toward standardized machine learning for flare forecasting and an application with a video-based convolutional neural network

michele piana

the MIDA group

dipartimento di matematica, università di genova

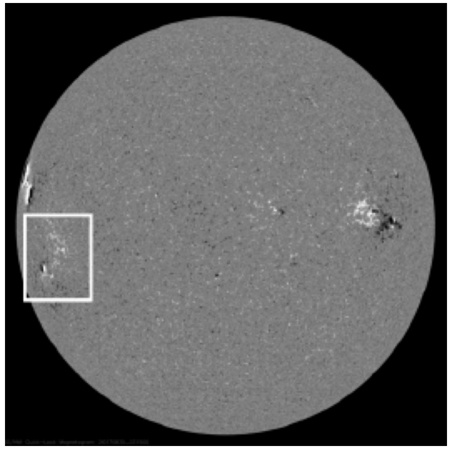
CNR – SPIN genova

17th european space weather week
25-29 october 2021

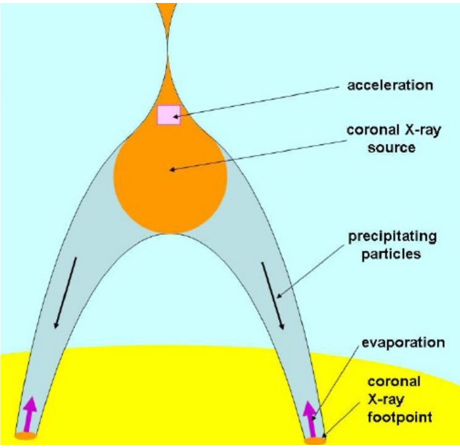
credits

- sabrina guastavino (dipartimento di matematica, università di genova)
- francesco marchetti (dipartimento di matematica, università di padova)
- federico benvenuto (dipartimento di matematica, università di genova)
- cristina campi (dipartimento di matematica, università di genova)

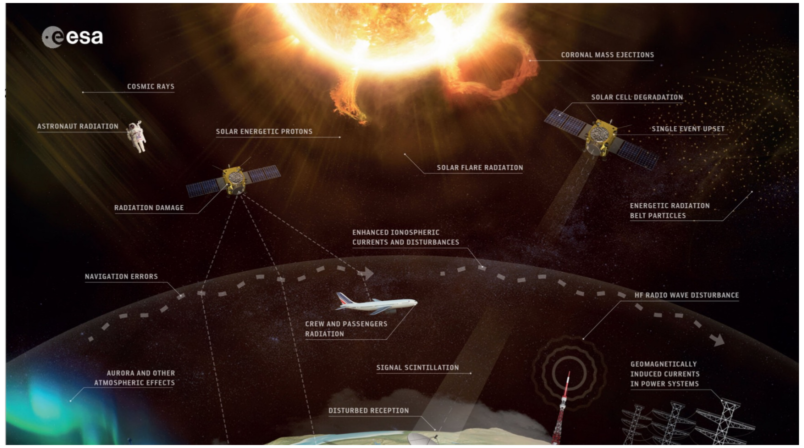
solar flares



active regions



standard model



impact on space weather

- each flare is originated by an active region (AR)
- not all ARs originate a flare

the flare forecasting problem is the one to predict whether an AR will produce a flare within a given time interval

- data: SDO/HMI provides AR images (magnetograms) every 12 minutes since february 2010
- labels: GOES detects solar flares and provides them with an energetic class (C, M, X)

model-driven approaches

MHD equations:

$$\frac{\partial \rho}{\partial t} + \nabla(\rho \mathbf{V}) = 0,$$

$$\frac{\mathrm{d}}{\mathrm{d}\,t}\left(\frac{P}{\rho^{\gamma}}\right) = 0,$$

$$\rho \frac{\mathrm{d}\,\mathbf{V}}{\mathrm{d}\,t} = -\nabla P - \frac{1}{\mu_0}\mathbf{B} \times (\nabla \times \mathbf{B}),$$

$$\frac{\partial \mathbf{B}}{\partial t} = \nabla \times (\mathbf{V} \times \mathbf{B}),$$

+

numerical methods

for the solution of

partial differential equations

data-driven approaches

$min_w L(\theta(w,x), y) + R(w)$	$\theta(w,\cdot): X \rightarrow [0,1]$	$\tau^* = argmax_\tau S_\tau(\theta(w,x), y)$
training phase	neural network	classification phase

