Automated Detection of Recent Mud Extrusions Using UAV Imagery and Deep Learning: A Comparative Analysis of Traditional and CNN-Based Approaches

Massimiliano Guastella [1,2], Antonino Pisciotta [3], Raffaele Martorana [2], Antonino D'Alessandro [3]

[1] Dept. of Civil, Construction and Environmental Engineering, Sapienza University, Rome, Italy - dicea@cert.uniroma1.it
[2] Dept. of Earth and Marine Sciences, University of Palermo, Palermo, Italy - dipartimento.distem@unipa.it
[3] National Institute of Geophysics and Volcanology (INGV), Rome, Italy - aoo.roma@pec.ingv.it

Keywords: Mud Volcano, UAV Imagery, Convolutional Neural Networks, Transfer Learning, Data Augmentation, Image Classification

Abstract

Mud volcanoes are geological formations resulting from the expulsion of mud, gases, and fluids from deep underground. Monitoring these formations provides critical insights into subsurface processes and geological hazards. This study focuses on detecting recent mud extrusions in mud volcano environments using high-resolution aerial imagery acquired by unmanned aerial vehicles (UAVs). Using UAV-based surveys instead of satellite imagery, we obtain finer spatial detail suitable for identifying subtle textural and chromatic variations in relatively small sites. A binary image classification pipeline was developed to distinguish recent mud from non-mud areas. Traditional machine learning algorithms, including Support Vector Machine (SVM), Random Forest, and Extreme Gradient Boosting (XGBoost), were compared with deep learning architectures such as Convolutional Neural Networks (CNNs), both leveraging transfer learning and custom models. Traditional algorithms rely on handcrafted features, while CNNs learn hierarchical representations directly from raw data. Feature extraction methods were selected based on their ability to distinguish between the two designated classes effectively. To enhance model robustness and generalization, a designed augmentation pipeline was applied before each training epoch or cross-validation fold. This strategy introduced controlled and random variations to simulate real-world imaging conditions, such as varying viewpoints and lighting, ensuring the models generalization, moreover it also minimized data leakage by presenting distinct image variations throughout training. CNNs achieved the highest accuracy, outperforming traditional algorithms and demonstrating the advantages of combining deep learning with effective data augmentation. These findings underscore the potential of CNNs for accurate and efficient monitoring of dynamic geological environments.

1. Introduction

Mud volcanoes are geological structures formed by pseudo-volcanic phenomena caused by over-pressured multiphase pore fluids, generally high-salinity water and methane gas, trapped in sedimentary basins by an impermeable top layer of rock. (Napoli et al. 2020). In Sicily, mud volcanoes are quite widespread both onshore and offshore (Etiope et al., 2002; Cangemi and Madonia, 2014). The main vents are located in the central-southern part of the island, with the exception of the groups located on the southwest flank of Mt Etna volcano, the main of which is named "Salinelle di Paternò" (Napoli et al., 2020). Understanding and monitoring recent mud extrusions are essential for several geoscientific and environmental objectives, including evaluating geological hazards, supporting environmental management, and informing hydrocarbon exploration strategies (Martinelli and Judd, 2004). Reliable and timely detection of fresh mud deposits can help researchers and policymakers assess risks, implement safety measures, and make informed decisions regarding resource development. Conventional methods for monitoring mud volcanoes have relied heavily on field surveys and satellite imagery analysis. While these approaches can yield valuable insights, they often involve certain limitations. Field surveys, for instance, are time-consuming, labour-intensive, and can pose logistical challenges, especially in remote or hazardous locations. Satellite imagery, although useful for large-scale observations, frequently lacks the spatial resolution needed to detect subtle features or small-scale changes in mud volcano environments (Laliberte et al., 2011). High spatial resolution is particularly crucial for identifying recent mud extrusions in relatively small and structurally complex sites, where minor textural differences between fresh mud and non-mud areas can be difficult to discern with coarse-scale imagery. To overcome these limitations, advancements in unmanned aerial vehicles (UAVs) and remote sensing technologies have provided new opportunities for high-resolution data acquisition (Colomina and Molina, 2014; Nex and Remondino, 2014). Of the revolutionized fields, topographic surveying is prominent because many low cost UAVs with on/board light weight optical payloads often deliver mapping products such as ortho-photos with centimetre level accuracy that had been exclusively bounded to the expensive field surveying methods earlier. (Perera and Nalani, 2022). UAVs can be deployed at low altitudes and flexible flight plans, enabling the capture of detailed aerial imagery that can reveal fine-scale geological features and variations both in texture and color. Nevertheless, the images taken during the surveys can provide valuable dataset to be used in computer vision tasks other than just for digital elevation model or orthomosaic. Machine learning and deep learning methods have shown remarkable potential in a variety of remote sensing and environmental monitoring applications, particularly in tasks like land use and land cover classification, object detection, and scene interpretation (Zhao et al., 2017; LeCun et al., 2015). Traditional machine learning algorithms, such as Support Vector Machines (SVM) (Cortes & Vapnik, 1995) and ensemble method including Random Forest (Breiman, 2001), and Extreme Gradient Boosting (XGBoost) (Chen & Guestrin, 2016), have long been employed in classification tasks involving handcrafted features. Such feature are extracted through methods that emphasizes edges, textures, frequencies or chromatic distributions, moreover their combination may provide inputs that can improve the performance of classifiers. However, traditional machine learning techniques face certain challenges. The design and selection of handcrafted features is task-dependent and may not capture all the nuances required for accurate classification, especially in a complex geological context like mud volcano

environments. In contrast, deep learning methods, particularly Convolutional Neural Networks (CNNs), learn features directly from raw pixel data, enabling them to discover hierarchical representations that can better capture subtle patterns (Krizhevsky et al., 2012; LeCun et al., 2015). As a result, CNN-based approaches have often outperformed traditional methods in various computer vision tasks, including those in remote sensing and environmental analysis.

2. Materials and Methods

2.1 Dataset

The dataset was constructed based on the INGV archive of drone surveys conducted around three selected mud volcano areas: Salinelle of Paternò, Maccalube of Aragona and Maccalube of Santa Barbara. Drone surveys for topographic purposes typically recommend an 80% front overlap and 70% side overlap. To minimize redundancy and ensure diversity among the instances, the images were manually searched, resulting in a total of 830 images being selected for inclusion in the study (Paternò 280 images: Mud 152, No Mud 128; Aragona 277 images: Mud 135, No Mud 142); Santa Barbara 273 images: Mud 128, No Mud 145). Each of the three mud volcano sites was chosen for its unique environmental context. Notably, the Aragona and Santa Barbara sites are situated far from inhabited areas, thereby eliminating the presence of buildings which could potentially confound the classification process due to the common gray hue shared between constructions and recent mud flows. In contrast, the Paternò site, which is embedded within urban settings, feature residential buildings and other constructions. The Paternò site, for example, is located within a stadium and surrounded by houses. Including images from these environments was a strategic choice aimed at assessing the model's performance in complex scenarios, with the ultimate goal of developing an autonomous system capable of accurate detection, despite potential challenges. The DJI Phantom 4 drone was employed to capture nadiral images using a 12.4-megapixel RGB camera equipped with a 1/2.3-inch sensor. To optimize data acquisition and enhance the accuracy of terrain reconstruction, a photogrammetric survey was conducted with a frontal overlap of 85% and a lateral overlap of 75%.

