JSWSC

Available online at:
www.swsc-journal.org

RESEARCH ARTICLE

OPEN ACCESS

Solar Energetic Particle Event occurrence prediction using Solar Flare Soft X-ray measurements and Machine Learning

Sigiava Aminalragia-Giamini[1,2,*] (ID), Savvas Raptis[3] (ID), Anastasios Anastasiadis[4] (ID), Antonis Tsigkanos[1] (ID), Ingmar Sandberg[1] (ID), Athanasios Papaioannou[4] (ID), Constantinos Papadimitriou[1,2] (ID), Piers Jiggens[5] (ID), Angels Aran[6] (ID), and Ioannis A. Daglis[2,7] (ID)

[1] Space Applications and Research Consultancy (SPARC), 10551, Athens, Greece
[2] Department of Physics, National and Kapodistrian University of Athens (NKUA), 15772, Athens, Greece
[3] Division of Space and Plasma Physics, School of Electrical Engineering and Computer Science, KTH Royal Institute of Technology, 10044, Stockholm, Sweden
[4] Institute for Astronomy, Astrophysics, Space Applications and Remote Sensing (IAASARS), National Observatory of Athens, 15236, Athens, Greece
[5] European Research and Technology Centre, European Space Agency (ESTEC-ESA), 2200AZ, Noordwijk, The Netherlands
[6] Dept. de Física Quàntica i Astrofísica, Institut de Ciències del Cosmos (ICCUB), University de Barcelona (UB-IEEC), 08028, Barcelona, Spain
[7] Hellenic Space Center, 15231, Athens, Greece

Abstract – The prediction of the occurrence of Solar Energetic Particle (SEP) events has been investigated over many years, and multiple works have presented significant advances in this problem. The accurate and timely prediction of SEPs is of interest to the scientific community as well as mission designers, operators, and industrial partners due to the threat SEPs pose to satellites, spacecrafts, and crewed missions. In this work, we present a methodology for the prediction of SEPs from the soft X-rays of solar flares associated with SEPs that were measured in 1 AU. We use an expansive dataset covering 25 years of solar activity, 1988–2013, which includes thousands of flares and more than two hundred identified and catalogued SEPs. Neural networks are employed as the predictors in the model, providing probabilities for the occurrence or not of a SEP, which are converted to yes/no predictions. The neural networks are designed using current and state-of-the-art tools integrating recent advances in the machine learning field. The results of the methodology are extensively evaluated and validated using all the available data, and it is shown that we achieve very good levels of accuracy with correct SEP occurrence prediction higher than 85% and correct no-SEP predictions higher than 92%. Finally, we discuss further work towards potential improvements and the applicability of our model in real-life conditions.

Keywords: Solar Energetic Particle Event / Solar Flare / Prediction / Machine Learning

1 Introduction

Solar Energetic Particle (SEP) events are major space weather agents, they consist of large populations of electrons, protons, and heavy-ion nuclei accelerated even up to very high energies well into the relativistic regime. These energetic particles can cause charging in spacecrafts, deterioration of materials, corruption, or loss of data, and importantly, they are dangerous for human health, see e.g., Hu et al. (2009). Solar flares (SFs) and Coronal Mass Ejections (CMEs) are both accepted to be

progenitors of SEPs (Belov et al., 2005; Klein & Dalla, 2017) and while SFs, and particularly the more intense ones, have a connection with CMEs (Youssef, 2012; Dierckxsens et al., 2015) their relation and separate contributions to SEP events is still unclear. Therefore, in spite of the importance of SEP events having been recognized for many years, their long-term accurate prediction, beyond a horizon of a few minutes to days, has been so far unattainable. This may continue to be the case as SEP events may be the result of Self-Organized Critical (SOC) processes (Aschwanden et al., 2016), which are inherently unpredictable, and/or be SOCs themselves (Xapsos et al., 2006; Jiggens & Gabriel, 2009). In this regard, over the past decades, several approaches have been investigated, e.g., Posner (2007),

*Corresponding author: sagiamini@sparc.gr; sagiamini@gmail.com; sagiamini@phys.uoa.gr

Balch (2008), Winter & Ledbetter (2015), and several systems developed for the forecasting of SEPs (Anastasiadis et al., 2019) such as the FORSPEF tool (Anastasiadis et al., 2017), the Proton Prediction System (PPS) (Kahler et al., 2007), the ESPERTA model (Laurenza et al., 2009; Laurenza et al., 2018; Alberti et al., 2017), the UMASEP system (Núñez, 2011; Núñez & Daniel, 2020) as well as the proton event component of the National Oceanic and Atmospheric Administration Space Weather Prediction Center (NOAA SWPC) (Bain et al., 2021). All such models and systems aim to strike a balance in the maximization of the advance warning time, the optimization of their correct predictions, and the minimization of false positives. More recently, Machine Learning (ML) methods are finding increased ground in space physics and space applications due to the advantages they offer, see for example, Balasis et al. (2019) and Azari et al. (2020), Nita et al. (2020) for useful discussions. ML methods have been shown over many years to outperform simpler approaches in multiple scientific and applied areas, and the use of ML in space weather has been discussed in depth in Camporeale (2019). ML is particularly useful when dealing with high dimensionality data, where multiple and sometimes unknown variables need to be taken into account, and/or unknown processes influence the output for a given input. ML methods can map such input-output relationships without the need for a pre-existing explicit description of the influence and inter-relationships of all factors that come into play. The prediction of SEPs is exactly such a problem where the physical processes of injection, acceleration, and propagation of solar particles in these eruptive events are complex and not fully characterized processes. In this work, we employ neural networks (NNs) for the prediction of the occurrence of SEPs using measurements of solar X-rays. X-ray measurements are available from NOAA GOES satellites (GOES I-M DataBook, 1996) over several decades, offering a very expansive dataset. Additionally, the primary operational purpose of GOES spacecrafts is to support forecasting operations, thereby providing real-time access to X-ray flux and proton flux measurements. This is an important consideration for a model such as the one presented here, which is aimed at real-time forecasting. Tens of thousands of solar flares have occurred over the duration of the GOES satellite's missions, and their X-ray enhancements have been readily identified in the measurements. As discussed, SFs are well established to have a key role in the generation of SEP events thus making them a good candidate for predicting SEP occurrences. Multiple works over the years have explored the relationships between SF X-rays and SEPs, and how well the former can be used to predict the latter, e.g., Cliver & D'Huys (2018), Papaioannou et al. (2016), Grayson et al. (2009). Recently Steyn et al. (2020) demonstrated the use of SF X-rays in a proof-of-concept physical model for the estimation of the injection of solar particles, while Kahler & Ling (2018) used two variables derived from the peak fluxes of GOES short-band [0.05–0.4 nm] and long-band [0.1–0.8 nm] X-rays as a simple method for the prediction of SEP occurrence. Laurenza et al. (2009) used logistic regression to provide probabilities of SEP occurrence using several solar variables, including the SF fluence of GOES X-rays, while Papaioannou et al. (2018) used Principal Component Analysis (PCA) on a set of solar variables, including the magnitude of SF GOES X-rays for the nowcasting of SEP events. Here, we have prepared and used a dataset of continuous

time-series of GOES X-rays and derived variables from several thousands of SFs defined in the NOAA GOES SF catalogue to create a parameter space used as input to NNs. We also utilize a SEP catalogue (Pacheco, 2019) with 257 identified solar proton events in 1988–2013 which includes their associated solar flares. These 257 enhancements directly correspond to 172 events in the SEPEM Reference Event List (REL) (Crosby et al., 2015), in which temporally close enhancements are compounded in a single Event, hence the smaller number. Using this data, we have designed and applied a robust methodology employing NNs for the prediction of the occurrence of SEPs. Our model provides probabilities of SEP occurrence readily transformed to categorical binary yes/no predictions. We train and evaluate the performance of the model and proceed to validate the results in a self-consistent and reliable manner. It is shown that the performance of the model is very good in correctly predicting SEPs while also crucially achieving low false-positive rates.

2 Datasets and Machine Learning methodology

2.1 X-ray data and SF catalogue

The GOES X-ray Sensor (XRS) (Garcia, 1994) provides solar X-ray fluxes for the wavelength bands of 0.05–0.4 nm and 0.1–0.8 nm, namely the "short" and "long" channels, respectively. X-ray Measurements have been made since 1974 by NOAA GOES satellites, in a configuration where at each time, a number of satellites provide real-time data while one is usually assigned to be primary and others as secondary. We have used flux data from XRS and performed pre-processing required to create a consistent homogeneous time series. Data files of flux measurements from NOAA (ftp://satdat.ngdc.noaa.gov/sem/goes/data/avg/) were used with a 1-minute cadence. The data files were parsed to create a large time series with all measurements, including the satellite number tag on each measurement. Using the list of time ranges given by NOAA (https://www.ngdc.noaa.gov/stp/satellite/goes/doc/GOES_XRS_readme.pdf) to determine which satellite is primary at a given time, the large time-series was pruned to keep measurements from primary satellites. For the years before 1986 – when the primary/secondary scheme was started – we retained the measurements from the highest numbered GOES satellite – a selection, however, not crucial for the needs of this work. The final product is time series starting at 1976 January 01 and ending in 2019 October 23, containing timestamps and the long and short X-ray channel measurements. With this continuous time-series, we use the NOAA GOES solar flare catalogue (ftp://ftp.ngdc.noaa.gov/STP/space-weather/solar-data/solar-features/solar-flares/x-rays/goes/). In particular, we focused on SXR flares in the period 1988–2013 because in this time span, we have the identification and association of SEP events as discussed below. It is noted that the SFs considered are of magnitude C1 and above (peak flux $\geq 10^{-6}$ W/m^2 at the long [0.1–0.8 nm] channel).

2.2 SEP events catalogue

The SEP event catalogue we have used (Pacheco, 2019) contains 263 SEP events, 257 of which have been associated with an SF for which the SF starting time, class in terms of

the long SXR peak flux, and location (heliolongitude and heliolatitude) are listed, among other information about the SEP events. This catalogue was built upon the 172 events of the SEPEM Reference Event List (REL) (http://sepem.eu/help/event_ref.html) which occurred from 1988 to 2013, with the first event starting on 1988 January 02 and the last one on 2013 March 15. In the SEPEM REL, an Event is defined to start when the differential proton flux in the [7.23–10.46] MeV channel rises above 0.01 p/(cm^2 s sr MeV) maintains equal or higher levels for at least 24 h and also achieves a peak flux of 0.5 p/(cm^2 s sr MeV). Therefore, any less intense or shorter SEP events not reaching these criteria are excluded from the catalogue. Also, it is noted that the event definition places emphasis on lower proton energies, and as such, it is not guaranteed that an appreciable rise in proton flux, above background values, has occurred for higher energies.

