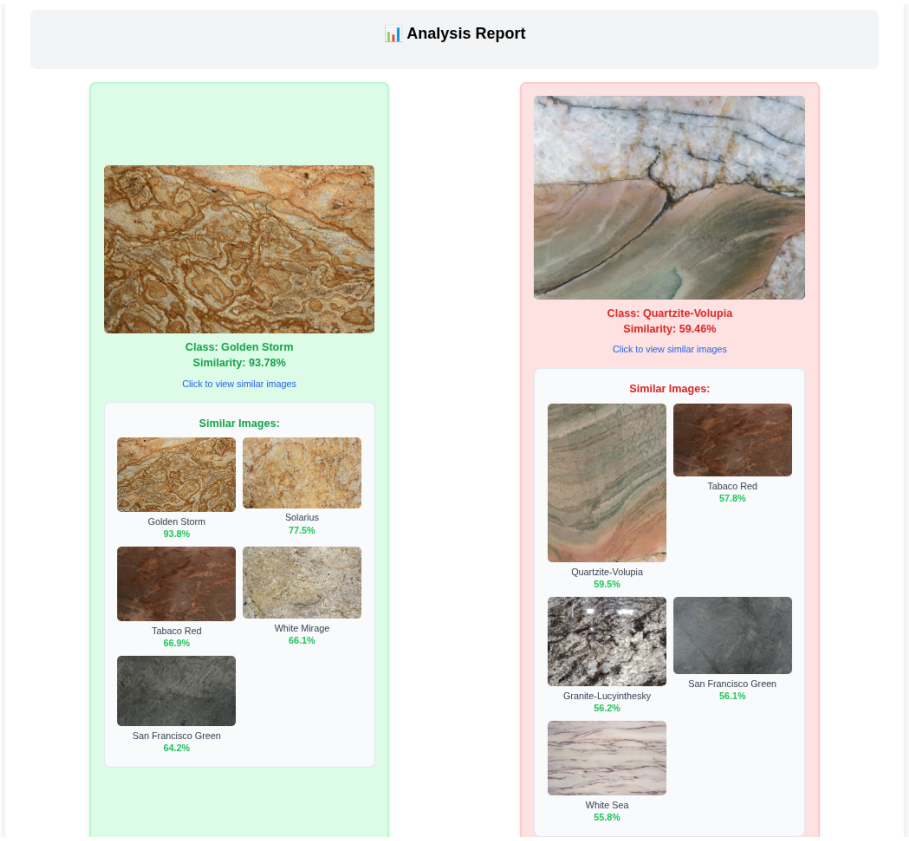

Figure 16. Color vs. Accuracy Mapping: Chromatic Variations in Inter-Class Similarity

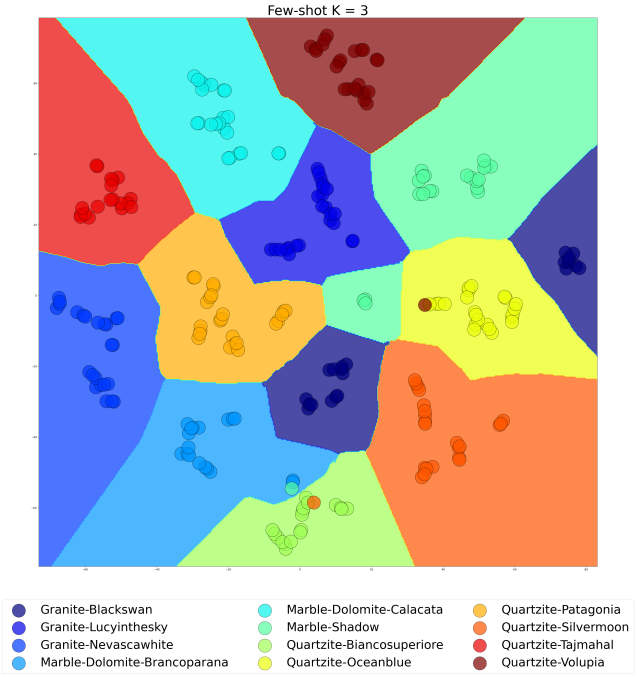

Figure 17. Classes Separability

Table 3. Similarity-Based Classification via Nearest Neighbor Search in the Embedding Space for 12 classes

