Mandal, Dipendra J.; Deborah, Hilda; Tobing, Tabita L.; Janiszewski, Mateusz; Tanaka, James W.; Lawrance, Anna

Comprehensive Evaluation of ImageNet-Trained CNNs for Texture-Based Rock Classification

Published in:
IEEE Access

DOI:
10.1109/ACCESS.2024.3424931

Published: 01/01/2024

Document Version
Publisher's PDF, also known as Version of record

IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 17 June 2024, accepted 5 July 2024, date of publication 8 July 2024, date of current version 18 July 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3424931

RESEARCH ARTICLE

Comprehensive Evaluation of ImageNet-Trained CNNs for Texture-Based Rock Classification

DIPENDRA J. MANDAL[1], HILDA DEBORAH[1], TABITA L. TOBING[2], MATEUSZ JANISZEWSKI[3], JAMES W. TANAKA[4], AND ANNA LAWRANCE[4]

[1]Department of Computer Science, Norwegian University of Science and Technology, 2802 Gjøvik, Norway
[2]Department of Information Security and Communication Technology, Norwegian University of Science and Technology, 2802 Gjøvik, Norway
[3]Department of Civil Engineering, School of Engineering, Aalto University, 02150 Espoo, Finland
[4]Department of Psychology, University of Victoria, Victoria, BC V8P 5C2, Canada

Corresponding author: Dipendra J. Mandal (dipendra.mandal@ntnu.no)

ABSTRACT Texture perception plays a vital role in various fields, from computer vision to geology, influencing object recognition, image segmentation, and rock classification. Despite advances in convolutional neural networks (CNNs), their effectiveness in texture-based classification tasks, particularly in rock classification, still needs exploration. This paper addresses this gap by evaluating different CNN architectures using diverse publicly available texture datasets and custom datasets tailored for rock classification. We investigated the performance of 38 distinct models pre-trained on the ImageNet dataset, employing both transfer learning and fine-tuning techniques. The study highlights the efficacy of transfer learning in texture classification tasks and offers valuable perspectives on the performance of different networks on different datasets. We observe that while CNNs trained on datasets like ImageNet prioritize texture-based features, they face challenges in nuanced texture-to-texture classification tasks. Our findings underscore the need for further research to enhance CNNs' capabilities in texture analysis, particularly in the context of rock classification. Through this exploration, we contribute insights into the suitability of CNN architectures for rock texture classification, fostering advancements in both computer vision and geology.

INDEX TERMS Image texture, convolutional neural network, transfer learning, rocks, image classification.

I. INTRODUCTION

For us humans, texture is an important visual cue that is exploited efficiently and almost effortlessly in our day-to-day lives, providing crucial details for organizing cohesive sections and identifying material characteristics. The perception and understanding of visual texture are also essential in many fields of expertise, albeit with varying terms being used, e.g., surface roughness in the Earth sciences [1] and spatial heterogeneity in medical imaging [2]. In geology, visual texture plays a crucial role in rock classification, helping geologists differentiate between rock types based on their unique surface patterns and features [3], [4]. In computer vision, texture has been exploited to aid object recognition and image segmentation [5], [6]. However, contrary to human vision, the classification of real natural textures is still considered a difficult problem. Natural texture is a complex class of texture, especially when compared to man-made objects.

Rock texture is a type of natural texture. Classifying rocks based on their texture can be crucial to deciphering the history of the Earth, geological processes, and environmental conditions. It provides valuable information on past environments and events, helps reconstruct geological events, assesses environmental conditions, and informs decision-making in engineering and resource exploration. Traditionally, geologists would perform physical and visual observation and laboratory testing to determine the type or category of rocks. Physical and visual characteristics being analyzed are, e.g., texture, structure, mineral composition, color, hardness, and density [7], [8], [9], [10]. Such analysis processes can be expensive and time-consuming.

Advances in computer vision have led to the development of algorithms that automate rock classification based on

The associate editor coordinating the review of this manuscript and approving it for publication was Wei He.

visual features extracted from images. Rock classification using deep learning has garnered considerable attention among researchers [11]. However, despite the availability of various CNN architectures [12], [13], [14], their effectiveness in classifying rock textures still needs to be fully explored. Many studies tend to focus on a limited range of CNN models for accuracy assessment, often overlooking the compatibility of datasets being used with the human visual perception of texture. As such, not only do results vary, but integrating computer vision into the practical day-to-day work in the field will also become challenging.

In this study, we aim to bridge this gap by systematically examining texture datasets and their potential applications across different CNN architectures, focusing on the specific case of rock classification. Taking inspiration from human perception, we used well-known texture datasets, from more generic to specific datasets tailored for rock classification. We used transfer learning and fine-tuning techniques on CNN models pre-trained with ImageNet to perform rock classification. By evaluating the performance of different models in texture analysis, we seek to summarize the effectiveness of these models and understand their suitability for texture analysis in the context of rock classification. A worthy mention is a previous study suggesting that CNNs trained on the ImageNet dataset often exhibit a bias towards texture-based features in image classification tasks [15]. However, we found that although these networks may prioritize texture over other image attributes, e.g., shape, they struggle when faced with more nuanced texture-to-texture classification tasks, highlighting the need for further investigation into their underlying mechanisms. We also used Grad-CAM [16] to observe the features that the network pays attention to when making classification decisions.

II. RELATED WORKS

Texture analysis within the realm of computer vision applications can be categorized into four primary domains [17], i.e., texture synthesis [18], classification [19], segmentation [20], and shape from texture [21]. This section comprehensively reviews texture analysis techniques applied to rock classification, highlighting the evolution from traditional methods to advanced machine learning and deep learning approaches.

A. TEXTURE ANALYSIS FOR ROCK CLASSIFICATION

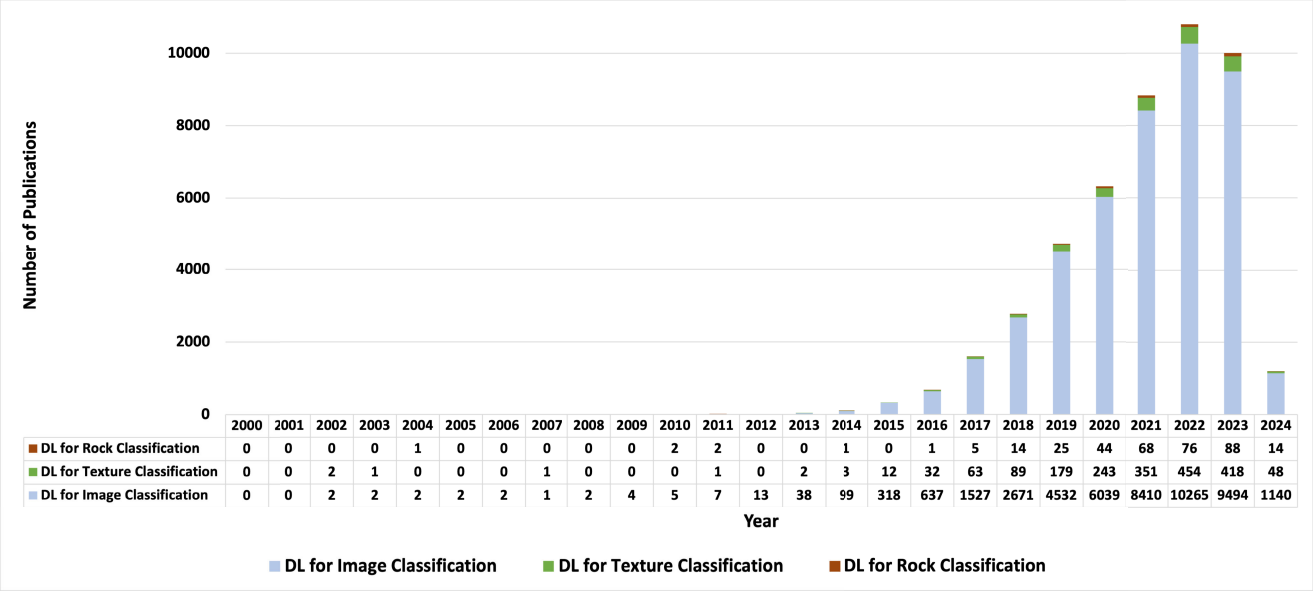
Visual analysis is one of the primary methods geologists use to assess rock characteristics. Tamura et al. [22] introduced six human visual perceptual texture features derived from psychological experiments, i.e., coarseness, contrast, directionality, line-likeness, regularity, and roughness. Harinie et al. [23] applied these features to classify three types of rock, i.e., igneous, sedimentary, and metamorphic, and compared their performance with commonly used methods such as the color co-occurrence matrices, gray-level co-occurrence matrices (GLCM) [24], and moments. The results showed that the methods based on the Tamura features achieved superior classification accuracy of >87%. It underscores the effectiveness of leveraging psychological insights into texture perception to improve rock classification; however, this study did not undertake subclass classification of the rocks. Lobos et al. [25] proposed a multi-scale transform method-based classification approach for analyzing and classifying natural rock textures. Multi-scale transform domain representations capture information at different levels of scale and orientation, allowing for a more comprehensive analysis of the rock textures. Their study incorporates both stationary patterns and structural information, leading to improved performance compared to conventional techniques.

In a study by Lepisto et al. [26], Gabor filtering-based feature extraction was combined with a k-nearest neighbor classifier for the classification of natural rock texture images. Here, textural features were combined with color information by applying Gaussian bandpass filtering to different color channels of the images, resulting in significantly improved classification accuracy. A similar observation was made by Bianconi et al. [27], where they compared the performance of different visual features and five classifiers over a set of 12 granite classes. Their findings suggest that classification based on both color and texture is highly effective and outperforms previous methods based solely on textural features.

Gonçalves and Leta [28] used Hierarchical Neuro-Fuzzy Class for the macroscopic classification of rock texture of four rock classes, i.e., gneiss, basalt, diabase, and rhyolite. Each of those classes has 2-5 subclasses, and a classification accuracy of 73% was achieved. The study used various texture descriptors, e.g., spatial variation coefficient [29], Hurst coefficient [30], entropy [31], and co-occurrence matrix [32]. As there is no fixed structure or a constant number of adjustable parameters for the neuro-fuzzy class, the learning algorithm requires more complex programming—presenting challenges and limitations for its usage. With the increasing complexity of rock data and texture analysis, many researchers are turning to convolutional neural networks (CNNs) as a solution, due to their superior performance in various applications, particularly image classification tasks.

B. CNN-BASED TEXTURE ANALYSIS FOR ROCK CLASSIFICATION

Deep learning (DL) based methods have existed for some time, but they did not receive much attention until 2012. Since then, these methods have been increasingly applied to a variety of problems related to computer vision, including texture analysis. FIGURE 1 illustrates the trends in the number of publications showing applications of deep learning over the last 20 years within the domains of image, texture, and rock classifications. Despite a significantly lower number of publications for rock classification using DL, the figure demonstrates a growing trend.



	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024
DL for Rock Classification	0	0	0	0	1	0	0	0	0	0	2	2	0	0	1	0	1	5	14	25	44	68	76	88	14
DL for Texture Classification	0	0	2	1	0	0	0	1	0	0	0	1	0	2	3	12	32	63	89	179	243	351	454	418	48
DL for Image Classification	0	0	2	2	2	2	2	1	2	4	5	7	13	38	99	318	637	1527	2671	4532	6039	8410	10265	9494	1140

FIGURE 1. Evolution of Deep Learning (DL) Applications over last 20 years (2004-2024), categorized by the trend in a specific application, i.e. for image classification, texture classification, and rock classification.

Zhang et al. [33] used a CNN based on the Inception-v3 architecture to classify rocks into three distinct classes. Using a transfer learning approach to train a pre-trained network, an accuracy of 90% was achieved. However, it was a limited assessment since the testing only involved three images from each set, and the accuracy varied with the increasing number of test images. Houshmand et al. [34] employed supervised machine learning (ML) algorithms to classify five rock types based on various measured (non-image based) parameters, e.g., compression wave velocities (P wave) and shear wave (S wave), Leeb hardness, and geochemical properties. In addition, image dataset of approximately 10,000 sliced images was also used with a CNN based on the ResNet-50 architecture the classification. Although the CNN-based classification accuracy outperformed the non-image based traditional ML, it varied between 89% and 94% for different classes due to the dataset imbalance. Notably, two classes of rocks with identical textural features were often misclassified with each other. A hybrid method combining CNN image analysis with non-image features was employed to address this, resulting in 99% classification accuracy.

