operational flare forecasting with video-based deep learning

MIDAgroup

the MIDA group
dipartimento di matematica
università di genova

michele piana

european space weather week (ESWW) 2022

SWR1 - solar sources of space weather

zagreb, croatia

october, 27th 2022

Università di Genova

how to make an algorithm operational

three crucial issues:

- define a validation strategy for machine/deep learning
- account for the solar cycle phase in the training and validation steps
- reduce the overall computational burden

validation

validation strategy: (guastavino et al, astronomy and astrophysics, 2022)

generation of well-balanced training, validation and test sets:

- chronological splitting introduces a bias due to the cyclicity of the solar cycle
- data generation process based on machine learning theory: **training, validation and test sets must be drawn from the same distribution** (vapnik, 1998)

bootstrap analysis:

- many classification tests performed by generating many triples of training, validation and test sets
- random extraction of AR images from the HMI archive (BUT while keeping AR separation in training, validation and test)
- confidence intervals for the skill scores

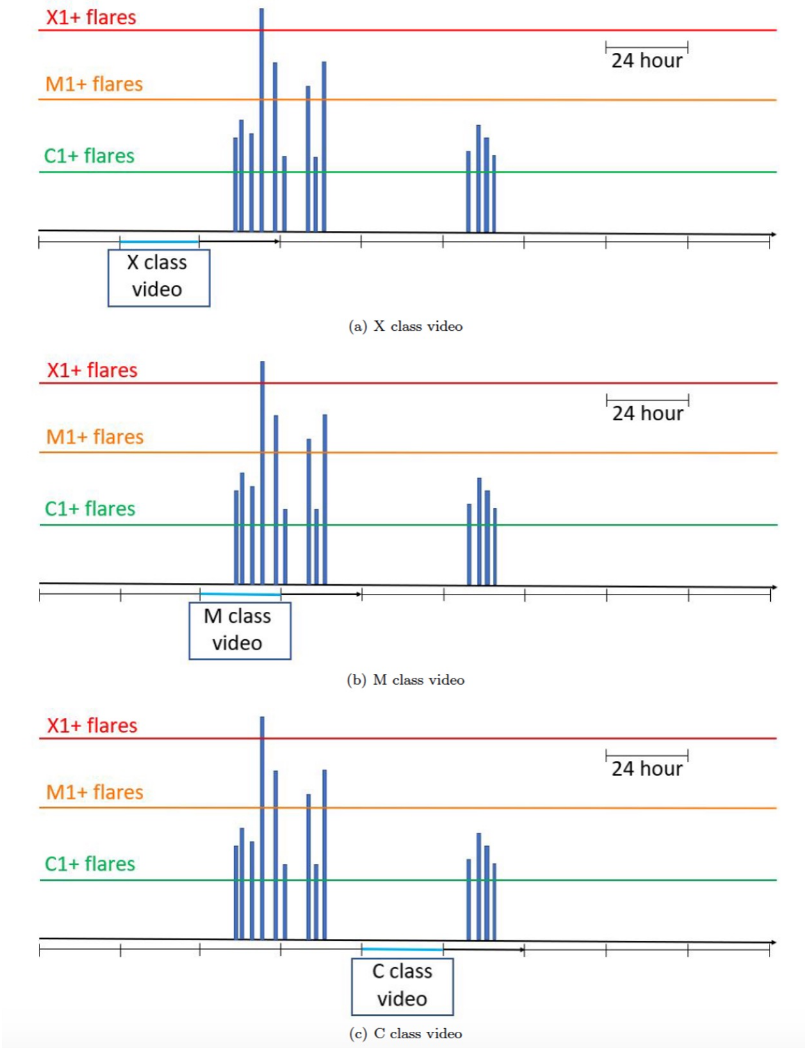
data set generation

definition of data samples:

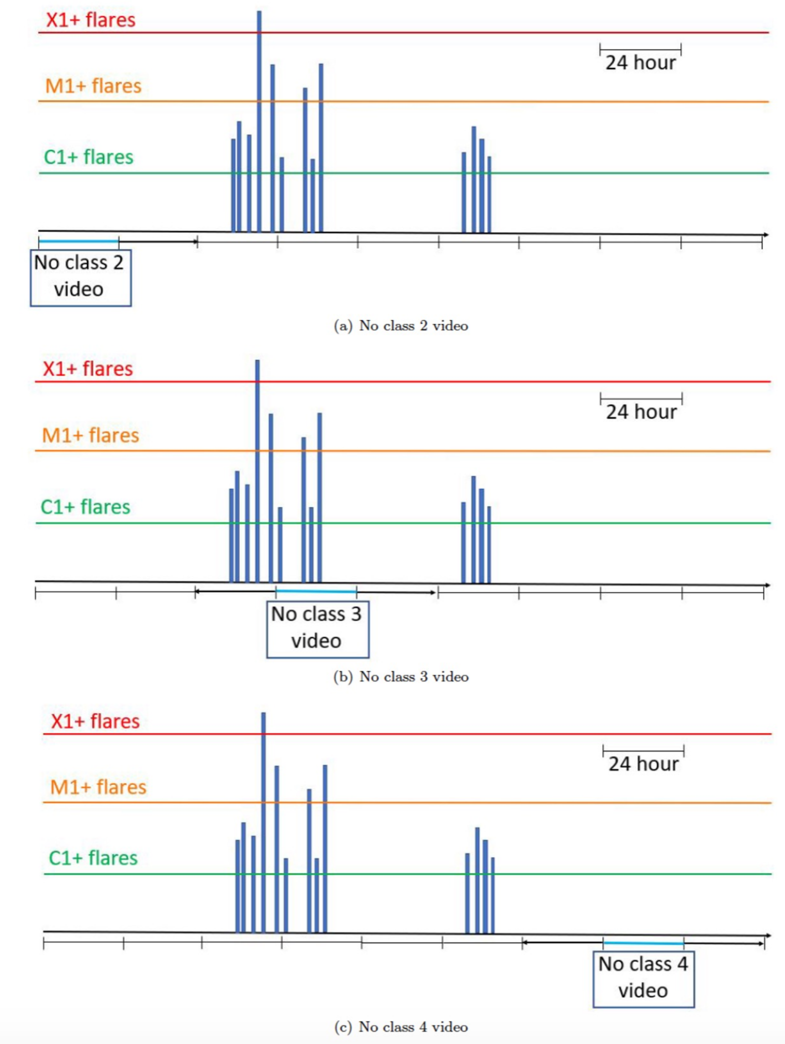
- X, M, C class
- NO1, NO2, NO3, NO4

well-balanced training and validation sets:

- **proportionality:** same rates of samples for each sample type
- **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

