An Innovative Solar Flare Metadata Collection for Space Weather Analytics

Jinsu Hong, Chetraj Pandey, Anli Ji, and Berkay Aydin
Department of Computer Science, Georgia State University, Atlanta, USA
{jhong36, cpandey1, aji1, baydin2 }@gsu.edu

***Abstract*—Space weather events can have a significant impact on electric systems and health, with solar flares being one of the central events in space weather forecasting. However, existing solar flare prediction tools heavily rely on the Geostationary Operational Environmental Satellites (GOES) classification system, using maximum X-ray flux measurements as proxies to label instances. This approach becomes problematic during solar minimum, where background X-ray flux fluctuations lead to false alarms and inaccurate predictions. To address this issue, we propose a new collection of solar flare intensity labels computed from GOES X-ray flux, introducing innovative labeling regimes that incorporate relative increases and cumulative measurements over prediction windows. Our goal is to improve the accuracy of flare prediction methods by reducing false positives and enhancing overall prediction performance. Throughout this paper, we introduce the concept of relative X-ray flux increase and explain how to derive relative X-ray flux increase metadata for generating new labels. Additionally, we present new cumulative indices and data-driven categorical labels designed for active region-based and full-disk flare prediction models. We then evaluate the effectiveness of our new labels when applied to established solar flare prediction models, demonstrating that they significantly enhance prediction capabilities and complement existing efforts. With our innovative data-driven labels, we aim to enhance flare forecasting capabilities and provide more accurate and reliable predictions for space weather phenomena.**

***Index Terms*—Solar Flares, Metadata, Space Weather Analytics**

I. Introduction

Space weather often refers to conditions and events occur in the geospace environment includes the area surrounding Earth and the Sun's magnetosphere, ionosphere, and thermosphere. Among these events, solar flares hold significant importance, especially when combined with other eruptions like coronal mass ejections (CMEs). The impacts of such events on Earth and near-Earth space are diverse and can range from power grid outages to disruption of navigation and positioning satellites. Additionally, increased radiation levels at high altitudes or during space missions pose health risks to astronauts. Consequently, accurate prediction of solar flares and associated events becomes crucial to mitigate potential negative effects caused by severe space weather events.

Solar flares are commonly identified using X-ray flux data obtained from Geostationary Operational Environmental Satellites (GOES). A heuristic approach is employed to detect such flares, triggered by a continuous and significant increase in the average X-ray flux series. This increase eventually reaches a plateau, and the flux measurements then return to background levels. The peak X-ray flux (*pxf*) of a flare is defined as the maximum X-ray flux measured between the start and end time of the event. The classification of the flares is based on their observed peak X-ray flux, measured in the 1-8Å passbands. Solar flares are categorized into five major classes based on GOES classification: X, M, C, B, and A [1]. These classes are defined in a logarithmic manner, with X representing the highest flare intensity and corresponding to GOES flux in excess of $10^{-4}Wm^{-2}$ at Earth. These classes are further divided into sub-classes, which are ranked from 1.0 to 9.9. For instance, a C5.7 flare corresponds to approximately $5.7 \times 10^{-6}Wm^{-2}$ X-ray flux.

Solar active regions are areas of the Sun with high concentrations of magnetic flux, and they play a crucial role in driving Earth-impacting events such as flares or coronal mass ejections [2] [3]. The active region-based categorical flare predictions often involve assigning one of two labels to each instance based on a user-defined threshold such as "flaring" or "non-flaring." Commonly, active regions associated with M- or X-class flares are considered as flaring, while regions with lower intensity flares or flare-quiet regions are considered non-flaring. While there are some extensions and alternatives [4], the most commonly adopted flare prediction schema uses discretized labels derived from maximum intensity flares and considers the flare productivity of individual active regions [5].