- a neural network is a parametric function that approximates the map connecting data to the event probability
- the training phase realizes parameter optimization:
  - ✓ the loss function measures the ability of the network to fit the data in the training set
  - ✓ the regularization term allows generalization
- forecasting pretends a binary outcome: the probabilistic outcome is turned into a classifier via skill score thresholding
- a neural network can be fed with:
  - ✓ point-in-time feature vectors
  - ✓ time series of feature vectors
  - ✓ point-in-time images of ARs
  - ✓ videos of ARs

Paper	Data	Multiple test realizations	AR separation	Epoch selection	Method	TSS (C1+ flares)	TSS (M1+ flares)
Bobra & Couvidat (2015)	Point in time features (SHARP)	Yes	No	Yes	SVM	-	0.74
Liu et al. (2017)	Point in time features (SHARP)	Yes	No	Yes	RF	-	0.76
Nishizuka et al. (2018)	Point in time features (SHARP + others)	No	Yes	No	MLP	0.63	0.80
Florios et al. (2018)	Point in time features (FLARECAST)	Yes	No	Yes	RF	0.60	0.74
Jonas et al. (2018)	Point in time features (SHARP+others)	Yes	Yes?	Yes	RF	-	0.74 – 0.81
Campi et al. (2019)	Point in time features (FLARECAST)	Yes	Yes	Yes	Hybrid Lasso	0.54	0.67
Liu et al. (2019)*	Time series features (SHARP)	Yes	Yes	Yes	LSTM	0.61	0.79
Wang et al. (2020)	Time series features (SHARP)	..	Yes	Yes	LSTM	0.55?	0.68?
Park et al. (2018)	HMI and MDI images	No	Yes	Yes	CNN	0.63	..
Huang et al. (2018)	HMI and MDI images	Yes	No	-	CNN	0.49	0.66
Li et al. (2020)	HMI images	Yes	Yes	No	CNN	0.68	0.75
Yi et al. (2021)	HMI images	No	Yes	Yes	CNN	0.65	-

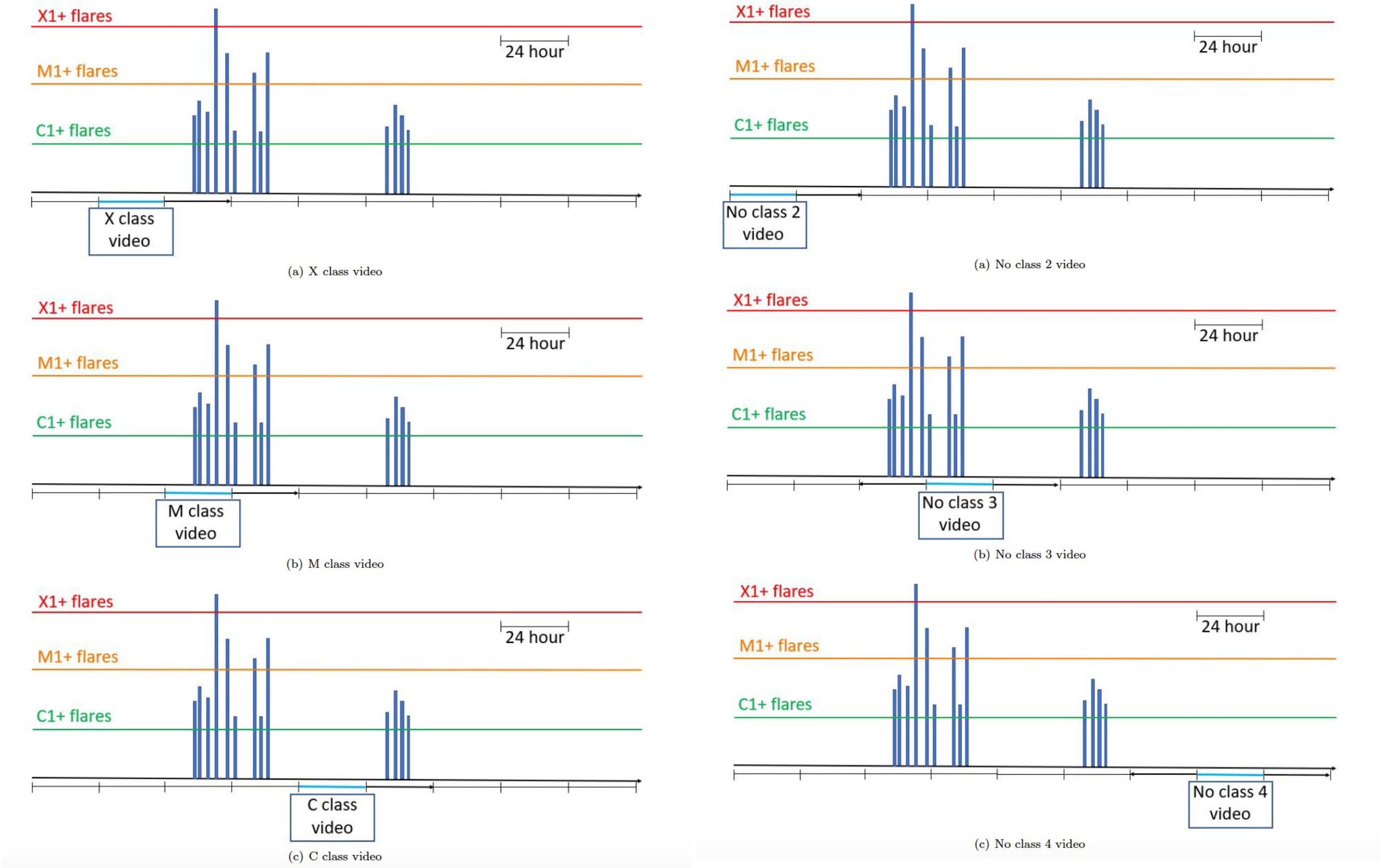
supervised flare forecasting: data generation

