Solar Energetic Particle Prediction in the Inner Heliosphere Using Deep Learning and PSP/IS⊙IS Data

Tate Hutchins[1,2], Spiridon Kasapis[2], Hameedullah Farooki[3], Manuel E. Cuesta[2], Lengying Khoo[2], Sungmin Pak[2], Robert Czarnota[3], Jamie Rankin[2], Jamey Szalay[2], Zigong Xu[4], Georgios Livadiotis[2], Nikolaos Sarlis[5], David McComas[2]

[1]Department of Computer Science, Princeton University, Princeton, NJ, USA
[2]Department of Astrophysical Sciences, Princeton University, Princeton, NJ, USA
[3]Department of Mathematics, Rowan University, Glassboro, NJ, USA
[4]Division of Physics, Mathematics and Astronomy, California Institute of Technology, Pasadena, CA, USA
[5]Physics Department, National and Kapodistrian University of Athens, Athens, Greece

Abstract

Solar events, such as coronal mass ejections and solar flares, accelerate large numbers of energetic charged particles, producing solar energetic particle (SEP) events that can harm astronauts, damage satellites, and potentially damage infrastructure on Earth. Previous machine learning (ML) SEP prediction models aimed to predict geoeffective events using data from Earth-orbiting or Sun-Earth Lagrange 1 (L1) point missions. However, predicting SEP events throughout the heliosphere is essential for supporting missions to Mars and other deep-space destinations. This study uses EUV images from the Atmospheric Imaging Assembly (AIA) onboard the Solar Dynamics Observatory (SDO) to predict particle intensity values observed by the Integrated Science Investigation of the Sun (IS⊙IS) onboard the Parker Solar Probe (PSP), over various points in the ecliptic plane between Sun and Earth, using deep learning. The ML models extract features from SDO/AIA 171 Å images through a series of convolutional layers, combine them with PSP's trajectory, and process the fused representation using transformer and dense layers. The output is a regression estimate of the energy-weighted average particle intensity (J_{linlin}), which is subsequently thresholded to yield a binary classification indicating whether $J_{\mathrm{linlin}} \geq 0.1\ \frac{\mathrm{counts}}{\mathrm{cm^2\,s\,sr\,keV}}$. The best-performing model achieves a probability of detection (POD) of 0.76, true skill score (TSS) of 0.47, F1 score of 0.47, false alarm rate (FAR) of 0.65 and precision of 0.35. These results are comparable to ML models that predict geoeffective events, opening avenues towards ML-aided predictions in the inner-heliosphere.

Plain Language Summary

Solar storms can accelerate large numbers of energetic particles that pose risks to astronauts, satellites, and future deep-space missions. Most existing prediction models focus on events that affect Earth, but missions to Mars and beyond require forecasts throughout the inner Solar System. This study uses solar images from NASA's Solar Dynamics Observatory (SDO) spacecraft together with measurements from the Parker Solar Probe (PSP) to predict energetic particle intensities at different locations between the Sun and Earth. A deep learning model analyzes extreme-ultraviolet images of the Sun along with the probe's position to estimate particle intensities and determine when a significant solar energetic particle (SEP) event is likely. The best performing model accurately identifies most events and achieves skill levels comparable to Earth-focused prediction efforts. These results show that machine learning can meaningfully connect what we see on the Sun to conditions experienced by spacecraft across the heliosphere, offering a promising step toward space-weather forecasting for deep-space exploration.

1 Introduction

SEP events are bursts of high-energy particles from the Sun, often linked to solar flares and coronal mass ejections (CMEs). They were detected as early as 1937, when S. E. Forbush observed sudden cosmic ray intensity increases following solar activity (Forbush, 1937; Hess & Demmelmair, 1937; Forbush, 1946). Since then, a substantial body of literature has been devoted to SEP events, specifically investigations into their sources (Reames, 2013), statistical properties (Hristopulos et al., 2026), arrival features (Xu et al., 2026), composition (e.g., Cane et al., 2010), among other properties, aimed to understand particle acceleration and transport (Desai et al., 1998; Giacalone, 2005; Desai & Giacalone, 2016; Klein & Dalla, 2017; Cuesta et al., 2025), as well as improve space weather prediction capabilities (Whitman et al., 2023).

Reliable SEP prediction is a key goal of space weather science, driven by the need to understand and predict these high-energy particle bursts that can disrupt satellite operations, pose radiation hazards to astronauts (Cucinotta & Pak, 2024; Pak & Cucinotta,

2024), and affect the Earth's magnetosphere and ionosphere (Boteler, 2001; Oughton et al., 2017). Recently, machine learning (ML) has emerged as a new tool that promises to aid with the prediction of space weather-related events, such as solar flares (Jiao et al., 2020), CMEs **(Pricopi et al., 2022; Chierichini et al., 2024; Zhou et al., 2026)**, and active regions (ARs; Kasapis et al., 2023; Kosovichev et al., 2025; Keegan et al., 2025; Kasapis, Kitiashvili, et al., 2025). For predicting SEP events specifically, a wide variety of models have been trained, ranging from low-level support vector machines or decision trees **(Laurenza et al., 2018; Kasapis et al., 2022, 2024; Papaioannou et al., 2025)**, to neural networks (Aminalragia-Giamini et al., 2021; Chatterjee et al., 2024) and recently even foundation models (Roy, Schmude, et al., 2025; Roy, Hegde, et al., 2025).

Given the critical importance of forecasting geoeffective space-weather events, most ML models developed to predict SEP events have relied on data from Earth-orbiting satellites—such as the Solar Dynamics Observatory (SDO; Pesnell et al., 2012) and the Geostationary Operational Environmental Satellites (GOES; Menzel & Purdom, 1994)—or from missions orbiting the Sun–Earth L1 point, including the Solar and Heliospheric Observatory (SOHO; Domingo et al., 1995), the Advanced Composition Explorer (ACE; Stone et al., 1998), and Wind (Harten & Clark, 1995). With upcoming missions to deep space, the ability to predict SEPs throughout the inner heliosphere has become increasingly important for protecting space assets and ensuring the safety of crewed spacecraft (Kuehnast et al., 2022). Although data coverage is more limited away from near-Earth space, recent missions such as Parker Solar Probe (PSP; Fox et al., 2016; McComas et al., 2019; Raouafi et al., 2023) and Solar Orbiter (SolO; Müller et al., 2020) are now providing valuable measurements from previously inaccessible vantage points.

This study aims to develop ML models that predict SEP events throughout the inner heliosphere, and are therefore not restricted to only geoeffective forecasting. Using extreme ultraviolet (EUV) imagery from the Atmospheric Imaging Assembly (AIA; Lemen et al., 2012) onboard Solar Dynamics Observatory (SDO; Pesnell et al., 2012) as inputs, and in-situ energetic proton intensity measurements from the Integrated Science Investigation of the Sun (IS⊙IS; McComas et al., 2016) onboard PSP as training targets, we train a Convolutional Neural Network (CNN) to predict the energy-weighted average particle intensity and test its capability for binary classification of whether an SEP event is likely to occur.

This manuscript is organized as follows. Section 2 describes in detail the datasets used to train the CNNs, and Section 3 outlines the training methodology. Section 4 presents the forecasting performance of the best-performing ML models, while Sections 5 and 6 provide a discussion and the main conclusions of this study. Given that PSP measurements remain underutilized in space-weather ML applications (Kasapis, Cuesta, et al., 2025), this work demonstrates that data from spacecraft operating away from near-Earth space can provide valuable information for SEP prediction frameworks (Sarlis et al., 2025).

2 Dataset

Launched in 2018, PSP is the first mission to directly sample the Sun's corona, providing in-situ measurements of the solar wind and energetic particles in the near-Sun environment (Fox et al., 2016; Raouafi et al., 2023). Among its four instrument suites, IS⊙IS (McComas et al., 2016) is designed to measure ions and electrons over a broad range of particle energies using two Energetic Particle Instruments: EPI-Lo (Hill et al., 2017; Wiedenbeck et al., 2017) and EPI-Hi (Wiedenbeck et al., 2021; Pak et al., 2025). EPI-Lo detects ions from $\sim$ 20 keV/nucleon–15 MeV total energy and electrons from $\sim$ 25–1000 keV, using 80 apertures distributed over a hemisphere to capture at least half of the pitch-angle distribution.

In this study, data from EPI-Lo are used to calculate the energy-weighted average of proton intensity (J_{linlin}, Cuesta et al., 2025) over the first 22 PSP orbits (January 2019 - January 2025). Specifically, J_{linlin} was calculated over 100 keV to 1 MeV, a subset of the available energy range that EPI-Lo measures. This is because the detector's efficiency drops significantly for protons over 1 MeV. The calculation of J_{linlin} for EPI-Hi was not included in this analysis as more careful treatment is needed to separate out contributions from anomalous and galactic cosmic rays in order to properly obtain the higher-energy baseline (see, e.g., Rankin et al., 2022). EPI-Lo particle intensities are calculated via the energy-averaged intensity in the linear space of intensity and energy width (e.g., Cuesta et al., 2025):

$$J_{\text{linlin}} = \frac{\Sigma_i^{N-1} j(E_{g,i})\Delta E_i}{\Sigma_i^{N-1} \Delta E_i}, \tag{1}$$

where N is the number of discrete energy bins to be combined, and $E_{g,i}$ is the log-centered energy of the i-th energy bin of width ΔE_i for i starting from zero. Figure 1 shows the IS⊙IS EPI-Lo J_{linlin} values for the first 22 orbits of the mission **(12-hour average in black and the 12-hour non-zero minimum in red)**. The discrete change in the minimum J_{linlin} value observed after 2021 is due to the difference in the EPI-Lo look-up tables. We note that for training of the ML models, J_{linlin} is downsampled to a cadence of three hours using a non-overlapping moving average in order to match the SDO EUV imagery data cadence.

Launched in 2010, the SDO was designed to observe the solar atmosphere at high spatial and temporal resolution, and across multiple wavelengths simultaneously (Pesnell et al., 2012). Among its instruments is AIA (Lemen et al., 2012), which provides imagery of the solar corona. AIA captures images spanning at least 1.3 solar diameters in several EUV and UV wavelengths nearly simultaneously, with a spatial resolution of about 1 arcsecond and a cadence of 12 seconds. In this study, the 171 Å channel images are used as they capture the upper coronal plasma, which is closer to the solar wind sampled by PSP. **Note here that although the AIA EUV channels are known to suffer sensitivity degradation over time, no degradation correction was applied to the images. D**ue to data transfer and storage limitations along with training time constraints, the SDO images are sampled every 3 hours, at 0:00, 3:00, 6:00, 9:00, 12:00, 15:00, 18:00, and 21:00 of each day. This yields over 17,000 images in total over the PSP mission timeline.