The current flare labeling techniques rely mainly on the peak X-ray flux of the largest flare in the prediction window, and they suffer from three main limitations. Firstly, quantifying the magnitude of X-ray flux for each active region is not feasible, as X-ray flux measurements are global [1]. This global representation can be misleading, as it does not accurately reflect the emitted radiation from individual active regions. Consequently, in a common binary flare prediction scheme, this misrepresentation may lead to active regions being misclassified. Secondly, relying solely on the maximum intensity of a flare in the prediction window disregards essential factors such as the background X-ray flux, integrated flux, and information on smaller flares. This approach overlooks critical contextual information that could significantly improve prediction accuracy. Furthermore, the use of empirical thresholds, such as $\geq$M1.0 or $\geq$C5.0 [6], to differentiate between flaring and non-flaring instances can further diminish the generalization capabilities of prediction models.

To address these issues, we expand upon the existing flare labeling methods by introducing novel, data-driven techniques that incorporate additional information, including the relative X-ray flux increase in relation to background X-ray flux, active region-based cumulative flare index, and full-disk cumulative flare index. The cumulative flare indices are generated based on absolute X-ray flux and relative X-ray flux increase values. Relative X-ray flux increase allows us to consider background X-ray flux, while cumulative indices enable us to take the effect of small flares (which can be shadowed by large flares) into consideration. These new labels are expected to reduce the number of false positives in our operational predictions, boost the existing forecasts, and improve the overall performance. To test the feasibility of the new flare labels, we trained and evaluated prediction models for active region-based and full-disk flare prediction approaches. We demonstrate that the novel labels complement the existing ones and provide a more comprehensive view of flare activity.

The rest of the paper is organized as follows: In Sec. II, we present related work in solar flare prediction and demonstrate existing approaches. In Sec. III, we present our methodology and describe in detail the process of generating new solar flare labels based on GOES X-ray flux data. In Sec. IV, we present a case study that demonstrates the feasibility of using these newly introduced labels in solar flare prediction tasks. We discuss the results obtained from the case study and analyze the performance of our proposed approach. Finally, in Sec. V, we conclude the paper with a summary of our findings and discuss potential avenues for future research in the field of solar flare prediction using data-driven labels.

II. Related Work

The National Oceanic and Atmospheric Administration (NOAA)/GOES has maintained a comprehensive catalog of detected flares since 1975, providing valuable features such as the GOES class, peak X-ray flux, spatial location on the solar disk (when available), NOAA active region number, and temporal information (start, peak, and end times) related to the flares [7]. When creating datasets for data-driven solar flare forecasting methods, a fixed-size interval known as the prediction window is often used to determine future flare events. Within this window, typically, the maximum intensity flare is selected and labeled, or if no flares are detected, a flare-quiet label is assigned. Binary classification tasks use flaring thresholds such as $\geq$M1.0 (or $\geq$C5.0) to differentiate between flaring (M- and X-class flares) and non-flaring instances (B- and C-class flares or flare-quiet instances).

In the field of solar flare forecasting, two main approaches are commonly used. The first one is referred to as active region-based models, which focus on data from active regions, represented as point-in-time vectors [8], time series [9], or images [10]. In this approach, time series classification techniques are often employed, and flares are labeled based on their association with specific active regions [11]. The second approach involves 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 [6]. Hybrid approaches also exist in the literature [12]. In these approaches, the primary objective is to predict the occurrence of significant flaring events within a defined prediction window, typically spanning 12, 24, or 48 hours.

In both active region-based [11] or full-disk [6] predictions, each instance is labeled with the maximum intensity flare in the prediction window, and previous studies in solar flare prediction primarily focus on labels derived from GOES classification, which can overlook pertinent information within the prediction windows. In [13], we presented a framework to create data-driven flare indices and introduced an extended dataset of flares with their associated data-driven labels. In this study, we employ these labels to predict flaring activity under both active region-based and full-disk flare prediction modes with different threshold settings. As part of our methodology, we integrate new machine learning-ready datasets and create new individual and cumulative labels associated with both active regions and full-disk data instances. We explore the performance and operational feasibility of these new labels with active region-based and full-disk predictions, considering a wide spectrum of threshold values and class weights.