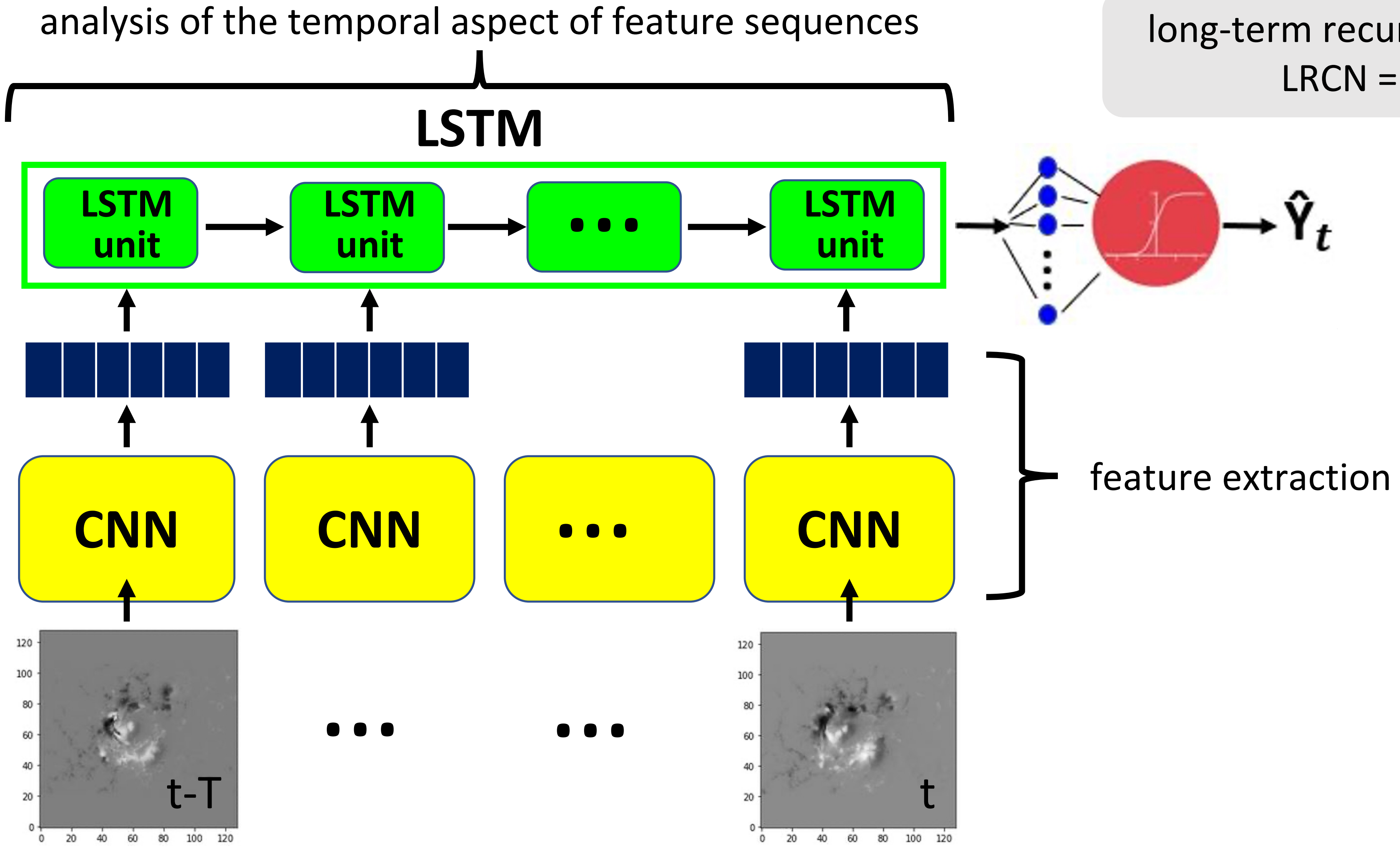
data

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

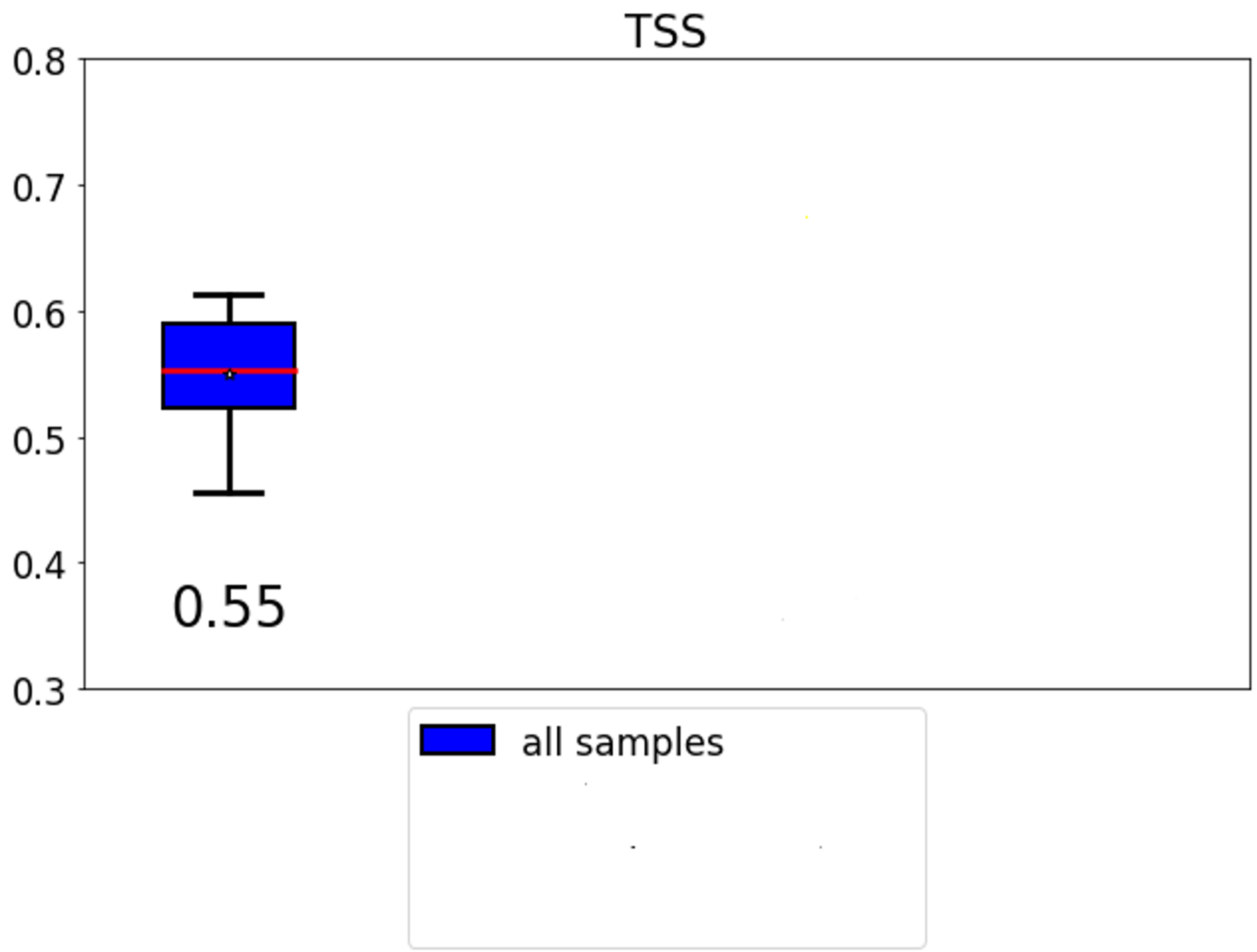
- for each AR, we considered the HMI magnetogram frames associated to it and we organized them in 24 hour long time series of HMI magnetogram frames
- we constructed a collection of data samples, each one represented by a video of HMI magnetogram frames associated to an AR
- data from the past: we annotated each time series with 0 if no flare occurred within 24 hours and with 1 if a flare occurred within 24 hours;
- we generated 10 training, validation and test sets according to the data generation process in order to assess the statistical robustness in results

deep neural network architecture

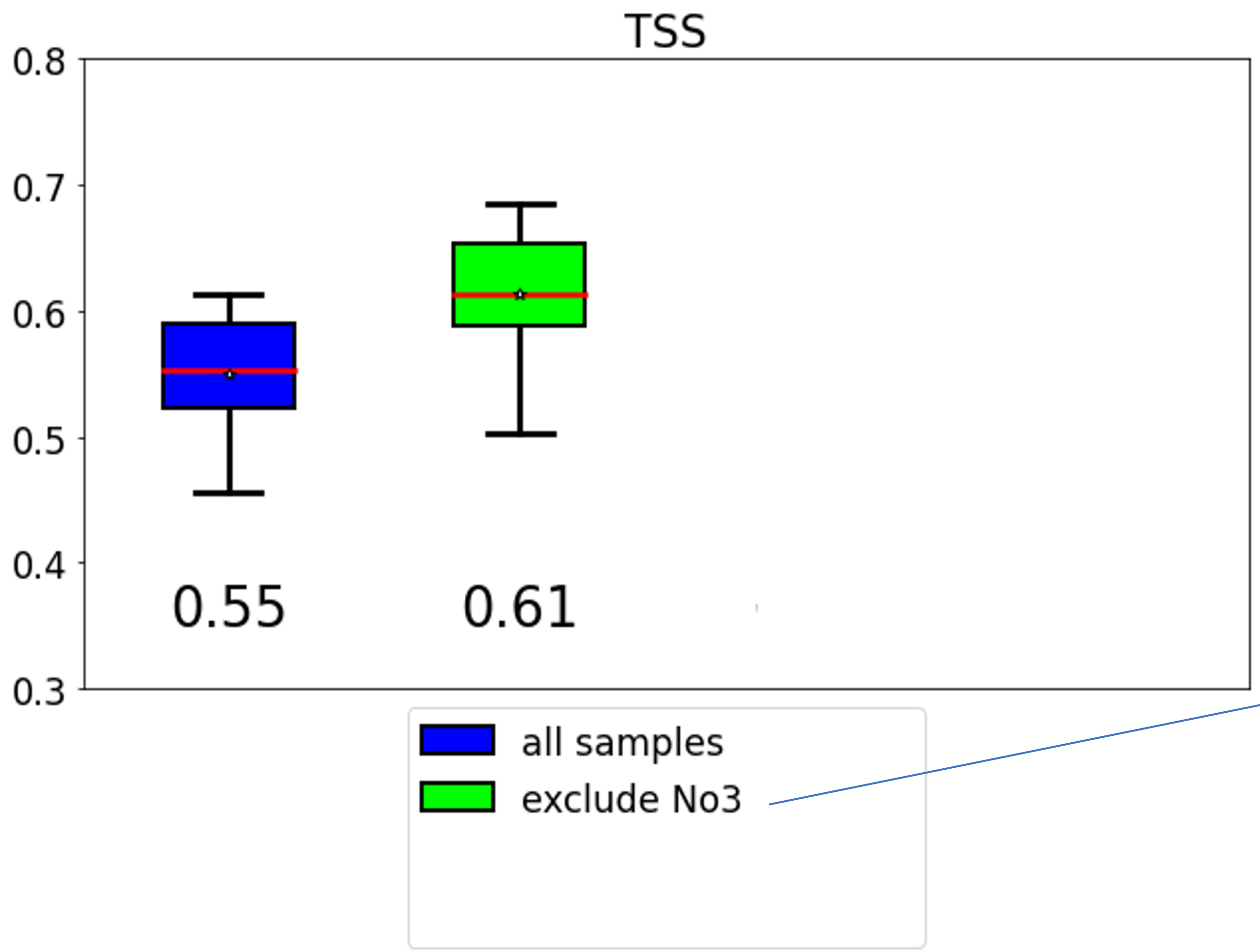
long-term recurrent neural network
LRCN = CNN + LSTM