Ran et al. [10] proposed a custom CNN architecture (RTCNNs) with fewer layers than pre-trained CNN architectures, consisting of two convolutional layers followed by a pooling layer and fully connected layers. They used a custom dataset comprising 24,315 images cropped from 2,290 high-resolution field rock photographs to classify six classes, i.e., mylonite, granite, conglomerate, sandstone, shale, and limestone. Using transfer learning on the same datasets, they also compared RTCNNs with pre-trained models VGG-16, AlexNet, and GoogLeNet Inception-V3. RTCNNs achieved a classification accuracy of 97.96%, outperforming VGG-16 (94.2%), AlexNet (92.78%), and GoogLeNet Inception-V3 (97.1%). The accuracy of sandstone and limestone classification was comparatively lower than that of the other rock classes, likely due to their similar visual features and possibly the uncontrolled environmental conditions inherent in field images. Li et al. [35] used transfer learning with pre-trained VGG-16, ResNet-50, and Inception-V3 architectures to classify 620 RGB rock images sourced and augmented from NASA's Mars Science Laboratory (MSL) datasets [36] into four distinct groups. Group 1 included dark-toned aphanitic effusive rocks, e.g., basalt, picro-basalt, trachy-basalt, and tephrite. Group 2 is of fine-grained intrusive rocks characterized by rounded edges and flat or curved facets. Group 3 is comprised of layered sedimentary rocks, while Group 4 includes conglomerates. The highest accuracy of 100% was achieved using the VGG-16, while Inception-v3 and ResNet-50 achieved 92.19% and 98.54%, respectively. Achieving the highest accuracy could be attributed to the notable visual distinctions among these four rock classes arising from variations in color, texture, and grain size.

Zheng et al. [37] Initially used three CNN architectures, i.e., EfficientNet-B2, MobileNet-V3, and ResNet-50, to classify six types of sedimentary rocks (quartz arenite, feldspathic arenite, lithic arenite, siltstone, oolitic packstone, and dolomite) and achieved an accuracy of 94%, 97%, and 97.7%, respectively. Despite this high accuracy, when the results were visualized using GradCAM, it was found that the model was basing its classification on features irrelevant for distinguishing sedimentary rocks, e.g., cracks, cements, and scale bars. An attention-based dual neural network was later used, which considered appropriate features, and achieved an accuracy of 99%. The study underscores the importance of

TABLE 1. Summary of deep learning architectures used for rock classification.

	DL Architectures	Datasets	Accuracy	Remarks
1	ResNet-50 [34]	10,000 sliced images	Varied (89% to 94%)	5 classes
2	Vgg16, ResNet-50, Inception-V3	620 Martin Rocks from MSL Notebook [36]	100%, 91.2%, 98.5%	4 classes, data augmentation, transfer learning
3	EfficientNet-B2, MobileNet-V3, ResNet-50, and SEDNet [37]	Sedimentary rock types, custom datasets	94%, 97%, 97.7%, 99%	6 classes, data augmentation, GradCAM
4	RTCNNs, AlexNet, VGG-16, GoogLeNet Inception-V3 [10]	Custom datasets with 2,290 field rock photographs	97.96%, 92.78%, 94.2%, 97.1%	6 classes, Flipping of images
5	Inception-V3 [33]	Custom datasets (571 images) compiled from diverse sources, including photographs, rock databases, and internet searches	90%	3 classes
6	Custom-designed CNN architecture [38]	Custom datasets with 4800 images	98.5%	3 classes (3 level of granularity in sandstone)
7	ResNet-50 [39]	Custom datasets (1000+ images) compiled from diverse open sources, including university archives and special collections	84%	10 classes (3 levels of granularity in sandstone)
8	Inception-ResNet-V2 [40]	Custom dataset: 42,000 images randomly cropped from 3,000 raw images of various tunnel faces	Avg. 95.96%	5 classes (rock structures i.e. mosaic, granular, layered, block and fragmentation)
9	Inception-v3, DenseNet-121, DenseNet-169, DenseNet-201, ResNet-50, ResNet-101, ResNet-152, VGG-16 and VGG-19 [41]	Custom datasets with 7000-104000 images	Avg. from 86 to 92%	7 classes of carbonate core images
10	MobileNet-V2 and Inception-v3 [42]	Custom dataset: 1521 smartphone-acquired images	98% and 97%	6 classes (e.g. Basalt, Garnet Schist and Granite)
11	ResNet-34 [43]	Custom dataset: 315 rock images (4,096 × 3,000 pixels), 382536 images (after augmentation)	99.1%	7 classes (black coal, gray black mudstone, gray argillaceous siltstone, gray fine sandstone, light gray fine sandstone, dark gray silty mudstone and dark gray mudstone)

incorporating feature visualization techniques such as Grad-CAM when using DL models. Such enables the assessment of whether a model is effectively capturing relevant features (aligning its focus with geological expertise), thus providing insights to enhance model performance.

Even with transfer learning, the dataset size can significantly impact the classification accuracy of CNN models. Larger datasets typically enable models to learn more diverse features, leading to better generalization and higher accuracy. For example, Chen et al. [43] had a limited number of 315 training images, resulting in lower classification accuracy. They then employed data augmentation, resulting in higher accuracy. Likewise, Dawson et al. [41] evaluated nine CNN architectures that were pre-trained on ImageNet weights and then fine-tuned using transfer learning on datasets of varying sizes (104,000, 42,000, and 7,000 images), emphasizing the impact of dataset size on model performance.

The study focused on textural information and defined seven classes: mudstone, wackestone, packstone, grainstone, floatstone, rudstone, and boundstone. Among the architectures tested (Inception-v3, DenseNet-121, DenseNet-169, DenseNet-201, ResNet-50, ResNet-101, ResNet-152, VGG-16, and VGG-19), Inception-v3 achieved the highest classification accuracy at 92%, while VGG-16 exhibited the lowest at 86%. The evaluation of the models on varying dataset sizes generally showed improved accuracy with larger datasets. Rudstone and floatstone were often misclassified with each other, likely due to their similar visual features. Both rudstone and floatstone are part of the Dunham classification system for carbonate rocks [44], commonly used to describe limestones with a coarse texture and large grain size. TABLE 1 summarizes various architectures utilized for rock classification, including information on the dataset, results obtained, number of classes, and types of rocks.

As discussed so far, various pre-trained CNN models have been used effectively for rock classification with good accuracy. However, none of these papers underscores the importance of their choice of a particular CNN model for this

task. This could be because deep learning models are often seen as black boxes, and usually, the focus is on achieving high performance for the task at hand with available resources and computational time. Appendix B summarizes the key attributes of these models, providing a general understanding of their complexity, potential advantages, and limitations.

C. TEXTURE DATASETS

The availability of well-labeled texture datasets with a sufficient number of images covering a diverse range of texture classes remains limited. This scarcity is particularly pronounced in the domain of rock texture datasets, where expert classification of texture properties by geologists is essential. Existing datasets often lack the depth and breadth required for a comprehensive analysis, hindering research efforts to understand geological textures and to sufficiently train deep learning models. There is a clear need for more comprehensive and specialized datasets in this domain. The following are some available texture datasets, including those existing for rock textures.

The GMSRI dataset [45] consists of approximately 30,000 images of Martian rocks, each sized at 560 × 500 pixels. The dataset is divided into five texture-based categories, i.e., igneous rocks, sedimentary rocks, cracked rocks, gravels, and sands; and is further subdivided into 28 distinct categories based on texture and shape. It also includes both real Martian images from the Mars32k datasets [46] and synthetic counterparts generated using generative adversarial networks (GAN). The Kylberg Texture Dataset version 1.0 [47] consists of 28 distinct texture categories, each containing 160 individual texture samples, totaling 4480 images. These images are standardized to a dimension of 576 × 576 pixels and are represented in 8-bit grayscale PNG format. The textures encompass a variety of surfaces, including fabrics and stone, captured within their local environments.

The texture dataset from the University of Illinois Urbana-Champaign (UIUC) [48] was comprised of 1000 uncalibrated images featuring 25 distinct textures, each with 40 samples. These images depict surfaces exhibiting texture variations primarily attributable to albedo differences (e.g., wood and marble), three-dimensional shapes (e.g., gravel and fur), and a blend of both characteristics (e.g., carpet and brick). While the original dataset link is no longer operational, an alternative dataset was introduced, i.e., Meta-Album Textures dataset [49]. This dataset serves as a preprocessed iteration of the original, consisting of four different datasets: KTH-TIPS [50], Kylberg, and UIUC. It includes a total of 8675 images, categorizing them into 64 classes, with each image standardized to a dimension of 128 × 128 pixels.

The publicly accessible dataset known as CoMMonS (Challenging Microscopic Material Surface Dataset) [51] consists of 6912 microscopic images capturing 24 fabric samples with fine details, each at a resolution of 2560 × 1920 pixels. This dataset mainly focuses on three key properties of fibers: length, smoothness, and toweling

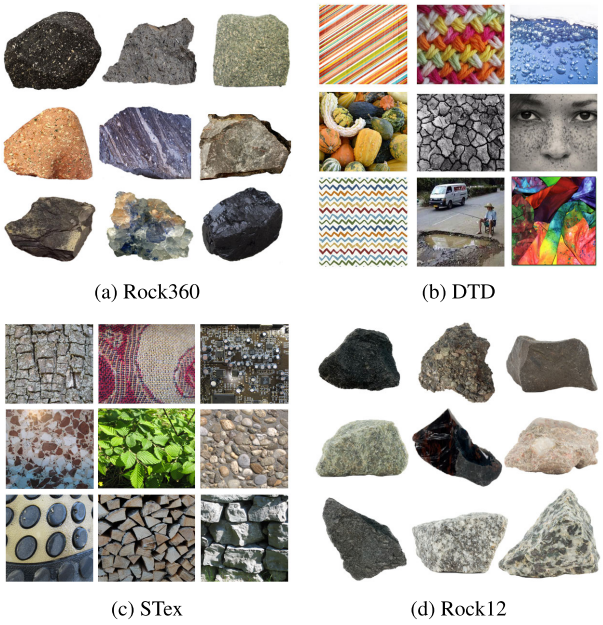


FIGURE 2. Example of images from the four texture datasets used in this study.

effect, facilitating fine-grained texture classification. Fekri-Ershad [52] introduced the Stone Texture Image (STI) dataset, designed for texture image analysis and surface defect detection. This dataset encompasses four distinct classes of stone texture images: hatchet, marble, orange travertine, and creamy travertine. In total, 60 images capture both defective and defect-free samples across these categories. For texture analysis, the dataset includes 20 images, all in JPEG format with a resolution of 72 dpi, with each class containing five samples.

Brodatz [53] is another texture dataset widely used in computer vision and image processing. It contains a total of 111 texture samples, each measuring 200 × 200 pixels. These samples cover a diverse range of natural and synthetic textures, including wood grains, stone surfaces, fabrics, and various other materials. Similarly, the USPtex dataset [54] is noteworthy, featuring 191 category of natural color textures. Each category consists of 12 images, resulting in a total of 2292 images, all standardized to dimensions of 128 × 128 pixels. The MIT Vistex dataset [55] offers a valuable alternative to the Brodatz dataset, which is not freely available for research use. It includes examples of many non-traditional textures, containing 640 texture images organized into 40 classes, with 16 images per class, each sized at 128 × 128 pixels. The Outex TC-00013 dataset [56] is a heterogeneous collection comprised of material texture such as paper, fabric, wool, stone, and more, with 68 texture classes. Each class contains 20 image samples, each of 128 × 128 pixels in size. Salzburg Texture Image Database (STex) was created for texture retrieval experiments. It is significantly larger than VisTex, and more homogeneous than Outex.

TABLE 2. Summary of the image datasets used in this study.

Dataset	Dimension	n classes	n images/ class
Rock360	$800 \times 500 - 800 \times 860$	30	12
DTD	$300 \times 300 - 600 \times 600$	47	120
STex	1024×1024	18	$10 - 77$
Rock12	4032×3024	12	$4 - 16$

Describable Textures Dataset (DTD) [57] is a collection of textural images annotated with human-centric attributes inspired by texture perception. The categories in DTD are not made based on the objects an image contains, as in a typical texture dataset, but based on textural adjectives assigned by human observers, e.g., stratified and dotted. There exists a rock dataset (Rock360) [58], [59] that was carefully examined by geologists to ensure that sufficient identifying features are reflected in the images. This dataset was further used in a psychovisual experiment for rock classification, resulting in human-centric dimensions [58].

III. MATERIALS AND METHODS

A. DATASETS

Four distinct sets of texture datasets are used to evaluate the performance of different CNN models, i.e., Rock360, DTD, STex, and Rock12. Several example images from each dataset are presented in FIGURE 2. These datasets were selected based on expert suggestion and their alignment with human texture perception, as well as their comprehensive representation of texture variations. The Rock360 dataset consists of 10 common rock types from the igneous, metamorphic, and sedimentary categories, totaling 30 rock types. Each rock type consists of 12 images, resulting in a dataset containing a total of 360 images. The second dataset, Describable Texture Dataset (DTD), it contains 5640 images grouped into 47 categories, with 120 images per category, featuring variable dimensions ranging between 300×300 and 640×640 pixels.

