SuryaBench: Benchmark Dataset for Advancing Machine Learning Applications in Heliophysics and Space Weather

Dinesha V. Hegde[1], Sujit Roy[2,3], Johannes Schmude[4], Rohit Lal[2], Vishal Gaur[2], Amy Lin[2], Kshitiz Mandal[2], Talwinder Singh[5], Andrés Muñoz-Jaramillo[6], Kang Yang[5], Chetraj Pandey[5], Jinsu Hong[5], Berkay aydin[5], Ryan McGranaghan[7], Spiridon Kasapis[8], Vishal Upendran[9,10], Shah Bahauddin[11,12], Daniel da Silva[13], Marcus Freitag[4], Iksha Gurung[2], Nikolai Pogorelov[1,14], Campbell Watson[4], Manil Maskey[3], Juan Bernabe-Moreno[4] & Rahul Ramachandran[3]

THE UNIVERSITY OF ALABAMA IN HUNTSVILLE

1. Center for Space Plasma and Aeronomic Research (CSPAR), the University of Alabama in Huntsville (UAH), Huntsville, 2. Earth System Science Center (ESSC), UAH, Huntsville. 3. NASA Marshall Space Flight Center (MSFC), Huntsville, 4. iBM Research, Yorktown Heights, 5. Georgia State University, Atlanta, 6. Southwest Research Institute, Boulder, 7. NASA Jet Propulsion Laboratory, Pasadena, 8. Princeton University, Princeton, 9. Seti institute, Mountain View, 10. Lockheed Martin Solar and Astrophysics Laboratory, 11. Laboratory for Atmospheric and Space Physics, University of colorado Boulder, Boulder, 12. Center for Astronomy, Space Science and Astrophysics, Independent University Bangladesh, Dhaka, Bangladesh. 13. NASA Goddard Space flight center, Greenbelt, 14. Department of Space Science, the UAH, Huntsville

Abstract

We introduce SuryaBench, an open-source, machine learning (ML)-ready heliophysics dataset suit designed to advance ML applications in solar physics and space weather forecasting. The core dataset is derived from NASA's Solar Dynamics Observatory (SDO) (Pesnell et al.,2012), including processed high-resolution imagery from the Atmospheric Imaging Assembly (AIA) and Helioseismic and Magnetic Imager (HMI), spanning May 2010 to December 2024. To ensure suitability for ML tasks, the data have been preprocessed through correction of spacecraft roll angles, orbital adjustments, exposure normalization, and degradation compensation. In addition to the core SDO data, SuryaBench includes auxiliary benchmark datasets for key heliophysics and space weather tasks, including active region segmentation, active region emergence forecasting, coronal magnetic field extrapolation, solar flare prediction, solar Extreme Ultraviolet (EUV) spectra prediction, and solar wind speed estimation. SuryaBench has also been used to train Surya (Roy et al., 2025), a heliophysics foundation model, demonstrating its utility as a large-scale resource for developing generalizable AI models. By establishing a unified, standardized data collection with task-specific benchmarks, SuryaBench aims to facilitate benchmarking, enhance reproducibility, and accelerate the development of AI-driven models for space weather prediction, bridging solar physics, machine learning, and operational forecasting (Roy et al.,2026).

Motivation

Challenges in Using SDO Data for Machine Learning:

- **Inconsistent data quality:** AIA and HMI observations differ in spatial and temporal resolution.
- **Data gaps and artifacts:** Missing, corrupted, or incomplete data can reduce temporal continuity and affect model training.
- **Exposure variation and instrument degradation:** AIA exposure-time variations and long-term sensitivity degradation introduce artificial inconsistencies.
- **Solar disk size variation:** Earth's elliptical orbit causes apparent solar disk size changes.
- **Massive data volume:** It is difficult to access, process, and organize large volume of SDO data for ML workflows.
- **Preprocessing barrier:** Standard SDO preprocessing depends on specialized, domain-specific and IDL-based workflows.
- **Limited existing ML-ready data:** The public SDO ML-ready dataset by Galvez et al. 2019 is available at **512 × 512** resolution, limiting native-resolution feature analysis.
- **Need:** A unified, standardized, native-resolution, and AI-ready solar dataset for heliophysics and space weather research.

