

Enhancing Solar Flare Prediction with Innovative Data-Driven Labels

Jinsu Hong, Anli Ji, Chetraj Pandey, and Berkay Aydin

Department of Computer Science

Georgia State University, Atlanta, USA

{ jhong36, aji1, cpandey1, baydin2 }@gsu.edu

Abstract—Solar flare prediction presents a significant challenge in space weather forecasting. Currently, existing solar flare prediction tools heavily rely on the GOES classification system. These tools commonly use the maximum X-ray flux measurement within a specific prediction window, often set at 24 hours, as a basis for labeling instances. However, the background X-ray flux experiences considerable fluctuations throughout the solar cycle, leading to misleading outcomes during solar minimum and an increase in false alarms. To address this issue, we propose a new set of solar flare intensity labels computed from GOES X-ray flux, with the aim of enhancing the accuracy of flare prediction methods. Our approach involves innovative labeling methods that take into account relative increases and cumulative measurements across prediction windows. In this paper, we introduce the concept of the 'relative X-ray flux increase' and provide an explanation of how this metadata is derived. Additionally, we present new cumulative indices and data-driven categorical labels specifically designed for active region-based and full-disk flare prediction models. We also assess the feasibility of integrating our new labels into established solar flare prediction models. Our findings demonstrate that these data-driven labels can be valuable supplements to existing techniques and their integration has the potential to enhance the effectiveness of solar flare prediction tools.

Index Terms—Solar Flares, Metadata, Space Weather Analytics

I. INTRODUCTION

Space weather encompasses various environmental conditions in the geospace, including the Sun, solar wind, magnetosphere, ionosphere, and thermosphere. Among these space weather events, solar flares play a crucial role and can have significant consequences on Earth and its surrounding environment, particularly when combined with other phenomena such as coronal mass ejections (CMEs). The impacts of solar flares extend to power grid outages, disruption of navigation and increased radiation levels, posing health risks to astronauts during space missions. Therefore, accurate prediction of solar flares and associated events is of utmost importance in order to provide timely alerts and mitigate the negative impacts resulting from severe space weather events.

Solar flares are typically identified using X-ray flux data collected by the Geostationary Operational Environmental Satellites (GOES). Solar flare detection involves a heuristic method that relies on a sustained and substantial increase in the average X-ray flux over time. This increase eventually levels off, and the flux measurements typically return to their

background levels. These flares are classified based on their highest X-ray flux (peak X-ray flux) readings within specific wavelength bands, between 1 and 8 Angstroms (0.1-0.8 nm). There are five main classes, denoted as A, B, C, M, and X, according to the GOES classification system. Each class is further divided into subclasses with values ranging from 1.0 to 9.9.

The current GOES flare index, however, has some limitations. Firstly, quantifying the magnitude of X-ray flux for individual active regions, which are areas on the sun with high magnetic flux, is not currently feasible. X-ray flux measurements are obtained globally from the Sun [1]. Relying solely on global X-ray flux values, particularly when using only the maximum (peak) values within a specific period, can be misleading since they do not accurately represent the radiation emitted from individual active regions. Moreover, the background X-ray flux fluctuates over the solar cycles, and the GOES classification system cannot reflect the variation of the Sun. For example, during solar maximum when solar activity is intense, the background X-ray flux is high due to the increased number of active regions. Solar flares are more likely to be categorized into higher classes (M- or X- classes) than flares during solar minimum. To address this issue, we propose a more comprehensive approach that takes into account background conditions and cumulative indices. This includes considering the sum of relative X-ray flux increase values and the absolute GOES subclass, thereby enhancing the predictive capabilities of traditional models.

In this study, we aim to enhance the current flare labeling methods by introducing innovative data-driven techniques that incorporate additional sources of metadata. These include the relative X-ray flux increase compared to background X-ray flux, as well as the active region-based cumulative flare index and full-disk cumulative flare index. The cumulative flare indices are derived from both absolute X-ray flux and relative X-ray flux increase values. To generate the new flare labels, we utilized a validated solar flare list from the SWAN-SF benchmark dataset [2]. These new labels complement the existing ones and provide a more comprehensive understanding of flare activity. Additionally, we developed and assessed new prediction models that can estimate flare intensities relative to background flux, moving beyond the reliance solely on absolute X-ray flux measurements and classifications.

