Data-driven explainable Solar Flare Forecasting Model

Anonymous Author(s)
Affiliation
Address
email

Abstract

Solar flares are the most violent activities in the solar system, which are generated by evolutions of magnetic field in solar active regions. However, the triggering mechanism of solar flares is still an active research area and many algorithms are proposed to forecast solar flares based on different models In this paper, we propose a novel data-driven and explainable algorithm, which could not only achieve over 90% precision rate and low false alarm rate to forecast solar flares, but also show the show the mechanism in forecasting solar flares. The algorithm proposed in this paper is built with CNNs and LSTM to percept continuous magnetic field observation data with 6 hours long. At the last layer, a Bayesian neural network is used to predict the probability of flares with different levels in the next 24 hours. We propose to use the class activation mapping (CAM) to analyze the mechanism in triggering of solar flares. Through the analysis of attention map, it is concluded that the prediction model mainly focuses on the strong magnetic field area, the polarity reversal line and the magnetic field conversion area, which is consistent to theory predictions

1 Introduction

The changes of Earth space environment due to solar activity, which will bring catastrophic effects to aerospace, communication navigation, oil pipelines and power grids etc. Solar flares are one of the most important forms of solar activity. 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.

So far, the specific physical mechanism of solar flare eruption has not been very clear, and the forecasting model of solar flare based on physics is still in the research stage [13, 11, 16]. The existing prediction model mainly relies on the statistical relationship between flare eruption records and historical observation data. Traditional flare forecasting models mainly rely on manual experience and domain knowledge to extract characteristic parameters related to solar flares from observed data to forecast solar flares[6, 2, 17, 1, 7]. It is difficult to make full use of the information related to solar flares contained in massive observational data.

In recent years, machine learning technology has been developed, especially the data modeling method represented by deep learning, which can automatically extract effective features from data, providing a new way for flare prediction research. Studies have shown that most of the solar activity parameters used for solar flare prediction are extracted from solar magnetogram, and there are many features related to solar flare eruption in solar magnetogram. In deep learning, convolutional neural networks can automatically extract effective features from images, so many researchers have applied convolutional neural networks [8, 14, 12, 9, 18] to the establishment of solar flare prediction models.

Submitted to NeurIPS 2021 AI for Science Workshop.

Meanwhile, the dynamic change of the solar active area over time is closely related to the eruption of solar flares, and the long short-term memory network is often used to process sequence data to analyze the correlation of features in the temporal dimension [3, 4, 15].

These methods perform well on classification problems, but they still have their own limitations in capturing the evolution of solar flares, mainly reflected in: 1.The characteristics of physical parameters of continuous solar magnetic field activity region over time are not considered; 2.The sloar flares prediction problem is regarded as a deterministic problem without considering the random and sudden characteristics of flares; 3.Input data needs to be preprocessed before being input into the model.

A Bayesian spatio-temporal connection solar flare forecasting model based on deep learning is proposed to predict the probability of various solar flares in the next 24 hours. Meanwhile, we use the class activation mapping technology to draw the focus area of the model and analyze the mechanism in triggering of solar flares. Moreover, our method does not require pre-processing of input data.

This article is orginzed 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 Forecasting model

The study shows that continuous solar magnetic field data can be used to predict the probability of future flares in the active region. Therefore, Therefore, solar flare prediction is regarded as a video prediction problem, and the occurrence of solar flare is random and unexpected. Bayesian spatio-temporal connection solar flare forecasting model based on deep learning is proposed to achieve the probability of projections for the future of all kinds of the occurrence of flares. The model mainly includes two parts: **Feature extractor** and **Forecasting model**. Feature extractor extracts the spatial features of each magnetogram frame and its correlation in the time dimension. Forecasting model to forecast the maximum brightness distribution, after by setting different threshold, to give the probability of all kinds of solar flares occur. The overall structure is shown in Figure 1.

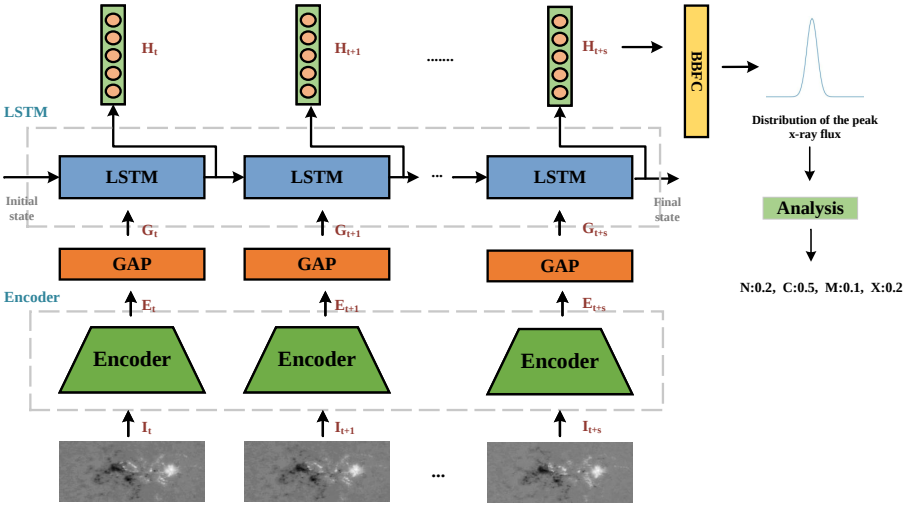

Figure 1: **Data-driven interpretable solar flare forecasting model**. The input and output of the model are respectively the continuous magnetic field data of prior time period, and the maximum brightness of prediction time period. **Feature extractor** extracts the spatial features of each frame of magnetic map and its correlation in the temporal dimension. **Forecasting model** uses the extracted features to predict the maximum brightness, and further analyzes the output results of the model to obtain the probability of various flares.

