

## AI-FLARES

### **WP5100-D1 (month 24): Report on the results of the validation process for the software tools developed within WP2100, WP3100, WP4100**

**For details, see the following AI-FLARES papers:**

- Krucker S, Hurford GJ, Grimm O, Kögl S, Gröbelbauer HP, Etesi L, Casadei D, Csillaghy A, Benz AO, Arnold NG, Molendini F. The spectrometer/telescope for imaging X-rays (STIX). *Astronomy & Astrophysics*. 2020 Oct 1;642:A15.
- Guastavino S, Benvenuto F. A mathematical model for image saturation with an application to the restoration of solar images via adaptive sparse deconvolution. *Inverse Problems*. 2020 Dec 15;37(1):015010
- Guastavino S, Piana M, Massone AM, Schwartz R, Benvenuto F. Desaturating SDO/AIA observations of solar flaring storms. *The Astrophysical Journal*. 2019 Sep 9;882(2):109
- Massa P, Hurford G, Volpara A, Kuhar M, Battaglia AF, Xiao H, Casadei D, Perracchione E, Garbarino S, Guastavino S, Collier H. STIX imaging I--Concept. *arXiv preprint arXiv:2303.02485*. 2023 Mar 4
- Guastavino S, Marchetti F, Benvenuto F, Campi C, Piana M. Operational solar flare forecasting via video-based deep learning. *arXiv preprint arXiv:2209.05128*. 2022 Sep 12.
- Guastavino S, Marchetti F, Benvenuto F, Campi C, Piana M. Implementation paradigm for supervised flare forecasting studies: A deep learning application with video data. *Astronomy & Astrophysics*. 2022 Jun 1;662:A105.
- Georgoulis MK, Bloomfield DS, Piana M, Massone AM, Soldati M, Gallagher PT, Pariat E, Vilmer N, Buchlin E, Baudin F, Csillaghy A. The flare likelihood and region eruption forecasting (FLARECAST) project: flare forecasting in the big data & machine learning era. *Journal of Space Weather and Space Climate*. 2021;11:39.
- Cicogna D, Berrilli F, Calchetti D, Del Moro D, Giovannelli L, Benvenuto F, Campi C, Guastavino S, Piana M. Flare-forecasting algorithms based on high-gradient polarity inversion lines in active regions. *The Astrophysical Journal*. 2021 Jul 1;915(1):38.
- Benvenuto F, Campi C, Massone AM, Piana M. Machine learning as a flaring storm warning machine: Was a warning machine for the 2017 September solar flaring storm possible? *The Astrophysical Journal Letters*. 2020 Nov 19;904(1):L7.

### **Section 1 - Validation process for the software tools developed within WP2100**

AI-FLARES contribution to image processing at the solar EUV regime is represented by the formulation and implementation of the Sparsity-Enhancing DESAT (SE-DESAT) method, a novel computational approach for the analysis of SDO/AIA saturated images able to recover the signal in the primary saturation region in a rapid fashion without using any other information but the one contained in the image itself. SE-DESAT is a modification of a previous algorithm developed in our group, named DESAT. As for DESAT, also in SE-DESAT the input data are represented by the diffraction fringes and therefore this is again an inverse diffraction algorithm. However, unlike for DESAT, this new approach realizes segmentation between the primary saturation region and blooming, background estimation and desaturation in the primary saturation region at the same time, without the need of any a priori information on the image background.