2.2 Feature Extraction Method

A critical step in the methodology was the comparative analysis of different feature extraction methods. While deep learning models can learn feature representations automatically, traditional machine learning algorithms require handcrafted features derived from the imagery. The feature extraction methods including Histogram of Oriented Gradients (HOG) (Dalal and Triggs, 2005), Gabor filters (Gabor, 1946; Granlund, 1978),Color histograms (Swain and Ballard, 1991), Local Binary Patterns (LBP) (Ojala et al. 1996, 2002), Scale-Invariant Feature Transform (SIFT) (Lowe 1999, 2004), Speeded-Up Robust Features (SURF) (Bay et al, 2008), Oriented FAST and Rotated BRIEF (ORB) (Rublee et al., 2011) and Canny Edge (Canny, 1986) were tested in order to evaluate their performance and the capabilities to discriminate between these two classes effectively. HOG captures local shape and texture by analyzing the distribution of intensity gradients, making it ideal for detecting nuanced patterns in mud flow textures. Gabor filters, renowned for their proficiency in texture analysis, are instrumental in identifying the fine details in surface textures due to their sensitivity to orientation and scale. Color histograms provide a robust analysis of color distribution, which varies significantly between fresh mud and its surrounding terrain, thereby aiding in distinguishing recent volcanic activity. Lastly, LBP is employed for its ability in texture classification, capturing local texture patterns that are prevalent in areas of recent extrusion versus older, settled formations. The parameters for each feature extraction method included are showed in table 1. The exclusion of SIFT, SURF, ORB, and Canny edge detection from our methodology was driven by their limited applicability to the specific textural and colorimetric nuances of mud volcanoes. While these techniques are highly effective in general image analysis applications, they lack the discriminative power necessary for accurately classifying the unique geological features of mud volcanoes. Specifically, keypoint-based methods such as SIFT, SURF, and ORB identified only a few points of interest in areas covered by recent mud, with the majority of keypoints detected in zones without recent extrusions. Additionally, the Canny edge detector was adept at recognizing the cracked textures typical of dried mud; however, such features are also present in the 'recent mud' class images in areas not affected by recent extrusions. Given these observations, we decided to omit these methods from our feature vector construction for traditional machine learning algorithms, as they did not contribute effectively to distinguishing between our designed classes.

2.2.1 Color Histograms: A color histogram is a graphical representation that illustrates the distribution of colors within an image (Swain and Ballard, 1991). It operates by dividing the image's color space into discrete bins and counting the number of pixels that fall into each bin. This process results in a histogram where the x-axis represents the color bins, and the y-axis indicates the frequency of pixels in each bin. Color histograms are versatile and can be constructed for various color spaces, as a matter of fact are widely used in image processing tasks, such as image retrieval, segmentation, and enhancement. Color histograms are effective tools in geological studies for classifying rock images. By analyzing color distributions and edge features, they enable precise differentiation between various rock types (Joseph et al. 2017). Combining color histograms with statistical and frequency-based methods further enhances the extraction of visual textural and colorimetric features (Vangah et al., 2023).

2.2.2 Histograms of Oriented Gradients (HOG): The histogram of oriented gradients (Dalal and Triggs, 2005) is a feature descriptor employed in computer vision and image processing for object detection and image classification tasks (Leonardis et al. 2006). This method shares similarities with edge orientation histograms, scale-invariant feature transform (SIFT) descriptors, and shape contexts. However, it differs by being calculated on a dense grid of uniformly spaced cells and incorporating overlapping local contrast normalization to enhance accuracy. The HOG method operates on the premise that the local appearance and shape of an object in an image can be effectively described by the distribution of local intensity gradients or edge orientations, which are inherently perpendicular to the gradient's direction. Hog was widely used in computer vision tasks in particular for hand gesture recognition (Freeman and Roth, 1994), human detection (Zhu et al, 2006), sketches for searching and indexing digital image libraries (Hu et al, 2010), interstellar molecular formation (Soler et al, 2019).

2.2.3 Local Binary Pattern (LBP): Local Binary Pattern (Ojala et al. 1996, 2002) is a robust texture descriptor that has gained prominence in image analysis due to its discriminative power and computational simplicity. LBP effectively captures local spatial patterns by comparing each pixel's intensity with its neighbors, encoding this relationship into a binary number. Specifically, it describes the pixels of an image by using a 3x3 neighbourhood area around each pixel. The central pixel subtracted from its eight neighbours. If the resulting value is negative, the pixel is set to '0', otherwise it is set to '1' which concatenate together to give an 8-bits code corresponding an interger ranging from 0 to 255 (Lizé et al, 2020) In geological and geophysical applications, LBP has been instrumental in enhancing the analysis and classification of complex textures inherent in geological formations, also in combination with color descriptors (Long et al., 2019; Vangah et al., 2023).

2.2.4 Gabor Filter: Gabor filtering (Gabor, 1946; Granlund, 1978) is a widely adopted computer vision technique for texture analysis. Gabor filters perform a local Fourier analysis and are essentially sine and cosine functions modulated by a Gaussian window (Idrissa & Acheroy, 2002). Their key properties include invariance to illumination, rotation, scale, and translation, which make them highly versatile in various applications. These characteristics are directly controlled by the parameters of the Gabor filters themselves (Kamarainen et al., 2006). However, this flexibility comes with a drawback: the high number of parameters that need to be carefully tuned, such as frequency, orientation, and the width of the Gaussian envelope, which can complicate optimization and increase computational costs (Bianconi & Fernández, 2007). Despite this challenge, the ability of Gabor filters to capture multi-scale and multi-orientation information makes them invaluable in domains such as texture classification, edge detection, and feature extraction in remote sensing and medical imaging.

2.2.5 Scale-Invariant Feature Transform (SIFT): The Scale-Invariant Feature Transform (Lowe 1999, 2004) is a computer vision algorithm, notable for its ability to robustly detect and describe local features invariant to scale, rotation, and moderate affine distortion. At its core, SIFT operates by convolving an image with Gaussian filters at multiple scales, then identifying extrema in the resulting difference of gaussian images. Each extremum undergoes a detailed characterization process that estimates its precise location, scale, and orientation, generating a set of highly distinctive keypoints. The orientation assignment is derived from the local gradient distribution, which allows the descriptor to maintain rotational invariance. These keypoints are then encoded into a signature (the SIFT descriptor), which captures the gradient magnitudes and orientations in a region around each keypoint; this descriptor is both discriminative and robust to photometric changes, making SIFT highly effective for tasks such as object recognition, image stitching, and 3D scene reconstruction (Hang Zhu et al 2022). Beyond its traditional domains, SIFT and its variants has also been applied to a variety of geological and geophysical problems, where the detection of scale and rotation-invariant features is critical for analyzing remote sensing data or high-resolution imagery of Earth's subsurface and surface structures (Rong, 2024; Yu, 2013).

