

Enhancing Lightweight SAR Image Classification Using MobileNetV3-Small Implemented with Convolutional Block Attention Module (CBAM)

Narimane Benouakta¹, Zohra Slimane², and Belkacem Benadda³

¹Departement of Electronic and Telecommunication, Belhadj Bouchaib University, SSL Laboratory, AinTémouchent, Algeria,

²Department of Telecommunication, Abou Bekr Belkaid University, STIC Laboratory, Tlemcen, Algeria,

³Departement of Electronic, Abou Bekr Belkaid University, Tlemcen, Algeria,

*Corresponding author: narimane.benouakta@univ-temouchent.edu.dz

ARTICLE INFO

ABSTRACT

Received: 30 Dec 2024

Revised: 12 Feb 2025

Accepted: 26 Feb 2025

The rapid worldwide advancement of environmental remote sensing technologies has created a growing need for automated, efficient, and high-performing analysis methods for Synthetic Aperture Radar (SAR) imagery. SAR is particularly valuable because it provides consistent imaging capabilities regardless of weather or lighting conditions, making it ideal for crucial applications such as environmental monitoring and disaster management. Despite these advantages, SAR image classification remains a challenging task due to the complex nature of SAR data, which often includes speckle noise, limited textural information, and high computational demands. To address these challenges, this study introduces an efficient and lightweight Convolutional Neural Network (CNN) architecture based on MobileNetV3-Small, integrated with a Convolutional Block Attention Module (CBAM). The attention mechanism enhances the extraction of meaningful features while reducing unwanted backscatter effects in SAR feature maps. The proposed hybrid model was trained and evaluated on a Sentinel-1 SAR dataset comprising four land cover categories: Agriculture, Barren Land, Grassland, and Urban areas. Implementing data augmentation and transfer learning techniques, the model was trained over 50 epochs, and achieved the high accuracy of 95.74%. This results, outperforming the original MobileNetV3-Small backbone, stems directly from the implementation of CBAM attention mechanism, allowing the proposed model to successfully emphasize relevant spatial and channel features. In this work, it was demonstrated that the proposed MobileNetV3-Small with CBAM architecture delivers a powerful balance between computational efficiency, robust accuracy, and generalization, capability, making it a suitable for integration in near-real-time applications and embedded systems. The results prove the potential of lightweight attention- augmented CNN architecture for a generation of autonomous and intelligent Earth observation systems, offering opportunities for further research in global ecological monitoring.

Keywords: SAR image classification, CNN, MobileNetV3, CBAM, Transfer Learning.

INTRODUCTION

Lorem Synthetic Aperture Radar (SAR) is an active radar system capable of producing high-resolution 2D images and 3D reconstructions of large areas (Li et al., 2021) under all weather and day-and-night conditions (Passah et al., 2022), making in a perfect candidate for consistent observation of ecosystems in pastoral areas and semi- arid zones (Zhang et al., 2021). This ability makes it particularly valuable for various environmental monitoring applications, including checking water pollution (Han & Ma, 2021), deforestation monitoring (Ygorra et al., 2021; Ygorra et al., 2022), floods monitoring, and other different water shapes like freshwater lake inundation, river water level (Chen

et al., 2021). Recent work on airborne radar technologies allowed advanced observation which facilitate disaster management such as earthquakes, floods, and hurricanes (Ardila et al., 2022; Xie et al., 2023; Moyano et al., 2016). SAR is also widely used in military and defence, including surveillance, reconnaissance, and target detection (Perry, 2017).

The integration of deep learning techniques has revolutionized SAR data analysis. Traditional machine learning algorithms such as Random Forests, Support Vector Machines (SVM), and K-Nearest Neighbors rely heavily on manually engineered features. This dependence limits their ability to generalize and accurately represent the complex, nonlinear, and heterogeneous backscattering characteristics of SAR data. Traditional classification methods therefore require intricate feature extraction processes, which often restrict their adaptability. Although deep learning models have demonstrated remarkable effectiveness, they typically demand substantial computational resources, which can hinder their application in real-time scenarios. Nonetheless, deep learning architectures possess the unique capability to automatically learn and extract discriminative features directly from raw data, leading to significant improvements in performance. In particular, Deep Convolutional Neural Networks (CNNs) have emerged as a powerful solution to overcome the limitations of manual feature engineering in SAR data analysis. CNNs can automatically capture hierarchical spatial and textural representations from radar imagery, eliminating the need for handcrafted features. These models have achieved state-of-the-art results across various SAR-related tasks, including change detection, image classification, and target recognition.

Despite their success, traditional Convolutional CNN architectures such as DenseNet, ResNet, and VGGNet are often computationally intensive and contain a large number of parameters, which limits their suitability for real-time or resource-constrained applications (Howard et al., 2017). In practical contexts such as unmanned aerial vehicles (UAVs), ground-based stations, and satellite systems achieving both high accuracy and computational efficiency is essential. To meet these demands, Lightweight Convolutional Neural Networks (L-CNNs) like MobileNet, EfficientNet, and ShuffleNet have been introduced (Zhang et al., 2018). These models leverage efficient design principles, including depthwise separable convolutions and inverted residual structures, allowing them to maintain strong performance while significantly reducing the number of parameters and energy consumption.

MobileNetV3 represents one of the most advanced lightweight architectures designed for mobile and embedded systems. It introduces several architectural improvements over its predecessors, such as the Hard Swish activation function, the integration of Squeeze and Excitation (SE) modules, and the use of an optimized Neural Architecture Search (NAS). Together, these enhancements enable MobileNetV3 to deliver strong feature extraction capabilities while maintaining a compact model size and fast inference speed. Such characteristics make it particularly well-suited for Synthetic Aperture Radar (SAR) applications in environmental monitoring and related tasks.

Although lightweight networks provide significant advantages in efficiency, they may still struggle to capture highly discriminative features when applied to SAR imagery, which is often characterized by surface roughness, speckle noise, and complex textural variations. To address these challenges, attention mechanisms have been incorporated into Convolutional CNN architectures to strengthen feature learning. These modules dynamically adjust the weighting of spatial and channel representations, enabling the network to focus on the most informative regions of the image while suppressing irrelevant or noisy information.

Among the various attention mechanisms developed in recent years, the Convolutional Block Attention Module (CBAM) has gained particular recognition for its strong performance and computational efficiency. CBAM applies channel attention and spatial attention in sequence, enabling the network to emphasize the most informative feature channels and spatial regions (Woo et al., 2018). This dual-attention mechanism makes CBAM especially well-suited for SAR imagery, where accurate representation of spatial structures and fine details is crucial for discriminative feature learning.

