Supplementary Information for:

Operational forecasting of solar energetic particle event and proton flux using multi-source solar observations and multi-task deep learning

Yian Yu[1], Yang Chen[1,*], Lulu Zhao[2], Kathryn Whitman[3], Ward Manchester[2], Tamas Gombosi[2]

[1]Department of Statistics, University of Michigan, Ann Arbor, MI 48109, USA
[2]Department of Climate and Space Sciences and Engineering, University of Michigan, Ann Arbor, MI 48109, USA
[3]NASA Space Radiation Analysis Group, Johnson Space Center, Houston, TX 77058, USA

[*]Corresponding author: ychenang@umich.edu

Testing results on the SEPVAL dataset

Table **??** summarizes model performance on the SEPVAL dataset. SEPVAL is a multi-year benchmark for SEP model validation, comprising 33 SEP events and 30 non-event periods from 2011 to 2023. The dataset is available on Zenodo [**?**] and through supplementary resources provided by the Community Coordinated Modeling Center at `https://ccmc.gsfc.nasa.gov/community-workshops/ccmc-sepval-2023/`.

The designated periods in SEPVAL were used to construct the testing set following the preprocessing procedure described in [**?**]. In our previous study, `SEPNET-O` showed improved performance over state-of-the-art pre-eruption models and general machine learning baselines when SHARP parameters or flare-related features were incorporated. The results in Table **??** further show that the proposed `SEPNET-PRISM` outperforms `SEPNET-O`, especially in terms of TSS and HSS.

Table 1: Performance on the SEPVAL dataset. Reported values are the median and target quantile (75% quantile) across the `S` and `S+F` input feature sets for each model. SoA denotes state-of-the-art pre-eruption models.

Metrics	SoA	`SEPNET-O`		`SEPNET-PRISM`	
		`S`	`S+F`	`S`	`S+F`
ACC	0.55, 0.56	0.7097, 0.7419	0.6048, 0.6290	0.7258, 0.7419	0.6935, 0.7258
AUC	0.56, –	0.7816, 0.8051	0.6583, 0.6850	0.7819, 0.8009	0.7544, 0.7879
FPR	0.46, 0.27	0.2069, 0.1724	0.3793, 0.3103	0.2759, 0.2155	0.3448, 0.2759
F1	0.42, 0.45	0.7097, 0.7333	0.6129, 0.6475	0.7317, 0.7575	0.7313, 0.7549
POD	0.56, 0.67	0.7273, 0.7576	0.6061, 0.6667	0.7576, 0.8182	0.7727, 0.8182
FAR	0.36, 0.33	0.2265, 0.1923	0.3611, 0.3333	0.2675, 0.2148	0.2971, 0.2390
TSS	0.07, 0.10	0.4274, 0.4901	0.2137, 0.2529	0.4493, 0.4932	0.3908, 0.4681
HSS	0.07, 0.10	0.4230, 0.4860	0.2122, 0.2535	0.4493, 0.4876	0.3883, 0.4595

Notes: `S` denotes SHARP parameters, and `F` denotes flare-related features. Performance metric include ACC (accuracy), AUC (area under the curve), FPR (false positive rate), F1 (F1 score), POD (probability of detection), FAR (false alarm ratio), TSS (true skill score), and HSS (Heidke skill score).

Supplementary figures

Figure **??** summarizes the temporal coverage of all observational data sources considered in this study, including both the original data products and the aligned datasets used in preprocessing.

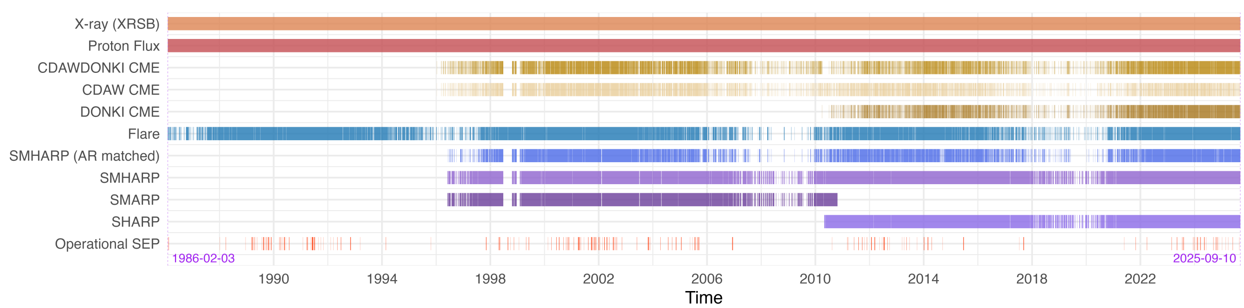

Figure 1: Timeline of the multi-source data used in this study from 3 February 1986 to 10 September 2025. Rows show GOES X-ray flux (XRSB), proton flux (>10 MeV), flares, the reconstructed CDAWDONKI CME event set, the original CDAW CME and DONKI CME catalogs, the SMHARP parameters, and the original SMARP and SHARP magnetic parameters. SMHARP (AR matched) denotes the subset of SMHARP observations whose active regions are matched to flare-associated NOAA active region numbers at the corresponding times. Vertical ticks indicate discrete events or observation times and illustrate temporal coverage across the different data sources. The bottom row shows operational SEP events, defined as proton flux >10 pfu in the >10 MeV channel.