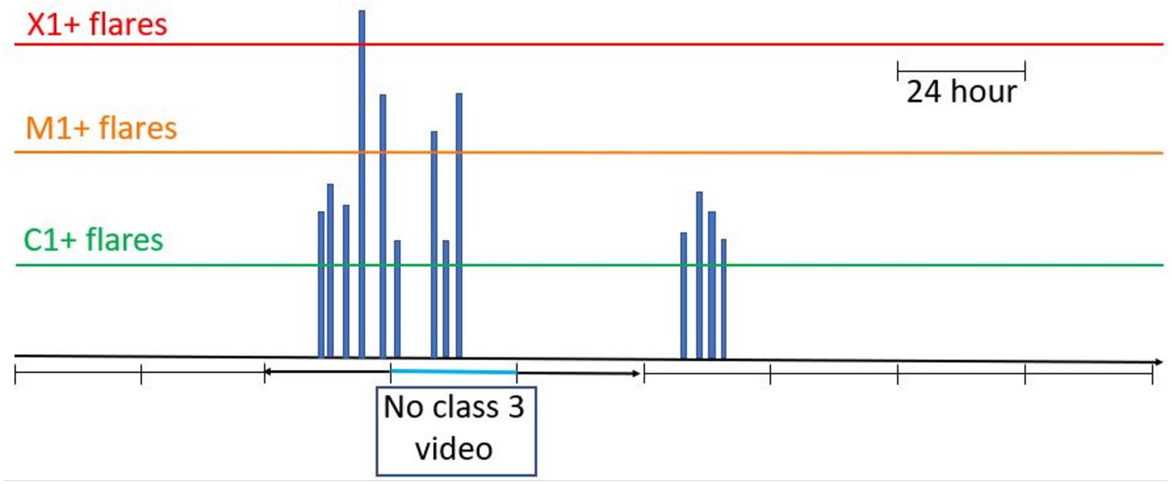


results on test sets: C+ flare prediction

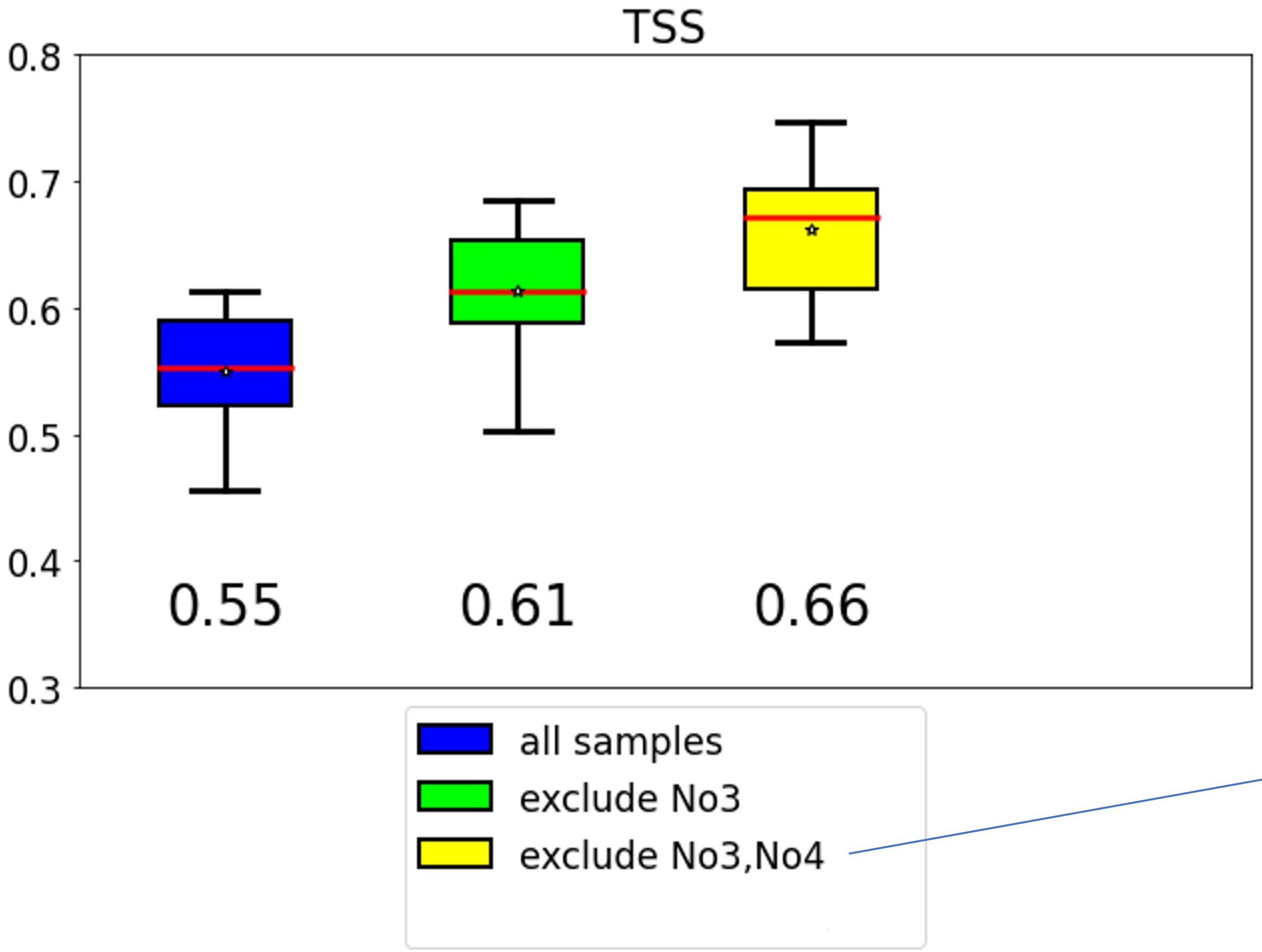


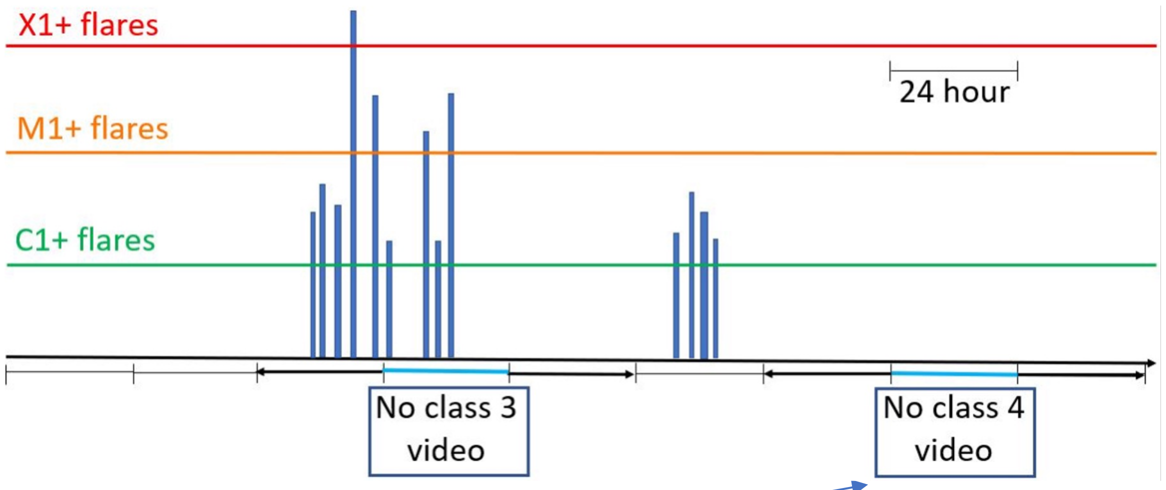
results on test sets: C+ flare prediction



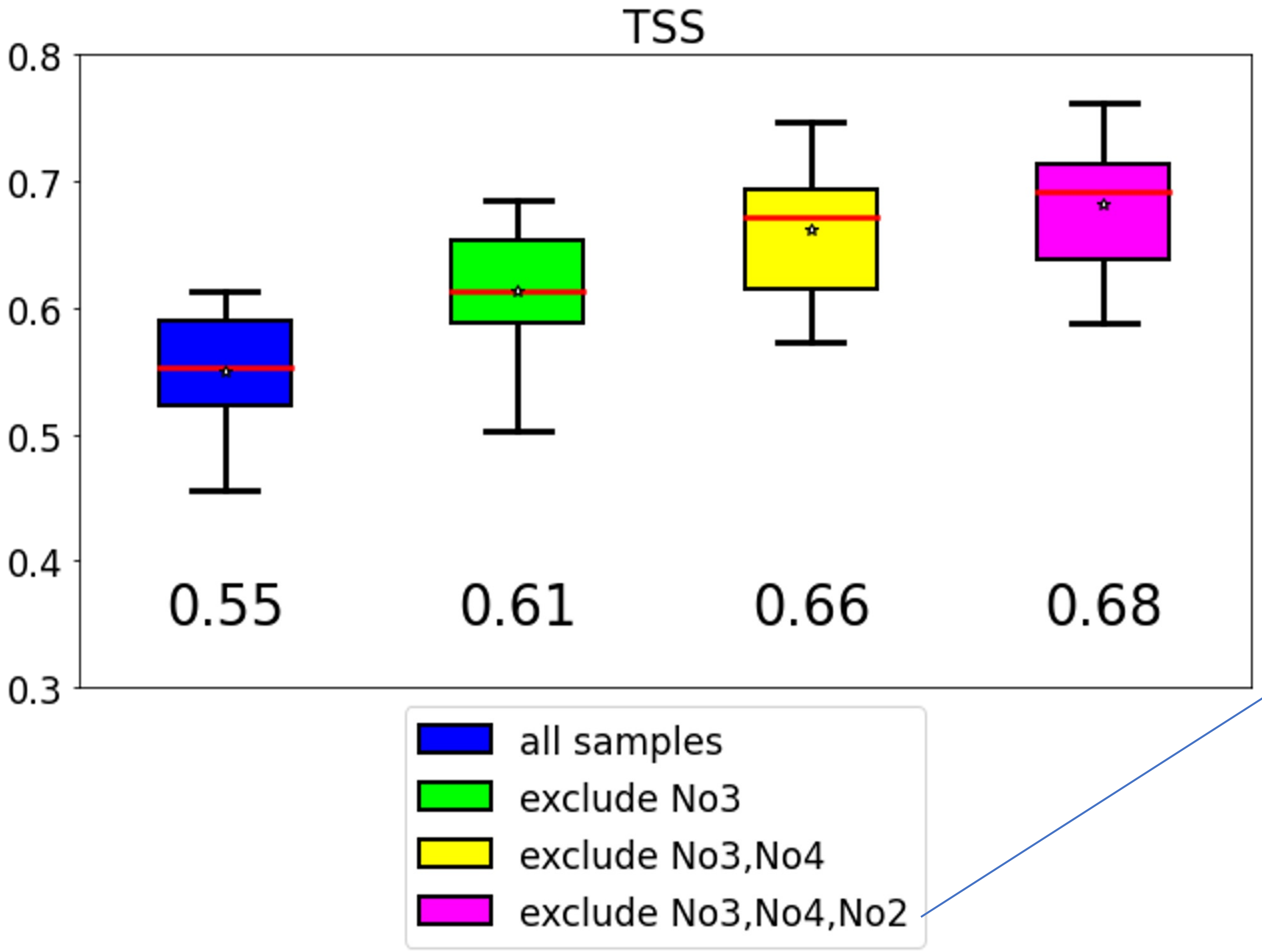


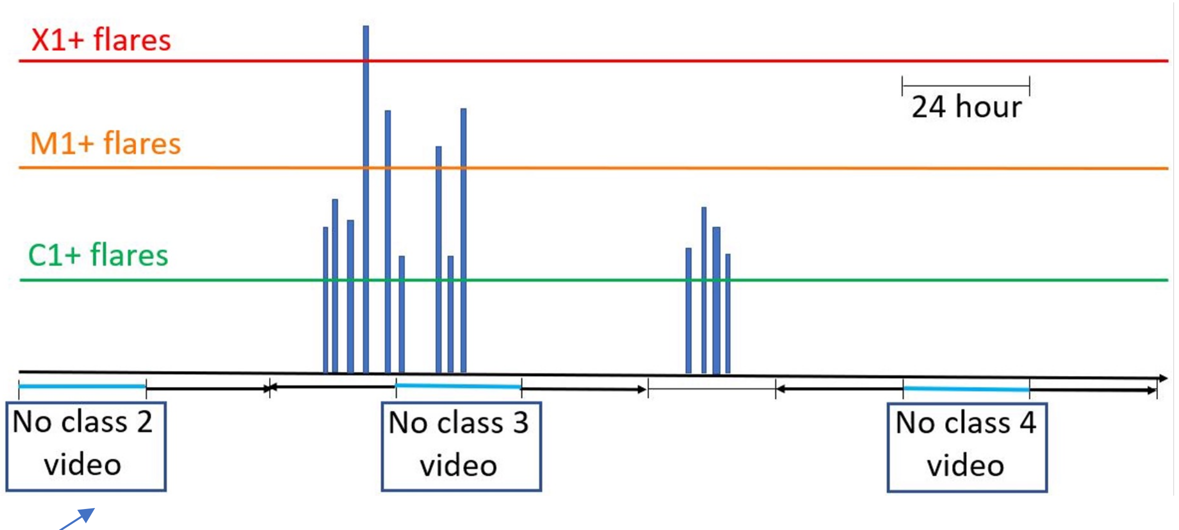
results on test sets: C+ flare prediction