2.2.6 Speeded-Up Robust Feature (SURF): The Speeded-Up Robust Features (Bay et al, 2008) algorithm is a fast and efficient method for detecting and describing keypoints in images, offering scale and rotation invariance. The SURF feature point detector utilizes a Hessian matrix approach, approximating the Laplacian of Gaussian with a difference of Gaussian. Key points are located as maxima in this determinant across scales, ensuring robustness to varying image sizes. Unlike SIFT, which uses Gaussian filters, SURF leverages box filters for speed, and its descriptors are simpler, resulting in faster computations. SURF assigns an orientation to each key point by analysing Haar wavelet responses in the surrounding area, enabling rotation invariance. The local neighbourhood of each key point is then divided into grids, and wavelet responses are used to create a compact descriptor vector for matching. Its speed and robustness make it ideal for applications like image matching, object recognition, and remote sensing, particularly when computational efficiency is critical. The advantages of the SURF algorithm find place in real-time UAV control systems, where being faster is a key aspect for such systems (Wang et al. 2021).

2.2.7 Oriented FAST and Rotated BRIEF (ORB): The Oriented FAST and Rotated BRIEF algorithm (Rublee et al., 2011) is a highly efficient feature descriptor, developed as an alternative to SIFT and SURF, with the goal of achieving comparable performance but at a significantly reduced computational cost. The algorithm builds upon the FAST keypoint detector (Rosten and Drummond, 2006) and the BRIEF descriptor (Calonder et al., 2010), introducing robustness to rotational variations and reducing susceptibility to noise. ORB enhances the FAST detector by incorporating an orientation component, which is computed using the intensity centroid method. This method evaluates the asymmetry of pixel intensities around the corner, providing a consistent orientation estimate. Additionally, ORB modifies BRIEF (creating rBRIEF) by adding a learning step to select the most uncorrelated binary tests, optimizing the descriptor for distinctiveness and efficiency. This combination enables ORB to perform real-time feature matching on low-power devices, achieving results that are on par with SIFT in terms of accuracy while being almost two orders of magnitude faster in execution. The ORB algorithm uses Fast to detect feature points and Brief to compute the descriptors of feature points, and its feature point performance is between SIFT and SURF, but it runs much faster than SURF (Qinjun et al., 2022). It is used in similar computer vision tasks as SIFT and SURF. Rarely employed in the geologic context unless for specific task involving both autonomous mapping of geologic features (Chen et al., 2021).

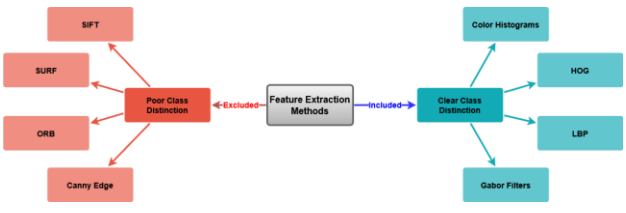

Figure 1. Feature extraction methods with poor class distinction were excluded, while those with clear separation were included.

2.2.8 Canny Edge Detector: The Canny edge detector is a feature detector in image processing, designed to identify edges by optimizing detection, localization, and minimizing multiple responses to a single edge (Canny, 1986). Edge detection is an essential technology for obtaining the edges of remote sensing images (Huang et al. 2017,) and the role it plays is of paramount importance in numerous Earth observation applications, and its extensive utilization can be observed in domains such as national defense and security, land use, urban planning, and geographic image retrieval, among others (Zhou et al., 2024, Cheng et al., 2020, Cheng et al., 2017). The process starts by applying a Gaussian filter to the image to reduce noise, which could interfere with the detection of edges. Then, the gradient magnitude and direction at each pixel are calculated using derivative filters, like the Sobel operator, to detect regions with significant intensity transitions. Non-maximum suppression is applied next, which eliminates non-maximum values in the gradient direction to refine the edges, retaining only the most important ones. Lastly, hysteresis thresholding is used: pixels with a gradient magnitude higher than a set high threshold are considered strong edges, while those below a low threshold are disregarded. Pixels that fall between the two thresholds are considered weak edges and are preserved only if they are connected to strong edges, which helps in maintaining the true edges while minimizing false positives.

2.3 Traditional Machine Learning Algorithms

2.3.1 Extreme Gradient Boosting (XGBoost): XGBoost, short for Extreme Gradient Boosting, is a highly efficient and scalable implementation of gradient boosting algorithms. At its core, gradient boosting combines multiple weak learners, decision trees, into a strong ensemble model by iteratively adding new trees that correct the residual errors of the previous ones. XGBoost distinguishes itself from traditional gradient boosting methods by incorporating a range of optimizations designed to improve both computational speed and predictive performance. One of its key innovations is the use of a sparsity-aware split finding algorithm, which leverages efficient data structures to handle missing values and sparse data in a more effective manner. Moreover, it employs an advanced tree learning approach called the weighted quantile sketch, enabling the algorithm to handle large datasets with high dimensionality (Chen & Guestrin, 2016). These optimizations make XGBoost particularly popular for large-scale machine learning tasks, where it balances speed, accuracy, and memory efficiency. Another critical aspect of XGBoost is its built-in regularization mechanisms, which help combat overfitting by penalizing the complexity of individual trees. The algorithm introduces two regularization parameters: one that controls the L2 norm of the leaf weights and another that controls the depth and structure of each tree. By carefully tuning these parameters, practitioners can strike a balance between model complexity and generalization performance, leading to robust and reliable predictive outcomes. XGBoost also provides flexibility by supporting various objective functions, ranging from regression and classification to ranking, and offers parallelization capabilities that take advantage of modern hardware architectures. Thanks to these features, XGBoost has become a go-to method in a wide array of industries, including finance, healthcare, and e-commerce, where the capacity to efficiently handle massive datasets without sacrificing accuracy is crucial.

2.3.2 Random Forest: Random Forest is a versatile and powerful ensemble learning method primarily deployed for classification and regression tasks. Developed by Breiman (2001), it operates by constructing a multitude of decision trees during the training phase and combining their predictions, typically through a majority vote or an average in the case of regression. This approach, known as bootstrap aggregation or "bagging," involves drawing multiple bootstrap samples from the training dataset and fitting a separate decision tree to each sample. By aggregating the outputs of numerous, slightly varied trees, Random Forest effectively reduces the variance of single decision-tree models, thereby improving predictive accuracy and robustness against overfitting. In addition to bagging, Random Forest introduces randomness in the feature selection process when determining splits. Rather than considering all available features at each node, the algorithm selects a random subset of features, thereby reducing the correlation between individual trees. This feature randomness further bolsters the model's capacity to generalize, making Random Forest well-suited for datasets with high dimensionality or complex feature relationships. Another practical strength lies in its ability to provide estimates of feature importance by measuring how each feature contributes to decreasing the impurity in the nodes of the ensemble trees. Consequently, Random Forest not only offers strong predictive performance but also yields valuable insights into the underlying structure and relevance of the input variables. These characteristics have made it a ubiquitous tool in applications ranging from biology and finance to computer vision and recommender systems.