definition of data samples:

- X, M, C class
- NO1, NO2, NO3, NO4

well-balanced training sets:

- **proportionality:** same rates of samples for each sample type for
- **parsimony:** each subset of samples made by as few ARs as possible (i.e., samples belonging to the same AR fall into the same data set)



(a) X class video
(b) M class video
(c) C class video
(a) No class 2 video
(b) No class 3 video
(c) No class 4 video

algorithm for data set generation must guarantee the generation of well-balanced sets

supervised flare forecasting: epoch selection

- the optimization of the weights in the network is an iterative process
- the stopping rule for the iterations (epochs) must rely on a validation set

algorithm for epoch selection:

- generate a well-balanced validation set (labels known)
- chose the weights corresponding to the epoch that provides that highest skill score

supervised flare forecasting: assessment of results

- skill score: $$CM(y,\hat{y}) = \begin{pmatrix} TN(y,\hat{y}) & FP(y,\hat{y}) \\ FN(y,\hat{y}) & TP(y,\hat{y}) \end{pmatrix}$$

$$TSS(CM(y,\hat{y})) = \frac{TP(y,\hat{y})}{TP(y,\hat{y}) + FN(y,\hat{y})} + \frac{TN(y,\hat{y})}{TN(y,\hat{y}) + FP(y,\hat{y})} - 1$$

- bootstrap analysis:
  - ✓ many classification tests should be performed by generating many triples of training, validation and test sets
  - ✓ bootstrap-like process: randomly extraction of AR images from the HMI archive
  - ✓ confidence interval for the skill score

supervised machine learning: optimization

the training phase relies on the minimization of a loss function (plus a penalty term)

score-oriented loss (SOL) functions:

- define a probabilistic skill score matrix where the threshold is a random variable

$$\overline{CM}(y,\theta(x)) = \begin{pmatrix} \overline{TN}(y,\theta(x)) & \overline{FP}(y,\theta(x)) \\ \overline{FN}(y,\theta(x)) & \overline{TP}(y,\theta(x)) \end{pmatrix}$$