The paper is structured as follows: Section II provides an overview of related work and discusses the the current labeling system. In Section III, we introduce our methodology and explain the process of generating new solar flare labels. Section IV presents a case study that evaluates the effectiveness of utilizing these novel flare labels. Lastly, Section V offers concluding remarks and outlines potential future work.

II. RELATED WORK

The National Oceanic and Atmospheric Administration (NOAA)/GOES has maintained a comprehensive record of detected flares since 1975. This record includes various attributes like the GOES class, peak X-ray flux, spatial location on the solar disk, NOAA active region number, and temporal details (start, peak, and end times) related to the flares [2]. The NOAA's method for detecting flares can be found in their historical documentation [3]. In the development of datasets for data-driven solar flare prediction methods, a fixed time period, often referred to as the prediction window or forecast horizon, is frequently used for labeling. If multiple flares occur within this prediction window, we typically select the flare with the highest intensity and label it. In situations where no flares are detected within this window, we assign a label of "flare-quiet." For binary classification tasks in solar flare prediction, we apply one or more thresholds to categorize the labels. Typically, a threshold of $\geq M1.0$ (or $\geq C5.0$) is applied, classifying M- and X-class flares as flaring instances, while considering the relatively weaker flares and instances labeled as "flare-quiet" as non-flaring instances.

In the field of solar flare forecasting, two main approaches are commonly utilized. The first approach is known as active region-based models, which focus on data from active regions and can represent them as point-in-time vectors [4], time series [5], or images [6]. This approach often employs time series classification techniques and labels flares based on their association with specific active regions [7]. The second approach is the full-disk models, which provide predictions for the entire solar disk. In full-disk models, all flares are considered, regardless of their association with active regions [8]. It's worth noting that hybrid approaches, which combine elements of both active region-based and full-disk models, also exist in the literature [9]. In these approaches, the primary goal is to predict the occurrence of significant flaring events within a defined prediction window, which typically spans 12, 24, or 48 hours

In both active region-based [7] and full-disk [8] prediction methods, each data instance is assigned the most intense flare observed within the prediction window. Previous research in solar flare prediction has primarily focused on labels generated from the GOES classification, potentially overlooking valuable information within these prediction windows. In our work [10], we introduced a framework for creating data-driven flare indices and expanded the dataset to include flares labeled using data-driven criteria. In this study, we integrate new machine learning-compatible datasets and create relative X-ray flux

increase and cumulative labels for both active regions and full-disk approaches. We evaluate the performance and practicality of these novel labels within active region-based models across a wide range of threshold values and class weight settings. Additionally, we introduce a model aggregation technique for predicting numerical solar indices.

III. METHODOLOGY

A. Relative Increase of Background X-ray Flux

In order to generate labels for the relative X-ray flux increase ($rxfi$) of solar flares, we utilize 1-minute averaged GOES X-ray flux data obtained from multiple GOES missions within the 0.1-0.8 nm passband. The determination of the background X-ray flux for a specific flare involves examining the preceding 24-hour period leading up to the flare's start time. This period spans from the flare start time to 24 hours before the start time. The background X-ray flux is computed as the mean of the X-ray flux measurements collected during that previous day. To ensure the accuracy and reliability of the data, we apply exclusion criteria to the intervals representing the background X-ray flux in three cases: (1) Intervals that fall between the start and end times of other flares are omitted. (2) X-ray flux measurements exceeding the initial X-ray flux at the start time of the flares are excluded. (3) Measurements identified as low quality by the instrument are also disregarded. These exclusion rules primarily aim to enhance the authenticity and veracity of the data. An example of this is the exclusion of increased X-ray flux measurements during flare occurrences, as they can significantly deviate, typically by 1-3 magnitudes, from the actual background.