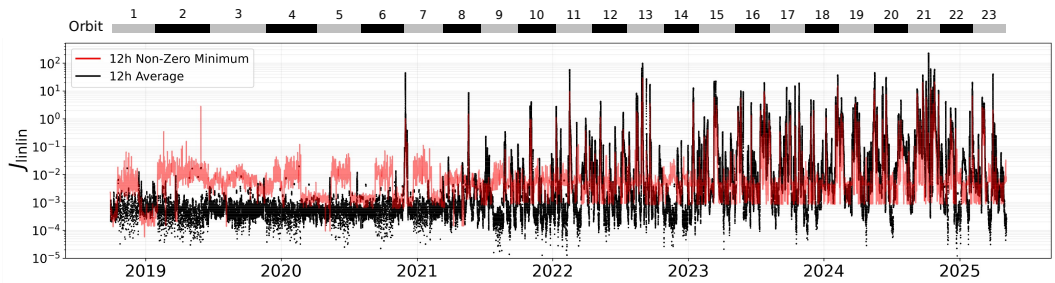

Figure 1. EPI-Lo energy-averaged (100 keV to 1 MeV) intensity J_{linlin} values in units of $\frac{\text{counts}}{\text{cm}^2\,\text{s}\,\text{sr}\,\text{keV}}$ on a logarithmic scale, from the beginning of the mission up to June, 2025. The timelines present the 12-hour non-zero minimum (red) and the 12-hour averaged J_{linlin} values. Solar cycle dependence and instrumental effects (change in EPI-Lo look-up tables) are observed.

The seven-year overlap between PSP and SDO observations has provided numerous opportunities to sample the inner heliosphere at distances significantly closer to the Sun than the Lagrange 1 (L1) point, enabling studies of how solar events observed by SDO affect the near-Sun environment (de Pablos et al., 2022; Cohen et al., 2021). Because SDO observes only the Earth-facing hemisphere of the Sun (the visible disk), energetic particles originating from coronal events seen by SDO are most likely to be detected by PSP when the spacecraft solar magnetic footprint's Stonyhurst longitude lies within the range of −90 deg to 90 deg, corresponding to the Earth-facing half of the heliosphere. In this study, we adopt a more conservative longitudinal window and use only EPI-Lo measurements acquired when PSP was located between −80 deg and 80 deg Stonyhurst longitude for training the ML models in order to ensure that the corresponding events are visible on the observable solar disc. Future studies can extend this longitude range to include coronal features that might be magnetically connected to the PSP. Figure 2 illustrates the regions sampled by PSP throughout its mission that fall within this selected longitudinal range.

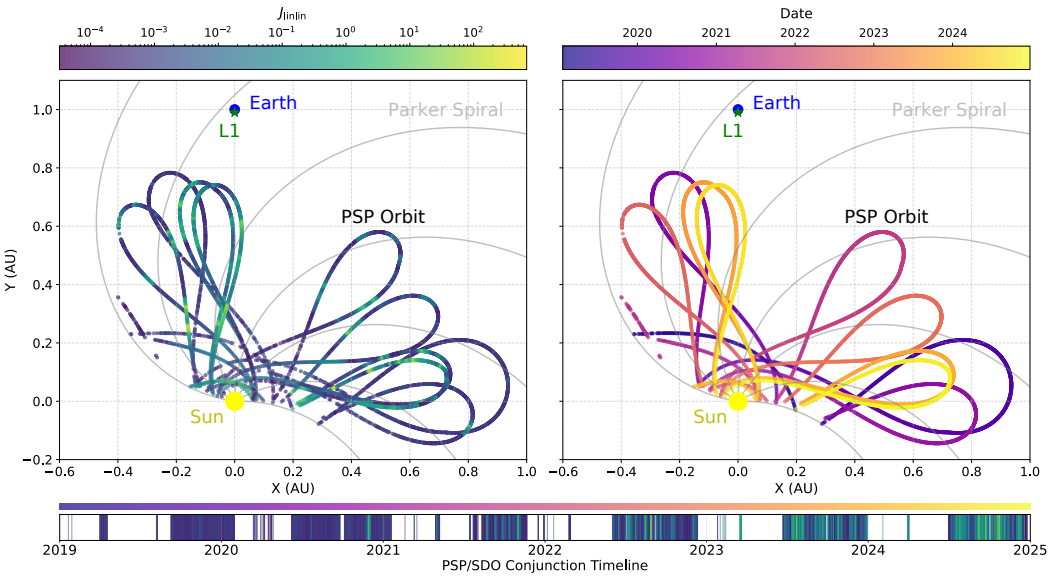

Figure 2. The locations for which data from PSP were used in this study, after the −80 deg and 80 deg Stonyhurst longitude coordinate threshold and the Parker spiral assumption were applied. The positions plotted are in a fixed Sun-Earth reference frame in the Stonyhurst coordinate system. **Each PSP Orbit point in the left and right panels represents the location for which a sample was used in this study, while the coloring represents the J_{linlin} value and the date of each sample, respectively.** The yellow and blue circles represent the Sun and Earth respectively. The Lagrange 1 point (L1), where many of the in-situ data of previous studies were observed, is represented by a green star. In grey are the Parker spiral lines starting at −80 deg and ending at 80 deg Stonyhurst longitude. The x and y coordinates are in AU, but the Sun and Earth are not to scale. **The timeline under the plots summarizes them by presenting the PSP and SDO conjunction times and their associated intensity.**

To date, all ML models developed for SEP prediction rely on in-situ measurements from spacecraft located either within Earth's orbit or at the Sun–Earth L1 point (Figure 2, blue dot and green star). After excluding J_{linlin} measurements obtained when PSP was located on the far side of the Sun relative to Earth, a total of 8102 data points re-

main available for constructing the training and **testing** datasets. **These data points are plotted in the left and right panels of Figure 2, and are colored based on the J_{linlin} value and date during which they occurred, respectively.** To synchronize SDO observations with the corresponding PSP measurements, we assume a constant energetic proton propagation speed of 4000 km s^{-1}, approximately a 100keV proton's speed (the lower boundary of the EPI-Lo instrument energy range). The spacecraft speed is considered to be negligible because it is most often a small fraction of the energetic protons velocities. For an SDO image acquired at time T, the time it takes for energetic protons to arrive at PSP, located at heliocentric distance R, is then estimated as:

$$t_{\text{offset}} = \frac{S_R}{4000 \text{ km s}^{-1}},$$

where S_R is the Parker spiral path-length (Laitinen & Dalla, 2019, Appendix A), so that the first particles associated with the observed solar event would reach PSP at $T{+}t_{\text{offset}}$.

During typical SEP events, the majority of particles arrive to PSP within a ≤ 12 **hour window (Wang & Qin, 2023)**, although this arrival is variable due to the highly-variable range of heliocentric radial distances. Therefore, the latest expected arrival time for particles originating from the event observed by SDO is approximated as $T{+}t_{\text{offset}}{+}12$ hr. Accordingly, for each SDO image taken at time T, the PSP measurements—including spacecraft coordinates, solar–wind speeds, and EPI-Lo J_{linlin} intensities—are averaged over the interval:

$$[\,T + t_{\text{offset}},\ T + t_{\text{offset}} + 12 \text{ hr}\,].$$

In this study, for a given SDO image, whether it captures a space–weather event or not, we aim to use ML models to predict the energy-averaged J_{linlin} intensity at the location of PSP, assuming magnetic connectivity between the spacecraft and the Sun as described by the Parker spiral (Parker, 1958; Smith, 2001). To inform the CNN of this magnetic connection, we use not only the SDO/AIA 171 Å images as inputs, but also PSP's heliocentric distance (R_{PSP}) and its magnetic footpoint on the solar surface. The magnetic footpoint for each time step is computed using the `SolarMACH` Python library (Gieseler et al., 2019), which requires the spacecraft identifier (PSP), the corresponding timestamp (from the PSP dataset), and the solar-wind speed, which is derived from measurements of the Solar Wind Electrons Alphas and Protons (SWEAP; Kasper et al., 2016) instrument on board of PSP. Here, we use the level3 SPAN-I solar wind proton speed magnitude (Livi et al., 2022), which is less prone to field-of-view issues. `SolarMACH` provides the magnetic footpoint at a particular time in Heliographic coordinates, so we transform to the Stonyhurst coordinate system to stay consistent with SDO/Earth's perspective. Table 1 presents a summary of the data used in this study.

The SDO/AIA images are provided at 4096×4096 resolution. To reduce computational cost and training time, the images are downsampled to 512×512 (as done in similar studies, e.g., Majid et al., 2024; Walsh et al., 2024), reducing the dataset size by a factor of 64 while also retaining the relevant coronal structure. Over the entire PSP/SDO overlap times and after the aforementioned longitudinal filtering, 8,102 SDO images were used at a 3-hour cadence. After this filtering, only 44% of PSP intervals fall inside the effective AIA field of view. Before being passed to the model, all inputs are normalized. The EUV images' pixel values span several orders of magnitude, from near zero to $\sim 10^4$ during intense events, so a logarithmic transformation is applied to compress the dynamic range while preserving contrast between low- and moderate-intensity pixels. The same transformation is applied to the EPI-Lo J_{linlin} intensities, which range from 10^{-5} to 10^2 $\frac{\text{counts}}{\text{cm}^2\,\text{s}\,\text{sr}\,\text{keV}}$. After the log transforms, the different inputs span different numerical ranges (e.g., footpoint longitude in $[-80°, 80°]$ and $R_{\text{PSP}} \leq 1$ AU). To prevent any single input from dominating the training, each quantity is independently normalized by its maximum absolute value (min-max normalization), ensuring that all inputs fall within $[0, 1]$ before being inputted into the ML models.

Table 1. Data from the PSP and SDO spacecraft used for ML training and testing. Physical quantities that have Aux under the Usage column indicate the data that were not directly used for ML training or testing, but rather as an auxiliary value in order to calculate a quantity that is seen by the model.

Physical Quantity	**Type**	**Usage**	**Spacecraft**	**Instrument**	**Units**
EUV 171 Å	Image	Input	SDO	AIA	DNs^{-1}
J_{linlin}	Point	Target	PSP	IS⊙IS/EPI-Lo	$\frac{\text{counts}}{\text{cm}^2\,\text{s}\,\text{sr}\,\text{keV}}$
Distance to Sun	Point	Input	PSP	Ephemeris	AU
S/C Coordinates	Point	Aux	PSP	Ephemeris	deg (Heliographic)
Solar Wind Speed	Point	Aux	PSP	SWEAP/SPAN-I	km/s
Magnetic Footpoint	Point	Input	PSP	Ephemeris & SWEAP	deg (Stonyhurst)

The dataset, which after the longitudinal filtering consists of just over 8,000 images (SDO EUV) and point data (PSP coordinates and J_{linlin}) pairs, is split into training and test sets**. Due to data scarcity a separate validation set was not used, and the hyperparameters were tuned based on the training loss to prevent data leakage. A standard random train-test split would also introduce data leakage, since the** SDO images and PSP parameters have a 3-hour cadence **and therefore** temporally adjacent samples are highly correlated**. A** random split would place near-duplicate conditions in both sets**, and to** prevent this, we use a blocked split. The dataset is divided chronologically into blocks of 80 images (10 days for a 3-hour cadence), and 80% of these blocks are randomly assigned to the training set while the remaining 20% form the test set, which the ML models never see during training. This block length is chosen to exceed the characteristic correlation time of the solar features sampled here. With a ∼27-day solar rotation period (Howe, 2009), images separated by 10 days generally correspond to different solar longitudes and are therefore effectively uncorrelated. Such a testing scheme makes the problem the ML model has to solve more difficult, but also more operationally realistic. For evaluating the model (Section 4) this 80-20 split was repeated randomly 10 times (repeated holdout validation) to evaluate the models' uncertainty.