2.3.3 Support Vector Machine (SVM): Support Vector Machines (SVMs) are supervised learning models widely used for both classification and regression problems, though they are perhaps most famous for high-performance classification tasks. The fundamental principle behind SVMs is to find an optimal hyperplane or set of hyperplanes in higher dimensions that maximizes the margin between the data points of different classes (Cortes & Vapnik, 1995). By focusing on the data points that lie closest to the decision boundary (known as support vectors), SVMs seek to create a robust separation that not only classifies current data accurately but also generalizes well to unseen data. This margin-maximization approach is critical to reducing overfitting, as it emphasizes the most challenging training examples. A notable strength of SVMs lies in their use of kernel functions to address non-linearly separable data. Kernels allow the model to project input data into higher-dimensional feature spaces, where a linear separating hyperplane may exist even if one does not in the original input space. Commonly used kernel functions include the linear kernel, polynomial kernel, and the radial basis function (RBF) kernel, each designed to capture different forms of data complexity. This flexibility makes SVMs particularly adept at handling intricate and high-dimensional datasets. However, they can be computationally expensive for very large datasets, and the selection of the optimal kernel and associated hyperparameters often requires careful tuning. Despite these challenges, SVMs remain a cornerstone of machine learning, prized for their theoretical foundations, strong empirical performance, and elegant mathematical framework.

2.4 Convolutional Neural Networks (CNNs)

Convolutional Neural Networks (CNNs) have emerged as a cornerstone in image recognition and classification tasks due to their ability to automatically learn spatial hierarchies of features from input images. This characteristic makes CNNs particularly effective in domains requiring high-level abstraction of visual

patterns (Krizhevsky et al., 2012). This kind of network operate by hierarchically extracting features from grid-like data, such as images, by using a combination of convolutional, activation, pooling, and fully connected layers. Convolutional layers play a central role by employing small, learnable filters that convolve across the input data, detecting spatially local patterns such as edges, textures, or gradients, which are vital for constructing hierarchical feature maps (LeCun et al., 1998; Krizhevsky et al., 2012). The non-linear activation functions, commonly ReLU, introduce essential non-linearity, enabling the network to model complex interactions and representational capabilities (Glorot et al., 2010). Pooling layers follow the convolutional layers to progressively reduce the spatial dimensions of feature maps, preserving the most critical features while reducing computational complexity, thus achieving translational invariance (Scherer et al., 2010). As the network deepens, higher-order layers aggregate these features into more abstract representations, which culminate in fully connected layers that combine all extracted features for classification or regression tasks (Krizhevsky et al., 2012; Simonyan & Zisserman, 2015).
The training of CNNs employs backpropagation and stochastic gradient descent, iteratively adjusting the weights to minimize the loss function. This powerful approach has demonstrated state-of-the-art performance across diverse domains, including image recognition, object detection, and medical imaging (He et al., 2016). Unlike traditional machine learning methods, which depend heavily on manual feature extraction, CNNs utilize convolutional layers to extract features hierarchically, capturing both local and global patterns (LeCun et al., 1998). One particularly notable architecture component is the Inception block, introduced by Szegedy et al. (2015), which employs parallel convolutional paths of different filter sizes and pooling operations to capture multi-scale information within the same layer. This design, sometimes referred to as the GoogLeNet module, effectively increases the width of the network while maintaining efficient computation, thus improving the representational power of CNNs (Szegedy et al., 2015).

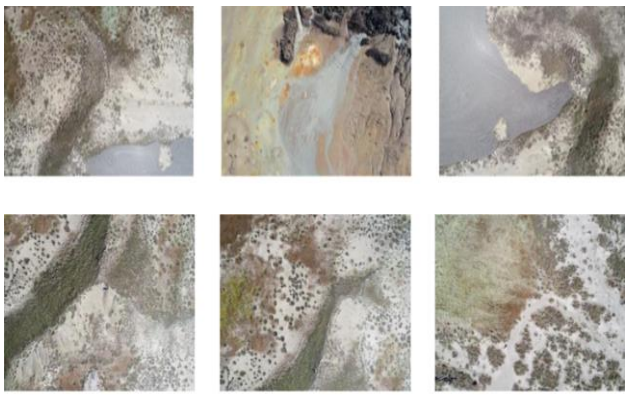

Figure 2. Representative samples from the two classes in the dataset. The first row contains images of recent mud extrusion, while the second row shows areas without recent mud deposits. These aerial images, captured by drone, highlight the visual differences used for classification in the study.

2.4.1 Transfer Learning (ResNet5050, VGG16, EfficientNet): Transfer learning, an extension of deep learning, leverages pre-trained models on large datasets, such as ImageNet, to improve the performance of models on domain-specific tasks with limited data (Pan and Yang, 2010). By reusing learned weights from pre-trained networks, transfer learning accelerates convergence and reduces the risk of overfitting in tasks where data scarcity is a challenge (Zhan et al., 2020). The practical implementation of CNNs often involves architectures designed to balance computational efficiency with feature extraction capabilities. For instance, residual networks (ResNet5050) address the vanishing gradient problem by introducing skip connections, enabling the training of very deep networks without degradation of performance (He et al., 2016). In the context of balancing computational efficiency with feature extraction capabilities, the VGG16 architecture small 3×3 filters to capture intricate features while maintaining manageable computational complexity (Simonyan and Zisserman, 2015). Similarly, architectures like EfficientNet employ compound scaling to optimize model size, accuracy, and computational cost (Tan and Le, 2019). These advancements underscore the adaptability and robustness of CNNs in tackling a wide array of image classification challenges. Transfer learning has proven to be invaluable in applications where domain-specific labeled datasets are scarce. By fine-tuning pre-trained CNN models on smaller datasets, is it possible to achieve state-of-the-art results with significantly reduced training time. This approach has been successfully applied in diverse fields, including medical imaging, remote sensing, and environmental monitoring (Yang et al., 2020). For instance, in UAV-based image classification tasks, transfer learning enables the adaptation of general-purpose visual features to specific contexts, such as detecting recent mud deposits in mud volcanoes. Nevertheless the importance of selecting appropriate models for practical applications is mandatory to obatin good performances (Zhuang et al. 2021). In our study all layers except the classification layer were frozen, while the classification head was replaced with a fully connected layer for binary classification.

2.4.2 Custom CNN Models: Parallel to the transfer learning experiments, four custom CNN architectures were trained from scratch: CNN_2, CNN_3, CNN_4, and Inception_CNN, were developed and trained from scratch to assess their performance in the specific context of mud volcano detection. Unlike the more complex networks such as ResNet50, VGG16, and EfficientNet, these models employ a simplified design aimed at capturing essential spatial and textural features with reduced computational complexity. CNN_2 provides a baseline with two convolutional layers employing ReLU activations and max pooling to extract basic features, while CNN_3 deepens the architecture through additional convolutional blocks with batch normalization to enhance stability and feature representation. CNN_4 refines this approach further by adding an extra convolutional layer and substituting standard ReLU with Leaky ReLU, thereby maintaining gradient flow in deeper layers. In contrast, the Inception_CNN incorporates inception-style modules that perform parallel convolutions with multiple kernel sizes (1×1, 3×3, and 5×5 convolutions), enabling simultaneous extraction of both fine and coarse details. This systematic progression in architectural design underscores the potential of tailored, less complex CNNs to effectively capture the subtle color and texture variations characteristic of fresh mud deposits in UAV imagery.