After applying the aforementioned cleaning steps, the background X-ray flux is calculated by averaging the valid X-ray flux values within a given period. Subsequently, the relative X-ray flux increase ($rxfi$) is calculated by dividing the peak X-ray flux (P_{xrf}) by the background X-ray flux. The

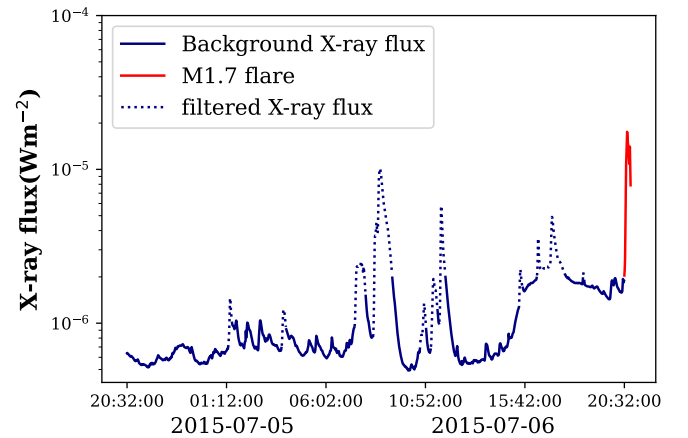


Fig. 1: The graph displays the GOES 1-8 Å Solar X-ray flux for the M1.7 flares, which occurred in active regions NOAA '12381'. The duration for the M1.7 flare on 2015-07-06 is [20:32 to 20:50].

specific equations for calculating the background X-ray flux and relative X-ray flux increase can be found in equation 1 and equation 2, respectively.

$$bgxf(t_{s.t.}) = \frac{\sum_{t_{s.t.}-24h}^{t_{s.t.}} xrf(t_i)}{N'} \quad (1)$$

$$rxfi = \frac{Pxf}{bgxf(t_{s.t.})} \quad (2)$$

To illustrate this process, we provide a running example in Figure 1, showcasing an M-class flare with a respective start time of 2015-07-05T20:32. The dotted blue line displays the filtered X-ray flux based on three rules. The cleaned and calibrated background X-ray flux is represented by the solid blue line. Please note that this example excludes Rule 3, which targets invalid values due to instrumental issues and removes them. The background X-ray flux is calculated by averaging the valid data points, as indicated by the blue line in Figure 1. In this particular example, we obtain a background X-ray flux of $8.8 \times 10^{-7} Wm^{-2}$ for the M1.7 flare. With the peak X-ray flux of the M1.7 flare at $1.7 \times 10^{-5} Wm^{-2}$, we calculate the relative X-ray flux increase ($rxfi$) as $rxfi = \frac{1.7 \times 10^{-5}}{8.8 \times 10^{-7}} = 19.32$. These updated relative X-ray flux increase values are applied to augment the solar flare catalog from NOAA, and the $rxfi$ values are utilized in our new labeling approaches.

B. Data-driven Labeling for Solar Flares

We have implemented a set of aggregated indices to create our data-driven labels. These indices include:

- 1) GC^{Max} : The GOES class of the flare with the maximum intensity within a given prediction window (24 hours in our study).
- 2) $rxfi^{Max}$: The flare with the highest $rxfi$ value within the prediction window.
- 3) GC^Σ : The weighted sum of the GOES subclass values within the prediction window, calculated as $\sum C_i + 10 \times \sum M_j + 100 \times \sum X_k$.
- 4) $rxfi^\Sigma$: The sum of the $rxfi$ values within the prediction window.