SuryaBench SDO Data Prep: JSOC -> Science Ready -> ML Ready

Atmospheric Imaging Assembly (AIA)
Observable: Coronal EUV & UV Intensity
Product/Channel: EUV- 094, 131, 171, 193, 211, 304, 335, UV-1600
Level: 1.0
Available cadence: 12s(EUV), 24s(UV)
Pixel size: ~0.6 arcsec/pixel
Unit: Data Number (DN)
Data Type: int16

Helioseismic Magnetic Imager (HMI)
Observable: Photospheric Mag Field & Velocity
Product: LOS & Vector- Magnetograms, LOS- Doppler Velocity
Level: 1.5
Available Cadence: 45s(LOS), 720s (LOS & Vec)
Pixel size: 0.5 arcsec/pixel
Unit: Gauss (G)
Data Type: int32

Data Description

Duration : May 13, 2010 - Dec 31, 2024
Image Size: 4096 x 4096 pixels
Downloaded Cadence: 12min
Data Format: Rice Compressed FITS

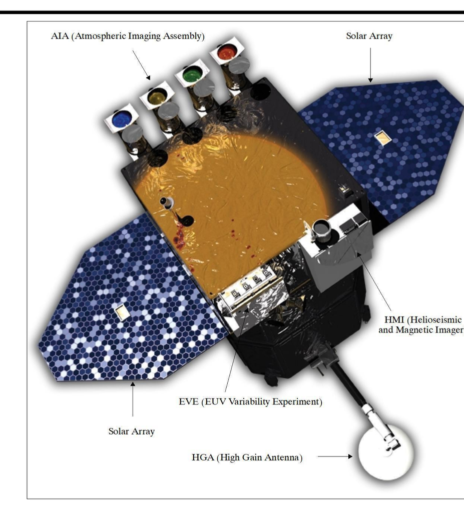
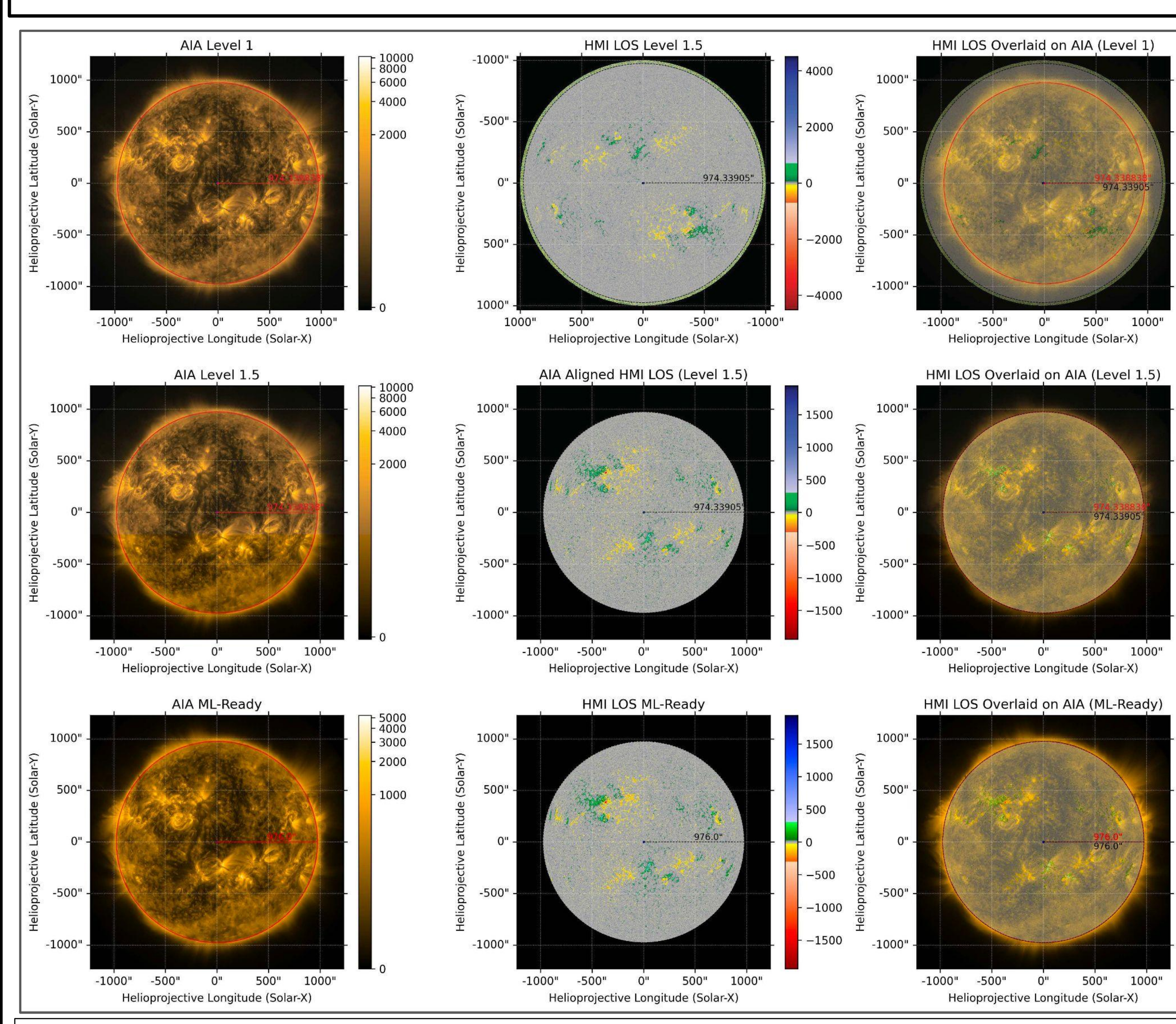


