Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Real-Time 24-hour Ahead Forecasting of Solar Energetic Particle Events with SEPNET-PRISM-R

Yian Yu
University of Michigan
Yang Chen
ychenang@umich.edu
University of Michigan
Lulu Zhao
University of Michigan
Kathryn Whitman
Johnson Space Center
Ward Manchester
University of Michigan
Tamas Gombosi
University of Michigan

Article

Keywords: solar energetic particle, real-time forecasting, multi-modal data, multi-task learning, feature importance, live website

Posted Date: November 6th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11433870/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Real-Time 24-hour Ahead Forecasting of Solar Energetic Particle Events with SEPNET-PRISM-R

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

*Corresponding author(s). E-mail(s): ychenang@umich.edu;

Abstract

Solar energetic particle (SEP) events, operationally defined as proton flux exceeding 10 pfu in the >10 MeV channel, threaten spacecraft systems, astronaut safety, and high-latitude aviation, yet continuous real-time forecasting with adequate lead time remains difficult. We present SEPNET-PRISM-R, a near-real-time system that issues hourly, 24-hour-ahead probabilistic SEP forecasts. The model integrates multimodal solar data—including active-region patches, flare features, coronal mass ejection (CME) characteristics, and historical proton flux—to concurrently predict SEP occurrences, peak proton flux, soft X-ray flux, and CME velocity, with the primary objective being SEP forecasting. Building on existing literature, SEPNET-PRISM-R enhances operational performance by incorporating a post-processing module that enables cross-validated decision-making and interpretable quantification of feature importance. The training dataset spans from February 1986 to September 2025, with subsequent forecasts generated sequentially through the present via a real-time space weather prediction platform that provides predictive probabilities, feature attributions, and historical analogs, enabling more reliable and informed operational decision-making. The proposed framework substantially suppresses false-alarm rates while maintaining robust event-detection capabilities, as evidenced by case studies and aggregate statistics.

Keywords: solar energetic particle, real-time forecasting, multi-modal data, multi-task learning, feature importance, live website

1 Introduction

Solar energetic particle (SEP) events are among the most hazardous manifestations of solar activity and represent a major component of space weather [1]. During solar eruptions, protons and heavier ions can be accelerated to relativistic energies through magnetic reconnection associated with solar flares and through shock waves driven by fast coronal mass ejections (CMEs) [2, 3]. These particles can propagate through the heliosphere and reach Earth within tens of minutes to several hours after eruption, and the flux may remain elevated for days. Operationally, NOAA (National Oceanic and Atmospheric Administration) defines an SEP event when the integral proton flux in the GOES $>$ 10 MeV channel exceeds 10 particle flux units (pfu), a threshold now standard in space weather monitoring and forecasting. SEP events threaten spacecraft electronics, satellite operations, astronaut radiation safety, high-latitude aviation, and radio communications [4, 5], making reliable forecasting an important objective for both scientific research and operational space weather services [6, 7].

However, SEP forecasting remains scientifically difficult, as event occurrence depends on a combination of solar eruption characteristics, the underlying particle acceleration mechanism, interplanetary transport, and the magnetic connectivity between the acceleration site and the observer [2]. Large X-class flares often fail to produce measurable SEP enhancements at Earth when the associated magnetic geometry is unfavorable, while moderate eruptions can occasionally generate intense particle events under efficient magnetic connection [8]. Fast CMEs are similarly important for large SEP events, yet CME speed alone is insufficient for reliable prediction, since shock formation and subsequent particle transport also depend on the ambient solar wind and the evolving heliospheric magnetic field [9].

A wide range of forecasting approaches has been proposed over the past several decades. Early empirical methods primarily relied on flare magnitude and location, and CME speed to estimate SEP probability [10]. Although physically interpretable, these methods are constrained by simplifying assumptions and struggle to capture complex nonlinear interactions among predictors. Machine learning techniques have become increasingly popular for SEP forecasting [11], various statistical and deep learning models incorporating magnetic field measurements, flare characteristics, CME properties, radio burst observations, and historical particle fluxes to improve skill [12, 13]. These data-driven approaches have delivered measurable gains in forecast performance, underscoring their promise for operational space weather forecasting.

Several obstacles nevertheless stand between these advances and a successful operational deployment [14]. Many published studies rely on retrospectively constructed datasets in which training and testing samples are randomly partitioned [13, 15]. Because neighboring observations are often highly correlated in time, such an evaluation protocol can substantially inflate the predictive skill scores. Furthermore, many studies focus primarily on event classification, while paying less attention to operational considerations, such as continuous hourly updates, forecast stability, false-alarm control, computational cost, and model interpretability [16]. Closing the gap between retrospective evaluation and real-time operation is especially important for SEP prediction, considering its potential impacts on near-Earth environments.

The previously developed SEPNET family of deep-learning forecasting models partially addressed these issues. The initial operational version, SEPNET-O [17], integrated SHARP magnetic parameters with flare characteristics to produce probabilistic forecasts of SEP events over the next 24 hours. The subsequent methodological refinement, SEPNET-PRISM [15], is a multitask extension trained on a substantially larger historical dataset spanning more than two solar cycles with additional auxiliary targets, including future proton flux, GOES soft X-ray flux, and CME speed. It was shown that the joint learning of these related quantities improved predictive performance and enriched the learned physical representation beyond binary classification. However, SEPNET-PRISM centered on model development and retrospective evaluation rather than real-time operation. While SEPNET-O was shown real-time capability and has been running live on NASA CCMC (Community Coordinated Modeling Center) [18], its outputs remained sensitive to short-term fluctuations from noisy observations, producing unstable warning behavior under sequential prediction. For instance, during the Artemis II space mission, the live forecasting produced by SEPNET-O issued false positive warnings of transient non-SEP soarl eruptions with a significantly perturbed prediction probability, which are primarily caused by noisy observations and random fluctuations. Furthermore, neither SEPNET-O nor SEPNET-PRISM had been paired with the temporal post-processing, independent post-training evaluation, or operational interpretability that a desirable operational system would have.

Recognizing the substantial gaps in methodology and the lack of practical real-time decision-supporting for operational forecasting, we propose SEPNET-PRISM-R, a thorough extension of SEPNET-PRISM built specifically for continuous near-real-time forecasting. With a trained model, forecasts are issued sequentially every hour using only the observations available at issuance time, closely mirroring true operational deployment. To improve temporal consistency, SEPNET-PRISM-R adopts the Hidden Semi-Markov Model (HSMM) [19] as a post-processing layer that explicitly models the persistence of SEP and quiet states, reducing isolated false alarms driven by short-term probability fluctuations. We evaluate the resulting system over a post-training period from September 2025 through July 2026 (that has no overlap with training data) to obtain a realistic picture of operational performance. We use SHapley Additive exPlanations (SHAP) [20, 21] to quantify global and forecast-specific feature contributions, improving transparency and offering physical insight into the drivers of each operational prediction. In our interactive online platform, we show real-time forecasting results, together with intuitive explanations of feature contributions and matched historical event visualizations, to facilitate interpretation of prediction results in terms of both reliability, stability and explanability.

2 Results

This section evaluates the forecasting performance of SEPNET-PRISM-R for hourly 24-hour-ahead operational SEP forecasting and compares it with the previous operational model SEPNET-O. We begin by evaluating the model's overall performance across the full test interval. Subsequently, we examine detailed case studies of SEP and non-SEP events, present the auxiliary regression results, and analyze feature importance using

SHAP values. Finally, we demonstrate the near-real-time hourly web dashboard that delivers these operational outputs alongside the analog-based context established in Section 4.5. Classification performance is evaluated using accuracy (ACC), the area under the receiver operating characteristic curve (AUC) [22], the false positive rate (FPR), the F1 score [23], the probability of detection (POD) [24], the false alarm ratio (FAR) [25], the true skill statistic (TSS) [26], and the Heidke skill score (HSS), in accordance with established practices in space weather forecast verification [27]. Regression performance is quantified using the root mean square error (RMSE), the mean absolute error (MAE), and the coefficient of determination (R^2).

