Solar Flare Forecasting with Data-driven Interpretable Model

Anonymous Author(s)
Affiliation
Address
email

Abstract

Solar flares are the most violent activities in the solar system, which are caused by the evolution of magnetic field in solar active regions. However, the mechanism which triggers solar flares is still an active research area and many algorithms based on different models are proposed to forecast solar flares. In this paper, we propose a novel data-driven method to forecast solar flares, which is built with convolutional neural network and long short term memory neural network. Our method could precept continuous magnetic field observation data with 6 hours long and predict the probability of flares of different classes in the next 24 hours with a Bayesian neural network. Comparing with traditional method, our method could not only forecast solar flares with high precision rate and low false alarm rate, but also highlight the region which would trigger solar flares with the class activation mapping (CAM). The inception obtained by the CAM could help scientists to dig deeper into physical mechanism which triggers solar flares. We use our method to process real observation data. Results show that our model mainly focuses on the region with strong magnetic field, the polarity reversal line and the magnetic field conversion area, which is consistent to theoretical predictions.

1 Introduction

The sun is the closest star to the earth, which is also the most important celestial objects to human kind. However, the sun is not a quiet star and solar activities would bring catastrophic effects to the space and the earth. Human built facilities, such as the space station, the communication satellite, the navigation satellite, the plane, the oil pipe or the power grid, would be seriously affected by solar activities. Solar flares are one of the most important solar activities, which would erupt a lot of energy within very short time and greatly affect electromagnetic environment of the Earth and the Space. Therefore, it is of great significance to study the triggering mechanism of solar flares and establish an accurate and reliable solar flare forecasting model, to avoid or reduce the impact of solar flares on human beings.

The sun is mainly composed of hydrogen and helium. In the centre of the sun, the nuclear fusion happens and emits energy to heat gas inside the sun. As the nuclear fusion continuously emits energy to increase the temperature of the gas, the gas would gradually become the plasma. The plasma has very high moving speed and the gravity would attract the plasma to form a stable celestial object. As a celestial object mainly composed by the plasma, the sun is controlled by the magnetic field. The heating process is not uniform and there would be some occasions that the plasma has very fast moving speed. The churning of plasma would push out through the surface and generate sunspots. In

Submitted to NeurIPS 2021 AI for Science Workshop.

some circumstances, the distance between lower regions (lower corona) of a magnetic loop would be too small and magnetic reconnection would happen to unleash a lot of energy, which would generate solar flares. The level of solar flare is defined by the maximum X-ray brightness of the solar flare.

As mentioned above, although the general theory about the mechanism which triggers solar flares is clear, it still lacks enough theory to predict when the solar flare happens. For quite a long time, scientists have proposed several different theories to explain mechanism of solar flares and predict solar flares [13, 11, 16]. Traditional solar flare forecasting models mainly rely on manual experience and domain knowledge to extract characteristic parameters related to solar flares from observed data [6, 2, 17, 1, 7]. It is difficult to make full use of the information related to solar flares contained in massive observational data, leaving along the interpretable theory behind these algorithms.

In recent years, machine learning algorithms have attracted a lot of attentions. Particularly with the help of deep neural networks, deep learning could automatically extract effective features from continuous data, providing a new way to predict solar flares. Since the magnetic field controls solar activities, solar magnetogram would contain important features to predict solar. Therefore, several studies have proposed to use the convolutioan neural network, which can automatically extract effective features from images [8, 14, 12, 9, 18], to extract features from single frame of magnetogram for solar flare forcasting. However, previous studies assume solar flare prediction as a classification problem and they have some limitations in real applications for the following reasons:
1. The characteristics of physical parameters of continuous solar magnetic field activity region over time are not considered;
2.The solar flare prediction problem is regarded as a deterministic problem without considering the random and sudden characteristics of flares.

Since the solar flare is a continuous process which is closely related to the dynamical variations of the solar active area, we would use the long short-term memory network [3, 4, 15] to merge features from continuous magnetogram. Besides, we use the Bayesian neural network to predict solar flares. Therefore, we build the Bayesian spatio-temporal connection model for solar flare forecasting. To better show the triggering mechanism of solar flares, we use the class activation mapping technology to draw the attention area of the model and analyze the mechanism in triggering of solar flares. This article is organized as follows: a forecasting model based on the deep learning method is proposed in Section 2; The interpretability of the model is investigated in Section 3; The performance of the prediction model and the interpretability of the model in Section 4; Finally, discussions and conclusions are provided in Section 5.

