SUPPLEMENTARY MATERIALS

Sentinel-1 SAR-based Globally Distributed Co-Seismic Landslide Detection by Deep Neural Networks

Lorenzo Nava[a,b,c], Alessandro Mondini[d], Kushanav Bhuyan[c,e], Chengyong Fang[e], Oriol Monserrat[f], Alessandro Novellino[g] and Filippo Catani[c]

[a] Department of Earth Science, University of Cambridge, Cambridge, UK; ln413@cam.ac.uk;

[b] Department of Geography, University of Cambridge, Cambridge, UK; ln413@cam.ac.uk;

[c] Machine Intelligence and Slope Stability Laboratory, Department of Geosciences, University of Padova, Padova, Italy; kushanav.bhuyan@phd.unipd.it; filippo.catani@unipd.it;

[d] National Research Council, Research Institute for Applied Mathematics and Information Technologies, Genova, Italy; alessandro.mondini@ge.imati.cnr.it;

[e] State Key Laboratory of Geohazard Prevention and Geoenvironment Protection, University of Technology, 610059, Chengdu, China; 2475372620@qq.com;

[f] Department of Remote Sensing, Geomatics Research Unit, Centre Tecnologic de Telecomunicacions de Catalunya (CTTC), 08860 Barcelona, Spain; omonserrat@cttc.cat;

[g] British Geological Survey, Nicker Hill, Keyworth, Nottingham, NG12 5GG, UK; alessn@bgs.ac.uk;

* Correspondence: ln413@cam.ac.uk;

1. Copernicus Global Land Cover Layers (CGLS-LC100) Collection 3

The Copernicus Global Land Cover Layers (CGLS-LC100) Collection 3 (https://developers.google.com/earth-engine/datasets/catalog/COPERNICUS_Landcover_100m_Proba-V-C3_Global#dois) provides a comprehensive land cover classification system with 23 main discrete classes. This classification is aligned with the UN Food and Agriculture Organization's (FAO) Land Cover Classification System (https://www.fao.org/documents/card/en/c/c41f08a4-e612-45d8-b569-b751f27a3542/).

2. Models Description and Comparison

2.1 ResNet-Based Model

The ResNet-based model employs residual learning to mitigate the vanishing gradient problem, enabling effective training of deep neural networks. Its architecture is built around Residual Blocks, which utilize skip connections to bypass intermediate layers and facilitate gradient flow. Each Residual Block contains two convolutional layers with 3×3 kernels, followed by Batch Normalization and ReLU activation. The output of the second convolutional layer is added to the block's input via the skip connection, allowing the network to learn residual mappings. A ReLU activation is then applied to the summed output, enhancing non-linear feature learning. The network begins with a convolutional layer for initial feature extraction, followed by a sequence of Residual Blocks interspersed with MaxPooling layers to reduce spatial dimensions. A Dropout layer is incorporated after the feature extraction process to prevent overfitting. The model concludes with a fully connected layer, which uses a sigmoid activation function to produce a binary classification output.

2.2 CBAM-Based Model

The CBAM-based model enhances feature learning by integrating the Convolutional Block Attention Module (CBAM) into a convolutional architecture. CBAM refines feature maps through a combination of Channel Attention and Spatial Attention, enabling the model to focus adaptively on the most relevant regions and channels of the input. This attention mechanism significantly improves the network's ability to capture salient patterns, particularly in tasks where feature importance varies spatially or across channels. The architecture begins with standard convolutional layers for hierarchical feature extraction, each followed by Batch Normalization and ReLU activation for training stability. MaxPooling layers are interspersed to downsample spatial dimensions while

retaining essential features. Intermediate feature maps are resized and concatenated to create a rich representation of the input, which is then passed through the CBAM module for adaptive refinement. Channel Attention emphasizes the most critical feature channels by computing global spatial statistics using average pooling, followed by dense layers to generate attention weights. These weights are applied to the input feature maps to enhance relevant channels while suppressing less important ones. Spatial Attention further refines these features by identifying important spatial regions using pooled information from all channels, which is processed through a 7×7 convolutional layer to produce spatial attention weights. The CBAM-refined feature maps are then passed through a Dropout layer for regularization, followed by dense layers to learn high-level abstractions. The final output layer uses a sigmoid activation for binary classification. The model is trained using the Adam optimizer and binary cross-entropy loss, ensuring efficient learning and robust generalization. The integration of CBAM introduces minimal computational overhead while substantially enhancing the model's representational power. This architecture is particularly effective in tasks requiring selective focus on significant features or regions, resulting in improved classification performance (Tang et al., 2021).

2.3 Performance Evaluation of Architectures

Table S1: Median and standard deviation of the accuracy, precision, recall, and F1-score of the different architectures trained on the 60-12 combination of the VV_VH dataset. They are derived from the hyperparameters and training imbalance tuning computed comparing the predictions against the unseen composite test set.