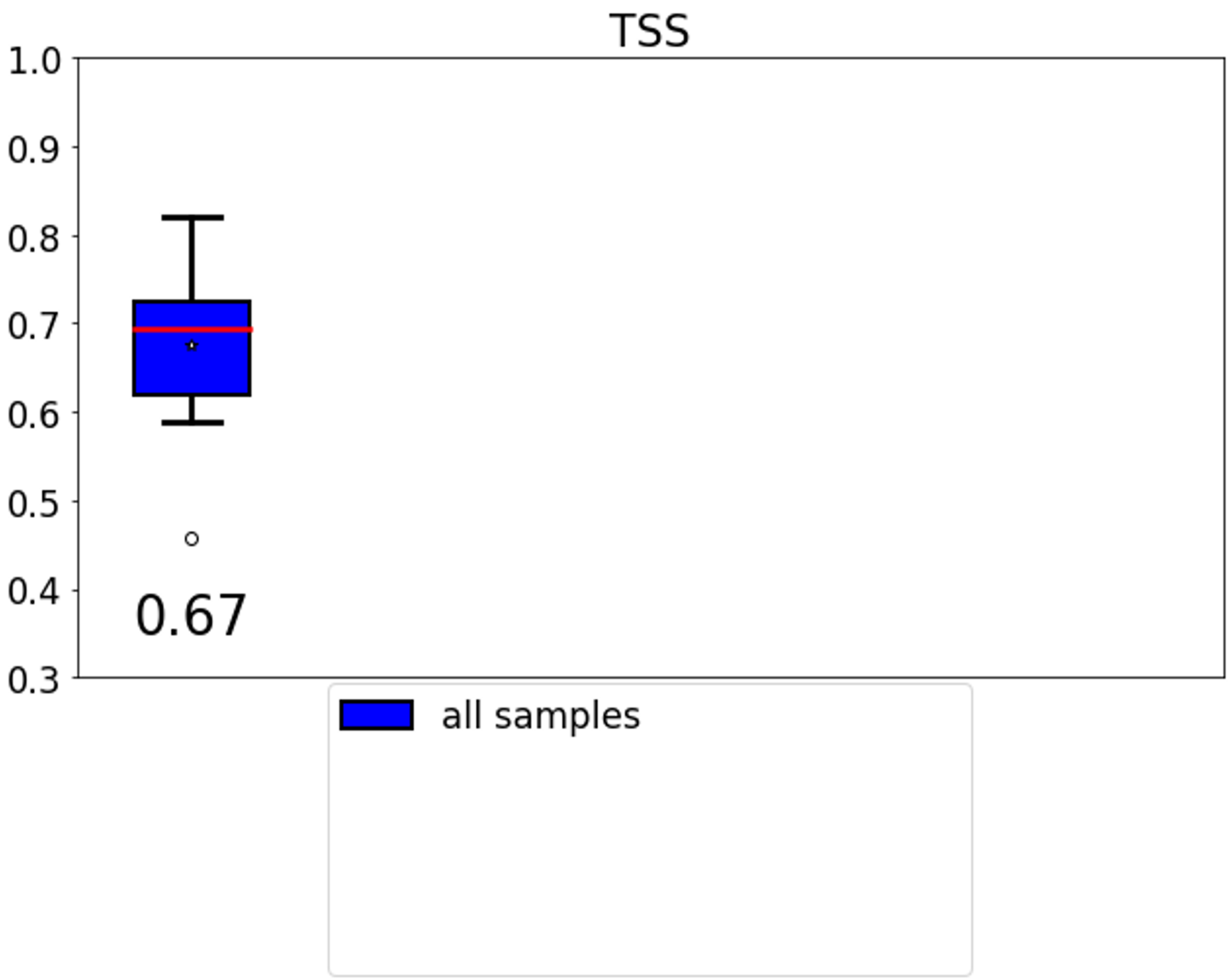


results on test sets: C+ flare prediction

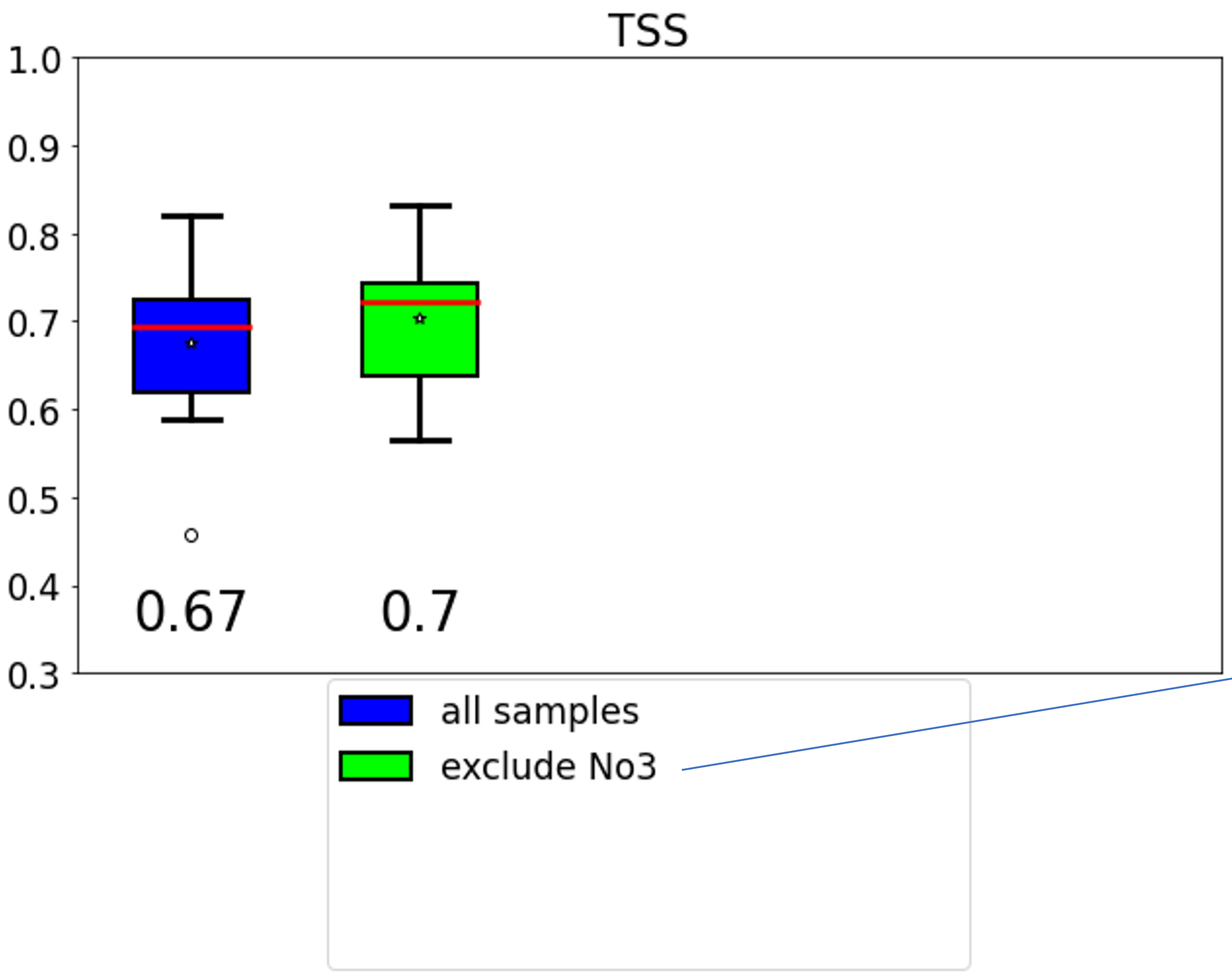


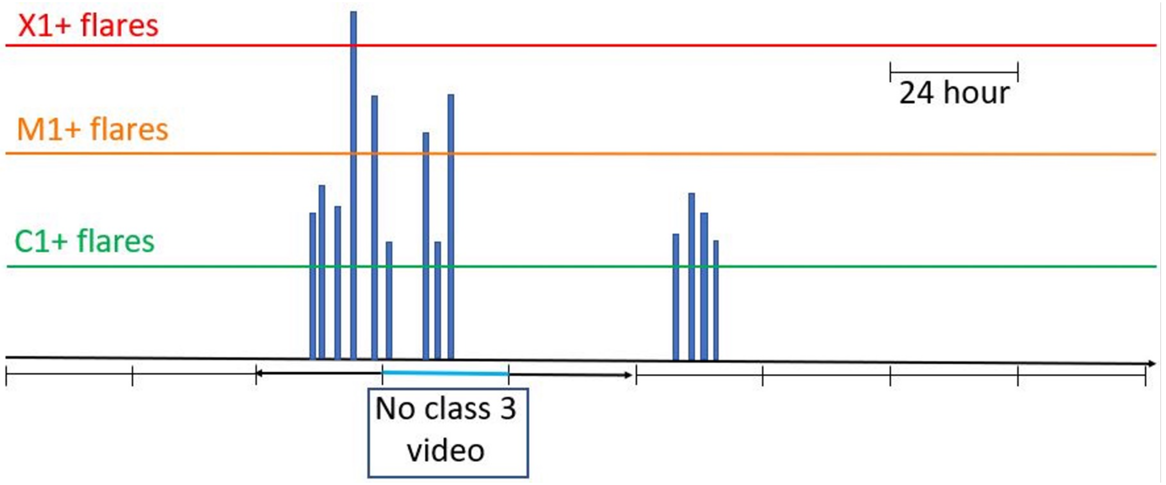


results on test sets: M+ flare prediction