2 Solar Flare Forecasting Model

In this paper, we assume continuous magnetogram frames could be used to predict the probability of solar flares with different classes in the active region. Therefore, solar flare prediction is regarded as a prediction problem with continuous frames of images and the level of solar flare is a random variable. We use the Bayesian spatio-temporal connection solar flare forecasting model to output the probability of solar flares of different classes, which mainly includes two parts: **Feature extractor** and **Forecasting model**. Feature extractor extracts the spatial features of each magnetogram frame and its correlation in the temporal dimension. Forecasting model is used to forecast the maximum X-ray brightness. If we set different threshold, we could output the probability of solar flares with different classes. The overall structure is shown in Figure 1.

2.1 The Feature Extractor

Extracting spatiotemporal features from continuous magnetogram frames is particularly important for the solar flare prediction task. The combination (LRCN) [5] of convolutional neural networks

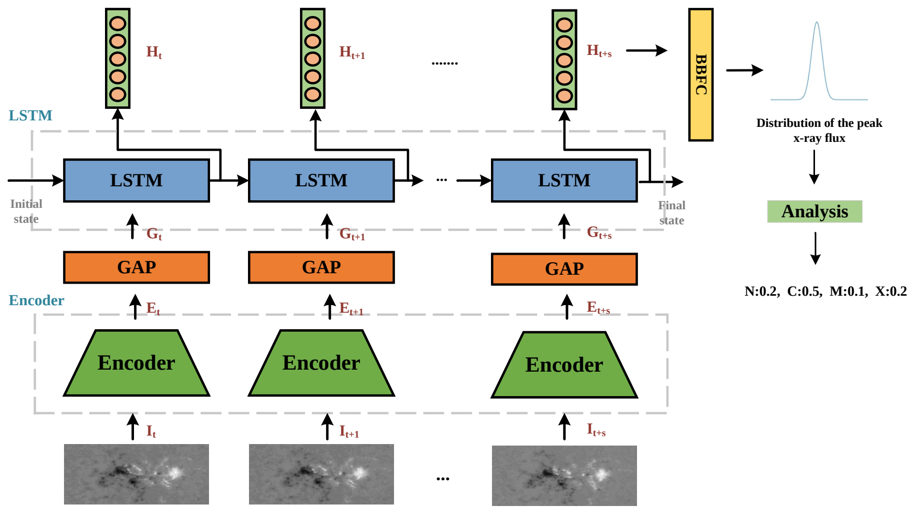

Figure 1: **Data-driven interpretable solar flare forecasting model**. The input of the model are continuous magnetogram frames (6 hours long) of previous time (24 hours before), and the output of the model is the maximum brightness of prediction time period. **Feature extractor** extracts the spatial features of each frame of magnetic map and its correlation in the temporal dimension. **Forecasting model** uses the extracted features to predict the maximum brightness, and we would output prediction results according to the maximum brightness later.

(CNNs) and long short-term memory neural networks (LSTM) is mainly used to fully consider the spatial features and its characteristics over time. The CNN is used to extract image features and the LSTM is used to analyze the correlation of feature sequences in the temporal dimension. The LRCN usually requires the size of input images to be the same. However the size of magnetogram frames would be different. Therefore, we have modified the LRCN by adding a GAP between the CNN and the LSTM. The structure of the feature extraction part includes the following part:

1. Convolutional neural networks (CNNs) are suitable for extracting the basic features of images and reducing the complexity of the model. Therefore, CNNs are used as the spatial feature extraction network of magnetograms, namely **Encoder**, to capture the spatial interaction, maintain the spatial continuity of the image, and extract the spatial features of magnetograms. The structure of the Encoder is shown in the Figure 2, which consists of four convolution block composed of convolution layer, instance normalization layer and Relu activation function.

