

EGU2020-7586

<https://doi.org/10.5194/egusphere-egu2020-7586>

EGU General Assembly 2020

© Author(s) 2023. This work is distributed under the Creative Commons Attribution 4.0 License.



## Selection of Reliable Machine Learning Algorithms for Geophysical Applications

**Octavian Dumitru**, Gottfried Schwarz, Dongyang Ao, Gabriel Dax, Vlad Andrei, Chandra Karmakar, and Mihai Datcu

Remote Sensing Technology Institute, German Aerospace Center, Münchener Strasse 20, 82234 Wessling, Germany  
(corneliu.dumitru@dlr.de)

During the last years, one could see a broad use of machine learning tools and applications. However, when we use these techniques for geophysical analyses, we must be sure that the obtained results are scientifically valid and allow us to derive quantitative outcomes that can be directly compared with other measurements.

Therefore, we set out to identify typical datasets that lend themselves well to geophysical data interpretation. To simplify this very general task, we concentrate in this contribution on multi-dimensional image data acquired by satellites with typical remote sensing instruments for Earth observation being used for the analysis for:

- Atmospheric phenomena (cloud cover, cloud characteristics, smoke and plumes, strong winds, etc.)
- Land cover and land use (open terrain, agriculture, forestry, settlements, buildings and streets, industrial and transportation facilities, mountains, etc.)
- Sea and ocean surfaces (waves, currents, ships, icebergs, coastlines, etc.)
- Ice and snow on land and water (ice fields, glaciers, etc.)
- Image time series (dynamical phenomena, their occurrence and magnitude, mapping techniques)

Then we analyze important data characteristics for each type of instrument. One can see that most selected images are characterized by their type of imaging instrument (e.g., radar or optical images), their typical signal-to-noise figures, their preferred pixel sizes, their various spectral bands, etc.

As a third step, we select a number of established machine learning algorithms, available tools, software packages, required environments, published experiences, and specific caveats. The comparisons cover traditional “flat” as well as advanced “deep” techniques that have to be compared in detail before making any decision about their usefulness for geophysical applications. They range from simple thresholding to k-means, from multi-scale approaches to convolutional networks (with visible or hidden layers) and auto-encoders with sub-components

from rectified linear units to adversarial networks.

Finally, we summarize our findings in several instrument / machine learning algorithm matrices (e.g., for active or passive instruments). These matrices also contain important features of the input data and their consequences, computational effort, attainable figures-of-merit, and necessary testing and verification steps (positive and negative examples). Typical examples are statistical similarities, characteristic scales, rotation invariance, target groupings, topic bagging and targeting (hashing) capabilities as well as local compression behavior.