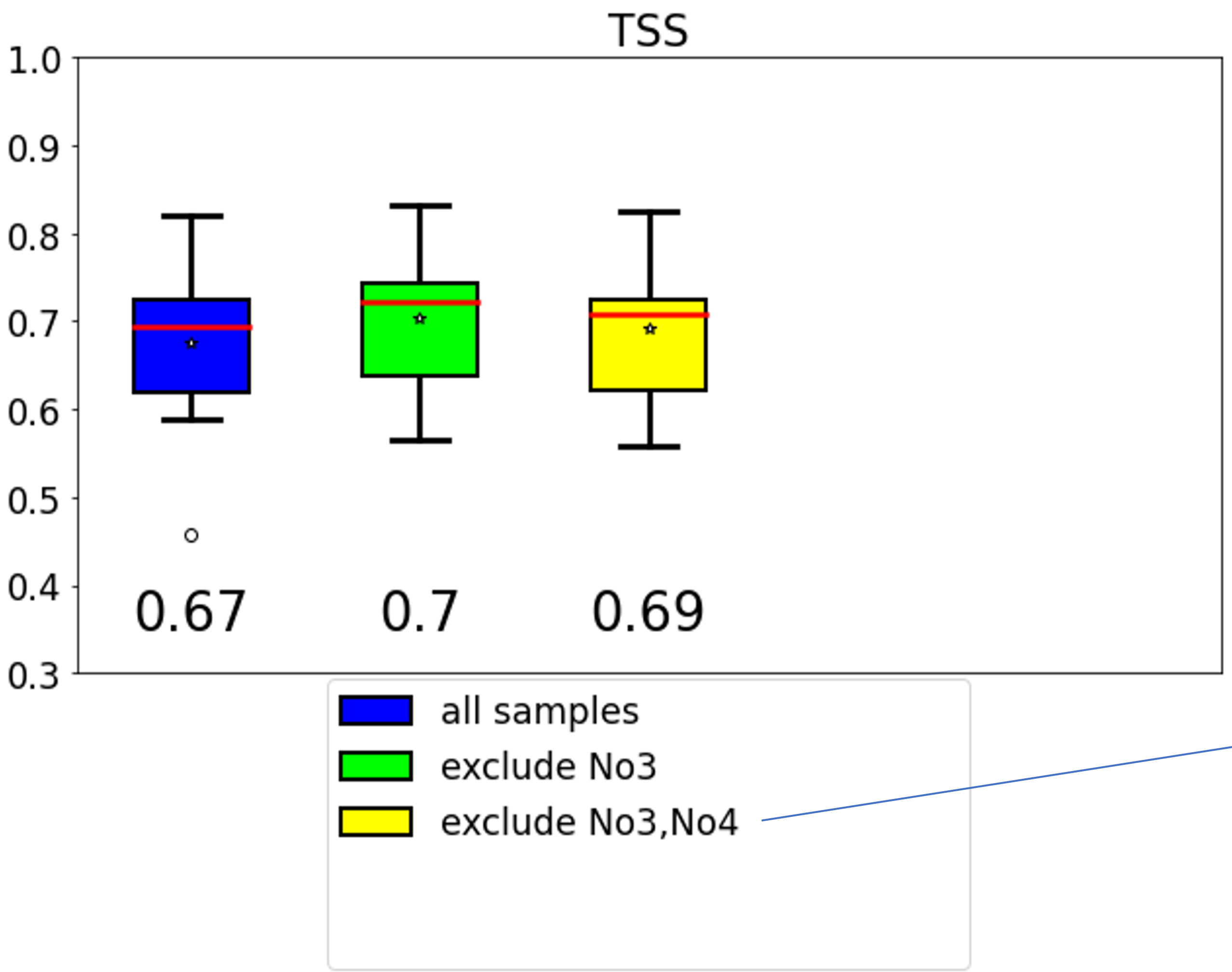


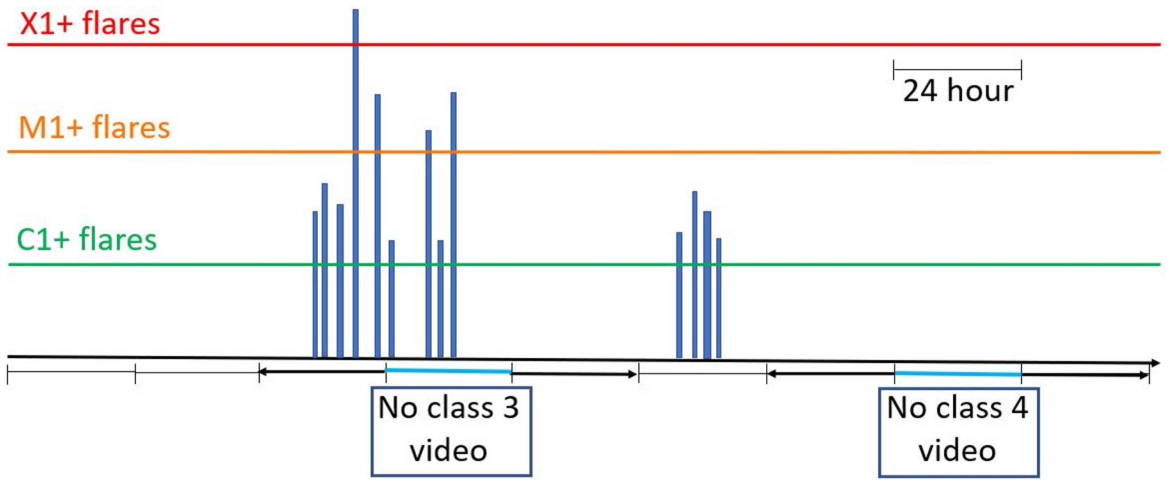
results on test sets: M+ flare prediction



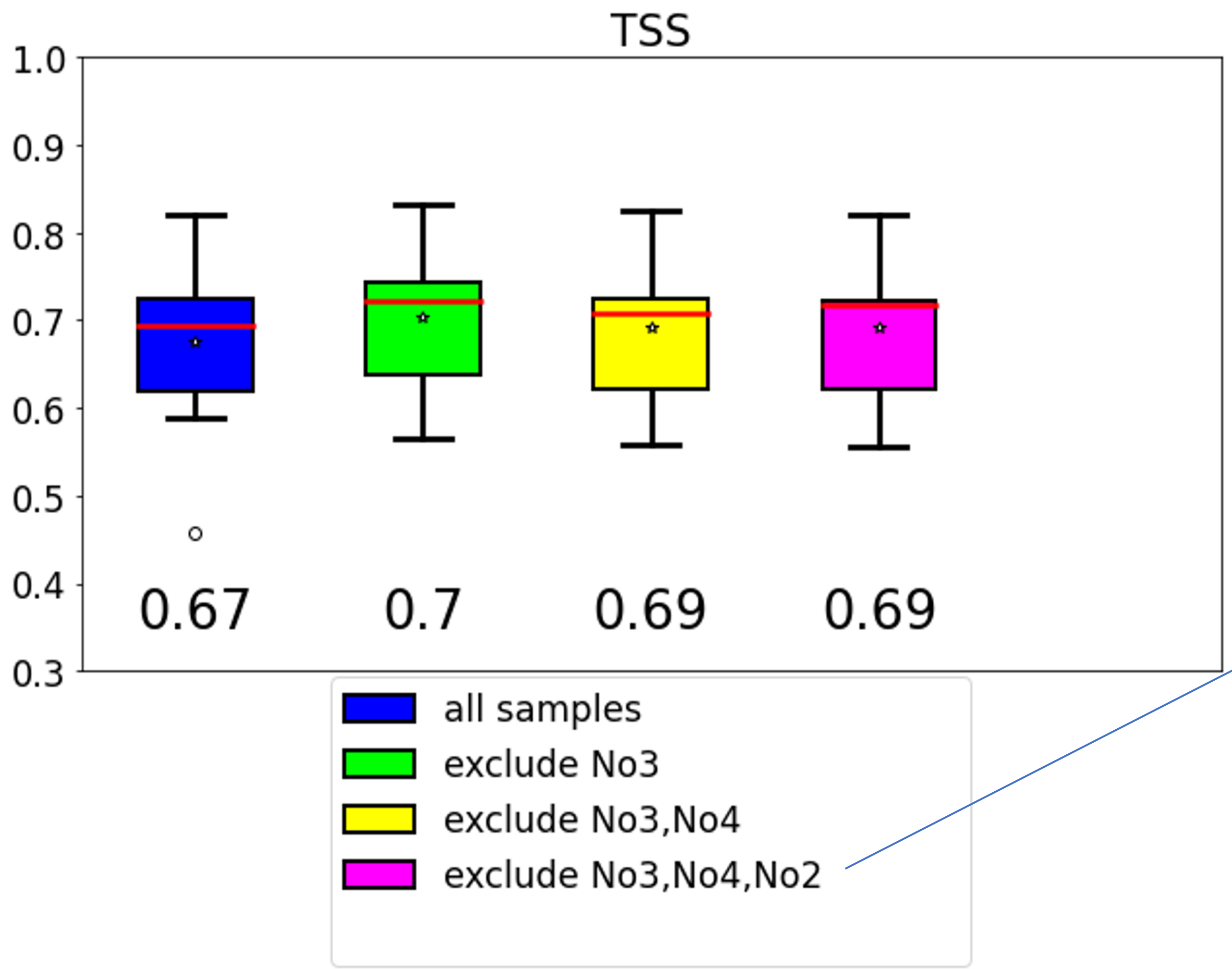


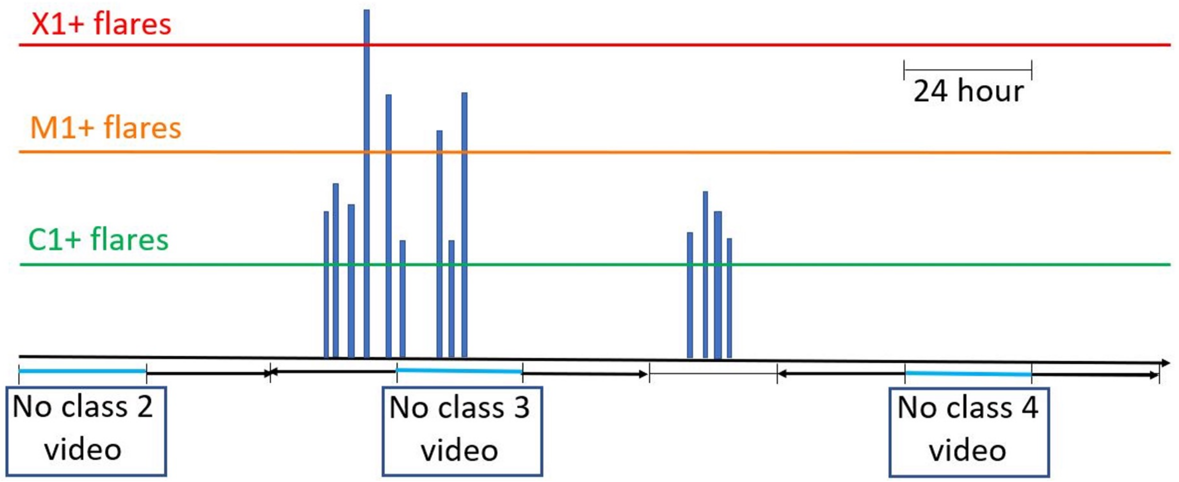
results on test sets: M+ flare prediction