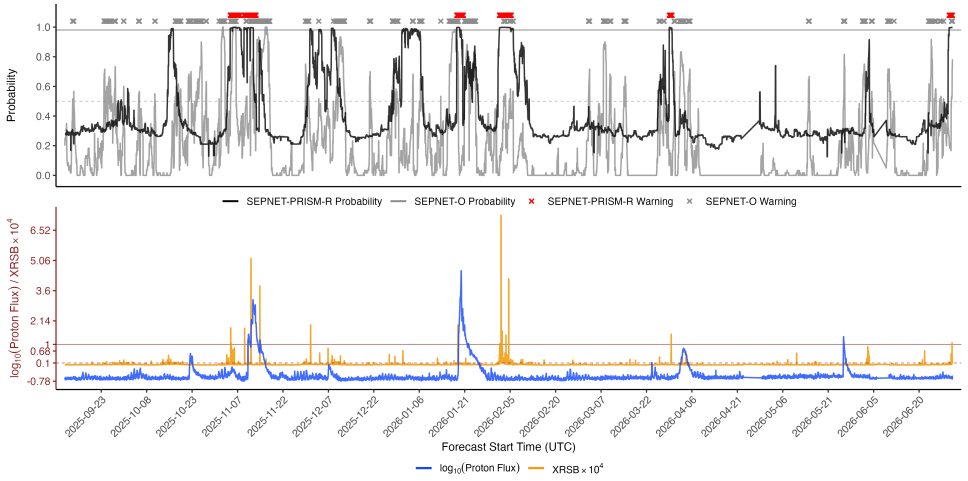
2.1 Overall Forecast Performance

We evaluated the models over the full test period from September 2025 to July 2026. Figure 1a) shows the temporal evolution of forecast probabilities and warning states from `SEPNET-PRISM-R` and `SEPNET-O`, together with the corresponding proton flux and soft X-ray XRSB observations. Over this period, `SEPNET-PRISM-R` generated substantially fewer false alarms while still identifying the major SEP intervals of interest, whereas `SEPNET-O` issued frequent and often persistent warnings outside confirmed SEP intervals. Within this test period, three proton events met the operational SEP criterion of proton flux exceeding 10 pfu in the >10 MeV channel, beginning at 11:25 UTC on November 10, 2025, 22:55 UTC on January 18, 2026, and 00:15 UTC on May 26, 2026. The first two events were intense, with peak proton fluxes on the order of 3×10^3 pfu, while the May 2026 event reached a much smaller peak of only about 20 pfu. It should be noted that the proton flux shown here is the unpreprocessed observational GOES measurement, and has not been background corrected [28].

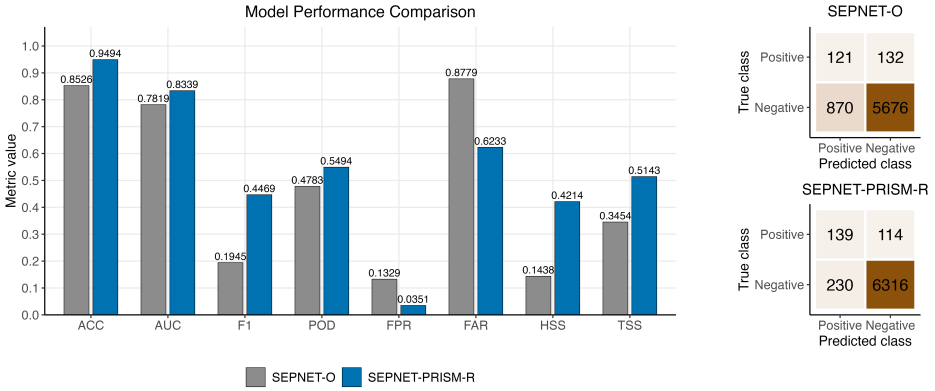
As shown in Figure 1a), `SEPNET-PRISM-R` successfully captured the first two high-intensity events and issued far fewer false warning periods than `SEPNET-O`. The remaining false alarms were associated with exceptionally strong soft X-ray flare activity, illustrating that intense solar eruptions do not always produce SEP events at Earth and underscoring the difficulty of forecasting rare SEP occurrences from precursor information alone. The aggregate verification statistics are summarized in Figure 1b). Relative to `SEPNET-O`, `SEPNET-PRISM-R` improved all major metrics, increasing probability of detection (POD) from 0.4783 to 0.5494, HSS from 0.1438 to 0.4214, and TSS from 0.3454 to 0.5143, while reducing FPR from 0.1329 to 0.0351 and FAR from 0.8779 to 0.6233. The confusion matrices further show that false alarms decreased from 870 for `SEPNET-O` to 230 for `SEPNET-PRISM-R`, while correctly identified positive samples increased from 121 to 139. Collectively, these results indicate enhanced event detection performance and markedly improved control of false alarms for `SEPNET-PRISM-R`.

2.2 SEP and Non-SEP Event Case Studies

Figures 2a) and 2b) show detailed views of the two successfully forecasted SEP events. In both cases, `SEPNET-PRISM-R` issued warnings several hours before proton flux exceeded the 10 pfu threshold and maintained elevated probabilities throughout the main event interval. The orange shaded region denotes the 24-hour input window used to generate the forecast, and the blue shaded region denotes the forecast window



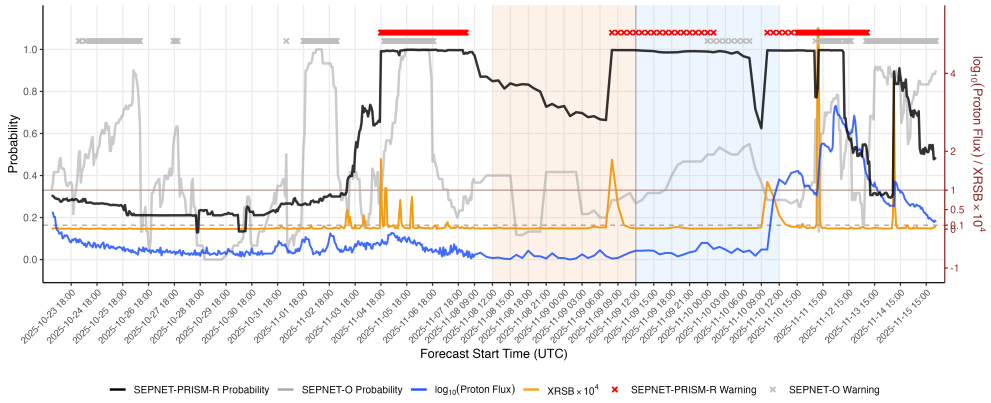
a). Operational forecasting results from September 10, 2025, to July 1, 2026. The upper portion shows forecast probabilities from `SEPNET-PRISM-R` (black) and `SEPNET-O` (gray), together with their corresponding warning indicators (red and gray crosses on the top). The lower portion shows the observed solar and particle context, including $\log_{10}$ proton flux (blue) and XRSB soft X-ray flux scaled by 10^4 (orange).