In this study, a novel hybrid model combining MobileNetV3 with CBAM is proposed for SAR image classification. By integrating attention modules into the feature extraction stages of the lightweight MobileNetV3 architecture, the model's ability to capture complex textural patterns is significantly enhanced while maintaining computational efficiency. The proposed framework effectively differentiates between multiple land-cover types namely agricultural

lands, barren lands, grasslands, and urban areas making it a promising approach for real-time environmental and land monitoring applications.

Extensive experiments were performed to evaluate and validate the MobileNetV3-CBAM model using a SAR dataset containing samples from four distinct terrain categories. To improve model robustness and generalization, transfer learning and data augmentation techniques were employed. Experimental results demonstrate that integrating CBAM substantially boosts classification accuracy with minimal additional computational cost, highlighting its potential for deployment in embedded and resource-constrained systems.

SAR IMAGE CLASSIFICATION WITH DEEP LEARNING

Synthetic Aperture Radar (SAR) has become one of the most important Earth observation technology, offering incomparable ability to provide 2D and 3D high resolution images unaffected by weather conditions or light availability. Its mechanism depends on effectively emitting microwave signals and calculate the amplitude and angles of backscatters reflected from earth surface, unlike optical imagery systems that rely on the sun availability to create data. This technique allows SA radars to continuously provide inputs regardless of sky clarity or nighttime issues, ensuring regular monitoring especially for environmental settings. These qualities make SAR effective in various applications such as flood surveillance, Earth flow, deforestation monitoring, and the surveillance of agricultural and pastoral areas (Balz & Zhang, 2020). Additionally, the capability of SAR backscatter to sense moisture variations, and surface texture makes it a strong tool for long term monitoring. (Ulaby et al., 2014; Lee & Pottier, 2017).

Recently, several studies havz utilized SAR image for different applications for instance agricultural activities and crop monitoring (Khan et al., 2022; McNairn et al., 2009), also it was used water pollution monitoring (Han & Ma, 2021), forest degradation (Ygorra et al., 2021; Ygorra et al., 2022), floods monitoring, and other different water bodies such as river water level (Chen et al., 2021).

However, the automated comprehensive of SAR images is still considered a challenge due to its complex nature and random geometric patterns. The presence of speckle noise, complex surface scattering shapes from different surface kinds, particularly random angle, introduces difficulties in classification and feature extraction. The traditional image classification methods, including statistical classifiers such as Maximum Likelihood and textural methods based on Gray-Level Co-occurrence Matrix (GLCM), rely on manual feature extraction. As noted by Moser et al. (2019), these manually features extraction shows strong sensitivity to noise and environmental variability, which limit generalization across wide or diverse SAR datasets.

With the intordution of deep learning, particurally Convolutional Neural Networks (CNNs), has completely change the analysis SAR data. CNNs zllowed learning of hierarchical feature extraction to be automated, using the ability to model complex spatial fetures and textual patterns from unpreprocessed data. This revolutionary methods improved the performance and allowed getting of high accuracy compared to traditional feature engineering techniques (Zhao et al., 2021).At first, applications of architectures such AlexNet and VGG to SAR image classification demonstrated the effectiveness of feature extraction frameworks. Nonetheless, their e computational demands and dependence on large parameter counts limits their use in embedded systems where resources and time are limited (Krizhevsky et al., 2012; Simonyan & Zisserman, 2014).

Table 1. Summary of Related Works on SAR Image Classification

Study	Model / Technique	Dataset	Parameter Count (M)	Accuracy (%)
Zhou et al., 2022	ResNet-18	Sentinel-1 Urban Scenes	11.7	91.0
Li et al., 2021	DenseNet-121	Gaofen-3 SAR Images	8.0	92.3

Chen et al., 2023	MobileNetV3-Large	Sentinel-1 + Sentinel-2 Pairs	5.4	93.2
Wang et al., 2023	ShuffleNet V2 + SE Module	Airborne SAR Scenes	3.8	94.0
Our proposed model	MobileNetV3-Small + CBAM	Sentinel-1 Terrain Dataset	2.7	95.75

To solve the problem of the computational effectiveness of the newly used deep learning models, novel architectures such Residual Networks (ResNets) (He et al., 2016) and Densely Connected Convolutional Networks (DenseNet) (Huang et al., 2017) introduced innovative linking techniques to improve feature extraction. ResNet's residual shortcut connections allowed deep networks to be trained effectively by preserving gradient flow. On the other hand, DenseNet introduced direct connection using dense inter-layer connectivity. These models achieved strong results in optical image classification and were adapted for SAR data. For instance, ResNet-based architectures have showed high performance in polarimetric SAR (PolSAR) imagery, particularly improving land-cover classification and urban feature detection (Gao et al., 2021), whereas DenseNet's design has proven computational efficiency by reducing vanishing gradients. Table 1 demonstrates a summary of related work on SAR image classification.

Although these deep architectures proved their effectiveness, they still are computationally demanding, making them unsuitable for real-time applications and embedded systems. To address these challenges, lightweight neural networks have been introduced to be an efficient replacement for the heavy SAR data.

The MobileNet architecture, proposed by Howard et al. (2017), marked an important step forward in designing lightweight convolutional neural networks. By introducing depthwise separable convolutions, the model size was reduced to about 4.2 million parameters, which significantly lowered the computational demand. Later, MobileNetV2 (Sandler et al., 2018) improved this design through the use of inverted residual blocks and linear bottlenecks, enabling better feature reuse and reducing information loss. Building on these advances, MobileNetV3 integrated squeeze-and-excitation (SE) blocks, a neural architecture search (NAS)-based optimizer, and the Hard-Swish activation function to further refine the architecture automatically (Howard et al., 2019). Figure 1 illustrates the overall structure of MobileNetV3. Altogether, these developments allow MobileNetV3 to approach state-of-the-art accuracy while keeping computational costs low, which makes it a practical choice for embedded or airborne applications.