2.5 Data Augmentation

Data augmentation is a fundamental strategy to enhance the performance of deep learning models by artificially expanding the size of training datasets. It introduces variability to the data, helping models generalize better and reducing the risk of overfitting. This technique has been demonstrated to improve classification accuracy significantly, particularly in cases of imbalanced or small datasets, as it creates synthetic data variations while preserving the original data's integrity (Shorten & Khoshgoftaar, 2019). In satellite imagery applications, data augmentation is crucial as the variability in lighting, angles, and atmospheric conditions can influence model predictions, necessitating transformations that maintain physical plausibility (Buslaev et al., 2020). Augmentation techniques can range from basic transformations, such as rotation, flipping, and scaling, to advanced approaches like domain-specific modifications or even the synthesis of entirely new images. For example, geometric and photometric augmentations are widely used to simulate variations observed in real-world scenarios (Zhong et al., 2020). Moreover, studies suggest that augmentation methods tailored to specific data domains, can significantly improve model robustness and accuracy (Perez & Wang, 2017). In this study a pipeline of random augmentation was employed before each training epoch or cross validation fold. Specifically, each image has a 50% probability of entering an augmentation stage. If selected, a series of operations are applied, each with a 50% chance of execution, ensuring at least one transformation per image and potentially multiple transformations in a single pass. Below are reported the details of augmentation steps:

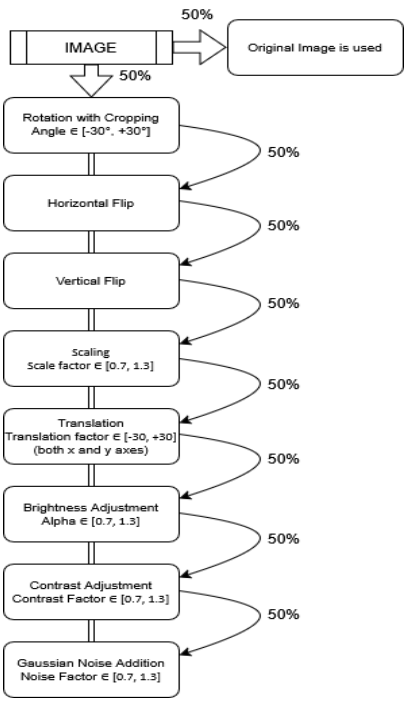

Figure 3. The diagram illustrates the augmentation pipeline applied for training the classification model. Each input image has a 50% probability of entering the pipeline. In that case at least one of the transformations is applied.

2.6 Training Parameter and Optimization Process

For traditional machine learning models, training employed a stratified K-Fold cross-validation with five splits to maintain class distribution consistency across folds. In stratified K-Fold cross-validation, the dataset is divided into *k* folds so that each fold has nearly the same percentage of minority and majority class samples as the entire set (Szeghalmy and Fazekas, 2023). Hyperparameter optimization relied on Bayesian search, running 30 iterations per fold to identify optimal model parameters. Most machine learning models have hyperparameters that require tuning via black-box optimization (Wu et al., 2020), and these kinds of optimization problems are often solved through Bayesian Optimization (Frazier, 2018). This approach uses a probabilistic surrogate model for the objective function to determine the most promising next evaluation point, where a popular criterion is expected improvement (Wu et al., 2020; Jones et al., 1998). Bayesian search optimization surpasses random and grid search by leveraging probabilistic models to prioritize promising regions in hyperparameter space, minimizing the number of evaluations required to find optimal configurations. Unlike exhaustive or purely random exploration, it focuses on regions most likely to contain ideal hyperparameters. The dataset was partitioned 70%–30% for training and validation. In deep learning models, training used the cross-entropy loss function to address binary classification tasks. Cross-entropy coincides with logistic loss when the softmax is used (Mao et al., 2023), measuring classification performance based on predicted probabilities between 0 and 1. All training employed the Adam optimizer with a learning rate of 0.001 and a weight decay of 10^{-4}. A step-based scheduler reduced the learning rate by an order of magnitude every seven epochs. The batch size was 64, and training was capped at 100 epochs, with an early stopping criterion (patience 20) triggered if validation performance failed to improve for several consecutive epochs. Adam is a first-order gradient-based optimization algorithm that computes adaptive learning rates for each parameter by estimating the first and second moments of gradients. It is computationally efficient, memory-efficient, invariant to gradient rescaling, and effective for large-scale problems with noisy or sparse gradients. By combining the strengths of AdaGrad for sparse gradients and RMSProp for non-stationary objectives, Adam employs bias-corrected moment estimates to ensure robust updates (Kingma and Ba, 2017). Batch normalization layers followed each convolution in deeper architectures to stabilize the learning process, and dropout (set to 0.5) was introduced to mitigate overfitting. Max-pooling operations were also placed after convolutional sequences to reduce spatial dimensions. A similar 70%–30% split was employed for transfer learning experiments. In this setup, all layers of the pre-trained models were frozen except for the classification layer, which was replaced by a fully connected layer configured for binary classification. This arrangement leveraged the pre-trained feature representations while refining only the task-specific output layer.

3. Results

Three pretrained architectures were initially explored (ResNet50, VGG16, and EfficientNet), each fine-tuned on the binary dataset by freezing the convolutional blocks of the original models and replacing the final classification layer. Data augmentation implemented randomly before each training epoch was integral in enhancing generalization and mitigating overfitting, as a matter of fact the train and validation loss are comparable. ResNet50 and EfficientNet both demonstrate stable training processes with minimal gap between training and

validation losses, which is indicative of good generalization capabilities. However, the VGG model shows an initial spike in validation loss, which quickly stabilizes. Among the pre-trained networks, VGG16 achieved the highest validation accuracy of 93%. It balanced predictions, with very few misclassifications of both mud and non-mud labels. On the test set, VGG likewise achieved a precision of 92% and a recall of 95% for mud, and a precision of 94% with a recall of 92% for no mud. ResNet50 stabilized at 85% accuracy, while EfficientNet reached 88%. Correspondingly, ResNet50's final test metrics show mud precision and recall at 85%, while EfficientNet reaches 90% precision and 86% recall for mud; both models maintain similarly high figures for no-mud predictions (84%/84% for ResNet50, 86%/89% for EfficientNet). These differences suggest that deeper or more specialized feature extractors do not necessarily guarantee higher performance on this particular dataset. Instead, VGG16's simpler stack of 3×3 convolutions, proved particularly effective for discerning the nuanced visual signatures of mud deposits.

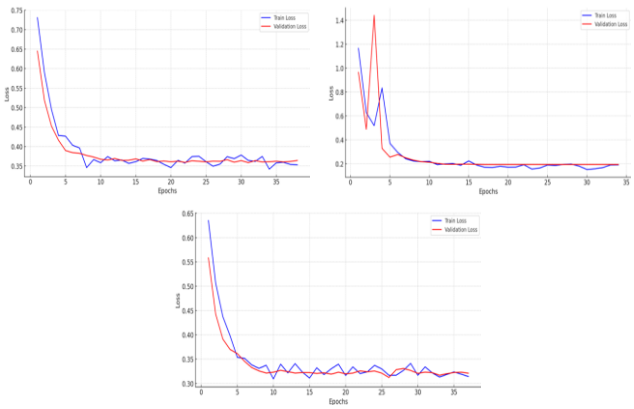

Figure 4. Training and validation loss curves for three transfer learning models: ResNet50 (top-left), VGG16 (top-right), and EfficientNet (bottom).