GC^{Max} represents the existing labels that we aim to enhance. Each time point or time series is assigned labels based on the four different criteria mentioned above. We employ the sliding window technique to create data instances and generate labels by iterating over the time series with a specified step size. In Figure 2, we present a diagram illustrating the sliding window with a 12-hour observation period and a 24-hour prediction window. The observation window stores a time series, and the prediction window is used to label the time series within the observation window. We generate new indices for both individual active regions (ARs) and the full disk. In the active region-based approach, for each time point (t_i), we check if there is a set of flares occurring within the prediction window (from t_i to $t_i + 24$ hours). Based on this set of flares, each time point is then labeled using either the maximum or cumulative indices. It's important to note that we use a 12-hour observation window in our case study (as

presented in Section IV). This means that for a time point t_i , we obtain multivariate time series instances from the time interval $[t_i - 12 \text{ hours}, t_i]$. The sliding windows iterate through the time series with a step size of 1 hour. To illustrate our new labels, we provide an example of a sliding window in Figure 2, encompassing three flares occurring between $T + 14h$ and $T + 25h$. In this example, we generate three distinct labels. For $rxfi^{Max}$, the first sequence is labeled with the highest value of 250.31 (derived from the M2.0 flare). Regarding GC^Σ , there is one M-class flare and one C-class flare in the prediction window; thus, the index for GC^Σ is 27.1 ($20 \times 10 + 1.6 \times 1 = 21.6$). Finally, for $rxfi^\Sigma$ at Seq 1, the calculated value is 262 ($250.3 + 11.7$). Moreover, the sliding window moves in 1-hour steps and labels the time intervals or points. We further utilized 10 different thresholds for binary classification. The thresholds for numerical labels $rxfi^{Max}$, GC^Σ , and $rxfi^\Sigma$ are listed in Table 1, with a step size of 10. For GC^{Max} , we used a step size based on the subclass value of 2.5. The purpose of varying thresholds is to identify the optimal ones and understand their impact on our models. These three distinct labels are applicable to either a time series of 12-hour observations or a solar full-disk image at a specific time point. The sliced time series dataset and full-disk labels are available in the data repository [11].

TABLE I: Imbalance ratio of the labels

THR($\geq$)	$rxfi^{Max}$	GC^Σ	$rxfi^\Sigma$	THR($\geq$)	GC^{Max}
10	62	79	46	C1	13
20	122	156	92	C2.5	32
30	187	218	137	C5.0	59
40	229	280	192	C7.5	80
50	331	328	263	M1.0	104
60	394	419	296	M2.5	239
70	620	657	405	M5.0	406
80	674	938	494	M7.5	1156
90	711	1030	565	X1.0	1622
100	909	1216	645	X2.5	8374

IV. CASE STUDY: FLARE PREDICTION WITH DATA-DRIVE LABELS

A. Data collection

To assess the impact of the newly introduced labels, we conducted an analysis using an interval-based time series classifier called Time Series Forest (TSF). Our baseline approach is active region-based and is implemented using the SWAN-SF dataset [2], which is a multivariate time series dataset comprising 24 magnetic field parameters, covering the period from 2010 to 2018. For our solar flare prediction case study, we selected six magnetic field parameters, namely USFLUX, TOTUSJZ, TOTUSJH, ABSNJZH, SAVNCP, and TOTPOT. To ensure robust evaluation, we partitioned the SWAN-SF dataset using the tri-monthly partitioning technique introduced in [12]. This led to the creation of four partitions, with each partition covering three months of data from the entire dataset: Partition 1 (January to March), Partition 2 (April to June), Partition 3 (July to September), and Partition 4 (October to December). In our study, we used Partition 4 for testing and