3 Methodology

Two different CNN architectures are explored in this study. The first is the "*Snapshot*" model, which takes a single AIA 171 Å image together with the corresponding PSP inputs (heliocentric distance, magnetic footpoint, etc.) and predicts the associated J_{linlin} intensity. This model provides an estimate of the SEP intensity expected at the PSP location based solely on the instantaneous state of the corona at the time of the image. The second architecture is the "*Video*" model, designed to incorporate temporal evolution. Instead of a single image, this model ingests a sequence of eight images spanning a full 24-hour interval (3-hour cadence). The target value time ($T+t_{offset}$) is defined with respect to the final image in the sequence (at time T), consistent with the snapshot model final image target. The motivation is that the temporal progression of coronal structures can contain information about the SEP that is not present in a single frame. By learning from full-day temporal patterns, the video model may capture the buildup and evolution of solar activity and, in principle, provide a more informed prediction than a single-image input.

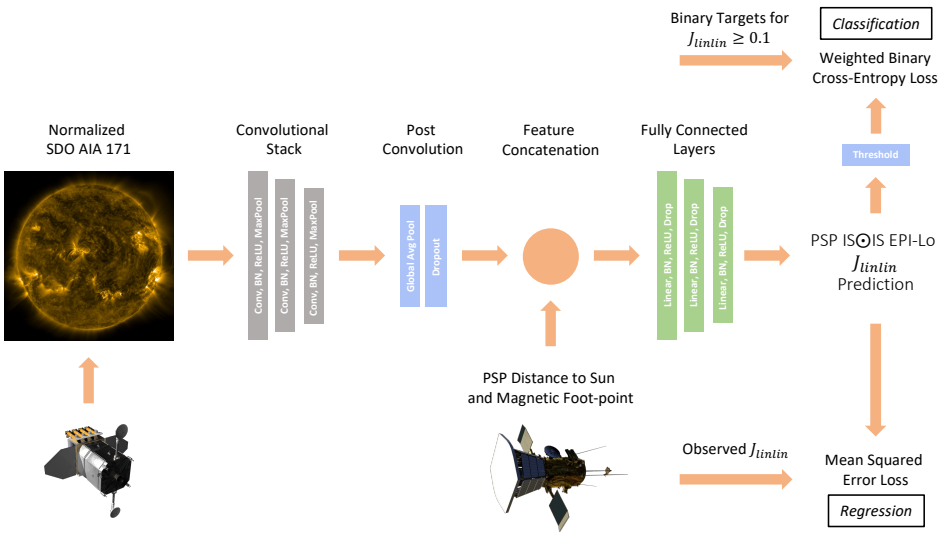

Figure 3. Schematic of the "*Snapshot*" model architecture. The model accepts as inputs normalized SDO/AIA 171 Å images, passes them through a series of convolutional layers (Convolutional Stack; grey), the image embeddings are then go through pooling and dropout layers (Post Convolution; blue), and then the PSP information is concatenated before getting passed through a series of dense layers (Fully Connected Layers; green). The resulting J_{linlin} prediction is compared with the observed value by PSP and the loss gradients are backpropagated through the network to train the model. Two types of predictions were used to train the "*Video*", regression and classification, yielding the two different versions in Table 2.

The "*Snapshot*" model begins with a convolutional encoder (*Convolutional Stack*, Figure 3) that learns spatial features in the AIA 171 Å images, such as large ARs, solar flares, coronal holes and other. The convolutional encoder consists of up to five convolution–normalization–activation–pooling blocks, each applying a 3×3 convolution with unit stride, batch normalization, a ReLU (Rectified Linear Unit; Hara et al., 2015) activation, and $2{\times}2$ max pooling. The number of filters increases with depth following $[16, 32, 64, 128, 256]$, allowing the network to capture increasingly abstract spatial features while reducing the resolution by a factor of two at each layer. This stage produces the image embedding, which is then passed through two *Post Convolution* layers: a Global Average Pooling layer and a Dropout layer which helps the models generalize and prevent them from overfitting.

To incorporate the spacecraft context, the PSP heliocentric distance and magnetic footpoint longitude are appended to this embedding (Figure 3, *Feature Concatenation*). Finally, the appended embedding is then processed through a dense layer (*Fully Connected Layers*), reducing this representation to a single output value. This output value corresponds to the predicted J_{linlin} intensity for the given image and PSP location, and therefore is compared, using Mean Squared Error (MSE) Loss to the observed $J_l inlin$. In every training iteration, this Loss is backpropagated through the CNN, changing the weights in the *Convolutional Stack* and the *Fully Connected Layers*. Although the main architecture of the ***Snapshot* model** remained the same, this study explores models with different sizes of convolutional and fully connected layers.

The "*Video*" model extends the snapshot architecture to operate on a sequence of eight AIA 171 Å images, allowing the network to use the temporal evolution of the solar corona rather than a single frame. As seen in Figure 4, a *ResNet-18 backbone* (He

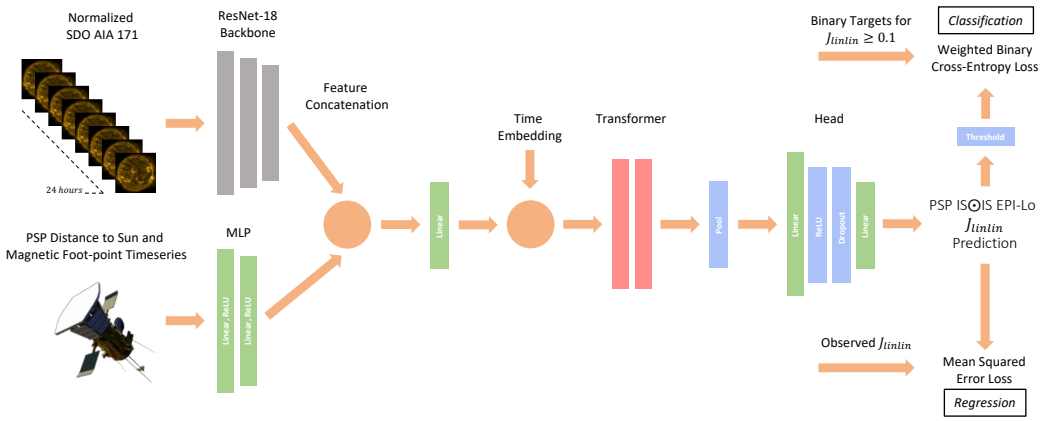

Figure 4. Schematic of the "*Video*" model architecture. The model accepts as inputs a) a stack of normalized SDO/AIA 171 Å images which passes through a series of pre-trained convolutional layers (ResNet-18 Backbone; grey) and b) the corresponding PSP J_{linlin} timeseries which it processes through a multilayer perceptron (MLP; green). The (a) and (b) embeddings are concatenated, passed through a linear layer (green) and time is embedded before the information passed through a series of transformer layers (Transformer; red), a pooling layer (blue) and a linear head. Two types of predictions were used to train the "*Video*", regression and classification, yielding the two different versions in Table 2.

et al., 2016) replaces the custom *Convolutional Stack* used in the "snapshot" model. This provides a more expressive feature extractor since the backbone begins with pretrained convolutional filters. In parallel, the PSP auxiliary inputs (heliocentric distance and magnetic footpoint longitude) are passed through a small two-layer *MLP* to produce a per-frame auxiliary embedding. The image and auxiliary embeddings for each timestamp are then concatenated (Figure 4, *Feature Concatenation*) and passed through a linear layer. To understand temporal structures, the model adds a learnable positional embedding (*Time Embedding*) and processes the resulting sequence using a *Transformer* encoder. The encoder applies multi-head self-attention across the timestamps, enabling the model to learn temporal relationships—such as sustained brightenings or the rise and decay of eruptive activity. The Transformer output is pooled and forwarded through a final MLP head that maps this embedding to a single scalar value corresponding to the predicted J_{linlin} intensity.

Most SEP prediction studies convert their output to a binary classification (SEP/no-SEP). To evaluate the models' performance in a similar setting, this study uses a J_{linlin} threshold of $10^{-1}\,\frac{\text{counts}}{\text{cm}^2\,\text{s}\,\text{sr}\,\text{keV}}$. If a J_{linlin} prediction is $\geq 10^{-1}\,\frac{\text{counts}}{\text{cm}^2\,\text{s}\,\text{sr}\,\text{keV}}$ then the model produces a positive SEP alarm, whereas if the opposite is true, the model predicts a quiet period. Although the "*Video*" **and "*Snapshot*"** models can be trained directly on the regression target J_{linlin}, applying the 0.1 threshold afterward does not necessarily yield optimal classification performance. Small errors near the threshold (e.g., a true value of 0.097 predicted as 0.107) produce correct regression behavior but incorrect class labels. To avoid this, targets can be one-hot encoded into a binary event label, where 1 indicates $J_{\text{linlin}} > 0.1$ and 0 otherwise. Training directly on these labels allows the model to focus on the classification task rather than the continuous intensity values.

A separate classification version of the "*Video*" **and "*Snapshot*"** models (top right of Figures 4 **and 3**, *Classification*) is therefore trained using this binary target in its loss function, instead of the intensity prediction itself. Because the dataset is imbalanced (approximately 87% non-events), the network would default toward predicting the domi-

nant class unless the loss function compensates for this. To address the imbalance, a weighted binary cross-entropy loss is applied, with the positive class weight set to the ratio of negative-to-positive samples **(5.47; calculated across the entire dataset, and therefore not recomputed for each k-fold)**. This is implemented using the *PyTorch* library's `BCEWithLogitsLoss`, which applies the sigmoid activation internally to the network's raw logit output $z \in (-\infty, \infty)$ and computes the weighted cross-entropy against the one-hot labels. This weighting encourages the model to learn both event and non-event predictions fairly despite the skewed class distribution. **Note that weighted loss is used only for the *Classification* models. Since the *Regression* models do not depend on the class labels and only on the target J_{linlin} values, standard MSE is used for them.**

4 Results

After developing the **four** models described above (***Snapshot Regression*, *Snapshot Classification***, *Video Regression*, and *Video Classification*), each model is evaluated to determine its predictive performance. All architectures were designed to allow configurable hyperparameters, so the first stage of evaluation (hyperparameter tunning) consisted of training multiple configurations to identify optimal settings. All models were tested across a range of hyperparameters including batch size, learning rate, number of convolutional blocks, number of dense layers, number of nodes per dense layer, and dropout rate. Each training run used the same training/test split, and hyperparameters were iteratively refined by sampling configurations near the best-performing settings from previous runs.

For the regression models, each input corresponds to a single scalar prediction, and performance is measured using the mean squared error (MSE). By applying a threshold of 0.1 J_{linlin} such that anything above is classified as an SEP event and everything else is not, we convert a regressive output to a binary SEP/no-SEP classification (using the $J_{\text{linlin}} \geq 10^{-1}$ threshold), following the convention used in existing SEP forecasting studies (Whitman et al., 2023). In the second stage of evaluation, the models are assessed using a comprehensive set of metrics, including the Heidke Skill Score (HSS), True Skill Score (Doswell et al., 1990, TSS), Probability of Detection (Wehling et al., 2011, (POD or Recall), F1 score (Lipton et al., 2014), False Alarm Ratio (Macmillan & Kaplan, 1985, FAR), overall accuracy (ACC), and Precision. Together with HSS and TSS, these metrics provide a broad assessment of model performance. For context, HSS and TSS both equal 0 for random predictions and 1 for perfect skill; POD, F1, ACC, and Precision also range from 0 (poor or random performance) to 1 (perfect), while FAR ranges from 0 (no false alarms) to 1 (all alarms false), so values well above the positive-class fraction (0.15 in this case) for POD and Precision and well below 1 for FAR indicate performance meaningfully better than chance. **The definitions of these metrics are prescribed in Appendix B.**

For the ***Snapshot Regression*** model, the hyperparameter tuning process (first stage; top right panel of Figure 5) converged to an optimal configuration consisting of 5 convolutional blocks, 2 dense layers, a batch size of 32, a dropout rate of 0.25, and a learning rate of 3×10^{-4}. Under this configuration, the model reduced the MSE by approximately 23% and performed particularly well on high-intensity J_{linlin} values. A scatterplot of the observed against the predicted J_{linlin} values for this model can be found on the top-right plot of Figure 5. After transforming the ***Snapshot Regression*** model's predictions to binary (second stage), it achieves ACC of 0.66 ± 0.10, Precision of 0.28 ± 0.05, POD of 0.75 ± 0.22, FAR of 0.72 ± 0.05, TSS of 0.39 ± 0.13, and HSS of 0.24 ± 0.07.