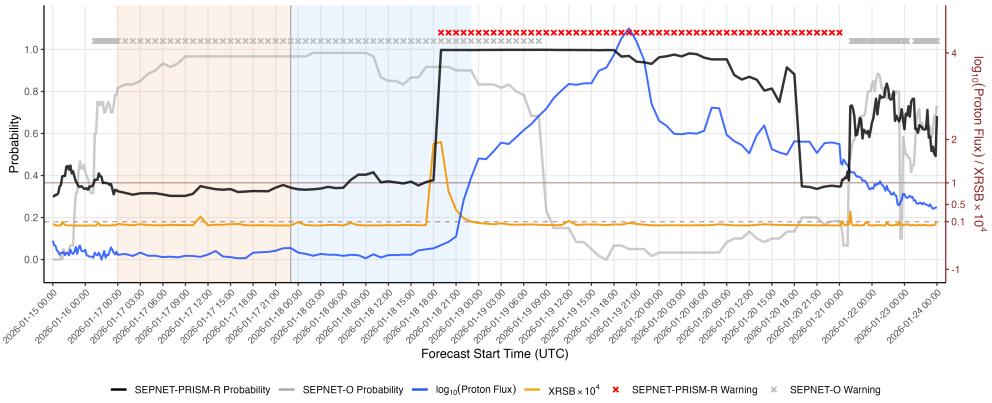
b). Summary of evaluation metrics and confusion matrices for `SEPNET-O` and `SEPNET-PRISM-R` over the full test period (September 2025 through July 2026).

Fig. 1: Overall performance of the operational SEP forecasting models.

preceding SEP onset. In the November 2025 event, the model responded to strong precursor activity, including a large soft X-ray enhancement, and issued a warning well before the proton flux crossed threshold. In the January 2026 event, the model again provided an early warning and sustained the alert throughout.



a). Forecast results for the November 2025 SEP event.



b). Forecast results for the January 2026 SEP event.

Fig. 2: Detailed visualizations of operational forecasts for two SEP events, November 2025 (top) and January 2026 (bottom). The left axis shows the forecast probabilities from `SEPNET-PRISM-R` (black) and `SEPNET-O` (gray), together with their corresponding warning indicators (red and gray crosses). The right axis shows $\log_{10}$ proton flux (blue) and soft X-ray flux XRSB scaled by 10^4 (orange). The orange shaded interval denotes the 24-hour precursor window used to construct the forecast, and the blue shaded interval marks the forecast period preceding SEP onset.

In addition to the verified SEP events, we examine two disturbed intervals to assess the model's robustness under challenging conditions. Figure 3 shows forecasting results during the Artemis II mission and the late-May 2026 event interval around Memorial Day. On March 30, 2026, an intense X-class flare occurred, and `SEPNET-PRISM-R` responded with elevated event probability (Figure 3a)) and gradually decayed over time. No SEP event was recorded during this interval. In contrast,

SEPNET-O issued several scattered SEP warnings, indicating a strong false-alarm tendency under disturbed but non-SEP conditions. This case indicates that the updated framework, SEPNET-PRISM-R, is responsive to strong eruptive solar activity while suppressing spurious warnings when the disturbance does not evolve into a detectable SEP enhancement at Earth. The late-May 2026 interval differs in that the proton enhancement was a true SEP event, likely driven by a CME, but it was not well represented by the short-horizon precursor variables used in both SEPNET-O and SEPNET-PRISM-R. Consequently, SEPNET-PRISM-R did not issue a timely warning, while SEPNET-O generated a delayed alert. This case suggests that CME-driven SEP events may require explicit CME-related inputs, motivating a more curated analysis of CME information in future work.

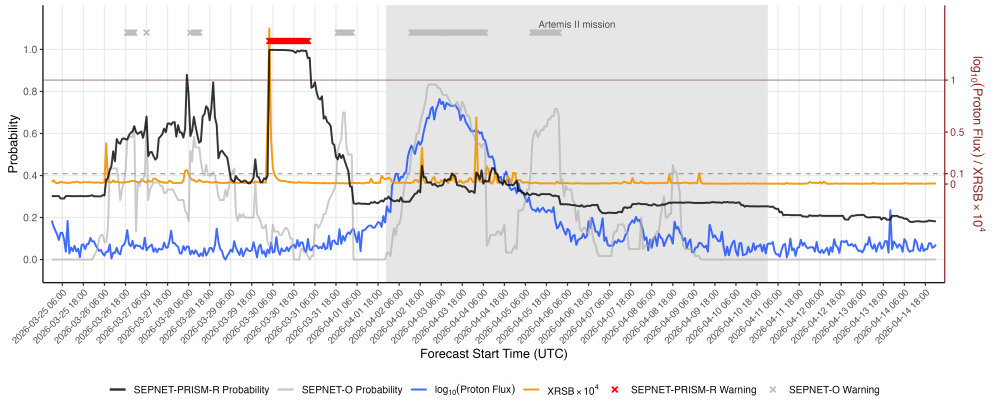
2.3 Auxiliary Regression Results

In addition to binary SEP event forecasting, SEPNET-PRISM-R predicts three auxiliary regression targets: the subsequent 24-hour maximums of logarithmic proton flux in the >10 MeV channel, X-ray flux XRSB, and the CME speed. These auxiliary tasks are intended to provide physically relevant information about the SEP environment, thereby enhancing feature learning and prediction accuracy.

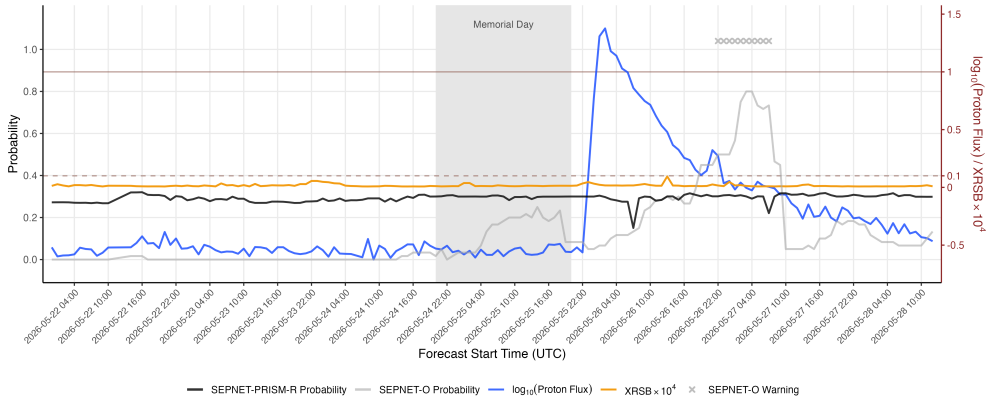
Figure 4 compares the predicted and observed values. The gray background curve indicates the estimated SEP event probability as a reference. In general, intervals of high estimated SEP probability are associated with elevated predicted proton flux, showing that SEPNET-PRISM-R has learned a physically meaningful relationship between SEP occurrence and enhanced proton intensity. At the same time, the predicted proton flux series remains somewhat elevated during some intervals in which the observed proton flux does not increase substantially, suggesting that accurate quantitative prediction of proton intensity remains challenging given the strongly skewed distribution of flux values. Moreover, the predicted X-ray flux XRSB follows the observed temporal variations more closely, whereas the skewed distributions resulting from the extremely high value are still not effectively captured. Accurately predicting the maximum CME speed remains challenging within the current framework, where CME speed is not the primary objective. This is also due to CME kinematics being influenced by complex, highly variable eruption dynamics, projection effects, and incomplete observability of input features [3, 29].