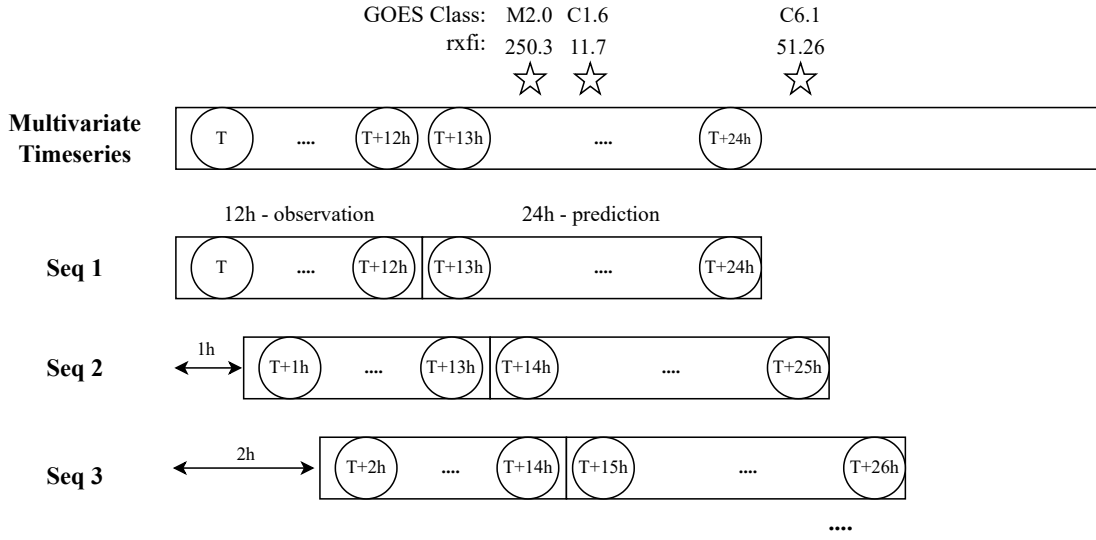


Fig. 2: Diagram of the sliding window method

the other partitions for training the TSF classifier (details of dataset preparation can be found in [7]).

B. Classification

Interval-based time series classification algorithms employ descriptive features, often in the form of statistical measures, derived from fixed or random intervals within each time series. These features, calculated over specific intervals, can capture crucial characteristics of the series. However, considering all possible intervals becomes impractical as the feature space grows exponentially, leading to computational challenges. To address this issue, Deng et al. [13] proposed the time series forest (TSF) approach, which employs a random forest ensemble to train multiple decision trees using a subset of statistical features from randomly selected intervals. These features include simple yet effective measures like mean, standard deviation, and slope. This approach reduces the high feature space, enabling efficient classification for time series data. Originally designed as a univariate classifier, TSF constructs random forests for each parameter individually. In our case, where we analyze multivariate time series from active region patches, we adopted a column ensemble technique to handle multiple parameters. Each parameter is independently fitted with a TSF classifier, and the outputs from these classifiers are aggregated to make the final prediction using equal voting based on prediction probabilities. Notably, we set the number of estimators to 50 and the maximum tree depth to 3 for each univariate time series forest. Further implementation details can be found in our project repository [14].

C. Model evaluation

To assess the performance of the models, we employed a 2x2 confusion matrix and utilized widely used forecast skill scores in solar flare prediction: True Skill Statistics (TSS,

shown in equation 3) and Heidke Skill Score (HSS, shown in equation 4).

$$TSS = \frac{TP}{TP + FN} - \frac{FP}{FP + TN} \quad (3)$$

$$HSS = 2 \times \frac{TP \times TN - FN \times FP}{((P \times (FN + TN) + (TP + FP) \times N))} \quad (4)$$

, where $P = TP + FN$ and $N = FP + TN$. In this scenario, P refers to strong flaring classes, while N refers to relatively small and flare quiet regions. TP , FP , FN , and TN stand for true positive, false positive, false negative, and true negative, respectively. To address the class imbalance issue, we explored ten different class weights (e.g., 1:1, 1:10, 1:15, ..., and 1:50) and 10 thresholds (e.g., 10, 20, ..., 100) to optimize the model's performance.

The results from the AR-based models using four different labels are presented in Figure 3. Our observations indicate that the utilization of GC^{Max} results in slightly higher TSS and HSS scores in comparison to using $rxfi^{Max}$, as illustrated in Figure 3. As the class weight increases, we observe an increment in the TSS scores, but a simultaneous decrease in the HSS scores across all labels. This phenomenon arises because the rise in true positives enhances the TSS scores, whereas the surge in false positives diminishes the HSS scores. Moreover, when we apply larger thresholds for labeling, it amplifies the imbalance ratio, rendering the prediction tasks more demanding and subsequently leading to a reduction in the overall skill scores.

