



OPEN

# Operational prediction of solar flares using a transformer-based framework

Yasser Abdullah<sup>1,2</sup>, Jason T. L. Wang<sup>1,2</sup>✉, Haimin Wang<sup>1,3,4</sup> & Yan Xu<sup>1,3,4</sup>

Solar flares are explosions on the Sun. They happen when energy stored in magnetic fields around solar active regions (ARs) is suddenly released. Solar flares and accompanied coronal mass ejections are sources of space weather, which negatively affects a variety of technologies at or near Earth, ranging from blocking high-frequency radio waves used for radio communication to degrading power grid operations. Monitoring and providing early and accurate prediction of solar flares is therefore crucial for preparedness and disaster risk management. In this article, we present a transformer-based framework, named SolarFlareNet, for predicting whether an AR would produce a  $\gamma$ -class flare within the next 24 to 72 h. We consider three  $\gamma$  classes, namely the  $\geq M5.0$  class, the  $\geq M$  class and the  $\geq C$  class, and build three transformers separately, each corresponding to a  $\gamma$  class. Each transformer is used to make predictions of its corresponding  $\gamma$ -class flares. The crux of our approach is to model data samples in an AR as time series and to use transformers to capture the temporal dynamics of the data samples. Each data sample consists of magnetic parameters taken from Space-weather HMI Active Region Patches (SHARP) and related data products. We survey flare events that occurred from May 2010 to December 2022 using the Geostationary Operational Environmental Satellite X-ray flare catalogs provided by the National Centers for Environmental Information (NCEI), and build a database of flares with identified ARs in the NCEI flare catalogs. This flare database is used to construct labels of the data samples suitable for machine learning. We further extend the deterministic approach to a calibration-based probabilistic forecasting method. The SolarFlareNet system is fully operational and is capable of making near real-time predictions of solar flares on the Web.

Solar flares are sudden explosions of energy that occur on the Sun's surface. They often occur in solar active regions (ARs), caused by strong magnetic fields typically associated with sunspot areas. Solar flares are categorized into five classes A, B, C, M, and X, with A-class flares having the lowest intensity and X-class flares having the highest intensity. Major flares are usually accompanied by coronal mass ejections and solar energetic particles<sup>1–7</sup>. These eruptive events can have significant and harmful effects on or near Earth, damaging technologies, power grids, space stations, and human life<sup>8–11</sup>. Therefore, providing accurate and early forecasts of solar flares is crucial for disaster risk management, risk mitigation, and preparedness.

Although a lot of effort has been devoted to flare prediction<sup>12–15</sup>, developing accurate, operational near-real-time flare forecasting systems remains a challenge. In the past, researchers designed statistical models for the prediction of flares based on the physical properties of active regions<sup>16–18</sup>. With the availability of large amounts of flare-related data<sup>14</sup>, researchers started using machine learning methods for flare forecasting<sup>3,19,20</sup>. More recently, deep learning, which is a subfield of machine learning, has emerged and showed promising results in predicting solar eruptions, including solar flares<sup>21,22</sup>.

For example, Nishizuka et al.<sup>23</sup> developed deep neural networks to forecast M- and C-class flares that would occur within 24 h using data downloaded from the Solar Dynamics Observatory (SDO)<sup>24</sup> and the Geostationary Operational Environmental Satellite (GOES). Sun et al.<sup>22</sup> employed three-dimensional (3D) convolutional neural networks (CNNs) to forecast  $\geq M$ -class and  $\geq C$ -class flares using Space-weather HMI Active Region Patches (SHARP)<sup>25</sup> magnetograms downloaded from the Joint Science Operations Center (JSOC) accessible at <http://jsoc.stanford.edu/>. Li et al.<sup>26</sup> also adopted a CNN model to forecast  $\geq M$ -class and  $\geq C$ -class flares using

<sup>1</sup>Institute for Space Weather Sciences, New Jersey Institute of Technology, University Heights, Newark, NJ 07102-1982, USA. <sup>2</sup>Department of Computer Science, New Jersey Institute of Technology, University Heights, Newark, NJ 07102-1982, USA. <sup>3</sup>Center for Solar-Terrestrial Research, New Jersey Institute of Technology, University Heights, Newark, NJ 07102-1982, USA. <sup>4</sup>Big Bear Solar Observatory, New Jersey Institute of Technology, 40386 North Shore Lane, Big Bear City, CA 92314-9672, USA. ✉email: wangj@njit.edu

SHARP magnetograms where the authors restructured the CNN layers in their neural network with different filter sizes. Deng et al.<sup>27</sup> developed a hybrid CNN model to predict solar flares during the rising and declining phases of Solar Cycle 24.

Some researchers adopted SHARP magnetic parameters in time series for flare prediction. Static SHARP parameters quantitatively describe the properties of ARs, especially their ability to produce flares, at a given time. On the other hand, dynamic information, such as the magnetic helicity injection rate, sunspot motions, shear flows, and magnetic flux emergence/flux cancellation, is more important for flare forecasting. Using time series of SHARP parameters allows a model to capture the relationship between the evolution of magnetic fields of ARs and solar flares, hence achieving more accurate flare predictions<sup>28,29</sup>. In an earlier study, Yu et al.<sup>30</sup> added the evolutionary information of ARs to a predictive model for the prediction of short-term solar flares. More recently, Chen et al.<sup>31</sup> designed a long short-term memory (LSTM) network to identify precursors of solar flare events using time series of SHARP parameters. LSTM is suitable for capturing the temporal dynamics of time series. Liu et al.<sup>9</sup> developed another LSTM network with a customized attention mechanism to direct the network to focus on important patterns in time series of SHARP parameters. Sun et al.<sup>32</sup> attempted to distinguish between ARs with strong flares ( $\geq M$ -class flares) and ARs with no flare at all. The authors showed that combining LSTM and CNN can better solve the “strong versus quiet” flare prediction problem, with data from both Solar Cycle 23 and Cycle 24. All of the aforementioned studies provided valuable models and algorithms in the field. However, the existing methods focused on short-term forecasts (usually within 24 h). Furthermore, the models were not used as operational systems.

In this paper, we propose a new deep learning approach to predicting solar flares using time series of SHARP parameters. Our approach employs a transformer-based framework, named SolarFlareNet, which predicts whether there would be a flare within 24 to 72 h, where the flare could be a  $\geq M5.0$ -,  $\geq M$ -, or  $\geq C$ -class flare. We further extend SolarFlareNet to produce probabilistic forecasts of flares and implement the probabilistic model into an operational, near real-time flare forecasting system. Experimental results demonstrate that SolarFlareNet generally performs better than, or is comparable to, related flare prediction methods.