Parallel to the transfer learning approach, four simpler custom CNN architectures were trained from scratch to evaluate the viability of dedicated networks. The most basic approach, CNN_2, highlighting the limitations of a shallow design with fewer convolutional layers and no batch normalization. In final testing, this approach reached 76% accuracy overall, with mud precision at 82% and recall at 68%, and no-mud precision at 71% and recall at 84%. CNN_3 improved on this performance, reaching 81%, by introducing multiple consecutive convolutions, batch normalization layers, and slightly deeper feature representations. Its test precision/recall also climbed to around 82%/80% for mud and 79%/82% for no mud, indicating more balanced classification. CNN_4, which incorporated four blocks of two convolutional layers each alongside leaky ReLU with a small negative slope to maintain nonzero gradients for negative inputs and thereby alleviate the "dying ReLU" problem, achieved 90% accuracy on the validation set. Notably, its test performance stands at 90% accuracy, with mud precision of 92% and recall of 89%, and no-mud precision of 89% and recall of 92%. This significant boost reflects the benefit of increased network depth, careful architectural choices, and adequate regularization. Further experiments led to the inception-style network (Inception_CNN), which adopted multi-branch feature extraction in each block by combining multiple filter sizes in parallel. This design sought to capture both broad and fine-grained spatial features within the same layer. The Inception_CNN converged to a performance equal to that of the best transfer learning approach, reaching a validation accuracy of 93%, showing equally robust detection for both classes, with only a handful of misclassifications. Its final test performance likewise mirrors VGG at 93% accuracy overall, supported by a mud precision of 93% and recall of 94%, plus a 93% precision and 93% recall for no mud. The Inception_CNN, despite its fluctuations, the overall variation is small relative to the range (0.2–0.6).

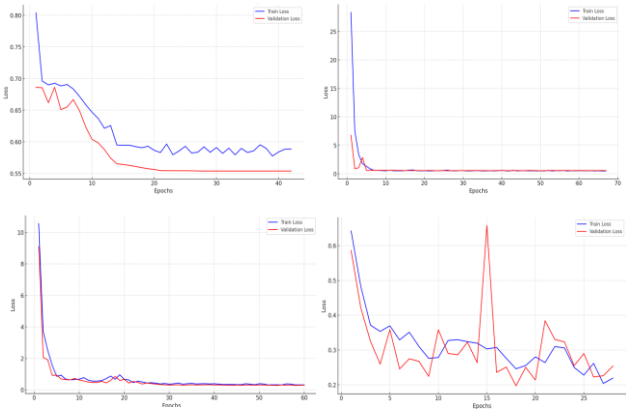

Figure 5. Training and validation loss curves for the four custom CNN models trained from scratch: CNN_2 (top-left), CNN_3 (top-right), CNN_4 (bottom-left), and Inception_CNN (bottom-right).

This indicates that the model isn't experiencing severe instability. Similarly, the CNN_4 and CNN_3 exhibit consistent downward trends in their loss curves, in spite of higher values at the beginning of the training. The simplest model (CNN_2) showed a gap between training and validation loss suggests some level of underfitting. Training loss doesn't go below 0.6, indicating that the model lack the complexity needed to capture all patterns in the training data. Additionally, classical machine learning models provide a useful comparative baseline: SVM achieves 67% accuracy (precision and recall in the mid-to-high 60s for both mud and no mud), Random Forest reaches 79% with balanced mud/no-mud metrics around 79–81%, and XGBoost improves further to 83%, featuring mud precision of 84% and recall of 83%, as well as 82%/83% for no mud.

Model	Test Acc.	Precision Mud	Recal l Mud	Precision No Mud	Recall No Mud
SVM	67	69	64	64	69
Random F.	79	79	81	79	78
XGBoost	83	84	83	82	83
ResNet50	85	85	85	84	84
VGG	93	92	95	94	92
Eff.Net	88	90	86	86	89
CNN_2	76	82	68	71	84
CNN_3	81	82	80	79	82
CNN_4	90	92	89	89	92
Inception_CNN	93	93	94	93	93

Table 1. The table presents the evaluation results of various classification models applied to the drone imagery dataset. Metrics include test accuracy (Test Acc.), precision and recall for both the "Mud" and "No Mud" classes.

4. Discussion

The study's performances across both traditional and advanced CNN models highlight the critical role of high-resolution UAV imagery and analytical methods in geological monitoring. The performance of traditional machine learning algorithms raises

important considerations. Despite their historical success in various classification tasks, algorithms like SVM, Random Forest, and XGBoost were less effective in handling the complex and subtle distinctions required for accurate mud classification in UAV imagery. The performance of traditional machine learning algorithms was found to be comparable to that of simpler CNN models employed, such as CNN_2 and CNN_3, demonstrating that even basic convolutional architectures can achieve the same results. This highlights a critical limitation of traditional methods: their dependence on handcrafted features and their accurate choice based on the application environment. The results thus emphasize the necessity for more adaptive and sophisticated analytical tools in environmental monitoring and geological assessments, pointing towards deep learning as a more capable approach in contexts characterized by high variability and complexity in visual data. The VGG16 and Inception_CNN models, which achieved top accuracies, demonstrate that both transfer learning and custom-designed CNN architectures can efficiently handling the specific challenges posed by mud volcano imagery. These findings underscore the potential of deep learning in enhancing geospatial analysis and risk assessment in volatile geological settings. However, the misclassification of certain samples, suggests room for improvement. This could be addressed by expanding the dataset to include more varied geological contexts or by integrating multimodal data sources that might provide additional discriminative features not visible in conventional RGB imagery. Future studies could explore the inclusion of thermal or multispectral imagery, which may capture subtle differences in material composition not discernible in visual-spectrum images. Furthermore, while data augmentation significantly bolstered model generalization, the approach might benefit from more sophisticated augmentation techniques such as generative adversarial networks (GANs) that can produce more realistic image variations. This could be particularly useful in environments where UAV access is limited or where environmental conditions rapidly change. The study also highlights a critical insight into the scalability of model architectures. While more complex models like EfficientNet offered some improvements, the simpler VGG16 and custom inception architecture's success points to the importance of model selection based on the specific characteristics of the dataset rather than the complexity alone. This suggests that for certain context and applications, simpler architectures trained from scratch may be preferable, especially when deploying models in real-time monitoring systems where computational resources are at a premium.

References

Bay, H., Ess, A., Tuytelaars, T., Van Gool, L., 2008: Speeded up robust features (SURF). Computer Vision and Image Understanding, 110(3), 346–359. doi.org/10.1016/j.cviu.2007.09.014

Bianconi, F., Fernández, A., 2007: Evaluation of the effects of Gabor filter parameters on texture classification. Pattern Recognition, 40(12), 3325–3335. doi.org/10.1016/j.patcog.2007.04.023

Bozkir, A., Nefeslioğlu, H. A., Kartal, O., Sezer, E., Gokceoglu, C., 2020: Geological strength index prediction by vision and machine learning methods.

Breiman, L., 2001: Random forests. Machine Learning, 45(1), 5–32. doi.org/10.1023/A:1010933404324

Buslaev, A., Parinov, A., Khvedchenya, E., Druzhkov, P., Kalinin, A. A., 2020: Albumentations: Fast and flexible image augmentations. Information, 11(2), 125. doi.org/10.3390/info11020125

Szegedy, C., et al., 2015: Going deeper with convolutions. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, pp. 1–9. doi.org/10.1109/CVPR.2015.7298594

Calonder, M., Lepetit, V., Strecha, C., Fua, P., 2010: BRIEF: Binary robust independent elementary features. In: European Conference on Computer Vision (ECCV). Springer.