However, in this catalogue, the events are not compounded as they are in the SEPEM REL, discussed above. It is usual in SEP catalogues if two or more enhancements have occurred very close in time to be compounded together and be considered a single event, especially if there is a partial overlap of the declining phase of an enhancement and the rise phase of the next, as is often the case in multiple-enhancement events. However, for the needs of accurately predicting SEPs, not compounding separate enhancements is optimal as more than one solar eruption can occur in a small period, and a model should be able to predict a new event even if there is an already ongoing one.

The 172 events in the SEPEM REL were analysed one by one to determine the most plausible (with the available information) parent solar eruption(s) and the different SEP event enhancements in compound instances (Crosby et al., 2015). The times of the SEP events were checked to include the 5–300 MeV energy range with a common start and end time across energies. To elaborate this catalogue, the SEPEM Reference Dataset 2.0 (e.g., Jiggens et al., 2018) proton flux intensity time profiles were plotted together with <5 MeV proton intensities either from the IMP-8 CPME detector (Armstrong, 1976) or the ACE/EPAM sensor (Gold et al., 1998) to help identify the parent solar sources and interplanetary (IP) shock crossings, together with interplanetary magnetic field and solar wind plasma measurements, when available. The division of compound events into single SEP events was performed by first considering the new high energy enhancements encompassed within the 7.23–10.46 MeV proton enhancement. Secondly, in the case that a new low-energy increase was detected, but the high energy intensities remained elevated due to a previous event, the detected IP shocks were used to look for the most probable associated coronal mass ejections (CMEs) and concomitant SF responsible for the new increase. To find the main solar eruptions generating each SEP events, many published associations of parent solar sources with SEP events were checked (e.g., Lario et al., 2001, 2013; Cane et al., 2010; Richardson et al., 2014) and confirmed by using different publicly available SF X-ray and H$_\alpha$ flare catalogues, CME catalogues, and the Extreme Ultraviolet event catalogue from STEREO. For the years prior to the SOHO spacecraft, when no-routinely CME observations were available, and for those events for which a published identification was not found, or it was doubtful, the NOAA's Solar Geophysical Data Comprehensive catalogues were used

to determine the more plausible SF associated with the SEP event. In those instances where CME observations were not available to help find the solar source site and more than one SF were plausible particle sources, the SF with the higher X-ray peak flux was associated with the SEP event. In the case of events associated with X-ray flares beyond the western limb, the heliolongitude determined in published studies was used, and in those few cases where such information was not found, the heliolongitude was set to W90.

The SF catalogue is limited to match the start and end time of the SEP catalogue so as not to include flares potentially associated with the SEPs that have occurred outside the 1988–2013 time span. For each SF, using its start and end time, the X-ray time-series are taken from the X-ray dataset. All non-appropriate values (NaNs, data gaps, negative fill values) are removed and a minimum of 5 usable measurements are imposed on the X-ray time series of the flares; if for a flare less than 5 usable measurements are available, it is not used. Also, all flares with heliolongitudes larger than 90 or lower than −90 are not used. Along with the cosine and sine of the flare's heliolongitude, a set of 24 parameters is extracted from the X-ray time-series, primarily involving the peak flux and fluence (time-integrated flux over the duration). Details on all the input parameters can be found in the Supplementary materials. The final dataset used contains ~18 000 (17 797) SFs not associated with an SEP and 220 SFs associated with an SEP. The dataset contains all SFs with SXR magnitude ≥ C1 from 1988 January to 2013 March spanning 25 years and covering the largest part of Solar Cycle 22, the whole Solar Cycle 23, and the rising phase of Solar Cycle 24. To our knowledge, this is one of the most extensive datasets used to date for such an approach, and it allows us the design of a methodology with real application potential. Figure 1 shows the Peak fluxes of all SFs used in this work, the solar cycle dependence is clearly seen, and the SFs associated with SEP events are marked red.

2.3 Machine Learning methodology

For the ML classifier, we used multilayer neural networks where the architecture, algorithm, and back-end training procedure were implemented in Python using the TensorFlow library version 2.1.0 (Abadi et al., 2016). The architecture used is that of a feedforward network with multiple layers, and it is a modified version of the one used in Raptis et al. (2020). The hyperparameters of the model were obtained through an empirical investigation on the performance and the stability of the model. For the model initialization and the training algorithm, we use the stochastic gradient descent (SGD) with momentum as implemented in the high-level API Keras framework (Chollet, 2018) using early stopping with internal validation subset from the training set. An adaption was required due to the imbalanced number of samples per class of the dataset, discussed below. Typically, a successful training of an NN requires at the very least hundreds, and preferably thousands, of samples. While this is the case here, there is the issue of the severe class imbalance between Flare-SEP (1 – SEP occurred) and Flare-noSEP (0 – SEP did not occur) samples, where the latter outnumber the former by a factor of ~80. Class imbalance is a non-trivial issue in machine learning optimization problems and often the solutions are highly case-dependent ranging from creating synthetic samples by
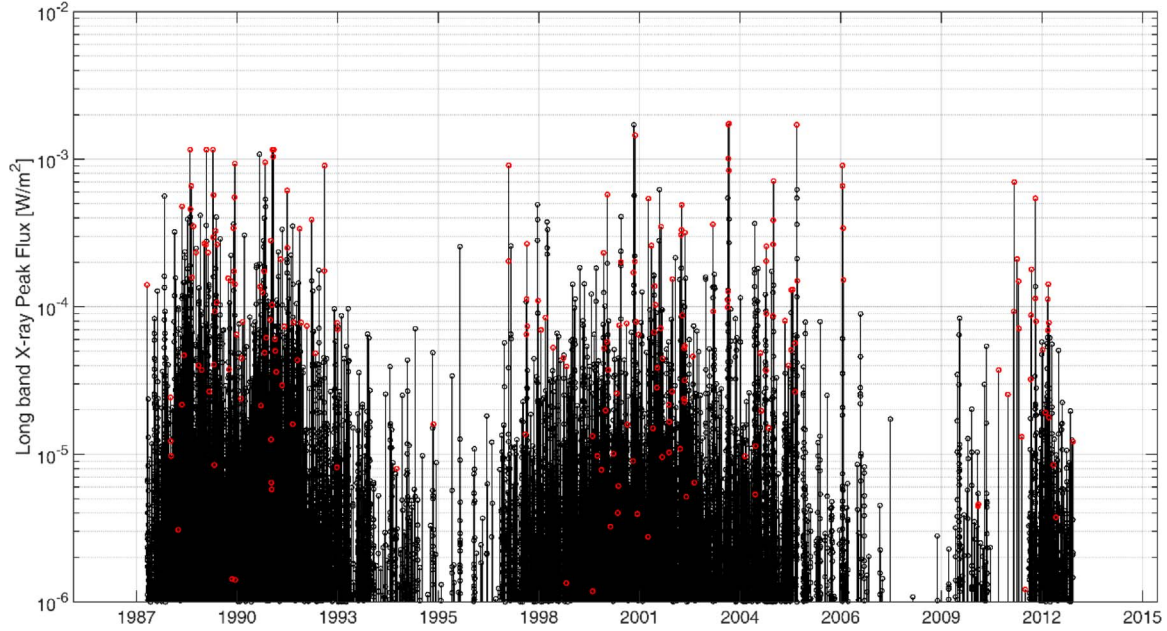
Fig. 1. X-ray Peak fluxes in the long band of all solar flares used in this work for the 25-year time span. Flares that are associated with SEPs are marked in red circles.

introducing small random perturbations in the actual data to more complex and sophisticated techniques (Goodfellow et al., 2016; Brownlee, 2020). In order to tackle this issue, we tested several under-sampling and up-sampling techniques (Lemaître et al., 2017) as well as the widely used SMOTE method (Chawla et al., 2002). When under-sampling, we decreased the majority class (Flare-noSEP) samples to be equal to those of the number of Flare-SEP class. However, such under-sampling, while solving the class imbalance problem, reduces significantly the size of the training dataset and also removes many samples that contain vital information for the successful classification of both classes. Conversely, over-sampling techniques focus on enlarging the underrepresented class by creating synthetic samples based on existing ones. However, such techniques may create samples with unphysical properties that could introduce unforeseen biases when treating complex physical-related parameters as is the case here; moreover in the present case over-sampling techniques did not increase the model's performance. The best performing and selected approach was error penalization. With error penalization, the NN is taught that it is more costly to misclassify the underrepresented class (here the Flare-SEP class) over the other. As a result, the internal parameters of the network (weights and biases) are updated differently for the samples of the two classes. Specifically, by applying error weighting (EW) factors related to the class imbalance ratio of the training set, the network preferentially changes its internal parameters (weights and biases) when trained with a Flare-SEP sample and is less sensitive to adapt its parameters for Flare-noSEP samples. The used error weight factors can be seen in equation (1):

$$EW_{\mathrm{noSEP}} = \frac{N_{\mathrm{SEPs}}}{N_{\mathrm{SEPs}} + N_{\mathrm{noSEPs}}},$$

$$EW_{\mathrm{SEP}} = \frac{N_{\mathrm{noSEPs}}}{N_{\mathrm{SEPs}} + N_{\mathrm{noSEPs}}}, \qquad (1)$$

where N_{SEPs} is the number of Flare-SEP samples, and N_{noSEPs} is the number of Flare-noSEP samples in the training set. We introduce the weight factors directly in the error function and therefore minimize a "weighted" cross-entropy loss function as seen in equation (2):

$$L = -EW \cdot \sum y_{i,\mathrm{true}} \cdot \log \left(y_{i,\mathrm{pred}} \right), \qquad (2)$$