In addition, we used the STex dataset [60] to cover more general texture images. It is preferred over other general texture databases for its size and homogeneity. STex consists of 476 color texture images grouped into 32 classes. STex is available in three different packages, with the second and third packages comprising downsampled and split versions of the original 476 color images, each with a resolution of 1024×1024 pixels. Finally, the Rock12 dataset, provided by both University of Victoria, Canada and by Aalto University, Finland, features images of rock samples available at the se institutions. It shares similarities in rock types with the Rock360 dataset. TABLE 2 provides a summary of the main characteristics of each dataset.

B. DATA PREPROCESSING

As specified in TABLE 2, the characteristics of each dataset vary. Rock12, Stex, and Rock360 feature high-resolution images. To prepare these images for analysis, they were sliced into non-overlapping zones. Furthermore, images with more than 50% white pixels (indicative of a white background) were excluded to prevent the model from learning background features, edges, or shapes. Additionally, since Stex dataset has 32 classes, each with a different number of images, classes with less than 10 image samples were eliminated. These images were also converted to PNG format for compatibility, from their original.pnm format. DTD images did not require modifications, since most of them were already smaller in dimensions. Finally, all datasets were split into training, validation, and testing subsets in a ratio of 60%, 30%, and 10%, respectively.

TABLE 3. Image Augmentation using ImageDataGenerator.

Parameters	Values	Description
Shear	0.2	shear transformation (skewed)
Zoom	0.2	zoom-in or zoom-out of 20% on an image
Horizontal flip	True	may be flipped horizontally
Vertical flip	True	may be flipped vertically
Rotation	20	randomly rotated by up to 20 degrees
Width shift	0.1	vertical translation of up to 10% of image height
Height shift	0.1	horizontal translation of up to 10% of image width
Fill mode	Reflect	fill with mirrored or reflected neighboring pixels

TABLE 4. Summary of preprocessed datasets.

Dataset	Dimension	n classes	n images/ class	n sliced images/ class
Rock360	300×300	30	12	$30 - 60$
DTD	$300 \times 300 - 600 \times 600$	47	120	120
STex	300×300	18	$10 - 77$	$200 - 1100$
Rock12	300×300	12	$4 - 16$	$250 - 950$

Deep learning networks require a substantial amount of data for effective feature learning during the training process. Acquiring such a large volume of data can be challenging for many application domains, including texture analysis. One way to deal with this is by using data augmentation techniques, a method used to increase the size of a training dataset by applying various transformations to existing data samples [61]. Several research studies [62], [63] have shown that data augmentation is effective for image classification tasks. The most used data augmentation approach in rock texture classification is flipping, rotating, scaling, etc. [10], [37], [64]. Cui et al. [65] used the CutMix data augmentation method in their study to classify fine-grained rocks. It involves randomly clearing some pixel values and replacing them with pixels from another image, this improves the localization ability of the model by directing its focus towards less discriminative parts of the classified object [66].

In this study, we used the Keras *ImageDataGenerator* tool for its automated loading and pre-processing for training purposes. A summary of the data augmentation settings

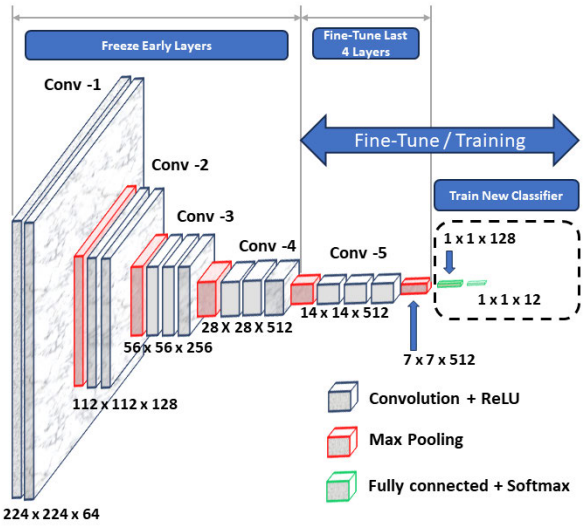


FIGURE 3. Visualization of VGG16 architecture, illustrating frozen and trainable layers during transfer learning and fine-tuning. Image adapted from Ref. [69].

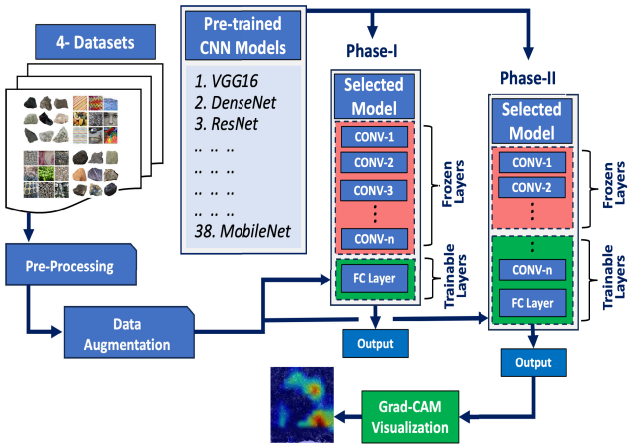


FIGURE 4. Workflow of the overall transfer learning methodology, with frozen layers drawn inside the red box. Only the layer highlighted inside the green box was used for training. Phase I demonstrates the transfer learning method, while Phase II illustrates transfer learning with fine-tuning.

used in this study is provided in TABLE 3. The resulting pre-processed datasets are also summarized in TABLE 4.

C. PRE-TRAINED CNN MODELS

While CNNs are widely used in many applications for their ability to automatically learn hierarchical features from the data, they require a large number of annotated images. This presents a significant challenge in scenarios where obtaining a substantial volume of labeled images is impractical or expensive. One solution to this challenge is transfer learning [67], where knowledge gained from training a model on one task/ domain is applied to another related task/ domain. It is computationally effective as it necessitates a smaller training dataset and significantly reduces training time compared to training deep learning models from scratch.

Many pre-trained CNN models have been developed for various applications, each with its unique architecture and characteristics, e.g., those available through Keras. Other well-known examples pre-trained on large-scale datasets, e.g., ImageNet, are also available, e.g., AlexNet, VGGNet, and DenseNet [68]. These models have been developed with increasing depth over time, leading to more sophisticated and powerful networks for, e.g., image classification and object detection. We refer readers to referenced papers for more details on these pre-trained networks.

In the first part of this study, we started by setting up all our models using pre-existing ImageNet weights. We selected 38 pre-trained CNN models, each of which had already learned to extract useful features from images through extensive training on the ImageNet dataset. The models were then evaluated on the four datasets described previously. To illustrate our application of transfer learning, let us consider the VGG16 model. VGG16 is a CNN architecture trained on the ImageNet dataset, consisting of 1.2 million images classified into 1000 categories (architecture shown in FIGURE 3). In its original form, the fully connected layer of VGG16 produces 1000 distinct output labels, whereas our specific dataset (e.g., Rock12) contains only 12 classes. To adapt VGG16 for our task, we first removed its fully connected layer (also referred to as the top layer), thus retaining only the convolutional and pooling layers. This ensured that the model's rich feature extraction capabilities, developed during the initial training on ImageNet, were preserved.

Next, we appended a new fully connected layer and an output layer customized to our dataset, ensuring the total number of outputs matched the 12 classes in our dataset. We froze all layers preceding the newly added fully connected layers to maintain the integrity of the pre-learned features during this process. We applied the same methodology to all 38 models used in this study. By doing so, we ensured that during the training phase, the weights of the newly added layers were updated through backpropagation while the weights of the pre-existing layers remained unchanged. This approach allowed us to leverage the sophisticated feature extraction abilities of the pre-trained CNN models while tailoring the final classification layer to our specific needs.

In the second phase of our experiment, we performed fine-tuning on the model that achieved optimal results. It is a specific technique within transfer learning which, in addition to adding new layers, partially unfreezes and further trains some of the pre-trained layers. This approach allows for greater customization of the pre-trained model to a specific dataset, often leading to improved performance compared to feature extraction alone. Thus, instead of keeping all pre-trained layers frozen and only changing the output layer (like phase one), we empirically selected to unfreeze the last four convolutional layers of the VGG16 model, in addition to the output layer. This selective unfreezing was based

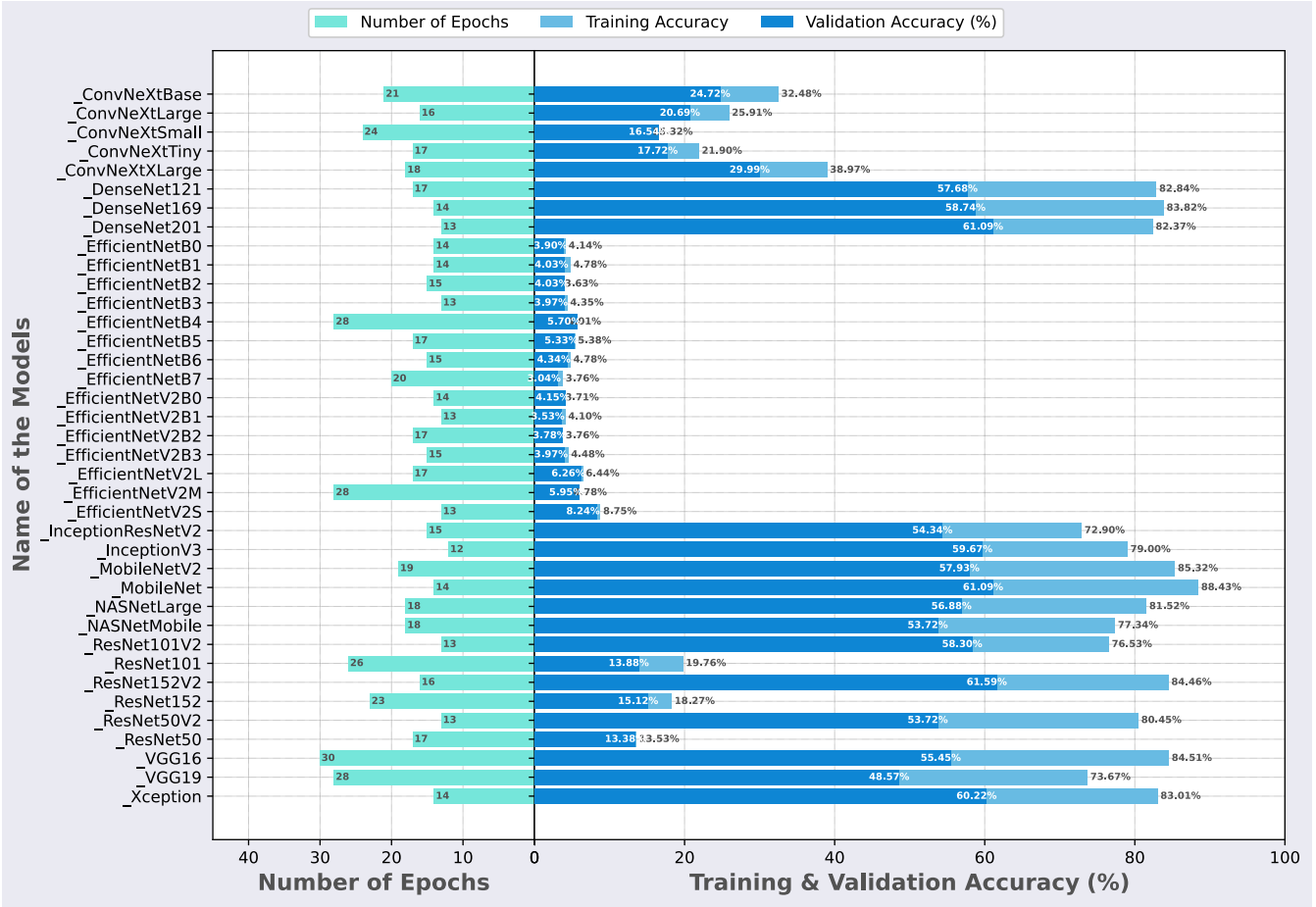


FIGURE 5. Comparative analysis of 38 pre-trained models on the Rock360 dataset, providing an overview of training and validation accuracy, along with its respective epoch count. Note also a limitation of this figure, where the bars for training accuracy are overlapped by the validation bars. As such, the light blue bars only show the difference between training and validation accuracy. Nevertheless, the training accuracy rates are shown to mitigate this data visualisation limitation.

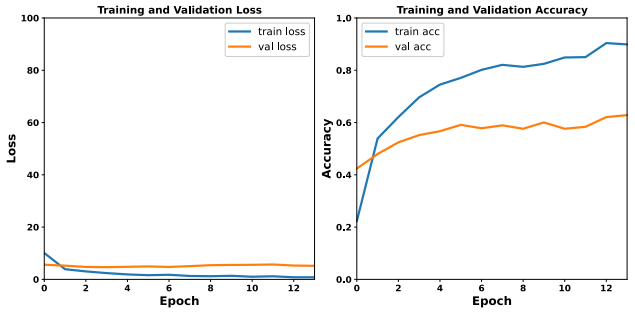


FIGURE 6. Learning curves showing the training and validation loss and accuracy across epochs for MobileNet architecture trained on the Rock360 dataset. MobileNet performed comparatively better than other architectures in FIGURE 5.

on the hypothesis that by updating the weights of these deeper convolutional layers and the fully connected output layers during training, the model could learn more relevant features, potentially leading to improved performance and generalization on our task. The overall methodology is also illustrated in FIGURE 4.