## Results

**Deterministic prediction tasks.** For any given active region (AR) and time point  $t$ , we predict whether there would be a  $\gamma$ -class flare within the next 24 h (48 h, 72 h, respectively) of  $t$  where  $\gamma$  can be  $\geq M5.0$ ,  $\geq M$ , or  $\geq C$ . A  $\geq M5.0$ -class flare means the GOES X-ray flux value of the flare is above  $5 \times 10^{-5} \text{ Wm}^{-2}$ . A  $\geq M$ -class flare refers to an X- or M-class flare. A  $\geq C$ -class flare refers to an X-class, M-class, or C-class flare. We focus on these three classes of flares due to their importance in space weather<sup>9,19,23,33</sup>. We developed three transformer models to tackle the three prediction tasks individually and separately. Notice that we did not consider  $\gamma$  to be  $\geq X$  due to the lack of samples for X-class flares. Instead, we use  $\geq M5.0$  as the most significant class, which contains sufficient samples.

**Comparison with previous methods.** We conducted a series of experiments to compare the proposed SolarFlareNet framework with closely related methods. All these methods perform binary classifications/predictions as defined above. We adopt several performance metrics. Formally, given an AR and a data sample  $x_t$  observed at time point  $t$ , we define  $x_t$  to be a true positive (TP) if the  $\geq M5.0$  ( $\geq M$ ,  $\geq C$ , respectively) model predicts that  $x_t$  is positive, i.e., the AR will produce a  $\geq M5.0$ - ( $\geq M$ -,  $\geq C$ -, respectively) class flare within the next 24 h of  $t$ , and  $x_t$  is indeed positive. We define  $x_t$  as a false positive (FP) if the  $\geq M5.0$  ( $\geq M$ ,  $\geq C$ , respectively) model predicts that  $x_t$  is positive while  $x_t$  is actually negative, i.e., the AR will not produce a  $\geq M5.0$ - ( $\geq M$ -,  $\geq C$ -, respectively) class flare within the next 24 h of  $t$ . We say  $x_t$  is a true negative (TN) if the  $\geq M5.0$  ( $\geq M$ ,  $\geq C$ , respectively) model predicts  $x_t$  to be negative and  $x_t$  is indeed negative;  $x_t$  is a false negative (FN) if the  $\geq M5.0$  ( $\geq M$ ,  $\geq C$ , respectively) model predicts  $x_t$  to be negative while  $x_t$  is actually positive. We also use TP (FP, TN, and FN, respectively) to represent the total number of true positives (false positives, true negatives, and false negatives, respectively). The TP, FP, TN, and FN for the 48-h and 72-h ahead predictions are defined similarly. The performance metrics are calculated as follows:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (1)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (2)$$

$$\text{Accuracy (ACC)} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}} \quad (3)$$

$$\text{Balanced Accuracy (BACC)} = \frac{1}{2} \left( \frac{\text{TP}}{\text{TP} + \text{FN}} + \frac{\text{TN}}{\text{TN} + \text{FP}} \right) \quad (4)$$

$$\text{True Skill Statistics (TSS)} = \frac{\text{TP}}{\text{TP} + \text{FN}} - \frac{\text{FP}}{\text{FP} + \text{TN}} \quad (5)$$

Table 1 compares SolarFlareNet with related methods for 24-h ahead flare predictions. The performance metric values of SolarFlareNet are mean values obtained from 10-fold cross-validation<sup>9</sup>. The metric values of the highest

Metric	Method	$\geq M5.0$ class	$\geq M$ class	$\geq C$ class
Recall	Huang et al. <sup>34</sup>	–	–	–
	Li et al. <sup>26</sup>	–	0.817	0.889
	Liu et al. <sup>9</sup>	<b>0.960</b>	0.885	0.773
	Sun et al. <sup>22</sup>	–	<b>0.925</b>	0.862
	Wang et al. <sup>35</sup>	–	0.730	0.621
	This work	0.853	0.842	<b>0.891</b>
Precision	Huang et al. <sup>34</sup>	–	–	–
	Li et al. <sup>26</sup>	–	<b>0.889</b>	0.906
	Liu et al. <sup>9</sup>	0.048	0.222	0.541
	Sun et al. <sup>22</sup>	–	0.595	0.878
	Wang et al. <sup>35</sup>	–	0.282	0.541
	This work	<b>0.977</b>	0.848	<b>0.949</b>
ACC	Huang et al. <sup>34</sup>	–	–	–
	Li et al. <sup>26</sup>	–	0.891	0.861
	Liu et al. <sup>9</sup>	0.921	0.907	0.826
	Sun et al. <sup>22</sup>	–	0.904	0.879
	Wang et al. <sup>35</sup>	–	<b>0.945</b>	0.883
	This work	<b>0.964</b>	0.928	<b>0.915</b>
BACC	Huang et al. <sup>34</sup>	–	–	–
	Li et al. <sup>26</sup>	–	–	–
	Liu et al. <sup>9</sup>	<b>0.940</b>	0.896	0.806
	Sun et al. <sup>22</sup>	–	–	–
	Wang et al. <sup>35</sup>	–	–	–
	This work	0.926	<b>0.919</b>	<b>0.917</b>
TSS	Huang et al. <sup>34</sup>	–	0.662	0.487
	Li et al. <sup>26</sup>	–	0.749	0.679
	Liu et al. <sup>9</sup>	<b>0.881</b>	0.792	0.612
	Sun et al. <sup>22</sup>	–	0.826	0.756
	Wang et al. <sup>35</sup>	–	0.681	0.553
	This work	0.818	<b>0.839</b>	<b>0.835</b>

**Table 1.** Performance comparison between SolarFlareNet and related methods for 24-h ahead flare predictions. Best values are in bold.

performance models in the related studies are taken directly from those studies and are displayed in Table 1. The symbol ‘–’ means that a method does not produce the metric value for the corresponding prediction task. The best metric values are highlighted in boldface. TSS is the primary metric used in the literature on flare prediction. It can be seen from Table 1 that SolarFlareNet outperforms the state-of-the-art methods in terms of TSS except that Liu et al.<sup>9</sup> perform better than SolarFlareNet in predicting  $\geq M5.0$  class flares.