2.4 SHAP Importance

To interpret the operational model beyond the aggregate skill scores, we quantified the contribution of individual predictors to the probability of the calibrated SEP event using SHAP (SHapley Additive exPlanations) analysis [20, 21]. For the preferred SHARP+flare (S+F,[15]) predictor configuration of SEPNET-PRISM-R, we computed gradient explainer attributions for each forecast window across the ensemble of full-train split models, which were subsequently averaged. For clarity of presentation, we pooled the windowed summaries (minimum, maximum, and average) for each physical quantity into a single base feature. Additionally, we aggregated individual predictors into block totals for all SHARP and flare features.



a). Forecast results during the Artemis II mission, including the extreme X-class flare that occurred on March 30, 2026.



b). Forecast results during the late-May 2026 interval around Memorial Day.

Fig. 3: Detailed visualizations of operational forecasts during two disturbed intervals. The left axis shows the forecast probabilities from `SEPNET-PRISM-R` (black) and `SEPNET-O` (gray), together with their corresponding warning indicators. The right axis shows $\log_{10}$ proton flux (blue) and XRSB soft X-ray flux scaled by 10^4 (orange).

The global feature importance summary is shown in Figure 5a). When importance is measured by the mean absolute attribution mass, the SHARP block dominates the flare block (mean $\Sigma|\text{SHAP}| \approx 0.1189$ versus ≈ 0.0311). Among individual base features, flare intensity (`Flare_log_Strength`) is the strongest single contributor (mean $|\text{SHAP}| \approx 0.0234$), followed by Carrington longitude and latitude of the observed region (`SHARP_CRLN_OBS`, `SHARP_CRLT_OBS`). Temporal consistency of these attributions is examined in Figure 5b). In 5b), the upper panel shows the ensemble SEP probability over the near-real-time test period; the middle panel compares the All-SHARP

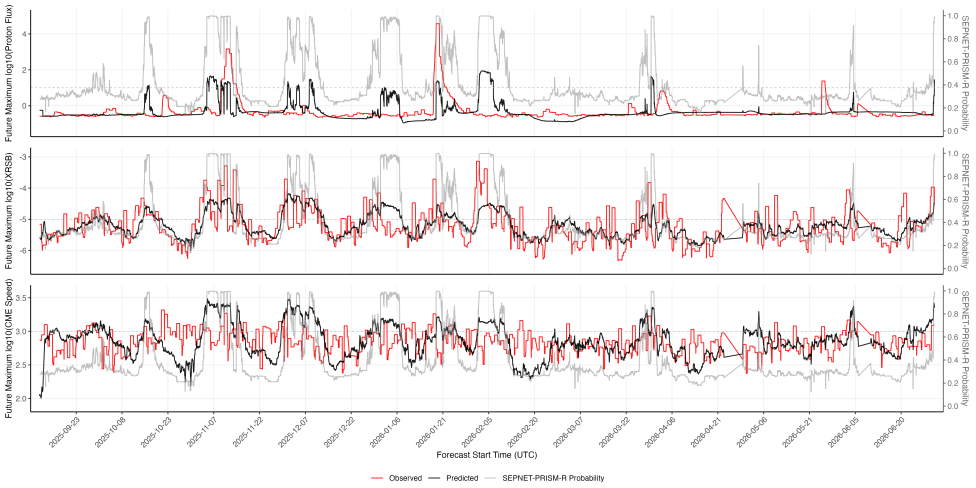


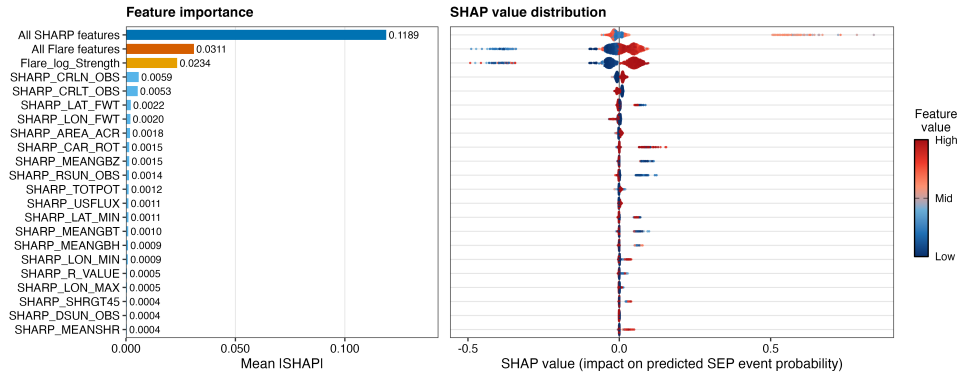
Fig. 4: Comparison between predicted and observed values for the two auxiliary regression targets. The upper panel presents the future maximum $\log_{10}$ proton flux in the $>$ 10 MeV channel, the middle panel presents the future maximum $\log_{10}$ XRSB flux, and the lower panel presents the future maximum $\log_{10}$ CME speed. The gray background curve indicates the estimated SEP-event probability from `SEPNET-PRISM-R` for reference.

and All-Flare attribution masses; and the lower panel tracks the leading individual base features. Elevated SEP probabilities generally coincide with increases in the aggregate SHARP contribution and, at times, with flare-related attributions, supporting the interpretation that magnetic-context and flare information jointly drive the operational forecast of SEP events.

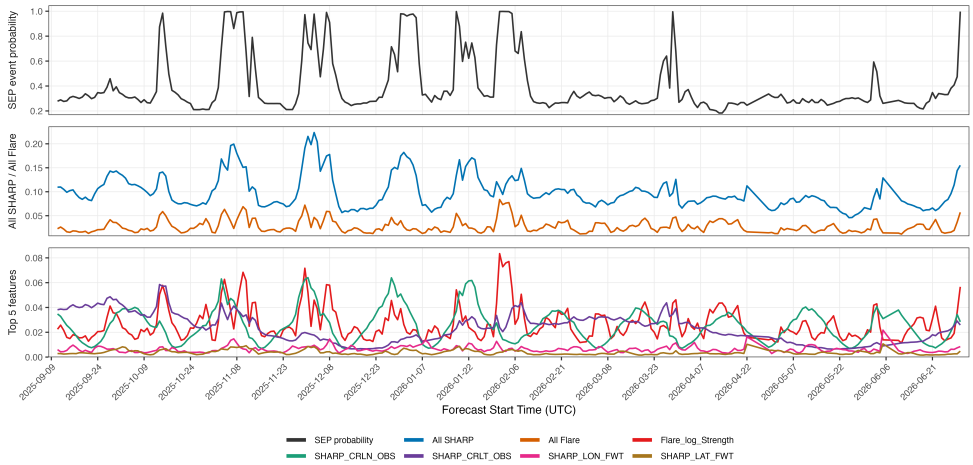
2.5 A Web-Based Live Dashboard

The forecasting framework `SEPNET-PRISM-R` is deployed as a publicly accessible, hourly updated web dashboard, `SEPNET-PRISM-R` (https://mlsw.engin.umich.edu/apps/SEPNET-PRISM-R), hosted within the University of Michigan Space Weather Interactive Lab (SWIL, spaceweatherlab.org) prediction-model portal. The dashboard exposes six linked views, "*Probability*, *Regression*, *SHAP over time*, *SHAP summary*, *Similar period*, and *Similar event*", that together provide continuous access to the model outputs discussed in the preceding subsections, together with the analog-based contextualization introduced in Section 4.5. The dashboard header reports the number of hourly forecasts, the number of active warnings, the covered forecast window, and the most recent update time. The underlying forecast archive is refreshed hourly and mirrored to a dedicated cloud storage location for reproducibility.