D. CNN INTERPRETABILITY USING GRAD-CAM

Grad-CAM, short for Gradient-weighted Class Activation Mapping [16], [70], is a technique used to visualize where in an image a DL network focuses to make its decision. To compute Grad-CAM, a prediction by, e.g., CNN, was initially generated by feeding an image into the network. Since CNN consists of multiple layers, each conducts computations during the forward pass, we will calculate the importance map α for a class C, at each spatial location (i, j) in feature map A:

$$\alpha(C, i, j) = \sum_i \sum_j \left(\frac{\partial P(C)}{\partial A} \right)_{ij}, \tag{1}$$

where $P(C)$ and ∂ represent the prediction probability of class C and prediction gradient, respectively. Note that A is the output of the final convolutional layer, as illustrated in FIGURE 4. The importance scores $\alpha(C, i, j)$ are then used to

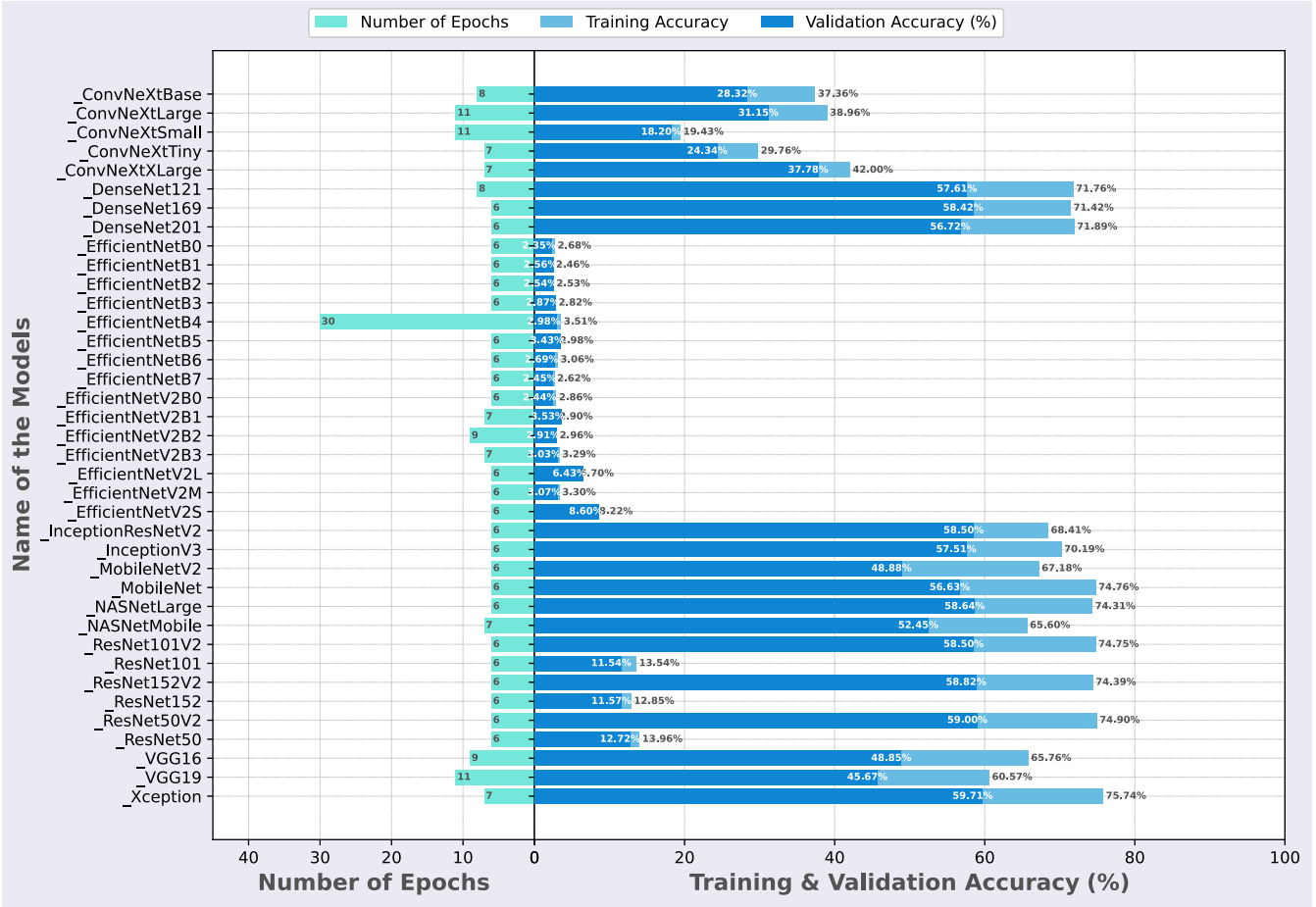


FIGURE 7. Comparative analysis of 38 pre-trained models on DTD dataset, providing an overview of training and validation accuracy, along with epoch count.

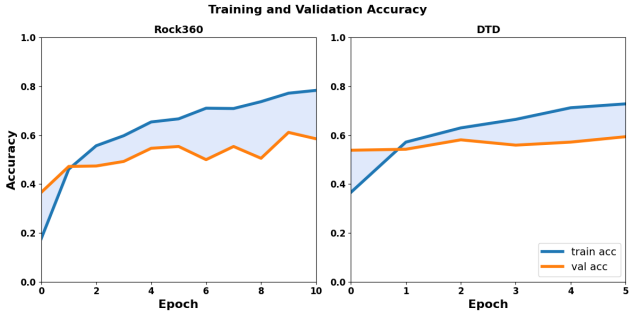


FIGURE 8. Training and validation accuracy across epochs for InceptionV3 architecture, trained on Rock360 and DTD datasets. InceptionV3 converged faster for DTD than Rock360 as revealed by differences in epoch count.

generate the Grad-CAM heatmap, calculated as follows:

$$\text{Grad-CAM}(C)(i,j) = \text{ReLU}(\alpha(C,i,j) \cdot A_{ij}), \qquad (2)$$

in which ReLU is the Rectified Linear Unit function (which sets negative values to zero), and A_{ij} is the feature map value at position i, j. Finally, we overlay the Grad-CAM heatmap on the original image to visualize which parts of the image the network paid attention to when making its decision for class C. Grad-CAM heatmap is then superimposed onto the original image, to visualize which image regions were given significant attention during its classification for class C.

IV. RESULTS AND DISCUSSION

In this section, we present the results derived from our transfer learning experiments on 38 pre-trained models across four datasets, i.e., Rock360, DTD, STex, and Rock12.

A. PHASE I: TRANSFER LEARNING

The results in FIGURE 5 highlight how well different models performed on the Rock360 dataset, with MobileNet showing comparatively better performance. As depicted, none of the pre-trained models performed satisfactorily. In fact, for certain architectures, accuracy was notably lower. Several factors could account for this, e.g., data characteristics. The efficacy of transfer learning on pre-trained models is another factor which may be affected by, e.g., model complexity (depth, width, or layer types). Furthermore, even in cases where accuracy appears higher, there is a substantial gap between training and validation accuracy across nearly

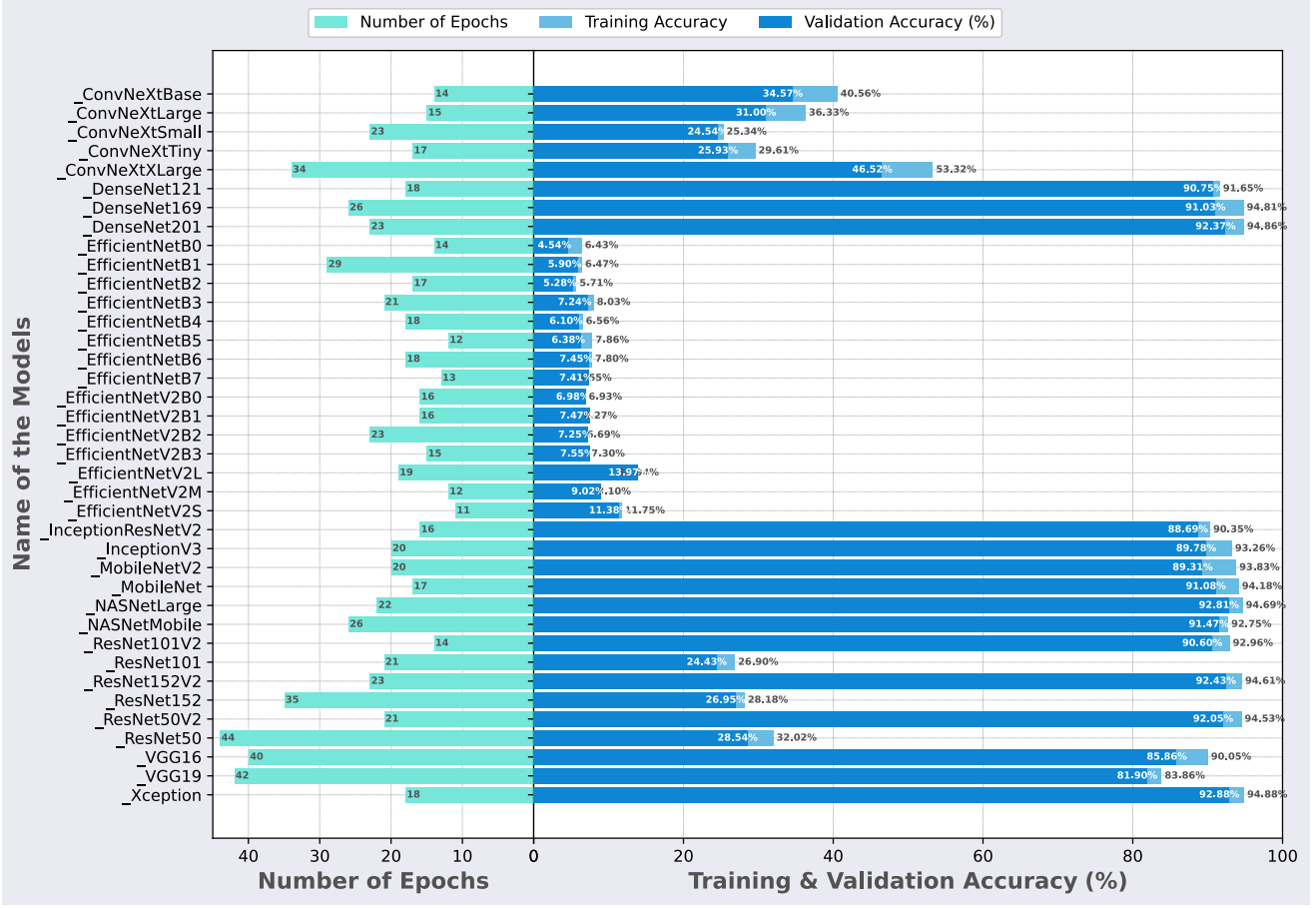


FIGURE 9. Comparative analysis of pre-trained models on the STex dataset, showcasing an overview of training and validation accuracy for 38 distinct models across epoch count. Here, Xception achieved the highest classification accuracy.

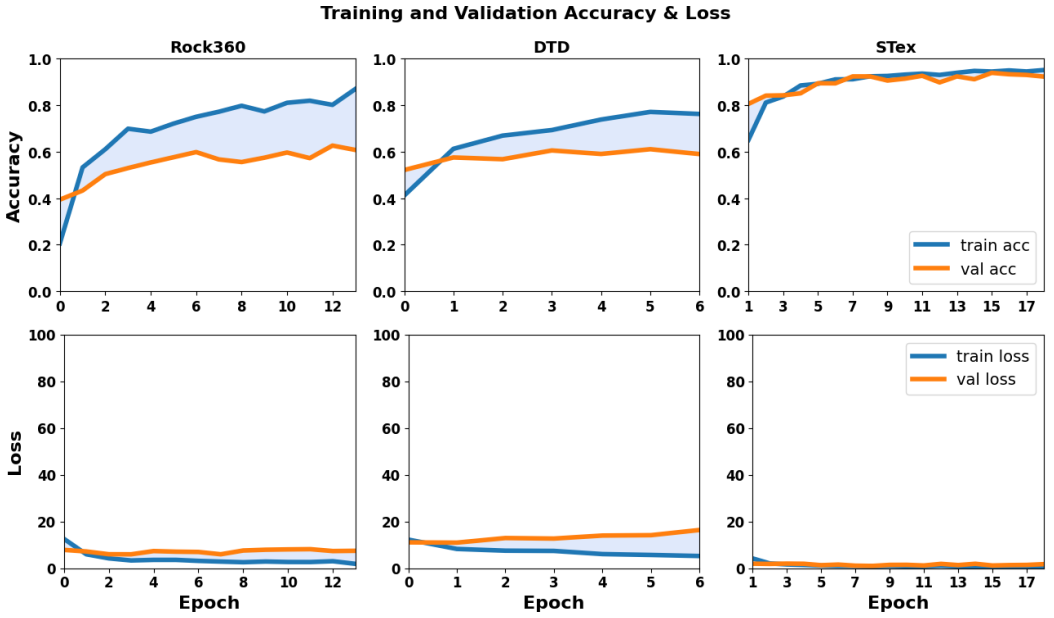


FIGURE 10. The learning curves of Xception architecture, showing the training and validation accuracy and loss across epochs and for the three different datasets, i.e., Rock360, DTD, and STex.