The hyperparameter tuning procedure (first stage) was performed using a single train–test split. To quantify the uncertainty associated with this choice, we subsequently trained and evaluated the ***Snapshot Regression*** model ten times using the same hy-

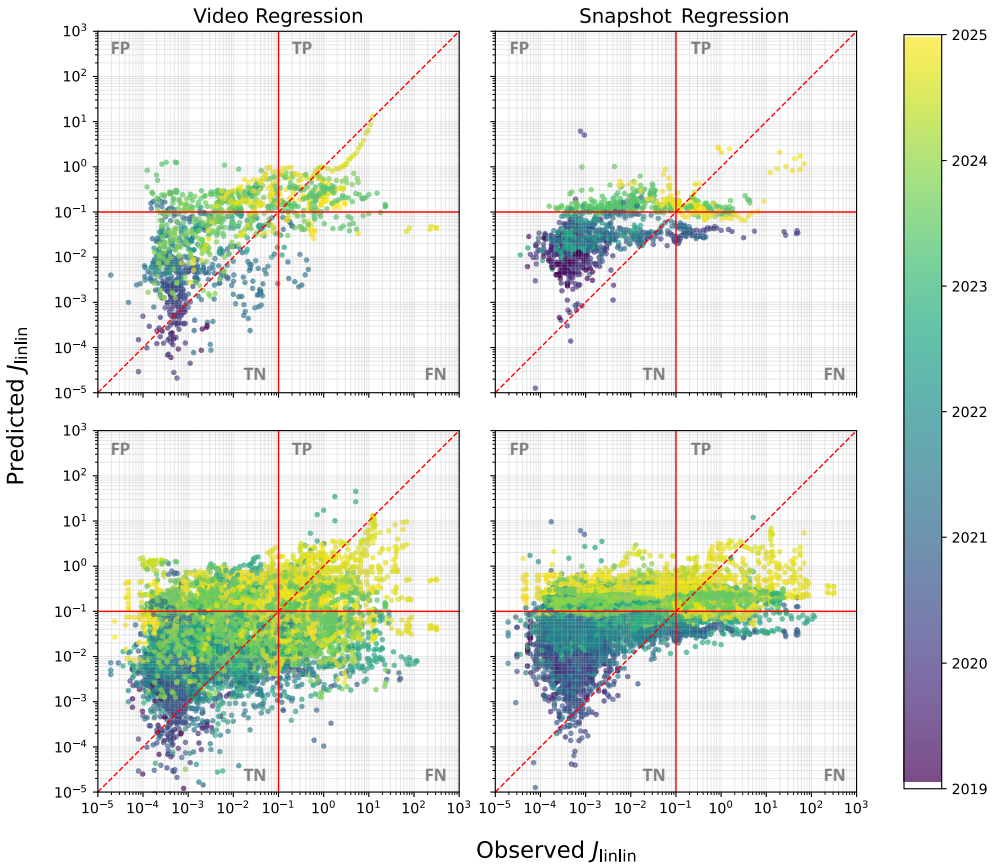

Figure 5. Scatterplots of the predicted J_{linlin} against the observed targets used during testing. The left column shows the predictions of the *Video Regression* model, while the right column predictions of the *Snapshot* model. The top row plots present the predictions of the models after hyperparameter tuning on a single train-test split. The bottom row plots present the predictions of the top row models but after training the same models, with the same parameters, but with a different train-test split. This process is repeated 10 times, therefore the bottom row plots contains 10 times more predictions. The points' colors indicate the date associated with the prediction, from the beginning of the PSP mission (2019; purple) to the most current orbits (2025; yellow). The red horizontal and vertical lines mark the threshold value of $0.1\,\frac{\text{counts}}{\text{cm}^2\,\text{s sr keV}}$ used to evaluate the models under a classification framework, while the diagonal $x = y$ line indicates a perfect prediction.

perparameters, but with different train–test splits and different random initialization seeds. This repeated random training/test splits, also known as repeated holdout validation, provides a statistical estimate of the model's robustness by revealing the variance in performance. Although the resulting metrics are often slightly worse than those obtained from the split used during hyperparameter optimization—a consequence of tuning the model to that specific partition—this repeated evaluation strategy serves as a form of cross-validation. This procedure often reduces the performance metrics but is essential as it ensures that the reported performance is not an artifact of a favorable split and that the model generalizes reliably across different realizations of the data. The results of all the 10 runs for the ***Snapshot Regression*** model are accumulated in the bottom-right plot of Figure 5.

The red horizontal and vertical lines in all scatterplots of Figure 5 mark the threshold value of 0.1 used to evaluate the models under a classification framework. These lines effectively partition each scatterplot into four quadrants, corresponding to the entries of a confusion matrix: the top-left quadrant represents False Positives (FP), the top-right represents True Positives (TP), the bottom-left corresponds to True Negatives (TN), and the bottom-right to False Negatives (FN). The diagonal dotted red line denotes the $x = y$ line, indicating perfect agreement between observed and predicted values. Finally, the points are color-coded by date according to the colormap shown to the right of each plot. Using such confusion matrices **(a summary of them is available in Figure B1 of Appendix B)**, the mean value and standard deviation of the POD, TSS, HSS, F1, FAR, ACC, Precision are calculated and presented in Table 2.

Table 2. Binary classification results for the **four** different models mentioned in the text **and the pixel intensity classification,** for the six most used by the community metrics. Each value represents the average of 10 runs, each one having a different train-test split, accompanied by the corresponding standard deviation. **The definitions of all classification matrices along with the confusion matrices for the four models and the pixel intensity classification are presented in Appendix B.**

	POD	TSS	HSS	F1	FAR	ACC	Precision
Video Class.	0.76 ± 0.23	0.47 ± 0.15	0.32 ± 0.11	0.47 ± 0.10	0.65 ± 0.10	0.73 ± 0.07	0.35 ± 0.10
Video Reg.	0.43 ± 0.20	0.29 ± 0.14	0.27 ± 0.08	0.39 ± 0.10	0.62 ± 0.05	0.80 ± 0.05	0.38 ± 0.05
Snapshot Class.	0.84 ± 0.11	0.44 ± 0.07	0.24 ± 0.05	0.42 ± 0.05	0.72 ± 0.04	0.64 ± 0.07	0.28 ± 0.04
Snapshot **Reg.**	0.75 ± 0.22	0.39 ± 0.13	0.24 ± 0.07	0.41 ± 0.08	0.72 ± 0.05	0.66 ± 0.10	0.28 ± 0.05
Pixel Intensity	0.56	0.34	0.27	0.404	0.683	0.747	0.317

Similarly, the *Video Regression* model (Table 2 and Figure 4, bottom right) includes several tunable hyperparameters. As with the ***Snapshot Regression*** model, its configuration involves selecting a batch size, learning rate, dropout rate, and dense-layer structure. However, due to its transformer-based architecture, additional parameters must also be tuned, including the number of attention blocks, number of attention heads, hidden-head size, and image-embedding dimension. Using the same strategy of randomly exploring the hyperparameter space and then refining the search toward regions with improved performance, the optimal configuration resulted in a batch size of 32, a dropout rate of 0.25, a learning rate of 10^{-4}, 4 attention heads, 3 attention blocks, a hidden-head size of 256, and an embedding size of 256. The performance of the *Video Regression* model under this configuration is shown in the top-left panel of Figure 5. As with the *Snapshot* model, classification predictions can be obtained by applying the $0.1\,J_{\mathrm{linlin}}$ threshold. Under this classification scheme, the *Video Regression* model achieves ACC of 0.80 ± 0.05, Precision of 0.38 ± 0.05, Recall of 0.43 ± 0.20, FAR of 0.62 ± 0.05, TSS of 0.29 ± 0.14, and HSS of 0.27 ± 0.08. The results from the ten repeated runs of the Video Regression model, along with their associated metrics, are presented in the bottom-left panel of Figure 5 and in the second row of Table 2.

Lastly, the *Video Classification* and ***Snapshot Classification*** models (Figure 4 **and Figure3**, top right) differs from the *Regression* models only in **their** output targets and loss function and therefore share the same set of tunable hyperparameters. After exploring the hyperparameter space using both the True Skill Score (TSS) and Heidke Skill Score (HSS) as selection criteria, the top-performing configuration **for the *Video Classification* model during the first k-fold** was obtained with a batch size of 32,

a dropout rate of 0.1, a learning rate of 5×10^{-4}, 4 attention heads, 3 attention blocks, a hidden-head size of 256, and an image-embedding size of 256. **By picking the hyperparameters only during the first k-fold, we prevent information leakage on the rest 9 k-folds along with keeping all the samples either for training or testing (no validation), to address our data scarcity problem.**

Based on the metrics in Table 2, the *Video Classification* model demonstrates the strongest overall classification performance. It achieves the highest mean TSS (0.47± 0.15), F1 score (0.47 ± 0.10) and HSS (0.32 ± 0.11). The *Video Regression* model has the second best precision (0.35±0.10) and FAR (0.65±0.10), and the *Snapshot* ***Classification*** model yields the highest POD (0.84±0.11) but with a relatively high FAR. FAR is especially important in operational SEP prediction, as a high rate of false alarms would interrupt space-based operations. Under this criterion, the *Video Classification* model's strong recall performance (0.76 ± 0.23), combined with its top TSS and HSS values**,** the lower FAR **and the second best Precision**, makes it the best-performing classifier among the **four** models.