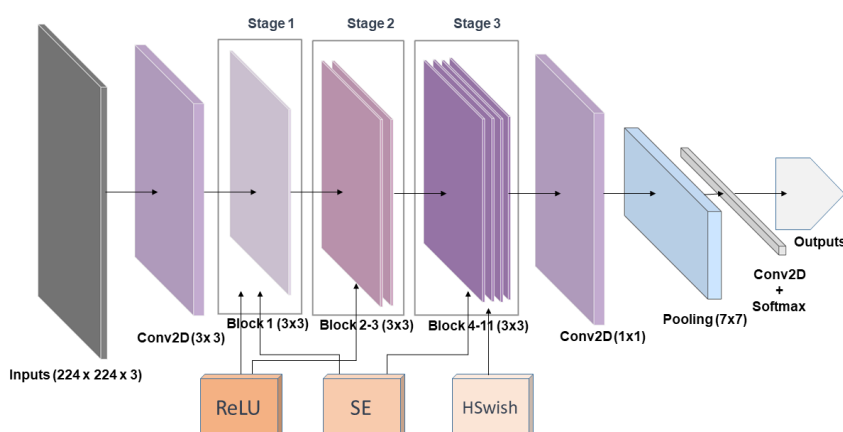


Figure 1. MobileNetV3-Small Architecture Overview

Nevertheless, applying lightweight neural networks directly to SAR imagery remains challenging. In contrast to optical images, SAR data are mainly affected by coherent microwave backscattering and the presence of speckle noise, while also lacking clear spectral gradients. Consequently, lightweight CNN models often have difficulties differentiating between land-cover classes with similar scattering properties, such as dry vegetation and bare soil (Zhang et al., 2023). Recent studies have adopted attention mechanisms designed to enhance their capability to capture relevant features.

LIGHTWEIGHT NETWORKS AND ATTENTION MECHANISM FOR SAR

Attention mechanism have recently become a transformative development in deep learning, improving feature representation capabilities. Inspired by how human vision focuses on important details, attention modules help convolutional networks emphasize the spatial regions and feature channels that matter most for decision-making, while reducing the influence of less informative areas. In practice, attention mechanisms such as the Squeeze-and-Excitation (SE) block (Hu et al., 2018) and the Convolutional Block Attention Module (CBAM) (Woo et al., 2018) have shown strong performance, particularly when dealing with complex data such as SAR imagery.

Building on this challenge, researchers have drawn inspiration from how human vision focuses on relevant details. Attention modules allow convolutional networks to highlight the spatial regions and feature channels that contribute most to decision-making, while downplaying less informative areas. In particular, mechanisms such as the Squeeze-and-Excitation (SE) block (Hu et al., 2018) and the Convolutional Block Attention Module (CBAM) (Woo et al., 2018) have demonstrated strong performance when applied to complex data, including SAR imagery. Consequently, this spatial blindness is a significant limitation when analyzing SAR data, where spatial distribution of backscatter, such as textural patterns and geometric shapes, is semantically meaningful.

The Convolutional Block Attention Module (CBAM), introduced by Woo et al. (2018), represents a more complete version of attention paradigm. CBAM uses two sequential approach which initially conclude inter-channel connections and then models intra-spatial components. In the first step, channel attention identifies “what” patterns are dominant by processing the connections between feature channels. The second one, the spatial attention shows a 2D attention map that precise “where” the important features are located. The dual techniques enable the network to filter less informative feature channels and block irrelevant background noise. [Figure 2] demonstrates the CBAM architecture, showing both channel and spatial attention branches and their sequential integration. Channel attention includes: Global Average Pooling 2D (GAP), Global Max Pooling 2D (GMP), Sigmoid activation, and Channel Attention Map; whereas Spatial Attention consist of Average Pooling (AP), Max Pooling (MP), Sigmoid activation, and Spatial Attention Map. CBAM is designed to be lightweight, easily deployed, and computationally undemanding, making it a powerful tool to improve the feature extraction of efficiency-conscious models like MobileNetV3, especially for complex data processing such as SAR classification.

Mathematically, CBAM technique works on an intermediate feature map $F \in \mathbb{R}^{C \times H \times W}$ through a sequential process of spatial attention and channels. First, it implements a channel attention map $\mathbf{M}_c(F) \in \mathbb{R}^{C \times 1 \times 1}$ and followed by a special map $\mathbf{M}_s(F') \in \mathbb{R}^{1 \times H \times W}$. The sequentially refined feature transformation is defining as the following:

$$\mathbf{F}' = \mathbf{M}_c(\mathbf{F}) \otimes \mathbf{F} \quad (1)$$

$$\mathbf{F}'' = \mathbf{M}_s(\mathbf{F}') \otimes \mathbf{F}' \quad (2)$$

Where $\otimes$ is the element- wise multiplication. The dual application of channel attention followed by spatial attention enable the model to dynamically focus on both the spectral and spatial feature extraction.

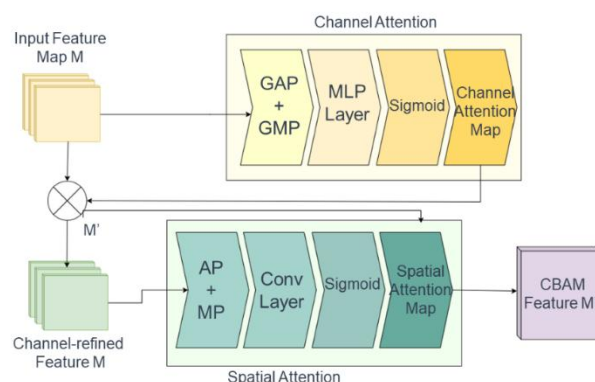


Figure 2. Convolutional Block Attention Module (CBAM) Architecture

Recent work has proven the effectiveness of the implementation of attention modules, particularly CBAM, in remote sensing applications. CBAM showed consistent performance in refining features representations in SAR data processing. For instance, Wang et al. (2023) showcased this by combining CBAM with ResNet to examine flood detection with Sentinel-1 dataset, achieving a 6% improvement in accuracy, proving CBAM's spatial attention efficiency to distinguish water zones. In the domain of PolSAR land cover classification, Gao et al. (2021) proposed the integration of CBAM to DenseNet model in order to improve performance in classifying difficult urban-vegetation areas. In another study, Zhang et al. (2022) presented a comparative study of CBAM and SE deployed in a lightweight CNN architecture, delivering consistently high accuracy and generalization for multi-temporal SAR processing. Collectively, these studies prove that hybrid techniques, which combine efficient backbones with lightweight attention mechanisms such as CBAM, are achieving strong performance while minimizing computational cost in SAR imagery.