where $y_{i,\mathrm{true}}$ is in one-hot encoding (e.g. class 1 = [1, 0], class 0 = [0, 1]), and $y_{i,\mathrm{true}}$ is the vector containing the probabilities that a sample belongs to each of the two classes outputted by the network. It is seen that the error (loss) for misclassifying Flare-SEP samples becomes much larger than that of misclassifying Flare-noSEP samples. It is worthy of note that error penalization as implemented here avoids altogether the problems of under- or up-sampling as discussed above. The deep neural network architecture can be seen in the left panel of Figure 2. It consists of five hidden layers (50 – 80 – 100 – 80 – 40 neurons) and uses the batch normalization technique. Batch normalization effectively normalizes the output of the neurons of a layer. As a result, it decreases slightly the computational time of the training and applies regularization to avoid over-fitting problems (Ioffe & Szegedy, 2015). In the hidden layers of our model the LeakyReLU activation function is used, which is a slightly modified version of a Rectified Linear Unit (ReLU). This version of ReLU solves a problem in the neural network, which is called "vanishing gradient". This problem may occur in ReLU units due to the zero-slope part of the function (Maas et al., 2013). The output of the NN is two probabilities that sum to unity, one that the sample is a Flare-SEP case and one that it is a Flare-noSEP case. Whichever is the highest is the final categorical prediction for the SEP occurrence by collapsing the probabilistic output using a threshold of 0.5. A further design we have implemented is the use of more than one network in the classifier, in a "committee" scheme, as seen in the right panel of Figure 2. In this
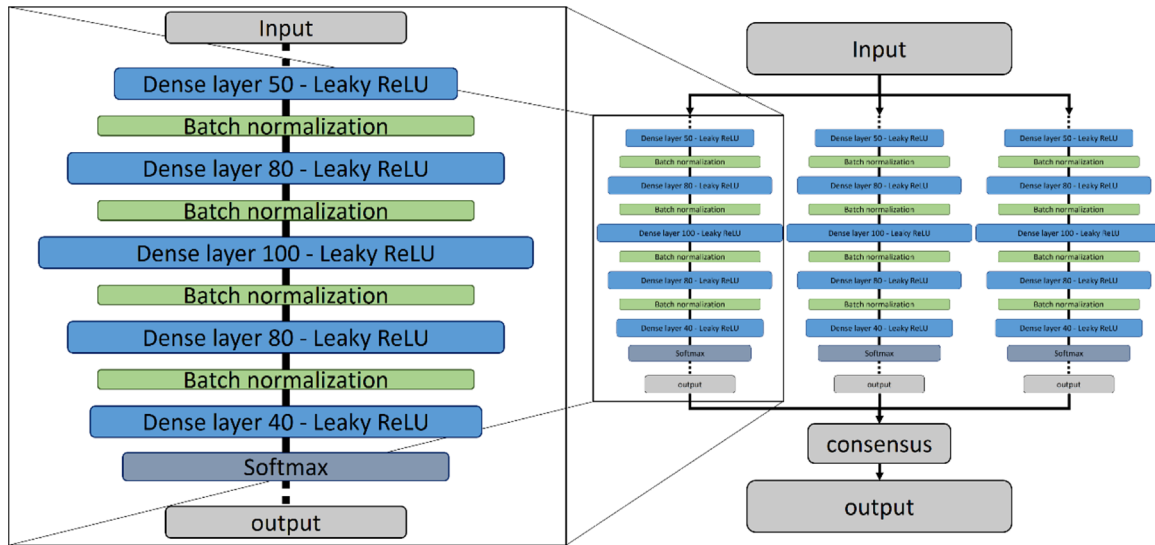
Fig. 2. Left panel: the architecture of the NNs showing the layers and numbers of neurons. Right panel: the structure which combines the outputs of the individual NNs.

approach, the networks are trained in parallel with the same training set, and their outputs on the testing set are combined, where the final output is the "consensus" of all committee members. Consensus is formulated as the arithmetic mean of the individual probability outputs. This is a modular approach that easily allows the addition or exclusion of classifiers-networks without changing the base architecture of the model. This method can improve results as it mitigates situations where the random initialization of a network's internal parameters leads to a less than optimal training, and the network may underperform by being trapped in local minima of the error space. This is more possible to occur when there are few training samples from one, or both classes and a random training set can differ non-negligibly from another. Here we have used three networks, which is the minimum meaningful number for a committee when applying the model to the full dataset and 10 networks in a limited version of the dataset. Details are discussed in the section below outlining the experimental results.

3 Results: evaluation and validation

Using the dataset and classifier scheme detailed above, the performance of the model is evaluated, and the results are further validated; for this two different methodologies are employed and discussed below.

3.1 Evaluation – random-subset methodology

The first methodology entails the creation of a training dataset with random sampling from both classes and the use of all remaining data as the test dataset. For Flare-SEP samples, we use a [0.8–0.2] training-testing ratio where 180 samples are used for training and the remaining 40 for testing. The training dataset is completed with the use of 3000 Flare-noSEP samples, and the remaining ~15 000 are used for testing. We use

3000 Flare-noSEP samples as firstly it is a large enough number to provide a representative training set for this class to the model, and furthermore, due to the weighted error function we have used, there is no real benefit to increasing the number of these samples in the training set. The Flare-noSEP samples are taken randomly but uniformly across the 25 years of data to avoid introducing any potential bias that may exist regarding the conditions of the different solar cycles. As the 40 Flare-SEP samples are very few to provide a good evaluation from a single test-run, we iterate this procedure of random sampling-training-testing 100 times to evaluate the performance and overall stability of our model. Figure 3 shows the results of 100 iterations in terms of the true positive (hit) rates (TPR – correct prediction of SEP occurrence), also called Probability of Detection (POD), false-negative (miss) rates (FNR = 1 − TPR − wrong prediction of SEP not occurring), false-positive rates (FPR – wrong prediction of SEP occurring), and true negative rates (TNR − correct prediction of SEP not occurring), TPR, FPR, and TNR definitions are seen below in equation (3):

$$\text{TPR} = \frac{TP}{TP + FN}, \text{FPR} = \frac{FP}{FP + TN}, \text{TNR} = 1 - \text{FPR}, \quad (3)$$

where *TP* and *FN* are the numbers of true positives (hits) and false negatives (misses), and *FP* and *TN* are the numbers of false positives and true negatives. It can be seen in Figure 3 that the classifier performs very well, with the mean SEP hit rate being ~89% and the mean true negative rate being ~92%. These results are a strong indication that the employed approach works well for the prediction of SEPs while at the same time having a very low false-positive rate. It is noted that the true negative rates show a tight clustering with a range of values ~4%, while the SEP hit rates have a wider range from 75% to 100%. This is understood by the fact that there are much fewer Flare-SEP samples, and the random division into training/testing sets can lead to this performance variance depending on which are used for the latter and the former in each run, especially since as it will be discussed further below, there are SEPs that are consistently missed.
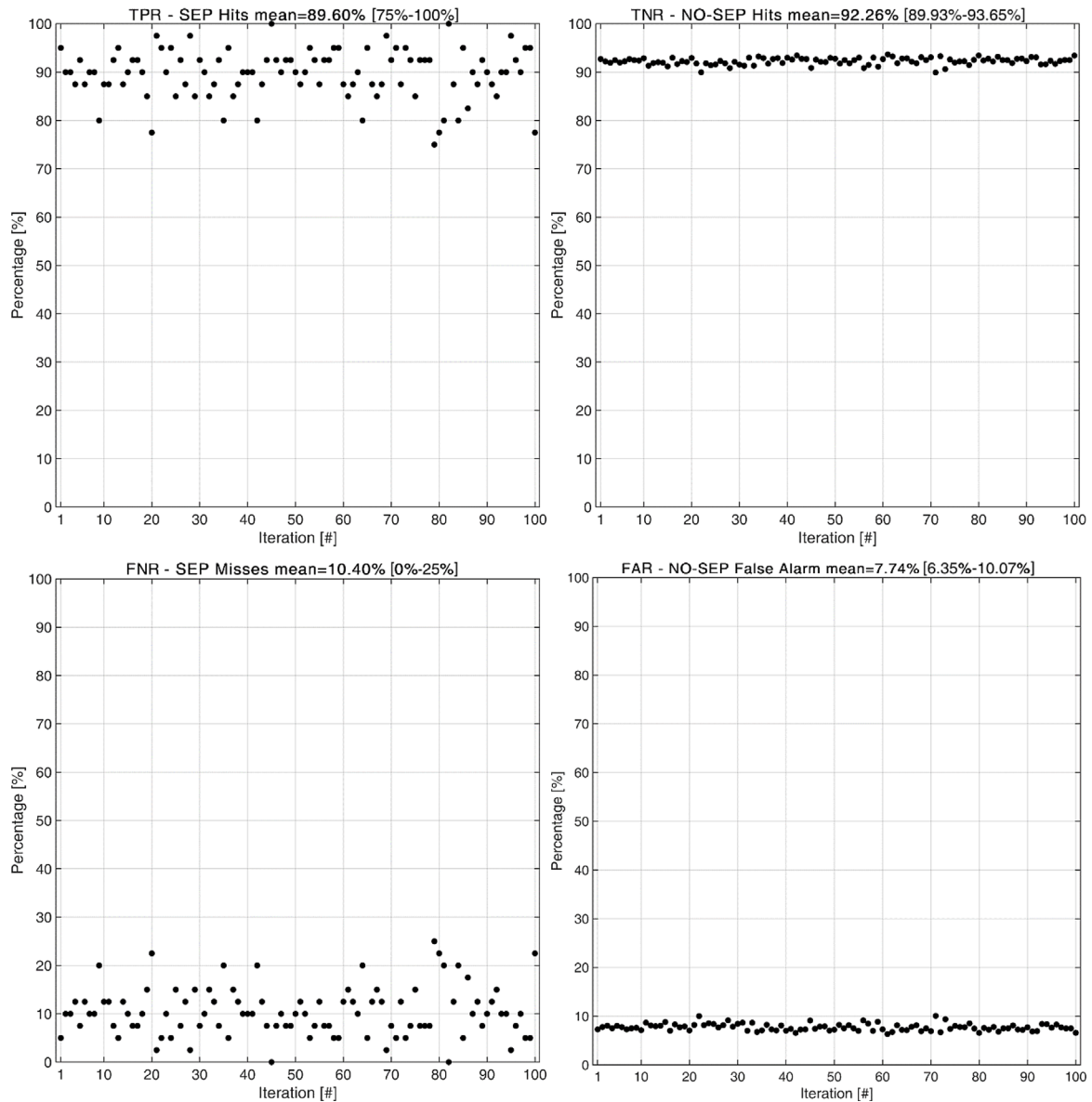
Fig. 3. Results from 100 iterations of random sampling-training-testing. Top panels: correct prediction rates for true positives (SEP occurred) and true negatives (no SEP). Bottom panels: wrong prediction rates for false negatives (missed SEP) and false alarms.