Since this is the first machine-learning study to attempt this type of prediction task, no established baseline exists against which to directly compare the *Video Classification* model. To address this, we adopt two complementary comparison strategies. First, we benchmark the model against a simple classification based on SDO/AIA $171, \text{Å}$ **pixel intensities (Table 2), in which each image is spatially summed, assigned a corresponding energy-weighted particle intensity (J_{linlin}), and evaluated using a confusion matrix with thresholds similar to those for the rest models, as shown in Figure 6. The *Video Classification* scores better than the *Pixel Intensity* benchmark in all metrics except accuracy (which is known to be a bad metric for rare event classification such as the SEP prediction problem). Second, we compare our results with those of existing geoeffective SEP prediction models (see Table C1 in Appendix C); although these models target a different forecasting task, we find that the performance of our approach lies within the same overall range.**

5 Discussion

Both the *Snapshot* and *Video* models produce predicted J_{linlin} values with a clear improvement over a random guess**, the *Pixel Intensity* test and some geoeffective SEP event ML models**. This conclusion is further supported by the classification results, where TSS, HSS and F1 are well above zero (random guess), while POD and Precision are well above 0.15, the positive class fraction. ACC is an often misleading metric for imbalanced datasets, but is still reported following the convention of similar studies. **All models trained here have reported high FAR, a problem which is often observed with SEP prediction models. Interpreting which regression model performs "best" requires caution, as different evaluation metrics emphasize different aspects of model behavior and can favor distinct operational trade-offs. For example, *Snapshot Regression* prioritizes detection-oriented metrics such as POD and TSS, whereas *Video Regression* favors reliability-oriented metrics, including lower FAR and higher Precision, reflecting different balances between sensitivity and false-alarm control.**

Examination of the regression scatterplots (Figure 5 **and Figure 6**) reveals shared strengths and weaknesses across the two architectures: while the models predict with relative success the moderate and especially the high J_{linlin} values, they struggle in the low-intensity regime, where a wide spread of predicted values persists. This behavior, which occurs even when evaluating on the training set (Figure D1 in Appendix D), suggests that predicting precise particle intensities during quiescent solar conditions is inherently difficult, likely due to the limited distinguishing features present in EUV images when

solar activity is minimal. Importantly, this limitation is far less consequential for the classification task since most low-intensity intervals remain below the decision threshold, meaning that bad estimation of quiet-time intensities does not degrade event forecasting skill. From an operational perspective, accurate predictions during non-event periods are also less critical, as the primary value of an SEP forecasting system lies in reliably identifying hazardous high-intensity events rather than characterizing baseline particle levels.

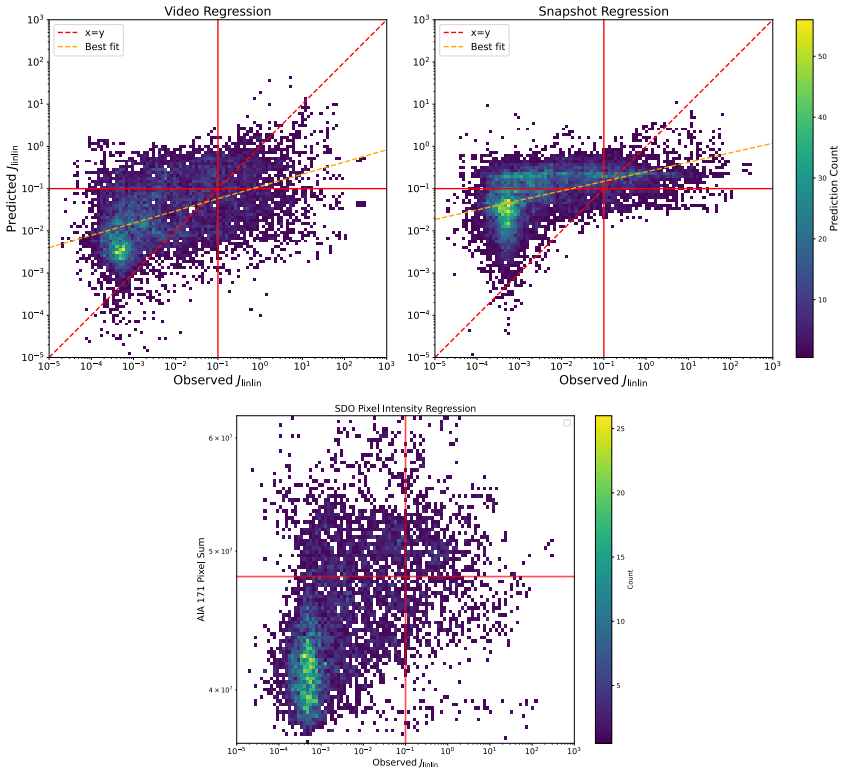

Figure 6. Two-dimensional histograms comparing observed and predicted solar energetic particle (SEP) flux J_{linlin} (particles $\geq 0.1\,\frac{\mathrm{counts}}{\mathrm{cm^2\,s\,sr\,keV}}$) for (top left) the *Video Regression* model and (top right) the *Snapshot Regression* model, aggregated across all random-seed ensemble members. Each panel uses a 100×100 logarithmically spaced bin grid and each bin is colored by the number of predictions falling within that region (Prediction Count colorbar on the right). The orange dashed line shows the best-fit power-law regression in log–log space. The red dashed diagonal line denotes perfect agreement. Vertical and horizontal red solid lines mark a flux threshold of $J = 10^{-1}$, delineating the four quadrants corresponding to true positives, false positives, true negatives, and false negatives relative to that threshold. The bottom histogram presents observed SEP flux J_{linlin} versus the summed AIA171 Å pixel intensity (the corresponding classification results are presented in the last row or Table 2 and the right-most confusion matrix of Figure B1 of Appendix B). Remember that similar to Figure 5, the *Regression* model results are inferior to those of the *Classification* models, but are being plotted since their *Classification* versions cannot.

Generally, the models tend to overfit on very high-intensity intervals, as a consequence of the pronounced class imbalance inherent in SEP datasets. Strong SEP events are comparatively rare ($\approx 15\%$ positive event intervals for this study), and the scarcity

of high-intensity samples makes it difficult for the models to generalize their behavior without overfitting to the few extreme cases available in the training set. Although the models still perform reasonably well on these intervals (Figure 5, TP quartiles), their performance could likely improve if overfitting at the upper end of the distribution were mitigated without weakening their ability to capture meaningful trends. This pattern was observed consistently throughout model development: architectures that performed poorly either severely overfit to the training set or underfit and failed to learn relevant structure. Class imbalance is a well-known challenge in SEP prediction and space-weather machine learning more broadly (M. A. Ali et al., 2025; Lavasa et al., 2021), and future work could explore loss reweighting, class-balancing strategies, or even synthetic data augmentation to counteract the disproportionate representation of non-event periods and improve model robustness at high intensities.

Whereas most existing models are designed to predict geoeffective SEP events at Earth, our framework extends forecasting capability to arbitrary heliospheric coordinates, enabling predictions at any location between the Sun and Earth. This fundamental difference in scope makes direct comparison with prior studies difficult, as earlier works employ Earth-centered observational inputs, distinct target variables, and often incompatible evaluation setups. It is also important to note that previous SEP prediction studies using ML methods typically rely on far longer observational windows, with most models trained on at least nine years of solar data (Sadykov et al., 2021), and some even training on more than four solar cycles—over 45 years of observations (Nedal et al., 2023). Note that in this analysis, only 44% of the 7 years ($\leq$ 3 years) worth of PSP EPI-Lo data are used as the rest of the PSP intervals fall outside AIA's effective field of view. Since ML models are inherently data-hungry, training on the comparatively small number of SEP events available during the first seven years of PSP operations presents a significant limitation of the present study. Future efforts should therefore seek to expand these datasets, for example by incorporating in-situ energetic particle measurements **and EUV images** from other Sun-orbiting missions such as Solar Orbiter (SolO; Müller et al., 2020; Marirrodriga et al., 2021) **and SOHO (Domingo et al., 1995), respectively**.

Nonetheless, when our predictions are reduced to a binary classification task, we can at least validate them using the same metrics reported for geoeffective SEP models (Chatterjee et al., 2024; Aminalragia-Giamini, Sigiava et al., 2021; M. A. Ali et al., 2025; Engell et al., 2017). In particular, the *Video Classification* model achieves a POD that is comparable to, and in some cases exceeds, that of existing ML-based geoeffective SEP forecasting efforts (A. Ali et al., 2024; Lavasa et al., 2021; Nedal et al., 2023; Hosseinzadeh et al., 2025)**, as presented in Appendix C**. However, the higher FAR produced by our models, along with the lower TSS and HSS relative to other studies, indicates that ML approaches trained solely on PSP targets are not yet reliable enough for operational use—especially in contexts where false alarms may trigger unnecessary protective actions. Still, it is encouraging that our model attains these performance levels despite being trained on relatively small dataset and under conservative, operationally realistic data-selection constraints, as described in Sections 2 and 3. Specifically, this study uses approximately 8,000 image–point data pairs for a total dataset size of $\sim$ 7GB, and training the model requires only 1–3 hours on NVIDIA A100 GPUs provided by the Princeton Stellar supercomputing cluster.

It should also be noted that by choosing a $\geq$ $0.1\frac{\text{counts}}{\text{cm}^2\,\text{s}\,\text{sr}\,\text{keV}}$ threshold, we detect SEP events that are of lower intensity (by about a decade) compared to the NOAA Space Weather Scale (S1 - S5) for solar radiation storms, which defines as the start of an S1 solar radiation storm the time when the flux of protons at energies $\geq$ 10 MeV equals or exceeds 10 pfu ($\frac{\text{counts}}{\text{cm}^2\,\text{s}\,\text{sr}}$). **Since the reported metrics are expected to be sensitive to the choice of threshold, future studies will explore the change in skill for a varied threshold, as well as additional input data from the PSP, the SDO, the SOHO and the Solar Orbiter**. The fact that ML models can learn from PSP data

in an SEP prediction setting is a promising result, highlighting the yet-untapped usefulness of PSP measurements for ML applications and suggesting that such data could complement existing datasets to produce more reliable forecasts, including geoeffective events.

Across both *Snapshot*- and *Video*-based architectures, reformulating the task as a classification problem leads to a pronounced improvement over regression-based training, indicating that the choice of objective function plays a central role in model performance. Although the *Video Classification* model achieves slightly better results than *Snapshot Classification* under otherwise identical conditions, the improvement is modest, and the extent to which temporal information alone provides additional predictive value remains inconclusive. Interpretation of these results is further limited by the present data volume and splitting strategy. A stricter temporal and event-level partitioning is needed, as the current, while necessary to maximize limited statistics, does not enforce strict temporal buffers between SEP events neither SEP event separation. Larger datasets will allow a more definitive assessment of the independent contribution of temporal context.

Part of the residual error—especially in the low-intensity regime—likely reflects limitations and mismatches in the observations themselves. The AIA and EPI-Lo instruments sense fundamentally different physical quantities with very different response functions and sampling geometries, meaning that weak particle enhancements near the EPI-Lo sensitivity limit will not map one-to-one to EUV features, and vice versa. Moreover, the 3 hour AIA cadence chosen here can attenuate (or not capture at all) short-lived brightenings, further complicating associations at low signal levels. Since, nearly 56% of PSP intervals fall outside AIA's effective field of view, a substantial fraction of PSP-detected events are invisible from SDO's vantage, and many AIA-visible transients will not yield in-situ SEPs at PSP's location. Finally, at low particle intensities the underlying counting statistics are intrinsically noisy, therefore it is expected that models will exhibit larger scatter during quiet solar conditions.

6 Conclusions

Most SEP prediction studies to date have focused on forecasting geoeffective events at or near Earth and therefore rely primarily on measurements from Earth-orbiting spacecraft or missions operating at the Sun–Earth L1 point. However, with ongoing efforts by NASA and the broader spaceflight community to support human and robotic exploration beyond near-Earth, forecasting SEP events at variable locations between Earth and Sun –such as cislunar space or the vicinity of Mars– has become increasingly important. Although a few recent studies have attempted to predict SEP intensities away from Earth, such as at Mars (Jolitz et al., 2021; Posner & Strauss, 2020), no machine-learning framework has yet been developed specifically to address this challenge.