The development of attention techniques has produced several efficient alternative mechanisms such as Efficient Channel Attention (ECA) and Coordinate Attention (CA). ECA (Wang et al., 2020) designed on avoiding dimensional reduction and replacing the fully-connected SE with local cross-channel connection method, making it extremely suitable for resource-limited embedded systems. On the other hand, Coordinate Attention (CA) (Hou et al., 2021) integrates spatial information into channel attention, allowing the network to capture precise long-range dependencies while maintaining computational efficiency. Although both methods are lightweight, CBAM—while slightly more complex—offers a more explicit and sequential analysis of channel and spatial attention, resulting in clearer and more interpretable feature maps. This makes CBAM a particularly strong tool for preserving the structural and geometric characteristics present in radar backscatter data.

In the context of pastoral and environmental monitoring, the need for efficient SAR automated models is critical. Lightweight attention-based architectures allow rapid, large-scale processing for SAR data, which required in near-real-time, resource-limited applications such flood detection, and vegetation degradation recognition. This paper proposes a hybrid architecture that embeds the Convolutional Block Attention Module (CBAM) to the MobileNetV3 backbone to target SAR land cover classification. Our technique is designed to improve feature extraction in SAR data, particularly through spatial and channel attention, while minimizing the computational cost.

The main contribution of this study is a practical demonstration that achieved an optimal accuracy and high performance using CBAM-MobileNetV3 framework, thereby offering a promising step toward real-time embedded SAR image classification for environmental monitoring.

MATERIALS

A) DATASET DESCRIPTION

To validate the proposed model, we used the publically available dataset 'Sentinel-1 & 2 Image Pairs (SAR & Optical) – Terrain – segregated' from Kaggle repository (RequiemOnk., 2021) The dataset consists of spatially aligned co-registered image pairs of C-band SAR and optical from Sentinel-1 satellite and Sentinel-2 satellite respectively, which

are classified and annotated by various terrain labels. In the context of this study, we exclusively used the SAR Sentinel-1 data to ensure radar-centric approach, focusing on four land cover classes including Urban, Agriculture, Grassland, and Bare Land. [Figure 3] shows samples of each land cover class of the Sanetinel-1 dataset used in this study.

Each SAR image in the dataset is pre-processed and stored as ‘.png’ file with a spatial resolution of 224 x 224 pixels. The dataset used contains 8,000 labeled samples in total evenly distributed into the four classes, where every class contains 2,000 image. The image labels, which link each image identifier with the corresponding land cover category, are accessible via a spreadsheet. The images are sourced from multiple geographic zones, allowing significant in sensing conditions, land-cover appearance, and backscatter responses.

Prior to training, the data was partitioned into 70% training, 15% validation, and 15% test samples, resulting in approximately 5,600, 1,200, and 1,200 images, respectively. Stratified random sampling was used to guarantee the original class distribution within each subset, thereby avoiding sampling bias and ensuring fair evaluation of generalization.

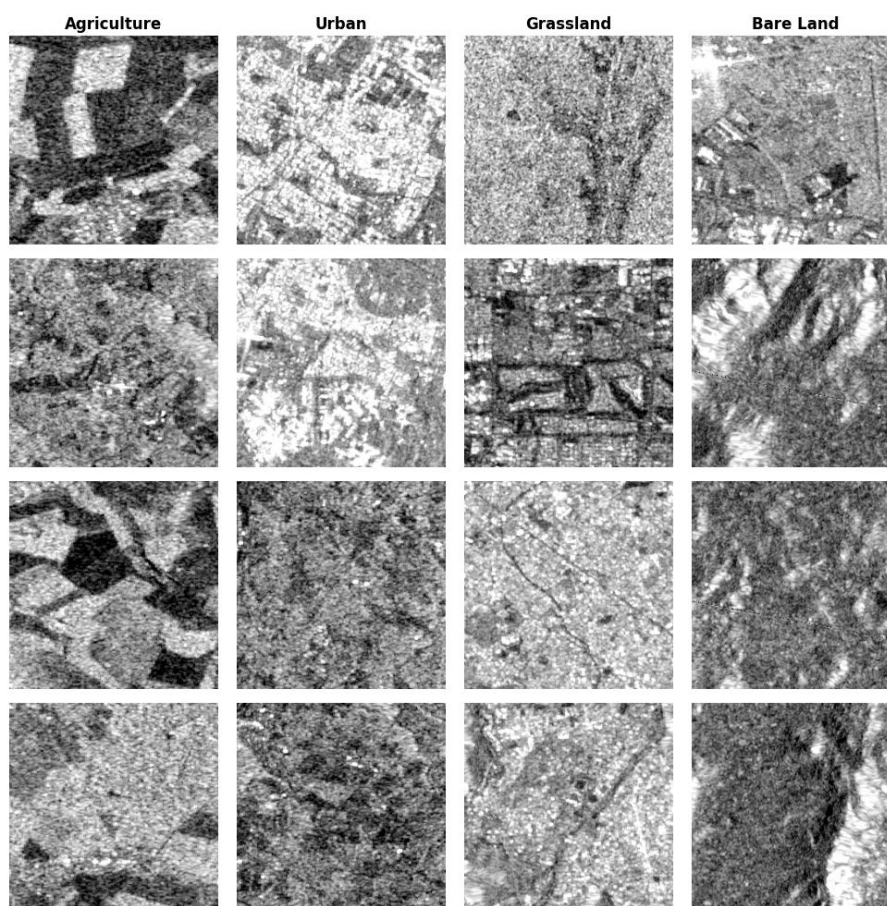


Figure 3. Representative Sentinel- 1 SAR Image Samples for Each Terrain Class

Due to SAR images nature that typically exhibit speckle noise, and to enhance model generalization performance and reduce the risk of overfitting to the training set, data augmentation is applied. Its pipeline includes: random horizontal/vertical flips and random rotation of up to $\pm 20^\circ$. To conform to the MobileNetV3 input dimensions and enable transfer learning from optical domain to the special characteristics of SAR data using pre-trained weights, all images were normalized based on the mean and standard deviation of ImageNet statistics (mean = [0.485, 0.456, 0.406], std = [0.229, 0.224, 0.225]).

B) DATASET CHARACTERISTICS AND CHALLENGES

The Synthetic Aperture Radar (SAR) works on a fundamentally different physical aspect than the optical imagery, capturing surface reflectance of the microwave backscatter intensity rather than reflected light. This results in unique texture patterns for different land covers, and high intra-class variance which happens due to significant variance in backscatter responses for each class. These scattering mechanisms distinctly influence class appearance: urban areas exhibit bright and chaotic returns caused by corner reflections; grasslands display relatively homogeneous volume scattering; bare land appears as dark and uniform regions; while agricultural fields present periodic, complex textures associated with crop type and growth stage.