2. Global average pooling (GAP)[10] can compress spatial features extracted by Encoder, pool images of different sizes to the same size, and reduce the number of parameters in the model to prevent overfitting. The specific operation is to perform global average pooling on the feature map of each channel to obtain a value, as shown in the Figure 2.

3.Long Short Term Memory(LSTM) extracts temporal variations of spatial features of magnetograms, so as to study the correlation of continuous magnetic field in the temporal dimension. This paper use a single-layer bidirectional LSTM structure to extract temporal features, which contains a hidden layer with 256 nodes and the input data are the spatial features of the continuous magnetic map extracted by Encoder.

2.2 The Solar Flare Prediction Model

The prediction model is used to forecast solar flare classes according to features extracted by the **Feature Extractor**. Since the solar flare occurrence and the level of solar flares are random variables, we propose to use the Bayesian neural network based on probability reasoning for solar flare prediction. The Bayesian neural network is regularized by introducing uncertainty into the weight of the neural network, which could achieve a balance between underfitting and overfitting

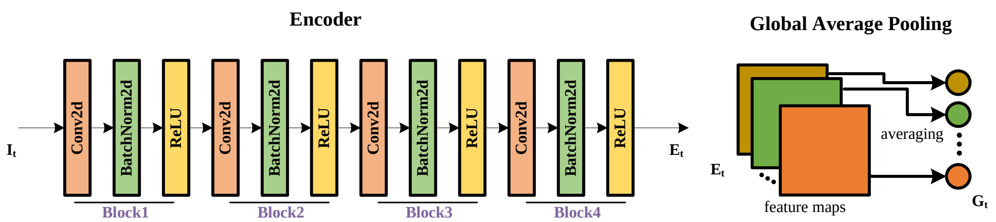

Figure 2: The structure of **Encoder** and **GAP**. After the input single frame image data (C, H, W) passes through four convolution blocks, the image size becomes ($\frac{H}{16}$, $\frac{W}{16}$), and the number of channels becomes 256. The parameter settings of the convolution kernel in each convolution block are respectively: Block1 (32, 5×5, 2),Block2 (64, 3×3, 2), Block3 (128, 3×3, 2), Block4 (256, 3×3, 2).

by learning the probability distribution over the weights of the neural network. Here, the Bayesian neural network is used as to predict the distribution of the maximum X-ray brightness, which uses two layers of bayesian full connection as the output block of the model. If we set the prediction model with different thresholds, we would get the probability of solar flares with different classes.

3 Interpretability analysis

The solar flare prediction model can learn the statistical relationship between continuous magnetograms and solar flares, and regions related to solar flares could be obtained according to the attention area of the feature extraction model. These regions could be further analyzed to obtain the mechanism that triggers solar flares. Grad-CAM [10] can help us analyze the attention area of the model according to input magnetograms.

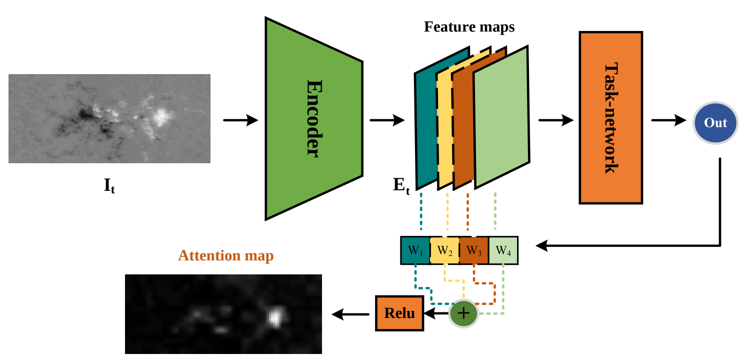

Figure 3: The structure of Grad-CAM.