In Figure 4, we present a comparison of models using four different labels. Unlike the traditional ROC curve, where each point represents a model, our representation includes four hundred points, each associated with specific class weights and thresholds. Models that utilize cumulative indices, such as GC^{Σ} and $rxfi^{\Sigma}$, appear to deliver superior overall performance. However, making a direct comparison between the four

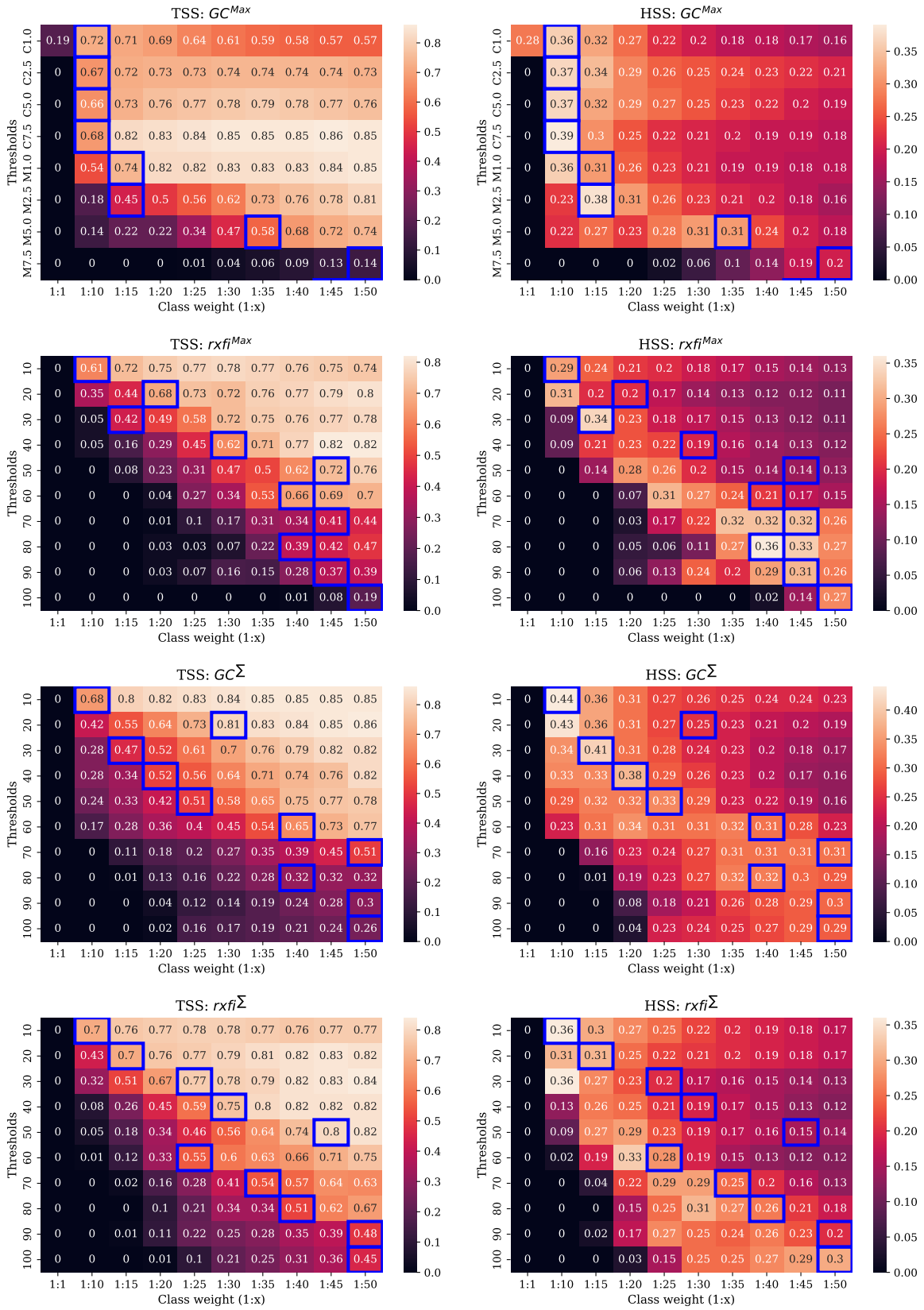


Fig. 3: The results of time series classifiers trained with different labels. Within each label, there are 10 different thresholds and class weights. From the score matrices, we select candidate models highlighted by blue squares. The geometric mean is used to choose the candidate models among different class weights.

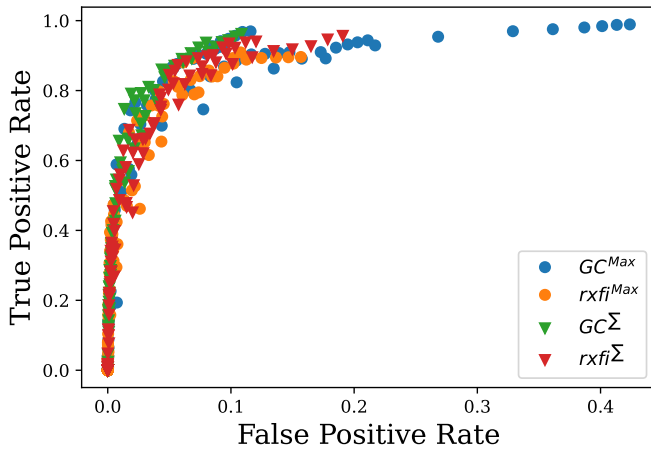


Fig. 4: A plot of the true positive and false positive rates for all the models. Each model, characterized by different labels, thresholds, and class weights, is represented by a single data point.

labels is challenging due to their differing imbalance ratios and threshold criteria. In the case of GC^{Max} , its threshold range spans 250, as X2.5 is 250 times larger than C1. On the other hand, the other labels have a threshold range of 10, from 10 to 100. This results in the right side of Figure 4 showcasing GC^{Max} with relatively high false positive rates. Models trained with GC^{Max} that use a threshold of C1.0 exhibit a high false positive rate. This is primarily because the threshold of C1.0 has a lower imbalance ratio, leading to a greater number of positive instances.

