Precise and Accurate Short-Term Forecasting of Solar Energetic Particle Events with Multivariate Time-Series Classifiers

Sumanth Rotti, Berkay Aydin and Petrus C. Martens

Georgia State University, Atlanta, GA, USA

Contact: srotti@gsu.edu

SUMMARY

- **We implement a data-driven supervised binary classification framework on multivariate time-series data from solar cycles 22-24.**
- **We use ensemble modeling to combine results from three proton (I_p) channels (E ≥10, ≥50, and ≥100 MeV) and the long-band X-ray (1–8 Å) channel of the GOES missions.**
- **The model aims to distinguish strong SEP events from non-events.**
- **SEP event definition is based on the GOES ≥10 MeV integral channel & SWPC 'S1' threshold:**
  - **Strong: $I_p \geq 10$ pfu**
  - **Weak: $0.5 \geq I_p < 10$ pfu**
- **We include "non-SEP" periods following ≥C6.0 flares to introduce a natural class imbalance in the data set.**
- **The 'positive' class comprises 244 strong SEP events. The 'negative' class has 189 weak events and 2,460 SEP-quiet samples.**
- **We experimented with summary statistics, one nearest neighbor, and supervised time-series forest (STSF; Cabello et al. 2020) classifiers.**
- **We compare their performances for prediction windows from 5 to 60 minutes.**
- **STSF performs well under all circumstances.**
- **For a 60-minute prediction window, we get:**
  - **True skill statistic (TSS) = 0.850**
  - **Heidke skill score (HSS) = 0.878**
- **Results assure confidence in our approach.**

DATA

- The strong and weak SEP samples are obtained from the Geostationary Solar Energetic Particle (GSEP) events data set (Rotti et al. 2022).
- For non-SEPs, large flares are identified that do not lead to significant variations in the GOES ≥10 MeV proton fluxes relative to the background.
- C6.0 is chosen as the bottom threshold for non-SEP flares (see Figure 1).

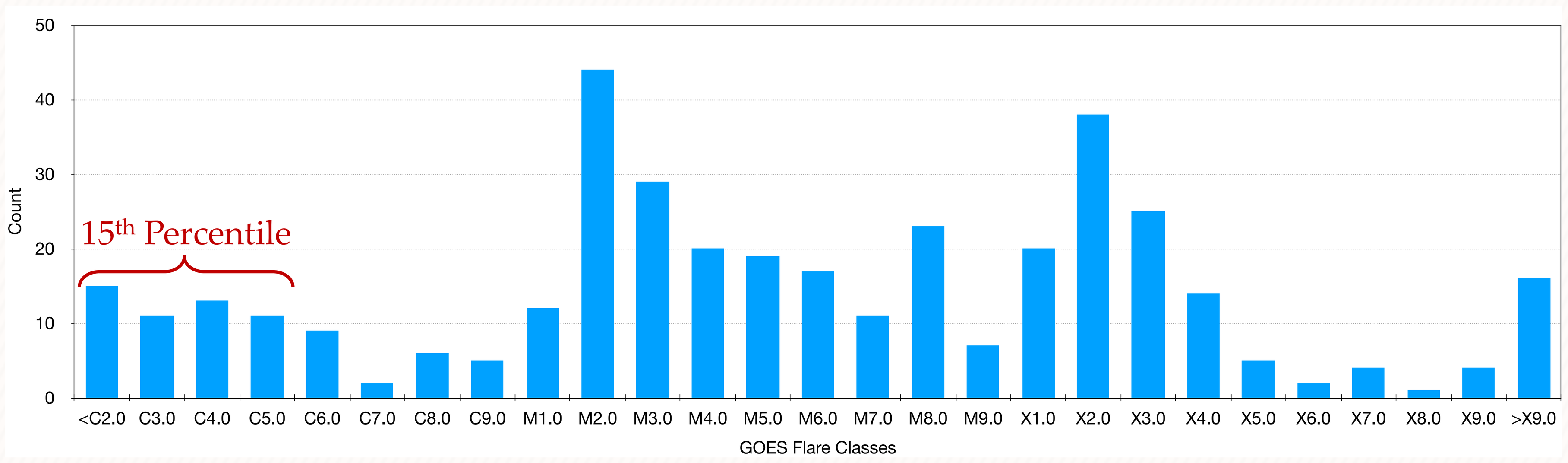

Figure 1: Distribution of 383 SEP-associated flares in the GSEP data set.

- Flares during an ongoing SEP event are not considered, and all consecutive flares within 11 hr of the onset of the flare at consideration are removed.
- 77(±4)% of SEP event onsets occur within 11 hr after the associated flare.
- Total size of the dataset is 2,893 samples, with a class imbalance of ~ 1:11.