Cangemi, M., Madonia, P., 2014: Mud volcanoes in onshore Sicily: A short overview.

Canny, J., 1986: A computational approach to edge detection. IEEE Transactions on Pattern Analysis and Machine Intelligence, 8(6), 679–698. doi.org/10.1109/TPAMI.1986.4767851

Chakravarti, R., Meng, X., 2009: A study of color histogram based image retrieval. In: 2009 Sixth International Conference on Information Technology: New Generations (ITNG), pp. 1323–1328. doi.org/10.1109/ITNG.2009.126

Chen, T., Guestrin, C., 2016: XGBoost: A scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 785–794. doi.org/10.1145/2939672.2939785

Chen, Z., Arrowsmith, J. R., Das, J., 2021: Autonomous robotic mapping of fragile geologic features. arXiv. doi.org/10.48550/arXiv.2105.01287

Cheng, G., Han, J., Lu, X., 2017: Remote sensing image scene classification: Benchmark and state of the art. Proceedings of the IEEE, 105, 1865–1883.

Cheng, G., Xie, X., Han, J., Guo, L., Xia, G. S., 2020: Remote sensing image scene classification meets deep learning: Challenges, methods, benchmarks, and opportunities. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 13, 3735–3756.

Colomina, I., Molina, P., 2014: Unmanned aerial systems for photogrammetry and remote sensing: A review. ISPRS Journal of Photogrammetry and Remote Sensing, 92, 79–97.

Cortes, C., Vapnik, V., 1995: Support-vector networks. Machine Learning, 20, 273–297. doi.org/10.1007/BF00994018

Dalal, N., Triggs, B., 2005: Histograms of oriented gradients for human detection. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), Vol. 1, pp. 886–893. doi.org/10.1109/CVPR.2005.177

Etiope, G., Caracausi, A., Favara, R., Italiano, F., Baciu, C., 2002: Methane emission from the mud volcanoes of Sicily (Italy). Geophys. Res. Lett., 29(8). doi.org/10.1029/2001GL014340

Farahbakhsh, E., Chandra, R., Olierook, H. K. H., Scalzo, R., Clark, C., Reddy, S. M., Muller, R. D., 2018: Computer vision-based framework for extracting geological lineaments from optical remote sensing data. International Journal of Remote Sensing, 41, 1–28. doi.org/10.1080/01431161.2019.1674462

Frazier, P. I., 2018: A tutorial on Bayesian optimization. arXiv Preprint arXiv:1807.02811. doi.org/10.48550/arXiv.1807.02811

Freeman, W. T., Roth, M., 1994: Orientation histograms for hand gesture recognition. Tech. Rep. TR94-03, MERL – Mitsubishi Electric Research Laboratories, Cambridge, MA.

Gabor, D., 1946: Theory of communication. Journal of the Institute of Electrical Engineers, 93, 429–457.

Glorot, X., Bordes, A., Bengio, Y., 2010: Deep sparse rectifier neural networks. Journal of Machine Learning Research, 15.

Granlund, G. H., 1978: In search of a general picture processing operator. Computer Graphics and Image Processing, 8(2), 155–173. doi.org/10.1016/0146-664X(78)90047-3

He, K., Zhang, X., Ren, S., Sun, J., 2016: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778.

Hu, R., Barnard, M., Collomosse, J., 2010: Gradient field descriptor for sketch based retrieval and localization. In: 2010 IEEE International Conference on Image Processing (ICIP), p. 1025. IEEE.

Huang, L., Yu, X., Zuo, X., 2017: Edge detection in UAV remote sensing images using the method integrating Zernike moments with clustering algorithms. International Journal of Aerospace Engineering, 2017, 1–7. doi.org/10.1155/2017/1793212

Idrissa, M., Acheroy, M., 2002: Texture classification using Gabor filters. Pattern Recognition Letters, 23(9), 1095–1102. doi.org/10.1016/S0167-8655(02)00056-9

Jain, A. K., Vailaya, A., 1995: Image retrieval using color and shape. Elsevier Science Ltd, Great Britain.

Jones, D. R., Schonlau, M., Welch, W. J., 1998: Efficient global optimization of expensive black-box functions. Journal of Global Optimization, 13(4), 455–492.

Joseph, S., Ujir, H., Hipiny, I., 2017: Unsupervised classification of intrusive igneous rock thin section images using edge detection and colour analysis. In: 2017 IEEE International Conference on Signal and Image Processing Applications (ICSIPA), pp. 90–95. doi.org/10.1109/ICSIPA.2017.8120669

Kamarainen, J.-K., Kyrki, V., Kalviainen, H., 2006: Invariance properties of Gabor filter-based features—Overview and applications. IEEE Transactions on Image Processing, 15(5), 1088–1099. doi.org/10.1109/TIP.2005.864174

Kingma, D. P., Ba, J., 2017: Adam: A method for stochastic optimization. arXiv:1412.6980. doi.org/10.48550/arXiv.1412.6980

Krizhevsky, A., Sutskever, I., Hinton, G. E., 2012: ImageNet classification with deep convolutional neural networks. Advances in Neural Information Processing Systems, 25, 1097–1105.

Laliberte, A. S., Goforth, M. A., Steele, C. M., Rango, A., 2011: Multispectral remote sensing from unmanned aircraft: Image processing workflows and applications for rangeland environments. Remote Sensing, 3(11), 2529–2551.

LeCun, Y., Bengio, Y., Hinton, G., 2015: Deep learning. Nature, 521(7553), 436–444.

LeCun, Y., Bottou, L., Bengio, Y., Haffner, P., 1998: Gradient-based learning applied to document recognition. Proceedings of the IEEE, 86(11), 2278–2324.

Leonardis, A., Bischof, H., Pinz, A., (Eds.), 2006: Lecture Notes in Computer Science (Vol. 3951). Springer.

Lizé, J., Débordès, V., Lu, H., Kpalma, K., Ronsin, J., 2020: Local binary pattern and its variants: Application to face analysis. In: Advances in Smart Technologies: Applications and Case Studies – Selected Papers from the First International Conference on Smart Information and Communication Technologies, pp. 94–102. Springer. doi.org/10.1007/978-3-030-53187-4_11

Long, Z., Alaudah, Y., Qureshi, M. A., Al Farraj, M., Wang, Z., Amin, A., Deriche, M., AlRegib, G., 2019: Characterization of migrated seismic volumes using texture attributes: A comparative study. arXiv:1901.10909. doi.org/10.48550/arXiv.1901.10909

Lowe, D. G., 1999: Object recognition from local scale-invariant features. In: Proceedings of the International Conference on Computer Vision, Vol. 2, pp. 1150–1157. IEEE.

Lowe, D. G., 2004: Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision, 60(2), 91–110.

Mao, A., Mohri, M., Zhong, Y., 2023: Cross-entropy loss functions: Theoretical analysis and applications. arXiv. doi.org/10.48550/arXiv.2304.07288

Martinelli, G., Judd, A., 2004: Mud volcanoes of Italy. Geological Journal, 39, 49–61.