Fig-1: SDO (Launch- Feb. 11, 2010)



Preprocessing : AIA Level 1

1) Removing bad quality images
2) Replacing missing files with nearest file
3) Update pointing
4) Register
5) Exposure normalization
6) Instrument degradation correction
7) Elliptical orbit variation correction
8) Convert the datatype to 'float32'

Preprocessing : HMI Level 1.5

1) Removing bad quality images
2) Flagging missing files
2) Align HMI with AIA
4) Elliptical orbit variation correction
5) Convert the datatype to 'float32

Save: AIA(8 channels) + HMI (5 channels)

Fig-2: An example of the ML-ready data preparation steps for AIA 171 Å and HMI LOS magnetogram on 2012-01-30 at 22:12 UT. Contours illustrate the image center, solar disk center, disk radius, and solar disk boundary. The top row shows the original AIA Level 1 image, HMI Level 1.5 magnetogram downloaded from JSOC, and HMI overlaid on AIA. The disk centers are misaligned with the image center (unregistered), and one dataset has a 180 degrees roll, with noticeable plate scale differences. The middle row displays the registered AIA Level 1.5 image, HMI aligned with AIA, and HMI overlaid on AIA, showing corrected disk centers and plate scales. The bottom row presents the final ML-ready AIA and HMI images after exposure time normalization and orbital corrections for AIA, with the overlaid image showing proper alignment and a fixed disk radius of 976".