Regression task is taken as the basic task of the model, and the specific flow of Grad-cam is shown in Figure 3. The model first conducts forward propagation to obtain the feature layer E and the network predictive value Out. Then the back propagation of Out can get the gradient information of the feature layer. By calculating the importance of each channel in the feature layer, and then weighted summing through $ReLU$, the final result is Grad-CAM, as shown from Equation 1 to 2:

$$L_{Grad-CAM} = ReLU(\sum_k w_k E^k), \tag{1}$$

Where E stands for the feature layer output by encoder, k represents the k channel in feature layer E, E^k stands for the data of channel k in feature layer E, and w_k stands for the weight of network output result for E^k.

$$a_k = \frac{1}{Z}\sum_i \sum_j \frac{\partial y}{\partial E_{ij}^k} \tag{2}$$

Where y stands for the predicted value of the network output, K_{ij}^k represents the data at ij of feature layer E in channel k, Z is equal to $H \times w$. The above equation is used to calculate attention map of a single image and we would carry out weighted summation of features to obtain Grad-CAM of the model for different frames.

4 Experiments and Results

4.1 Introduction to the Data Set

At present, solar active region data mainly is obtained from the SDO/HMI data and the SOHO/MDI data. Since 1996, these two types of data have provided consistent, high-quality solar active region data. Among them, the SOHO/MDI data starts and ends from January 1st. 1996 to April 12th. 2011. The SDO/HMI project is the successor of the SOHO/MDI project. It has accumulated more than 10 years of data since it started daily observation on April 30th. 2010, and its sampling frequency is 12 minutes. According to the peak value of soft X-ray flux observed by the Geostationary Operational Environmental Satellite System (GOES), solar flares can be classified into A, B, C, M, and X classes, with the energy released increasing successively. The values of each class represent the specific values of the peak X-ray flux. The data can be downloaded from (https://www.ngdc.noaa.gov/stp/space-weather/solardata/solar-features/solar-flflares/x-rays/goes/xrs/).

In this paper, SDO/HMI data with high image data quality are used as magnetogram data. The data can be downloaded from the SHARPs data sequence in joint scientific operations center (JSOC). We have downloaded 2010 data. The data of 5-2021.5 (the time interval is 1h) is used as the magnetic field data of the active area. During the loading process, the SHARP parameter of the active area is used to check the quality of observation data. The data used in our model has to be 1) disambiguated with a version of the disambiguation module greater than 1.1, 2) taken while the orbital velocity of the spacecraft is less than 3500 m/s, 3) of a high quality and 4) within 70 degrees of central meridian.

According to GOES, there are 839 active flares larger than A1.0 that meet these requirements between May 2010 and May 2021. Since solar flares of A and B classes are not significant, we study the flares with classes greater than C in the training data. In total, there are 5252 flares greater than C1.0 selected in the dataset. The data amount of all kinds of data is shown in Table 1.

Table 1: The classes of solar flares, sample sizes, and label of each type of solar flare

Flare Thresholds	Actual number	label	Train\Test(6,24)
N	5252	0	2299 \2 55
$10^-6 \leq C < 10^-5$	4752	100×n	1890 \ 210
$10^-5 \leq M < 10^-4$	468	1000×n	979 \ 109
$10^-4 \leq X$	4752	100 × n	472 \ 52

Previous studies assume flare prediction as a classification task, and there is no relationship between solar flares of different classes. In fact, solar flares of different classes have particular relations and we assume it as a regression problem here. According to the peak value of flux in soft X-ray band to generate label, different classes of flares set different magnitudes, the specific values are shown in Table 1 , where n represents the value of the X-ray peak flow. Meanwhile, this label setting makes large flares with small amount of data have a bigger loss, which can prevent the overfitting problem caused by data imbalance.

According to the number of solar flares of different levels, multiple time points are randomly selected from each active area as the starting point of our prediction time,The label is generated according to whether a flare will occur in the predicted time period, and the data is loaded according to the prior time period. In the data set, we randomly selected 10% of the active area data as the test set, and the rest as the training set.

4.2 Experiment Result

Solar flares forecast model based on data driven, using a large amount of observation data with labels to have supervision and training of forecast model, study the magnetic figure and the statistical relationship between solar flares occur, thus to forecast the flare. We also further used the class activation mapping (CAM) to analyze the triggering mechanism of solar flares. In this paper, we investigate the performance of the network using 6 hours of continuous magnetic field observations to predict the probability of various flares occurring within the next 24 hours. The network is trained using the training set, and when the training process is completed, the performance of the prediction model can be estimated through the test set.