As a statistic skill score, we use the True Skill Score (TSS) as seen in equation (4) below:

$$
\begin{aligned}
\text{TSS} &= \frac{TP}{TP + FN} - \frac{FP}{FP + TN} = \text{TPR} - \text{FPR} \\
&= \text{TPR} - (1 - \text{TNR}).
\end{aligned} \tag{4}
$$

TSS takes values in the $[-1, 1]$ range, with 1 being a perfect score, -1 being completely wrong, and 0 being a random classifier. As seen, it is essentially the true positive rate (first fraction) minus the false positive rate (second fraction), and thus, it is insensitive to the relative number of samples in the two classes. This is a major consideration in the objective performance evaluation of forecast methodologies where samples from one class heavily outnumber those of the other, as is the

case here. TSS is the only widely used statistical skill score that is unbiased with respect to class imbalance, unlike other metrics, e.g., the often used Heidke Skill Score. Additionally, and for the same reason, TSS is very appropriate for objective comparative evaluations of models that use different methodologies as well as datasets with potentially very different numbers of samples. For an in depth, discussion, see Bloomfield et al. (2012) and Bobra & Couvidat (2015), who examined the use of such metrics in the similarly imbalanced problem of solar flare prediction. Figure 4 shows the calculated True Skill Scores (TSS) from the 100 iterations which ranges [0.675–0.927], and the cross-plot of true positives rates (Flare-SEP hits) versus true negative rates (Flare-noSEP hits). The latter exhibits the typical trend in binary classification problems where the two scores are anti-correlated due to the competing categorization in two classes. A strong anti-correlation is an indication that a model
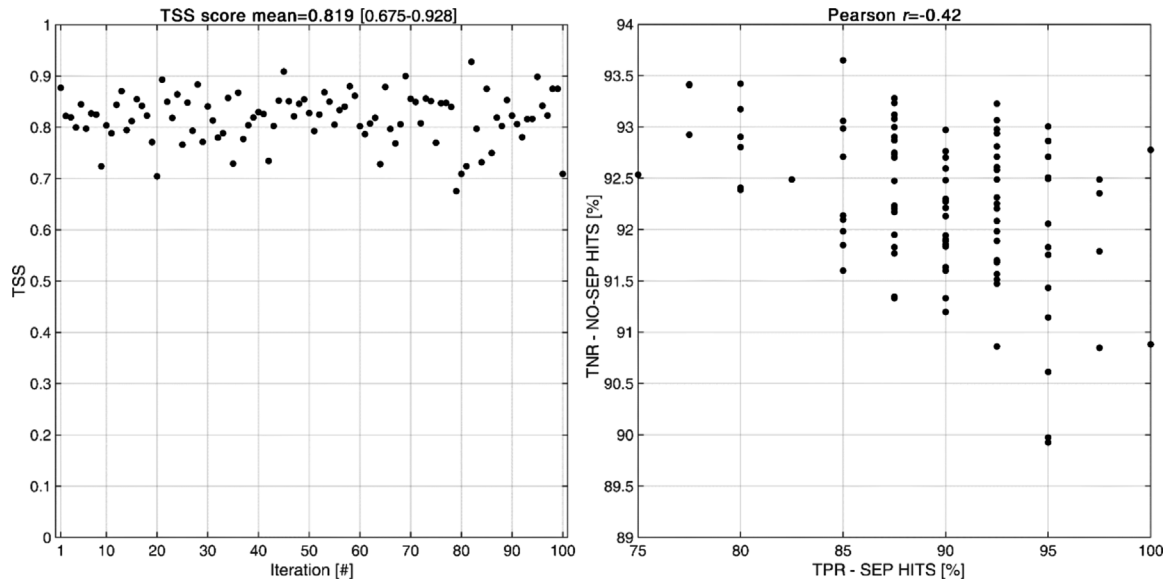
Fig. 4. Results from 100 iterations of random sampling-training-testing. Left panel: calculated True Skill Score with a high mean value. Right panel: cross-plot of TPR versus TNR showing a weak anticorrelation.

Table 1. Contingency matrix with mean percentages from the 100 evaluation iterations.

Contingency matrix	SEP predicted YES	SEP predicted NO
SEP occurred YES	**89.6%** [75%–100%]	**10.4%** [25%–0%]
SEP occurred NO	**7.74%** [10.2%–6.4%]	**92.26%** [89.8%–93.6%]

is underperforming, which is not the case here. Table 1 shows the contingency matrix with the mean values and ranges from the 100 runs of the evaluation process.

3.2 Leave-one-out methodology

3.2.1 Full dataset

The evaluation discussed above shows the very good results achieved on the prediction of SEPs while providing a general picture of the performance of our model. However, a robust approach with more definitive results is needed in order to validate our model. For this purpose, we have used a leave-one-out process, see e.g. Aminalragia-Giamini et al. (2020), Raptis et al. (2020), also referred to as *k*-fold cross-validation. In this process, all but one of the 220 Flare-SEP samples are used for training along with a randomly sampled subset of 3000 Flare-noSEP samples. After training, the singular Flare-SEP sample which was set apart, plus the ~15 000 Flare-noSEP samples, are tested. If a correct prediction for the Flare-SEP sample occurs, it is marked as a hit. If the prediction is wrong, then it is identified as a miss. Additionally, we apply an even stricter requirement in order to accurately test our model's robustness. We require that the singular Flare-SEP sample is correctly predicted in *k* = 10 repetitions, using in each repetition a different Flare-noSEP training subset of 3000 flares, which are again randomly sampled. This essentially removes any bias from favorable cases where the Flare-noSEP training subset happens to be optimal and requires that there is a correct prediction regardless. Finally,

we consider the sample to be a successful hit only if the prediction is correct in all 10 repetitions. We note that this requirement is very strict and, in essence, shows the lowest possible attainable scores. This process goes iteratively over all the 220 available Flare-SEP samples and provides objective and true predictions for each one of the historically occurred SEPs as if each one was the next to occur. Figure 5 shows the absolute numbers of SEPS that were predicted correctly, those that were never predicted correctly, and the in-between cases. Table 2 shows the results from the validation process.

As it is seen, 191 of the 220 SEPs are always predicted correctly ("perfect" samples), and the average of correct prediction for Flare-noSEPs from all 2200 runs (10 for each of the 220 events) is 92.23% which corresponds to an absolute number of 13 648 of the total 14 797 Flare-noSEP samples in the test set of each run. These results are in agreement with those of the evaluation runs and validate the capability of our approach to quite accurately predict the occurrence of SEP events from solar flare X-ray measurements while achieving a very low false-positive rate of ~8%. Finally, using the derived average false positive rate for Flare-noSEPs, a TSS value of 0.790 can be calculated, which indicates the good performance of the model. It is interesting to note that 10 of the events were predicted correctly between 1 and 9 out of 10 times (in-between cases), and the remaining 19 events were never correctly predicted. The existence of events that are consistently never predicted along with the otherwise high hit rate for events always predicted correctly is not typical of an expected variance in the performance of such a system which can be explained
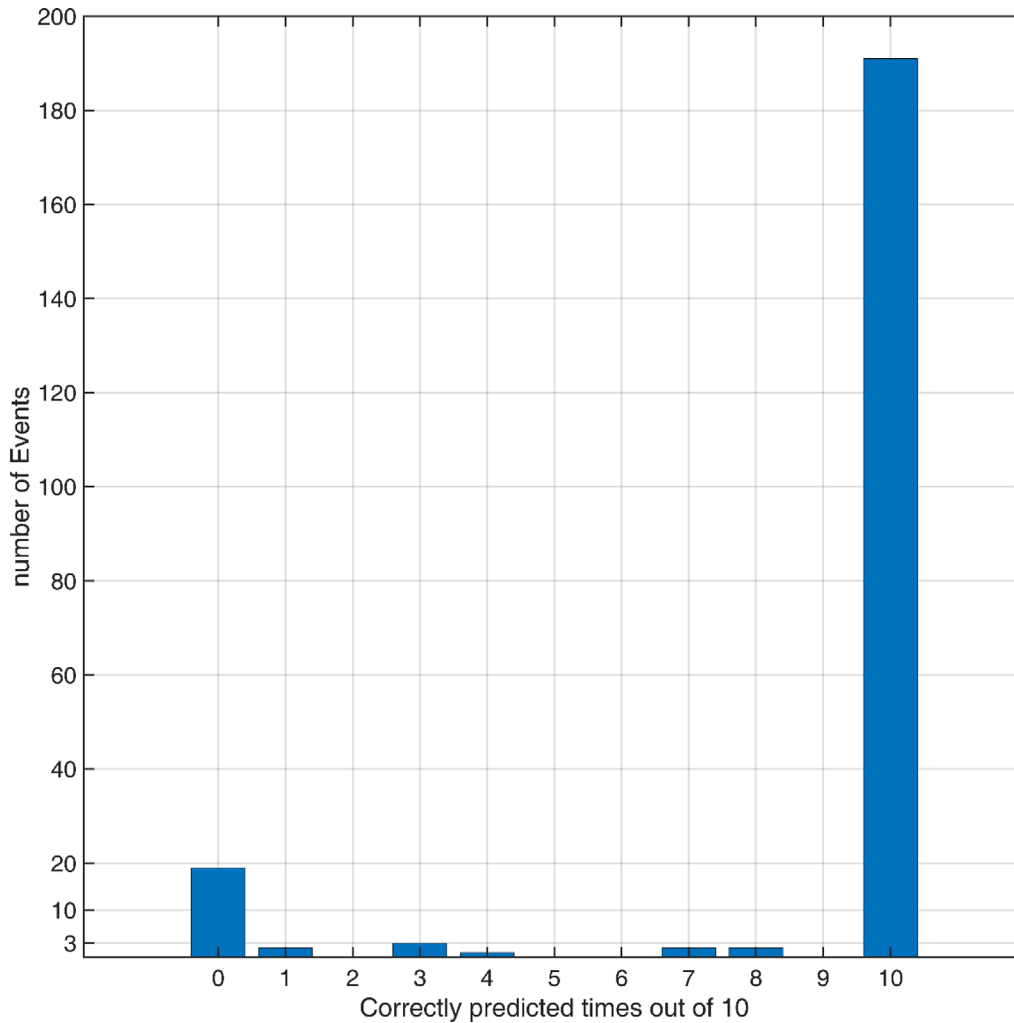
Fig. 5. Flare-SEP samples with the numbers of their correct predictions in the $k = 10$ cross-validation process.

Table 2. Results from the k-fold validation process.