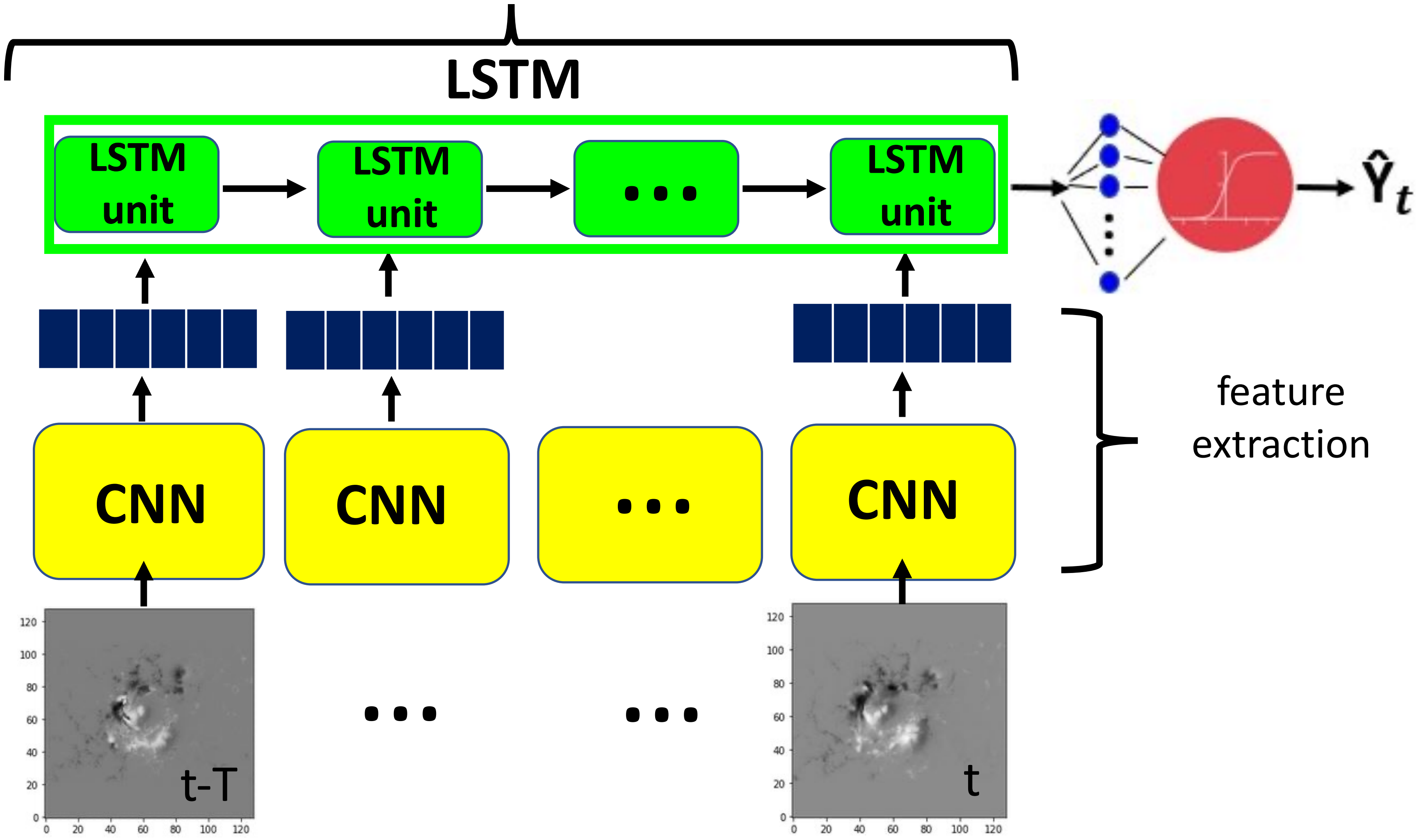
- define the corresponding probabilistic TSS
- define the (differentiable) loss function

$$L(\theta(x),y) = -\,TSS\left(\overline{CM}\big(y,\theta(x)\big)\right)$$

the minimization of this TSS-oriented loss function corresponds to the maximization of the TSS (no a posteriori threshold tuning needed)

deep neural network architecture

analysis of the temporal aspect of feature sequences



long-term

recurrent

neural

network

data

SDO/HMI images recorded in the time range between 2012 September 14 and 2017 September 30:

- for each AR, we considered the HMI magnetogram frames associated to it and organized them in 24-hour long-time series of HMI magnetogram frames

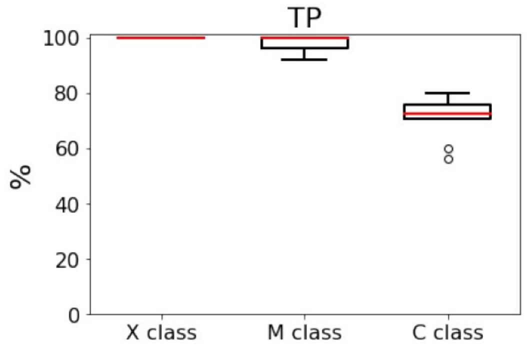
- each data sample is a video of HMI magnetogram frames associated to an AR.

- data from the past: each video is labelled by means of GOES data, depending on whether a flare either occurred or did not occur within 24 hours after the end of the video