The problem of class imbalance is a common issue in solar flare prediction. Nevertheless, the evaluation metrics utilized in this study demonstrate that the introduced labels can improve the performance of flare prediction models beyond random chance. Furthermore, when equipped with suitable thresholds and class weights, models trained with these new labels could potentially surpass current labeling methods. Thus, the results indicate that these proposed labels can be valuable supplements to existing techniques and their integration has the potential to enhance the effectiveness of solar flare prediction tools.

V. CONCLUSION AND FUTURE WORK

In this study, we introduced a novel set of flare labels, which includes relative X-ray flux increase ($rxfi$), maximum relative X-ray flux value ($rxfi^{Max}$), sum of X-ray flux increase values ($rxfi^{\Sigma}$), and weighted sum of GOES subclass values (GC^{Σ}). These new labels were applied to active region-based classification models, and in a preliminary case study, we assessed their effectiveness in solar flare prediction. Our results show that these new labels yield skill scores comparable to well-established techniques, demonstrating their potential as an alternative approach in solar flare forecasting.

Looking ahead, our future research will focus on several areas. Firstly, we aim to investigate the relationship between

these labels and features of coronal mass ejections. Additionally, we plan to gain deeper insights into our models, moving beyond treating them as black boxes to understand their behavior in specific scenarios. This approach will enhance model interpretability and enable more informed decision-making in space weather forecasting. Overall, by expanding our research in these directions, we aim to improve the accuracy and reliability of space weather forecasting systems, leading to more dependable predictions for space weather phenomena.