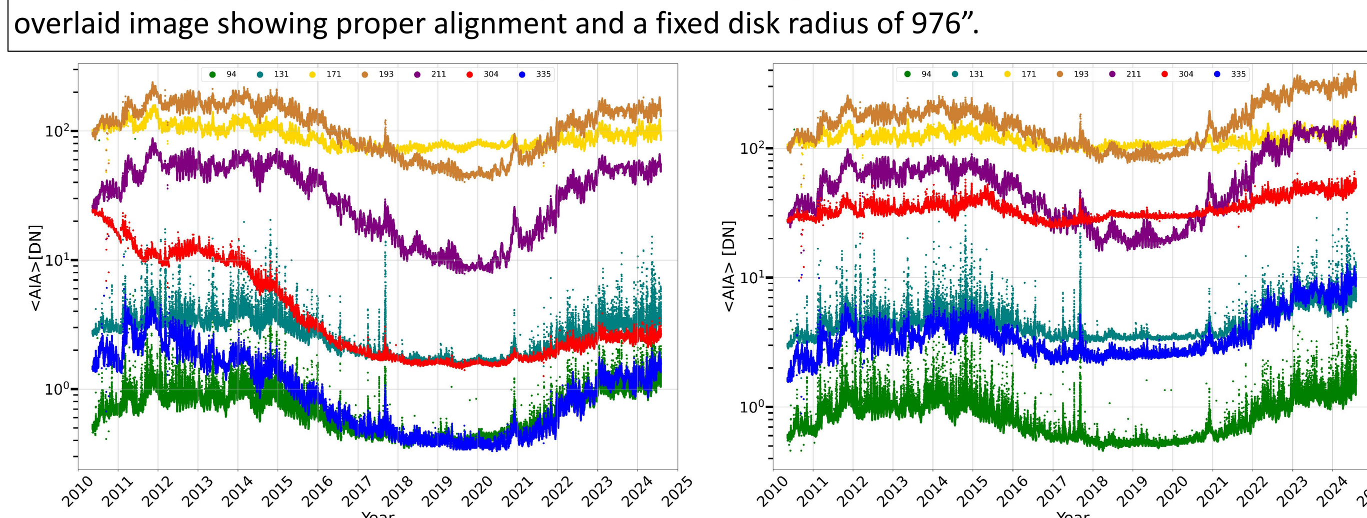
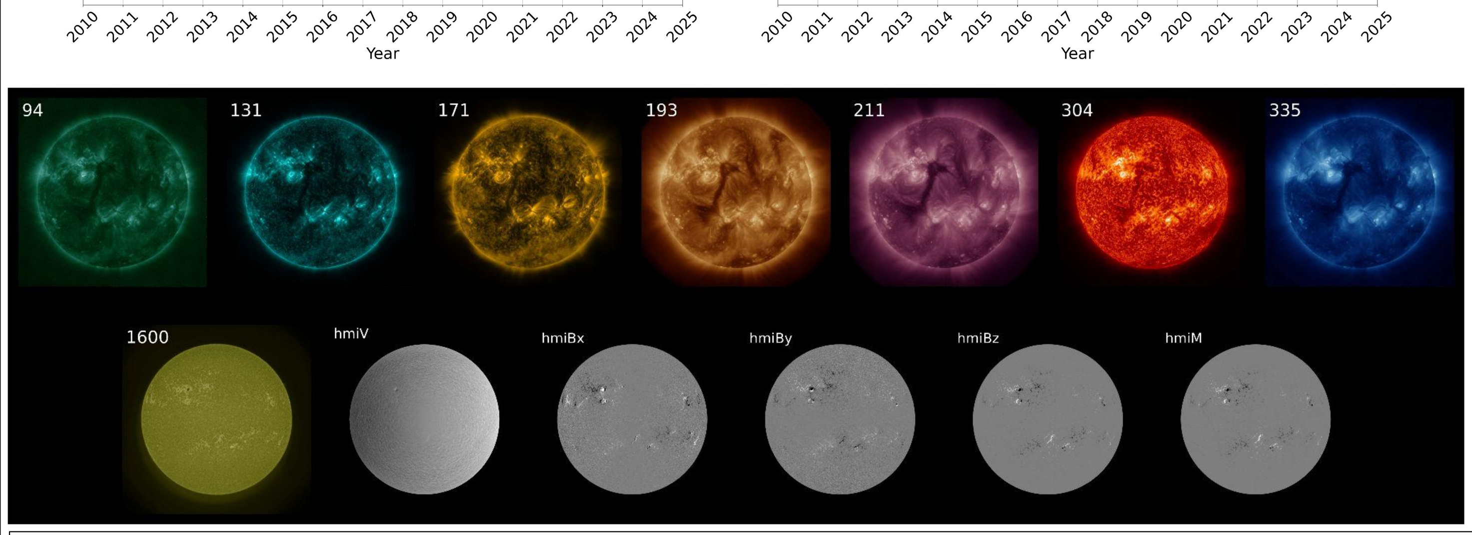


Fig-3: Mean pixel value of full-disk AIA images (extreme ultraviolet channels 94–335Å, DN/sec) over time before (left panel) and after (right panel) the degradation correction and exposure normalization.