The notable features of the web dashboard are its high interpretability and support for real-time decision-making. Typical real-time space weather forecasting websites issue warnings based on predicted event probability and pre-trained or preset decision



a). Global SHAP summary for `SEPNET-PRISM-R` with feature configuration `S+F`. Left: mean absolute SHAP importance for All-SHARP and All-Flare blocks and for the leading base features (pooling `min`/`max`/`avg`). Right: signed SHAP beeswarm colored by the corresponding feature value.



b). Time series of the ensemble SEP predicted probability (top), aggregate All-SHARP and All-Flare SHAP masses (middle), and leading individual base-feature contributions (bottom) over the near-real-time evaluation period. Values are daily means of $\Sigma|\text{SHAP}|$ over `min`/`max`/`avg`; a few pathological post-outage days are omitted from the line plot.

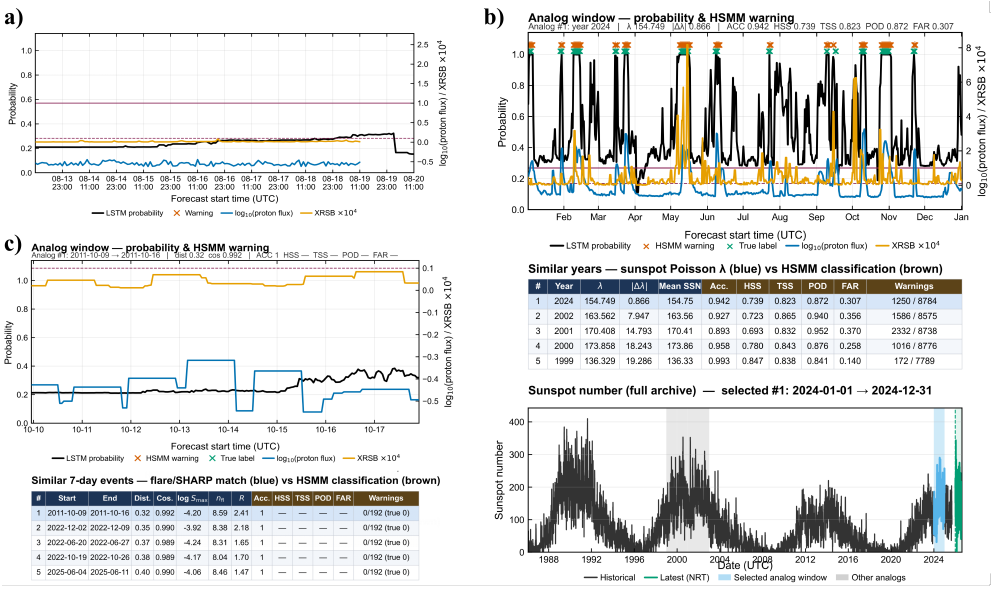
Fig. 5: SHAP interpretation of `SEPNET-PRISM-R` under the operational `S+F` feature set. a) Global feature importance and signed contribution distributions. b) Temporal co-evolution of forecast probability and SHAP attributions.

thresholds. In our dashboard, we offer researchers a feature contribution decomposition to materialize the forecasting outcome in terms of physical quantities. Moreover, the raw model output, probability of a forthcoming event, does not directly give the reliability and trustworthiness of the forecast, and a preset threshold makes the decision-making arbitrary and sensitive to small perturbations of the system and noise contamination. We match the current solar condition with historical windows and historical events in the web interface, showing (i) how good/bad our model performed in similar solar cycle years and (ii) whether our model correctly/incorrectly predicted historical events that enjoyed similar features as the current event. For each one, we provide summary statistics so operators can easily rely on them when making operational decisions.

Figure 6 illustrates representative views from a live snapshot of the dashboard that gets continuously updated as new observations arrive each hour. The *Probability view* (Figure 6a) gives the real-time forecast probability and the HSMM warning state alongside observed proton flux and soft X-ray flux, in the same style as Figure 1a). The *SHAP summary view* shows the global feature-importance ranking and beeswarm distribution described in Section 2 in an interactive form, alongside a complementary *SHAP-over-time view* that tracks the same attributions hour by hour. The *Similar period view* (Figure 6b) ranks historical years by $|\Delta\lambda|$ relative to the current year's sunspot-number Poisson rate; in a representative snapshot from mid-August 2026 (Figure 6c), 2024 ($\lambda = 154.7$, $|\Delta\lambda| = 0.866$), 2002, 2001, 2000, and 1999 were returned as the top five analog years, with historical HSMM warning skill ranging from HSS = 0.72 to 0.85 across these analogs. The *Similar event view* ranks historical non-overlapping 7-day windows by z-scored Euclidean distance in flare-and-SHARP feature space; for the same representative snapshot, the closest matched windows (e.g., 9–16 October 2011, 2–9 December 2022, 20–27 June 2022, 19–26 October 2022, and 4–11 June 2025) all had distances below 0.4 and cosine similarities around 0.99, and none coincided with an operational SEP event, consistent with the quiescent forecast probability observed during that period. Together, these analog views comparing historical events allow forecasters to cross-check the current model probability and its dominant SHAP drivers against both the broader solar-cycle context and specific historical analogs of the current precursor signature, supporting more informed operational decision-making than a point-probability forecast alone.

3 Discussions on Model Performance

The primary contribution of this study is the transition of the `SEPNET` framework from retrospective model development to a realistic operational forecasting system. Unlike previous studies [15], `SEPNET-PRISM-R` was evaluated in a strictly post-training setting, where model parameters remained fixed, and forecasts were generated sequentially as new observations became available. This evaluation more closely represents how an operational forecasting system would perform in practice. The results show that the proposed framework substantially improves forecast reliability by reducing false alarms while maintaining strong detection capability, suggesting that temporal consistency is as important as classification accuracy for operational SEP forecasting.



#	Year	λ	|Δλ|	Mean SSN	Acc.	HSS	TSS	POD	FAR	Warnings
1	2024	154.749	0.866	154.75	0.942	0.739	0.823	0.872	0.307	1250 / 8784
2	2002	163.562	7.947	163.56	0.927	0.723	0.865	0.940	0.356	1586 / 8575
3	2001	170.408	14.793	170.41	0.893	0.693	0.832	0.952	0.370	2332 / 8738
4	2000	173.858	18.243	173.86	0.958	0.780	0.843	0.876	0.258	1016 / 8776
5	1999	136.329	19.286	136.33	0.993	0.847	0.838	0.841	0.140	172 / 7789

#	Start	End	Dist.	Cos.	log S_{max}	n_{fl}	R	Acc.	HSS	TSS	POD	FAR	Warnings
1	2011-10-09	2011-10-16	0.32	0.992	-4.20	8.59	2.41	1	—	—	—	—	0/192 (true 0)
2	2022-12-02	2022-12-09	0.35	0.990	-3.92	8.38	2.18	1	—	—	—	—	0/192 (true 0)
3	2022-06-20	2022-06-27	0.37	0.989	-4.24	8.31	1.65	1	—	—	—	—	0/192 (true 0)
4	2022-10-19	2022-10-26	0.38	0.989	-4.17	8.04	1.70	1	—	—	—	—	0/192 (true 0)
5	2025-08-04	2025-08-11	0.40	0.990	-4.06	8.46	1.47	1	—	—	—	—	0/192 (true 0)

Fig. 6: Representative views of the `SEPNET-PRISM-R` web dashboard. a) Real-time probability, HSMM warning state, and observed proton and soft X-ray flux. b) Similar solar-cycle periods, ranked by $|\Delta\lambda|$ in year-to-date sunspot-number Poisson rate, with historical HSMM classification skill for each analog year. c) Similar precursor events, ranked by z-scored Euclidean distance and cosine similarity in flare-and-SHARP feature space over non-overlapping 7-day windows.