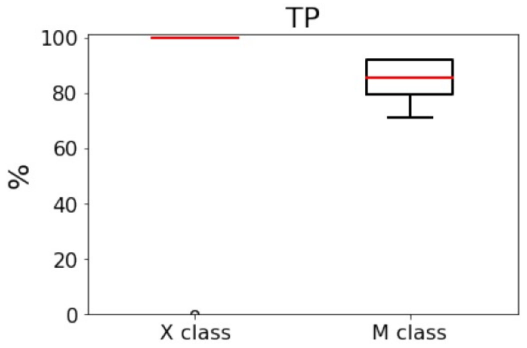
results

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

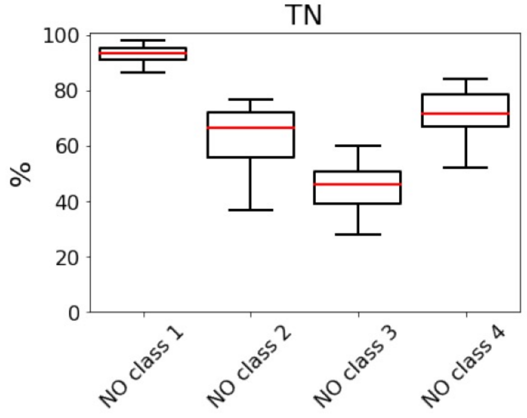
results



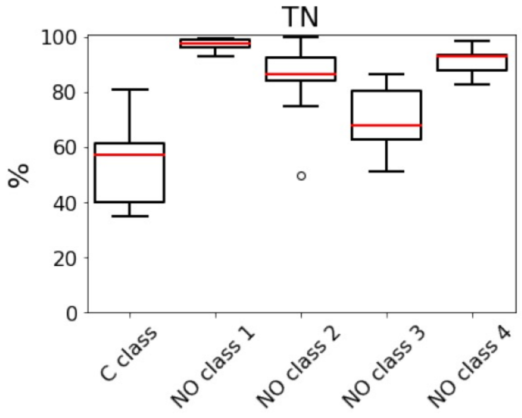
(a) C1+ flares prediction



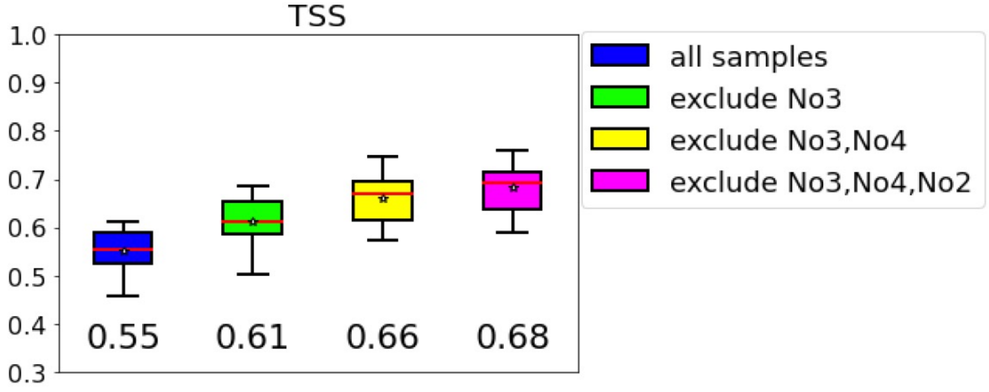
(b) M1+ flares prediction



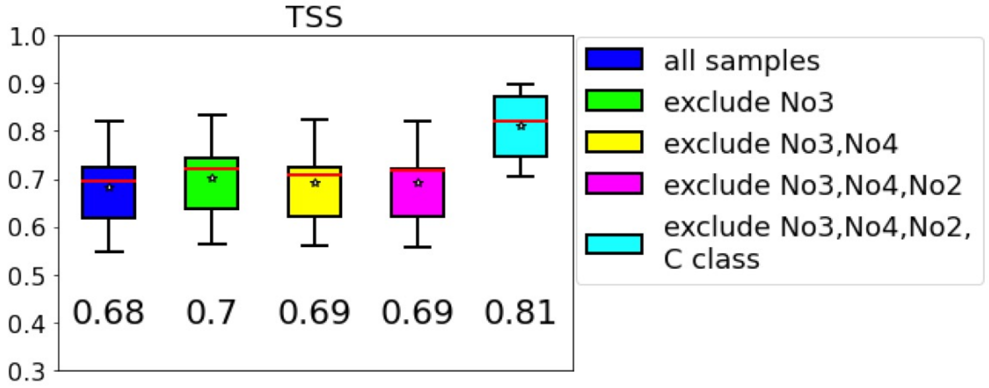
(c) C1+ flares prediction



(d) M1+ flares prediction



(a) C1+ flares prediction



(b) M1+ flares prediction

conclusions

- deep learning fed by HMI videos is feasible and provides skill scores comparable to the ones obtained by means using point-in-time images and feature vectors
- the way the training set is generated plays a crucial role with respect to the prediction performances of the supervised network
- score-oriented loss functions allow an automated optimization of the network (no a posteriori thersholding of the skill score needed)