Table 2 presents the mean performance metric values with standard deviations enclosed in parentheses for the 48- and 72-h forecasts made by SolarFlareNet. None of the existing methods in Table 1 provides predictions

Metric	Hour	$\geq M5.0$ class	$\geq M$ class	$\geq C$ class
Recall	48	0.739 (0.108)	0.735 (0.089)	0.722 (0.089)
	72	0.717 (0.100)	0.708 (0.078)	0.702 (0.089)
Precision	48	0.890 (0.210)	0.823 (0.092)	0.812 (0.072)
	72	0.872 (0.045)	0.812 (0.089)	0.809 (0.051)
ACC	48	0.923 (0.003)	0.907 (0.007)	0.896 (0.047)
	72	0.906 (0.002)	0.883 (0.005)	0.863 (0.040)
BACC	48	0.864 (0.054)	0.857 (0.045)	0.848 (0.040)
	72	0.856 (0.039)	0.843 (0.048)	0.834 (0.029)
TSS	48	0.736 (0.112)	0.728 (0.090)	0.719 (0.079)
	72	0.729 (0.108)	0.714 (0.095)	0.709 (0.058)

**Table 2.** Performance metric values of SolarFlareNet for 48- and 72-h ahead flare predictions.

in 48 or 72 h in advance and, therefore, they are not listed in Table 2. Overall, SolarFlareNet performs well for the 48- and 72-h forecasts. However, the metric values of the tool in Table 2 are lower than those in Table 1. This is understandable due to the longer range of predictions in Table 2.

**Probabilistic forecasting with calibration.** SolarFlareNet is essentially a probabilistic forecasting method, producing a probability between 0 and 1. The method compares the probability with a predetermined threshold, which is set to 0.5. Given an AR and a data sample  $x_t$  at time point  $t$ , if the predicted probability is greater than or equal to the threshold, then the AR will produce a flare within the next 24 (48, 72, respectively) hours of  $t$  (i.e.,  $x_t$  belongs to the positive class); otherwise, the AR will not produce a flare within the next 24 (48, 72, respectively) hours of  $t$  (i.e.,  $x_t$  belongs to the negative class). We can turn SolarFlareNet into a probabilistic forecasting method by directly outputting the predicted probability without comparing it with the threshold. Under this circumstance, the output is interpreted as a probabilistic estimate of how likely the AR will produce a flare within the next 24 (48, 72, respectively) hours of  $t$ . We employ a probability calibration technique with isotonic regression<sup>36,37</sup> to adjust the predicted probability and avoid the mismatch between the distributions of the predicted and expected probabilistic values<sup>5</sup>. We add a suffix “-C” to SolarFlareNet to denote the network without calibration.

To evaluate the performance of a probabilistic forecasting method, we use the Brier score (BS) and Brier skill score (BSS), defined as follows<sup>4,5,38</sup>:

$$BS = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2 \quad (6)$$

$$BSS = 1 - \frac{BS}{\frac{1}{N} \sum_{i=1}^N (y_i - \bar{y})^2} \quad (7)$$

where  $N$  is the number of data samples in a test set;  $y_i$  denotes the observed probability and  $\hat{y}_i$  denotes the predicted probability of the  $i$ th test data sample, respectively;  $\bar{y} = \frac{1}{N} \sum_{i=1}^N y_i$  denotes the mean of all the observed probabilities. BS values range from 0 to 1, with 0 being a perfect score. BSS values range from  $-\infty$  to 1, with 1 being a perfect score.

Table 3 compares SolarFlareNet, used as a probabilistic forecasting method, with a closely related method<sup>9</sup>. The BS and BSS values in the table are mean values (with standard deviations enclosed in parentheses) obtained from 10-fold cross-validation. The metric values for the existing method are taken directly from the related work<sup>9</sup>. The best BS and BSS values are highlighted in bold. Notice that the existing method did not make 48-h or 72-h forecasts in advance. Table 3 shows that there is no definitive conclusion regarding the relative performance of SolarFlareNet and the existing method. The existing method is better in terms of BS, while SolarFlareNet is better in terms of BSS. On the other hand, the calibrated version of a model is better than the model without calibration. Notice also that the results of the 48-h and 72-h forecasts are worse than those of the 24-h forecasts. This is understandable since the longer the prediction window, the worse the performance a model achieves due to data deviation over time.

Hour	Metric	Method	$\geq M5.0$ class	$\geq M$ class	$\geq C$ class
24	BS	Liu et al. <sup>9</sup>	<b>0.090</b> (0.011)	<b>0.090</b> (0.009)	<b>0.133</b> (0.007)
		SolarFlareNet	0.226 (0.024)	0.244 (0.013)	0.285 (0.034)
		SolarFlareNet-C	0.263 (0.024)	0.281 (0.050)	0.313 (0.033)
	BSS	Liu et al. <sup>9</sup>	-21.576 (2.956)	-2.241 (0.319)	0.152 (0.047)
		SolarFlareNet	<b>0.584</b> (0.022)	<b>0.521</b> ( <b>0.042</b> )	<b>0.409</b> (0.062)
		SolarFlareNet-C	0.504 (0.026)	0.491 (0.031)	0.349 (0.055)
48	BS	Liu et al. <sup>9</sup>	-	-	-
		SolarFlareNet	<b>0.272</b> (0.091)	<b>0.312</b> (0.101)	<b>0.361</b> (0.091)
		SolarFlareNet-C	0.315 (0.049)	0.336 (0.033)	0.378 (0.104)
	BSS	Liu et al. <sup>9</sup>	-	-	-
		SolarFlareNet	<b>0.569</b> (0.045)	<b>0.524</b> (0.021)	<b>0.502</b> (0.033)
		SolarFlareNet-C	0.457 (0.062)	0.424 (0.091)	0.411 (0.056)
72	BS	Liu et al. <sup>9</sup>	-	-	-
		SolarFlareNet	<b>0.313</b> (0.062)	<b>0.327</b> (0.063)	<b>0.344</b> (0.049)
		SolarFlareNet-C	0.329 (0.094)	0.369 (0.088)	0.376 (0.102)
	BSS	Liu et al. <sup>9</sup>	-	-	-
		SolarFlareNet	<b>0.549</b> (0.067)	<b>0.524</b> (0.089)	<b>0.501</b> (0.093)
		SolarFlareNet-C	0.514 (0.077)	0.469 (0.095)	0.447 (0.059)

**Table 3.** Performance comparison between SolarFlareNet and an existing method for probabilistic flare predictions (24 to 72 h in advance). Best values are in bold.