III. Methodology

A. Relative Increase of Background X-ray Flux

Relative X-ray flux increase (*rxfi*) refers to the relative enhancement of X-ray flux with respect to the background X-ray flux. This index is created for a flare event and represents the ratio between the peak (maximum) X-ray flux (*Pxrf* and background X-ray flux (*bgxf*). To generate the *rxfi* index for solar flares, we use 1-minute averaged GOES X-ray flux data collected from multiple GOES missions and define the background X-ray flux for a specific flare. We consider the X-ray flux measurements from the previous 24 hours prior to the flare's start time. The calibrated background X-ray flux is calculated as the mean of the 24-hour X-ray flux measurements before the flare's start time. The formulas for calculating the background X-ray flux and *rxfi* are as follows:

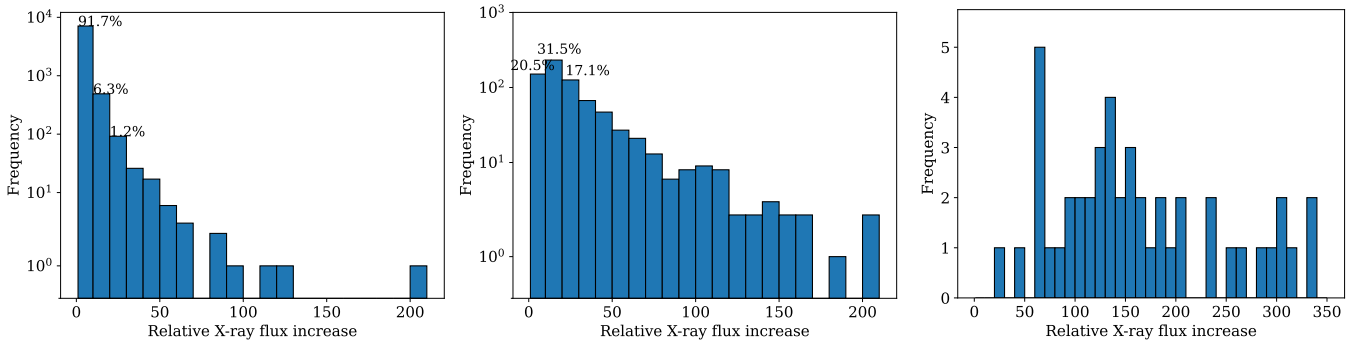
$$bgxf(t_{s.t.}) = \frac{\sum_{t_{s.t.}-24h}^{t_{s.t.}} xrf(t_i)}{N'} \tag{1}$$

$$rxfi = \frac{Pxrf}{bgxf(t_{s.t.})} \tag{2}$$

In Figure 1, we compare the distributions of *rxfi* values for major GOES-classes (C, M and X) in Solar Cycle 24 (between 2010 to 2019). Majority of the C-class flares have relatively lower *rxfi* values (1-30). We observe a shift towards higher *rxfi* values for M-, and especially, X-class flares as expected.

B. Data-driven Labeling for Solar Flares

As mentioned earlier, we define relative X-ray flux increase (*rxfi*) as the ratio between the peak X-ray flux to background X-ray flux. For each C-, M-, and X-class flare in our list, we generated new *rxfi* labels from their peak X-ray flux and developed cumulative indices for absolute X-ray flux



(a) *rxfi* values in **C** class flares (b) *rxfi* values in **M** class flares (c) *rxfi* values in **X** class flares

Fig. 1: Distribution of relative X-ray flux increase values in each GOES class.

measurements and relative X-ray flux increase. We did not include A or B-class flares as they are severely under-detected due to the background X-ray flux in solar maximum. For each instance, using a 24h prediction window, we integrate relevant flare events and define our labels for data instances as follows:

1) GC^{Max}: The GOES class of the flare with the maximum intensity within a given prediction window.
2) $rxfi^{Max}$: The flare with the highest *rxfi* value within the prediction window.
3) $GC^{\sum}$: 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) $rxfi^{\sum}$: The sum of the *rxfi* values within the prediction window.

GC^{Max}, serves as a baseline and represents the traditional labels, while $rxfi^{Max}$, $GC^{\sum}$, and $rxfi^{\sum}$ are the new labels introduced in this study. To calculate the cumulative GOES class flare indices, we apply a multiplication factor of 10 or 100 to the subclass values of M- or X-class flares, respectively. This adjustment is based on the fact that an M-class flare is 10 times stronger than a C-class flare (for the same subclass), while an X-class flare is 10 times stronger than an M-class flare. Consequently, we multiply the numeric subclass values of X-, M-, and C-class flares by 100, 10, and 1, respectively, and then combine them to obtain the cumulative indices. The cumulative indices serve as valuable indicators of solar activity, taking into account multiple smaller flares. We also note that the label integration is performed for both active region instances and full-disk instances. For full-disk, all flares are considered, while for active region-based predictions, we only consider the associated flares.