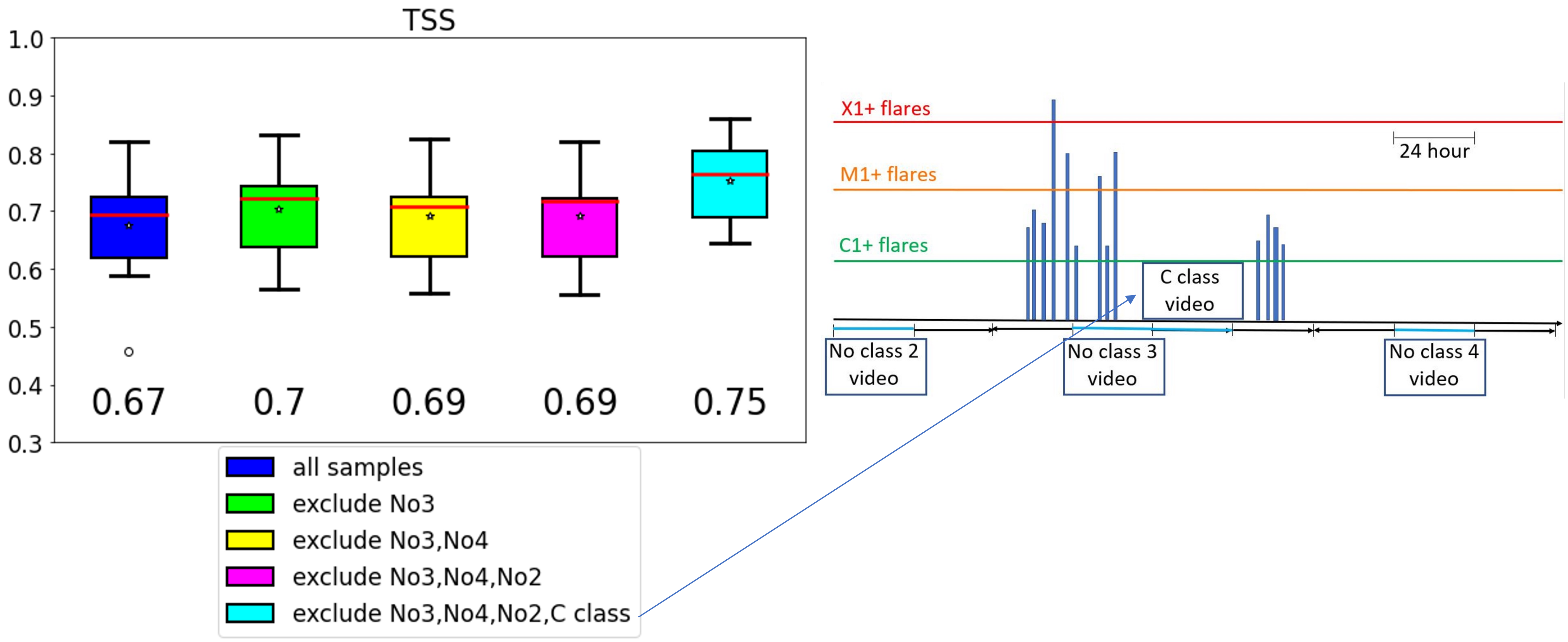


results on test sets: M+ flare prediction





results on test sets: M+ flare prediction

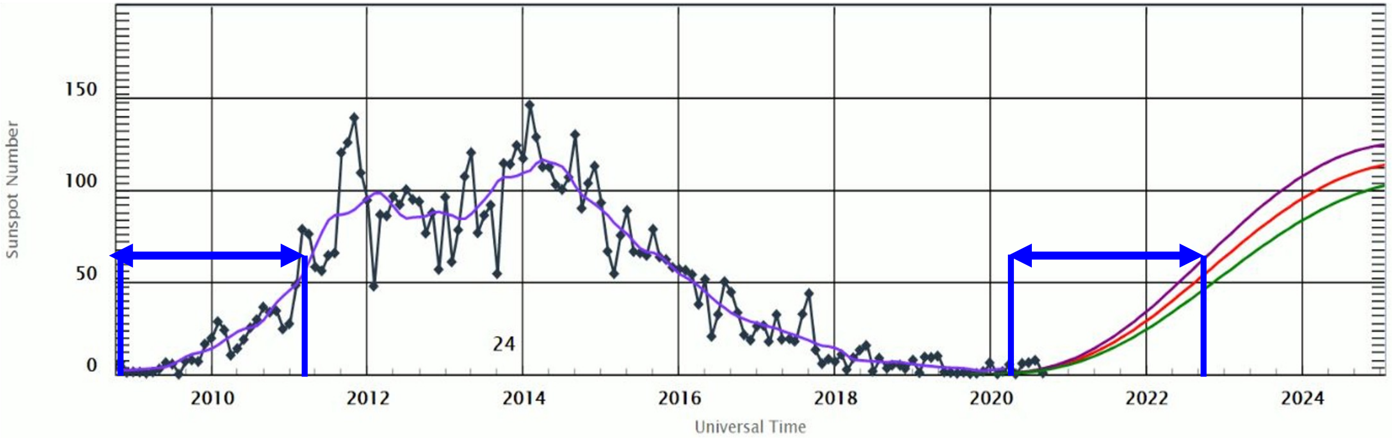


focus on solar cycle

generalization of the operational flare forecasting process

- identification of three phases in the current solar cycle: increasing, plateau, decreasing

- given a time window in the current solar cycle, the corresponding phase is identified.



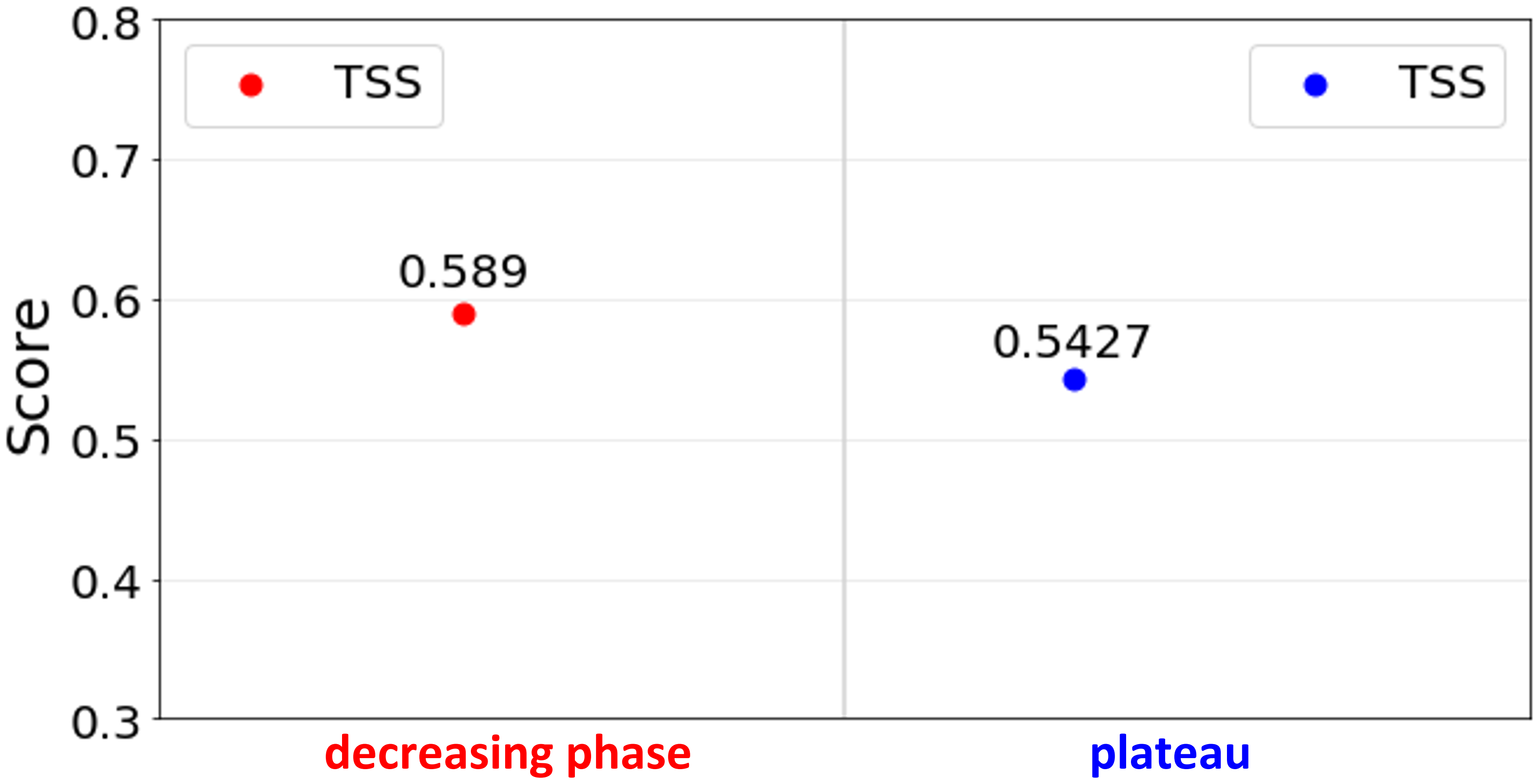
- for the same phase in the previous solar cycle the data set generation algorithm computes the rates of the different sample types

- the training and validation sets are generated according to the sample rates from the whole data archive at disposal

results: C+ flare prediction (guastavino et al, arXiv:2209.05128, 2022)

test window: march-december 2015 (decreasing phase)