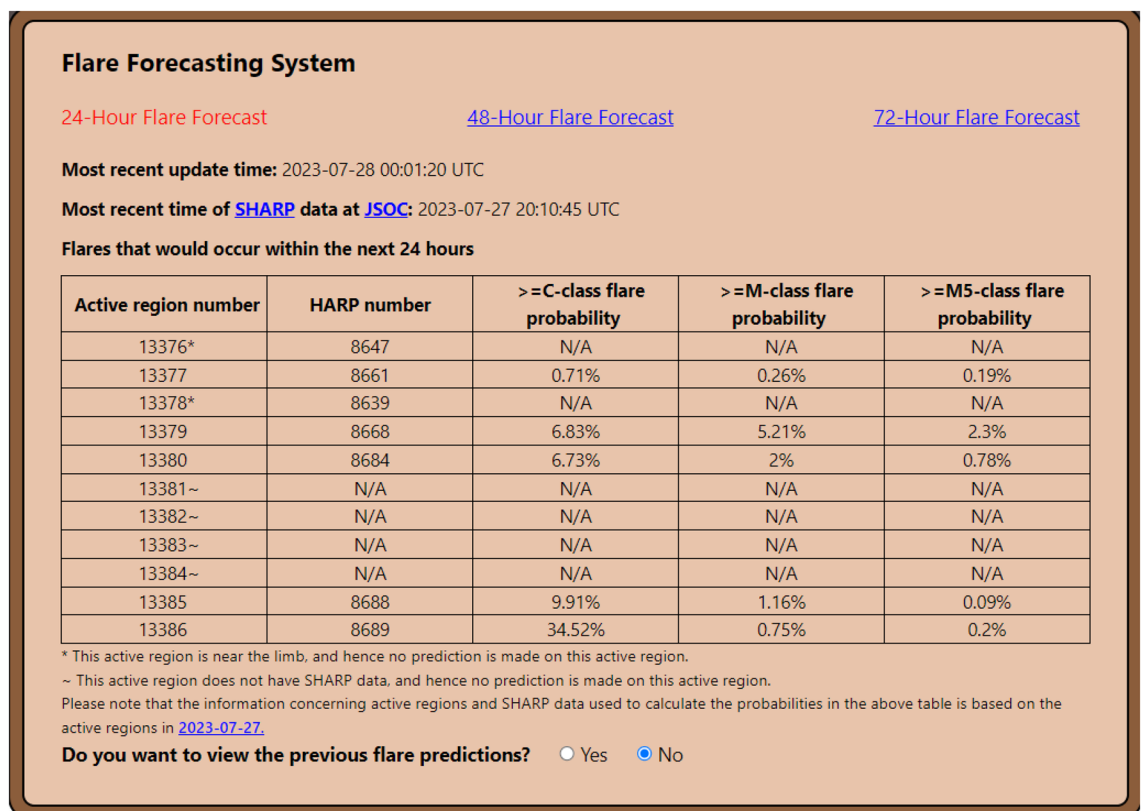
**The SolarFlareNet system.** We have implemented the probabilistic forecasting method described above into an operational, near real-time flare forecasting system. To access the system, visit the SolarDB website at <https://nature.njit.edu/solardb/index.html>. On the website, select and click the menu entry “Tools” and then select and click “Flare Forecasting System.” Figure 1 shows the graphical user interface (GUI) of the system. It displays a probabilistic estimate of how likely an AR will produce a flare within the next 24, 48, and 72 h of the time point at which the system is updated each day. No prediction is made for an AR marked with a special character \*, #, or ~ where

- \* means the AR is near the limb,
- # means the AR is spotless with the number of spots being zero,
- ~ means no SHARP data is available for the AR.

The system provides daily predictions based on the data obtained from the previous day. When the user clicks the link to the previous day, the user is led to the SolarMonitor site that is accessible at <https://www.solarmonitor.org/index.php> where detailed AR information for that day is available. The system also provides previous forecasting results since the operational system came online. We compare the previous forecasting results with the true flare events in the GOES X-ray flare catalogs provided by NCEI. The SolarFlareNet system achieves 89% (76%, 71%, respectively) accuracy for 24-h (48-h, 72-h, respectively) ahead predictions.

## Discussion and conclusion

In this article, we present a novel transformer-based framework to predict whether a solar active region (AR) would produce a  $\gamma$ -class flare within the next 24 to 72 h where  $\gamma$  is  $\geq M5.0$ ,  $\geq M$ , or  $\geq C$ . We use three transformer models to handle the three classes of flares individually and separately. All three transformer models perform binary predictions. We collect ARs with flares that occurred between 2010 and 2022 from the GOES X-ray flare catalogs provided by the National Centers for Environmental Information (NCEI). In addition, we downloaded SHARP magnetic parameters from the Joint Science Operations Center (JSOC). Each data sample contains SHARP parameters suitable for machine learning. We conducted experiments using 10-fold cross-validation<sup>9</sup>. Based on the experiments, our transformer-based framework generally performs better than closely related methods in terms of TSS (true skill statistics), as shown in Table 1. We further extend our framework to produce probabilistic forecasts of flares and implement the framework into an operational, near real-time flare forecasting system accessible on the Web. The probabilistic framework is comparable to a closely related method<sup>9</sup> in terms of BS (Brier score) and BSS (Brier skill score) when making 24-h forecasts, as shown in Table 3, although the



**Figure 1.** The graphical user interface of the SolarFlareNet system.

existing method did not make 48- or 72-h forecasts. Thus, we conclude that SolarFlareNet is a feasible tool for producing flare forecasts within 24 to 72 h.

## Methods

**Data collection.** In this study we used SHARP magnetic parameters<sup>2,9,25</sup> downloaded from the Joint Science Operations Center (JSOC) accessible at <http://jsoc.stanford.edu/>. Specifically, we collect data samples, composed of SHARP parameters, at a cadence of 12 minutes where the data samples are retrieved from the `hmi.sharp_cea_720s` data series on the JSOC website using the Python package `SunPy`<sup>39</sup>. We selected nine SHARP magnetic parameters as suggested in the literature<sup>2-4,9,19</sup>. These nine parameters include the total unsigned current helicity (TOTUSJH), total unsigned vertical current (TOTUSJZ), total unsigned flux (USFLUX), mean characteristic twist parameter (MEANALP), sum of flux near polarity inversion line (R\_VALUE), total photospheric magnetic free energy density (TOTPOT), sum of the modulus of the net current per polarity (SAVNCPP), area of strong field pixels in the active region (AREA\_ACR), and absolute value of the net current helicity (ABSJZH). Table 4 presents an overview of the nine parameters. The SHARP parameters' values are in different scales and units; therefore, we normalize each parameter's values using the min-max normalization method<sup>4,5</sup>. Formally, let  $p_i^k$  be the original value of the  $i$ th parameter of the  $k$ th data sample. Let  $q_i^k$  be the normalized value of the  $i$ th parameter of the  $k$ th data sample. Let  $min_i$  be the minimum value of the  $i$ th parameter. Let  $max_i$  be the maximum value of the  $i$ th parameter. Then