IV. Case study: Flare prediction with new labels

A. Data collection

To assess the impact of the newly introduced labels, we conducted a thorough analysis using both the active region-based and full-disk approaches for solar flare forecasting. For the active region-based approach, we employed the SWAN-SF dataset [7]. The SWAN-SF dataset consists of 24 magnetic field parameters, covering the period from 2010 to 2018. In our study, we focused on six magnetic field parameters: USFLUX, TOTUSJZ, TOTUSJH, ABSNJZH, SAVNCPP, and TOTPOT, which are relevant for solar flare prediction. We use a 12-hour observation window that moves in 1-hour steps over the multivariate time series. During each iteration, we check if there is a group of flares associated with the active region within the upcoming 24-hour prediction window. We label each slice with maximum or cumulative indices based on the observation. Regarding the full-disk approach, we utilized twenty-four HMI solar magnetogram images per day, selected hourly, spanning from May 1, 2010, to Dec 30, 2018. Each image is labeled by the subsequent twenty-four hour window. The 24-hour prediction windows are marked by four different labels. To ensure a robust evaluation, we divided our data into four partitions using the tri-monthly partitioning technique introduced in [14]. Each partition covers three months of data over the entire dataset. Specifically, Partition 1 contains data from January to March, Partition 2 from April to June, Partition 3 from July to September, and Partition 4 from October to December. In our study, we used Partition 4 for testing our models, while the other partitions were used for training. The sliced time series dataset and full-disk labels are both available in the data repository [15].

B. Classification

1) Active-region based approach: Time series forest (TSF) approach employs a random forest ensemble technique [16]. TSF trains multiple decision trees with a subset of statistical features derived from randomly selected intervals, including measures like mean, standard deviation, and slope. By utilizing this method, high dimensional feature spaces are effectively reduced, enabling efficient classification of time series data. Originally designed as a univariate classifier, TSF constructs a random forest for each parameter independently. However, in our case of analyzing multivariate time series from active region patches, we utilize a column ensemble technique to work with multiple parameters simultaneously. This involves fitting each parameter individually with a TSF classifier, and then aggregating the outputs of classifiers using equal voting based on prediction probabilities to form a final prediction.

2) Full-disk approach: The Vision Transformer (ViT) is a deep learning architecture used to analyze full-disk images. It treats the image as a sequence of fixed-size non-overlapping patches, which are then transformed into a 1D vector for input to the Transformer [17]. Unlike traditional convolutional neural networks (CNNs), ViT does not use convolutional layers or pooling operations. Instead, it relies on self-attention mechanisms in the Transformer [18] to capture relationships between different patches and understand the global context of the image. To exploit the pre-trained ViT models and map the magnetogram images (original size – 512x512), a convolutional layer is added to reduce their resolution to 224x224. Additionally, a final linear layer with two output values is appended to the last layer of ViT for binary classification purposes. Additionally, in our ViT-based model, we use Stochastic Gradient Descent (SGD) as the optimizer and focal loss as the loss function. Focal loss is a modified version of cross-entropy loss that puts more emphasis on learning from hard misclassified examples [19]. This setup allows ViT to efficiently analyze and classify full-disk images. For further implementation details, readers can refer to our project repository [20].

C. Model evaluation

To evaluate the models, we utilized a 2×2 confusion matrix and computed forecast skill scores commonly used in solar flare prediction: the True Skill Statistics (TSS, shown in Eq. (3)) and the Heidke Skill Score (HSS, shown in Eq. (4)).

$$TSS = \frac{TP}{TP+FN} - \frac{FP}{FP+TN} \tag{3}$$

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

where P = TP + FN and N = FP + TN. In these equations, P represents instances of strong flaring classes, while N represents relatively smaller and flare-quiet regions. TP, FP, FN, and TN correspond to 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 ten thresholds (e.g., 10, 20, ..., 100) for new labels to optimize the models' performance. For GC^{Max}, we used a step size based on the subclass value of 2.5. The imbalance ratios present in our experiments are shown in Table 1.

