Solar flare nowcasting using multi-wavelength SDO/AIA data

Paolo Massa[1]

Emslie A. G.[1], Novikov I.[1], Boils J.[1], Packwood J.[1], Liang G.[2], Guastavino S.[3], Marchetti F.[4], Piana M.[3]

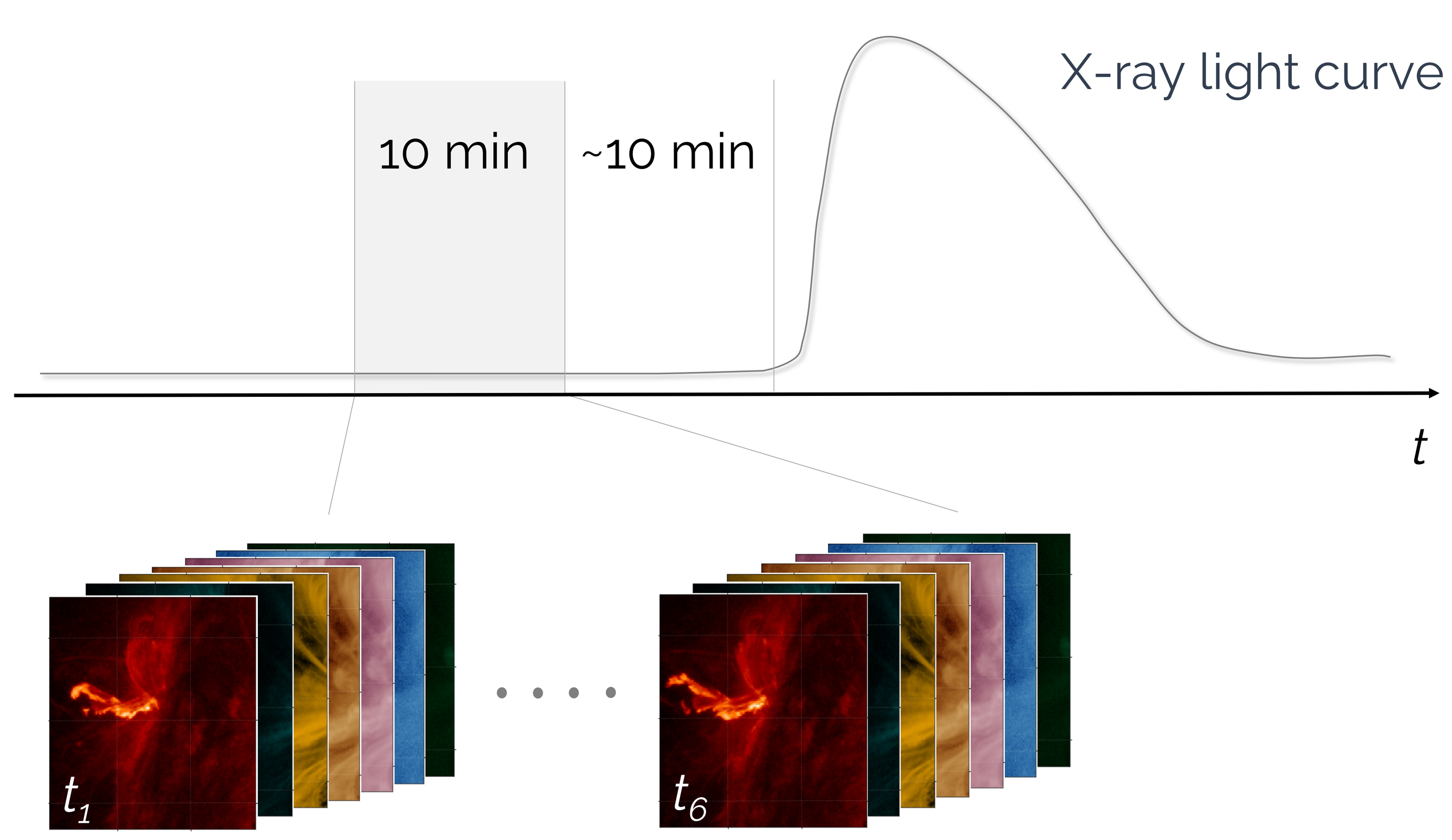
1. Department of Physics & Astronomy, Western Kentucky University, KY
2. Department of Computing and Cyber Security, Texas A&M University–San Antonio, TX
3. Department of Mathematics, University of Genova, Italy
4. Department of Mathematics “Tullio Levi Civita”, University of Padova, Italy

Introduction

The goal of this project is to develop machine learning methods for forecasting the occurrence of solar flares ~10 min in advance. Such methods can be used to support the upcoming NASA sounding rocket campaign scheduled for Spring 2024, during which three instruments (FOXSI, Hi-C, and SNIFS) will be launched on board rockets in order to observe the evolution of a solar flare [1].