This work aims to close this gap by training CNN-based models to predict SEP events throughout the heliosphere, leveraging the unique in-situ energetic particle measurements provided by the IS⊙IS instrument onboard PSP. More specifically, using as inputs SDO/AIA 171 Å images together with PSP location and magnetic footpoint information, and as targets the PSP/IS⊙IS EPI-Lo J_{linlin}, we train CNN-based models with different architectures to make predictions at any given point in the inner heliosphere. The best-performing model is the *Video Classification* architecture, which combines a ResNet-18 image backbone with an MLP-processed PSP timeseries and a transformer module operating on the fused representation to produce either regression or classification outputs. Through this approach, we show that the models successfully learn meaningful physical patterns linking AIA 171 Å coronal structure to energetic particle intensities throughout the inner heliosphere. More importantly, our results demonstrate that PSP data can be useful for

SEP prediction and should be incorporated into future ML-based SEP forecasting frameworks—both for inner heliosphere predictions and to complement the existing datasets for traditional geoeffective event prediction efforts.

Future work should incorporate additional in-situ **particle and EUV** measurements from Solar Orbiter and **SOHO, and** from later PSP orbits, as well as merge the current dataset with historical geoeffective SEP datasets to increase the number of high-intensity samples. **Increasing the size of the training and testing datasets is crucial, as it would enable evaluations across a range of heliographic distances—where current statistics are limited—and allow for the inclusion of a dedicated validation set to further mitigate information leakage. The performance of different AIA wavelengths can be explored and multi-spacecraft training should also be investigated to enable generalization across different heliolongitudes and latitudes.** The data cadence should be reduced from 3 hours to a shorter interval to increase the training sample size, as PSP's low cadence permits substantially denser sampling. Finally, the physical context of the modeling setup should be improved, for example by implementing a more accurate SEP speed calculation in the model.

Acknowledgments

We thank the IS⊙IS team and everyone that made the PSP mission possible. The IS⊙IS data and visualization tools are available to the community at `https://spacephysics.princeton.edu/missions-instruments/PSP`. PSP was designed, built, and is operated by the Johns Hopkins Applied Physics Laboratory as part of NASA's Living with a Star (LWS) program (contract NNN06AA01C).

Conflict of Interest

The authors declare no conflicts of interest relevant to this study.

Data Availability Statement

All figures were generated using Matplotlib version 3.2.1 (Hunter, 2007), available under the Matplotlib license at https://matplotlib.org/. The ML methods were implemented using the PyTorch library, which is publicly available at `https://pytorch.org/`. For the implementation of the above research, all relevant code can be found at `https://github.com/thutch17/PSP-SEP-Event-Prediction`. All of the computing and data storage for the model were done on Princeton University's Stellar cluster `https://researchcomputing.princeton.edu/systems/stellar`. The IS⊙IS data are publicly available and can be downloaded at `https://spp-isois.sr.unh.edu/`. The authors have made all necessary efforts to comply with the NASA Transform to Open Science (TOPS) directive. More information about the NOAA Space Weather Scale used in this work can be found in `https://www.swpc.noaa.gov/phenomena/solar-radiation-storm`.

References

Ali, A., Sadykov, V., Kosovichev, A., Kitiashvili, I. N., Oria, V., Nita, G. M., ... others (2024). Predicting solar proton events of solar cycles 22–24 using goes proton and soft-x-ray flux features. *The Astrophysical Journal Supplement Series*, *270*(1), 15.

Ali, M. A., Abdelkawy, A. G., Shaltout, A. M., & Beheary, M. (2025). Forecasting solar energetic particles using multi-source data from solar flares, cmes, and radio bursts with machine learning approaches. *Scientific Reports*, *15*(1), 9546.

Aminalragia-Giamini, Sigiava, Raptis, Savvas, Anastasiadis, Anastasios, Tsigkanos,

Antonis, Sandberg, Ingmar, Papaioannou, Athanasios, ... Daglis, Ioannis A. (2021). Solar energetic particle event occurrence prediction using solar flare soft x-ray measurements and machine learning. *J. Space Weather Space Clim.*, *11*, 59. Retrieved from `https://doi.org/10.1051/swsc/2021043` doi: 10.1051/swsc/2021043

Aminalragia-Giamini, S., Raptis, S., Anastasiadis, A., Tsigkanos, A., Sandberg, I., Papaioannou, A., ... Daglis, I. A. (2021). Solar energetic particle event occurrence prediction using solar flare soft x-ray measurements and machine learning. *Journal of Space Weather and Space Climate*, *11*, 59.

Boteler, D. (2001). Space weather effects on power systems. *Geophysical Monograph-American Geophysical Union*, *125*, 347–352.

Boubrahimi, S. F., Aydin, B., Martens, P., & Angryk, R. (2017). On the prediction of¿ 100 mev solar energetic particle events using goes satellite data. In *2017 ieee international conference on big data (big data)* (pp. 2533–2542).

Cane, H., Richardson, I., & Von Rosenvinge, T. (2010). A study of solar energetic particle events of 1997–2006: Their composition and associations. *Journal of Geophysical Research: Space Physics*, *115*(A8).

Chatterjee, S., Dayeh, M. A., Muñoz-Jaramillo, A., Bain, H. M., Moreland, K., & Hart, S. (2024). Mempsep-i. forecasting the probability of solar energetic particle event occurrence using a multivariate ensemble of convolutional neural networks. *Space Weather*, *22*(9), e2023SW003568.

Chierichini, S., Liu, J., Korsós, M. B., Del Moro, D., & Erdélyi, R. (2024). Cme arrival modeling with machine learning. *The Astrophysical Journal*, *963*(2), 121.

Cohen, C., Christian, E., Cummings, A., Davis, A., Desai, M., De Nolfo, G., ... others (2021). Psp/is is observations of the 29 november 2020 solar energetic particle event. *Astronomy & Astrophysics*, *656*, A29.

Cucinotta, F. A., & Pak, S. (2024). Cancer and circulatory disease risks for the largest solar particle events in the space age. *Life Sciences in Space Research*, *40*, 1–7.

Cuesta, M., Khoo, L., Livadiotis, G., Shen, M., Szalay, J., McComas, D., ... others (2025). Comparing methods for calculating solar energetic particle intensities: Rebinning versus spectral binning. *The Astrophysical Journal*, *980*(2), 235.

Cuesta, M. E., Fraschetti, F., Livadiotis, G., Farooki, H. A., Shen, M. M., Khoo, L. Y., ... Livi, R. (2025, November). Distinct Solar Energetic Particle Shock IntensityDiffusion Coefficient Relationships in the Inner Heliosphere. , *993*(1), L15. doi: 10.3847/2041-8213/ae109c

de Pablos, D., Samanta, T., Badman, S. T., Schwanitz, C., Bahauddin, S. M., Harra, L., ... others (2022). Searching for a solar source of magnetic-field switchbacks in parker solar probe's first encounter. *Solar Physics*, *297*(7), 90.

Desai, M., & Giacalone, J. (2016). Large gradual solar energetic particle events. *Living Reviews in Solar Physics*, *13*(1), 3.

Desai, M., Marsden, R., Sanderson, T., Balogh, A., Forsyth, R., & Gosling, J. (1998). Particle acceleration at corotating interaction regions in the three-dimensional heliosphere. *Journal of Geophysical Research: Space Physics*, *103*(A2), 2003–2014.

Domingo, V., Fleck, B., & Poland, A. (1995). Soho: the solar and heliospheric observatory. *Space Science Reviews*, *72*(1), 81–84.

Doswell, C., Davies-Jones, R., & Keller, D. L. (1990). On summary measures of skill in rare event forecasting based on contingency tables. *Weather and forecasting*, *5*(4), 576–585.

Engell, A., Falconer, D., Schuh, M., Loomis, J., & Bissett, D. (2017). Sprints: A framework for solar-driven event forecasting and research. *Space Weather*, *15*(10), 1321–1346.

Forbush, S. E. (1937). On the effects in cosmic-ray intensity observed during the re-

cent magnetic storm. *Physical Review*, *51*(12), 1108.

Forbush, S. E. (1946). Three unusual cosmic-ray increases possibly due to charged particles from the sun. *Physical Review*, *70*(9-10), 771.

Fox, N., Velli, M., Bale, S., Decker, R., Driesman, A., Howard, R., ... others (2016). The solar probe plus mission: humanity's first visit to our star. *Space Science Reviews*, *204*(1), 7–48.

Giacalone, J. (2005). Particle acceleration at shocks moving through an irregular magnetic field. *The Astrophysical Journal*, *624*(2), 765.

Gieseler, J., Dresing, N., & Klassen, A. (2019). Solar-mach: An interactive tool for comparing solar events with multiple spacecraft. *Solar Physics*, *294*, 32. doi: 10.1007/s11207-019-1421-5

Hara, K., Saito, D., & Shouno, H. (2015). Analysis of function of rectified linear unit used in deep learning. In *2015 international joint conference on neural networks (ijcnn)* (pp. 1–8).

Harten, R., & Clark, K. (1995). The design features of the ggs wind and polar spacecraft. *Space Science Reviews*, *71*(1), 23–40.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. , 770–778. doi: 10.1109/CVPR.2016.90

Hess, V. F., & Demmelmair, A. (1937). World-wide effect in cosmic ray intensity, as observed during a recent magnetic storm. *Nature*, *140*(3538), 316–317.

Hill, M. E., Mitchell, D. G., Andrews, G. B., Cooper, S. A., Gurnee, R. S., Hayes, J. R., ... Westlake, J. H. (2017, February). The Mushroom: A half-sky energetic ion and electron detector. *Journal of Geophysical Research (Space Physics)*, *122*(2), 1513-1530. doi: 10.1002/2016JA022614

Hosseinzadeh, P., Boubrahimi, S. F., & Hamdi, S. M. (2025). An end-to-end ensemble machine learning approach for predicting high-impact solar energetic particle events using multimodal data. *The Astrophysical Journal Supplement Series*, *277*(2), 34.

Howe, R. (2009). Solar interior rotation and its variation. *Living Reviews in Solar Physics*, *6*(1), 1–75.

Hristopulos, D., Khoo, L., Cuesta, M., Kasapis, S., Pak, S., Livadiotis, G., ... Cohen, C. (2026). Statistical anatomy of a solar energetic particle event. *The Astrophysical Journal*. (in press) doi: 10.3847/1538-4357/ae5ba1

Hunter, J. D. (2007). Matplotlib: A 2d graphics environment. *Computing in Science & Engineering*, *9*(3), 90–95. doi: 10.1109/MCSE.2007.55

Jiao, Z., Sun, H., Wang, X., Manchester, W., Gombosi, T., Hero, A., & Chen, Y. (2020). Solar flare intensity prediction with machine learning models. *Space weather*, *18*(7), e2020SW002440.

Jolitz, R., Dong, C., Rahmati, A., Brain, D., Lee, C., Lillis, R., ... Jakosky, B. (2021). Test particle model predictions of sep electron transport and precipitation at mars. *Journal of Geophysical Research: Space Physics*, *126*(8), e2021JA029132.

Kasapis, S., Cuesta, M. E., Khoo, L. Y., Farooki, H. A., Pak, S., Szalay, J. R., ... others (2025). Reconstructing psp/isois pitch-angle-resolved intensities using unsupervised learning. *The Astrophysical Journal*, *995*(1), 72.

Kasapis, S., Kitiashvili, I. N., Kosovich, P., Kosovichev, A. G., Sadykov, V. M., O'Keefe, P., & Wang, V. (2024). Forecasting solar energetic particle events during solar cycles 23 and 24 using interpretable machine learning. *The Astrophysical Journal*, *974*(1), 131.