Hugging Face

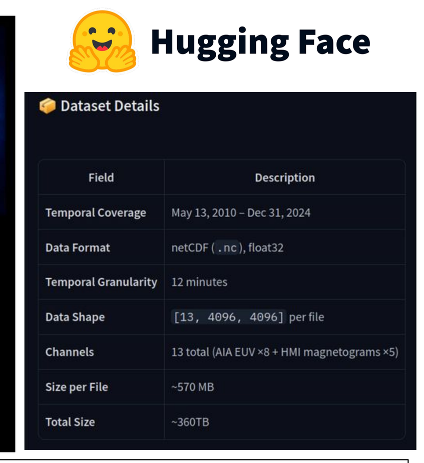


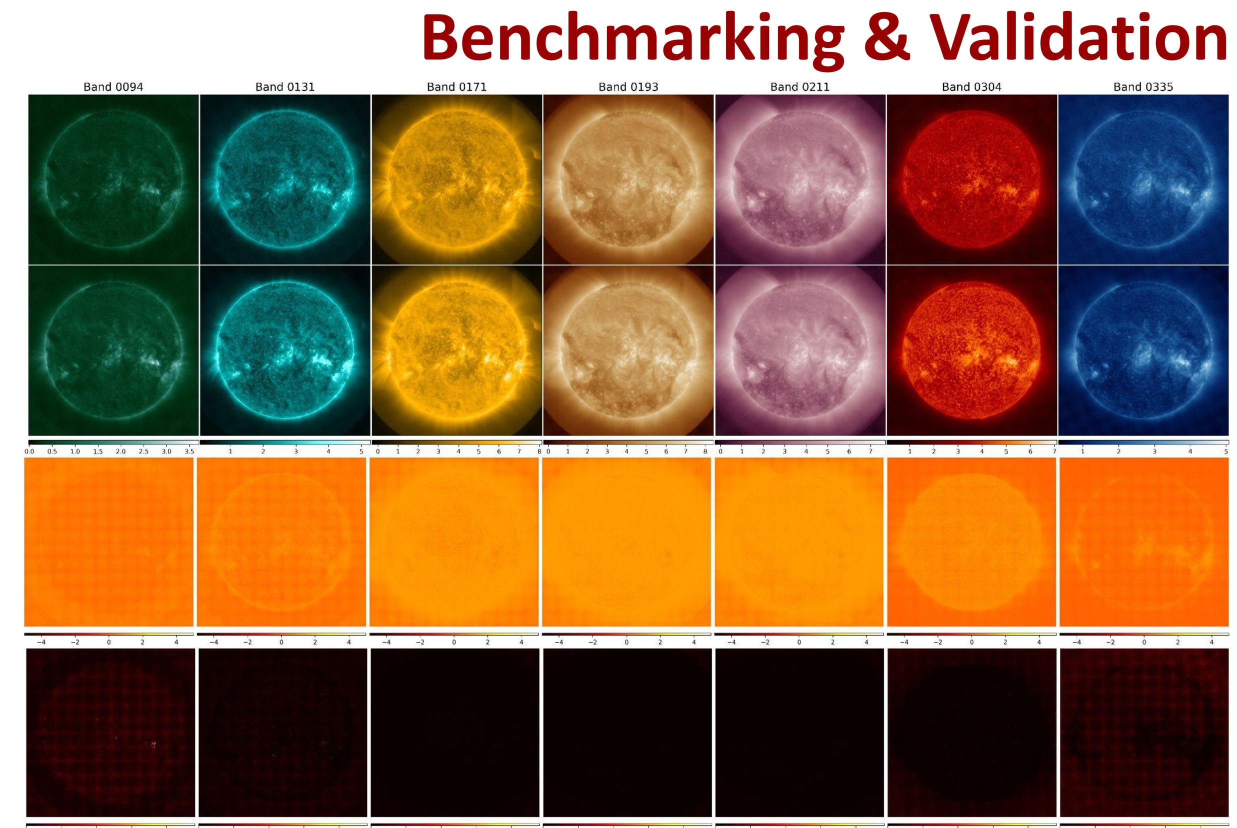
Fig-4: SuryaBench Core-SDO dataset: Solar coronal EUV/UV images from AIA (8 channels: 94, 131, 171, 193, 211, 304, 335, 1600 Å) and solar surface velocity and magnetic field maps from HMI (5 channels: hmiV, hmiBx, hmiBy, hmiBz, hmiM)

