

AlignSAR: An open-source toolbox of SAR benchmark dataset creation for machine learning applications

Ling Chang¹, Xu Zhang¹, Bingquan Li¹, Anurag Kulshrestha², Milan Lazecky³, Serkan Girgin¹, and Jose Manuel Delgado Blasco⁴

¹ University of Twente, The Netherlands ² ICEYE, Finland ³ University of Leeds, United Kingdom ⁴ Microgeodesia Research Group, University of Jaen, Spain ¶ Corresponding author

DOI: [10.xxxxxx/draft](https://doi.org/10.xxxxxx/draft)

Software

- [Review](#)
- [Repository](#)
- [Archive](#)

Editor: [✉](#)

Submitted: 02 September 2025

Published: unpublished

License

Authors of papers retain copyright and release the work under a Creative Commons Attribution 4.0 International License ([CC BY 4.0](#)).

Summary

Benchmark datasets are essential for the development, evaluation, and comparison of machine learning models. However, in the field of Synthetic Aperture Radar (SAR) remote sensing (Curlander & McDonough, 1991), there is a notable lack of large and standardized benchmark datasets, as reported by Long et al. (2021) and Zhu et al. (2021). Moreover, there is currently no open reference library of methods and tools for creating SAR benchmark datasets. To address these gaps, we developed AlignSAR, a comprehensive toolbox designed to create benchmark datasets specifically from SAR imagery, optimized for machine learning applications.

AlignSAR enables users to process both legacy and contemporary SAR datasets, including those from Envisat, Radarsat, and Sentinel-1, facilitating the extraction of representative SAR signatures. Additionally, the toolbox supports the alignment of other geospatial observations, such as LiDAR and optical satellite data, onto SAR reference grids, thereby enriching each SAR pixel with additional signatures. The SAR benchmark datasets created with AlignSAR, constructed from these enriched and aligned signatures, are then ready for use in a variety of machine-learning tasks and can be shared easily on the Cloud, e.g. by using STAC (SpatioTemporal Asset Catalogs).

Statement of Need

SAR techniques, including InSAR (Interferometric SAR) (Bamler & Hartl, 1998) and PolSAR (Polarimetric SAR) (Lee & Pottier, 2017), are well-established and are employed in natural and anthropogenic hazard monitoring, as well as land use and land cover classification. These techniques typically employ conventional methods such as persistent scatterer interferometry and the small baseline subset approach (Ferretti et al., 2001; Lanari et al., 2004), recently in conjunction with machine learning techniques. A variety of commercial and open-source processing tools have been developed to support different SAR applications. For instance, The ESA SentiNel Application Platform SNAP, a JAVA-based tool, which contains the Sentinel-1 Toolbox that is commonly used for standard InSAR and PolSAR decomposition analysis (European Space Agency (ESA), 2014). Doris, stands for Delft object-oriented radar interferometric software, and is another open-source tool designed for InSAR analysis, though it is somewhat outdated and has been partially deprecated (Delft University of Technology (TUDelft), 2008). On the commercial side, Gamma (GAMMA Remote Sensing, 1995) offers a robust solution for InSAR analysis, also integrated into platforms like ASF Services – On Demand Processing (NASA Distributed Active Archive Center (DAAC), 2017) and COMET LiCSAR System (Lazecky et al., 2020).

As an increasing number of dedicated SAR missions are launched, the community of SAR users is expanding, with growing interest in applying SAR data within machine learning

frameworks (Kulshrestha et al., 2022; Ma et al., 2024; Zhu et al., 2021). Despite this trend, the aforementioned tools are not specifically designed for machine learning applications. Furthermore, directly applying methods developed for other types of signals and images to SAR data, without considering SAR representative signatures, can lead to biased and inaccurate machine learning outcomes.

AlignSAR toolbox developed new methods and tools upon existing tools such as Doris and SNAP, and enhances the functionalities, optimizes key processes, and extends the capabilities of the tools used, see the details in Main Functionalities. It standardizes the extraction procedure of representative SAR signatures, and provides radar- and geo-coded geospatial observations that are all aligned onto SAR reference coordinates and then associates additional signatures from these geospatial data with the corresponding SAR pixels. The SAR signatures are classified into three categories:

Category 1: Single polarimetric signatures, like amplitude, intensity, backscatter coefficient, interferometric and SLC (Single Look Complex) phase and coherence;

Category 2: Multi polarimetric signatures, such as co-, dual polarization cross product, and summation, difference and ratio of the co-, dual, quad polarization intensities, and entropy, scattering mechanisms, SAR vegetation index (RVI), and SAR soil moisture.