TABLE I: Imbalance ratio of the labels

THR($\geq$)	***rxfi*** ^{Max}	***GC*** $^{\Sigma}$	***rxfi*** $^{\Sigma}$
10	62	79	46
20	122	156	92
30	187	218	137
40	229	280	192
50	331	328	263
60	394	419	296
70	620	657	405
80	674	938	494
90	711	1030	565
100	909	1216	645

THR($\geq$)	***GC*** ^{Max}
C1	13
C2.5	32
C5.0	59
C7.5	80
M1.0	104
M2.5	239
M5.0	406
M7.5	1156
X1.0	1622
X2.5	8374

1) AR-based Models: The results of the AR-based models using four different labels are visualized as a set of heatmaps in Figure 2. Our findings indicate that using GC^{Max} leads to slightly higher TSS and HSS scores compared to using $rxfi^{Max}$. As the class weight increases, the TSS scores increase, but the HSS scores decrease for all the labels. This is because an increase in true positives boosts the TSS scores, while an increase in false positives lowers the HSS scores. Moreover, larger thresholds for labeling create higher imbalance ratios, making the prediction tasks more challenging and causing overall skill scores to drop.

This case study highlights the challenge of class imbalance in solar flare prediction. Note that direct comparison between the data instances and labels used in this study may not be entirely appropriate. The evaluation metrics used suggest that the proposed labels can enhance the performance of flare forecasting models. Moreover, with optimal thresholds and class weights, the models trained with the new labels have the potential to outperform existing labeling techniques. Therefore, the findings suggest that the proposed labels can be valuable additions to existing techniques, and their combination can improve the capabilities of flare prediction .

2) Full-disk classification: The results of the full-disk models are shown in Figure 3. This figure presents the TSS and HSS scores at the last epoch for models trained with four different labels. In all labels, there is a consistent trend where HSS scores decrease as thresholds increase, as higher thresholds lead to more false positives. We observe the highest nominal TSS scores with *GC* . However, cumulative labels (specifically the GC^{Σ}) provide an overall more flexible prediction performance, where we see a more robust set of forecast skills as the threshold is increasing. Note that, with higher thresholds in GC^{Σ} we capture either a very large event or multiple large events or a combination of the two. In addition, while there are some plausible results, for full-disk models with *rxfi*-based indices, the overall results show relatively lower performance. This is even more evident for the *rxfi* index as we observe fluctuating TSS performance and consistently low HSS.

Considering full-disk models makes use of all visible active regions, the overall higher performance of cumulative indices are expected and they can be used for full-disk predictions. Nevertheless, we note that the results from different labels are not directly comparable, due to background flux fluctuations across different phases of solar cycles.

V. Conclusion and Future Work

In this research, we introduced a novel set of flare labels, including 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 labels serve as valuable additions to existing ones in solar flare prediction. We integrated these new labels to both active region-based and full-disk classification models. Through a preliminary case study, we evaluated the effectiveness of these new labels in solar flare prediction. Our findings