Downstream Application Dataset

	Application	Broader Topic	N	C	T	H	W	Dynamic Range
DS1	AR Segmentation	Magnetic energy in solar atmosphere	121,963	1	1	4,096	4,096	{0, 1}
DS2	AR Emergence Forecasting	Magnetic energy in solar atmosphere	56	6	240	9	9	-1.7×10^4 to 4.0×10^3
DS3	Coronal Field Extrapolation	Coronal structure & magnetism	5,347	1	12	1	4,095 (2)	-4.3×10^3 to 4.3×10^3
DS4	Flare Forecasting	Space weather forecasting	128,328	2	1	1	1	{0, 1}
DS5	Solar Wind Forecasting	Solar forcing of magnetosphere	119,225	5	1	1	1	2.4×10^2 to 1.096×10^3
DS6	EUV Forecasting	Solar forcing of thermosphere	189,344	1,343	1	1	1	1.0×10^{-9} to 1.1×10^{-2}

Table-2: Statistical summary of auxiliary datasets. The shape of all datasets has been standardized with the following dimensions: N → Number of datapoints, C → Number of channels, T → Number of timestamps, H → Height, and W → Width.

Benchmarking & Validation Results



Benchmarking Experiment:

- **Problem:** Forecasting the next time step using the present and previous time step as inputs.
- **Model:** The modified long-short Spectral Transformer (LSST) model (Roy et al., 2024).
- **Data:** Four years' (2011-2014) worth of SuryaBench core SDO data [AIA(8 EUV channels) + (HMI (1 LOS Magnetogram)].
- **Training:** 20 epochs in 24 to 36 hours on with16 nodes, 4 A100 80GB GPUs per node.

Fig-5: Forecast of AIA channels for next time steps, using 2 input timesteps. The first row is the ground truth (GT) for the model. The second row is the prediction from the model. The third row is the SSIM map between the ground truth and the prediction image. The fourth row is the squared error between the ground truth and the prediction image.

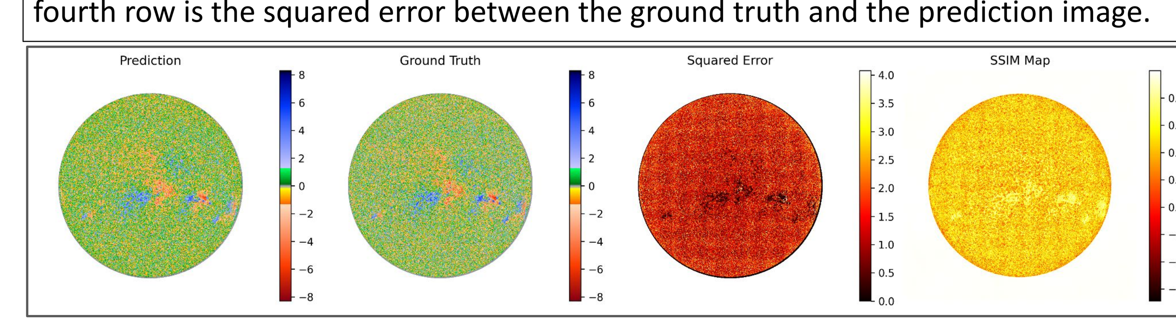


Fig-6: Same as Fig-4 but for HMI LOS magnetogram. [From left] the prediction by the model, GT, squared error, and SSIM map respectively. It is evident that active regions are well reproduced compared to the quiet regions.