Category 3: Inherited attributes from additional geospatial observations, e.g. land use land cover type, cadastral features, temperature, atmospheric phase, and geological information.

Source Code and Tutorial

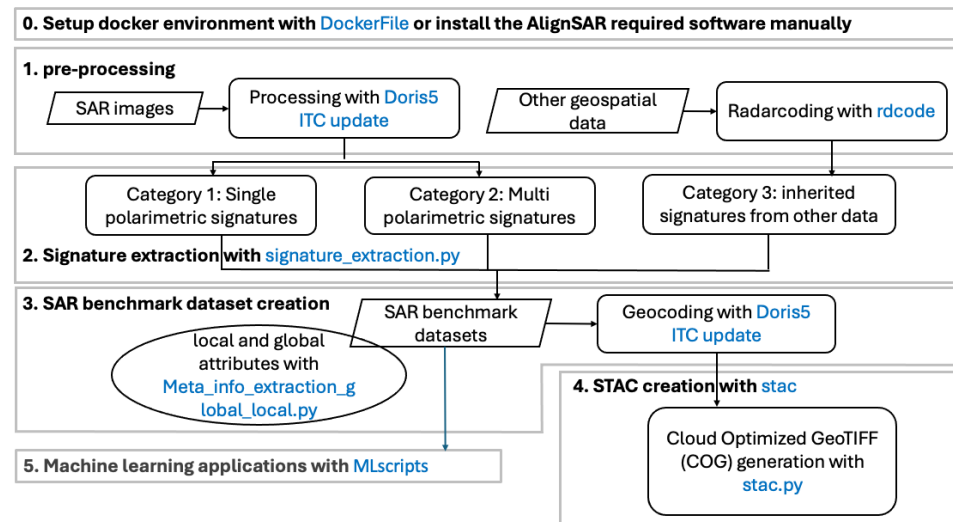
AlignSAR is available as an Apache-2.0-licensed open-source toolbox hosted on [GitHub](#). The [AlignSAR tutorial](#) outlines the methodology and functionality structure employed in this project, describes all the open-source tools developed within this toolbox, and uses a use case with [Sentinel-1 SAR](#) and [TOP10NL](#) data for demonstration.

Main Functionalities and Processing Flowchart

1. [DockerFile](#): docker file with operating system and software setup. The docker file can automatically install Doris-5 and relevant software.
2. [Doris5ITCupdate](#): an updated version of Doris that is used to read raw SAR (Single Look Complex format), coregister all SAR images to the same reference grids, resample them, remove the reference and topographic phase using external Digital Elevation Model (DEM) (e.g. [SRTM](#) or [Copernicus DEM](#)), and generate co-registered amplitude, SLC phase and interferometric phase data.
3. [rdcode](#): the radarcoding scripts to align other geospatial data onto SAR reference grid. For method details refer to Kulshrestha et al. (2024).
4. [signature_extraction.py](#): a script to extract representative SAR signatures. For method details refer to Zhang et al. (2025).
5. [Meta_info_extraction_global_local.py](#): a script to extract global/local attributes from Sentinel-1 SAR metadata.
6. [stac](#): python scripts to create STAC (SpatioTemporal Asset Catalogs). Our STAC examples are available via [STAC Browser](#) by specifying the STAC Catalog with 'https://data.crib.utwente.nl/stac/alignsar/catalog.json'.
7. Post analysis demonstration: [MLscripts](#): scripts for machine learning analysis. [Yolov8](#), ANN (Artificial Neural Network) and Siamese are separately used for Object Detection (India), Land Use Land Cover classification (Netherlands), and Change Detection (Poland). And [jupyter_notebook_demo](#): a jupyter notebook to demonstrate how to

87 extract, visualize and analyse SAR signatures, with the sample output generated by
88 Doris.

89 Figure 1 illustrates the processing flowchart for creating SAR benchmark datasets, with
90 key functionalities highlighted in blue. The flowchart consists of six main stages: en-
91 vironment and software setup, pre-processing, signature extraction, SAR benchmark
92 dataset generation, STAC creation, and machine learning application demonstration.



93
94 Figure 1: Processing flowchart.

95 Acknowledgements

96 The authors extend sincere gratitude to the ESA (European Space Agency) for funding
97 the development of AlignSAR through the Open SAR Library project (4000140600/23/I-
98 DT) and the ESA Network of Resources. We also thank all project members, listed on
99 <https://www.alignsar.nl/alignsar-team/> for their invaluable contributions to the development,
100 testing and demonstration of the methods in this project.