Kasapis, S., Kitiashvili, I. N., Kosovichev, A. G., & Stefan, J. T. (2025). Prediction of intensity variations associated with emerging active regions using helioseismic power maps and machine learning. *The Astrophysical Journal Supplement Series*, *280*(2), 64.

Kasapis, S., Kitiashvili, I. N., Kosovichev, A. G., Stefan, J. T., & Apte, B. (2023).

Predicting the emergence of solar active regions using machine learning. *Proceedings of the International Astronomical Union*, *19*(S365), 311–319.

Kasapis, S., Zhao, L., Chen, Y., Wang, X., Bobra, M., & Gombosi, T. (2022). Interpretable machine learning to forecast sep events for solar cycle 23. *Space Weather*, *20*(2), e2021SW002842.

Kasper, J. C., Abiad, R., Austin, G., Balat-Pichelin, M., Bale, S. D., Belcher, J. W., ... others (2016). Solar wind electrons alphas and protons (sweap) investigation: Design of the solar wind and coronal plasma instrument suite for solar probe plus. *Space Science Reviews*, *204*(1), 131–186.

Keegan, K., Bonaventura, N., Guzmán, N., Karna, N., Hess-Webber, S., Kasapis, S., ... Muñoz-Jaramillo, A. (2025). Data-driven solar surface flux transport modeling with uncertainty quantification. In *The reach and limits of ai for scientific discovery workshop, neurips.*

Klein, K.-L., & Dalla, S. (2017). Acceleration and propagation of solar energetic particles. *Space Science Reviews*, *212*(3), 1107–1136.

Kosovichev, A. G., Basu, S., Bekki, Y., Buitrago-Casas, J. C., Chatzistergos, T., Chen, R., ... others (2025). Structure and dynamics of the sun's interior revealed by the helioseismic and magnetic imager. *Solar Physics*, *300*(5), 70.

Kuehnast, T., Abbott, C., Pausan, M. R., Pearce, D. A., Moissl-Eichinger, C., & Mahnert, A. (2022). The crewed journey to mars and its implications for the human microbiome. *Microbiome*, *10*(1), 26.

Laitinen, T., & Dalla, S. (2019). From sun to interplanetary space: what is the path-length of solar energetic particles? *The Astrophysical Journal*, *887*(2), 222.

Laurenza, M., Alberti, T., & Cliver, E. (2018). A short-term esperta-based forecast tool for moderate-to-extreme solar proton events. *The Astrophysical Journal*, *857*(2), 107.

Lavasa, E., Giannopoulos, G., Papaioannou, A., Anastasiadis, A., Daglis, I. A., Aran, A., ... Sanahuja, B. (2021). Assessing the predictability of solar energetic particles with the use of machine learning techniques. *Solar Physics*, *296*(7), 107. doi: 10.1007/s11207-021-01837-x

Lemen, J. R., Title, A. M., Akin, D. J., Boerner, P. F., Chou, C., Drake, J. F., ... others (2012). The atmospheric imaging assembly (aia) on the solar dynamics observatory (sdo). *Solar Physics*, *275*(1), 17–40.

Lipton, Z. C., Elkan, C., & Narayanaswamy, B. (2014). Thresholding classifiers to maximize f1 score. *stat*, *1050*, 14.

Livi, R., Larson, D. E., Kasper, J. C., Abiad, R., Case, A. W., Klein, K. G., ... others (2022). The solar probe analyzer—ions on the parker solar probe. *The Astrophysical Journal*, *938*(2), 138.

Macmillan, N. A., & Kaplan, H. L. (1985). Detection theory analysis of group data: estimating sensitivity from average hit and false-alarm rates. *Psychological bulletin*, *98*(1), 185.

Majid, H. A., Sittoni, P., & Tudisco, F. (2024). Solaris: a foundation model for the sun. In *Workshop: Foundation models for science: Progress, opportunities and challengesat neurips 2024.*

Marirrodriga, C. G., Pacros, A., Strandmoe, S., Arcioni, M., Arts, A., Ashcroft, C., ... others (2021). Solar orbiter: Mission and spacecraft design. *Astronomy & Astrophysics*, *646*, A121.

McComas, D., Alexander, N., Angold, N., Bale, S., Beebe, C., Birdwell, B., ... others (2016). Integrated science investigation of the sun (isis): Design of the energetic particle investigation. *Space Science Reviews*, *204*, 187–256.

McComas, D., Christian, E., Cohen, C., Cummings, A., Davis, A., Desai, M., ... others (2019). Probing the energetic particle environment near the sun. *Nature*, *576*(7786), 223–227.

Menzel, W. P., & Purdom, J. F. (1994). Introducing goes-i: The first of a new

generation of geostationary operational environmental satellites. *Bulletin of the American Meteorological Society*, *75*(5), 757–782.

Müller, D., Cyr, O. S., Zouganelis, I., Gilbert, H. R., Marsden, R., Nieves-Chinchilla, T., ... others (2020). The solar orbiter mission-science overview. *Astronomy & Astrophysics*, *642*, A1.

Müller, D., St. Cyr, O. C., Zouganelis, I., Gilbert, H. R., Marsden, R., Nieves-Chinchilla, T., ... Williams, D. (2020, October). The Solar Orbiter mission: Science overview. *Astronomy & Astrophysics*, *642*, A1. Retrieved 2025-10-01, from `https://www.aanda.org/10.1051/0004-6361/202038467` (867 citations (NASA ADS/DOI) [2025-10-01]) doi: 10.1051/0004-6361/202038467

Nedal, M., Kozarev, K., Arsenov, N., & Zhang, P. (2023). Forecasting solar energetic proton integral fluxes with bi-directional long short-term memory neural networks. *Journal of Space Weather and Space Climate*, *13*, 26.

Oughton, E. J., Skelton, A., Horne, R. B., Thomson, A. W., & Gaunt, C. T. (2017). Quantifying the daily economic impact of extreme space weather due to failure in electricity transmission infrastructure. *Space weather*, *15*(1), 65–83.

Pak, S., & Cucinotta, F. A. (2024). Organ dose equivalents of albedo protons and neutrons under exposure to large solar particle events during lunar human landing missions. *Life Sciences in Space Research*, *42*, 133–139.

Pak, S., Cuesta, M., Farooki, H., Khoo, L., Xu, Z., Davis, A., ... others (2025). Species-dependent variability in the energy spectra of intense solar energetic particle events observed by psp/isis/epi-hi/let. *The Astrophysical Journal Supplement Series*, *281*(1), 21.

Papaioannou, A., Strauss, R. D. T., Lario, D., Vainio, R., Wijsen, N., Afanasiev, A., ... Kouloumvakos, A. (2025). Predicting solar energetic particles: Solar storm watch-preparing for space odyssey. *Space Science Reviews*, *221*(6), 82.

Parker, E. N. (1958). Dynamics of the interplanetary gas and magnetic fields. *Astrophysical Journal, vol. 128, p. 664*, *128*, 664.

Pesnell, W. D., Thompson, B. J., & Chamberlin, P. (2012). The solar dynamics observatory (sdo). In *The solar dynamics observatory* (pp. 3–15). Springer.

Posner, A., & Strauss, R. (2020). Warning time analysis from sep simulations of a two-tier release system applied to mars exploration. *Space Weather*, *18*(4), e2019SW002354.

Pricopi, A.-C., Paraschiv, A. R., Besliu-Ionescu, D., & Marginean, A.-N. (2022). Predicting the geoeffectiveness of cmes using machine learning. *The Astrophysical Journal*, *934*(2), 176.

Rankin, J. S., McComas, D. J., Leske, R. A., Christian, E. R., Cohen, C., Cummings, A. C., ... others (2022). Anomalous cosmic-ray oxygen observations into 0.1 au. *The Astrophysical Journal*, *925*(1), 9.

Raouafi, N. E., Matteini, L., Squire, J., Badman, S., Velli, M., Klein, K., ... others (2023). Parker solar probe: Four years of discoveries at solar cycle minimum. *Space Science Reviews*, *219*(1), 8.

Reames, D. V. (2013). The two sources of solar energetic particles. *Space Science Reviews*, *175*(1), 53–92.

Roy, S., Hegde, D. V., Schmude, J., Lin, A., Gaur, V., Lal, R., ... others (2025). Suryabench: Benchmark dataset for advancing machine learning in heliophysics and space weather prediction. *arXiv preprint arXiv:2508.14107*.

Roy, S., Schmude, J., Lal, R., Gaur, V., Freitag, M., Kuehnert, J., ... others (2025). Surya: Foundation model for heliophysics. *arXiv preprint arXiv:2508.14112*.

Sadykov, V., Kosovichev, A., Kitiashvili, I., Oria, V., Nita, G. M., Illarionov, E., ... Ali, A. (2021). Prediction of solar proton events with machine learning: Comparison with operational forecasts and" all-clear" perspectives. *arXiv preprint arXiv:2107.03911*.

Sarlis, N., Livadiotis, G., McComas, D., Cuesta, M., Khoo, L., Shen, M., ...

Schwadron, N. (2025). Persistent behavior of low-energy solar energetic particles observed by parker solar probe during orbit 15. *The Astrophysical Journal*, *992*(1), 95.

Smith, E. J. (2001). The heliospheric current sheet. *Journal of Geophysical Research: Space Physics*, *106*(A8), 15819–15831.

Stone, E. C., Frandsen, A., Mewaldt, R., Christian, E., Margolies, D., Ormes, J., & Snow, F. (1998). The advanced composition explorer. *Space Science Reviews*, *86*(1), 1–22.

Torres, J., Chan, P. K., Zhao, L., & Zhang, M. (2025). A machine learning approach to predicting sep proton intensity and events using time series of relativistic electron measurements. *Space Weather*, *23*(2), e2024SW003921.

Torres, J., Zhao, L., Chan, P. K., & Zhang, M. (2022). A machine learning approach to predicting sep events using properties of coronal mass ejections. *Space Weather*, *20*(7), e2021SW002797.

Walsh, J., Gass, D. G., Pollan, R. R., Wright, P. J., Galvez, R., Kasmanoff, N., ... Baydin, A. G. (2024). A foundation model for the solar dynamics observatory. *arXiv preprint arXiv:2410.02530*.

Wang, Y., & Qin, G. (2023). The crucial role of perpendicular diffusion in the longitude distribution of¿ 10 mev solar energetic protons. *The Astrophysical Journal*, *954*(1), 81.

Wehling, P., LaBudde, R. A., Brunelle, S. L., & Nelson, M. T. (2011). Probability of detection (pod) as a statistical model for the validation of qualitative methods. *Journal of AOAC International*, *94*(1), 335–347.

Whitman, K., Egeland, R., Richardson, I. G., Allison, C., Quinn, P., Barzilla, J., ... others (2023). Review of solar energetic particle prediction models. *Advances in Space Research*, *72*(12), 5161–5242.