FIGURE 11. Comparative analysis of 38 pre-trained models on the Rock12 dataset, showcasing an overview of training and validation accuracy across epoch count.

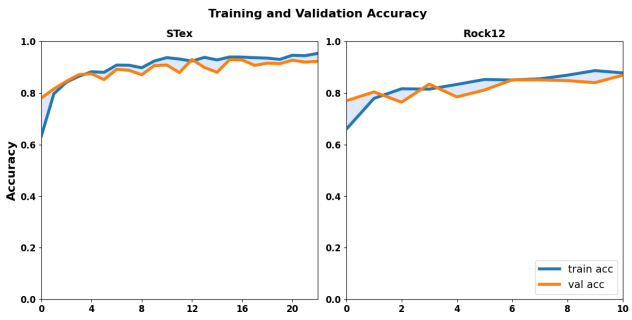


FIGURE 12. Training and validation accuracy across epochs for DenseNet201 architecture, trained on STex and Rock12.

all models. For example, consider the learning curve of MobileNet shown in FIGURE 6. With an increase in the number of epochs, the difference between training and validation accuracy widens. This trend strongly suggests that the given model lacks generalizability or that the model was overfitted.

The performance of the same 38 pre-trained models on the DTD dataset is summarized in FIGURE 7. Similar to the observations made for Rock360, several models consistently showed poor accuracy on DTD dataset, e.g., EfficientNet. However, despite the overall lower accuracy compared to when Rock360 dataset was used, we observed an improvement in the generalizability of these models. This highlights the unique challenges posed by DTD for transfer learning. For example, InceptionV3 exhibited relatively better performance compared to other models, demonstrating its potential for transfer learning specifically on this dataset. This was evidenced by an increase in accuracy immediately at the start of the epoch compared to when the Rock360 dataset was used, see FIGURE 8. Then, there is a reduced gap between the training and validation accuracy in the DTD dataset, see the shaded area in the figure. Moreover, the validation accuracy is relatively more stable and achieved by a fewer number of epochs. The higher number of classes and larger number of images per class in DTD may have contributed to this improvement, possibly enhancing the regularization capabilities of the models.

FIGURE 10 shows the performance of 38 pre-trained models on STex dataset. In general, most of the models exhibited better accuracy compared to the previous two datasets, i.e., Rock360 and DTD. However, similar to the observations made for previously, certain models consistently displayed poor accuracy on STex dataset, e.g., EfficientNet.

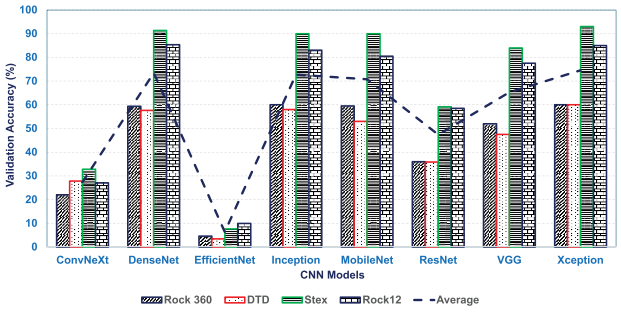


FIGURE 13. Validation accuracy of various CNN models across different datasets, with DenseNet and Xception showing the relatively highest accuracy on all four datasets.

An outstanding performance is shown by the Xception model, with 92.88% and 94.88% accuracy in the validation and training sets, respectively. Compared to its performance on Rock360 and DTD, Xception demonstrates closely aligned validation and training loss curves in FIGURE 9, indicating its improved generalizability and reduced overfitting tendencies when using STex dataset. Additionally, the gap between the training and validation accuracy is also significantly reduced.

Finally, the performance of the same set of pre-trained models was also assessed on the Rock12 dataset. As shown in FIGURE 11, DenseNet201 emerged as a standout performer with 87.84% and 85.27% accuracy in training and validation, respectively. Similar to the results for STex in FIGURE 10, the gaps between training and validation accuracy are low for most models. However, the overall accuracy for training and validation across all models was also lower. FIGURE 12 illustrates the learning curve for DenseNet201 for Rock12, showing a trend similar to the STex dataset with lower overall accuracy. Notably, in this case, the validation accuracy curve fluctuates slightly more, which could be attributed to the lower number of images in the dataset.

Given that CNN models are often considered black boxes and providing exact reasons for their performance can be challenging, we have offered a probable explanation based on factors such as model depths, architectures, complexities, and dataset properties. FIGURE 13 shows the overall comparison based on the validation accuracy of various base models evaluated in this study across four datasets. We averaged the validation accuracy for each family across different variants (e.g., EfficientNet-B0 to B7, ResNet-50 to ResNet-152) to obtain a generalized performance measure for each architecture. As shown in FIGURE 13, the performance across four datasets reveals distinct trends in how different models handle rock image data. Models like ResNet, DenseNet, Inception, and Xception have more versatile and robust architectures for capturing complex and varied features.

These models include elements such as residual connections (ResNet), dense connectivity (DenseNet), multi-scale feature extraction (Inception), and depthwise separable convolutions (Xception) that help in learning intricate patterns and details present in specialized datasets like rock and texture images. The ability to capture and process features at multiple scales (Inception), ensure gradient flow and feature reuse (DenseNet), and leverage deep residual connections (ResNet) allows these models to better adapt to and learn from specialized datasets. Similarly, Xception's use of depthwise separable convolutions also enhances its ability to separate spatial and cross-channel features, which is crucial for texture analysis.

The poor performance of EfficientNet models can be attributed to several factors, including the model's architecture and the specific characteristics of the datasets used in this study. EfficientNet models are often pretrained on large-scale datasets like ImageNet, which consists primarily of object-centric images. Consequently, the learned features from such datasets are more geared towards object recognition rather than texture analysis, which often requires capturing fine details and subtle pattern variations. Although transfer learning was applied by modifying the output layers to fit the new datasets, the inherent feature extraction and representation capabilities of EfficientNet may be less effective for these custom datasets. Furthermore, EfficientNet employs compound scaling to balance depth, width, and resolution, and this scaling might be less effective for textures that require very high resolution and detailed feature extraction. Texture recognition tasks often benefit from high-resolution input images and detailed feature maps, which EfficientNet's scaling strategy might not fully utilize. Additionally, the training and optimization of EfficientNet are geared towards achieving a balance between accuracy and computational efficiency. This focus might compromise learning highly detailed features specific to texture datasets.

B. PHASE II: TRANSFER LEARNING WITH FINE-TUNING

To further improve the performance of the pre-trained models, we conducted fine-tuning by training the weights of the last convolutional layer with the Rock12 dataset. Although a higher validation accuracy of 85% was previously achieved with DenseNet201, we chose to fine-tune VGG16 instead, see FIGURE 3. With only a slightly lower accuracy of 81.23% compared to DenseNet201, VGG16 has simpler architecture and lower computational demands. We consider this a worthy trade-off. After fine-tuning the model, a significant improvement in performance was observed with accuracy of 91%. As illustrated in FIGURE 14, the training and validation accuracy is closely aligned across epochs, indicating a robust performance. This alignment suggests effective generalization, demonstrating consistent accuracy on both seen and unseen data. Although there is some fluctuation in the validation curve, it is likely attributed to the use of a dataset containing a relatively small number of images.

C. FIRST RESULTS FROM GRAD-CAM VISUALISATION

To delve deeper into the model's performance at the class level, gain insights into the accuracy of individual classes, and identify instances of misclassification among classes to

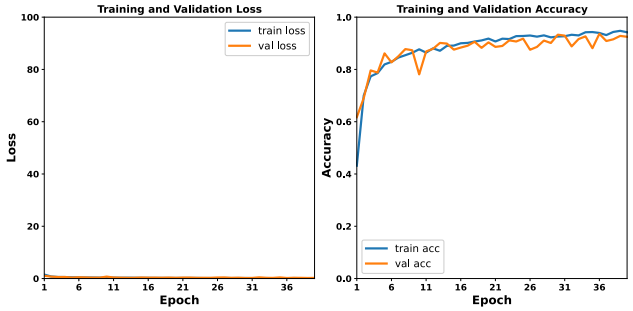


FIGURE 14. Learning curve illustrating the training and validation loss and accuracy progression across the number of epochs for VGG16 architecture with fine-tuning on the Rock12 dataset.

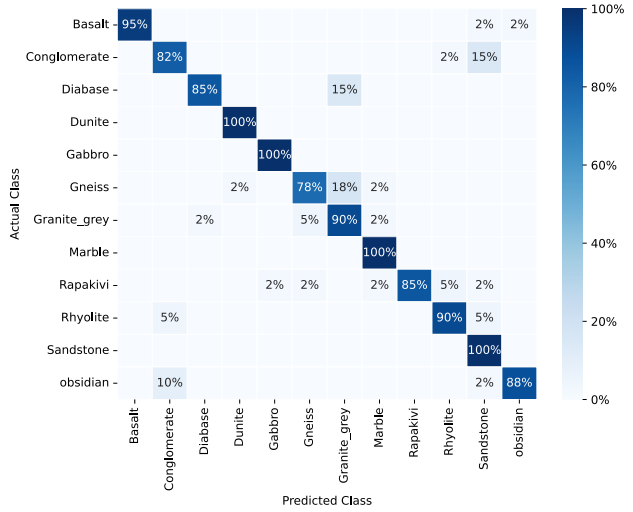
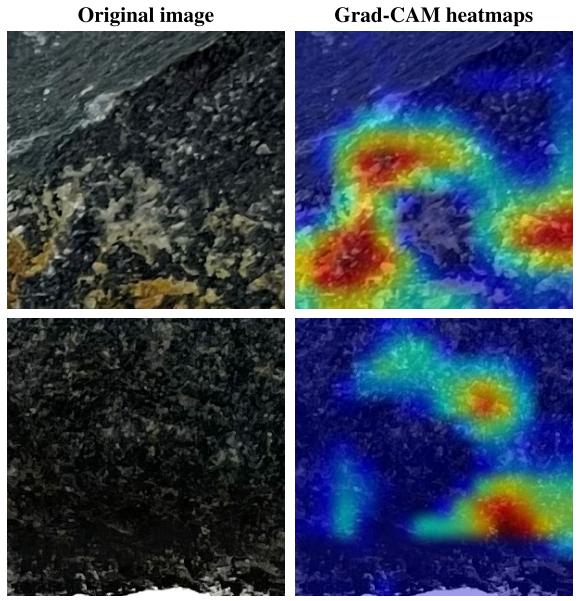
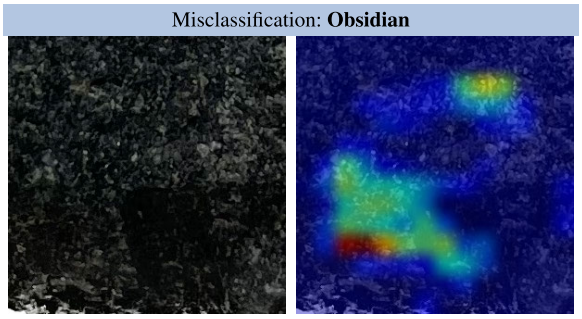


FIGURE 15. Confusion matrix illustrating the classification results of 12 different rock types using the VGG16 architecture with transfer learning and fine-tuning; the matrix's main diagonal represents the percentage of correctly classified images, while the other elements indicate the percentage of images in one category that are incorrectly classified into other categories.



(a) Correctly classified Basalt



(b) Misclassified Basalt

FIGURE 16. Correctly classified and misclassified basalt rock images, with their respective Grad-CAM generated visualisations. Image regions overlaid by red are considered as having important features for classifying the image to its respective class, while the blue regions were ignored.

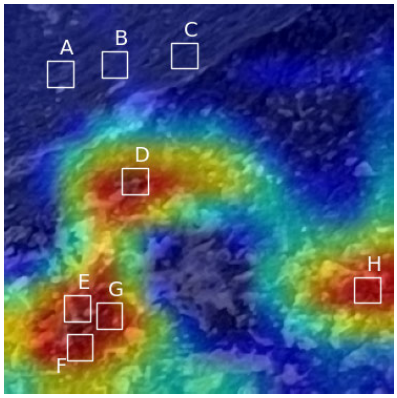


FIGURE 17. Heatmap image from Grad-Cam with boxes indicating the patches from low-weighted (A to C) and high-weighted (D to H) regions; each patch is 20 × 20 pixels.

pinpoint areas for improvement, we computed the confusion matrix in FIGURE 15 for the fine-tuned VGG16 architecture. We observe that for most classes, accuracy is relatively high, with some reaching 100%. In this figure, we can also observe the common misclassifications, e.g., gneiss being misclassified as $\text{granite}_{\text{grey}}$. Similarly, we also note instances of conglomerate being misclassified as sandstone and diabase as $\text{granite}_{\text{grey}}$. Other details of misclassified images are listed in TABLE 5 in the appendix. To further analyze and understand the results, we employed Grad-CAM for interpretability. Using Grad-CAM, we pinpointed image regions that contributed to the misclassification, thus gaining initial insights into the behaviour of VGG16.