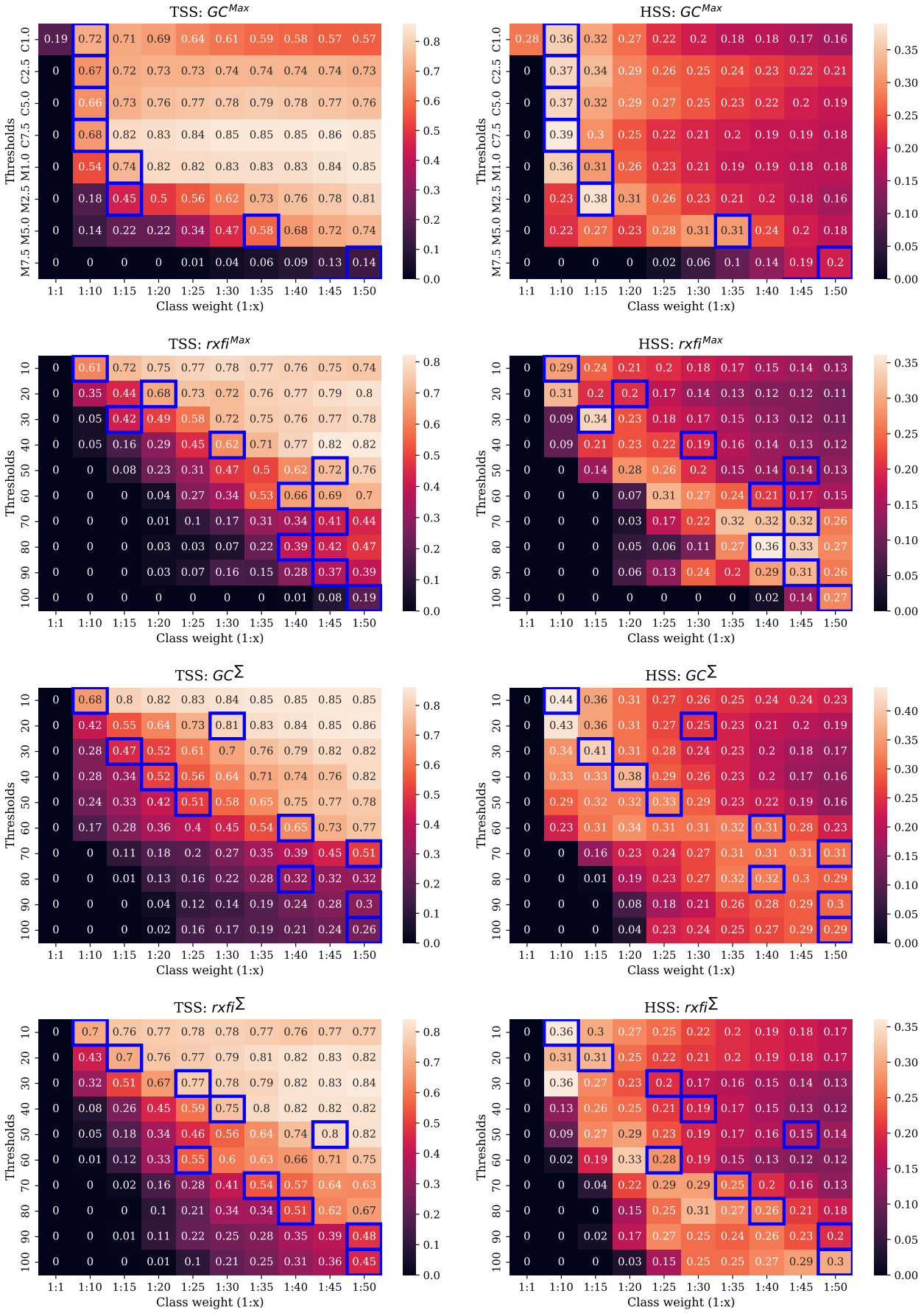
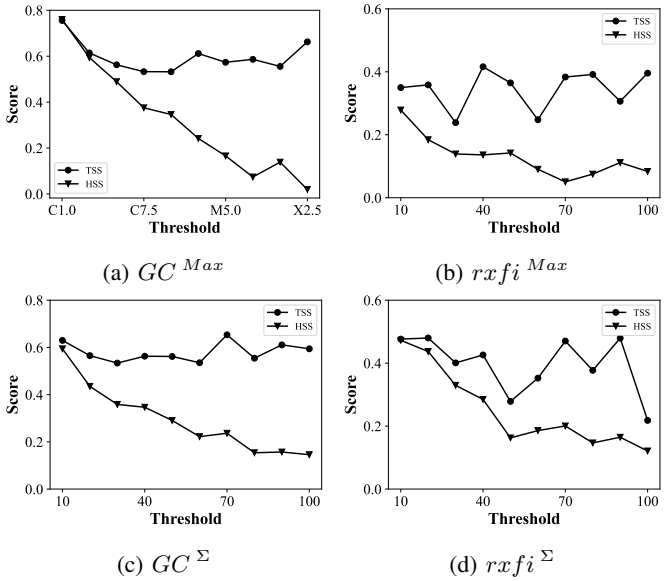


Fig. 2: Results of the AR-based model with four different labels are presented. Candidate models are highlighted with blue boxes. The geometric means between TSS and HSS with the same class weight are calculated to select the best class weight, so the scores marked with blue boxes have the highest geometric mean among the ten class weights.