4.2.1 Prediction Results of the maximum X-ray brightness

We use the Bayesian neural network as the last layer of the network to predict the distribution of the highest brightness in the next 24 hours. Since the Bayesian network will learn the probability distribution as the weight of the neural network, we will get different highest brightness values for a single sample, forming a distribution, as shown in Figure 4. We show the results of four types of N, C, M and X prediction, and calculate the corresponding probability of flare occurrence based on the threshold value defined in Table 1. The title is the time range of the prior time period entered, the upper left corner is the probability of occurrence in the next 24 hours, the mean and standard deviation. It should be noted that since the output result of the model is the distribution of the maximum X-ray brightness, we can set different thresholds to study the prediction results.

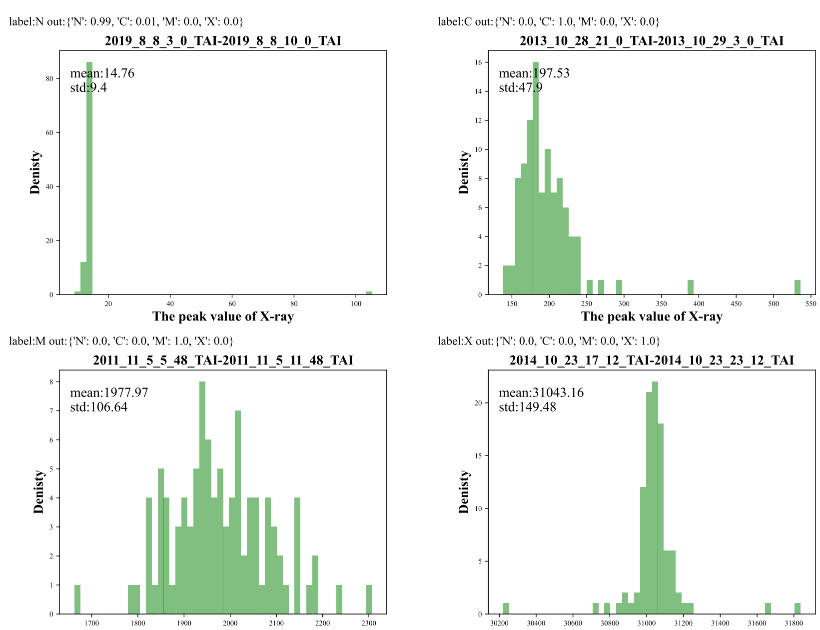

Figure 4: The results of four types of N, C, M and X prediction

4.2.2 Classification results

We randomly select 50 solar flares of different classes in the test set to investigate the overall performance of the prediction model. For each sample, the result of the model output is the distribution of the highest brightness in the next 24 hours. We choose the class of the solar flare with the highest flare probability as the classification result of our neural network.

For the binary classification task of whether a solar flare occurs or not, we set different thresholds to study the prediction results of the model. As shown in the following table 2, TP rate is the percentage of positive instances correctly classified, TN rate is the percentage of negative instances correctly classified. Pre is the percentage of how many positive predictions are actually positive. Allacc is the percentage of all correctly classified samples in the total sample. The table shows the comparison between our algorithm and other traditional algorithms, from which it can be seen that our model can achieve more than 90 $\%$ accuracy for different threshold divisions, and maintain a low false positive rate.

Table 2: The binary classification result of model

	label	TPrate	TNrate	Pre	Allacc	Train	Test
Huang^{2018}	$\geq C$	0.73	0.76	0.35		2010-2015	1996-2010
	$\geq M$	0.85	0.81	0.1			
	$\geq X$	0.87	0.85	0.015			
Park^{2018}	$\geq C$			0.84	0.83	2009-2017	1996-2008
Tang^{2021}	$\geq C$	0.878	0.82	0.131			2010-2015(20$\%$)
	$\geq M$	0.817	0.84	0.464			
This work(BBFC)	$\geq C$	0.917	0.959	0.986	0.949		2010-2015(10$\%$)
	$\geq M$	0.910	0.858	0.867	0.884		
	$\geq X$	0.920	0.993	0.979	0.979		
This work(FC)	$\geq C$	0.916	0.877	0.945	0.914		2010-2015(10$\%$)
	$\geq M$	0.830	0.858	0.855	0.844		
	$\geq X$	0.910	0.926	0.863	0.919		