Examples of heatmap images obtained by Grad-CAM are shown in FIGURE 16, both for correctly classified and misclassified basalt rocks. These heatmaps were generated by overlaying the original image with false color mask ranging from blue, green, yellow, to red. Pixel regions overlaid with color towards red receive an increasing emphasis in importance for the classification. This means that important

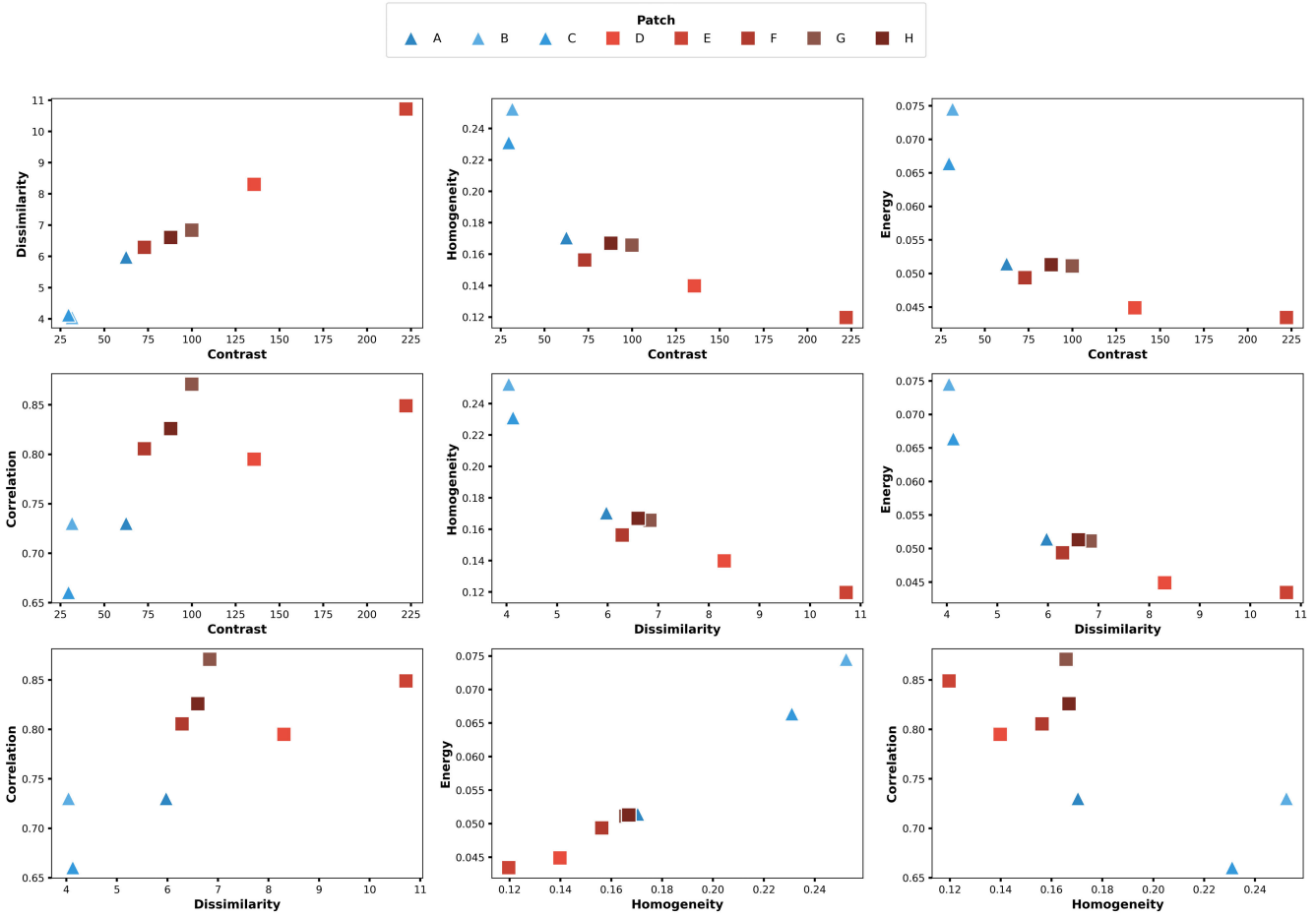


FIGURE 18. Scatter plot analysis of GLCM statistics measures; each point on the plot represents a correlation between combinations of five GLCM statistics measures; triangles denote data points from low or minimal-weighted regions, while square boxes represent data points from the high-weighted regions.

basalt-related features were found in the red pixels region in FIGURE 16a, while the blue pixels region were ignored. On the other hand, obsidian features were found in a basalt image in FIGURE 16b, a clear misclassification case.

As it is, the Grad-CAM heatmap analyses do not offer sufficient detail to justify or understand both the correct and incorrect classifications. To provide further understanding, we employed a quantitative approach to explore these heatmaps using GLCM texture features. Eight patches were selected from high-weighted (maximum attention) and low-weighted (zero or minimal attention) regions from the heatmap, as shown in FIGURE 17, and their GLCM texture features computed. GLCM is a method for analyzing an the texture of an image by examining the relationships between neighboring pixels. It captures the spatial arrangement of textures and considers factors such as the distance between pixels, the angle of pixel pairs, and the discrete intensity levels present in the image. Derived from the GLCM, various statistics describing texture can be obtained, e.g., contrast, dissimilarity, and homogeneity. Further details on these statistical texture, refer to Ref. [24].

The distribution of GLCM features of the eight patches are visualized using scatter plots to detect any distinct clustering patterns between patches from high-attention and low-attention regions. This statistical analysis can determine whether the areas identified by Grad-CAM also exhibit unique texture characteristics that could explain the focus and decision-making process of the network. A clear distinction between patches from high-attention and low-attention regions can be observed in FIGURE 18, indicating that the CNN models rely on specific texture features in the high-attention regions to make classification decisions. This separation suggests that the neural network relies on specific texture features in the high-attention regions to make classification decisions. Additionally, it is also important to remember that what the model determines as features does not necessarily correspond to features used by human experts to perform rock identification. Therefore, further research is necessary to analyze both statistical properties and expert-derived features in rock images, and to examine such features using methods like Grad-CAM to visualize whether the model consistently focuses on the same

TABLE 5. List of misclassified images with their true and predicted class names.

Image filename	True class	Predicted class
Basalt/patch_167_Basalt3b.png	Basalt	Obsidian
Basalt/patch_98_Basalt3c.png	Basalt	Sandstone
Conglomerate/patch_540_Conglomerate_sample11 (3).png	Conglomerate	Rhyolite
Conglomerate/patch_676_Conglomerate_sample05 (4).png	Conglomerate	Sandstone
Conglomerate/patch_851_Conglomerate_sample05 (4).png	Conglomerate	Sandstone
Conglomerate/patch_854_Conglomerate_sample09 (4).png	Conglomerate	Sandstone
Conglomerate/patch_936_Conglomerate_sample03 (3).png	Conglomerate	Sandstone
Conglomerate/patch_946_Conglomerate_sample09 (4).png	Conglomerate	Sandstone
Conglomerate/patch_982_Conglomerate_sample03 (3).png	Conglomerate	Sandstone
Diabase/patch_593_Diabase_sample09 (4).png	Diabase	Granite_grey
Diabase/patch_632_Diabase_sample05 (3).png	Diabase	Granite_grey
Diabase/patch_718_Diabase_sample11 (4).png	Diabase	Granite_grey
Diabase/patch_765_Diabase_sample05 (3).png	Diabase	Granite_grey
Diabase/patch_863_Diabase_sample10 (3).png	Diabase	Granite_grey
Diabase/patch_906_Diabase_sample10 (3).png	Diabase	Granite_grey
Gneiss/patch_596_Gneiss_sample05 (3).png	Gneiss	Granite_grey
Gneiss/patch_637_Gneiss_sample10 (4).png	Gneiss	Granite_grey
Gneiss/patch_677_Gneiss_sample07 (3).png	Gneiss	Granite_grey
Gneiss/patch_680_Gneiss_sample10 (4).png	Gneiss	Granite_grey
Gneiss/patch_732_Gneiss_sample07 (3).png	Gneiss	Granite_grey
Gneiss/patch_776_Gneiss_sample05 (3).png	Gneiss	Granite_grey
Gneiss/patch_819_Gneiss_sample07 (3).png	Gneiss	Marble
Gneiss/patch_989_Gneiss_sample05 (3).png	Gneiss	Granite_grey
Gneiss/patch_992_Gneiss_sample05 (3).png	Gneiss	Dunite
Granite_grey/patch_1034_Granite_grey_sample04_04.png	Granite_grey	Diabase
Granite_grey/patch_508_Granite_grey_sample09_04.png	Granite_grey	Gneiss
Granite_grey/patch_555_Granite_grey_sample09_04.png	Granite_grey	Gneiss
Granite_grey/patch_765_Granite_grey_sample04_04.png	Granite_grey	Marble
Rapakivi/patch_500_Rapakivi-granite_sample01 (2).png	Rapakivi	Marble
Rapakivi/patch_501_Rapakivi-granite_sample05 (4).png	Rapakivi	Rhyolite
Rapakivi/patch_507_Rapakivi-granite_sample01 (2).png	Rapakivi	Gabbro
Rapakivi/patch_545_Rapakivi-granite_sample05 (4).png	Rapakivi	Rhyolite
Rapakivi/patch_814_Rapakivi-granite_sample12 (4).png	Rapakivi	Sandstone
Rapakivi/patch_850_Rapakivi-granite_sample01 (2).png	Rapakivi	Gneiss
Rhyolite/patch_459_Rhyolite_sample12 (4).png	Rhyolite	Sandstone
Rhyolite/patch_679_Rhyolite_sample12 (4).png	Rhyolite	Sandstone
Rhyolite/patch_720_Rhyolite_sample10 (1).png	Rhyolite	Conglomerate
Rhyolite/patch_775_Rhyolite_sample12 (4).png	Rhyolite	Conglomerate
Obsidian/patch_166_Obsidian3a.png	Obsidian	Conglomerate
Obsidian/patch_179_Obsidian3a.png	Obsidian	Conglomerate
Obsidian/patch_182_Obsidian3a.png	Obsidian	Conglomerate
Obsidian/patch_193_Obsidian3a.png	Obsidian	Conglomerate
Obsidian/patch_211_Obsidian3a.png	Obsidian	Sandstone

features during classification. This additional consideration will help to improve the interpretability of the model and to find a correlation with geological features by human experts.

V. CONCLUSION

We have conducted an in-depth investigation of transfer learning using 38 ImageNet pre-trained CNN models for texture analysis across four distinct texture datasets, i.e., Rock360, DTD, STex and Rock12. The study highlights the efficacy of transfer learning in texture classification tasks and offers valuable perspectives on the performance of different CNN architectures on different datasets. Although certain models demonstrated promising performance on specific datasets, e.g., MobileNet on Rock360 and Xception on STex dataset, others generally struggled to achieve satisfactory accuracy. This indicates the influence of model complexity and dataset characteristics on transfer learning efficacy.

In this study, we have also found that the different models performed better across different datasets when there were more images for each category. Fine-tuning

TABLE 6. Summary of strengths and limitations of various CNN architectures.