(a) $GC^{\,Max}$ (b) $rxfi^{\,Max}$

(c) $GC^{\,\Sigma}$ (d) $rxfi^{\,\Sigma}$

Fig. 3: Full-disk model results

indicate that these new labels produce plausible skill scores to well-established techniques, demonstrating their potential as an alternative approach in solar flare forecasting.

Moving forward, there are several avenues for future research. Firstly, we plan to investigate multi-class classification and regression tasks using new indices. Furthermore, understanding eruptive flares, which are flares associated with coronal mass ejections, will be a crucial aspect of our future work. Additionally, we aim to delve deeper into our models, understanding their behavior in specific cases rather than treating them as black boxes. This will enable us to improve model interpretability and make more informed decisions in space weather forecasting. Overall, by extending our research in these directions, we aim to enhance the accuracy and robustness of space weather forecasting systems, creating more reliable predictions for space weather phenomena.

ACKNOWLEDGMENTS

This work is supported in part under two grants from NSF (Award #2104004) and NASA (SWR2O2R 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. [Online]. Available: https://doi.org/10.1007/s11214-010-9701-8

[2] B. J. Thompson, J. Qiu, N. Lugaz, and D. F. Webb, "Solar flares and coronal mass ejections," pp. 179–220, Apr. 2021. [Online]. Available: https://doi.org/10.1002/9781119815600.ch5

[3] L. M. Winter and K. Balasubramaniam, "Using the maximum x-ray flux ratio and x-ray background to predict solar flare class," *Space Weather*, vol. 13, no. 5, pp. 286–297, May 2015. [Online]. Available: https://doi.org/10.1002/2015sw001170

[4] H. Zhang, Q. Li, Y. Yang, J. Jing, J. T. L. Wang, H. Wang, and Z. Shang, "Solar flare index prediction using SDO/HMI vector magnetic data products with statistical and machine-learning methods," *The Astrophysical Journal Supplement Series*, vol. 263, no. 2, p. 28, Dec. 2022. [Online]. Available: https://doi.org/10.3847/1538-4365/ac9b17

[5] K. Leka, S.-H. Park, K. Kusano, J. Andries, G. Barnes, S. Bingham, D. S. Bloomfield, A. E. McCloskey, V. Delouille, D. Falconer *et al.*, "A comparison of flare forecasting methods. ii. benchmarks, metrics, and performance results for operational solar flare forecasting systems," *The Astrophysical Journal Supplement Series*, vol. 243, no. 2, p. 36, 2019.

[6] 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.

[7] R. A. Angryk, P. C. Martens, B. Aydin, D. Kempton, S. S. Mahajan, S. Basodi, A. Ahmadzadeh, X. Cai, S. F. Boubrahimi, 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. [Online]. Available: https://doi.org/10.1038/s41597-020-0548-x

[8] 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.

[9] 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.

[10] 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.

[11] 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.

[12] 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.

[13] 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

[14] 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. [Online]. Available: https://doi.org/10.1007/978-3-031-04447-2_26

[15] J. Hong, A. Ji, C. Pandey, and B. Aydin, "A data-driven Labels for solar flare predictions," 2023. [Online]. Available: https://doi.org/10.7910/DVN/1U2Q3C

[16] 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. [Online]. Available: https://doi.org/10.1016/j.ins.2013.02.030

[17] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An image is worth 16x16 words: Transformers for image recognition at scale," *CoRR*, vol. abs/2010.11929, 2020. [Online]. Available: https://arxiv.org/abs/2010.11929

[18] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," in *Proceedings of the 31st International Conference on Neural Information Processing Systems*, ser. NIPS'17. Red Hook, NY, USA: Curran Associates Inc., 2017, p. 6000–6010.

[19] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," 2018.

[20] "Source code." [Online]. Available: https://bitbucket.org/gsudmlab/data_driven_labels/src/main/