ACKNOWLEDGMENTS

This work is supported in part under two grants from NSF (Award #2104004) and NASA (SWR202R Grant #80NSSC22K0272).

REFERENCES

- [1] L. Fletcher, B. R. Dennis, H. S. Hudson, S. Krucker, K. Phillips, A. Veronig, M. Battaglia, L. Bone, A. Caspi, Q. Chen, P. Gallagher, P. T. Grigis, H. Ji, W. Liu, R. O. Milligan, and M. Temmer, "An observational overview of solar flares," *Space Science Reviews*, vol. 159, no. 1-4, pp. 19–106, Aug. 2011.
- [2] R. A. Angryk, P. C. Martens, B. Aydin, D. Kempton, S. S. Mahajan, S. Basodi, A. Ahmadzadeh, X. Cai, S. F. Boubrabimi, S. M. Hamdi, M. A. Schuh, and M. K. Georgoulis, "Multivariate time series dataset for space weather data analytics," *Scientific Data*, vol. 7, no. 1, Jul. 2020.
- [3] S. C. Janet Machol and C. Peck, *User's Guide for GOES-R XRS L2 Products*, 2022 [Online].
- [4] A. Ahmadzadeh, M. Hostetter, B. Aydin, M. K. Georgoulis, D. J. Kempton, S. S. Mahajan, and R. Angryk, "Challenges with extreme class-imbalance and temporal coherence: A study on solar flare data," in *2019 IEEE international conference on big data (Big Data)*. IEEE, 2019, pp. 1423–1431.
- [5] A. Ji, J. Wen, R. Angryk, and B. Aydin, "Solar flare forecasting with deep learning-based time series classifiers," in *2022 26th International Conference on Pattern Recognition (ICPR)*. IEEE, 2022, pp. 2907–2913.
- [6] A. K. Abed, R. Qahwaji, and A. Abed, "The automated prediction of solar flares from sdo images using deep learning," *Advances in Space Research*, vol. 67, no. 8, pp. 2544–2557, 2021.
- [7] A. Ji, B. Aydin, M. K. Georgoulis, and R. Angryk, "All-clear flare prediction using interval-based time series classifiers," in *2020 IEEE International Conference on Big Data (Big Data)*. IEEE, 2020, pp. 4218–4225.
- [8] C. Pandey, R. A. Angryk, and B. Aydin, "Solar flare forecasting with deep neural networks using compressed full-disk hmi magnetograms," in *2021 IEEE International Conference on Big Data (Big Data)*. IEEE, 2021, pp. 1725–1730.
- [9] C. Pandey, A. Ji, R. A. Angryk, M. K. Georgoulis, and B. Aydin, "Towards coupling full-disk and active region-based flare prediction for operational space weather forecasting," *Frontiers in Astronomy and Space Sciences*, vol. 9, p. 897301, 2022.
- [10] J. Hong, A. Ji, C. Pandey, and B. Aydin, "Beyond traditional flare forecasting: A data-driven labeling approach for high-fidelity predictions," in *Big Data Analytics and Knowledge Discovery*. Springer Nature Switzerland, 2023, pp. 380–385. [Online]. Available: https://doi.org/10.1007/978-3-031-39831-5_34
- [11] Hong, Jinsu and Ji, Anli and Pandey, Chetraj and Aydin, Berkay, "A data-driven Labels for solar flare predictions." [Online]. Available: <https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/1U2Q3C>
- [12] C. Pandey, R. A. Angryk, and B. Aydin, "Deep neural networks based solar flare prediction using compressed full-disk line-of-sight magnetograms," in *Information Management and Big Data*. Springer International Publishing, 2022, pp. 380–396.
- [13] H. Deng, G. Runger, E. Tuv, and M. Vladimir, "A time series forest for classification and feature extraction," *Information Sciences*, vol. 239, pp. 142–153, Aug. 2013.
- [14] "Source code." [Online]. Available: https://bitbucket.org/gsudmlab/data_driven_labels/src/main/