Database	Acc(%)	Approach	Author
[Dias et al. 2024a]	98.36	Softmax Activation	[Dias et al. 2024a]
[Dias et al. 2024a]	99.23	Few-Shot @1	Present Work
[Dias et al. 2024a]	**99.62**	**Few-Shot @5**	**Present Work**
[Dias et al. 2024a]	80.60	Zero-Shot @1	Present Work
[Dias et al. 2024a]	95.81	Zero-Shot @5	Present Work

signed to the class of its closest neighbors in the database. The system's performance is then measured using Top-K (@K) Accuracy, where the classification is considered correct if the true class is found within the K most similar results. This methodology applies to both Few-Shot and Zero-Shot scenarios presented in the tables.

As detailed in Table 3, the performance comparison reveals that the proposed similarity-based classification framework consistently outperforms the Softmax activation baseline 98.36% Acc. The highest overall result, achieved by the Few-Shot @5 configuration at 99.62%, highlights the system's reliability in presenting the correct class within the top five most similar categories. Even under the strict Top-1 Accuracy metric, the Few-Shot @1 result 99.24% surpasses the baseline, confirming that the learned metric space is highly discriminatory. Furthermore, the results illustrate the challenge of classifying unseen data; the Zero-Shot @1 accuracy 80.60% is significantly lower, but improves drastically when the system is allowed to consider the top five predictions 95.81% Acc. for Zero-Shot @5. The marked improvement between the @1 and @5 metrics underscores the value of using this ranking ability in our application. This feature is leveraged to present the closest classes to the user, allowing the client to view and understand the most similar alternatives.

The experimental results, detailed in Table 4, show the superiority of the proposed similarity-based classification framework over the traditional Softmax baseline 92.00% Acc. even in the more challenging scenario involving 45 distinct classes. The framework proved highly effective: the Few-Shot @1 (Top-1) accuracy of 95.76% successfully surpassed the traditional baseline, confirming the highly discriminatory nature of the Siamese Network's metric. Furthermore, the Few-Shot @5 result of 97.98% demonstrates that the correct class is consistently present within the top five most similar categories. The system also maintains strong performance in the Zero-Shot regime, where entirely unseen classes are queried, achieving 88.09% accuracy at @5. This performance improvement when moving from @1 75.16% to @5 underscores the practical utility of the similarity-based ranking as a viable method for rock identification and substitution, particularly when managing large inventories.

The experimental results presented in Table 3 and Table 4 demonstrate that the proposed approach surpasses the current state-of-the-art for few-shot scenarios. This is achieved through a clear differential technique: replacing the traditional Softmax classification head with a Siamese network architecture. This shift transitions the model from direct class mapping to discriminative metric learning, allowing the tool to identify and classify images based on feature similarity distances rather than fixed labels. This indicates a significant methodological advancement in rock image classification, providing a

Table 4. Similarity-Based Classification via Nearest Neighbor Search in the Embedding Space for 45 classes

Database	Max Acc(%)	Approach	Author
[Araujo 2022]	92.00	Softmax Activation	[Dias et al. 2024a]
[Araujo 2022]	95.76	Few-Shot @1	Present Work
[Araujo 2022]	**97.98**	**Few-Shot @5**	**Present Work**
[Araujo 2022]	75.16	Zero-Shot @1	Present Work
[Araujo 2022]	88.09	Zero-Shot @5	Present Work

viable solution where data scarcity is a limiting factor.