101 References

- 102 Bamler, R., & Hartl, P. (1998). Synthetic aperture radar interferometry. *Inverse Problems*,
103 14(4), R1. <https://doi.org/10.1088/0266-5611/14/4/001>
- 104 Curlander, J. C., & McDonough, R. N. (1991). *Synthetic aperture radar: Systems and signal*
105 *processing* (Vol. 11). Wiley, New York. ISBN: 9780471857709
- 106 Delft University of Technology (TUDelft). (2008). *DORIS: The delft object-oriented radar*
107 *interferometric software*. <https://doris.tudelft.nl/>
- 108 European Space Agency (ESA). (2014). *Sentinel Application Platform (SNAP)*. <https://step.esa.int/>
- 109
110 Ferretti, A., Prati, C., & Rocca, F. (2001). Permanent scatterers in SAR interferometry. *IEEE*
111 *Transactions on Geoscience and Remote Sensing*, 39(1), 8–20. <https://doi.org/10.1109/36.898661>
- 112
113 GAMMA Remote Sensing. (1995). *GAMMA software*. <https://www.gamma-rs.ch/software>
- 114 Kulshrestha, A., Chang, L., & Stein, A. (2022). Use of LSTM for sinkhole-related anomaly

- 115 detection and classification of InSAR deformation time series. *IEEE Journal of Selected*
116 *Topics in Applied Earth Observations and Remote Sensing*, 15, 4559–4570. <https://doi.org/10.1109/JSTARS.2022.3180994>
117
- 118 Kulshrestha, A., Chang, L., & Stein, A. (2024). Radarcoding reference data for SAR training
119 data creation in Radar coordinates. *IEEE Geoscience and Remote Sensing Letters*, 21.
120 <https://doi.org/10.1109/LGRS.2024.3376992>
- 121 Lanari, R., Mora, O., Manunta, M., Mallorquí, J. J., Berardino, P., & Sansosti, E. (2004). A
122 small-baseline approach for investigating deformations on full-resolution differential SAR
123 interferograms. *IEEE Transactions on Geoscience and Remote Sensing*, 42(7), 1377–1386.
124 <https://doi.org/10.1109/TGRS.2004.828196>
- 125 Lazecký, M., Spaans, K., González, P. J., Maghsoudi, Y., Morishita, Y., Albino, F., Elliott, J.,
126 Greenall, N., Hatton, E., Hooper, A., & others. (2020). LiCSAR: An automatic InSAR
127 tool for measuring and monitoring tectonic and volcanic activity. *Remote Sensing*, 12(15),
128 2430. <https://doi.org/10.3390/rs12152430>
- 129 Lee, J. S., & Pottier, E. (2017). *Polarimetric radar imaging: From basics to applications*. CRC
130 press. <https://doi.org/10.1201/9781420054989>
- 131 Long, Y., Xia, G.-S., Li, S., Yang, W., Yang, M. Y., Zhu, X. X., Zhang, L., & Li, D. (2021).
132 On creating benchmark dataset for aerial image interpretation: Reviews, guidances, and
133 million-aid. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote*
134 *Sensing*, 14, 4205–4230. <https://doi.org/10.48550/arXiv.2006.12485>
- 135 Ma, Y., Ke, P., Hossein, A., Chang, L., & Wei, J. (2024). Despeckling SAR images with
136 Log-Yeo-Johnson transformation and conditional diffusion models. *IEEE Transactions on*
137 *Geoscience and Remote Sensing*. <https://doi.org/10.1109/TGRS.2024.3419083>
- 138 NASA Distributed Active Archive Center (DAAC). (2017). *Alaska Satellite Facility (ASF)*.
139 <https://search.asf.alaska.edu>
- 140 Zhang, X., Chang, L., Girgin, S., & Stein, A. (2025). Creating and leveraging SAR benchmark
141 datasets to facilitate machine learning application. *International Journal of Applied Earth*
142 *Observation and Geoinformation*, 142, 104722. <https://doi.org/10.1016/j.jag.2025.104722>
- 143 Zhu, X. X., Montazeri, S., Ali, M., Hua, Y., Wang, Y., Mou, L., Shi, Y., Xu, F., & Bamler,
144 R. (2021). Deep learning meets SAR: Concepts, models, pitfalls, and perspectives. *IEEE*
145 *Geoscience and Remote Sensing Magazine*, 9(4), 143–172. <https://doi.org/10.1109/MGRS.2020.3046356>
146