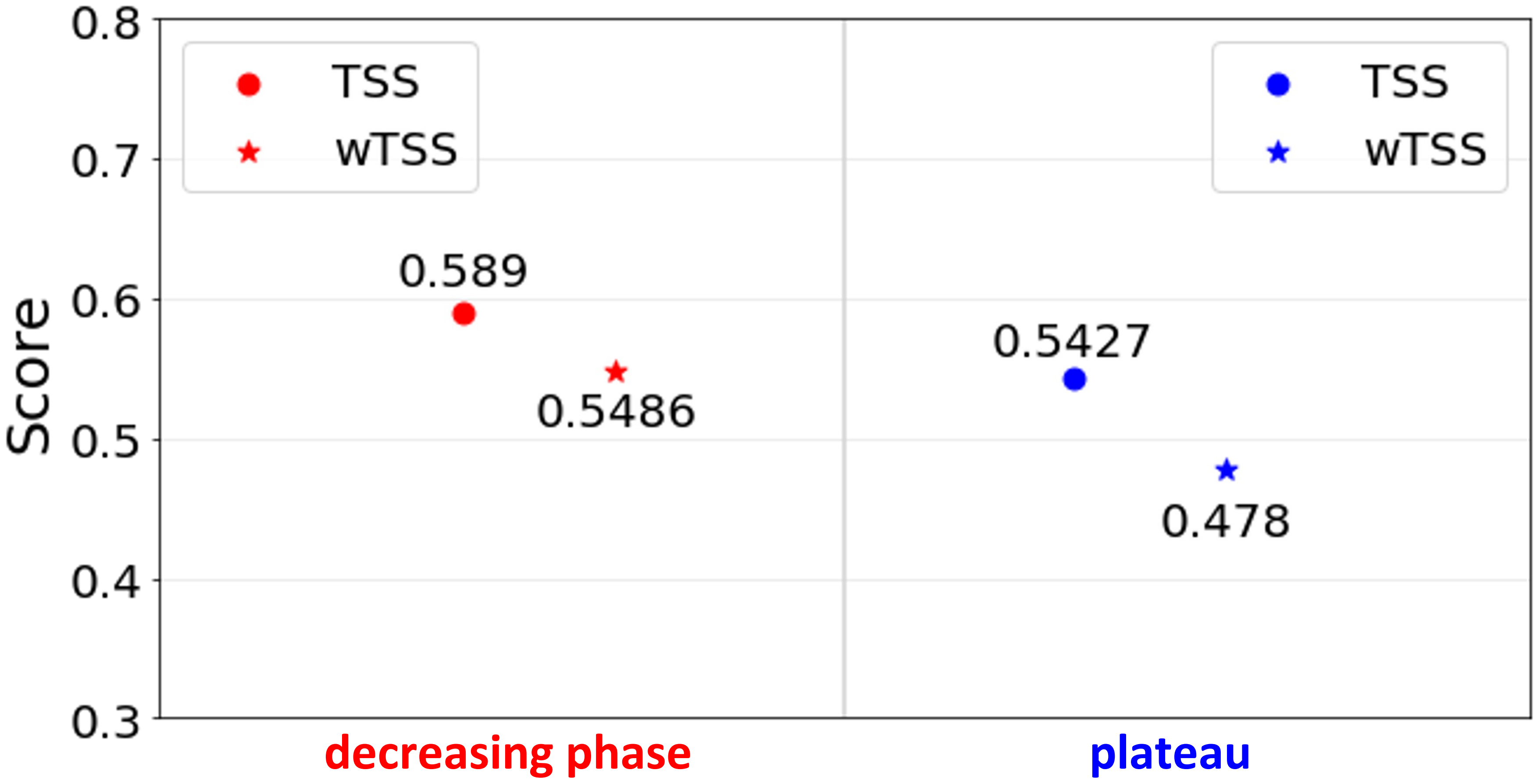
training and validation sets are generated with the sample rates computed on the **decreasing phase** and on the **plateau**



results: C+ flare prediction (guastavino et al, arXiv:2209.05128, 2022)

test window: march-december 2015

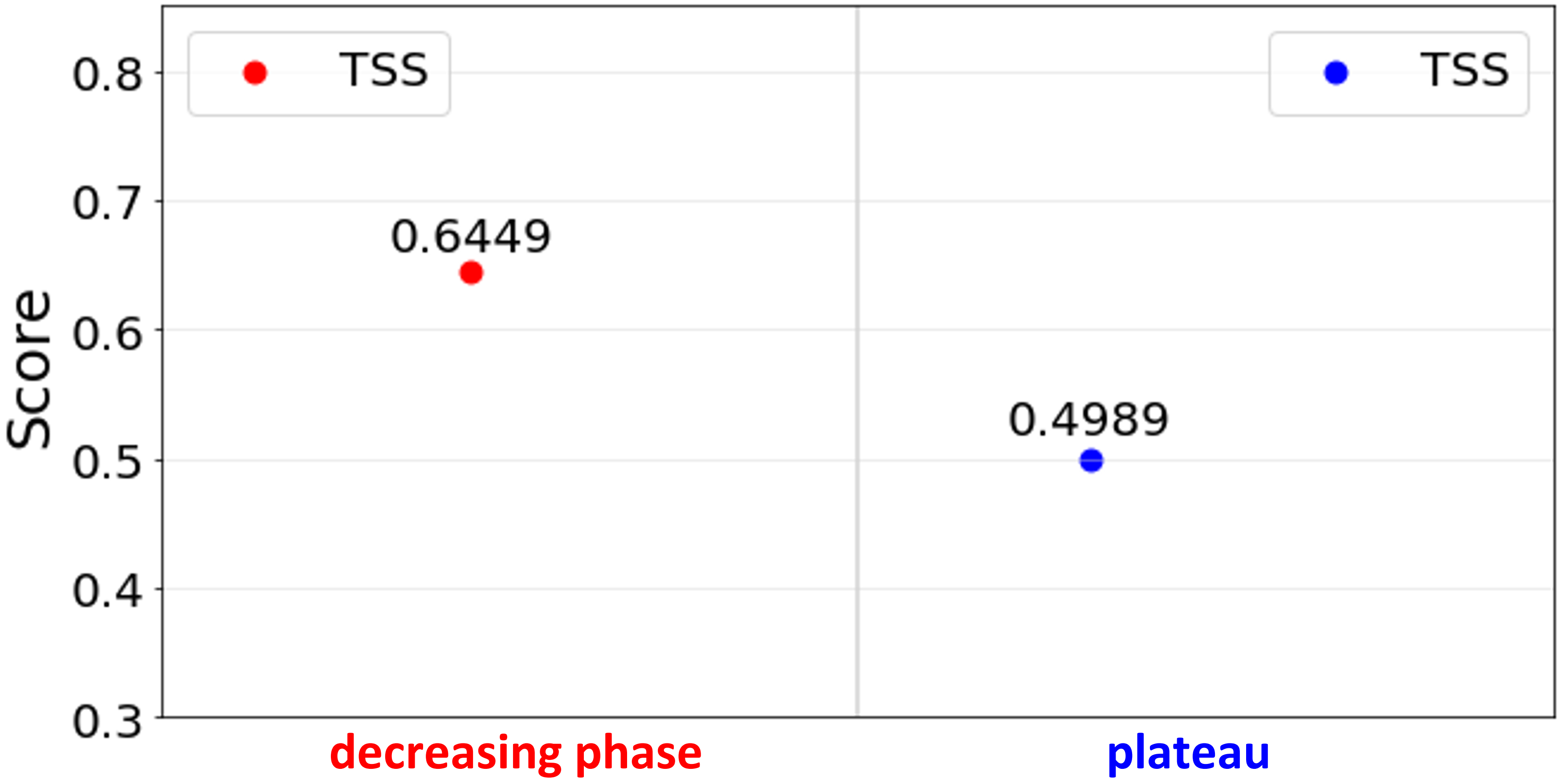
training and validation sets are generated with the sample rates computed on the **decreasing phase** and on the **plateau**



results: M+ flare prediction (guastavino et al, arXiv:2209.05128, 2022)

test window: March-December 2015

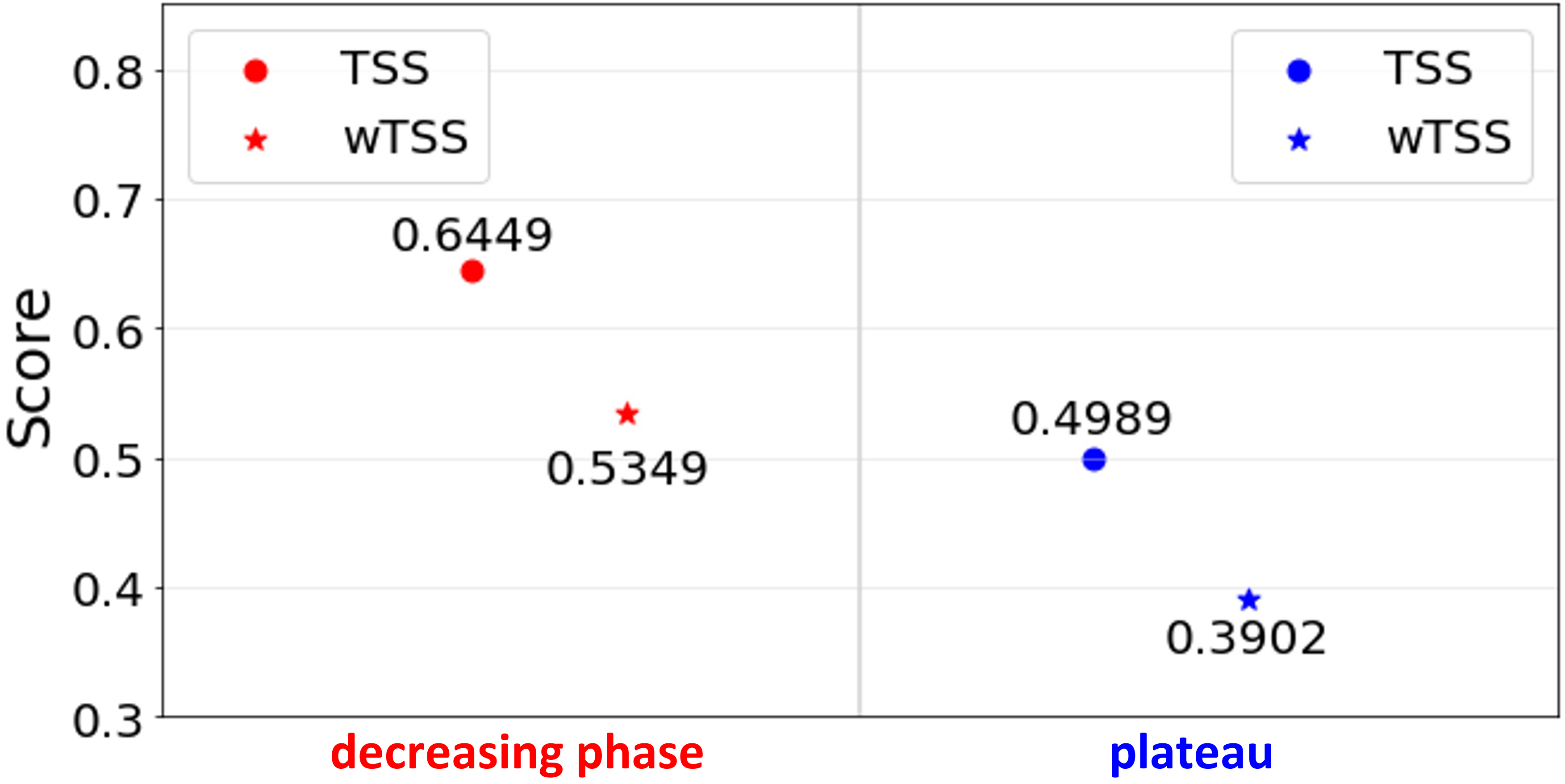
training and validation sets are generated with the sample rates computed on the same phase or different phase



results: M+ flare prediction (guastavino et al, arXiv:2209.05128, 2022)

test window: march-december 2015

training and validation sets are generated with the sample rates computed on the **decreasing phase** and on the **plateau**



computational issues

online training: machine learning vs CNNs

- feature-based machine learning is computationally more effective for online training than image-based deep learning
- sparsity enhancing algorithms allow the identification of the image features that mostly impact the flare forecasting performances

nice piece of news (campi et al, astrophysical journal, 2019):