The application of the t-SNE (t-distributed Stochastic Neighbor Embedding) in the Figure 17 technique to the dataset [Dias et al. 2024a], comprising 263 samples, evidenced high efficacy in preserving neighborhood structures, resulting in the formation of cohesive clusters for most categories, which validates the discriminatory power of the input features. Classes such as quartzite-tajmahal and granite-nevascawhite formed isolated and compact groupings, which indicates an inherent distinction and a high probability of successful classification. Categories that previously exhibited higher variance, such as granite-lucyinthesky, also showed a reduction in dispersion, confirming that the current projection successfully mapped the majority of materials into an ideal state of visual separation for a supervised modeling process.

However, the visual analysis of the projection reveals the presence of heterogeneous internal structures, which represent a methodological limitation. The granite-blackswan class exhibited pronounced bimodality, segmenting into two spatially distant sub-clusters in the plot, strongly suggesting that this single label encompasses two sample populations with intrinsically different physicochemical or spectral characteristics. Furthermore, the presence of isolated outliers was confirmed in the quartzite-volupia and quartzite-silvermoon classes. The occurrence of these discrepant points suggests a possible presence of noise within the dataset, underscoring the need for a thorough reevaluation of data quality and label homogeneity. These structural complexities must be addressed via pre-processing to maximize the accuracy and the final classifier.

4.1. Impact of the Solution on the Rock Sector

One of our results shows that it is possible to identify images of the same class among different categories without the need for retraining. This is particularly relevant for the scenario in which new rock types are introduced, aiding the creation of a robust repository. Within this repository, the user, regardless of their knowledge level, can find the rocks they are looking for based solely on an image.

This technological advancement not only enables faster and more precise rock identification but also allows for the suggestion of similar images and their respective origins. By analyzing the similarity with rocks from a specific region of interest, it becomes viable to promote greater recognition for that area, thereby integrating quarries into a unified platform.

Finally, there are many scenarios and situations wherein the customer needs to find the most similar rock to a given image. For example, if there is a room with very

expensive ornamental rocks and one of them breaks, it might be useful to be able to find another rock that does not differ from the others, as previously illustrated in Figure 2

4.2. Storage and Indexing Patterns for Ornamental Rock Images

The database must contain real and available rock samples from merchants, following the same standard, preferably polished, as demonstrated in Figures 3 and 4. If the image is captured outside of this standard, it will directly impact the results. For example, in Figure 18, we see that the photograph was taken in a non-standard way. However, they should ideally be horizontal, as the images are stored horizontally. We also must avoid objects that are not part of the image scenario, such as this object, similar to wood, in front of the rock. This type of foreign object has a direct impact on the prediction, since the system will look for an image that contains the same object, which does not make sense because it is not part of the rock itself.

Figure 18. Bad images samples.

4.3. Limitations

Due to the necessity for interested parties to add data, it is important that those responsible ensure the quality of the images regarding illumination, resolution, and size, while also striving to maintain quality in the environment's organization, aiming to remove objects that may be obstructing the rocks to guarantee the prediction quality. As shown in the results, the model can generalize to unseen images, but it is advisable that the model be retrained as many new classes are acquired to ensure its predictive quality.

During the experiments, it was observed that the complexity of the similarity model's inference is higher than that of the traditional classification methods used in the work, where comparisons were made. This characteristic resulted in a considerable increase in processing time for the classification step, given that the duration is directly proportional to the number of query images submitted by the user ($\mathbf{M}$) and the number of reference images stored in the dataset ($\mathbf{N}$). The computational complexity for the similarity-based classification step when applied to Few-Shot and Zero-Shot scenarios that require searching across the entire dataset is $\mathcal{O}(\mathbf{M} \times \mathbf{N})$. This means that the total time required for the search grows linearly with the number of query images ($\mathbf{M}$) and the number of images in the reference dataset ($\mathbf{N}$).

5. Conclusion

The developed classification model demonstrated a remarkable capacity for generalization, confirming the initial hypothesis that it could identify new classes without the need for exhaustive re-training. This adaptability provides the system with the necessary flexibility to handle the constant addition of new types of rocks in a dynamic commercial environment. The model performed better in majority cases, and in few-shot classification scenarios, the present work surpassed the benchmark on the [Dias et al. 2024a] database, achieving an accuracy of 99.23% with the kNN (k=3) approach, compared to 98.36% for the referenced work.