Figure **??** presents predictor-level feature importance results corresponding to the grouped summary shown in the main text. Whereas the main-text figure emphasizes the aggregate contribution of each data source and variable family, Figure ?? shows the detailed permutation-importance values for individual predictors together with their variability across repeated, class-balanced random forest resamples. Each predictor is represented by summary statistics computed over the preceding non-overlapping 24-hour window, namely the minimum, mean, and maximum values.

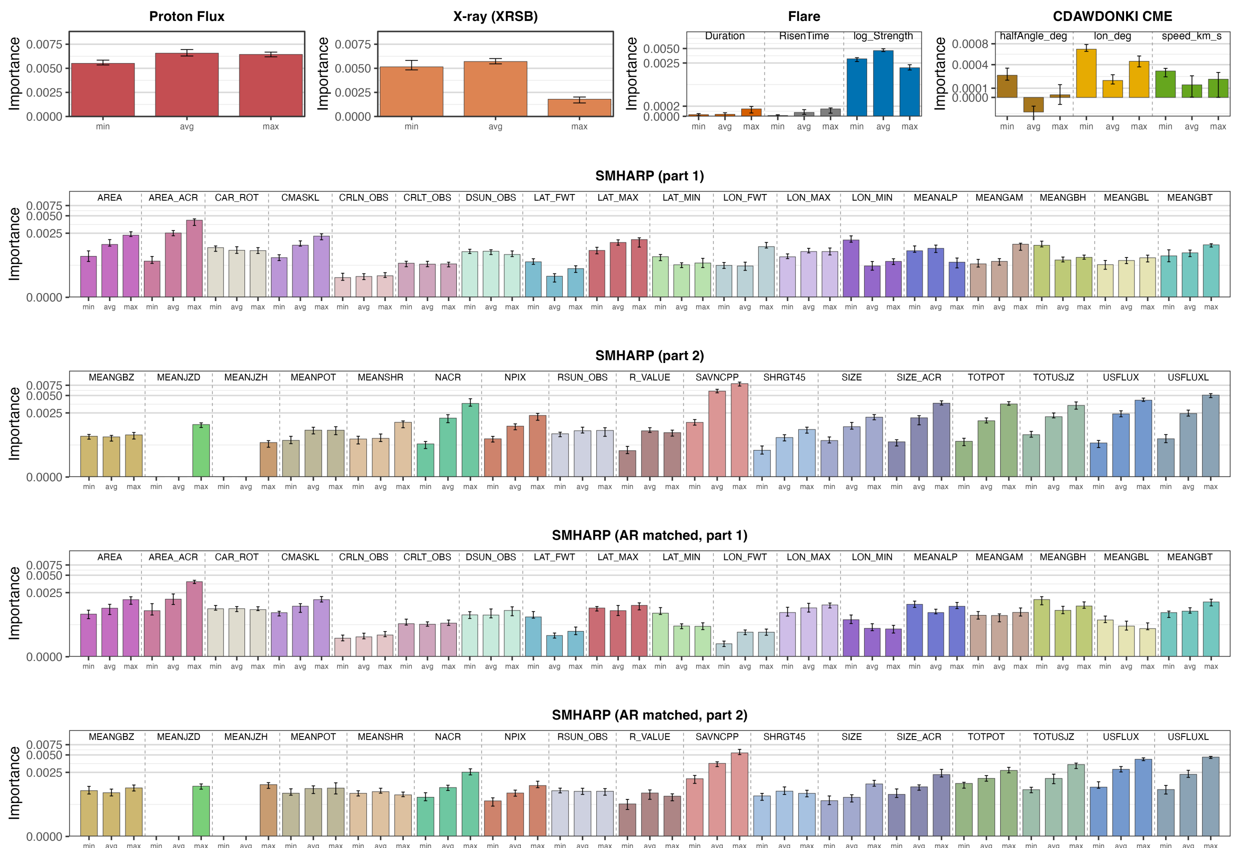

Figure 2: Predictor-level feature importance for operational SEP forecasting estimated using permutation importance from repeated, class-balanced random forest models. Panels are grouped by data source, including proton flux, GOES XRSB flux, flare properties, CME features from the CDAWDONKI catalog, SMHARP parameters, and SMHARP restricted to flare-matched NOAA active regions. For each predictor, the three bars denote the minimum, mean, and maximum values computed over the preceding non-overlapping 24-hour window. Bars show mean permutation importance across resamples, and error bars indicate variability across resamples.

References

[1] Whitman, K. SEPVAL 2023 challenge event lists (v2), DOI: 10.5281/zenodo.15555244 (2025).

[2] Yu, Y. *et al.* Solar energetic particle forecasting with multi-task deep learning: SEPNET. *Journal of Geophysical Research: Machine Learning and Computation* **3**, e2026JH001247 (2026).