We further investigate the performance of our algorithm in prediction solar flares of different levels. The confusion matrix shows the difference between the actual and predicted values, with the elements on its main diagonal corresponding to the correct classification, while other elements show how many samples in a category are incorrectly assigned to other categories as a proportion of the total number of categories. It shows us intuitively how many solar flares are correctly predicted. It can be seen from the table 3 that the model makes good predictions for no flares (N) and large flares (X), while for solar flares with class of C is more likely to be predicted as solar flares with classes of M, and solar flares with class of M is more likely to be predicted as solar flares with class of C.

Table 3: Multiple classifications results of BBfc

label\ label	N	C	M	X
N	0.92	0.08	0.00	0.00
C	0.10	0.58	0.32	0.00
M	0.00	0.12	0.86	0.00
X	0.00	0.00	0.02	0.98

In order to verify the performance of Bayesian neural network (BNN), we also transform bayesian full connection into full connection neural network for comparison, as shown in the table 4. The results show that Bayesian neural network has better prediction performance, can better distinguish C and M class flares, improve the accuracy of prediction model.

Table 4: Multiple classifications results of FC

label\ label	N	C	M	X
N	0.90	0.10	0	0.00
C	0.16	0.56	0.24	0.00
M	0.00	0.34	0.60	0.06
X	0.00	0.00	0.06	0.94

4.3 Attention Area of the Solar Flare Forecasting Model

The solar flare prediction model automatically extracts features related to solar flares. If the model can correctly classify solar flares, it can be considered that the model has extracted effective features. Here, we use Grad-CAM to study the features extracted by the model. Figure 5 shows a frame of magnetograms in AR12192, the attention area of the model on the map, and the multiplication result of these two images. It can be seen that the attention area of the model is mainly in the polarity reversal region and the strong magnetic field region. Figure 6 shows the variation of the magnetogram within 24 hours before the occurrence of a large solar flare (the time interval is 3h). This figure shows variations of the model's attention area, which is consistent to theoretical predictions.

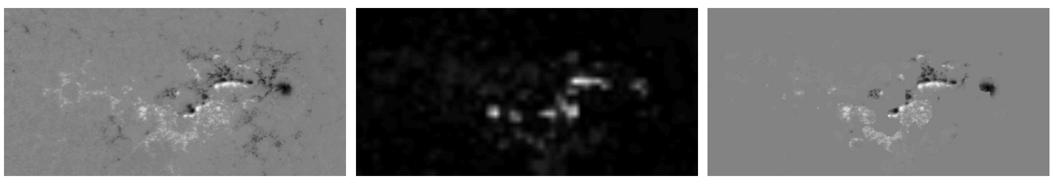

Figure 5: Magnetogram, Heatmap, and Heatmap $\times$ Magnetogram for active regions 12192.

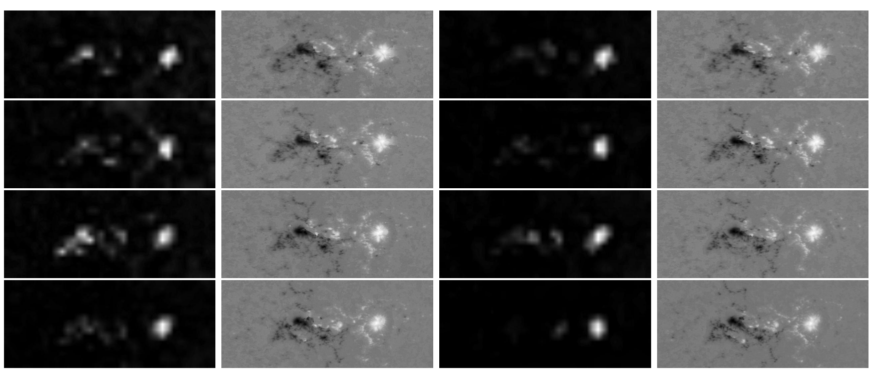

Figure 6: The change of the Magnetogram and Heatmap, within 24 hours before the occurrence of a large flare (the time interval is 3h).