The unique characteristics of SAR imagery—such as complex textures, speckle noise, and limited labeled data—pose significant challenges for lightweight CNN models. This underscores the need for architectures capable of emphasizing relevant channel-wise representations and capturing informative spatial patterns. This directly motivates the deployment of an attention technique such as CBAM, which allow guiding the network to avoid irrelevant activations, improve filter refinement, and selectively amplify informative regions and channel-wise features; in order to enhance performance without increasing model computational complexity.

METHODS

A) MODEL ARCHITECTURE

This study proposes a hybrid architecture for SAR image classification that enhance the performance of MobileNetV3-Small backbone with the Convolutional Block Attention Module (CBAM). Chosen for its efficiency and lightweight foundation, MobileNztV3 is a powerful tool for real-time embedded deployment or airborne radar applications. It uses depthwise separable convolutions and built-in Squeeze-and-Excitation (SE) blocks. [Figure 4] demonstrates the proposed architecture of the hybrid MobileNetV3- Small + CBAM diagram. The integration of CBAM further enhance the performance of this model, allowing stronger extraction of spatial and channel-wise features from complex SAR inputs.

The CBAM module is deployed sequentially subsequent to the convolutional layers following the feature extraction stages of the MobileNetV3 architecture. The channel attention block enables the model to scale feature maps along the channel dimension by focusing on relevant meaningful filters. On the other hand, the spatial attention block generates a two-dimensional spatial map that highlights the most informative regions. This dual attention mechanism enables the network to allocate its capacity toward learning terrain-specific features and diverse textural patterns, such as the diffuse scattering observed in agricultural areas.

The proposed CBAM-enhanced model preserves the computational efficiency of the lightweight MobileNetV3-Small architecture, with only a negligible 5% increase in parameters due to CBAM integration. The feature maps extracted from the original MobileNetV3-Small are fed into a compact classifier consisting of a batch normalization layer, a ReLU activation, a 40% dropout layer, and a final softmax dense layer for all classes (Urban, Agriculture, Grassland, and Bare Land). Despite having only 3.4 million trainable parameters, the model retains its lightweight nature while achieving high performance. This demonstrates that incorporating CBAM significantly enhances generalization and accuracy while maintaining minimal computational cost, making it an ideal choice for resource-constrained, real-time environmental monitoring applications.

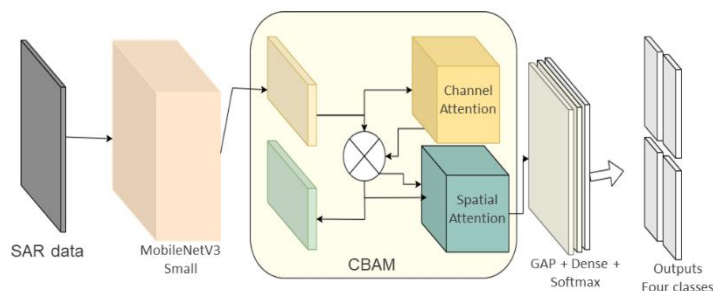


Figure 4. The Architecture of the Proposed MobileNetV3-Small + CBAM

B) TRAINING SETUP

The training and validation of the model were carried out on Google Colab, utilizing GPU acceleration to provide sufficient computational power. The network was developed in PyTorch 2.0, based on the MobileNetV3-Small architecture, initialized with pre-trained ImageNet weights to serve as the primary backbone. This transfer learning setup allowed the model to benefit from already established low-level visual features, facilitating stronger generalization and making it adaptable to the specific characteristics of SAR imagery.

For experimentation, a total of 8,000 Sentinel-1 image tiles were employed, divided into 70% for training, 15% for validation, and 15% for testing. All images underwent normalization using the ImageNet mean and standard deviation values, and were resized to 224×224 pixels to ensure compatibility with the transfer learning framework. Given the relatively limited dataset, several data augmentation techniques, such as rotations, contrast modification, and random horizontal/vertical flips, were applied to improve robustness and reduce overfitting.

Model training was performed using the Adam optimizer with an initial learning rate of 1×10^{-3} , a batch size of 32, and the cross-entropy loss function. To maintain generalization and stable convergence, both early stopping and a learning rate scheduler were employed. The model was configured to train for up to 50 epochs, but the early stopping mechanism halted the process at epoch 47 when no further improvement was observed on the validation set. Throughout training, checkpoints were periodically saved, and both loss and accuracy metrics were recorded for subsequent evaluation and analysis.

RESULTS AND DISCUSSION

A) TRAINING RESULTS

The training process exhibited consistent and stable convergence throughout. During the initial epochs, both training and validation losses decreased rapidly, indicating effective early learning. As training progressed, the model entered a stage of gradual refinement, with validation accuracy surpassing 90% between epochs 10 and 20. The convergence process stabilized in the later stages, guided by the automated learning rate scheduler, which facilitated smooth fine-tuning. Ultimately, the model reached a peak validation accuracy of 95.75% at epoch 47, confirming both effective optimization and strong generalization performance. This result shows a strong performance and robust generalization.

The accuracy and loss curves demonstrate that the training and validation metrics follow an almost parallel paths, providing a clear image on the model's robust generalization ability. This stability confirms the success of the CBAM attention module's integration to the lightweight MobileNetV3, which enhance its feature extraction without causing destabilizing to the architecture efficiency. The result is a clear prove that model exhibits a strong performance while avoiding overfitting problems, making it well- suited for lightweight embedded environments. The early stopping of training process at the epoch 47, even before reaching the maximum epoch, validates the architecture's computational efficiency with achieving peak performance.

The progressive increasing the accuracy's gain captured by the checkpoint analysis, evolving from 89% on epoch 10 to 93.5% on epoch 16 to reaching the peak of 95.4% on epoch 24, confirms the model's consistent capability to successfully learn to decode and extract the informative spatial-spectral features from SAR inputs, resulting to a classifier that achieve high performances for all target land cover classes.