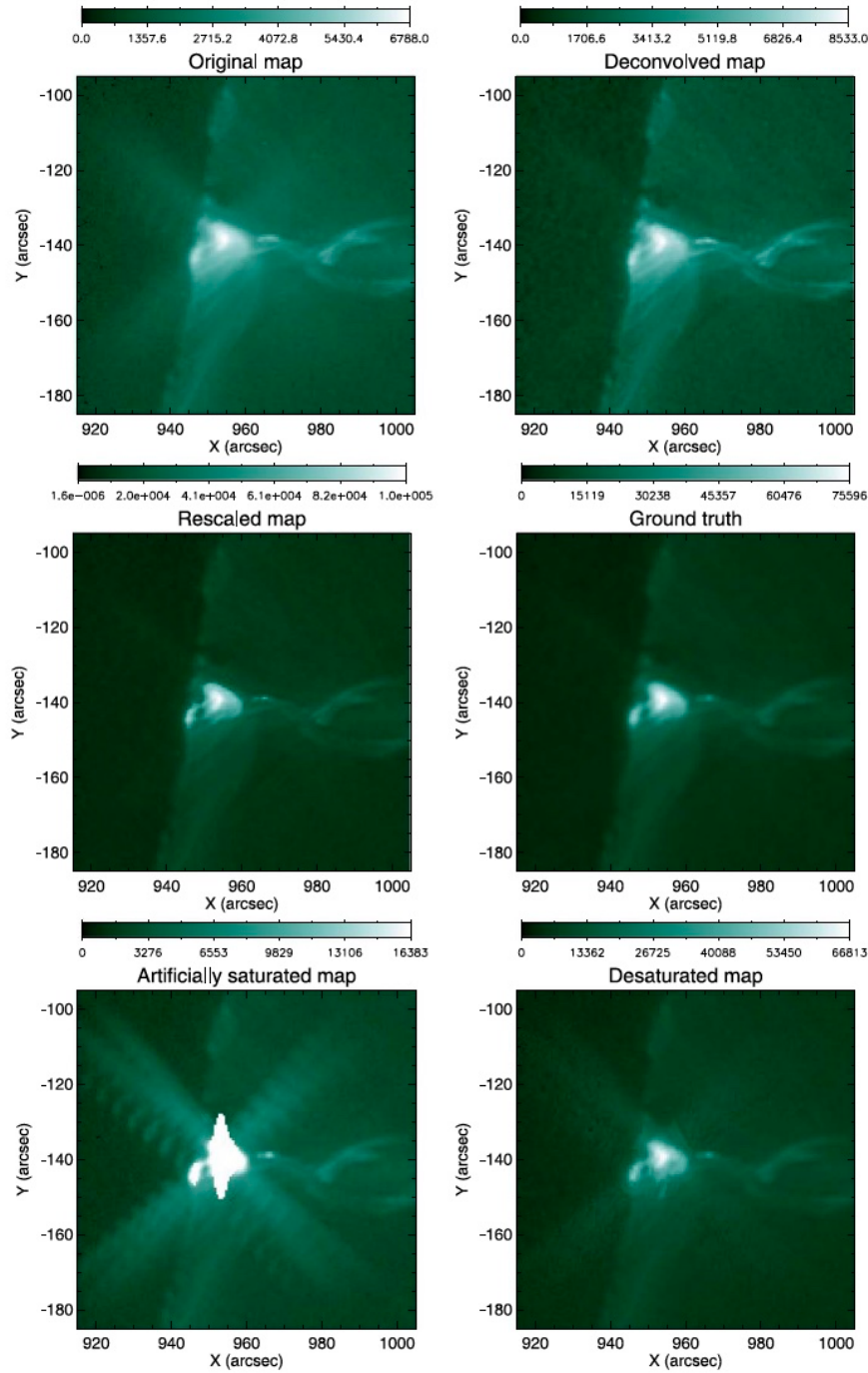
The validation process of SE-DESAT relied on two approaches, one via comparison with synthetic ground-truth, and the other one via comparison with the performance of DESAT in the case of experimental data.

Simulated data. We realized the simulation of a synthetic AIA image affected by both primary and secondary saturation according to a scheme relying on the approach followed by Torre et al. (2015). Specifically:

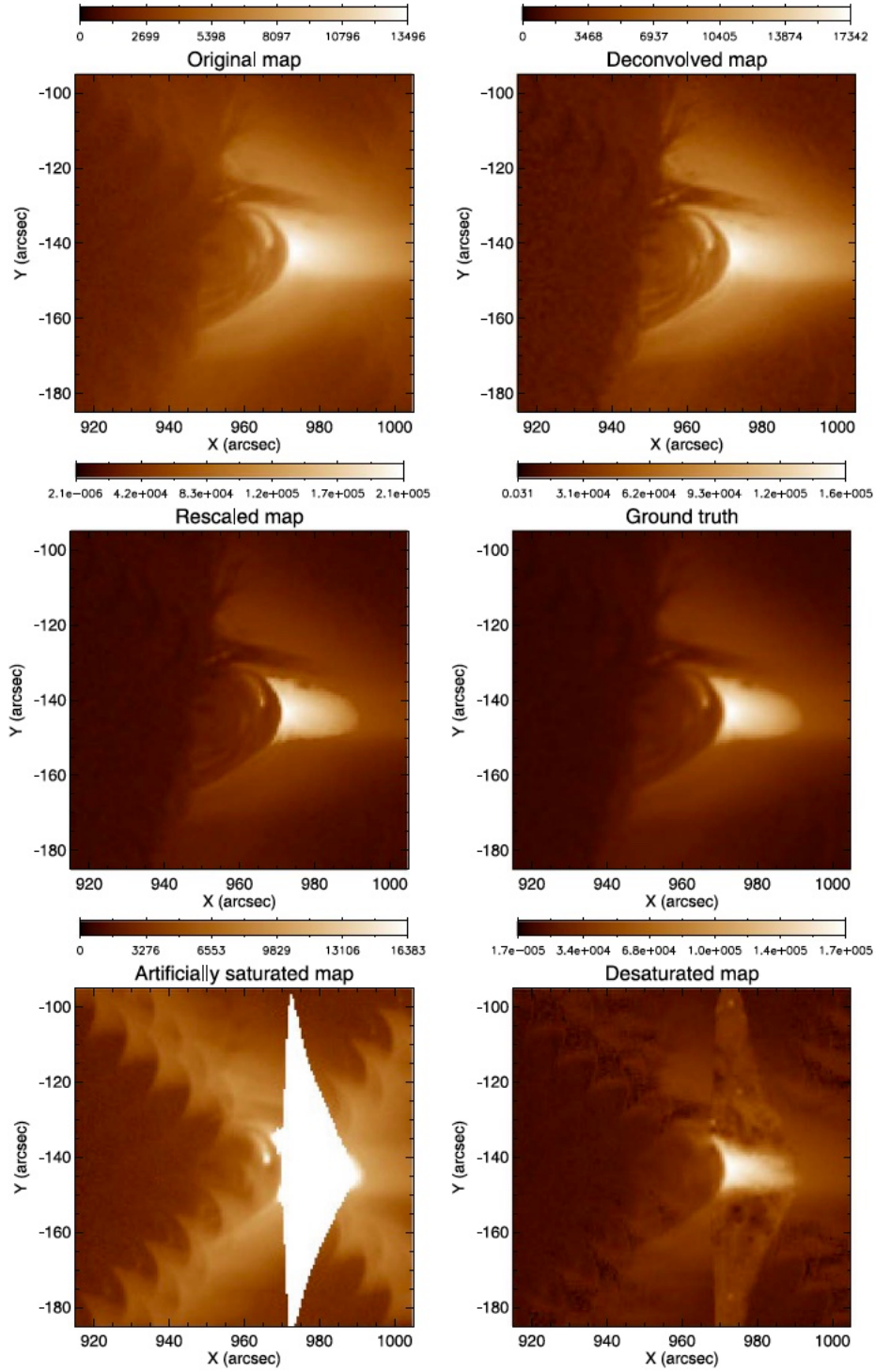
1. The original unsaturated image is deconvolved by means of EM using the global AIA point spread function (PSF).
2. The deconvolved image is rescaled in order to lift the most intense pixels over the threshold value.
3. The ground-truth image is constructed as the result of the application of the diffusion core PSF on the deconvolved and rescaled image.
4. The deconvolved and rescaled image is also convolved by means of the global AIA PSF, and the result is perturbed with Poisson noise.
5. We saturated this latter image by:
  - a. Setting the over-threshold values at the threshold values, thus producing the primary saturation effect.
  - b. Simulating the blooming effect by assuming that the spill-over charge expands column-wise in a symmetric way.

The resulting saturated image is given as input to our desaturation algorithm, with results illustrated in Figure 1 and Figure 2. Specifically, Figure 1 corresponds to the case where the original unsaturated image is recorded by the 94 Å filter and the saturated image resembles the case of a mild saturation effect. Figure 2 starts from an image recorded by the 193 Å filter and mimics the case of intense blooming. The quantitative validation of the results provided by our algorithm is illustrated in Table 1, which reports the relative reconstruction error in the primary saturation region, the C-statistic in the diffraction fringes, and the relative error in the determination of the number of bloomed and primarily saturated pixels.