CNN Architectures	Year	Strengths	Limitations
AlexNet [72]	2012	• **Historical significance**: achieved state-of-the-art recognition accuracy, surpassing traditional ML and computer vision methods. • **Simplicity**: relatively simple architecture, making it easier to understand and implement.	• **Outdated**: modern architectures significantly outperform AlexNet in terms of accuracy and efficiency. • **Size**: larger and more computationally intensive compared to recent models.
VGG [13]	2014	• **Simplicity**: simple architecture. • **Depth**: increased depth (16,19) allows for learning complex features.	• **Computationally intensive**: a large number of parameters leads to high computational and memory requirements. • **Overfitting**: prone to overfitting on smaller datasets.
Inceptions [73]	2015	• **Efficiency**: captures multi-scale features through parallel convolution. • **Performance**: balanced depth and computational cost effectively.	• **Complexity**: architecture and hyperparameters are complex, making it challenging to modify and tune. • **Implementation**: more challenging to implement compared to simpler models.
ResNet [14]	2016	• **Residual connections**: these connections help to mitigate the vanishing gradient problem, thereby allowing for very deep networks. • **Performance**: high performance on various image classification tasks.	• **Training time**: increased depth can lead to increased training time. • **Complexity**: implementing residual connections.
DenseNet [74]	2017	• **Features reuse**: dense connections promotes feature reuse, improving efficiency. • **Parameter efficiency**: lower parameters count compared to traditional architecture with similar depth. • **Performance**: strong performance in terms of both accuracy and speed.	• **Memory use**: dense connectivity can lead to memory issues during training. • **Complexity**: complex due to the interconnected layers.
MobileNet [75]	2018	• **Efficiency**: designed for mobile and embedded vision applications, uses depth-wise separable convolutions to reduce complexity and size. • **Lightweight**: fewer parameters and lower computational cost compared to traditional CNNs. • **Performance**: a good balance between accuracy and speed especially in resource-constrained environments.	• **Accuracy**: slightly lower accuracy compared to larger, more complex models. • **Capacity**: may struggle with very large or complex datasets due to its lightweight architectures.
EfficientNet [76]	2019	• **Scalability**: uses a compound scaling method to balance network depth, width, and resolution, providing better performance with fewer parameters. • **Performance**: high accuracy on Image-Net with significantly fewer parameters. • **Efficiency**: optimized for both speed and accuracy, making it suitable for a wide range of applications.	• **Complexity**: more complex architectures due to compound scaling. • **training time**: can require longer training time due to sophisticated scaling strategy.

the model significantly improved classification accuracy, as shown by the accuracy achieved by VGG16 architecture on the Rock12 dataset. Despite several misclassifications, particularly among similar rock classes, the fine-tuned model exhibited consistent accuracy and demonstrated potential for practical applications in rock classification. In conclusion, our study contributes to advancing the understanding of transfer learning in texture classification and provides valuable information for researchers and practitioners in fields such as geology, computer vision, and image processing. Future research directions may involve exploring hybrid approaches that integrate domain-specific knowledge with deep learning techniques to effectively address the complexities of texture analysis and classification.

APPENDIX A
CLASSIFICATION RESULTS FOR VGG16 WITH FINE-TUNING

See Table 5.

APPENDIX B
SUMMARY OF SELECTED CNN ARCHITECTURES

With AlexNet in 2012, CNN architectures have been continuously modified and upgraded; most upgrades were performed on network layers or depth, and several parameter optimization strategies were implemented. These CNN models use a multi-layered architecture, with initial layers extracting low-level features and final layers extracting high-level features. Some CNN architectures discussed in

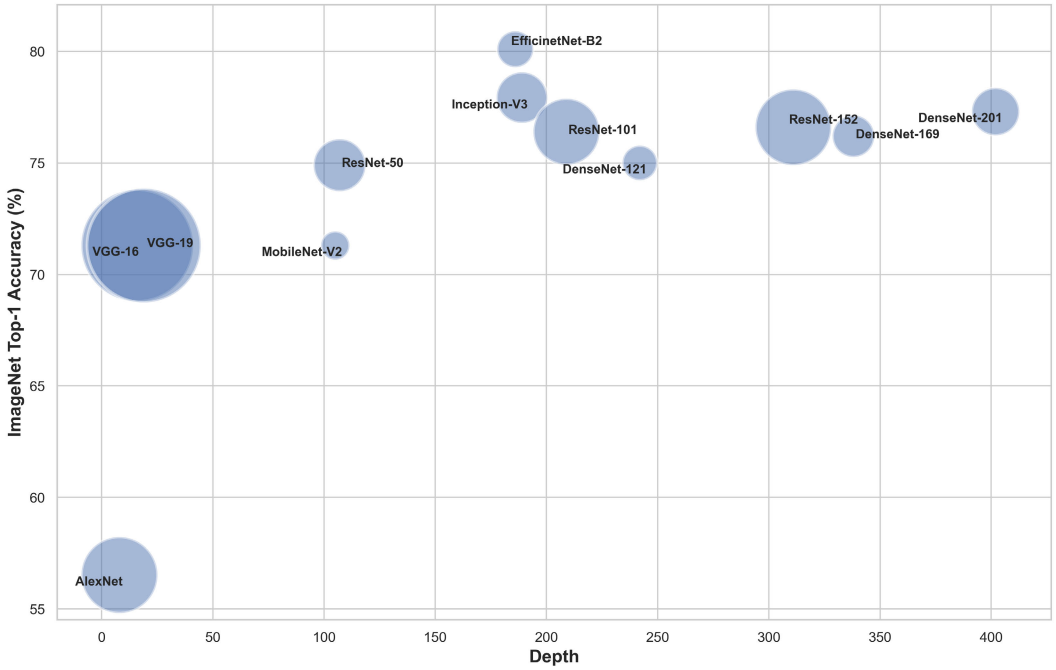


FIGURE 19. Bubble plot illustrating the depth and top-1 accuracy of various pre-trained models on the ImageNet dataset. Each bubble represents a different model, bubble size is relative to the number of parameters in millions, and depth refers to the topological depth of the network (including activation layers, batch normalization layers, etc).

the related work section are summarized in Table 6, and their key parameters are plotted in Figure 19. For a more comprehensive understanding of the topic, we recommend exploring the research conducted by Alzubaidi et al. [71] along with the cited papers on the models.

REFERENCES

[1] M. W. Smith, "Roughness in the earth sciences," *Earth-Sci. Rev.*, vol. 136, pp. 202–225, Sep. 2014.

[2] L. Wang, N. Xu, and J. Song, "Decoding intra-tumoral spatial heterogeneity on radiological images using the Hilbert curve," *Insights Imag.*, vol. 12, no. 1, pp. 1–10, Dec. 2021.

[3] H. R. Wenk, *Preferred Orientation in Deformed Metal and Rocks: An Introduction to Modern Texture Analysis*. Cambridge, MA, USA: Academic Press, 1985.

[4] B. Leiss, K. Ullemeyer, K. Weber, H. G. Brokmeier, H.-J. Bunge, M. Drury, U. Faul, F. Fueten, A. Frischbutter, H. Klein, W. Kuhs, P. Launeau, G. E. Lloyd, D. J. Prior, C. Scheffzük, T. Weiss, K. Walther, and H.-R. Wenk, "Recent developments and goals in texture research of geological materials," *J. Struct. Geol.*, vol. 22, nos. 11–12, pp. 1531–1540, Nov. 2000.

[5] J. Malik, S. Belongie, T. Leung, and J. Shi, "Contour and texture analysis for image segmentation," *Int. J. Comput. Vis.*, vol. 43, no. 1, pp. 7–27, Jun. 2001.

[6] M. Sonka, V. Hlavac, and R. Boyle, *Image Processing, Analysis, and Machine Vision*. Belmont, CA, USA: Wadsworth, 2013.

[7] E. F. Mcbride, "A classification of common sandstones," *SEPM J. Sedimentary Res.*, vol. 33, no. 3, pp. 664–669, 1963.

[8] L. Lepistö, "Colour and texture based classification of rock images using classifier combinations," Ph.D. thesis, Dept. Inf. Technol., Tampere Univ. Technol., Tempere, Finland, 2006.

[9] C. Pellant and H. Pellant, *Rocks and Minerals*. London, U.K.: Dorling Kindersley, 2021.

[10] X. Ran, L. Xue, Y. Zhang, Z. Liu, X. Sang, and J. He, "Rock classification from field image patches analyzed using a deep convolutional neural network," *Mathematics*, vol. 7, no. 8, p. 755, Aug. 2019.

[11] W. Zhang, H. Li, Y. Li, H. Liu, Y. Chen, and X. Ding, "Application of deep learning algorithms in geotechnical engineering: A short critical review," *Artif. Intell. Rev.*, vol. 54, no. 8, pp. 5633–5673, Dec. 2021.

[12] X. Wang, "Deep learning in object recognition, detection, and segmentation," *Found. Trends Signal Process.*, vol. 8, no. 4, pp. 217–382, 2016.

[13] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. 3rd Intl. Conf. Learning Represent. (ICLR)*, 2015, pp. 1–14.

[14] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[15] R. Geirhos, P. Rubisch, C. Michaelis, M. Bethge, F. A. Wichmann, and W. Brendel, "ImageNet-trained CNNs are biased towards texture; Increasing shape bias improves accuracy and robustness," in *Proc. Int. Conf. Learn. Represent.*, 2019, pp. 1–22.

[16] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual explanations from deep networks via gradient-based localization," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 618–626.

[17] M. Tuceryan and A. K. Jain, *Texture Analysis*. Singapore: World Scientific, 1993, pp. 235–276.

[18] A. A. Efros and T. K. Leung, "Texture synthesis by non-parametric sampling," in *Proc. 7th IEEE Int. Conf. Comput. Vis.*, Sep. 1999, pp. 1033–1038.

[19] M. Crosier and L. D. Griffin, "Using basic image features for texture classification," *Int. J. Comput. Vis.*, vol. 88, no. 3, pp. 447–460, Jul. 2010.

[20] T. Reed, "A review of recent texture segmentation and feature extraction techniques," *Comput. Vis. Image Understand.*, vol. 57, no. 3, pp. 359–372, May 1993.

[21] P. S. Heckbert, "Survey of texture mapping," *IEEE Comput. Graph. Appl.*, vol. CGA-6, no. 11, pp. 56–67, Nov. 1986.

[22] H. Tamura, S. Mori, and T. Yamawaki, "Textural features corresponding to visual perception," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-8, no. 6, pp. 460–473, Jun. 1978.

[23] T. Harinie, I. J. Chellam, S. B. S. Bama, S. Raju, and V. Abhaikumar, *Classification of Rock Textures*. Berlin, Germany: Springer, 2012, pp. 887–895.

[24] R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-3, no. 6, pp. 610–621, Nov. 1973.

[25] R. Lobos, J. F. Silva, J. M. Ortiz, G. Díaz, and A. Egaña, "Analysis and classification of natural rock textures based on new transform-based features," *Math. Geosci.*, vol. 48, no. 7, pp. 835–870, Oct. 2016.

[26] L. Lepistö, "Rock image classification using color features in Gabor space," *J. Electron. Imag.*, vol. 14, no. 4, Oct. 2005, Art. no. 040503.