McConnell, R., 1986: Method of and apparatus for pattern recognition (U.S. Patent No. 4,567,610). U.S. Patent and Trademark Office.

Napoli, R., Currenti, G., Giammanco, S., Greco, F., Maucourant, S., 2020: Imaging the Salinelle mud volcanoes (Sicily, Italy) using integrated geophysical and geochemical surveys. Annals of Geophysics, 63(4), PE442. doi.org/10.4401/ag-8215

Nex, F., Remondino, F., 2014: UAV for 3D mapping applications: A review. Applied Geomatics, 6(1), 1–15.

Ojala, T., Pietikäinen, M., Harwood, D., 1996: A comparative study of texture measures with classification based on featured distributions. Pattern Recognition, 29(1), 51–59.

Ojala, T., Pietikainen, M., Maenpaa, T., 2002: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Transactions on Pattern Analysis and Machine Intelligence, 24(7), 971–987. doi.org/10.1109/TPAMI.2002.1017623

Pan, S. J., Yang, Q., 2010: A survey on transfer learning. IEEE Transactions on Knowledge and Data Engineering, 22(10), 1345–1359.

Perera, S., Nalani, H., 2022: UAVs for a complete topographic survey. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLIII-B2-2022, 441–447. doi.org/10.5194/isprs-archives-XLIII-B2-2022-441-2022

Perez, L., Wang, J., 2017: The effectiveness of data augmentation in image classification using deep learning. arXiv. doi.org/10.48550/arXiv.1712.04621

Qiu, Q., Tan, Y., Ma, K., Tian, M., Xie, Z., Tao, L., 2023: Geological symbol recognition on geological map using convolutional recurrent neural network with augmented data. Ore Geology Reviews, 153, 105262. doi.org/10.1016/j.oregeorev.2022.105262

Rong, H., 2024: Exploration of geological hazards using improved SIFT algorithm based aerial surveying and mapping technology. In: 2024 International Conference on Data Science and Network Security (ICDSNS). IEEE. doi.org/10.1109/ICDSNS62112.2024.10691130

Rosten, E., Drummond, T., 2006: Machine learning for high-speed corner detection. In: European Conference on Computer Vision (ECCV), Vol. 1.

Rublee, E., Rabaud, V., Konolige, K., Bradski, G., 2011: ORB: An efficient alternative to SIFT or SURF. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 2564–2571. doi.org/10.1109/ICCV.2011.6126544

Shapiro, L. G., Stockman, G. C., 2003: Computer vision. Prentice Hall. (ISBN 0-13-030796-3)

Shorten, C., Khoshgoftaar, T. M., 2019: A survey on image data augmentation for deep learning. Journal of Big Data, 6(1), 1–48. doi.org/10.1186/s40537-019-0197-0

Simonyan, K., Zisserman, A., 2015: Very deep convolutional networks for large-scale image recognition. arXiv Preprint arXiv:1409.1556. doi.org/10.48550/arXiv.1409.1556

Soler, J. D., Beuther, H., Rugel, M., Wang, Y., Clark, P. C., Glover, S. C. O., Goldsmith, P. F., Heyer, M., Anderson, L. D., Goodman, A., Henning, T., Kainulainen, J., Klessen, R. S., Longmore, S. N., McClure-Griffiths, N. M., Menten, K. M., Mottram, J. C., Ott, J., Ragan, S. E., … Schilke, P., 2019: Histogram of oriented gradients: A technique for the study of molecular cloud formation. Astronomy & Astrophysics, 622, A166. doi.org/10.1051/0004-6361/201834300

Swain, M. J., Ballard, D. H., 1991: Color indexing. International Journal of Computer Vision, 7(1), 11–32.

Szeghalmy, S., Fazekas, A., 2023: A comparative study of the use of stratified cross-validation and distribution-balanced stratified cross-validation in imbalanced learning. Sensors, 23(4), 2333. doi.org/10.3390/s23042333

Tan, M., Le, Q., 2019: EfficientNet: Rethinking model scaling for convolutional neural networks. In: Proceedings of the International Conference on Machine Learning, pp. 6105–6114.

Vangah, J. W., Ouattara, S., Ouattara, G., Clément, A., 2023: Global and local characterization of rock classification by Gabor and DCT filters with a color texture descriptor. arXiv:2302.08219. doi.org/10.48550/arXiv.2302.08219

Wang, X., Kealy, A., Li, W., Jelfs, B., Gilliam, C., May, S. L., Moran, B., 2021: Toward autonomous UAV localization via aerial image registration. Electronics, 10(4), 435. doi.org/10.3390/electronics10040435

Wu, J., Cao, M., Shan, L., Ying, Q., 2020: Higher performance for AutoML: The benefit of various ensemble Bayesian optimization strategy. Retrieved from https://valohaichirpprod.blob.core.windows.net/papers/duxiaoman.pdf

Yang, Q., Zhang, Y., Dai, W., Pan, S. J., 2020: Transfer Learning. Cambridge University Press.

Yu, X., Lyu, Z., Hu, D., Xu, J., 2013: Scale-invariant feature transform based on the frequency spectrum and the grid for remote sensing image registration. GIScience & Remote Sensing, 50(5), 543–561. doi.org/10.1080/15481603.2013.827370

Zhan, Z., Liu, B., Sang, X., Xue, L., 2020: Accelerate fine-scale geological mapping with UAV and convolutional neural networks. IOP Conference Series: Materials Science and Engineering, 768(7), 072082. doi.org/10.1088/1757-899X/768/7/072082

Zhang, X., Yun, L., Zheng, Y., Wang, D., Hua, L., 2023: Enhance the accuracy of landslide detection in UAV images using an improved Mask R-CNN model: A case study of Sanming, China. Sensors, 23(9), 4287. doi.org/10.3390/s23094287

Zhao, W., Du, S., Qi, J., 2017: Learning representative features from land use/land cover change using deep learning. Remote Sensing, 9(12), 1279.

Zhong, Z., Zheng, L., Kang, G., Li, S., Yang, Y., 2020: Random erasing data augmentation. In: Proceedings of the AAAI Conference on Artificial Intelligence, 34(7), 13001–13008. doi.org/10.1609/aaai.v34i07.6985

Zhou, M., Zhou, Y., Yang, D., Song, K., 2024: Remote sensing image classification based on Canny operator enhanced edge features. Sensors, 24(12), 3912. doi.org/10.3390/s24123912

Zhu, H., Jiang, Y., Zhang, C., Liu, S., 2022: Research on Mosaic Method of UAV Low-altitude Remote Sensing Image based on SIFT and SURF. Journal of Physics: Conference Series, 2203, 012027. doi.org/10.1088/1742-6596/2203/1/012027

Zhu, Q., Yeh, M.-C., Cheng, K.-T., Avidan, S., 2006: Fast human detection using a cascade of histograms of oriented gradients. In: 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06), Vol. 2, pp. 1491–1498. IEEE.

Zhuang, F., Qi, Z., Duan, K., Xi, D., Zhu, Y., Zhu, H., Xiong, H., He, Q., 2021: A comprehensive survey on transfer learning. Proceedings of the IEEE, 109(6), 43–66. doi.org/10.1109/5.0037924