$$q_i^k = \frac{p_i^k - min_i}{max_i - min_i} \quad (8)$$

We collected A-, B-, C-, M- and X-class flares that occurred between May 2010 and December 2022, and their associated active regions (ARs) from the GOES X-ray flare catalogs provided by the National Centers for Environmental Information (NCEI). Flares without identified ARs were excluded. This process yielded a database of 8 A-class flares, 6571 B-class flares, 8973 C-class flares, 895 M-class flares, and 58 X-class flares. Also, we collected 10 nonflaring ARs<sup>40</sup>. We collected data samples that were 24 (48, 72, respectively) hours before a flare. Furthermore, we collected data samples that were 24 (48, 72, respectively) hours after the start time of each nonflaring AR. The data was then cleaned as follows<sup>2,5,9</sup>.

We discard ARs that are outside  $\pm 70^\circ$  of the central meridian. These ARs are near the limb and have projection effects that render the calculation of the ARs' SHARP parameters incorrect. In addition, we discard a data sample if (i) its corresponding flare record has an absolute value of the radial velocity of SDO greater than  $3500 \text{ m s}^{-1}$ , (ii) the HMI data have low quality<sup>41</sup>, or (iii) the data sample has missing values or incomplete SHARP parameters. Thus, we exclude low-quality data samples and keep qualified data samples of high quality in our study.

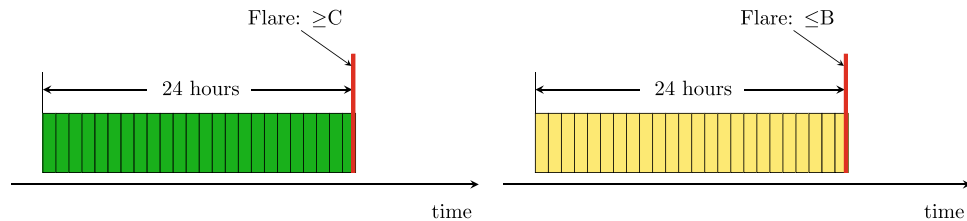
**Data labeling.** Data labeling is crucial in machine learning. To predict  $\geq C$ -class flares, suppose that a C-, M-, or X-class flare occurs at time point  $t$  on an AR (more precisely, the start time of the flare is  $t$ ). Data samples between  $t$  and  $t - 24$  h (48, 72 h, respectively) in the AR are labeled positive. If the flare occurs at time point  $t$  is an A-class or B-class flare, the data samples between  $t$  and  $t - 24$  h (48, 72 h, respectively) in the AR are labeled negative. Figure 2 illustrates the labeling scheme to predict whether a  $\geq C$ -class flare would occur within 24 h. In predicting  $\geq M$ -class flares, we use  $\geq M$ -class flares to label positive data samples; use  $\leq C$ -class flares to label negative data samples. In predicting  $\geq M5.0$ -class flares, we use  $\geq M5.0$ -class flares to label positive data samples; use  $\leq C$ -class flares as well as M1.0- through M4.0-class flares to label negative data samples. If there are recurring flares whose corresponding data samples overlap, we give priority to the largest flare and label the overlapped data samples based on the largest flare. In all three prediction tasks, the data samples in the nonflaring ARs are labeled negative.

Table 5 shows the total numbers of positive and negative data samples in each class for 24-, 48-, and 72-h ahead flare predictions. The numbers in the table are lower than expected. This is because we discarded/removed many low-quality data samples as described above. If a gap occurs in the middle of a time series due to removal, we use a zero-padding strategy<sup>5,9</sup> to create a synthetic data sample to fill the gap. The synthetic data sample has

Keyword	Description	Formula
TOTUSJH	Total unsigned current helicity	$H_{total} \propto \sum  B_z \cdot J_z $
TOTUSJZ	Total unsigned vertical current	$J_{ztotal} = \sum  J_z  dA$
USFLUX	Total unsigned flux	$\Phi = \sum  B_z  dA$
MEANALP	Mean characteristic twist parameter, $\alpha$	$\alpha_{total} \propto \frac{\sum J_z B_z}{\sum B_z^2}$
R_VALUE	Sum of flux near polarity inversion line	$\Phi = \sum  B_{Los}  dA$ within $R$ mask
TOTPOT	Total photospheric magnetic free energy density	$\rho_{tot} \propto \sum (\mathbf{B}^{Obs} - \mathbf{B}^{Pot})^2 dA$
SAVNCPP	Sum of the modulus of the net current per polarity	$J_{zsum} \propto  \sum^{B_z^+} J_z dA  +  \sum^{B_z^-} J_z dA $
AREA_ACR	Area of strong field pixels in the active region	Area = $\sum$ Pixels
ABSJZH	Absolute value of the net current helicity	$H_{abs} \propto  \sum B_z \cdot J_z $

**Table 4.** Overview of the nine SHARP parameters used in our study.





**Figure 2.** Illustration of positive and negative data samples used in predicting  $\geq C$ -class flares. In the left panel, the red vertical line indicates the start time of a  $\geq C$ -class flare. The data samples collected in the 24 h prior to the red vertical line are labeled positive (in green color). In the right panel, the red vertical line indicates the start time of an A-class or B-class flare. The data samples collected in the 24 h prior to the red vertical line are labeled negative (in yellow color).