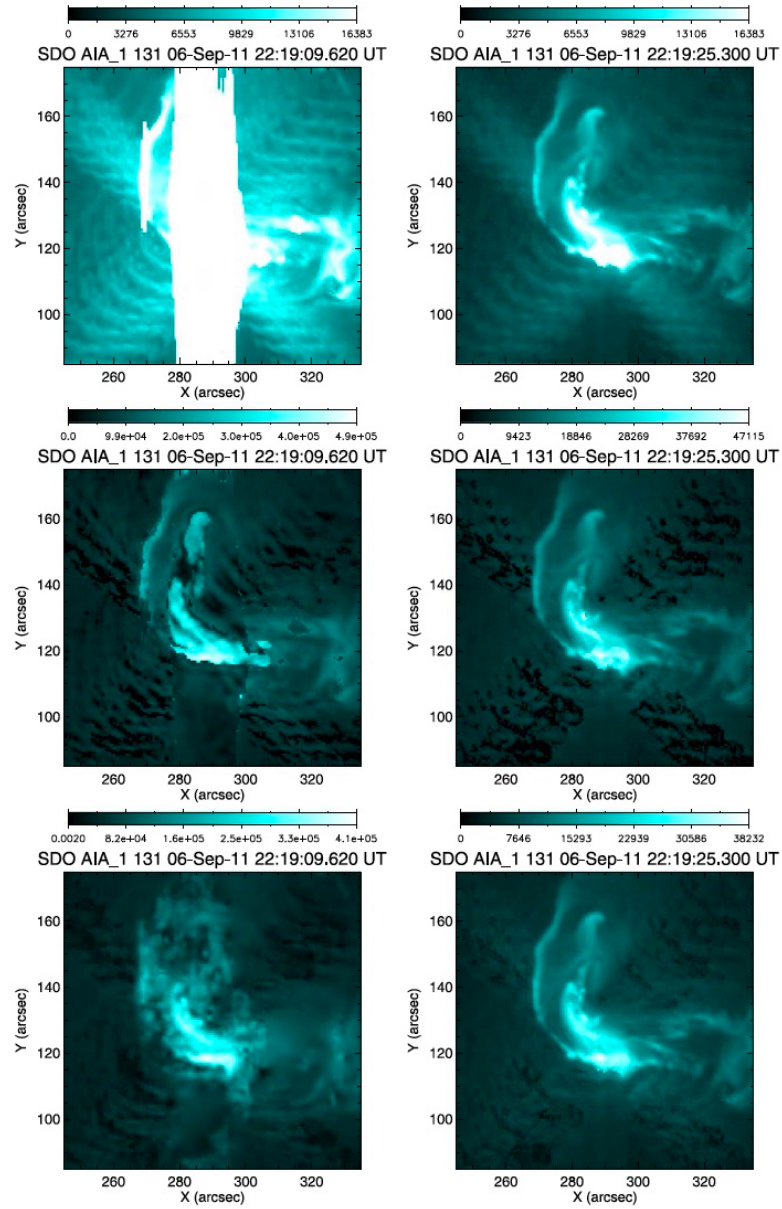
Experimental data. DESAT has been conceived to perform desaturation when a reliable estimate of the background is available. In that case, a correlation analysis combined with the use of Expectation Maximization (EM) allows the solution of the AIA inverse diffraction problem and therefore the signal restoration in the primary saturation region. The most straightforward way to obtain an estimate of the background is probably applying interpolation between the frame immediately before and the frame immediately after the saturated one. This limits the applicability of DESAT just to cases when saturation affects a single frame at a specific time, while the previous and following images are unsaturated. This is the case, for example, of the 2011 September 6 event described in Figures 3 and 4 for the 131 Å bandwidth, at time intervals 22:19:09 UT and 22:19:25 UT. These results compare the behavior of SE-DESAT with respect to DESAT. In particular, Figure 3 contains the desaturated images, while Figure 4 compares the profiles of the desaturated flux integrated along the image columns. In terms of C-statistic, SE-DESAT achieves 0.84 for the 22:19:09 UT frame and 0.63 for the 22:19:25 UT frame; DESAT achieves 14.03 for the 22:19:09 UT frame and 1.97 for the 22:19:25 UT frame (Torre et al. 2015).