	SEP predicted always YES	SEP predicted always NO	SEP predicted sometimes
SEP occurred YES	**86.81**% [191/220]	**8.63**% [19/220]	**4.54**% [10/220]
SEP occurred NO	**7.77**% [1149/14797]	**92.23**% [13648/14797]	–

from the randomization of training samples (for Flare-noSEPs) as well as the random initial values of the networks' internal parameters as previously discussed. It would be much more typical for the non-"perfect" samples, to be all sometimes correctly predicted and sometimes not. This could entail some interesting possibilities, discussed further below.
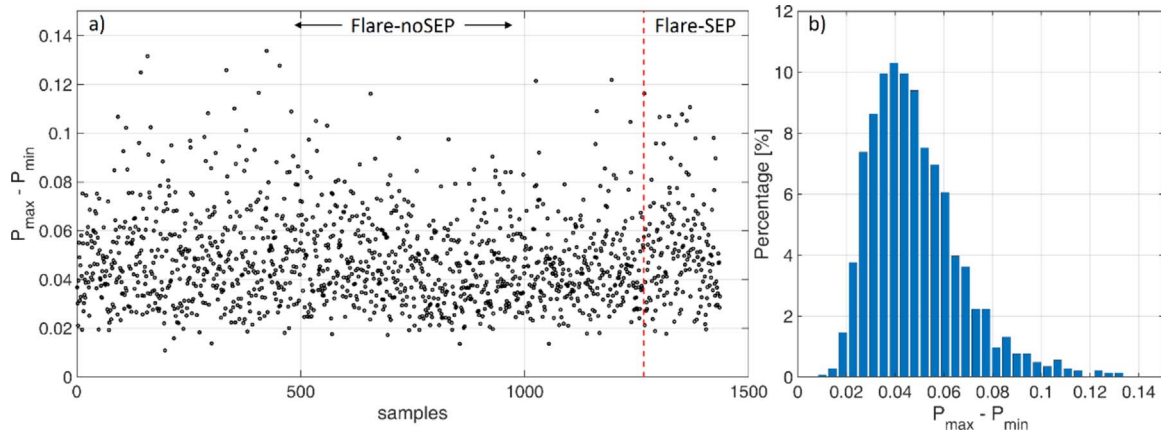
3.2.2 Solar flares with class ≥ M2

We further investigate the performance of our model by limiting the dataset to flares of class equal or higher than M2 (X-ray peak flux of 2×10^{-5} [W/m^2] in the long channel). In real-life operational conditions often, this or similar limits are imposed on the flares that are considered as candidates for SEP occurrence in order to avoid a potentially large number of false alarms. The reduction to class ≥ M2 reduces the Flare-SEP samples to 171 and the Flare-noSEP samples to 1265. While the former is an important but smaller reduction in a number of samples, the latter is a much more dramatic reduction of more than one order of magnitude. The total number of available 1436 (1265 + 171) samples is close to the limit of what is considered reasonable for training an ML model such as the one used here. Even so, we show that our methodology still achieves good scores. The same leave-one-out process as described in the previous section is performed, but now all available data are used for training except the one investigated, i.e., when a Flare-SEP sample is examined, all the 1265 Flare-noSEP samples are used for training, rather than a random subset. Additionally, the relatively limited number of Flare-noSEP samples allows us to apply the leave-one-out process to this class as

Table 3. Results from the leave-one-out *k*-fold validation process for solar flare class≥M2.

	SEP predicted always YES	SEP predicted always NO	SEP predicted sometimes
SEP occurred YES	**78.36%** [134/171]	**17.71%** [31/171]	**3.51%** [6/171]
SEP occurred NO	**15.81%** [200/1265]	**80.39%** [1017/1265]	**3.79%** [48/1265]

Fig. 6. Maximum probability minus minimum probability of the model from the *k* = 10 repetitions for each sample. (a) Values for all 1436 Flare-noSEP and Flare-SEP samples; (b) Distribution of values from (a) showing the majority clustered in the [0.02–0.08] range.

well; when such a sample is examined, all the Flare-SEP samples are included in the training set. Due to the reduced number of available samples, we take advantage of the discussed modularity of our model, and for the optimization of its performance, the network committee members are increased to 10. Table 3 below shows the results from the validation process

It is seen that even with the severe reduction of training samples, the model performs quite well, albeit with reduced scores. For the Flare-SEP class, a true positive rate of 78.36% is achieved, and the true negative rate is 80.39%. This is a reduction of approximately 8.5% and 11.8% respectively compared to using the full dataset. It is noteworthy that there are 31 never-predicted events which indicate that at least some of these false predictions are possibly due to limitations in the approach, and this is aggravated by the limiting of the training data. The application of the leave-one-out process for all samples also allows the investigation of the probabilistic outputs. Figure 6 shows the range of probabilities derived from the *k* = 10 repetitions for each sample. The range is calculated as the maximum minus the minimum probability given by the model that a sample belongs to the Flare-SEP class. It is seen that the range is typically small, indicating that there is good stability in the outputs of the model. The distribution of the probability ranges shows that the majority is found in the [0.02–0.08] span, which are very small values.

The probability outputs are also used to derive the receiver operating characteristic (ROC) curve and the respective area under the curve (AUC) value, two very widely used statistical tools. The ROC curve shows the evolution of the true positive rate versus the false positive rate as different threshold values are applied on the probabilities the model outputs; the integral of the ROC curve, the AUC value, is a metric on the performance of a model. A random classifier will have an AUC = 0.5 and a perfect classifier AUC = 1; AUC values above

0.75 are considered good or better. Figure 7 shows the derived ROC curve for the threshold value range [0, 1] as well as the evolution of the calculated TSS for the same threshold range. Both the SF ≥ C1 and SF ≥ M2 cases are shown, and the model achieves AUC values of 0.938 and 0.857, respectively, while for threshold = 0.5, the TSS values are 0.79 and 0.5876.

4 Discussion and future work

Besides the good performance shown in the results, as discussed, there are a number of missed SEPs and false positives from non-SEP flares. The simplest explanation is, of course, possible limitations in our approach. Such limitations may be related to the basic premise employed, the use of SF X-rays as the SEP predictor, the otherwise limited number of occurred SEPs, design parameters such as the inputs or the NN architecture, or even caveats in the X-ray data itself such as signal saturation that occurred predominantly in older GOES satellites. However, these answers do not offer much new information as they are inherent in this work. Following, we discuss our results, what could be inferred from them if taken at face value, and offer alternative explanations and potential future steps. We consider here the results from the full dataset as it can offer the most information. Figure 8 shows the empirical cumulative distributions of the Peak fluxes and Fluences of the proton component from the SEP events used in this work, at 6 MeV, 38 MeV, and 115 MeV, the never-predicted (NP) events are marked in orange. It is seen that the NP events lie in the lower ends of the distributions having comparatively weak Peak fluxes and low Fluences; even more so when examining the higher energies, where in the majority of cases, they are mostly background. An example of such an NP event can be seen in the
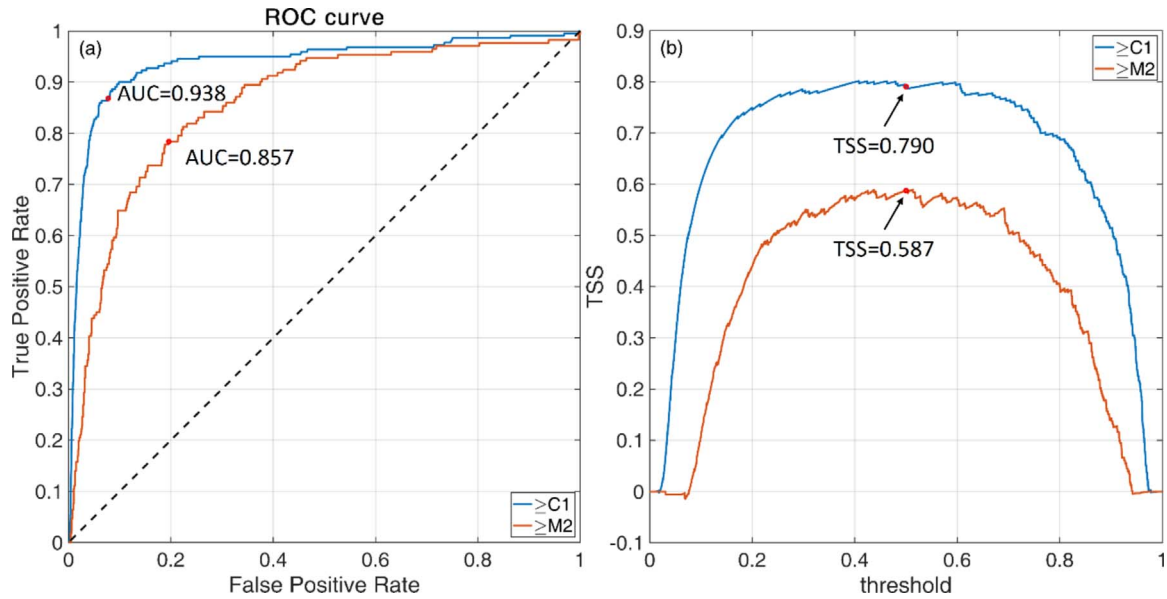
Fig. 7. (a) The ROC curve derived from the probabilistic outputs of the model showing the relation of true positive rates and false-positive rates for all thresholds values $\in$ [0, 1]. Two curves are shown for both the SF $\geq$ C1 and SF $\geq$ M2 cases. The dashed line denotes the random classifier behaviour. (b) Evolution of the TSS score for the same range of applied thresholds. Red circles mark the points for a threshold equal to 0.5.
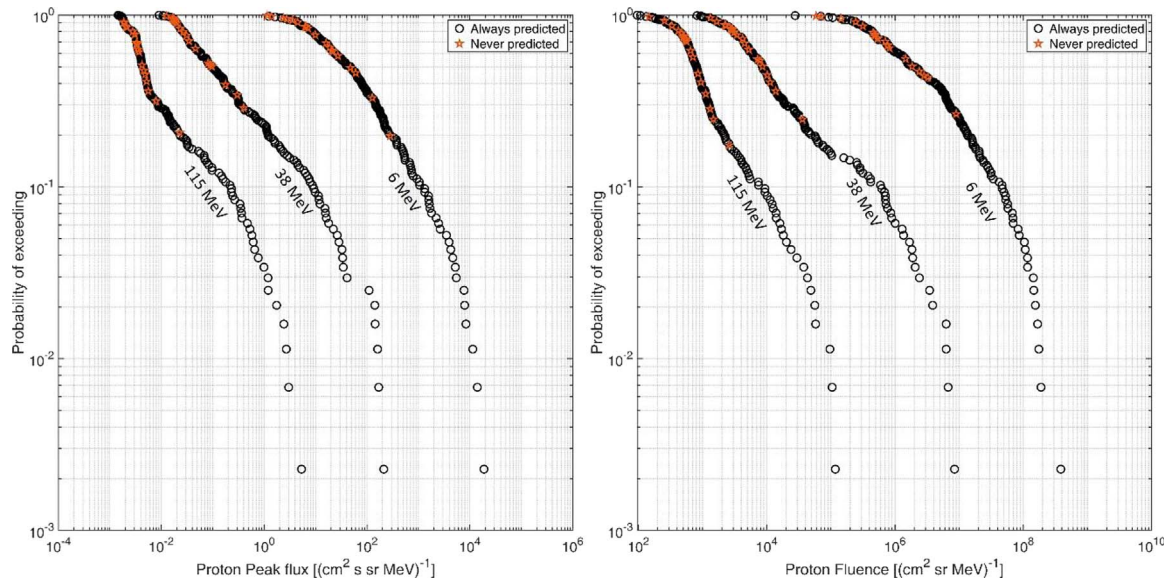

Fig. 8. Experimental cumulative distributions of the Peak Fluxes and Fluences of the SEP proton components at 6 MeV, 38 MeV, and 115 MeV. Events always predicted are in black circles, and the never predicted are in orange stars.