Table 2. Performance Metrics for the Proposed Model

Class	Precision	Recall	F1-score	Support
Agriculture (0)	0.939	0.917	0.927	300
Urban (1)	0.898	0.913	0.906	300
Grass Land (2)	0.944	0.957	0.950	300
Bare Land (3)	0.973	0.967	0.970	300
Overall Accuracy	–	–	–	93.8%
Macro Avg	0.939	0.938	0.938	–
Weighted Avg	0.939	0.938	0.938	–

B) PERFORMANCE METRICS

To comprehensively evaluate the effectiveness of the proposed hybrid lightweight model that combines MobileNetV3 and CBAM attention module, various quantitative performance metrics were applied. The initial indicator was the classification accuracy as an overall assessment for the model's performance. However, due to dealing imbalanced classes' dataset, the incorporation of recall, precision, the F1-score into the analysis is necessary to assess the model's weaknesses and strength more discriminatively. As shown in [Figure 5], the model achieved high accuracy across all land cover classes. The highest misclassification, although minimal, was observed between "Urban" and "Bare Land". The metrics mentioned were all derived from a detailed analysis of the confusion matrix and detailed in [Table 2], demonstrating a high precision, F1-score, and recall across all categories, where each class reaching over 0.90, providing an insight of the network's high predictions, and allowing a per- class performance assessment and error prediction.



Figure 5. Confusion Matrix

C) COMPARISON AND ANALYSIS

To comprehensively evaluate the lightweight hybrid MobileNetV3 and CBAM model propose an effective approach for SAR image classification, achieving a high validation accuracy of 95.4%. The deployment of the Convolutional Block Attention Module (CBAM) to the lightweight MobileNetV3 model, allows efficient spatial-channel feature representations while minimizing computationally low demand compared to traditional convolutional methods. These results highlight that the proposed architecture successfully combines high performance with computational efficiency, making it a practical solution for real-time and embedded environmental monitoring applications.

To gain a deeper understanding of the model's behavior, the training and validation performance were systematically analyzed and compared across multiple epochs. The training and validation accuracy curves is shown in [Figure 6], whereas [Figure 7] illustrates the corresponding loss curves, collectively demonstrated that both accuracy and loss performance stabilized after the 20th epoch. The minimal accuracy gap and the parallel nature of these curves at the fine-tuning phase prove that the proposed architecture maintained strong performance and generalization capability.

Further quantitative measures were obtained from the validation set performance to evaluate the model's effectiveness, the results are detailed in [Table 2]. All classes showed promising results, with recall, precision, and F1-scores consistently exceeding 0.93, indicating the strong performance of the model's predictive ability across the different terrain classes. A minor increasing in recall was observed for certain classes such as "Urban Land" class. This is likely due to the complex scattering mechanisms and the overlapping of textural patterns of urban zones in SAR inputs, which can result to confusion with other classes such as "Bare Land", a common challenge in SAR image classification.

The confusion matrix presented in [Figure 5] offers an overall visualization of the model's classification performance of each land- cover category. The results confirm the overall correct prediction of the model; however, predominant misclassification arise between classes with similar backscattering properties, notably between "Urban" and "Bare Land" areas, due to the possibility of overlapping of textural patterns in SAR images. Despite these expected limitations, the architecture proved its effectiveness by achieving high true positive performance rate across the confusion matrix for all terrain categories, particularly dealing with the complex spectral patterns and the inherent noise of the SAR dataset. As illustrated in [Figure 8], the per-class ROC analysis prove the model's high performance across all classes including Agriculture (0), Urban (1), Grass Land (2), and Bare Land (3); this is supported by AUC indicators of 0.94, 0.92, 0.98, and 0.89 for classes respectively.