Hour	Data samples	$\geq M5.0$ class	$\geq M$ class	$\geq C$ class
24	Positive	2,125	13,989	244,968
	Negative	461,060	449,196	218,517
48	Positive	2,255	16,709	316,149
	Negative	615,708	602,154	304,714
72	Positive	2,375	18,505	356,219
	Negative	704,997	689,567	350,953

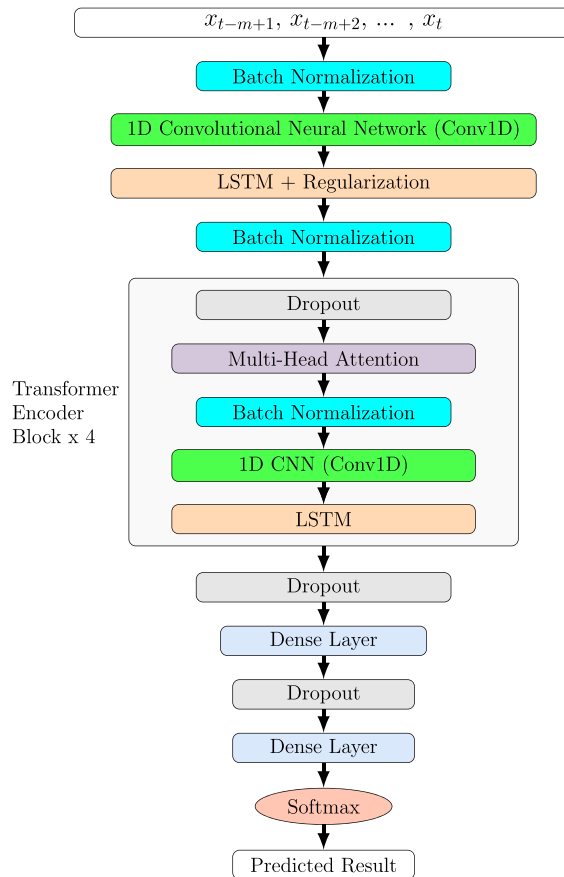
**Table 5.** Total numbers of positive and negative data samples in each class for 24-, 48-, and 72-h ahead flare predictions.

zero values for all nine SHARP parameters. The synthetic data sample is added after normalization of the values of the SHARP parameters, and therefore the synthetic data sample does not affect the normalization procedure.

For each prediction task, we divide the corresponding data samples into 10 equal sized distinct partitions/folds that are used to perform 10-fold cross-validation experiments. In the run  $i$ , where  $1 \leq i \leq 10$ , we use fold  $i$  as the test set and the union of the other nine folds as the training set. The data samples of the same AR are placed in the training set or the test set, but not both. This scheme ensures that a model is trained with data different from the test data and makes predictions on the test data that it has never seen during training. There are 10 folds and, consequently, 10 runs. The means and standard deviations of the performance metrics' values over the 10 runs are calculated and recorded.

**Data augmentation.** The data sets used in this study to predict flares of the  $\geq M$ - and  $\geq M5.0$ -class are imbalanced as shown in Table 5 where negative data samples are much more than positive data samples. Imbalanced data pose a challenge in model training and often result in poor model performance. One may use data augmentation to combat the imbalanced data. Data augmentation is an important technique that enriches training data and increases the generalization of the model<sup>42</sup>. Here, we adopt the Gaussian white noise (GWN) data augmentation scheme because it has shown a significant improvement in model performance<sup>43,44</sup>. GWN assumes that any two values are statistically independent, regardless of how close they are in time. The stationary random values of GWN are generated using the zero mean and 5% of the standard deviation. During training, the data augmentation is applied to the minority (positive) class, leaving the majority (negative) class as is. During testing, the data are left without any augmentation so that the model predicts only on the actual test data to avoid any misleading performance assessment.

**The SolarFlareNet architecture.** Figure 3 presents the architecture of SolarFlareNet. It is a transformer-based framework that combines a one-dimensional convolutional neural network (Conv1D), long short-term memory (LSTM), transformer encoder blocks (TEBs), and additional layers that include batch normalization (BN) layers, dropout layers, and dense layers. The first layer is the input layer, which takes as input a time series of  $m$  consecutive data samples  $x_{t-m+1}, x_{t-m+2}, \dots, x_{t-1}, x_t$  where  $x_t$  is the data sample at time point  $t$ <sup>5</sup>. (In the study presented here,  $m$  is set to 10.) The input layer is followed by a BN layer. BN is an additional mechanism to stabilize SolarFlareNet, make it faster, and help to avoid overfitting during training<sup>45</sup>. We applied BN after the input layer, the LSTM layer, and within the TEBs to make sure that SolarFlareNet is stable throughout the training process. The BN layer is followed by the Conv1D layer because time series generally have a strong 1D time locality that can be extracted by the Conv1D layers<sup>46</sup>. Then, the LSTM layer is used, which is equipped with regularization to also avoid overfitting. LSTM is suitable for handling time series data to capture the temporal correlation and dependency in the data. Adding an LSTM layer after a Conv1D layer has shown significant improvement in time series prediction<sup>47–49</sup>. The LSTM layer passes the learned features and patterns to a BN layer to stabilize the network before the data go to the TEBs.



**Figure 3.** Architecture of SolarFlareNet.

We use transformer encoders without decoders because we process time series here, rather than performing natural language processing where the decoders are required to decode the words for sentence translation. The number of TEBs is set to 4. This number has a significant effect on the overall performance of the model<sup>50</sup>. When we use less than 4 TEBs, the model is not able to learn useful patterns and is under-fitted. When we use more than 4 TEBs, the large number of TEBs causes overhead on the encoder processing while the model tends to do excessive overfitting and lean toward the majority class (i.e., negative class) in the data, ignoring the minority class (i.e., positive class) entirely. Each TEB is configured with a dropout layer, multi-head attention (MHA) layer, a BN layer, a Conv1D layer, and an LSTM layer. The MHA layer is the most important layer in the encoder because it provides the transformation on the sequence values to obtain the different metrics. The MHA layer is configured with 4 heads and each attention head is also set to 4. The dropout layer is mainly used to overcome the overfitting caused by the imbalanced data. It drops a percentage of the neurons from the architecture, which causes the internal architecture of the model to change, allowing for better performance and stability. Finally, the softmax function is used as the final activation function, which produces a probabilistic estimate of how likely a flare will occur within the next 24 (48, 72, respectively) hours of  $t$ .