- rather few features contribute to the prediction process
- such features depend on neither the machine learning algorithm nor the issuing time

feature-based experiment

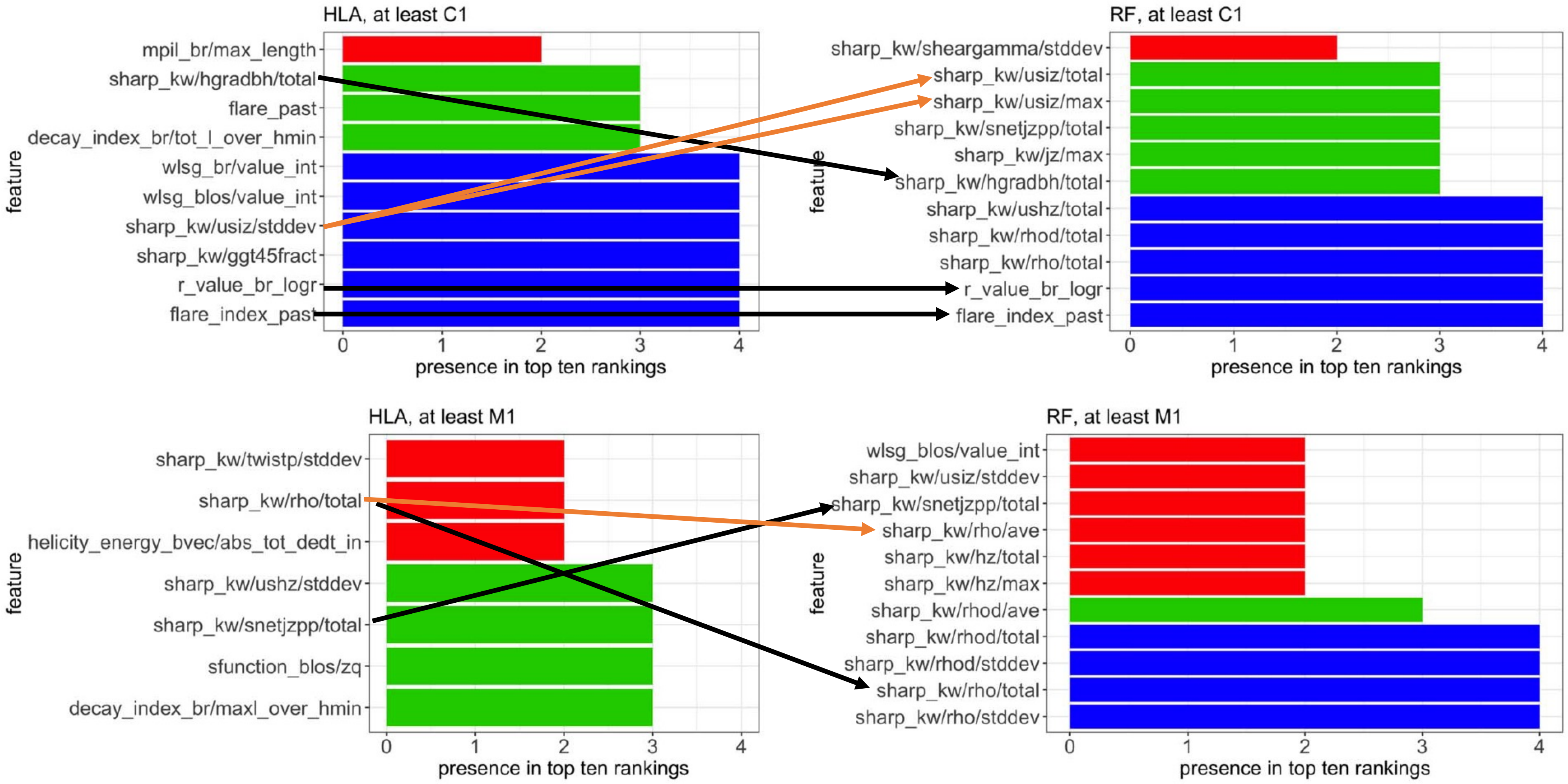
point-in-time SHARP images:

- time range: 09/14/2012 – 04/30/2016
- four issuing times: 00:00 UT – 06:00 UT – 12:00 UT – 18:00 UT
- cadence: 24 hours

features (for each AR):

- 171 features identified in each active region:
  - 167 extracted with a specific pattern recognition algorithms
  - longitude and latitude of the AR
  - binary label encoding the presence of a flare in the past
  - flare class (if occured)
- overall 4442 sets of 171-dimension feature vectors (one AR may last for more than one HMI image)

different algorithms



different issuing times

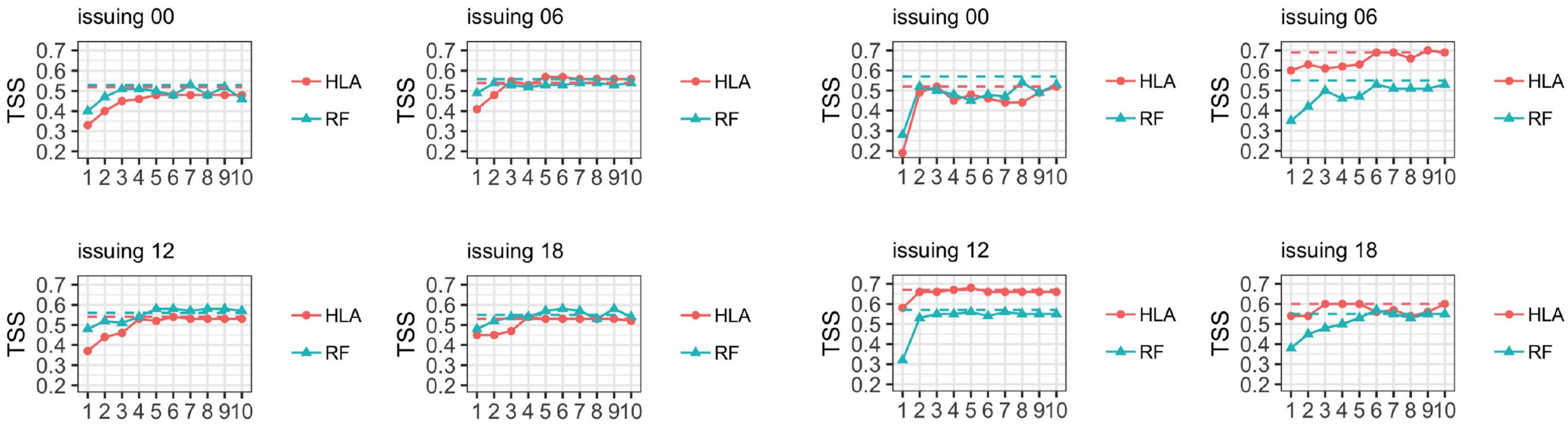
	at least C1 flares						
	Hybrid Lasso	Hybrid Logit	SVC	Random Forest		average	std
flare_index_past	13,98	28,84	19,91	3,51		16,56	10,63
sharp_kw/hgradbh/total	3,47	37	18,59	16,57		18,95	13,87
wlsg_br/value_int	3,74	14,43	22,86	43,14		21,04	16,68
sharp_kw/jz/max	26,05	28	16,94	18,58		22,27	5,29
sharp_kw/usiz/max	24,2	36,75	34,79	18,37		28,53	8,73
wlsg_blos/value_int	3,96	45	46,61	25,39		30,24	20,00
r_value_br_logr	3,52	2,81	128,91	7,47		35,68	62,19
sharp_kw/ggt45fract	14,99	32	57,6	49,3		38,35	19,00
sharp_kw/usiz/stddev	17,15	49,46	54,26	45,89		41,69	16,72
sharp_kw/gamma/total	61,65	20,76	52,95	34,67		42,51	18,35

issuing time: 12:00:00

	at least C1 flares						
	Hybrid Lasso	Hybrid Logit	SVC	Random Forest		average	std
wlsg_br/value_int	3,23	5	23,95	30,02		15,55	13,45
flare_index_past	13,89	31,13	33,92	5,47		21,10	13,68
sharp_kw/usiz/total	36,44	15,8	11,39	26,1		22,43	11,19
sharp_kw/hgradbh/total	29,06	7,55	24,87	39,45		25,23	13,29
sharp_kw/ggt45fract	5,25	17,14	50,68	28,25		25,33	19,33
ising_energy_br/ising_energy	24,14	19,67	26,49	55,1		31,35	16,08
wlsg_blos/value_int	12,83	35,08	41,12	51,11		35,04	16,21
r_value_br_logr	6,52	4,13	126,06	16,87		38,40	58,70
sharp_kw/usiz/stddev	4,67	22,41	69,63	57,61		38,58	30,21
sharp_kw/usiz/max	31,77	44,87	80,57	19,66		44,22	26,33

issuing time: 00:00:00

redundancy of information



TSS scores obtained by using just the 10 top-ten features added one at a time

references

- **guastavino s et al** 2022 operational solar flare forecasting via video-based deep learning *arXiv:2209.05128*
- **marchetti f et al** 2022 score-oriented loss (sol) functions *pattern recognition* **132** 108913
- **guastavino s et al** 2022 bad and good errors: value-weighted skill scores in deep ensemble learning IEEE transactions on neural networks and learning systems
- **campi c et al** 2019 feature ranking of active region source properties in solar flare forecasting and the uncompromised stochasticity of flare occurrence *astrophysical journal* **883** 150
- **guastavino s et al** 2022 implementation paradigm for supervised flare forecasting studies: A deep learning application with video data *astronomy and astrophysics* **662** A105
- **cicogna et al** 2021 flare forecasting algorithms based on high-gradient polarity inversion lines in active regions *astrophysical journal* **915** 38
- **the FLARECAST team** 2021 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* **11** 39
- **benvenuto f et al** 2018 a hybrid supervised/unsupervised machine learning approach to solar flare prediction *astrophysical journal* **853** 90
- **florios k et al** 2018 forecasting solar flares using magnetogram-based predictors and machine learning *solar physics* **293** 1

thank you!

e-mail: **piana@dima.unige.it**

MIDA Group | università di genova, Italy