METHOD

- Data splitting on nonoverlapping years (# of samples shown in Table 1):
  - **Training** – 1986 to 1992
  - **Validation** – 1993 to 2002
  - **Test** – 2003 to 2018

Table 1. Data Set Partitioning

	Train	Validation	Test
Positive	80	80	84
Negative	918	894	837

- Proton fluxes are interpolated to 1 minute to match with the X-ray flux.
- Each input series has a fixed length of 11 hr.
- The problem is framed as a time-series binary classification task, and we address it through multivariate time-series classification approaches.
- The model (late fusion) schema is obtained from Rotti et al. (2024) for short-term prediction windows of 5, 15, 30, 45, and 60 minutes.
- It is a parameter-wise ensemble of columns in which individual classifiers are applied to every parameter (column).
- The classifiers extract statistical features from input time-series intervals.
- Model Evaluation Metric:

1) $$\text{TSS} = \frac{TP}{(TP+FN)} - \frac{FP}{(FP+TN)}$$

2) $$\text{HSS} = \frac{2\left((TP\,.\,TN)-(FN\,.\,FP)\right)}{(TP+FN)\,(FN+TN)+(TN+FP)\,(TP+FP)}$$

RESULT

- We find optimal classification thresholds for each model and assess their performance. Figure 2 compares TSS model performance.

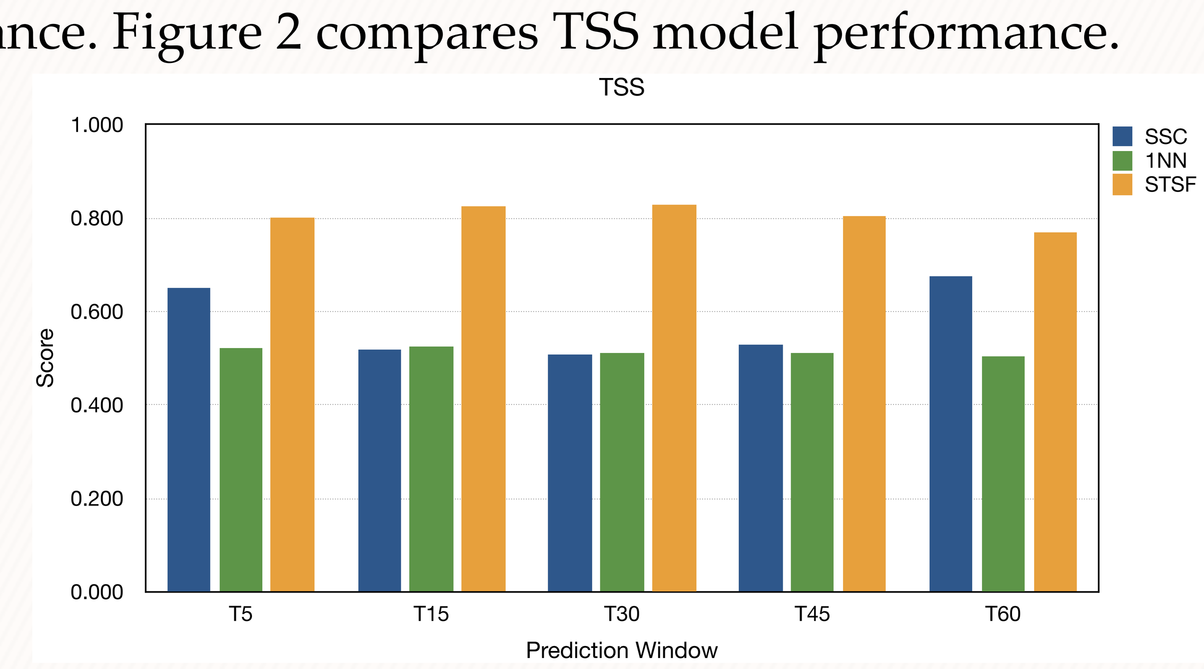

Figure 2: Model comparison for TSS variation on validation set.

- Thresholds are determined by analyzing ROC curves for each model. Figure 3 illustrates the ROC curve for the STSF classifier.