The key validation lies in the Zero-Shot classification capability, where the model demonstrated functionality in identifying entirely novel classes on the *Ornamental Stone Slabs* [Araujo 2022] database. The Top-1 accuracy of 75.15% achieved by our Zero-Shot approach for 45 classes validates the knowledge transfer capability of our Siamese network to new rock classes. This result is particularly noteworthy because the network was trained only on an initial dataset of 12 classes, characterized by its limited number of images. Furthermore, the result was achieved without any parameter adjustment or prior exposure to the specific target dataset, although this value is lower than the 92.00% Top-1 accuracy of the ViT network benchmark.

However, when analyzing the performance across the top five classifications, which is our goal when presenting the user with the 5 closest images, we achieved 88.00% in Top-5 accuracy. This means that, without specific training, the model is capable of presenting the correct class among the five main suggestions. This Top-5 performance closely approaches the benchmark's result, falling only 4 percentage points below the reference model. This performance reinforces the practical viability of the system for identifying unknown rocks with high reliability for the end user. This performance validates the effectiveness and precision of the implemented approach, confirming that the results remain close to the state of the art when considering the combined metrics of few-shot performance and Zero-Shot generalization capability.

The implementation of this system adds substantial value to commercialized products, offering better customization of the customer offering. Furthermore, the automated nature of the system modernizes the rock searching and identification procedure. By analyzing the entire stock without the need for physical movement of samples, the system reduces the exclusive dependence on the vendor's knowledge and the time required for the process. This practicality and efficiency significantly accelerate commercialization, transforming a historically laborious process into an agile and modern operation.

Future development involves integrating the classification and search system with sales and inventory platforms to align technical analysis with commercial workflows. This expansion includes developing search modules based on visual references, such as user drawings or structural patterns, to identify inventory items beyond standard classification. Additionally, a mobile application will enable real-time rock identification within storage environments, enhancing portability during the inventory and sales process. Technical improvements aim to reduce system inference time and increase scalability by optimizing similarity searches through a One-Shot strategy. This approach involves limiting comparisons to pre-selected class prototypes rather than the entire dataset, effectively reducing algorithmic complexity. Further research will explore advanced Siamese architectures,

such as Twin Siamese Networks, to refine embedding quality and classification accuracy. Finally, applying Explainable AI (XAI) methods will identify the visual features prioritized by the network, providing domain experts with interpretable insights.

Acknowledgments

The authors thank CNPq for grant No. 407742/2022-0. Professor Komati thanks CNPq for the DT-2 grant (302726/2023-3) and FAPES for project 1023/2022 (P:2022-8TZV6). In compliance with SBC's Code of Conduct, the authors acknowledge the use of ChatGPT and Gemini for grammatical revision and text polishing. The authors remain fully responsible for the final content and integrity of this work.

References

Apex (2024). Brazil ends 2024 with an increase in natural stone exports and strengthens its global leadership. https://apexbrasil.com.br/content/apexbrasil/br/pt/conteudo/noticias/Brasil-encerra-2024-com-alta-nas-exportacoes-de-rochas-naturais-e-reforca-lideranca-global.html. Accessed on September 13, 2025.

Araujo, J. V. C. (2022). Rede neural convolucional para classificação de chapas polidas de rochas ornamentais. Monografia de graduação (Bacharelado em Sistemas de Informação), Instituto Federal do Espírito Santo, Cachoeiro de Itapemirim.

Bhende, N., Sheth, S., and Reddy, M. (2025). Siamese network embeddings and knn classifier for robust acne image classification: A hybrid approach. In *2025 International Conference on Information, Implementation, and Innovation in Technology (I2ITCON)*, pages 1–7.

CETEM (2020). Request characterization of ornamental stones. https://www.gov.br/pt-br/servicos/caracterizar-rochas-ornamentais. Accessed on September 13, 2025.