Channel	AIA-94	AIA-131	AIA-171	AIA-193	AIA-211	AIA-304	AIA-335	HMI-LOS
RMSE	0.43	0.29	**0.11**	**0.09**	**0.10**	0.27	0.47	0.65
SSIM	0.44	0.48	**0.84**	**0.90**	**0.86**	0.49	0.24	0.74

Table-3: RMSE & SSIM scores for all channels from the baseline experiments. AIA-171,193,211 show strong performance over others.

(a) Solar Wind Forecasting

Model	RMSE (km/s)	MAE (km/s)
AlexNet	123.77	108.99
MobileNet	125.75	111.04
ResNet18	123.11	106.60
ResNet34	154.58	127.25
ResNet50	118.34	102.65

(b) EVE Prediction

Model	MSE	MAE
AlexNet	0.0001	0.007
MobileNet	0.0792	0.253
ResNet18	0.0078	0.069
ResNet34	0.0902	0.236
ResNet50	0.0339	0.139

(c) Solar Flare Prediction

Model	TSS	HSS	F1	Precision	Recall
AlexNet	0.16	0.21	0.29	0.85	0.17
MobileNet	0.53	0.44	0.65	0.51	0.89
ResNet50	0.05	0.08	0.11	0.85	0.06

(d) Active Region Segmentation

Model	IoU	Dice
UNet	0.62	0.76
Attention UNet	0.52	0.68

Table-4: Results from baseline evaluation experiments for (a) solar wind forecasting, (b) solar irradiance prediction, (c) solar flare forecasting, and (d) active region segmentation using standard deep learning architectures on held-out test datasets.

Summary

- The first ML ready dataset collection, **SuryaBench**, leveraging high-resolution SDO data and multiple heliophysics application dataset has been created and publicly available.
- Each dataset is benchmarked using the state-of-the-art models in the community.
- As a first major application, the first high resolution heliophysics foundation model, **Surya,** is being trained on SuryaBench dataset.
- Baseline evaluation for various downstream applications showed promising potential in the utility of extensive, unlabeled SuryaBench dataset.
- With efficient data handling and robust training strategies, SuryaBench can evolve into a versatile reusable resource for the heliophysics community.

References

- Pesnell, W. D., Thompson, B. J. & Chamberlin, P. C. (2012). Solar Physics 275, 3–15.
- Galvez, R. et al. (2019), The Astrophysical Journal Supplement Series 242, 7.
- Roy, S., Hegde, D. V., et al. (2026), Suryabench: Benchmark dataset for advancing machine learning in heliophysics and space weather prediction. *Scientific Data*, *13*(1), 712.
- Roy, S, et al. (2025), Surya: Foundation model for heliophysics. *arXiv preprint arXiv:2508.14112*.
- Roy, S., et al. (2024), Ai foundation model for heliophysics: Applications, design, and implementation, *arXiv preprint arXiv:2410.1084.*

Data Availability

The datasets are publicly available on Hugging Face (**https://huggingface.co/collections/nasa-ibm-ai4science/suryabench**), that includes Coronal Extrapolation, Solar Flare Forecasting, Solar Wind, Active Region Segmentation, Active Region Emergence, EUV Spectra, and the Core SDO dataset.

Code Availability

The complete dataset preparation and creation pipelines are open-sourced at **https://github.com/NASA-IMPACT/SuryaBench**. Additionally, the baseline models and training scripts used for evaluation are available at **https://github.com/NASA-IMPACT/Surya/tree/main/downstream_examples.**

Acknowledgment

This work is supported by NASA Grant 80MSFC22M004. The Authors acknowledge the National Artificial Intelligence Research Resource (NAIRR) Pilot and NVIDIA for providing support under grant no. NAIRR240178. DVH is grateful for the support provided by the NASA FINESST grant 80NSSC22K0058. Vishal Upendran would like to acknowledge NASA for support under award number 80NSSC25K7956

SHINE 2026, Madison, WI, June 28- July 3, 2026 , Email: dinesha.hegde@uah.edu