**Ablation study.** We performed ablation tests to assess each component of SolarFlareNet. We consider four variants of SolarFlareNet, denoted SolarFlareNet-Conv, SolarFlareNet-L, SolarFlareNet-ConvL, and SolarFlareNet-T, respectively. Here, SolarFlareNet-Conv (SolarFlareNet-L, SolarFlareNet-ConvL, SolarFlareNet-T, respectively) represents the subnet of SolarFlareNet in which the Conv1D layer (LSTM layer, Conv1D and LSTM layers, transformer network with the 4 TEBs, respectively) is removed while keeping the remaining components of the SolarFlareNet framework. Table 6 compares the TSS values of the five models for the 24-, 48-, and 72-h ahead flare prediction. It can be seen from Table 6 that the full model, SolarFlareNet, outperforms the four subnets in terms of the TSS metric. This happens because the SolarFlareNet-Conv model captures the temporal correlation of the test data, but does not learn additional characteristics of the data to build a stronger relationship between the test data. SolarFlareNet-L captures the properties of the test data, but lacks knowledge of the temporal correlation patterns in the data to deeply analyze the sequential information in the test data. It can also be seen from Table 6 that the SolarFlareNet-ConvL model is not as good as the full model, indicating that the transformer network alone is not enough to produce the best results. Lastly, SolarFlareNet-T has the least performance among the four subnets, demonstrating the importance of the transformer network. In conclusion, our ablation study indicates that the performance of the proposed SolarFlareNet framework is not dominated



Hour	Method	$\geq M5.0$ class	$\geq M$ class	$\geq C$ class
24	SolarFlareNet	<b>0.818</b> (0.021)	<b>0.839</b> (0.030)	<b>0.835</b> (0.048)
	SolarFlareNet-Conv	0.780 (0.036)	0.759 (0.052)	0.822 (0.023)
	SolarFlareNet-L	0.779 (0.022)	0.737 (0.041)	0.713 (0.039)
	SolarFlareNet-ConvL	0.742 (0.029)	0.719 (0.041)	0.728 (0.037)
	SolarFlareNet-T	0.716 (0.101)	0.704 (0.093)	0.712 (0.078)
48	SolarFlareNet	<b>0.736</b> (0.112)	<b>0.728</b> (0.090)	<b>0.719</b> (0.079)
	SolarFlareNet-Conv	0.729 (0.049)	0.715 (0.055)	0.695 (0.035)
	SolarFlareNet-L	0.694 (0.066)	0.689 (0.012)	0.675 (0.021)
	SolarFlareNet-ConvL	0.681 (0.063)	0.676 (0.054)	0.673 (0.048)
	SolarFlareNet-T	0.662 (0.061)	0.647 (0.032)	0.641 (0.033)
72	SolarFlareNet	<b>0.729</b> (0.108)	<b>0.714</b> (0.095)	<b>0.709</b> (0.058)
	SolarFlareNet-Conv	0.703 (0.042)	0.696 (0.011)	0.658 (0.023)
	SolarFlareNet-L	0.688 (0.046)	0.666 (0.039)	0.658 (0.016)
	SolarFlareNet-ConvL	0.665 (0.026)	0.643 (0.031)	0.632 (0.030)
	SolarFlareNet-T	0.635 (0.028)	0.624 (0.046)	0.619 (0.033)

**Table 6.** TSS values of the five methods considered in the ablation study. Best values are in bold.

by any single component. In fact, all components have made contributions to the overall performance of the proposed framework.

### Data availability

The trained SolarFlareNet model and datasets used in this study can be downloaded from <https://nature.njit.edu/solardb/solarflarenet>.

Received: 8 June 2023; Accepted: 17 August 2023

Published online: 22 August 2023

### References

- Qahwaji, R., Colak, T., Al-Omari, M. & Ipson, S. Automated prediction of CMEs using machine learning of CME - flare associations. *Sol. Phys.* **248**, 471–483 (2008).
- Bobra, M. G. & Ilonidis, S. Predicting coronal mass ejections using machine learning methods. *Astrophys. J.* **821**, 127 (2016).
- Liu, C., Deng, N., Wang, J. T. L. & Wang, H. Predicting solar flares using SDO/HMI vector magnetic data products and the random forest algorithm. *Astrophys. J.* **843**, 104 (2017).
- Liu, H., Liu, C., Wang, J. T. L. & Wang, H. Predicting coronal mass ejections using SDO/HMI vector magnetic data products and recurrent neural networks. *Astrophys. J.* **890**, 12 (2020).
- Abduallah, Y. *et al.* Predicting solar energetic particles using SDO/HMI vector magnetic data products and a bidirectional LSTM network. *Astrophys. J. Suppl. Ser.* **260**, 16 (2022).
- Moreland, K. *et al.* A machine-learning oriented dataset for forecasting SEP occurrence and properties. In *44th COSPAR Scientific Assembly*, vol. 44, 1151 (2022).
- Singh, T., Benson, B., Raza, S., Kim, T. & Pogorelov, N. Improving the arrival time prediction of coronal mass ejections using magnetohydrodynamic ensemble modeling, heliospheric imager data and machine learning. In *EGU General Assembly Conference Abstracts*, EGU22–13167 (2022).
- Daglis, I., Baker, D., Kappenman, J., Panasyuk, M. & Daly, E. Effects of space weather on technology infrastructure. *Space Weather* **2**, S02004 (2004).
- Liu, H., Liu, C., Wang, J. T. L. & Wang, H. Predicting solar flares using a long short-term memory network. *Astrophys. J.* **877**, 121 (2019).
- Zhang, H. *et al.* Solar flare index prediction using SDO/HMI vector magnetic data products with statistical and machine-learning methods. *Astrophys. J. Suppl. Ser.* **263**, 28 (2022).
- He, X.-R. *et al.* Solar flare short-term forecast model based on long and short-term memory neural network. *Chin. Astron. Astrophys.* **47**, 108–126 (2023).
- Huang, X., Zhang, L., Wang, H. & Li, L. Improving the performance of solar flare prediction using active longitudes information. *Astron. Astrophys.* **549**, A127 (2013).
- Panos, B. & Kleint, L. Real-time flare prediction based on distinctions between flaring and non-flaring active region spectra. *Astrophys. J.* **891**, 17 (2020).
- Georgoulis, M. K. *et al.* The flare likelihood and region eruption forecasting (FLARECAST) project: flare forecasting in the big data & machine learning era. *J. Space Weather Space Clim.* **11**, 39 (2021).
- Tang, R. *et al.* Solar flare prediction based on the fusion of multiple deep-learning models. *Astrophys. J. Suppl. Ser.* **257**, 50 (2021).
- Gallagher, P. T., Moon, Y. J. & Wang, H. Active-region monitoring and flare forecasting I. Data processing and first results. *Solar Phys.* **209**, 171–183 (2002).
- Leka, K. D. & Barnes, G. Photospheric magnetic field properties of flaring versus flare-quiet active regions. IV. A statistically significant sample. *Astrophys. J.* **656**, 1173–1186 (2007).
- Mason, J. P. & Hoeksema, J. T. Testing automated solar flare forecasting with 13 years of Michelson Doppler Imager magnetograms. *Astrophys. J.* **723**, 634–640 (2010).
- Bobra, M. G. & Couvidat, S. Solar flare prediction using SDO/HMI vector magnetic field data with a machine-learning algorithm. *Astrophys. J.* **798**, 135 (2015).
- Abduallah, Y., Wang, J. T. L., Nie, Y., Liu, C. & Wang, H. DeepSun: Machine-learning-as-a-service for solar flare prediction. *Res. Astron. Astrophys.* **21**, 160 (2021).