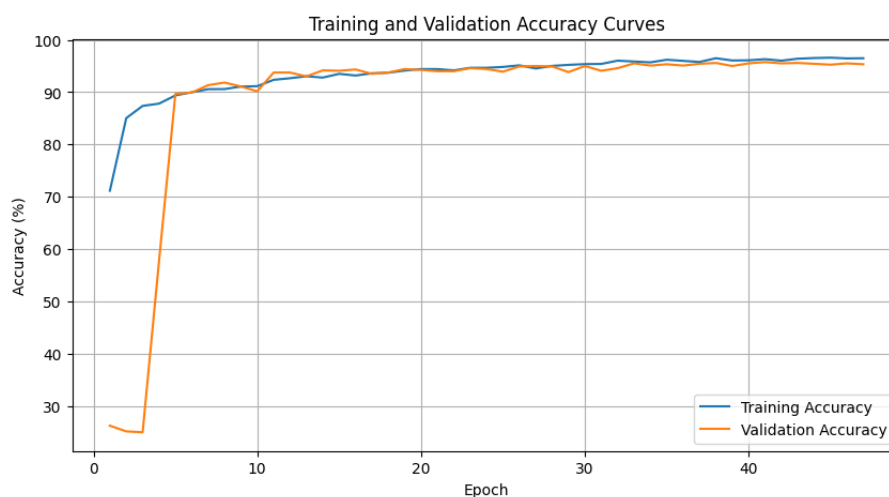


Figure 6. Training vs Validation Accuracy Curves

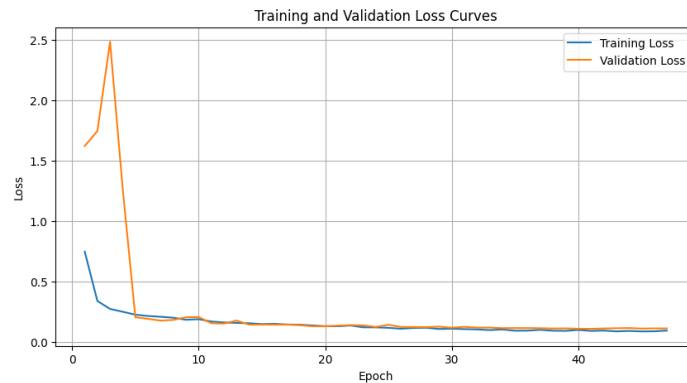


Figure 7. Training vs Validation Loss Curves

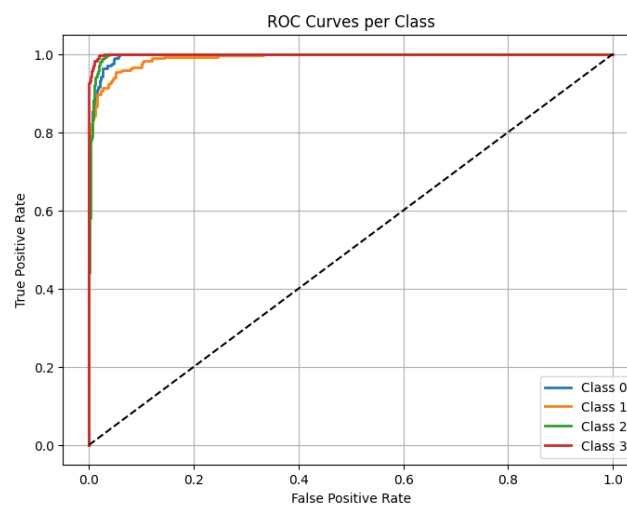


Figure 8. ROC curves per Class

A comparative evaluation was executed against the original baselines MobileNetV3- Small ResNet18, to further evaluate the proposed model. The comparison revealed a clear improvement in performance, presented in [Table 3]. The proposed hybrid architecture MobileNetV3 – CBAM achieved a 2.75% higher validation accuracy compared to the original baseline MobileNetV3- Small and a 4.5% higher validation accuracy than the ResNet model. The model achieved higher performance while minimizing computational parameter by 35% less than ResNet model. The obtained results confirm the effectiveness the integration of a lightweight attention technique in a model enhance its feature extraction without causing increasing in the computational complexity.

Table 3. Comparison of lightweight CNN models on the Sentinel- 1 SAR dataset

Model Architecture	Parameters (M)	Precision	F1-score	Recall	Validation Accuracy (%)	Model Size (MB)
ResNet18	11.7	0.89	0.89	0.90	91.2	44.3
MobileNetV3-Small	2.5	0.92	0.91	0.91	93	11.2
MobileNetV3-	2.7	0.939	0.938	0.94	95.75	12.0

Small- CBAM (proposed)						
---------------------------------------	--	--	--	--	--	--

To further examine the model's performance, the training and validation losses curves across the full 50 epochs, highlighting early stopping at epoch 47. The plots demonstrate a stable and synchronous smooth convergence, with no sign of divergence, proving the efficiency of the learning rate scheduling and regularization techniques, and preventing the overfitting. Consequently, the proposed architecture achieved strong performance reflecting an ideal balance between variance and bias. Furthermore, comprehensive measurements including training and validation accuracy, and training and validation loss were systematically saved at epochs 10, 16, 24, 30, offering detailed analyses of models development and variation, see [Table 4].

Table 4. Accuracy and Loss Evolution across Epochs

Epoch	Train Accuracy (%)	Val Accuracy (%)	Train Loss	Val Loss
10	90.84	92.17	0.1905	0.1711
16	93.18	93.83	0.1502	0.1415
24	95.61	94.33	0.1013	0.1249
30	96.20	94.75	0.0882	0.1211

The analysis of computational demand of the model was conducted to measure training time and model footprint. The hybrid model required approximately 47 seconds per epoch when using a Tesla T4 GPU on Google Colab, with a compact model size of less than 12 MB, which is quarter the size of the traditional convolutional CNN architectures that can exceed 50 MB. This minimal computational requirement makes the proposed model well-suited for implementation in resource-limited embedded application, such as environmental land cover onboard monitoring systems.

In conclusion, the proposed hybrid MobileNetV3 and CBAM model confirms its efficient framework for SAR image classification. The results obtained prove that attention-based lightweight architectures achieve stronger performance compared to deeper CNN models, and still remain computationally efficient. The conducted evaluation of the proposed combination technique and its learning behavior suggests that this architecture can be an ideal foundation for future work, including the implementation of multi-temporal SAR inputs for change monitoring, the fusion of high-resolution optical images with SAR data for more detailed land cover maps, and improving a mobile-friendly transformer-based lightweight backbone for better generalization performance.

CONCLUSION AND FUTURE WORK

In this study, a deep learning model for SAR image classification was developed, ensuring high performance and computational efficiency. It is a hybrid lightweight model which combines a Convolutional Block Attention Module (CBAM) and the MobileNetV3-Small backbone. The aim of this study was to balance strong accuracy with minimal computational requirements, a key point for implementation in real-time monitoring applications such as drone-based and embedded systems. Experimental results achieved a validation accuracy of 95.4%, ensuring a performance gain without the need to sacrifice the model's lightweight nature.

This proposed model confirms that implementing attention techniques into lightweight network is a strong method to enhance extraction features for the limited and complex nature of SAR images. On the lightweight foundation provided by MobileNetV3's depthwise separable convolutions, CBAM's adaptive feature recalibration works on

successfully direct the network to emphasize informative spatial and spectral features within the complex SAR data. This method introduces an efficient and practical solution for embedded systems and land cover monitoring classification, settings where energy and time are limited.

Despite the promising outcomes, this study has certain limitations remaining, particularly three major ones. The dataset utilized with the proposed architecture, while diverse, is still considered modest in scale compared to large-scale Earth observation databases. Additionally, the dataset was limited to four land cover classes: agricultural land, bare land, urban, grassland areas. The model was trained on a stable spatial resolution which indicates that its performance could further improve by implementing multi-temporal SAR stacks and sequences from other sensors. These elements give a room for improvement opportunities such as implementing dataset expansion, time-series and optical data, and the development scale invariant architecture.

Future work will be pursued along different promising objectives. Firstly, the dataset's size and diversity will be expanded by collecting SAR data from wider range of spatial regions and learning parameters to improve model stability. Second, we will experiment with a cross-modal dataset that combine SAR images with high-resolution optical imagery using transfer learning techniques, which will allow to adapt from contextual and spectral patterns. Third, we will investigate lightweight transformer-based attention module and hybrid convolutional transformer architecture, in order to increase accuracy without compromising efficiency. Finally, a key point will be to practically implement and test on embedded systems and UAV platforms to consider a real-time environmental monitoring application.