Supplementary materials. We note that as discussed in Section 2.2, the Event definition used is based on the [7.23–10.46] MeV channel, and therefore it is entirely possible that a weak enhancement at low energies will not show any appreciable fluxes at higher ones and therefore be mostly or entirely in the background.

Importantly, the fact that the NP events are mostly in the low ends of the distributions implies that even if our approach missed an event, it would not be a major one, something that would be detrimental to any model aiming to predict SEP occurrences. Secondly, it is interesting to note that, limitations of our model notwithstanding, these results can be interpreted as going against the classical paradigm of the division of SEPs in impulsive and gradual events. In this binary division, which has been however often questioned or revised see e.g. (Reames, 2020), impulsive events are typically attributed to being driven by SFs accelerating heated flare material, and they exhibit low proton fluxes and short durations, whereas gradual SEPs have more intense fluxes, longer durations, and are mainly attributed to coronal mass ejections, being accelerated in coronal and
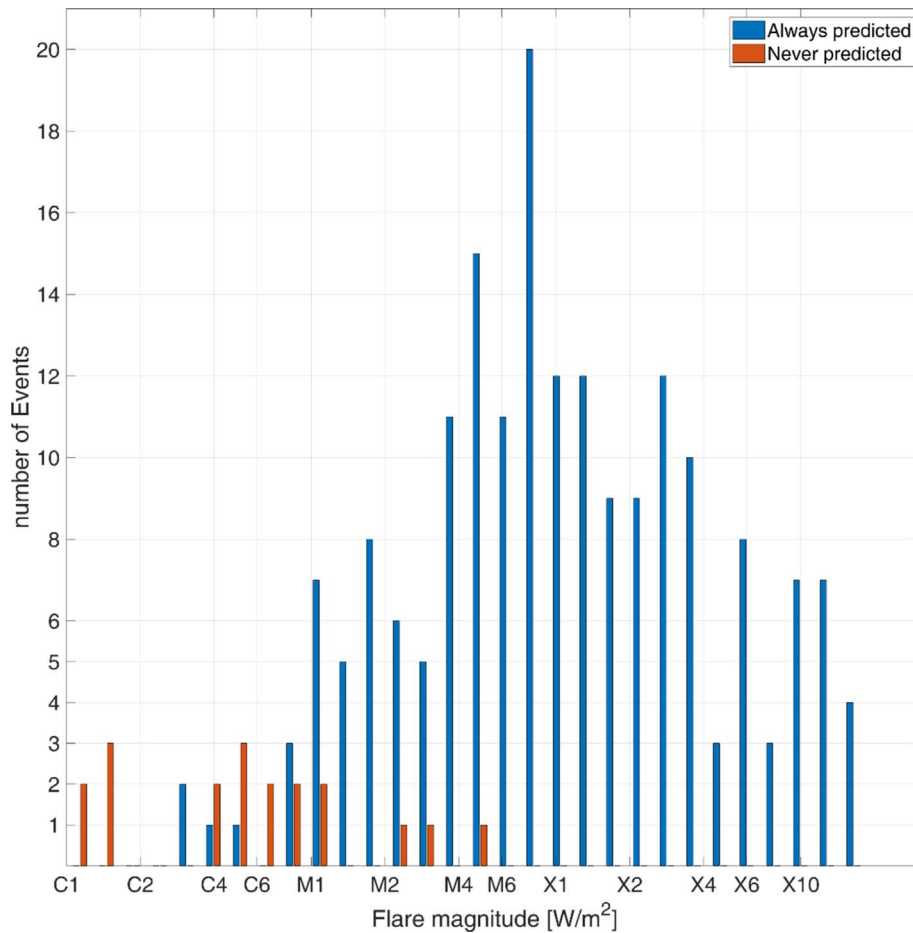
Fig. 9. The number of events in 30 logarithmic bins of the magnitude of their associated SFs. The NP events mainly occupy the lower end of the flare magnitudes.

interplanetary shocks. In our approach, we use only characteristics of SFs to predict the occurrence of SEPs; one, therefore, could ostensibly expect impulsive, and thus generally weaker, events to be preferentially predicted. Our results do not show such bias, if anything, the contrary is true. However, it must be noted that this can be possibly attributed to "big flare syndrome" where the more intense flares are more correlated with subsequent SEP events, either gradual or impulsive. An explanation for the NP events we explored was whether they could be attributed to Filament Eruptions (FEs) (Gopalswamy et al., 2015). FEs are eruptive events that are posited to accelerate particles outside solar active regions and display only weak X-ray enhancements. This is in agreement with the fact that the associated flare magnitudes of the NP events are indeed in the lower end of the Flare-SEP magnitude distribution, as seen in Figure 9. Gopalswamy et al. (2016) reported on a SEP catalogue starting from 1997, listing eight SEPs indicating that they may be attributed to FEs. Of these eight, five are common with the events used in this work, and three of them are always predicted correctly. Of the other two, one is indeed an NP Event (start time 2011 November 26), and another is predicted correctly only 3 out of 10 times (start time 2004 April 11). This could, therefore, potentially offer a partial explanation for some of the NP events. Finally, we note that the divide in the flare magnitude distributions seen in Figure 9 points at initial misassociations of

Flare-SEP pairs as a probable cause for the existence of the NP events. Small magnitude flares are much more common and can occur in very close temporal proximity making the matching with a SEP that much more challenging. If that is the case, even a small number of corrected misassociations would have a non-negligible impact in improving the performance of the NNs given the otherwise low number of SEPs. The Flare-SEP associations may be revisited in future work with particular emphasis on the NP events.

As previously discussed, our methodology in the presented configuration uses the cosine and sine of the SF heliolongitude along with the 24 parameters extracted from the X-ray time series. It is worth noting that, of the 19 NP events, 11 occurred for eastern longitudes ($\theta < 0$) and 8 for western ($\theta > 0$), not showing any strong trend with regards to poor or stronger magnetic field connection. Additionally, while suitable for proof-of-concept, in an operational manner, the identification of the location of the flare on the solar disk, and therefore the knowledge of the heliolongitude, is not necessarily trivial. Therefore, we also tested our methodology without including the two heliolongitude parameters in the NN input. Interestingly the performance was very similar and only slightly reduced, 1–2% less for correct SEP prediction and ~1% less for correct true negative prediction. This is interesting in and of itself and merits future investigation as it seems to intrinsically incorporate that the
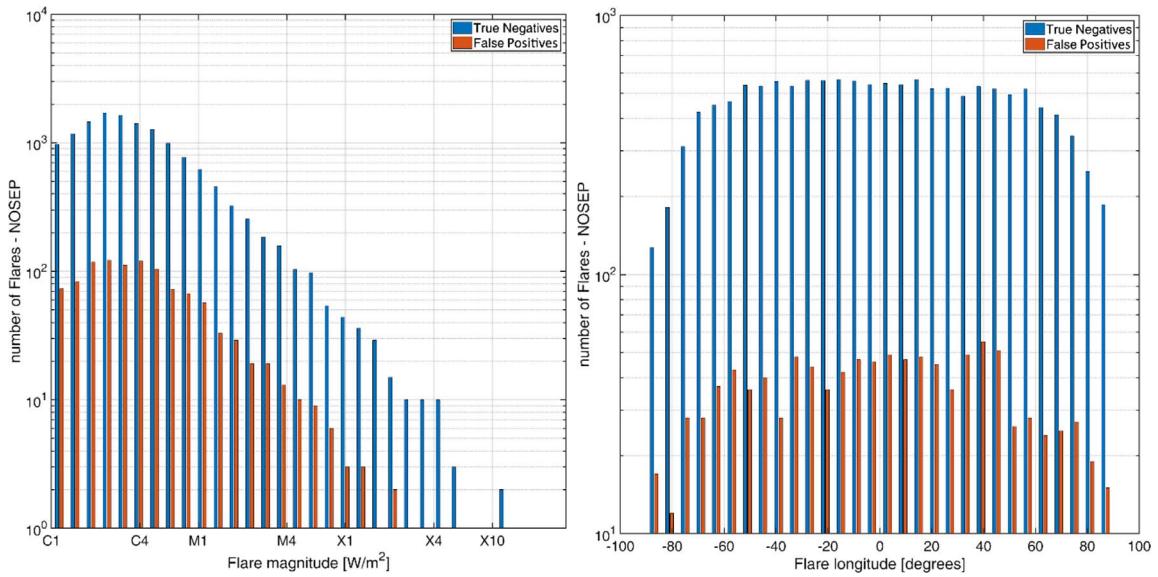
Fig. 10. The number of true negative flares and false positives. Left panel: in 30 logarithmic bins of the magnitude of their associated SFs. Right panel: in 30 bins of heliongitude.
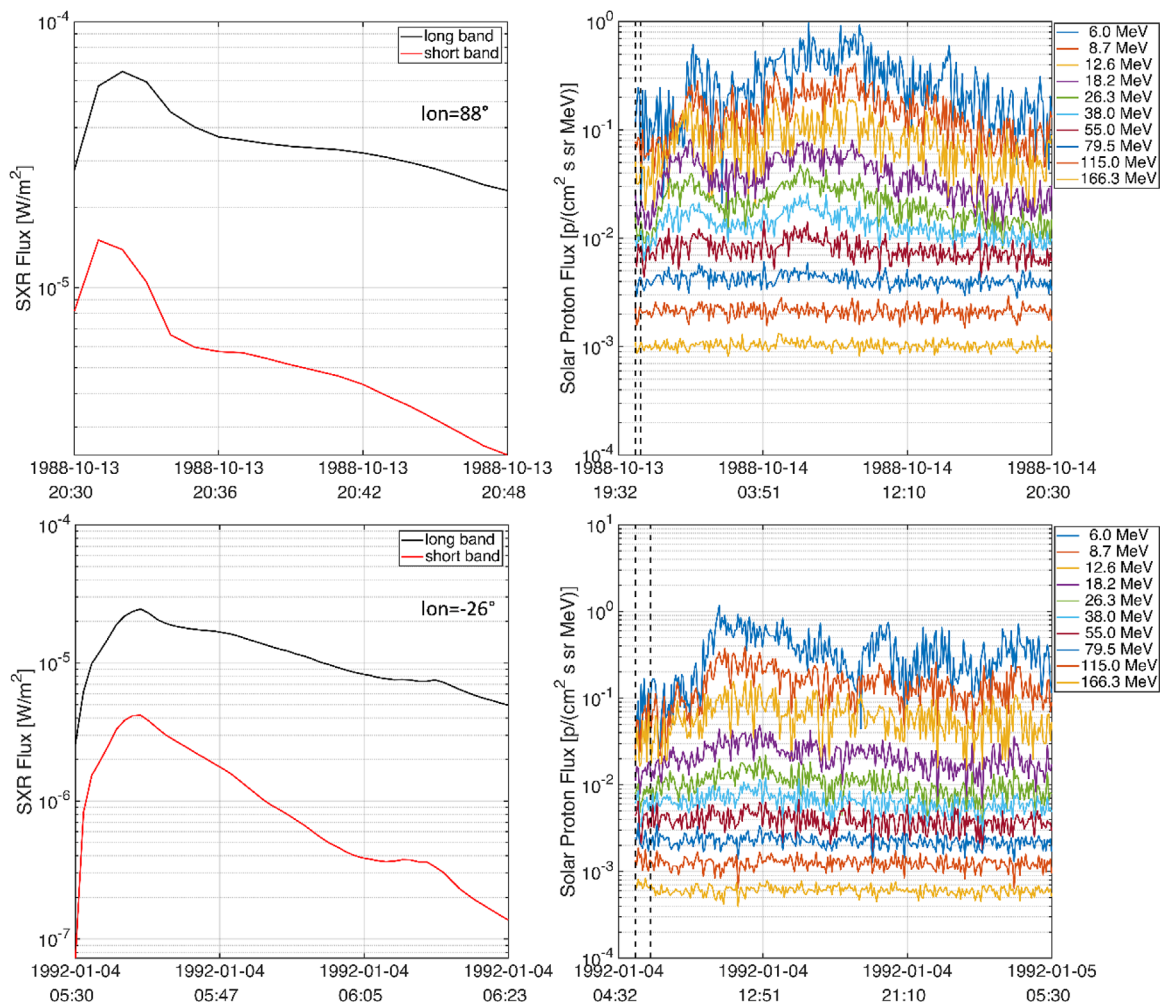