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
CNN	96.04 ± 0.2	81.58 ± 2.4	80.17 ± 2.5	80.94 ± 1.3
CBAM	95.92 ± 0.2	80.75 ± 2.8	80.41 ± 3.1	80.47 ± 0.8
ResNet	96.43 ± 0.2	84.17 ± 3.4	81.94 ± 2.3	82.53 ± 0.9

Table A1 presents the median and standard deviation of accuracy, precision, recall, and F1-score for the CNN, CBAM, and ResNet architectures, evaluated on the 60-12 combination of the VV_VH dataset. These metrics were derived from hyperparameter-tuned models and adjusted for training imbalances, with predictions compared against the unseen composite test set. Among the architectures, ResNet achieved the highest performance across all metrics, with an accuracy of 96.43 ± 0.2%, precision of 84.17 ± 3.4%, recall of 81.94 ± 2.3%, and an F1-score of 82.53 ± 0.9%. The CNN, however, performed competitively, with an accuracy of 96.04 ± 0.2%, precision of 81.58 ± 2.4%, recall of 80.17 ± 2.5%, and an F1-score of 80.94 ± 1.3%, slightly outperforming CBAM in both accuracy and F1-score. CBAM achieved an accuracy of 95.92 ± 0.2%, precision of 80.75 ± 2.8%, recall of 80.41 ± 3.1%, and an F1-score of 80.47 ± 0.8%. Despite ResNet's superior overall performance, the decision to use the baseline CNN for iterative sensitivity experiments was driven by several factors. First, the simpler architecture of the CNN makes it computationally more efficient for large-scale experiments, requiring less training time and memory compared to CBAM and ResNet. This efficiency is particularly advantageous when conducting extensive parameter sweeps or testing on diverse datasets, as it allows for rapid iteration and experimentation without significant resource overhead. Additionally, the CNN provides a strong baseline with consistent and competitive performance across all metrics, demonstrating its ability to generalize well without the additional complexity of attention mechanisms or residual learning. By utilizing the CNN, we ensure that our experimental framework remains accessible and reproducible while maintaining robust performance. The comparative analysis highlights the incremental benefits of advanced architectures such as CBAM and ResNet, particularly for tasks requiring high precision and F1-scores. However, for the purposes of this study, the balance between simplicity, computational efficiency, and reliable performance positions the baseline CNN as the preferred choice for the majority of experiments.