The improved performance primarily stems from incorporating temporal information via the HSMM. Hourly forecasts are naturally correlated because consecutive prediction windows overlap considerably. Treating each forecast independently can therefore produce isolated probability fluctuations and fragmented warnings, like in `SEPNET-O` [17]. By explicitly modeling state persistence, the HSMM suppresses short-lived fluctuations while preserving sustained periods of elevated probability that are characteristic of true SEP events. This post-processing strategy is independent of the underlying neural network architecture and could therefore be combined with future forecasting models as a general framework for improving operational stability.

Beyond the point-probability forecast, the operational dashboard described in Section 2.5 is intended to broaden the basis on which forecasters interpret a given SEP alert. SHAP attributions indicate which physical drivers, chiefly SHARP magnetic-field parameters and flare intensity, are shaping a specific forecast, while the similar-period and similar-event analogs situate the current forecast within the broader solar-cycle context and against concrete historical precedents with known outcomes. This case-based layer does not alter the underlying probability but is intended to help forecasters assess the plausibility and historical precedent of a given warning, in a manner analogous to how ensemble and analog techniques are used to contextualize probabilistic forecasts in terrestrial weather prediction [30, 31].

The case studies also highlight several directions for future improvement. Although SEPNET-PRISM-R successfully forecasted the two major SEP events during the evaluation period, it failed to provide an early warning for the relatively weak event in May 2026. This event appears to have been more strongly associated with CME evolution than with the magnetic and flare precursors emphasized in the current model. Our supplementary analysis indicates that the available CME variables provide limited incremental predictive information beyond SHARP and flare parameters, likely because their current representation does not adequately capture eruption-specific characteristics. Future work shall therefore investigate more event-centered CME descriptions, together with longer operational evaluation over additional years of Solar Cycle 25, to further assess model robustness and generalization.

In summary, our numerical results indicate that combining multitask representation learning with explicit temporal post-processing and SHAP-based interpretability yields a forecasting system whose reliability and transparency are well-suited to operational deployment. By leveraging physically meaningful magnetic-field and flare-related predictors alongside an operational web dashboard that provides real-time probabilities, feature attributions, and historical analogs, this system represents a critical advancement toward reliable, interpretable, and continuous machine-learning architectures for space weather forecasting.

4 Methods

This section describes the methodology for developing the hourly near-real-time SEP forecasting framework and the SHAP-based feature importance analysis.

4.1 Forecasting Design and Sample Construction

In the intended real-time forecasting setting, SEP predictions are issued hourly using predictor information collected over the preceding 24 hours. This design inherently produces input samples with substantial temporal overlap, such that consecutive samples share highly similar predictor histories. Constructing a full training dataset using one-hour-ahead sliding windows, therefore, induces strong dependence among samples. Under these conditions, a conventional random split into training and validation sets is not appropriate because neighboring samples are nearly identical, and can place highly similar instances in both sets, leading to information leakage. This leads to overly optimistic estimates of predictive performance and compromises the validity of generalization assessment. To address this issue, the training dataset for model development is constructed using non-overlapping 24-hour windows, as illustrated in Figure 7, following the procedure in [15] and [17], which enforces a clearer separation among samples and yields more reliable estimates of out-of-sample predictive skill.

The training data is constructed using observations from February 1986 through September 2025, consistent with the temporal coverage of the SEP benchmark dataset [32]. During this period, models are trained and validated across 50 random seeds to account for variability due to data partitioning and stochastic optimization. The resulting fitted models are then fixed and analyzed to evaluate the stability and reproducibility of predictive performance. To emulate near-real-time forecasting, a

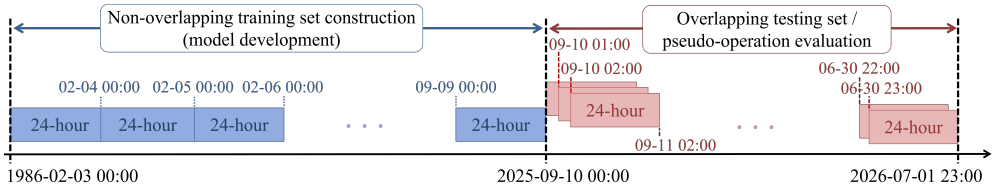


Fig. 7: Illustration of the training and evaluation sample construction. Model development adopts non-overlapping 24-hour windows from February 1986 to September 2025 to avoid dependence and information leakage, while pseudo-operational evaluation uses hourly forecasts over an independent post-training period from September 10, 2025, to July 1, 2026.

separate post-training interval beginning immediately after the end of the training period and continuing through 1 July 2026. This interval is excluded from all model training and tuning procedures and is treated as a pseudo-operational test period. Performance over this period is therefore intended to reflect how the forecasting system would operate if the model were to be adopted for real-time use.

The forecasting framework is built on the enhanced multitask deep learning model `SEPNET-PRISM` [15]. At each forecast issue time, the model ingests solar observations from the preceding 24 hours, including active-region patch information, SHARP parameters, and solar-flare characteristics. In addition to predicting SEP occurrence, `SEPNET-PRISM` is designed as a multitask learning model that jointly forecasts three auxiliary quantities: the maximum logarithmic proton flux in the >10 MeV channel, the GOES soft X-ray flux, and the CME speed within the next 24 hours. These auxiliary targets are selected to capture key physical processes associated with SEP generation and to support shared representation learning across related tasks.

4.2 Hidden Semi-Markov Model (HSMM) for Post-Processing

At each forecast time, the previous model `SEPNET-PRISM` estimates the probability of an SEP event within the next 24 hours using observations from the previous 24-hour window. In near-real-time operation, forecasts are updated hourly using a sliding window, resulting in overlapping inputs and temporal dependence between successive predictions. While the neural network generates each probability independently, this overlap induces autocorrelation, leading to short-term fluctuations that are inconsistent with the typically continuous evolution of the process during SEP events. Consequently, probabilities near decision thresholds may produce fragmented or unstable event identification.

To mitigate this issue, `SEPNET-PRISM-R` incorporates a post-processing step based on an HSMM [19], which explicitly models temporal dependence via Markov transitions between states and state persistence via a sojourn-time distribution. Let S_t denote the latent state at time t, where $S_t = 0$ represents a quiet state and $S_t = d$, for $d = 1, \ldots, D$, represents an SEP event with elapsed duration d hours. Unlike standard hidden Markov models, the HSMM accounts for duration in each state, enabling a more realistic representation of event continuity. Integrating information over time

produces a smoothed, temporally coherent sequence of event probabilities. To prevent information leakage, the HSMM is trained using only the original development period. An auxiliary dataset is constructed using the same hourly sliding-window procedure, and the trained and frozen `SEPNET-PRISM` model is then applied to generate predicted SEP probabilities. These serve as observations for HSMM training, allowing the model to learn the temporal dependence structure and uncertainty characteristics of the neural network outputs. Importantly, no data from the post-training evaluation period are used in this process, ensuring a fair assessment of the out-of-sample performance.