Fig. 11. Top row: on the left are SXR time-series from a western flare at 88° identified as a false positive, and on the right the solar proton flux time-series in the next 24 h after the start of the flare. The black dashed lines denote the time of the flare. Bottom row: similarly for an eastern flare at −26°.

heliolongitude becomes a more significant factor for SEP occurrence in 1 AU, mainly for weaker events.

Moreover, the false-positive results are also interesting. Figure 10 shows the distribution of flare magnitude and heliolongitude for the correct true negative and false positive predictions from one of the 100 runs in Section 3.1.

It can be seen that the distributions show no strongly discernible difference (apart from the absolute number) and can be considered effectively identical. We offer, therefore, a potential explanation that can account for some of the false positives. Recently, McComas et al. (2019) reported some of the first results from the Parker Solar Probe where it was shown that relatively weak energetic particle events occurred that were nevertheless not detected by any spacecraft at 1 AU; it was posited that such events might actually be much more common than previously thought. This prompted us to perform a preliminary investigation and test the hypothesis that some of the false-positive flares may be associated with small enhancements in the proton flux intensity levels that register but are far from being considered events by any criteria. Proton flux time-series from the SEPEM Reference Dataset (RDS) 2.0 (http://sepem.eu/help/SEPEM_RDS_v2-00.zip) were examined in a window of 24 h after the start time of false-positive flares while ensuring that they are temporally distant from the listed events. For several false positives, there are indeed weak proton enhancements occurring mainly at low energies. Figure 11 has two such examples for an eastern and a western flare showing the X-ray time-series and the solar protons time-series taken from SEPEM RDS2.0 in the next 24 h after the flares.

It can be seen that there is indeed a signal in the proton flux, which is elevated above the background, but that the enhancements are very weak. Of course, the temporal proximity is by no means conclusive and such associations would require thorough verification. However, it is a very interesting result that hints at the possibility that at least part of the false alarms in our results may have not been false alarms at all and that the model detected the X-ray signatures of such minor enhancements/SEPs.

5 Synopsis and conclusions

We have developed a methodology and a model for the prediction of the occurrence of SEPs. We have used solar flare SXR measurements in a methodology incorporating neural networks and demonstrated that our model has a very good performance being able to effectively predict the large majority of SEPs, crucially including the most intense ones, while also maintaining a very low false-positive rate. We have evaluated our results and validated them in a rigorous manner with strict criteria. The model was built, and the results were obtained using data that cover 25 years of solar activity, leaving little room for uncertainty regarding case-specific biases that can sometimes occur in limited datasets. Application-wise, the design and the way the model operates offer the capability to function in real-time with minimal pre-processing of data. Therefore, it could be easily applied either as a stand-alone module or in tandem with other methodologies in current and future space weather warning systems to provide independent outputs or in a co-advisory role to a final human operator, as is often operationally done. Additionally, the implementation presented uses state-of-the-art software tools that have been proven to be reliable and are being actively supported and developed, rendering the possibility of software obsoleteness very unlikely. Finally, the methodology is easily extendable and can be further refined in the future as new data, and scientific insights come from current and future missions like the Parker Solar Probe and the Solar Orbiter.

Supplementary materials

The supplementary materials of this article are available at https://www.swscjournal.org/10.1051/swsc/2021043/olm.

Table S1. The 26 parameters derived from the heliolongitude and the SXR time-series of each flare used as input in the neural network model.

Figure S1. One of the more typical NP Events with a minor enhancement in lower proton energies and mainly background noise in higher energies.

Figure S2. Similarly to S1, another typical NP Event with slightly higher intensities at low energies.

Figure S3. A medium intensity NP Event which exhibits enhancement up to the 166 MeV channel.

Figure S4. The most intense NP Event with the highest Peak Flux reaching ~300 p/(cm^2·s·sr·MeV) for the 6 MeV channel. However, higher energies are again mostly if not entirely background.

References

Abadi M, Barham P, Chen J, Chen Z, Davis A, et al. 2016. Tensorflow: A system for large-scale machine learning. In: Proceedings of the 12th USENIX conference on Operating Systems Design and Implementation (OSDI'16), USENIX Association, USA, pp. 265–283.

Alberti T, Laurenza M, Cliver EW, Storini M, Consolini G, Lepreti F. 2017. Solar Activity from 2006 to 2014 and Short-term Forecasts of Solar Proton Events Using the ESPERTA Model. *ApJ* **838**: 59. https://doi.org/10.3847/1538-4357/aa5cb8.

Aminalragia-Giamini S, Jiggens P, Anastasiadis A, Sandberg I, Aran A, et al. 2020. Prediction of Solar Proton Event Fluence spectra from their Peak flux spectra. *J Space Weather Space Clim* **10**: 1. https://doi.org/10.1051/swsc/2019043.

Anastasiadis A, Papaioannou A, Sandberg I, Georgoulis M, Tziotziou K, Kouloumvakos A, Jiggens P. 2017. Predicting Flares and Solar Energetic Particle Events: The FORSPEF Tool. *Sol Phys* **292**: 134. https://doi.org/10.1007/s11207-017-1163-7.

Anastasiadis A, Lario D, Papaioannou A, Kouloumvakos A, Vourlidas A. 2019. Solar energetic particles in the inner heliosphere: status and open questions. *Phil Trans R Soc A* **377**: 20180100. https://doi.org/10.1098/rsta.2018.0100.

Armstrong TP. 1976. *Handbook and reference manual for charged particle measurement experiment data from explorer 47 and 50.* Johns Hopkins Univ. Press, Laurel, MD.

Aschwanden MJ, Crosby NB, Dimitropoulou M, Georgoulis MK, Hergarten S, et al. 2016. 25 years of self-organized criticality: solar and astrophysics. *Space Sci Rev* **198**: 47–166. https://doi.org/10.1007/s11214-014-0054-6.

Azari AR, Lockhart JW, Liemohn MW, Jia X. 2020. Incorporating physical knowledge into machine learning for Planetary Space Physics. *Front Astron Space Sci* **7**: 36. https://doi.org/10.3389/fspas.2020.00036.

Bain HM, Steenburgh RA, Onsager TG, Stitely EM. 2021. A summary of National Oceanic and Atmospheric Administration Space Weather Prediction Center proton event forecast performance and skill. *Space Weather* **19**: e2020SW002670. https://doi.org/10.1029/2020SW002670.

Balasis G, Aminalragia-Giamini S, Papadimitriou C, Daglis IA, Anastasiadis A, Haagmans R. 2019. A machine learning approach for automated ULF wave recognition. *J Space Weather Space Clim* **9**: A13. https://doi.org/10.1051/swsc/2019010.

Balch CC. 2008. Updated verification of the Space Weather Prediction Center's solar energetic particle prediction model. *Space Weather* **6**: S01001. https://doi.org/10.1029/2007SW000337.

Belov A, Garcia H, Kurt V, Mavromichalaki E. 2005. Proton events and X-ray Flares in the last three solar cycles. *Cosmic Res* **43**: 165–178. https://doi.org/10.1007/s10604-005-0031-7.

Bloomfield DS, Higgins PA, McAteer RTJ, Gallagher PT. 2012. Toward reliable benchmarking of solar flare forecasting methods. *ApJ* **747**: 2. https://doi.org/10.1088/2041-8205/747/2/L41.