21. Liu, S. *et al.* Deep learning based solar flare forecasting model. II. Influence of image resolution. *Astrophys. J.* **941**, 20 (2022).
22. Sun, P. *et al.* Solar flare forecast using 3D convolutional neural networks. *Astrophys. J.* **941**, 1 (2022).
23. Nishizuka, N., Sugiura, K., Kubo, Y., Den, M. & Ishii, M. Deep flare net (DeFN) model for solar flare prediction. *Astrophys. J.* **858**, 113 (2018).
24. Pesnell, W. D. Solar Dynamics Observatory (SDO). In *Handbook of Cosmic Hazards and Planetary Defense*, 179–196 (Springer, 2015).
25. Bobra, M. G. *et al.* The Helioseismic and Magnetic Imager (HMI) vector magnetic field pipeline: SHARPs - Space-weather HMI Active Region Patches. *Sol. Phys.* **289**, 3549–3578 (2014).
26. Li, X., Zheng, Y., Wang, X. & Wang, L. Predicting solar flares using a novel deep convolutional neural network. *Astrophys. J.* **891**, 10 (2020).
27. Deng, Z. *et al.* Fine-grained solar flare forecasting based on the hybrid convolutional neural networks. *Astrophys. J.* **922**, 232 (2021).
28. Tian, Y. Relationship between the evolution of solar magnetic field and flares in different active regions. *J. Phys: Conf. Ser.* **2282**, 012025 (2022).
29. van Driel-Gesztelyi, L. & Green, L. M. Evolution of active regions. *Living Rev. Sol. Phys.* **12**, 1 (2015).
30. Yu, D., Huang, X., Wang, H. & Cui, Y. Short-term solar flare prediction using a sequential supervised learning method. *Sol. Phys.* **255**, 91–105 (2009).
31. Chen, Y. *et al.* Identifying solar flare precursors using time series of SDO/HMI images and SHARP parameters. *Space Weather* **17**, 1404–1426 (2019).
32. Sun, Z. *et al.* Predicting solar flares using CNN and LSTM on two solar cycles of active region data. *Astrophys. J.* **931**, 163 (2022).
33. Jonas, E., Bobra, M., Shankar, V., Todd Hoeksema, J. & Recht, B. Flare prediction using photospheric and coronal image data. *Sol. Phys.* **293**, 48 (2018).
34. Huang, X. *et al.* Deep learning based solar flare forecasting model. I. Results for line-of-sight magnetograms. *Astrophys. J.* **856**, 7 (2018).
35. Wang, X. *et al.* Predicting solar flares with machine learning: Investigating solar cycle dependence. *Astrophys. J.* **895**, 3 (2020).
36. Kruskal, J. B. Nonmetric multidimensional scaling: A numerical method. *Psychometrika* **29**, 115–129 (1964).
37. Sager, T. W. & Thisted, R. A. Maximum likelihood estimation of isotonic modal regression. *Ann. Stat.* **10**, 690–707 (1982).
38. Wilks, D. S. Sampling distributions of the Brier score and Brier skill score under serial dependence. *Q. J. R. Meteorol. Soc.* **136**, 2109–2118 (2010).
39. The SunPy Community *et al.* SunPy—Python for solar physics. *Computational Science and Discovery* **8**, 014009 (2015).
40. Hazra, S., Sardar, G. & Chowdhury, P. Distinguishing between flaring and nonflaring active regions. *Astron. Astrophys.* **639**, A44 (2020).
41. Hoeksema, J. T. *et al.* The Helioseismic and Magnetic Imager (HMI) vector magnetic field pipeline: Overview and performance. *Sol. Phys.* **289**, 3483–3530 (2014).
42. Deng, Y. *et al.* Deep transfer learning and data augmentation improve glucose levels prediction in type 2 diabetes patients. *Nat. Portf. J. Digit. Med.* **4**, 1013345 (2021).
43. Um, T. T. *et al.* Data augmentation of wearable sensor data for Parkinson's disease monitoring using convolutional neural networks. In *Proceedings of the 19th Association for Computing Machinery International Conference on Multimodal Interaction* (2017).
44. Li, K., Daniels, J., Liu, C., Herrero, P. & Georgiou, P. Convolutional recurrent neural networks for glucose prediction. *IEEE J. Biomed. Health Inform.* **24**, 603–613 (2020).
45. Zerveas, G., Jayaraman, S., Patel, D., Bhamidipaty, A. & Eickhoff, C. A transformer-based framework for multivariate time series representation learning. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining* (2021).
46. Kravchik, M. & Shabtai, A. Detecting cyber attacks in industrial control systems using convolutional neural networks. In *Proceedings of the 2018 Workshop on Cyber-Physical Systems Security and Privacy* (2018).
47. Abdullah, Y. *et al.* Reconstruction of total solar irradiance by deep learning. In *Proceedings of the 34th International Florida Artificial Intelligence Research Society Conference* (2021).
48. Abdullah, Y., Wang, J. T. L., Xu, C. & Wang, H. A transformer-based framework for geomagnetic activity prediction. In *Proceedings of the 26th International Symposium on Methodologies for Intelligent Systems* (2022).
49. Abdullah, Y. *et al.* Forecasting the disturbance storm time index with Bayesian deep learning. In *Proceedings of the 35th International Florida Artificial Intelligence Research Society Conference* (2022).
50. Vaswani, A. *et al.* Attention is all you need. In *Proceedings of the Annual Conference on Neural Information Processing Systems* (2017).

## Acknowledgements

The authors thank members of the Institute for Space Weather Sciences for helpful discussions. We also thank the reviewers for valuable comments and suggestions.

## Author contributions

J.W. and H.W. conceived the study. Y.A. implemented the operational, near real-time flare forecasting system. Y.X. evaluated and tested the system. All authors reviewed the manuscript.

## Funding

This work was supported in part by US NSF grants AGS-1927578, AGS-2149748, and AGS-2228996.

## Competing interests

The authors declare no competing interests.

## Additional information

**Correspondence** and requests for materials should be addressed to J.T.L.W.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023