Wiedenbeck, M. E., Angold, N. G., Birdwell, B., Burnham, J. A., Christian, E. R., Cohen, C. M. S., ... White, M. L. (2017, October). Capabilities and Performance of the High-Energy Energetic-Particles Instrument for the Parker Solar Probe Mission. In *Proceedings of 35th International Cosmic Ray Conference — PoS(ICRC2017)* (p. 016). Bexco, Busan, Korea: Sissa Medialab. Retrieved 2023-09-06, from `https://pos.sissa.it/301/016` (16 citations (NASA ADS/DOI) [2025-09-08]) doi: 10.22323/1.301.0016

Wiedenbeck, M. E., Burnham, J. A., Cohen, C. M. S., Cook, W. R., Crabill, R. M., Cummings, A. C., ... Schwadron, N. A. (2021, June). Thin silicon solid-state detectors for energetic particle measurements. Development, characterization, and application on NASA's Parker Solar Probe mission. , *650*, A27. doi: 10.1051/0004-6361/202039754

Xu, Z., Cohen, C., Leske, R., Muro, G., Cummings, A., Romeo, O., ... others (2026). Parker solar probe observations of solar energetic particle (sep) events with inverse velocity arrival (iva) features. *arXiv preprint arXiv:2602.12475*.

Yu, Y., Chen, Y., Zhao, L., Whitman, K., Manchester, W., & Gombosi, T. (2025). Solar energetic particle forecasting with multi-task deep learning: Sepnet. *arXiv preprint arXiv:2512.12786*.

Zhou, W., Zhou, W., Yang, Y., Cao, J., Ge, Z.-Y., Xu, T., ... others (2026). Coronal mass ejection transit time prediction based on spatiotemporal modeling using 3d convolutional neural networks. *The Astrophysical Journal*, *999*(1), 23.

Appendix A Parker Spiral Path-length Calculation

Given the solar rotation rate Ω and the solar wind velocity v_r, the path-length of a Parker spiral at a given distance r from the Sun is:

$$S_r = \int_0^r \sqrt{1 + \left(\frac{\Omega}{v_r} r\right)^2}\, dr \tag{A1}$$

Assuming a constant solar rotation of $\Omega = 2.8 \times 10^{-6}$ rad/s and a constant solar wind speed of $v_r = 400$ km/s, the analytic solution used in this work for the Parker spiral path-length S_r is:

$$S_r = 0.5\sqrt{4.9 \cdot 10^{-17} r^2 + 1}\, r + 7.14 \cdot 10^7 \operatorname{arcsinh}(7 \cdot 10^{-9} r) \tag{A2}$$

Appendix B Confusion Matrices and Metrics Definitions

In this appendix, we provide the formal definitions of the classification performance metrics reported in Table 2. All metrics are expressed in terms of the elements of the binary confusion matrix—true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN)—which are also shown for each model.

Probability of Detection (POD)

$$\text{POD} = \frac{TP}{TP + FN} \tag{B1}$$

True Skill Statistic (TSS)

$$\text{TSS} = \frac{TP}{TP + FN} - \frac{FP}{FP + TN} \tag{B2}$$

Heidke Skill Score (HSS)

$$\text{HSS} = \frac{2\,(TP \cdot TN - FP \cdot FN)}{(TP + FN)(FN + TN) + (TP + FP)(FP + TN)} \tag{B3}$$

F1 Score (F1)

$$\text{F1} = \frac{2\,TP}{2\,TP + FP + FN} \tag{B4}$$

False Alarm Rate (FAR)

$$\text{FAR} = \frac{FP}{TP + FP} \tag{B5}$$

Accuracy (ACC)

$$\text{ACC} = \frac{TP + TN}{TP + TN + FP + FN} \tag{B6}$$

Precision (Precision)

$$\text{Precision} = \frac{TP}{TP + FP} \tag{B7}$$

The cumulative TP, FP, TN, FN values for the ten k-fold runs of each model are presented in Figure B1.

Appendix C Comparison with other Studies

The Video Classification model presented in this work aims in predicting SEP events in the area of the heliosphere that spans from the solar corona up to 0.8 AU. It is the only ML model in literature that does not predict geoeffective SEP events (i.e. events

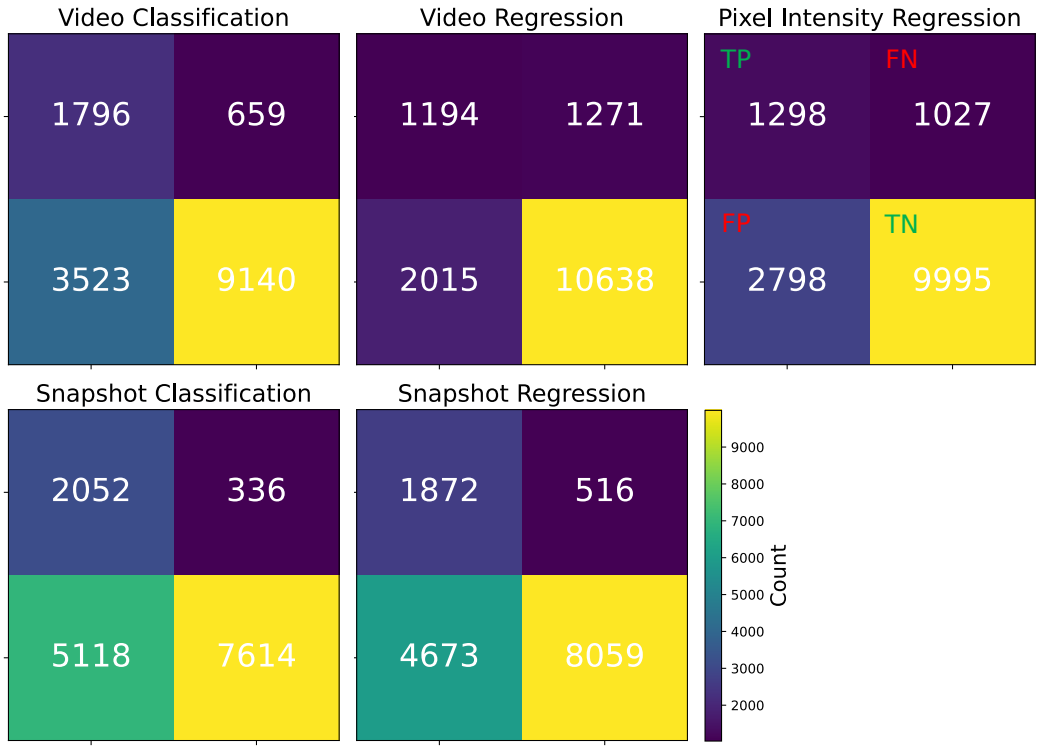

Figure B1. Confusion matrices showing the cumulative numbers of TP, TN, FP, and FN summed over all ten k-fold runs for the four machine-learning models in Table 2. For the pixel intensity regression, the total number of hits and misses has been scaled to match those of the machine-learning models to enable a direct visual comparison.

close to Earth). Despite the difference between prediction task, ML model architectures and data the SEP prediction models have been trained in, Table C1 presents a comparison between this work and others in literature.

More specifically, other than the Video Classification model's results, Table C1 presents the results for the SMARP-SHARP Kasapis et al. (2022), Kasapis et al. (2024), CART Boubrahimi et al. (2017), SEP-C Torres et al. (2022) , SEP-E Torres et al. (2025), UNSPELL Aminalragia-Giamini et al. (2021), SEPNET Yu et al. (2025) and BiLSTM-SEP Nedal et al. (2023) models. We note that not all comparison studies report the full set of evaluation metrics considered here; consequently, some entries in Table C1 are unavailable and appear as gaps. This limitation reflects differences in reporting practices among prior works.

Overall, the Video Classification model demonstrates competitive performance relative to several SEP forecasting approaches, achieving a POD comparable to CART, SEP-E, BiLSTM-SEP and SEPNET, and a TSS exceeding that of SMARP-SHARP and SEPNET. While SEP-C reports higher detection rates, its results are accompanied by substantially higher FAR. The Video Classification model maintains a balanced trade-off between detection capability and FAR (slightly higher than most studies), yielding F1 and HSS values that are comparable to those reported by SMARP-SHARP, SEPNET and SEP-E.

Appendix D Training Scatterplot

Table C1. Comparison of the Video Classification model with previously published SEP prediction methods, including SMARP–SHARP, CART, SEP–C, SEP–E, UNSPELL, SEPNET, and BiLSTM–SEP. Reported values correspond to the published metrics in each study; entries without quoted uncertainties do not include ± values because standard deviations were not reported by the corresponding authors.

Model	POD	TSS	HSS	F1	FAR
Video Class.	0.76 ± 0.23	0.47 ± 0.15	0.32 ± 0.11	0.47 ± 0.10	0.65 ± 0.10
SMARP-SHARP		0.39 ± 0.19	0.37 ± 0.19	0.67 ± 0.11	0.30 ± 0.14
CART	0.73			0.82	
SEP-C	0.92	0.91 ± 0.04	0.25 ± 0.06	0.25 ± 0.06	0.88
SEP-E	0.70			0.76	
UNSPELL	0.86	0.78			0.08
SEPNET	0.64	0.43	0.42	0.71	0.23
BiLSTM-SEP	0.62	0.53	0.73		0.09

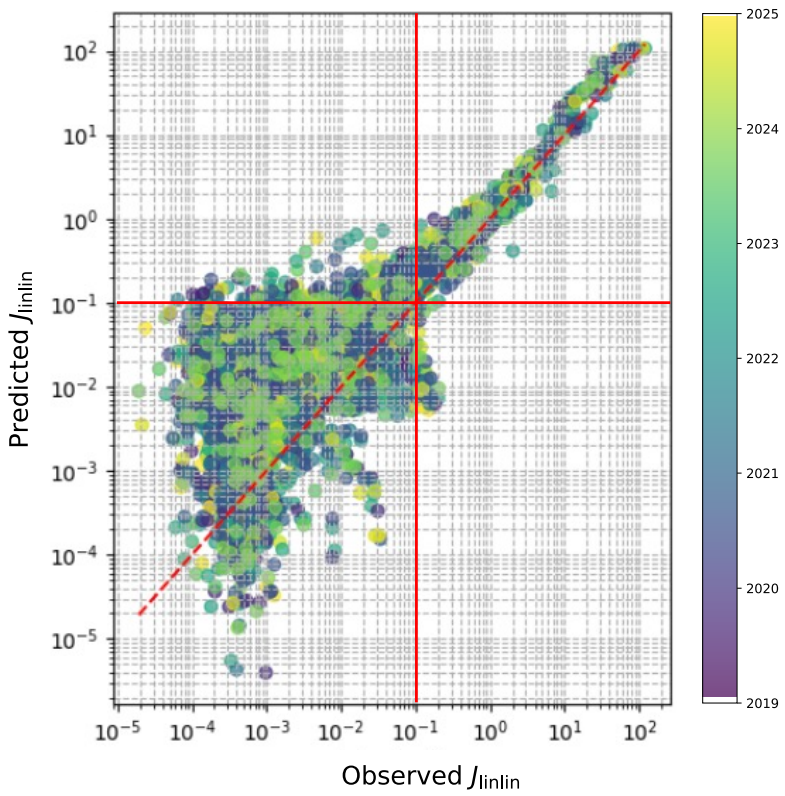

Figure D1. Scatterplots of the predicted J_{linlin} against the observed targets for the 80% of training samples (not the testing, unseen samples). Even for the AIA EUV 171 Å seen during training, the *Video Regression* model cannot produce reliable predictions for $J_{\text{linlin}} \leq 10^{-1} \frac{\text{counts}}{\text{cm}^2\,\text{s}\,\text{sr}\,\text{keV}}$. This serves as evidence that during times of quiet solar activity, ML models have difficulties identifying space weather event features in AIA EUV 171 Å images that are relevant to the particle intensity levels of the heliosphere.