[27] F. Bianconi, E. González, A. Fernández, and S. A. Saetta, "Automatic classification of granite tiles through colour and texture features," *Expert Syst. Appl.*, vol. 39, no. 12, pp. 11212–11218, Sep. 2012.
[28] L. B. Gonçalves and F. R. Leta, "Macroscopic rock texture image classification using a hierarchical neuro-fuzzy class method," *Math. Problems Eng.*, vol. 2010, no. 1, Jan. 2010, Art. no. 163635.
[29] A. E. Gelfand, H.-J. Kim, C. F. Sirmans, and S. Banerjee, "Spatial modeling with spatially varying coefficient processes," *J. Amer. Stat. Assoc.*, vol. 98, no. 462, pp. 387–396, Jun. 2003.
[30] J. R. Parker, *Algorithms for Image Processing and Computer Vision*. Hoboken, NJ, USA: Wiley, 2010.
[31] T. M. Cover, *Elements of Information Theory*. Hoboken, NJ, USA: Wiley, 1999.
[32] C. C. Gotlieb and H. E. Kreyszig, "Texture descriptors based on co-occurrence matrices," *Comput. Vis., Graph., Image Process.*, vol. 51, no. 1, pp. 70–86, 1990.
[33] Y. Zhang, M. Li, and S. Han, "Automatic identification and classification in lithology based on deep learning in rock images," *Yanshi Xuebao/Acta Petrologica Sinica*, vol. 34, no. 2, pp. 333–342, 2018.
[34] N. Houshmand, S. GoodFellow, K. Esmaeili, and J. C. Ordóñez Calderón, "Rock type classification based on petrophysical, geochemical, and core imaging data using machine and deep learning techniques," *Appl. Comput. Geosci.*, vol. 16, Dec. 2022, Art. no. 100104.
[35] J. Li, L. Zhang, Z. Wu, Z. Ling, X. Cao, K. Guo, and F. Yan, "Autonomous Martian rock image classification based on transfer deep learning methods," *Earth Sci. Informat.*, vol. 13, no. 3, pp. 951–963, Sep. 2020.
[36] Geosciences Node of NASA's Planetary Data System. (Mar. 1, 2024). *Analyst's Notebook*. [Online]. Available: https://an.rsl.wustl.edu
[37] D. Zheng, H. Zhong, G. Camps-Valls, Z. Cao, X. Ma, B. Mills, X. Hu, M. Hou, and C. Ma, "Explainable deep learning for automatic rock classification," *Comput. Geosci.*, vol. 184, Feb. 2024, Art. no. 105511.
[38] G. Cheng and W. Guo, "Rock images classification by using deep convolution neural network," *J. Phys., Conf.*, vol. 887, Aug. 2017, Art. no. 012089.
[39] C. Pham and H.-S. Shin, "A feasibility study on application of a deep convolutional neural network for automatic rock type classification," *Tunnel Underground Space*, vol. 30, no. 5, pp. 462–472, 2020.
[40] J. Chen, T. Yang, D. Zhang, H. Huang, and Y. Tian, "Deep learning based classification of rock structure of tunnel face," *Geosci. Frontiers*, vol. 12, no. 1, pp. 395–404, Jan. 2021.
[41] H. L. Dawson, O. Dubrule, and C. M. John, "Impact of dataset size and convolutional neural network architecture on transfer learning for carbonate rock classification," *Comput. Geosci.*, vol. 171, Feb. 2023, Art. no. 105284.
[42] R. P. de Lima, A. Bonar, D. Duarte Coronado, K. Marfurt, and C. Nicholson, "Deep convolutional neural networks as a geological image classification tool," *Sedimentary Rec.*, vol. 17, no. 2, pp. 4–9, Jun. 2019.
[43] W. Chen, L. Su, X. Chen, and Z. Huang, "Rock image classification using deep residual neural network with transfer learning," *Frontiers Earth Sci.*, vol. 10, Jan. 2023, Art. no. 1079447.
[44] R. J. Dunham, "Classification of carbonate rocks according to depositional textures," in *Proc. Classification Carbonate Rocks Symp.* Tulsa, OK, USA: American Association of Petroleum Geologists, 1962, pp. 108–121.
[45] C. Wang, Z. Zhang, Y. Zhang, R. Tian, and M. Ding, "GMSRI: A texture-based Martian surface rock image dataset," *Sensors*, vol. 21, no. 16, p. 5410, Aug. 2021.
[46] J. F. Bell, A. Godber, S. McNair, M. A. Caplinger, J. N. Maki, M. T. Lemmon, J. Van Beek, M. C. Malin, D. Wellington, K. M. Kinch, M. B. Madsen, C. Hardgrove, M. A. Ravine, E. Jensen, D. Harker, R. B. Anderson, K. E. Herkenhoff, R. V. Morris, E. Cisneros, and R. G. Deen, "The Mars science laboratory curiosity rover mastcam instruments: Preflight and in-flight calibration, validation, and data archiving," *Earth Space Sci.*, vol. 4, no. 7, pp. 396–452, Jul. 2017.
[47] G. Kylberg. (2011). *The Kylberg Texture Dataset V. 1.0*. Centre for Image Analysis, Swedish University of Agricultural Sciences and Uppsala University, External report (Blue Series). [Online]. Available: http://www.cb.uu.se/~gustaf/texture/
[48] S. Lazebnik, C. Schmid, and J. Ponce, "A sparse texture representation using local affine regions," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 8, pp. 1265–1278, Aug. 2005.
[49] I. Ullah, D. Carrion, S. Escalera, I. M. Guyon, M. Huisman, F. Mohr, J. N. van Rijn, H. Sun, J. Vanschoren, and P. A. Vu, "Meta-Album: Multi-domain meta-dataset for few-shot image classification," in *Proc. Neural Inf. Process. Syst.*, 2022, pp. 1–16.
[50] P. Mallikarjuna, A. T. Targhi, M. Fritz, E. Hayman, B. Caputo, and J.-O. Eklundh, "The KTH-TIPS2 database," *Comput. Vis. Act. Perception Lab., Stockholm, Sweden*, vol. 11, p. 12, Jun. 2006.
[51] Y. Hu, Z. Long, A. Sundaresan, M. Alfarraj, and G. AlRegib, "Commons: Challenging microscopic material surface dataset," IEEE Dataport, 2020, doi: 10.21227/zzsw-3w48.
[52] S. Fekri-Ershad, "A new benchmark dataset for texture image analysis and surface defect detection," 2019, *arXiv:1906.11561*.
[53] P. Brodatz, *Textures: A Photographic Album for Artists and Designers*. New York, NY, USA: Dover, 1966.
[54] A. R. Backes, D. Casanova, and O. M. Bruno, "Color texture analysis based on fractal descriptors," *Pattern Recognit.*, vol. 45, no. 5, pp. 1984–1992, May 2012.
[55] M-Vision and M-Group, *Vision Texture*, Media Lab., MIT, Cambridge, MA, USA, 2002.
[56] T. Ojala, T. Maenpaa, M. Pietikainen, J. Viertola, J. Kyllonen, and S. Huovinen, "Outex—New framework for empirical evaluation of texture analysis algorithms," in *Proc. Object Recognit. Supported User Interact. Service Robots*, 2002, pp. 701–706.
[57] M. Cimpoi, S. Maji, I. Kokkinos, S. Mohamed, and A. Vedaldi, "Describing textures in the wild," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 3606–3613.
[58] R. M. Nosofsky, C. A. Sanders, B. J. Meagher, and B. J. Douglas, "Toward the development of a feature-space representation for a complex natural category domain," *Behav. Res. Methods*, vol. 50, no. 2, pp. 530–556, Apr. 2018.
[59] C. A. Sanders, "Using deep learning to automatically extract psychological representations of complex natural stimuli," Ph.D. thesis, Dept. Psychol. Brain Sci., Indiana Univ., Bloomington, India, 2018.
[60] R. Kwitt and P. Meerwald. (2012). *Salzburg Texture Image Database*. [Online]. Available: http://www.wavelab.at/sources/STex
[61] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *J. Big Data*, vol. 6, no. 1, pp. 1–48, Dec. 2019.
[62] L. Perez and J. Wang, "The effectiveness of data augmentation in image classification using deep learning," 2017, *arXiv:1712.04621*.
[63] A. Mikolajczyk and M. Grochowski, "Data augmentation for improving deep learning in image classification problem," in *Proc. Int. Interdiscipl. PhD Workshop (IIPhDW)*, May 2018, pp. 117–122.
[64] Z. Xu, W. Ma, P. Lin, H. Shi, D. Pan, and T. Liu, "Deep learning of rock images for intelligent lithology identification," *Comput. Geosci.*, vol. 154, Sep. 2021, Art. no. 104799.
[65] Y. Liang, Q. Cui, X. Luo, and Z. Xie, "Research on classification of fine-grained rock images based on deep learning," *Comput. Intell. Neurosci.*, vol. 2021, pp. 1–11, Sep. 2021.
[66] S. Yun, D. Han, S. Chun, S. J. Oh, Y. Yoo, and J. Choe, "CutMix: Regularization strategy to train strong classifiers with localizable features," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 6022–6031.
[67] L. Torrey and J. Shavlik, "Transfer learning," in *Handbook of Research on Machine Learning Applications and Trends: Algorithms, Methods, and Techniques*. Hershey, PA, USA: IGI Global, 2010, pp. 242–264.
[68] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, "ImageNet large scale visual recognition challenge," *Int. J. Comput. Vis.*, vol. 115, no. 3, pp. 211–252, Dec. 2015.
[69] B. Kromydas. (Mar. 2023). *Unlock the Power of Fine-Tuning Pre-Trained Models in Tensorflow and Keras*. Accessed: Jul. 1, 2023. [Online]. Available: https://learnopencv.com/fine-tuning-pre-trained-models-tensorflow-keras/
[70] H. Wang, Z. Wang, M. Du, F. Yang, Z. Zhang, S. Ding, P. Mardziel, and X. Hu, "Score-CAM: Score-weighted visual explanations for convolutional neural networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2020, pp. 111–119.
[71] L. Alzubaidi, J. Zhang, A. J. Humaidi, A. Al-Dujaili, Y. Duan, O. Al-Shamma, J. Santamaría, M. A. Fadhel, M. Al-Amidie, and L. Farhan, "Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions," *J. Big Data*, vol. 8, no. 1, pp. 1–74, Mar. 2021.
[72] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 25, 2012, pp. 1–9.
[73] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2818–2826.

[74] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2261–2269.
[75] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted residuals and linear bottlenecks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 4510–4520.
[76] M. Tan and Q. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in *Proc. Int. Conf. Mach. Learn.*, 2019, pp. 6105–6114.

DIPENDRA J. MANDAL received the B.E. and M.S. degrees in electrical and electronic engineering from Kathmandu University, Nepal, in 2010 and 2017, respectively, and the Ph.D. degree in computer science from Norwegian University of Science and Technology (NTNU), Norway, in 2023, under the Marie Curie Fellowship (CHANGE-ITN).

He was an Exchange Student under the Erasmus Mundus Program, University Lumiere Lyon 2, France, from 2017 to 2018. He is currently a Postdoctoral Fellow with Colorlab, Department of Computer Science, NTNU. His research interests include image processing, spectral imaging, cultural heritage digitization, image quality, texture analysis, and machine learning.

Dr. Mandal is a member of the Society for Imaging Science and Technology (IS&T).

HILDA DEBORAH received the B.Sc. degree in computer science from the University of Indonesia, in 2010, the joint M.Sc. degree from the Erasmus Mundus Color in Informatics and Media Technology Master's Program, University Saint Étienne, France, University of Granada, Spain, and Gjøvik University College (now NTNU—Norwegian University of Science and Technology), Norway, in 2013, the Ph.D. degree in signal and image processing from the University of Poitiers, France, through the Co-Tutelle Program, and the Ph.D. degree in computer science from NTNU, in 2016.

From 2016 to 2018, she was a Researcher with Colourlab, Department of Computer Science, NTNU. From 2018 to 2020, she was a Marie Curie Postdoctoral Fellow through the Marie Curie co-funding of the FRICON Mobility Program by the Research Council of Norway, with mobility with the University of Iceland, Reykjavik. Since 2020, she has been a Senior Researcher with the Colourlab, Department of Computer Science, NTNU. Her research interests include fundamental image processing, texture analysis and perception, mathematical morphology, and spectral imaging.

Dr. Deborah was a recipient of the Outstanding Academic Fellows Program, NTNU, from 2022 to 2026. She is a Board Member of NOBIM, a Norwegian professional forum for image processing and machine learning, and a member of the Society for Imaging Science and Technology (IS&T).

TABITA L. TOBING received the B.Sc. degree in physics from the University of Indonesia, in 2016, and the M.Sc. degree in electrical engineering from the Institute Technology of Bandung, Indonesia, in 2021. She is currently pursuing the Ph.D. degree in information security and communication technology with Norwegian University of Science and Technology, Norway, under the context of the Lying Pen of Scribes Project (ToppForsk, Research Council of Norway).

Her research interests include image processing, document analysis and recognition, and machine learning.

MATEUSZ JANISZEWSKI received the B.Sc. degree in mining engineering and geology from Wroclaw University of Technology, Poland, in 2012, the joint M.Sc. (Tech.) degree in minerals and environmental engineering from Aalto University, Finland, and Delft University of Technology, The Netherlands, in 2014, and the D.Sc.(Tech.) degree in geoengineering from Aalto University, in 2019.

From 2019 to 2022, he was a Postdoctoral Researcher with the Mineral-Based Materials and Mechanics Group, Aalto University. From 2022 to 2023, he was with the Structures and Architecture Research Group. He is currently an University Lecturer with the Civil Engineering Department, Aalto University. He is also a Rock Mechanics Specialist with Fractuscan Ltd. His current research interests include applying photogrammetry and extended reality technologies in rock engineering, mining, and geology, both from the engineering and educational perspectives.

Dr. Janiszewski was a recipient of the Teaching Achievement Award, in 2019, awarded by the Dean of School of Engineering, Aalto University. He is a member of the International Society for Rock Mechanics and Rock Engineering (ISRM) and the Federation of European Mineral Programs (FEMP).

JAMES W. TANAKA received the Ph.D. degree in psychology from the University of Oregon, in 1989.

He was a NIH Postdoctoral Fellow with Carnegie Mellon University. He was an Assistant Professor and then an Associate Professor with Oberlin College, Oberlin, OH, USA from 1993 to 2003. He is currently a Professor in psychology with the Cognition and Brain Sciences Program, University of Victoria, BC, Canada. His research interest includes how experience influences the way we perceive objects in the world. He has studied the perceptual processes involved in expert object recognition and different kinds of experts including birdwatchers, dog judges, radiologists, and geologists. His work has investigated the cognitive and neural processes of face recognition, a domain in which we are all experts. With colleagues with the Yale Child Study Centre, he developed the Let's Face It! software, a program designed to improve the face processing skills of children with autism. He is also developing an intervention to help people recognize other-race faces.

Dr. Tanaka is a fellow of the Association for Psychological Science and an elected member of the Royal Society of Canada, Canada's national academy for the arts, humanities, and sciences.

ANNA LAWRANCE is currently pursuing the B.Sc. degree (Hons.) in psychology with the University of Victoria, BC, Canada.

Since 2022, she has been conducting research with Dr. James Tanaka with the University of Victoria. As a Research Associate of the Different Minds Laboratory, she has examined perceived categorical structure using PsiZ, a psychological embedding technique developed by Dr. Brett Roads. Much of her work probes how conceptual knowledge and relevant perceptual experiences contribute to category restructuring.

Mrs. Lawrance's research has been recognized by the Natural Sciences and Engineering Research Council of Canada (NSERC) through multiple Undergraduate Student Research Awards, in addition to receiving high-academic-honours scholarships.