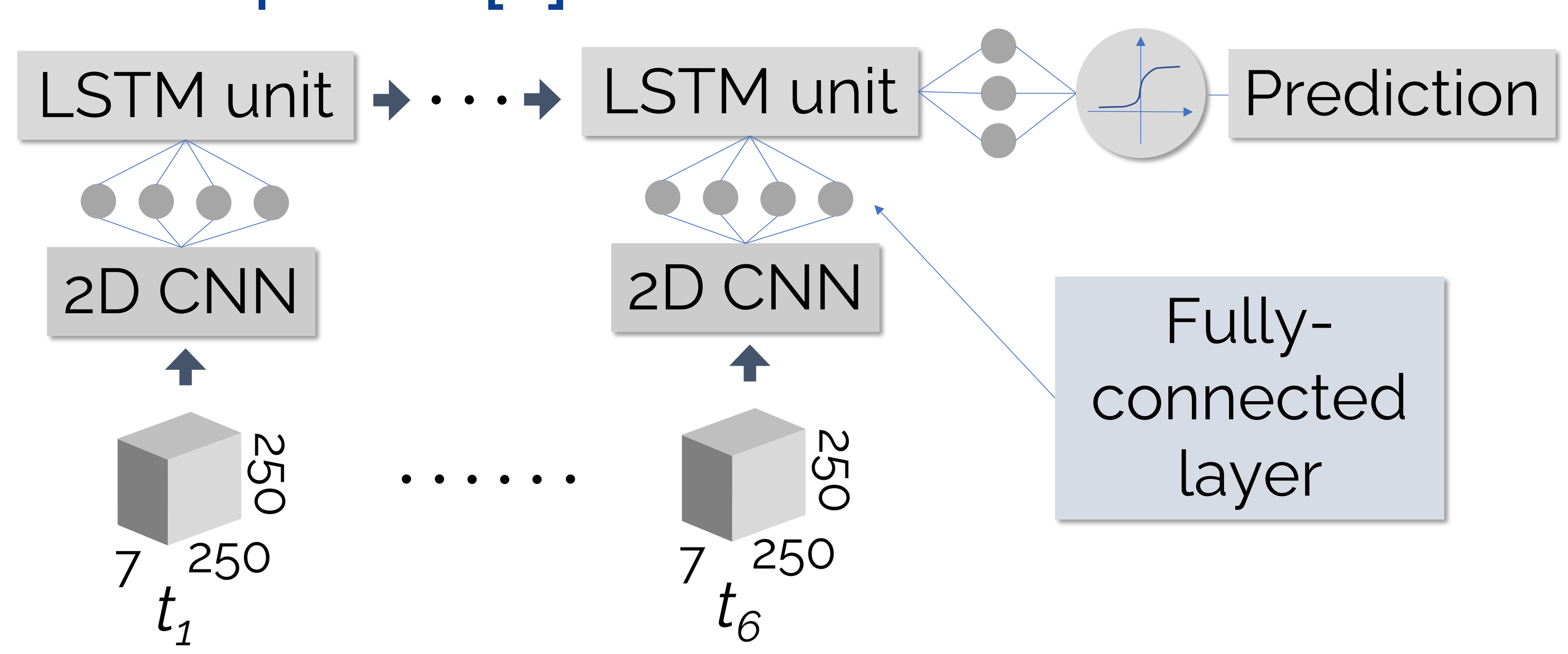
AIA data

The data we consider are Extreme Ultraviolet (EUV) images of active regions recorded by the Atmospheric Imaging Assembly on board the Solar Dynamics Observatory (SDO/AIA). We constructed “datacubes” containing 250 x 250 pixel images, recorded at seven different wavelengths (94 Å, 131 Å, 171 Å, 193 Å, 211 Å, 304 Å, and 335 Å), in 6 consecutive timesteps (spanning a range of 10 min, with a cadence of 2 min). The AIA datacubes we constructed are recorded between -20 and -10 min before flares with different intensity, according to the flare list defined in [2].



Machine learning approach

We trained and tested a Long-term Recurrent Convolutional Network (LRCN) developed in [3].



Our dataset consisted of

- 3240 datacubes in the training set;
- 400 datacubes in the validation set;
- 400 datacubes in the test set.

We labelled as positive the datacubes recorded before flares with GOES class larger than C5.

Results on the test set

We selected the seven epochs with largest True Skill Statistic (TSS; [4]) on the validation set and we defined an ensemble method as follows. We evaluated the selected seven models on each example of the test set, and we associated each example with the most frequent prediction. The results are reported in the tables below.

GOES class	# examples	# correct predictions
< B5	55	55
B5-C1	100	94
C1-C5	190	152
C5-M1	33	16
M1-M5	18	10
M5-X1	3	2
>X1	1	1

Confusion matrix

Predicted \ Actual	P	N
P	29	44
N	26	301

Skill scores

TSS	HSS
0.40	0.35

Conclusions and future developments

The skill scores values obtained so far are rather low. Therefore, for improving the model performances, we are planning to:

1. Reconstruct Differential Emission Measure (DEM) datacubes from the original AIA datacubes using a newly developed Regularized Maximum Likelihood (RML; [5]) method and train the LRCN model on the DEM datacubes. Thermodynamic changes preceding flaring activity should be more obviously revealed in DEM datacubes; hence, DEM datacubes should be more suited for addressing the flare forecasting problem.
2. Compress the AIA/DEM datacubes using a Fourier based technique described in [6] in order to extract useful information while reducing the size of the dataset.

References

[1] Savage S., et al., American Astronomical Society Meeting Abstracts. Vol. 53. No. 6. (2021).
[2] Angryk R.A., et al., Sci Data 7, 227 (2020).
[3] Guastavino S., et al., A&A 662, A105 (2022).
[4] Bloomfield D.S., et al., The Astrophysical Journal Letters, 747(2), L41 (2012).
[5] Massa P., et al., accepted for publication in A&A (2023).
[6] Massa P. and Emslie G., Frontiers in Astronomy and Space Sciences, 9 (2022).

Acknowledgements

This research is supported by the NASA KY EPSCoR award 80NSSC21M0362, with a match from the Commonwealth of Kentucky. The work is carried out in collaboration with researchers at UK and Stanford University, using the computer cluster at the UK Center for Computational Sciences.