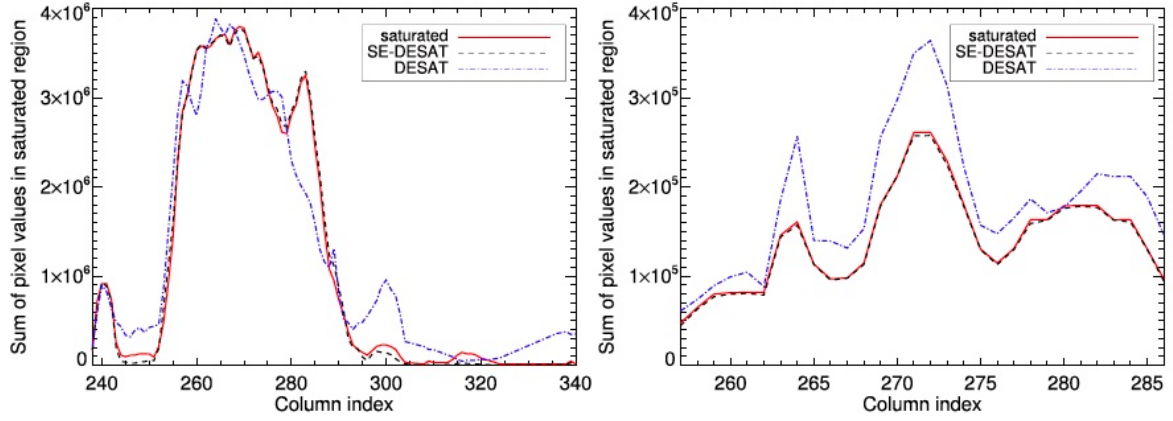
**Figure 1: Simulation study. First row: original image (2017 September 10 at the acquisition time 15:48:11 UT at the waveband  $94 \text{ \AA}$ ) and deconvolved image. Second row: rescaled image and the ground truth (i.e., the result of the convolution between the rescaled map and the core component of the AIA PSF). Third row: saturated image with Poisson noise added and the reconstruction obtained by SE-DESAT.**



**Figure 2. Simulation study. First row: original image (2017 September 10 at the acquisition time 15:48:11 UT at 193 Å) and deconvolved image. Second row: rescaled image and ground truth. Third row: saturated image with Poisson noise added and the reconstruction obtained by SE-DESAT.**



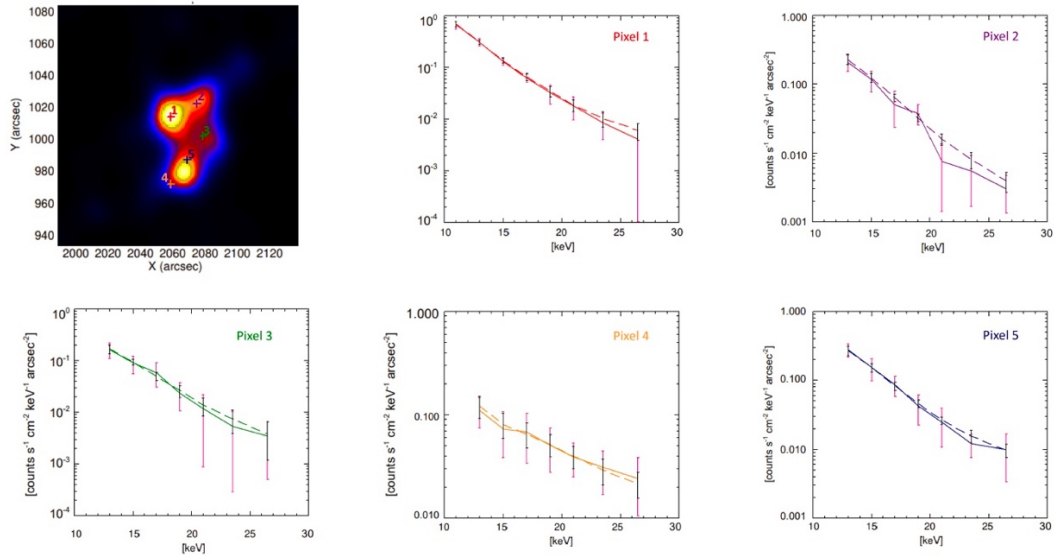
**Figure 3. Comparison with DESAT in the case of the 2011 September 6 event acquired at 131 Å. First column, from top to bottom: the saturated 22:19:09 UT frame; DESAT desaturation; SE-DESAT desaturation. Second column, from top to bottom: the saturated 22:19:25 UT frame; DESAT desaturation; SE-DESAT desaturation.**