REFERENCES

- [1] Ardila, J., Laurila, P., Kourkouli, P., & Strong, S. (2022). Persistent monitoring and mapping of floods globally based on the Iceye SAR imaging constellation. In *IGARSS 2022 - 2022 IEEE International Geoscience and Remote Sensing Symposium* (pp. 6296-6299). Kuala Lumpur, Malaysia. <https://doi.org/10.1109/IGARSS46834.2022.9883587>
- [2] Balz, T., & Zhang, L. (2020). SAR for land use and land cover classification: A review. *ISPRS Journal of Photogrammetry and Remote Sensing*, 168, 208-224. <https://doi.org/10.1016/j.isprsjprs.2020.08.010>
- [3] Chen, X., Zheng, Y., Peng, J., & Floris, M. (2021). Monitoring river water level using multiple bounces of bridges in SAR images. *Advances in Space Research*, 68(10), 4016-4023. <https://doi.org/10.1016/j.asr.2021.08.006>
- [4] Han, K., Wang, Y., Tian, Q., Guo, J., Xu, C., & Xu, C. (2020). GhostNet: More features from cheap operations. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (pp.1580-1589). <https://doi.org/10.1109/CVPR42600.2020.00165>
- [5] Han, Y., & Ma, W. (2021). Automatic monitoring of water pollution based on the combination of UAV and USV. In *2021 IEEE 4th International Conference on Electronic Information and Communication Technology (ICEICT)* (pp. 420-424). Xi'an, China. <https://doi.org/10.1109/ICEICT53123.2021.9531204>
- [6] Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., ... & Adam, H. (2017). MobileNets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint*. <https://arxiv.org/abs/1704.04861>
- [7] Howard, A., Sandler, M., Chu, G., Chen, L.-C., Chen, B., Tan, M., ... & Adam, H. (2019). Searching for MobileNetV3. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)* (pp. 1314-1324). <https://doi.org/10.1109/ICCV.2019.00140>
- [8] Hu, J., Shen, L., & Sun, G. (2018). Squeeze-and-excitation networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 7132-7141). <https://doi.org/10.1109/CVPR.2018.00745>
- [9] Khan, A., Vibhute, A. D., Mali, S., & Patil, C. H. (2022). A systematic review on hyperspectral imaging technology with a machine and deep learning methodology for agricultural applications. *Ecological Informatics*, 69, 101678. <https://doi.org/10.1016/j.ecoinf.2022.101678>
- [10] Lee, J.-S., & Pottier, E. (2017). *Polarimetric radar imaging: From basics to applications* (2nd ed.). CRC Press. <https://doi.org/10.1201/9781420054989>
- [11] Li, H., Huang, H., Chen, L., Peng, J., Huang, H., Cui, Z., ...Wu, G. (2021). Adversarial examples for CNN-based SAR image classification: An experience study. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 14, 1333-1347. <https://doi.org/10.1109/JSTARS.2020.3038683>

- [12] McNairn, H., Duguay, C., Brisco, B., & Pultz, T. J. (2009). The sensitivity of C-band polarimetric SAR to crop condition. *International Journal of Remote Sensing*, 23(14), 826-843. <https://doi.org/10.1080/01431160210129466>
- [13] Moser, G., Serpico, S. B., & Zerubia, J. B. (2019). SAR image analysis for land-cover classification. *IEEE Transactions on Geoscience and Remote Sensing*, 57(5), 2531-2549. <https://doi.org/10.1109/tgrs.2018.2875782>
- [14] Moyano, G., Parizzi, A., & Eineder, M. (2016). Exploiting SAR data for earthquake damage assessment. *IEEE Transactions on Geoscience and Remote Sensing*, 54(10), 6038-6048. <https://doi.org/10.1109/TGRS.2016.2582384>
- [15] Passah, A., Sur, S., Paul, B., & Kandar, D. (2022). SAR image classification: A comprehensive study and analysis. *IEEE Access*, 10, 26319-26342. <https://doi.org/10.1109/ACCESS.2022.3151089>
- [16] Perry, T. (2017). SAR and military applications: An overview. *Defense & Security Analysis*, 33(3), 230-242. <https://doi.org/10.1080/14751798.2017.1340150>
- [17] RequiemOnk. (2021). *Sentinel-1/2 image pairs segregated by terrain* [Data set]. Kaggle. Retrieved from <https://www.kaggle.com/datasets/requiemonk/sentinel12-image-pairs-segregated-by-terrain>
- [18] Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L.-C. (2018). MobileNetV2: Inverted residuals and linear bottlenecks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 4510-4520). <https://doi.org/10.1109/CVPR.2018.00474>
- [19] Ulaby, F. T., Moore, R. K., & Fung, A. K. (2014). *Microwave remote sensing: Active and passive, Volume III: From theory to applications*. Artech House.
- [20] Wang, Y., Hernández, H. H., Albrecht, C. M., & Zhu, X. X. (2023). SAR flood mapping using attention-based lightweight CNNs. *Remote Sensing Letters*, 14(3), 295-305. <https://doi.org/10.1080/2150704X.2022.2162335>
- [21] Woo, S., Park, J., Lee, J.-Y., & Kweon, I. S. (2018). CBAM: Convolutional block attention module. In *Proceedings of the European Conference on Computer Vision (ECCV)* (pp. 3-19). Springer. https://doi.org/10.1007/978-3-030-01234-2_1
- [22] Xie, C., Zhuang, L., Guo, J., & Lei, Z. (2023). Flood monitoring from Sentinel-1 SAR images based on convolutional neural networks: A case study in Xinxiang City. In L. Yan, H. Duan, & Y. Deng (Eds.), *Advances in guidance, navigation and control: Proceedings of the ICGNC 2022* (Vol. 845, pp. 613-623). Springer. https://doi.org/10.1007/978-981-19-6613-2_60
- [23] Ygorra, B., El Harti, A., Lhissou, R., & Mougin, E. (2021). Deforestation monitoring using Sentinel-1 SAR images in humid tropical areas. In *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS* (pp. 5957-5960). Brussels, Belgium. <https://doi.org/10.1109/IGARSS47720.2021.9554698>
- [24] Ygorra, B., El Harti, A., Lhissou, R., & Mougin, E. (2022). Classification and deforestation monitoring using Sentinel-1 C-SAR images in a temperate exploited pine forest. In **IGARSS 2022 - 2022 IEEE International Geoscience and Remote Sensing Symposium** (pp. 691-694). Kuala Lumpur, Malaysia. <https://doi.org/10.1109/IGARSS46834.2022.9884389>
- [25] Zhang, X., Wang, Y., & Xu, Q. (2022). Lightweight attention networks for SAR classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 15, 3215-3227. <https://doi.org/10.1109/JSTARS.2021.3135566>
- [26] Zhang, X., Wang, Y., & Xu, Q. (2021). Synthetic aperture radar for Earth observation: An overview of applications. *Remote Sensing*, 13(4), 713. <https://doi.org/10.3390/rs13040713>
- [27] Zhang, X., Zhou, X., Lin, M., & Sun, J. (2018). ShuffleNet: An extremely efficient convolutional neural network for mobile devices. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 6848-6856). <https://doi.org/10.1109/CVPR.2018.00716>