5 Future work

We propose a data-driven solar flare prediction model, which can extract effective features from continuous Magnetograms to predict distribution of maximum soft X-ray brightness in active regions over the next 24 hours. Results show that our model could obtain statistical prediction of solar flares with more than 90 % accuracy, and compared with other algorithms, our algorithm could predict solar flares with lower false positive rates and higher ture positive rates. Besides, our model could extract effective features for solar flare predictions, which include polar reversal regions, strong magnetic field regions and the magnetic field conversion area. Our model could help scientists to discover important features that would trigger solar flares to promote the study of the mechanism that would trigger solar flares.

References

[1] Yoshua Bengio. Practical recommendations for gradient-based training of deep architectures. In *Neural networks: Tricks of the trade*, pages 437–478. Springer, 2012.

[2] PL Bornmann and D Shaw. Flare rates and the mcintosh active-region classifications. *Solar physics*, 150(1):127–146, 1994.

[3] Yang Chen, Ward B Manchester, Alfred O Hero, Gabor Toth, Benoit DuFumier, Tian Zhou, Xiantong Wang, Haonan Zhu, Zeyu Sun, and Tamas I Gombosi. Identifying solar flare precursors using time series of sdo/hmi images and sharp parameters. *Space Weather*, 17(10):1404–1426, 2019.

[4] Zhou Chen, Mingwu Jin, Yue Deng, Jing-Song Wang, Heng Huang, Xiaohua Deng, and Chun-Ming Huang. Improvement of a deep learning algorithm for total electron content maps: Image completion. *Journal of Geophysical Research: Space Physics*, 124(1):790–800, 2019.

[5] Jeffrey Donahue, Lisa Anne Hendricks, Sergio Guadarrama, Marcus Rohrbach, Subhashini Venugopalan, Kate Saenko, and Trevor Darrell. Long-term recurrent convolutional networks for visual recognition and description. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2625–2634, 2015.

[6] Ronald Gordon Giovanelli. The relations between eruptions and sunspots. *The Astrophysical Journal*, 89:555, 1939.

[7] Michael L Goodman, Chiman Kwan, Bulent Ayhan, and Eric L Shang. A new approach to solar flare prediction. *Frontiers of Physics*, 15(3):1–27, 2020.

[8] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4700–4708, 2017.

[9] Xin Huang, Huaning Wang, Long Xu, Jinfu Liu, Rong Li, and Xinghua Dai. Deep learning based solar flare forecasting model. i. results for line-of-sight magnetograms. *The Astrophysical Journal*, 856(1):7, 2018.

[10] Min Lin, Qiang Chen, and Shuicheng Yan. Network in network. *arXiv preprint arXiv:1312.4400*, 2013.

[11] ZongJun Ning. Power conversion factor in solar flares. *Chinese Science Bulletin*, 57(12):1397–1404, 2012.

[12] Eunsu Park, Yong-Jae Moon, Seulki Shin, Kangwoo Yi, Daye Lim, Harim Lee, and Gyungin Shin. Application of the deep convolutional neural network to the forecast of solar flare occurrence using full-disk solar magnetograms. *The Astrophysical Journal*, 869(2):91, 2018.

[13] AV Stepanov and VV Zaitsev. The challenges of the models of solar flares. *Geomagnetism and Aeronomy*, 56(8):952–971, 2016.

[14] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1–9, 2015.

[15] Rongxin Tang, Fantao Zeng, Zhou Chen, Jing-Song Wang, Chun-Ming Huang, and Zhiping Wu. The comparison of predicting storm-time ionospheric tec by three methods: Arima, lstm, and seq2seq. *Atmosphere*, 11(4):316, 2020.

[16] JingXiu Wang. Solar activity studies: From a magnetohydrodynamics description to a plasma perspective. *Chinese Science Bulletin*, 57(12):1362–1368, 2012.

[17] MS Wheatland. A bayesian approach to solar flare prediction. *The Astrophysical Journal*, 609(2):1134, 2004.

[18] Yanfang Zheng, Xuebao Li, and Xinshuo Wang. Solar flare prediction with the hybrid deep convolutional neural network. *The Astrophysical Journal*, 885(1):73, 2019.