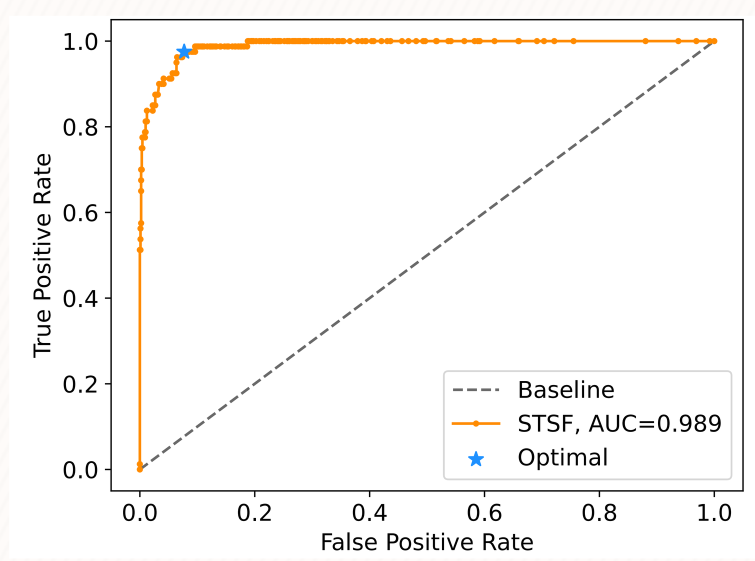

Figure 3: ROC curve for the best model.

Table 2. Contingency Table for STSF.

		Predicted	
		Strong	Weak
True	Strong	72	12
	Weak	6	831

- Table 2 summarizes the STSF model's contingency table on the test set.
- A comparison of the skill scores with different prediction windows on the unseen test set for STSF is presented in Figure 4.

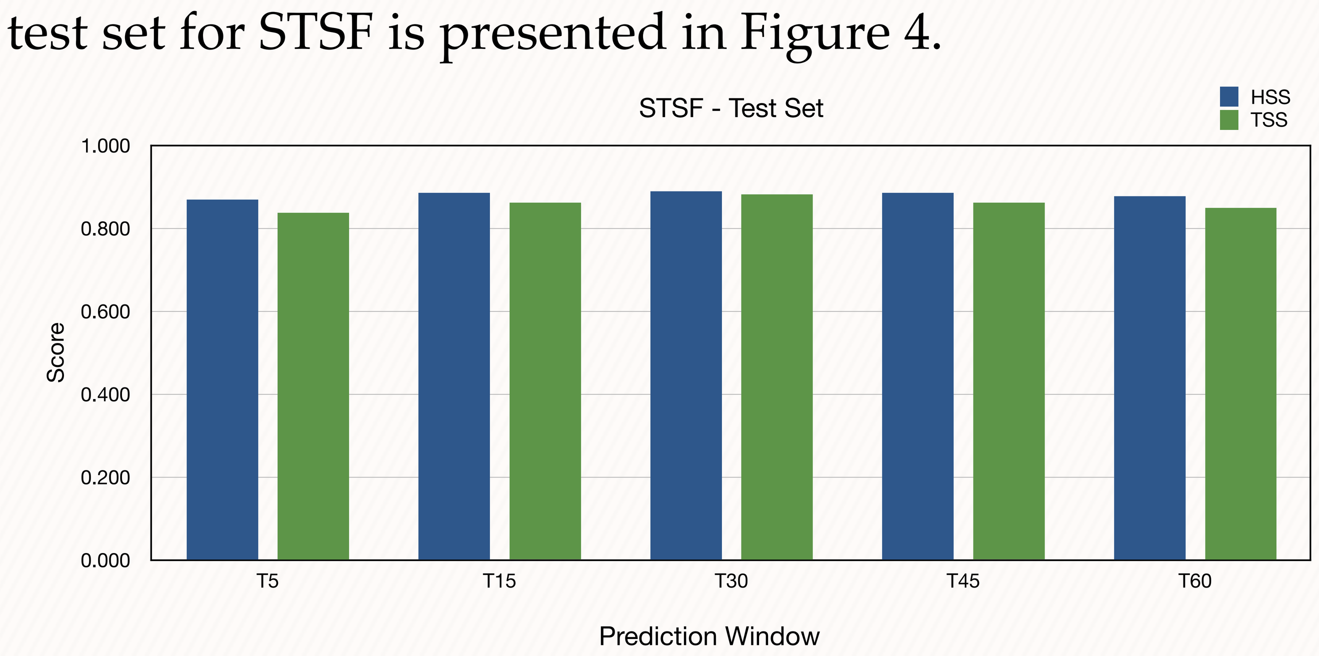

Figure 4: Comparison of TSS and HSS on the test set.

- Scores for T_{60} marginally reduced (<2%) compared to T_5.
- A detailed discussion of the results has been published in ApJ and can be accessed via the QR code here.

Acknowledgments

We acknowledge the use of data from the NOAA-GOES missions. Our work was supported by NASA's FINESST (80NSSC21K1388) and SWR2O2R (80NSSC22K0272) grants.

References

Cabello N., Naghizade E., Qi J. and Kulik L. 2020 *2020 IEEE Int. Conf. on Data Mining (ICDM)* (Piscataway, NJ: IEEE) 948
Rotti S., Aydin B., Georgoulis M. K. and Martens P. C. 2022 *ApJS* 262 29
Rotti S. A., Aydin B. and Martens P. C. 2024 *ApJ* 966 165