The HSMM takes as input the `SEPNET-PRISM` output of SEP event probability and physically motivated covariates derived from the preceding one hour before the forecast issue time, including the logarithmic maximum proton flux and GOES soft X-ray flux. Specifically, let $x_{\mathrm{prob},t}$ denote the standardized logit-transformed `SEPNET-PRISM` probability at time t, where the logit transform maps a probability $p_t \in (0,1)$ to the real line via $\log\{p_t/(1-p_t)\}$. Let $x_{\mathrm{pf},t}$ and $x_{\mathrm{cme},t}$ denote the logarithmic maximum proton flux and XRSB flux over the preceding one hour, respectively. These covariates are then used to estimate two time-varying hazards: an entry hazard that governs transitions from quiet to event, and an exit hazard that governs transitions from event back to quiet. The estimated entry hazard is given by

$$\mathrm{logit}\{P(S_t > 0 \mid S_{t-1} = 0, \mathbf{x}_t)\} = -8.06 + 1.43\, x_{\mathrm{prob},t} - 0.74\, x_{\mathrm{pf},t} - 0.07\, x_{\mathrm{xrs},t},$$

which indicates that higher `SEPNET-PRISM` probabilities $x_{\mathrm{prob},t}$ increase the likelihood of transitioning from a quiet state to an event state, whereas the proton flux $x_{\mathrm{pf},t}$ and XRSB $x_{\mathrm{xrs},t}$ covariates exert weaker negative effects. The large negative intercept implies a low baseline entry probability unless the covariate evidence is sufficiently strong. The estimated exit hazard is

$$\mathrm{logit}\{P(S_t = 0 \mid S_{t-1} = d, \mathbf{x}_t)\} = 0.54\, d_t - 2.58 - 0.40\, x_{\mathrm{prob},t} - 0.15\, x_{\mathrm{pf},t} - 0.31\, x_{\mathrm{xrs},t},$$

where d_t is the elapsed duration of the current event state. The positive duration coefficient indicates that longer event runs are increasingly likely to terminate, whereas stronger evidence from event-related covariates decreases the exit probability. This duration-dependent structure helps prevent unrealistically long event sequences and promotes more physically plausible event segmentation.

A weight parameter $w \in [0,1]$ is introduced to control the influence of the neural network output in the emission model (we selected $w = 0.44$). Larger values of w increase sensitivity but may raise false positives, while smaller values favor conservative predictions. The parameter is selected from the training set to maximize true positives while constraining false positives relative to the baseline model. In implementation, the HSMM expands event states into duration-specific substates up to a fixed maximum duration (72 hours). Forward–backward inference is used to compute smoothed posterior probabilities at each time step, and final event predictions are obtained by thresholding these probabilities.

4.3 SHAP-Based Feature Importance Analysis

To interpret the trained forecasting model, we conduct a SHAP-based feature importance analysis [20, 21]. SHAP values quantify the marginal contribution of each predictor to the model output within a game-theoretic framework, enabling evaluation of individual features in a consistent, additive manner. We use SHAP analysis at both the global and local levels. At the global level, aggregated absolute SHAP values are used to rank predictors according to their overall contribution to the SEP forecasting model. At the local level, SHAP values are examined for individual forecasts to determine how specific input features increase or decrease the predicted probability of SEP. This analysis provides insight into the relative importance of active-region properties, magnetic-field parameters, and flare-related variables, thereby improving the interpretability and operational transparency of the proposed framework.

4.4 Data Preparation

The primary prediction target in this study is the occurrence of operational SEP events, defined as intervals during which the proton flux exceeds 10 pfu in the > 10 MeV energy channel. Event labels and events are obtained from the CLEAR SEP benchmark dataset through the `FetchSEP` python module [28], which provides event information spanning 3 February 1986 to 10 September 2025. A comprehensive description of data sources, preprocessing procedures, and quality control is provided in the dataset descriptor [33], with the dataset publicly archived on Zenodo [34]. Following prior work [15], several groups of candidate predictors were considered, including magnetic-field parameters, flare-related variables, CME-related variables, soft X-ray flux, and historical proton flux. Comparative analyses indicate that magnetic-field parameters derived from SHARP-consistent data products, together with flare-related variables, provide the most stable predictive performance. In contrast, CME-related variables contribute less consistently, as detailed in the Supplementary Information. Although CME features are physically relevant to SEP production, our analysis shows that they are substantially redundant with the SHARP-based magnetic and flare predictors and provide only limited incremental value for the operational classification task. Historical proton flux can improve short-term event detection but can encourage the model to rely on immediate precursor signatures, thereby reducing the effective forecast lead time. Accordingly, the final predictor set is restricted to magnetic-field and flare-related features, aligning with the `SEPNET-O` model in [17].

All predictors are drawn from the curated, multi-source dataset and are represented in a temporally aligned and quality-controlled format. To construct supervised learning samples, predictor variables are aggregated over the 24-hour period preceding each forecast issuance time. For each variable, summary statistics (minimum, mean, and maximum) are computed to characterize recent solar conditions. These predictors are paired with targets defined over the subsequent 24-hour forecast horizon, including a binary indicator of SEP occurrence and auxiliary regression targets for proton flux, soft X-ray flux, and CME speed. This formulation yields a consistent dataset for joint classification and regression modeling of SEP-related activity.

4.5 Analog-Based Contextualization: Similar Solar-Cycle Periods and Similar Precursor Windows

To complement the forecast and support situational interpretation, the deployment of `SEPNET-PRISM-R` also provides two forms of analog-based context: (i) historical years with a comparable level of solar-cycle activity to the ongoing year, and (ii) historical 7-day precursor windows whose flare and magnetic-field characteristics closely resemble the current forecast window. Both analogs are automatically recomputed at each hourly dashboard refresh and are intended to provide forecasters with an empirical, case-based reference for interpreting current model output, complementing the model-based probability and the SHAP attributions described above.

The first analog situates the current forecast within the broader solar cycle. For the current year-to-date, a Poisson rate λ is fit to the daily sunspot number, providing a compact summary of the ambient level of solar activity, and each historical year in the training archive (1986 to present) is assigned an analogous λ estimated over the same fitting window. Candidate years are ranked in ascending order of $|\Delta\lambda|$, the absolute difference between the current year's λ and each historical year's λ, and the five years with the smallest $|\Delta\lambda|$ are reported as similar periods. For each analog year, the realized classification performance of the HSMM-based warning system, evaluated against the operational SEP label for that year (ACC, HSS, TSS, POD, and FAR), is also reported, allowing users to assess how the forecasting system historically performed during a comparably active/inactive solar-cycle period.

A second, independent analog examines the shape of the immediate-precursor input features. The dashboard identifies historical, non-overlapping 7-day windows whose flare-related features and SHARP parameters summaries most closely resemble those of the current 7-day forecast window. Candidate windows are drawn from the full historical archive and ranked using a z-scored Euclidean distance computed over flare-activity summaries (e.g., flare count and maximum and mean logarithmic peak flux) and SHARP magnetic-field summaries (e.g., total unsigned flux, the R-value, and mean shear angle); cosine similarity between the same feature vectors is reported alongside the distance as a complementary, scale-invariant measure of directional agreement. The five nearest historical windows are displayed as similar events, along with the realized SEP outcome and HSMM-based classification performance for each matched window, providing an event-level, case-based complement to the aggregate SHAP-based interpretation and allowing users to examine specific historical analogs of the current precursor configuration.

Supplementary information. Supplementary Information contains additional details on the CME-related predictor comparison referenced in Section 4.4.

Acknowledgements. The authors thank Ian Richardson (University of Maryland, NASA GSFC) and A. Steve Johnson (Leidos, NASA JSC SRAG) for their efforts in acquiring and preparing the dataset used in this study.

Declarations

Funding. This work is supported by the NASA Space Weather Center of Excellence program under award No. 80NSSC23M0191 and No. 80NSSC23M0192.

Conflict of interest/Competing interests. The authors declare no competing interests.

Ethics approval and consent to participate. Not applicable. This study did not involve human participants, human data, or animal subjects.

