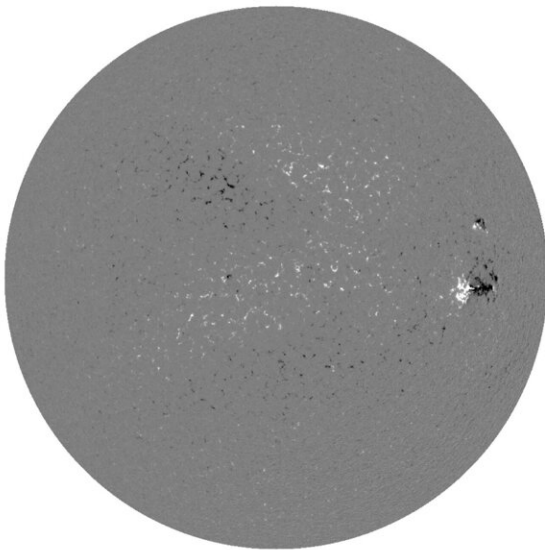


# Scientists demonstrate machine learning tool to efficiently process complex solar data

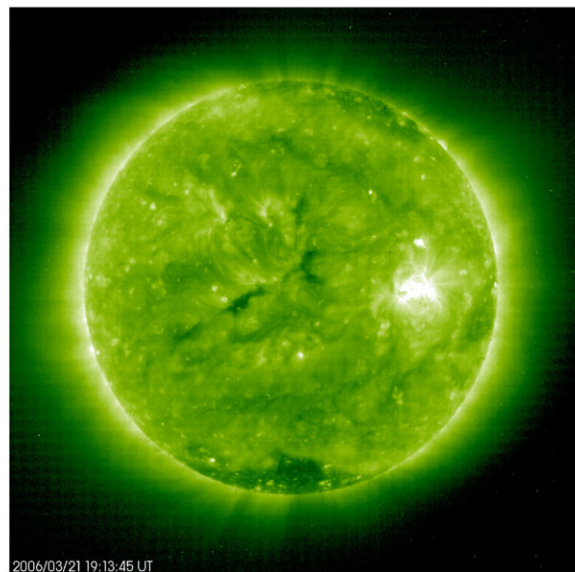
July 6 2022

---

SOHO/MDI Magnetic Image



SOHO/EIT Lower Coronal Image



Credit: Southwest Research Institute

Big data has become a big challenge for space scientists analyzing vast datasets from increasingly powerful space instrumentation. To address this, a Southwest Research Institute team has developed a machine learning tool to efficiently label large, complex datasets to allow deep learning models to sift through and identify potentially hazardous solar events. The new labeling tool can be applied or adapted to address other

challenges involving vast datasets.

As [space](#) instrument packages collect increasingly [complex data](#) in ever-increasing volumes, it is becoming more challenging for scientists to process and analyze relevant trends. Machine learning (ML) is becoming a critical tool for processing large complex datasets, where algorithms learn from existing data to make decisions or predictions that can factor more information simultaneously than humans can. However, to take advantage of ML techniques, humans need to label all the data first—often a monumental endeavor.

"Labeling data with meaningful annotations is a crucial step of supervised ML. However, labeling datasets is tedious and time consuming," said Dr. Subhamoy Chatterjee, a postdoctoral researcher at SwRI specializing in solar astronomy and instrumentation and lead author of a paper about these findings published in the journal *Nature Astronomy*. "New research shows how [convolutional neural networks](#) (CNNs), trained on crudely labeled astronomical videos, can be leveraged to improve the quality and breadth of data labeling and reduce the need for human intervention."

Deep learning techniques can automate processing and interpret large amounts of complex data by extracting and learning complex patterns. The SwRI team used videos of the solar magnetic field to identify areas where strong, complex magnetic fields emerge on the solar surface, which are the main precursor of space weather events.

"We trained CNNs using crude labels, manually verifying only our disagreements with the machine," said co-author Dr. Andrés Muñoz-Jaramillo, an SwRI solar physicist with expertise in [machine learning](#). "We then retrained the [algorithm](#) with the corrected data and repeated this process until we were all in agreement. While flux emergence labeling is typically done manually, this iterative interaction between the

human and ML algorithm reduces manual verification by 50%."

Iterative labeling approaches such as active learning can significantly save time, reducing the cost of making [big data](#) ML ready. Furthermore, by gradually masking the videos and looking for the moment where the ML algorithm changes its classification, SwRI scientists further leveraged the trained ML algorithm to provide an even richer and more useful database.

"We created an end-to-end, deep-learning approach for classifying videos of magnetic patch evolution without explicitly supplying segmented images, tracking algorithms or other handcrafted features," said SwRI's Dr. Derek Lamb, a co-author specializing in evolution of magnetic fields on the surface of the Sun. "This database will be critical in the development of new methodologies for forecasting the emergence of the complex regions conducive to space weather events, potentially increasing the lead time we have to prepare for space weather."

**More information:** Subhamoy Chatterjee et al, Efficient labelling of solar flux evolution videos by a deep learning model, *Nature Astronomy* (2022). [DOI: 10.1038/s41550-022-01701-3](https://doi.org/10.1038/s41550-022-01701-3)

Provided by Southwest Research Institute

Citation: Scientists demonstrate machine learning tool to efficiently process complex solar data (2022, July 6) retrieved 25 April 2024 from <https://phys.org/news/2022-07-scientists-machine-tool-efficiently-complex.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.