Additional Figures
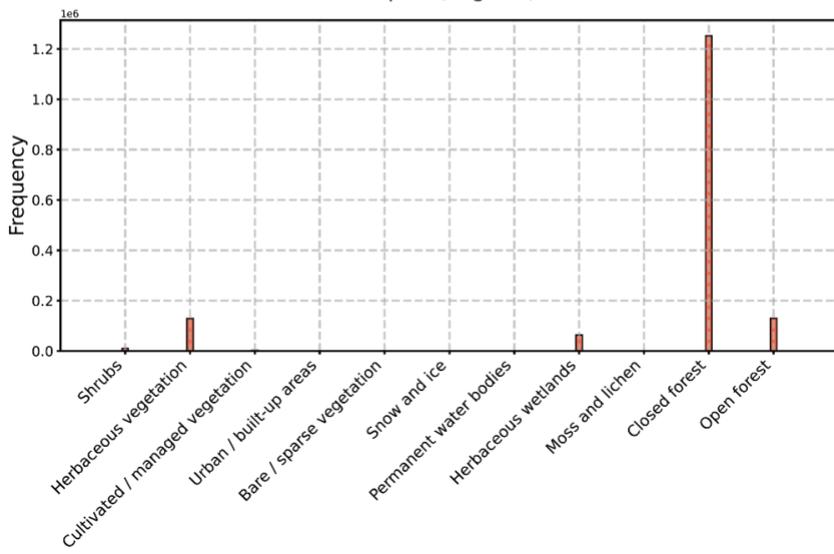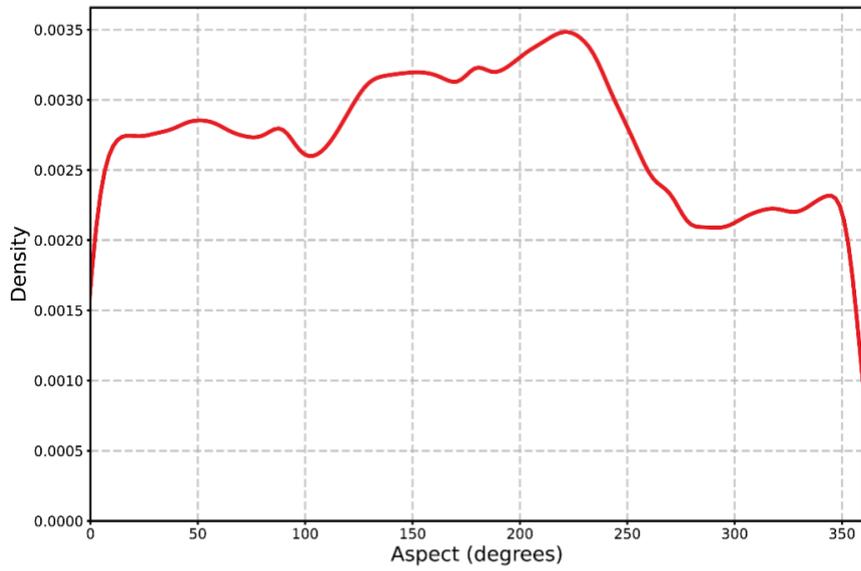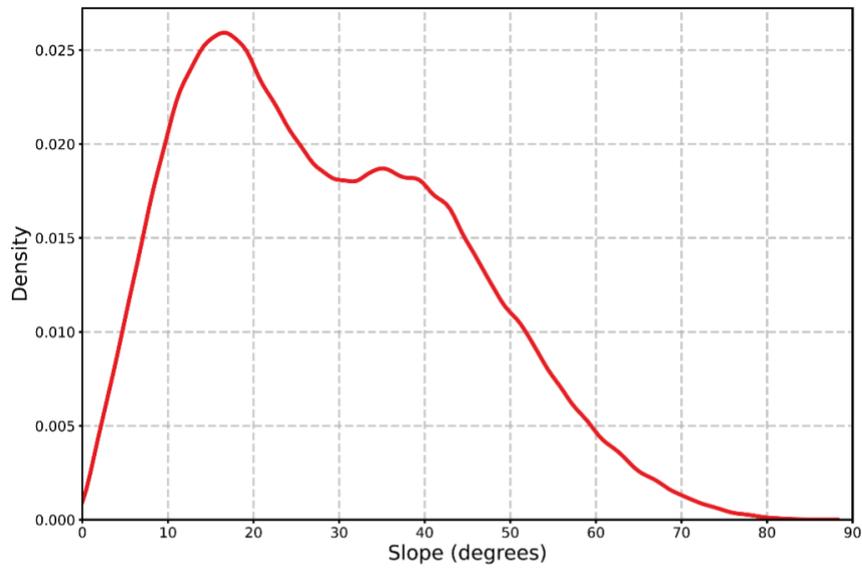
Figure S2: Distribution of majority class of landcover in the test dataset for the classification results of the 60_12 temporal stack combination in the six study areas used to perform the comparison between the VV and VV_VH combinations.
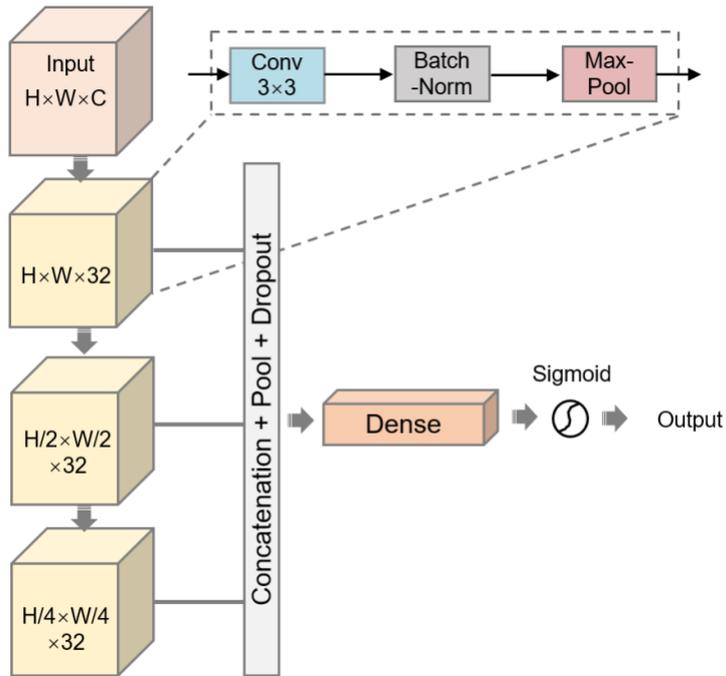
Figure S3: *CNN model architecture used.*

References

X. Tang, M. Liu, H. Zhong, Y. Ju, W. Li, and Q. Xu, "Mill: Channel attention-based deep multiple instance learning for landslide recognition," ACM Trans. Multimedia Comput. Commun. Appl., vol. 17, 2021. DOI: 10.1145/3454009.