Chiodo Filho, C. (2020). Identificação de minerais por meio de Redes Neurais Convolucionais: um estudo comparativo entre Inteligência Artificial e o Sistema Visual Humano. *Anais do Brazilian e-Science Workshop (BreSci)*.

Dias, D., Komati, K., and Gazolli, K. (2024a). Automating rock classification: A vision transformer approach in Brazil's ornamental stone. *Ibero-Latin American Congress on Computational Methods in Engineering (CILAMCE)*.

Dias, D. F., Komati, K. S., and de Souza Gazolli, K. A. (2024b). Comparative evaluation of image classification models for ornamental rock classification. In *2024 L Latin American Computer Conference (CLEI)*, pages 1–4.

Dong, L., Sun, C., Yu, X., Zhang, X., Chen, M., and Xu, M. (2025). Hybrid architecture for tight sandstone: Automated mineral identification and quantitative petrology. *Minerals*, 15(9).

Eppel, S., Li, J. Y., Drehwald, M. S., and Aspuru-Guzik, A. (2024). Infusing synthetic data with real-world patterns for zero-shot material state segmentation. In *The Thirty-eight Conference on Neural Information Processing Systems Datasets and Benchmarks Track*.

FINDES (2024). Espírito santo: a world reference in the ornamental stone sector. https://findes.com.br/es-referencia-mundial-no-setor-de-rochas-ornamentais/. Accessed on September 13, 2025.

Hinton, G. E. and Roweis, S. (2002). Stochastic neighbor embedding. *Advances in neural information processing systems*, 15.

Hinton, G. E., Srivastava, N., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R. R. (2012). Improving neural networks by preventing co-adaptation of feature detectors.

Nesteruk, S., Agafonova, J., Pavlov, I., Gerasimov, M., Latyshev, N., Dimitrov, D., Kuznetsov, A., Kadurin, A., and Plechov, P. (2023). Mineralimage5k: A benchmark for zero-shot raw mineral visual recognition and description. *Computers & Geosciences*, 178:105414.

Ouzounis, A., Sidiropoulos, G., Papakostas, G., Sarafis, I., Stamkos, A., and Solakis, G. (2021). Interpretable deep learning for marble tiles sorting. In *DeLTA*, pages 101–108.

Pal, A., Xue, Z., Befano, B., Rodriguez, A. C., Long, L. R., Schiffman, M., and Antani, S. (2021). Deep metric learning for cervical image classification. *IEEE Access*, 9:53266–53275.

Rajpoot, A. and K.R., S. (2023). Enhancing rare retinal disease classification: a few-shot meta-learning framework utilizing fundus images. *Multimedia Tools and Applications*, 83:1–19.

Serrano, N. and Bellogín, A. (2023). Siamese neural networks in recommendation. *Neural Computing and Applications*, 35(19):13941–13953.

Sidiropoulos, G. K., Ouzounis, A. G., Papakostas, G. A., Lampoglou, A., Sarafis, I. T., Stamkos, A., and Solakis, G. (2022). Hand-crafted and learned feature aggregation for visual marble tiles screening. *Journal of Imaging*, 8(7).

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15(56):1929–1958.

Vignesh Baalaji, S., Sandhya, S., Sajidha, S. A., Nisha, V. M., Vimalapriya, M. D., and Tyagi, A. K. (2023). Autonomous face mask detection using single shot multibox detector, and ResNet-50 with identity retrieval through face matching using deep siamese neural network. *Journal of Ambient Intelligence and Humanized Computing*, 14(8):11195–11205.

Wang, F. and Liu, H. (2021). Understanding the behaviour of contrastive loss. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2495–2504.

Zheng, D., Zhong, H., Camps-Valls, G., Cao, Z., Ma, X., Mills, B., Hu, X., Hou, M., and Ma, C. (2024). Explainable deep learning for automatic rock classification. *Computers & Geosciences*, 184:105511.