**Figure 4. Comparison with DESAT in the case of the 2011 September 6 event at 131 Å: profiles of the flux integrated along the desaturated image columns for the 22:19:09 UT and 22:19:25 frames.**

## Section 2 - Validation process for the software tools developed within WP3100

In order to assess the reliability of regularized imaging spectroscopy, we studied both regularized spatially-resolved and regularized spatially-integrated spectra obtained from the AI-FLARES methodology. Figure 5 considers the September 29 2022 event and, in the five highlighted locations, compares the count spectra provided by the original not regularized images with the ones provided by the regularized count images. Regularization does not modify the overall shape of the spectra, while it significantly increases their stability.



**Figure 5. Top left panel: selected pixels. Other panels: regularized and not-regularized local count spectra corresponding to the five selected pixels.**

Table 1 compares the spectral indices provided by spatially integrating the count images, the electron flux spectral images, and the regularized count images with the ones provided by OSPEX using standard spectroscopy. We considered five events characterized by similar morphologies. The entries in the table show that 1) the count spectral indices generated by the regularized count images reproduce the ones provided by OSPEX; and 2) the approximate thin-target relation  $\gamma = \delta + 1$  are both well satisfied by the regularized electron flux and count spectra.



Event	OSPEX	electron flux spectral images	count images	regularized count images
May 08, 2021	$\gamma = 5.28 \pm 0.18$	$\delta = 4.53 \pm 0.07$	$\gamma = 4.94 \pm 0.24$	$\gamma = 5.50 \pm 0.04$
August 26, 2021	$\gamma = 5.65 \pm 0.12$	$\delta = 4.59 \pm 0.11$	$\gamma = 5.27 \pm 0.17$	$\gamma = 5.57 \pm 0.06$
January 20, 2022	$\gamma = 6.51 \pm 0.26$	$\delta = 5.07 \pm 0.03$	$\gamma = 6.25 \pm 0.14$	$\gamma = 6.36 \pm 0.01$
August 28, 2022	$\gamma = 6.74 \pm 0.18$	$\delta = 4.97 \pm 0.04$	$\gamma = 6.88 \pm 0.24$	$\gamma = 6.81 \pm 0.03$
September 29, 2022	$\gamma = 4.54 \pm 0.05$	$\delta = 3.68 \pm 0.01$	$\gamma = 4.24 \pm 0.11$	$\gamma = 4.42 \pm 0.01$

**Table 1. First column: event date. Second column: photon power-law index  $\gamma$  provided by OSPEX. Third column: electron power-law index  $\delta$  obtained by fitting spatially integrated regularized electron flux spectral images. Fourth and fifth columns: photon power-law index  $\gamma$  obtained by fitting spatially integrated count images and spatially integrated regularized count images, respectively.**

### Section 3 - Validation process for the software tools developed within WP4100

AI-FLARES efforts concerning flare forecasting included two aspects: the design and implementation of machine learning algorithms for the interpretation of features extracted from full-disk magnetograms, and deep learning networks that takes magnetograms' images and videos as input.

Validation of feature-based machine learning algorithms. AI-FLARES focused on the role of two novel topological descriptors in flare forecasting: an appropriate generalization of the morphological metric R, originally proposed by Schrijver (2007); and a new topology-based metric D related to the magnetic complexity of active regions (ARs) (essentially, this counts the number of polarity inversion lines above an appropriate magnetic field threshold). The machine learning approach for the validation of the role of these two descriptors in flare forecasting relied on the following experiment exploiting the SDO/HMI archive in the time range between 2012 September 14 and 2016 April 30.

