**Reading #1**

1. **Citation**: Bergen, K.J., Johnson, P.A., de Hoop, M.V., & Beroza, G.C. (2019). *Machine learning for data-driven discovery in solid Earth geoscience*. Science, 363(6433). <https://doi.org/10.1126/science.aau0323>
2. **Scientific Motivation**: This paper sets out to prove that Earth scientists can and should be leveraging ML to wrangle unwieldy datasets that are too big, too messy, or too nonlinear for traditional methods to handle alone.
3. **Data Source and Type**: The authors review a broad range of data types—primarily seismic, geodetic, and remote sensing—used in solid Earth studies. each with its own flavor of high-dimensional, noisy data.
4. **Method**: The paper provides a high-level overview of how machine learning techniques are being applied to geoscience questions, with emphasis on supervised, unsupervised, and deep learning approaches.
5. **Key Findings**: ML approaches can reveal patterns and predictive relationships in high-dimensional data that are inaccessible to traditional techniques, provided issues of training, interpretability, and reproducibility are addressed.
6. **Data**: No original datasets are shared, but the emphasis on open-access repositories is not subtle.
7. **Workflow**: No specific workflow is described, but the paper encourages transparent practices and community-curated benchmarks.
8. **Code**: No accompanying code/not applicable.
9. **Open Access**: In an act of true irony, this is trapped behind the Science paywall and requires a subscription or institutional login to access.

**Reading #2**

1. **Citation**: Gil, Y., et al. (2016). *Toward the geoscience paper of the future: Best practices for documenting and sharing research from data to software to provenance*. Earth and Space Science, 3(10), 388–415. <https://doi.org/10.1002/2016EA000201>
2. **Scientific Motivation**: This paper advocates for a fundamental rethinking of how Earth science research is documented and published, with an emphasis on transparency, reproducibility, and long-term accessibility.
3. **Data Source and Type**: No original dataset—just excellent examples from across Earth science fields that showcase best practices in digital transparency.
4. **Method**: The authors propose a framework for digital scholarship that integrates datasets, code, workflows, and provenance tracking directly into scientific publications.
5. **Key Findings**: The paper of the future includes not just plots and prose, but intentional design of metadata, persistent identifiers, and documented workflows from the outset.
6. **Data**: The paper outlines data-sharing best practices and recommends institutional and discipline-specific repositories but does not provide datasets.
7. **Workflow**: Though conceptual, the paper outlines an ideal scientific workflow architecture and offers a model for how these elements should be integrated into publication.
8. **Code**: No code or specific implementation is shared.
9. **Open Access**: Fully open-access via Wiley.

**Reading #3**

1. **Citation**: Wilkinson, M.D., et al. (2016). *The FAIR Guiding Principles for scientific data management and stewardship*. Scientific Data, 3, 160018. <https://doi.org/10.1038/sdata.2016.18>
2. **Scientific Motivation:** The authors respond to the increasing need for formalized, scalable principles that enable better data reuse, especially as data volumes and diversity expand.
3. **Data Source and Type:** This is a framework paper—no specific datasets are introduced.
4. **Method:** The authors define and justify four principles—Findability, Accessibility, Interoperability, and Reusability (FAIR)—with recommendations for implementation in scientific data systems.
5. **Key Findings:** Applying the FAIR principles ensures that data is machine-readable, discoverable, and reusable, enhancing the rigor and efficiency of scientific collaboration.
6. **Data**: No new data, just strong opinions on what yours should look like.
7. **Workflow**: Conceptual workflows are discussed in terms of data lifecycle and repository infrastructure.
8. **Code**: No code or computational workflow is included.
9. **Open Access**: Fully open-access via Nature's Scientific Data journal.

**Reading #4**

1. **Citation**: Goetz, J.N., et al. (2015). *Evaluating machine learning and statistical prediction techniques for landslide susceptibility modeling*. Computers & Geosciences, 81, 1–11. <https://doi.org/10.1016/j.cageo.2015.04.007>
2. **Scientific Motivation**: The paper evaluates the comparative performance of ML algorithms in predicting landslide susceptibility, addressing a persistent challenge in geohazard assessment.
3. **Data Source and Type**: The study uses a spatially distributed landslide inventory and terrain data from a mountainous region in Austria.
4. **Method**: A suite of classifiers—random forest, support vector machines, logistic regression, and naive Bayes—are trained and tested using standard performance metrics.
5. **Key Findings**: Random forest and SVM reliably outperform the others in accuracy and resilience to class imbalance.
6. **Data**: The dataset is not publicly linked, but the paper includes sufficient metadata and preprocessing descriptions to enable reproduction with similar data.
7. **Workflow**: The model selection, feature extraction, and validation pipeline is clearly outlined.
8. **Code**: Sadly, no repo or script, but the methodology is reproducible if you’re fluent in scikit-learn.
9. **Open Access**: Not open-access; requires Elsevier access.

**Reading #5**

1. **Citation**: Unglert, K., Radić, V., & Jellinek, A.M. (2016). *PCA vs. SOM + hierarchical clustering for pattern recognition in volcano seismic spectra*. J. Volcanology and Geothermal Research, 320, 58–74.
2. **Scientific Motivation**: This paper examines how dimensionality reduction and clustering techniques can be used to better classify volcanic seismic activity, particularly in the context of eruption monitoring.
3. **Data Source and Type**: Continuous seismic spectral data from Mount Spurr volcano (Alaska) were processed and analyzed.
4. **Method**: Compares PCA (classic and clean) to a more flexible two-step approach using Self-Organizing Maps + hierarchical clustering.
5. **Key Findings**: The SOM approach captures more spectral nuance than PCA alone, and when combined with hierarchical clustering, it improves classification of eruption-related signals.
6. **Data**: Not directly linked but likely retrievable from the Alaska Volcano Observatory, and processed methods and parameter ranges are well described.
7. **Workflow**: Preprocessing, feature extraction, and clustering workflow is clearly laid out and modifiable.
8. **Code**: Code is not provided, but the methods are reproducible using common clustering librarieson MATLAB or Python.
9. **Open Access**: Not open-access; requires Elsevier access.

**Reading #9**

1. **Citation**: Shoji, D., et al. (2018). *Classification of volcanic ash particles using a convolutional neural network and probability*. Scientific Reports, 8, 1–12. <https://doi.org/10.1038/s41598-018-26200-2>
2. **Scientific Motivation**: This paper aims to improve the accuracy and speed of volcanic ash particle classification using deep learning.
3. **Data Source and Type**: Te dataset consists of thousands of manually labelled SEM images of volcanic ash particles collected from multiple eruptions.
4. **Method**: A convolutional neural network (CNN) is trained to classify images into morphological categories with a probabilistic confidence metric.
5. **Key Findings**: The CNN reached >90% classification accuracy, outperforming rule-based image classifiers in every category.
6. **Data**: Images not hosted publicly, but the classes, labeling scheme, and examples are well illustrated.
7. **Workflow**: The CNN architecture and training-validation pipeline are clearly explained, including dropout and optimization parameters.
8. **Code**: No code is linked, but the model is replicable using common CNN frameworks (e.g., TensorFlow, Keras).
9. **Open Access**: Fully open-access via Nature’s Scientific Reports.