Consent for publication. Not applicable.

Data availability. The training dataset underlying `SEPNET-PRISM-R` is archived on Zenodo [34], and the SEP event benchmark used to construct it is described in [28] and accessible through the `FetchSEP` python module. The operational forecasts generated by `SEPNET-PRISM-R`, including real-time event probabilities, auxiliary regression outputs, SHAP-based feature attributions, and the similar-period and similar-event analogs described in Methods, are made publicly available through an hourly updated web dashboard at https://mlsw.engin.umich.edu/apps/SEPNET-PRISM-R, hosted by the University of Michigan Space Weather Interactive Lab (SWIL).

Materials availability. Not applicable. This study did not generate new physical materials.

Code availability. The code implementing the `SEPNET-PRISM-R` forecasting framework, including model training, the HSMM temporal post-processing layer, and the SHAP-based interpretability analysis, is publicly available at https://github.com/yuyian/SEPNET-PRISM-R.

Author contribution. Y.Y. and Y.C. conceived the study. Y.Y. conducted the data preparation, model development, and experiments. Y.Y. and Y.C. analyzed the results and wrote the manuscript. L.Z., K.W., W.M., and T.G. contributed to the scientific interpretation and revision of the manuscript. All authors reviewed the manuscript.

References

[1] Reames, D. V. Particle acceleration at the Sun and in the heliosphere. *Space Science Reviews* **90**, 413–491 (1999).

[2] Desai, M. & Giacalone, J. Large gradual solar energetic particle events. *Living Reviews in Solar Physics* **13**, 3 (2016). URL https://doi.org/10.1007/s41116-016-0002-5.

[3] Zhao, L. *et al.* Solar wind with field lines and energetic particles (SOFIE) model: Application to historical solar energetic particle events. *Space Weather* **22**, e2023SW003729 (2024).

[4] Schwadron, N. A. *et al.* Earth-Moon-Mars radiation environment module framework. *Space Weather* **8** (2010).

[5] Shea, M. & Smart, D. Space weather and the ground-level solar proton events of the 23rd solar cycle. *Space science reviews* **171**, 161–188 (2012).

[6] Whitman, K. *et al.* Review of solar energetic particle prediction models. *Advances in Space Research* **72**, 5161–5242 (2023). URL https://www.sciencedirect.com/science/article/pii/S0273117722007244.

[7] Kasapis, S. *et al.* Review of machine learning models for solar energetic particle prediction (2026). URL https://arxiv.org/abs/2606.19539. arXiv:2606.19539.

[8] Kahler, S. W. The role of the big flare syndrome in correlations of solar energetic proton fluxes and associated microwave burst parameters. *Journal of Geophysical Research: Space Physics* **87**, 3439–3448 (1982).

[9] Gopalswamy, N. *et al.* Interacting coronal mass ejections and solar energetic particles. *The Astrophysical Journal Letters* **572**, L103–L107 (2002).

[10] Laurenza, M. *et al.* A technique for short-term warning of solar energetic particle events based on flare location, flare size, and evidence of particle escape. *Space Weather* **7** (2009).

[11] Camporeale, E., Wing, S. & Johnson, J. *Machine Learning Techniques for Space Weather* (Elsevier, 2018). URL https://books.google.com/books?id=zaRBDwAAQBAJ.

[12] Papaioannou, A., Anastasiadis, A., Sandberg, I. & Jiggens, P. Nowcasting of Solar Energetic Particle Events using near real-time Coronal Mass Ejection characteristics in the framework of the FORSPEF tool. *Journal of Space Weather and Space Climate* **8**, A37 (2018).

[13] Kasapis, S. *et al.* Interpretable machine learning to forecast SEP events for solar cycle 23. *Space Weather* **20**, e2021SW002842 (2022). URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2021SW002842.

[14] Camporeale, E. The challenge of machine learning in space weather: Nowcasting and forecasting. *Space Weather* **17**, 1166–1207 (2019).

[15] Yu, Y. *et al.* Realtime forecasting of solar energetic particle event and proton flux using multi-source solar observations and multi-task deep learning. *Scientific Reports* (2026).

[16] Roberts, D. R. *et al.* Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography* **40**, 913–929 (2017).

[17] 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).

[18] NASA Community Coordinated Modeling Center (CCMC). SEPNET 1 model page. https://ccmc.gsfc.nasa.gov/models/SEPNET~1/ (2026). Accessed: 2026-09-24.

[19] Yu, S.-Z. Hidden semi-markov models. *Artificial Intelligence* **174**, 215–243 (2010). Special Review Issue.

[20] Iong, D. *et al.* New findings from explainable sym-h forecasting using gradient boosting machines. *Space Weather* **20**, e2021SW002928 (2022). URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2021SW002928.

[21] Wang, H., Liang, Q., Hancock, J. T. & Khoshgoftaar, T. M. Feature selection strategies: A comparative analysis of SHAP-value and importance-based methods. *Journal of Big Data* **11**, 44 (2024).

[22] Muschelli III, J. ROC and AUC with a binary predictor: A potentially misleading metric. *Journal of classification* **37**, 696–708 (2020).

[23] Lipton, Z. C., Elkan, C. & Narayanaswamy, B. Thresholding classifiers to maximize F1 score (2014). URL https://arxiv.org/abs/1402.1892. arXiv:1402.1892.

[24] Wehling, P., LaBudde, R. A., Brunelle, S. L. & Nelson, M. T. Probability of detection (pod) as a statistical model for the validation of qualitative methods. *Journal of AOAC International* **94**, 335–347 (2011).

[25] Macmillan, N. A. & Kaplan, H. L. Detection theory analysis of group data: Estimating sensitivity from average hit and false-alarm rates. *Psychological bulletin* **98**, 185 (1985).

[26] Doswell, C., Davies-Jones, R. & Keller, D. L. On summary measures of skill in rare event forecasting based on contingency tables. *Weather and forecasting* **5**, 576–585 (1990).

[27] Leka, K. D. *et al.* A comparison of flare forecasting methods. II. Benchmarks, metrics, and performance results for operational solar flare forecasting systems. *The Astrophysical Journal Supplement Series* **243**, 36 (2019). URL https://doi.org/10.3847/1538-4365/ab2e12.

[28] Whitman, K. FetchSEP: CLEAR benchmark dataset version 1.0 (2025). URL https://github.com/ktindiana/fetchsep/releases/tag/CLEAR_Benchmark_v1.0.

[29] Vourlidas, A. & Howard, R. A. The proper treatment of coronal mass ejection brightness: A new methodology and implications for observations. *The Astrophysical Journal* **642**, 1216 (2006).

[30] Van den Dool, H. M. A new look at weather forecasting through analogues. *Monthly Weather Review* **117**, 2230–2247 (1989).

[31] Delle Monache, L., Nipen, T., Liu, Y., Roux, G. & Stull, R. Probabilistic weather prediction with an analog ensemble. *Monthly Weather Review* **141**, 3498–3518 (2013).

[32] Whitman, K., Egeland, R., Allison, C., Quinn, P. & Stegeman, L. Validation of solar energetic particle forecasting models for space radiation operations with SPHINX and VIVID. Tech. Rep. NASA/TP-20260000463, National Aeronautics and Space Administration (2026). URL https://ntrs.nasa.gov/citations/20260000463.

[33] Yu, Y. *et al.* SEP-PRISM Data: A multi-source dataset for solar energetic particle forecasting (2026). URL https://arxiv.org/abs/2607.16160. arXiv:2607.16160.

[34] Yu, Y. *et al.* SEP-PRISM Data: A multi-source dataset for solar energetic particle forecasting (2026).