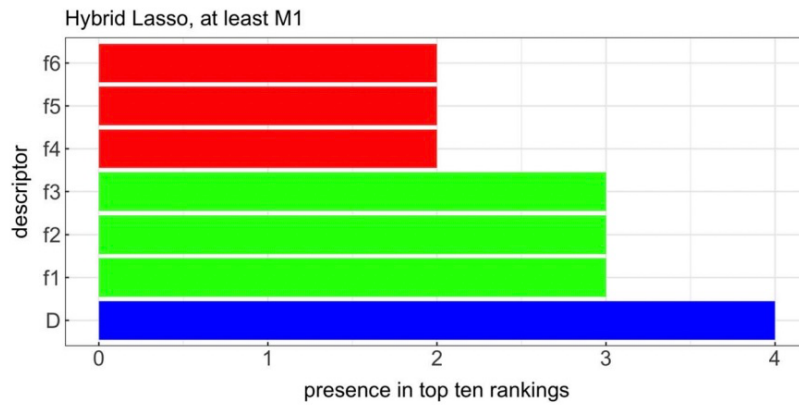
- The corresponding magnetograms have been grouped into four subsets belonging to the four issuing times 00:00, 06:00, 12:00, and 18:00 UT. Of course, it may happen that an AR lasted more than 1 day: in that case, each one of the four subsets will contain as many samples associated to that AR, as the days the AR lasts.
- For each subset, i.e., for each issuing time, we randomly extracted two-thirds of ARs and constructed the training, while the remaining one-third was used to populate the corresponding test set.
- Each sample in each training set has been labelled by using GOES data (label “1” corresponds to an event occurrence within the 24 hour from the issuing time)

This process was repeated 100 times to obtain 100 realizations of the training set and 100 realization of the test set for each issuing time. From each AR we extracted D and log R and fed a hybrid version of LASSO with the corresponding two-dimensional feature vectors. In order to assess the forecasting performances we computed TPR, TSS, and HSS values and obtained the results in Table 2.

Issuing Time	Test Set—at Least M1 Flares		
	$\log R^*$ & $D$		
	TPR	TSS	HSS
00:00:00UT	$0.47 \pm 0.13$	$0.40 \pm 0.12$	$0.22 \pm 0.06$
06:00:00UT	$0.54 \pm 0.09$	$0.47 \pm 0.08$	$0.31 \pm 0.05$
12:00:00UT	$0.49 \pm 0.09$	$0.44 \pm 0.08$	$0.33 \pm 0.05$
18:00:00UT	$0.43 \pm 0.04$	$0.38 \pm 0.04$	$0.30 \pm 0.03$

**Table 2. Average TPR, TSS, and HSS values associated to the outcomes of Hybrid Lasso (100 random realizations of the training/test sets at four specific issuing times)**

The score values automatically obtained in this table are comparable with the ones obtained by heuristically tuning the thresholds on  $\log R$  and  $D$ . We also quantified the impact on the forecasting performance of the two descriptors when used in combination with many more predictors. We used the 171 features extracted within the Horizon 2020 FLARECAST project and arranged them again in the four training sets corresponding to the four issuing times (00:00, 06:00, 12:00, and 18:00 UT); then, we applied again hybrid Lasso in order to both realize binary prediction within the next 24 hr for each issuing time and identify the features that mostly impact such prediction; and we did this same experiment by including also the descriptors  $R^*$  and  $D$  for the prediction of M1 or more intense events within the next 24 hr. The results of binary classification are shown in Table 3.



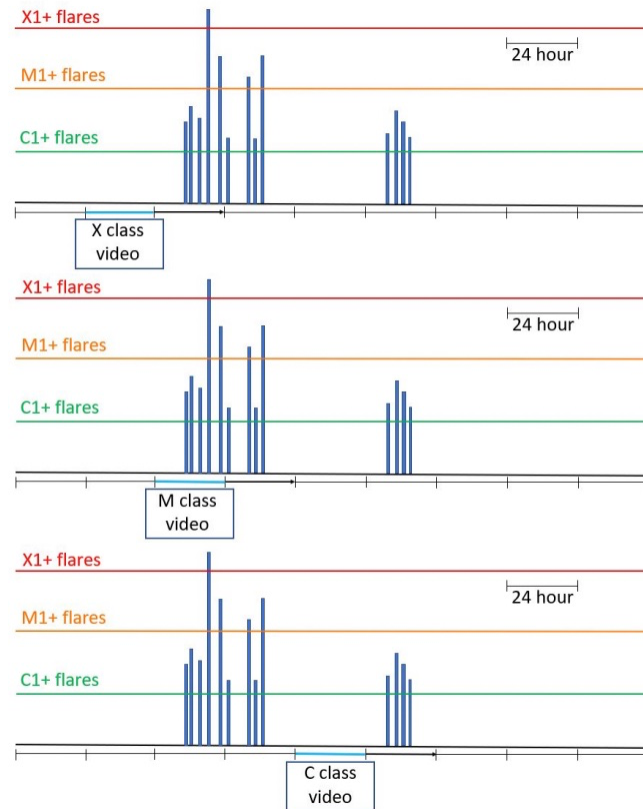
**Figure 6. Histograms counting the number of times predictors are in the top 10 rankings, on average over the 100 random realizations of the training and test set; f1-f6 are six FLARECAST features.**