Brownlee J. 2020. *Imbalanced classification with python: better metrics, balance skewed classes, cost-sensitive learning, machine learning mastery.* Available online at: https://books.google.be/books?id=jaXJDwAAQBAJ.

Bobra MG, Couvidat S. 2015. Solar Flare Prediction using SDO/HMI vector magnetic field data with Machine-Learning algorithm. *ApJ* **798**: 2. https://doi.org/10.1088/0004-637X/798/2/135.

Camporeale E. 2019. The challenge of machine learning in Space Weather: Nowcasting and forecasting. *Space Weather* **17**: 1166–1207. https://doi.org/10.1029/2018SW002061.

Cane HV, Richardson IG, von Rosenvinge TT. 2010. A study of solar energetic particle events of 1997–2006: Their composition and associations. *J Geophys Res* **115**: A08101. https://doi.org/10.1029/2009JA014848.

Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. 2002. SMOTE: Synthetic minority over-sampling technique. *J Artif Intell Res* **16**: 321–357. doi: 10.1613/jair.953.

Chollet F. 2018. *Keras: The python deep learning library.* Astrophysics Source Code Library: ascl-1806.

Cliver EW, D'Huys E. 2018. Size distributions of solar proton events and their associated soft X-Ray flares: application of the maximum likelihood estimator. *ApJ* **864**: 48. https://doi.org/10.3847/1538-4357/aad043.

Crosby N, Heynderickx D, Jiggens P, Aran A, Sanahuja B, et al. 2015. SEPEM: a tool for statistical modeling the solar energetic particle environment. *Space Weather* **13**: 406–426. https://doi.org/10.1002/2013SW001008.

Dierckxsens M, Tziotziou K, Dalla S, Patsou I, Marsh MS, Crosby NB, Malandraki O, Tsiropoula G. 2015. Relationship between solar energetic particles and properties of flares and CMEs: Statistical analysis of solar cycle 23 events. *Sol. Phys.* **290**: 841–874. https://doi.org/10.1007/s11207-014-0641-4.

Garcia HA. 1994. Temperature and emission measure from goes soft X-ray measurements. *Sol Phys* **154**: 275–308. https://doi.org/10.1007/BF00681100.

GOES I-M DataBook. 1996. *Prepared for NASA DRL 101–08, GSFC Specification S-480-21A, Contract NAS5-29500, Reference S-415-19, Revision 1.* Greenbelt, MD.

Gold RE, Krimigis SM, Hawkins SE III, Haggerty DK, Lohr DA, Fiore E, Armstrong TP, Holland G, Lanzerotti LJ. 1998. Electron, proton, and alpha monitor on the advanced composition explorer spacecraft. *Space Sci Rev* **86**: 541–562. https://doi.org/10.1007/978-94-011-4762-0_19.

Goodfellow I, Bengio Y, Courville A. 2016. *Deep learning*, Vol **1, No. 2**, MIT Press, Cambridge. ISBN: 978-0262035613

Gopalswamy N, Tsurutani B, Yan Y. 2015. Short-term variability of the Sun-Earth system: an overview of progress made during the CAWSES-II period. *Prog Earth Planet Sci* **2**: 13. https://doi.org/10.1186/s40645-015-0043-8.

Gopalswamy N, Yashiro S, Thakur N, Mäkelä P, Xie H, Akiyama S. 2016. The 2012 July 23 backsed eruption: An extreme energetic particle event? *ApJ* **833**: 2. https://doi.org/10.3847/1538-4357/833/2/216.

Grayson JA, Krucker S, Lin RP. 2009. A Statistical study of spectral hardening in solar flares and related solar energetic particle events. *ApJ* **707**: 1588. doi: 10.1088/0004-637X/707/2/1588.

Hu S, Kim M-HY, McClellan GE, Cucinotta FA. 2009. Modeling the acute health effects of astronauts from exposure to large solar particle events. *Health Phys* **96(4)**: 465–476. https://doi.org/10.1097/01.HP.0000339020.92837.61.

Ioffe S, Szegedy C. 2015. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In: *Proceedings of the 32nd International Conference on Machine Learning*, Vol. **37**, pp. 448–456.

Jiggens PTA, Gabriel SB. 2009. Time distributions of solar energetic particle events: Are SEPEs really random? *J Geophys Res* **114**: A10105. https://doi.org/10.1029/2009JA014291.

Jiggens P, Heynderickx D, Sandberg I, Truscott P, Raukunen O, Vainio R. 2018. Updated model of the solar energetic proton environment in space. *J Space Weather Space Clim* **8**: A31. https://doi.org/10.1051/swsc/2018010.

Kahler SW, Cliver EW, Ling AG. 2007. Validating the proton prediction system (PPS). *J Atmos Sol-Terr Phys* **69**: 1–2. https://doi.org/10.1016/j.jastp.2006.06.009.

Kahler SW, Ling AG. 2018. Forecasting Solar Energetic Particle (SEP) events with Flare X-ray peak ratios. *J Space Weather Space Clim* **8**: A47. https://doi.org/10.1051/swsc/2018033.

Klein KL, Dalla S. 2017. Acceleration and propagation of solar energetic particles. *Space Sci Rev* **212**: 1107–1136. https://doi.org/10.1007/s11214-017-0382-4.

Lario D, Decker RB, Armstrong TP. 2001. Major solar proton events observed by IMP-8 (from November 1973 to May 2001). *ICRC Proc* **3254**: 1–4.

Lario D, Aran A, Gómez-Herrero R, Dresing N, Heber B, Ho GC, Decker RB, Roelof EC. 2013. Longitudinal and radial dependence of solar energetic particle peak intensities: STEREO, ACE, SOHO, GOES, and MESSENGER observations. *ApJ* **767**: 41. https://doi.org/10.1088/0004-637X/767/1/41.

Laurenza M, Cliver EW, Hewitt J, Storini M, Ling AG, Balch CC, Kaiser ML. 2009. 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**: S04008. https://doi.org/10.1029/2007SW000379.

Laurenza M, Alberti T, Cliver EW. 2018. A Short-term (ESPERTA)-based Forecast Tool for Moderate-to-extreme Solar Proton Events. *ApJ* **857**: 107. https://doi.org/10.3847/1538-4357/aab712.

Lemaître G, Nogueira F, Aridas CK. 2017. Imbalanced-learn: a python toolbox to tackle the curse of imbalanced datasets in machine learning. *J Mach Learn Res* **18**: 559–563Available online at: http://jmlr.csail.mit.edu/papers/v18/16-365.html.

Maas AL, Hannun AY, Ng AY. 2013. Rectifier nonlinearities improve neural network acoustic models. *Proc ICML* **30**, 1.

McComas DJ, Christian ER, Cohen CMS, Cummings AC, Davis AJ, et al. 2019. Probing the energetic particle environment near the Sun. *Nature* **576**: 223–227. https://doi.org/10.1038/s41586-019-1811-1.

Nita G, Georgoulis M, Kitiashvili I, Sadykov V, Camporeale E, et al. 2020. *Machine learning in heliophysics and space weather forecasting: a white paper of findings and recommendations.* arXiv:2006.12224v1 [astro-ph.SR].

Núñez M. 2011. Predicting solar energetic proton events ($E > 10$ MeV). *Space Weather* **9**: S07003. https://doi.org/10.1029/2010SW000640.

Núñez M, Daniel P-P. 2020. Predicting >10 MeV SEP events from solar flare and radio burst data. *Universe* **6(10)**: 161. https://doi.org/10.3390/universe6100161.

Pacheco D. 2019. Analysis and modelling of the solar energetic particle radiation environment in the inner heliosphere in preparation for solar orbiter, *Ph.D. Thesis*, Dep. Física Quàntica i Astrofísica, Universitat de Barcelona, Barcelona, Spainhttp://hdl.handle.net/10803/667033.

Papaioannou A, Sandberg I, Anastasiadis A, Kouloumvakos A, Georgoulis MK, Tziotziou K, Tsiropoula G, Jiggens P, Hilgers A. 2016. Solar flares, coronal mass ejections and solar energetic particle event characteristics. *J Space Weather Space Clim* **6**: A42. https://doi.org/10.1051/swsc/2016035.

Papaioannou A, Anastasiadis A, Kouloumvakos A, Paassilta M, Vainio R, et al. 2018. Nowcasting solar energetic particle events using principal component analysis. *Sol Phys* **293**: 100. https://doi.org/10.1007/s11207-018-1320-7.

Posner A. 2007. Up to 1-hour forecasting of radiation hazards from solar energetic ion events with relativistic electrons. *Space Weather* **5**: S05001. https://doi.org/10.1029/2006SW000268.

Raptis S, Aminalragia-Giamini S, Karlsson T, Lindberg M. 2020. Classification of magnetosheath jets using neural networks and High Resolution OMNI (HRO) Data. *Front Astron Space Sci* **7**: 24. https://doi.org/10.3389/fspas.2020.00024.

Reames DV. 2020. Four distinct pathways to the element abundances in solar energetic particles. *Space Sci Rev* **216**: 20. https://doi.org/10.1007/s11214-020-0643-5.

Richardson I, Rosenvinge T, Cane H, Christian E, Cohen C, Labrador A, Leske R, Mewaldt R, Wiedenbeck M, Stone E. 2014. > 25 MeV proton events observed by the high energy telescopes on the STEREO A and B spacecraft and/or at Earth during the first ~seven years of the STEREO mission. *Sol Phys* **289**: 3059–3107. https://doi.org/10.1007/s11207-014-0524-8.

Steyn R, Strauss DT, Effenberger F, Pacheco D. 2020. The soft X-ray Neupert effect as a proxy for solar energetic particle injection - A proof-of-concept physics-based forecasting model. *J Space Weather Space Clim* **10**: 64. https://doi.org/10.1051/swsc/2020067.

Winter LM, Ledbetter K. 2015. Type II and Type III Radio Bursts and their correlation with Solar Energetic Proton Events. *ApJ* **809**: 105. https://doi.org/10.1088/0004-637X/809/1/105.

Xapsos MA, Stauffer C, Barth JL, Burke EA. 2006. Solar particle events and self-organized criticality: Are deterministic predictions of events possible? *IEEE Trans Nuc Sci* **53**: 4. https://doi.org/10.1109/TNS.2006.880576.

Youssef M. 2012. On the relation between the CMEs and the solar flares. *NRIAG J Astron Geophys* **1(2)**: 172–178. https://doi.org/10.1016/j.nrjag.2012.12.014.