Comparing Tables 2 and 3 shows that the use of more descriptors allows machine learning to produce higher scores; however, Table 3 also suggests that the use of the two new descriptors does not significantly increase the score values. The results of the feature ranking analysis illustrated in Figure 6 shows that the descriptor  $D$  plays a significant role in the realization of the prediction performances.



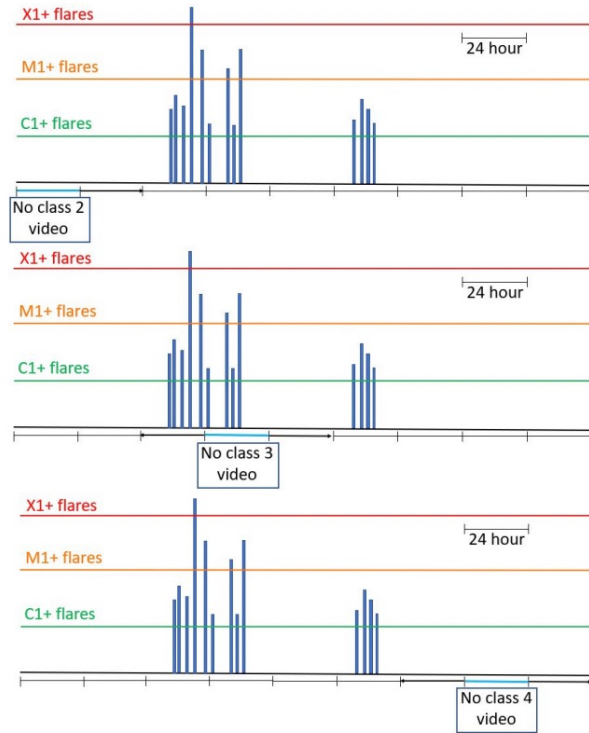
Validation of deep learning. We considered the AI-FLARES Long-term Recurrent Convolutional Network (LRCN) and validated it against the Near Realtime Space Weather HMI Archive Patch (SHARP) data products associated with the line-of-sight components in the time range between 2012 September 14 and 2017 September 30. More specifically, we used 24-h-long videos made of 40 SHARP images of an AR, with 36 minutes cadence.

Figures 7 and 8 are iconographical representations of how these videos are categorized, where Figure 7 represents the C- M- and X-class flares and Figure 8 contains three possible types of null events (NO1 corresponds to ARs that never originated a flare and are not included in the figure).



**Figure 7. C-, M-, and X-class events**

We generated a training set, a validation set, and a test set made of 3000, 750, and 750 data samples, respectively, and we used data augmentation to increase the cardinality of these sets up to 15 000, 3750, and 3750, respectively. We repeated this set generation process ten times in order to create ten random realizations of these three sets. Table 3 shows the prediction results obtained by the LRCN in the case of the realization of the validation and test sets.



**Figure 8. Three types of null events**

TSS (C1+ flares)							
	Mean	Std	Min	25th perc	Median	75th perc	Max
Validation test	0.57	0.02	0.55	0.56	0.57	0.59	0.61
	0.55	0.05	0.46	0.52	0.54	0.60	0.61
TSS (M1+ flares)							
	Mean	Std	Min	25th perc	Median	75th perc	Max
Validation test	0.76	0.07	0.65	0.67	0.77	0.82	0.85
	0.68	0.09	0.55	0.61	0.69	0.72	0.82

**Table 3. Outcomes of AI-FLARES LRCN: TSS values, the 25th percentile values, the median values, the 75th percentile values and the maximum value for the prediction of both C1+ and M1+ flares**

Both the mean and the median values for the test sets are close to the ones provided by the network when applied to the validation sets, and in all cases the standard deviations are nicely small. either M or X are more distinguishable from ARs associated with NO-class flares (this is particularly true for all cases where flares are close to strong B events)