2.1 Feature extractor

The extraction of spatiotemporal features of continuous solar magnetic field data is particularly important because continuous field observations are the number of frames of video sequences. In the existing deep learning methods, the combination (LRCN)[5] of convolutional neural networks(CNNs) and long short-term memory(LSTM) neural network is mainly used to fully consider the spatial features in the video and its characteristics over time. CNNs is used to extract image features, and LSTM is used to analyze the correlation of feature sequences in the temporal dimension. LRCN usually requires the size of the input picture to be fixed to ensure the link between CNNs and LSTM. However, we did not want to do anything with the magnetic field data, so we modified the LRCN by adding a GAP between the CNN and the LSTM. The specific feature extraction process is as follows:

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 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, pooling the data of different scales to the same size, and reducing the number of parameters and calculation 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**

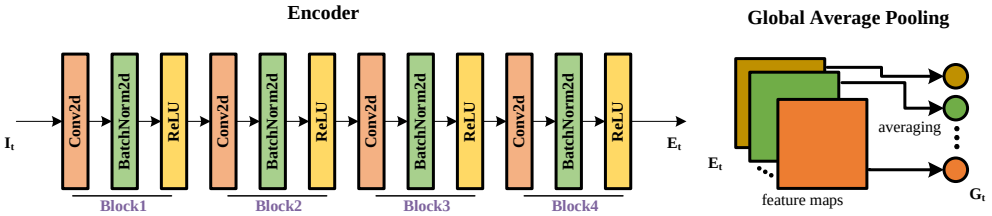

Figure 2: The structure of **Encoder** and **GAP**. After the input single frame image data (C, H, W) passes through four convolution blocks, the image size becomes ($\frac{H}{16}$, $\frac{W}{16}$), and the number of channels becomes 256. The parameter settings of the convolution kernel in each convolution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

Memory(LSTM) extracts the change information of spatial features of magnetic maps, so as to study the correlation of continuous magnetic maps in the dimension of time. This paper use a single-layer bidirectional LSTM structure, 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 Prediction model

The prediction model is required to have the ability of uncertainty modeling and inference when predicting flare, because of the randomness and suddenness of flare occurrence and the existence of general uncertainty. Bayesian network based on probability reasoning has great advantages in solving the uncertainty. It is a neural network model regularized by introducing uncertainty into the weight of the neural network. This model achieves a balance between underfitting and overfitting by learning the probability distribution over the weights of the neural network. Here, Bayesian neural network is used as the maximum X-ray brightness prediction model to introduce uncertainty into the flare prediction model, predict the distribution of the maximum X-ray brightness. And then the probability of each type of flare is given by setting different thresholds.

3 Interpretability analysis

In a data-driven way, the flare prediction model can learn the statistical relationship between continuous magnetic field observation data and solar flares, and the region related to the solar flares can be obtained according to the focus area of the feature extraction model, so as to further analyze the characteristics of flare triggering. Grad-CAM [10] can help us analyze the focus area of the model for the output results. Through the focus area of the model, we can in turn analyze whether the model learns the right features or information, or find some rules from it.

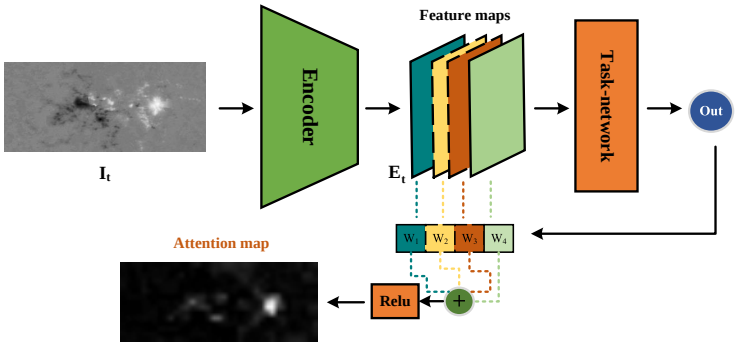

Figure 3: The structure of Grad-CAM.

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 in Figure 2. The formula is shown below:

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

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 is for the case that a single image is taken as input, but the data we input is video data, so when carrying out weighted summation, it is necessary to carry out weighted summation of features of the same frame to obtain Grad-CAM of the model for different frames.

4 Experiment design and result analysis

4.1 Production of data sets

At present, solar active region data mainly come from SDO/HMI data and SOHO/MDI data. Since 1996, these two types of data have provided consistent, high-quality data on solar active regions. Among them, the SOHO/MDI data starts and ends from January 1 1996 to April 12 2011. SDO/HMI is the successor of SOHO/MDI. It has accumulated more than 10 years of data since it started daily observation on April 30 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 following 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 for continuous magnetic field data. The data can be from joint scientific operations center (JSOC) database SHARPs data sequence. We downloaded 2010. 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 of space weather HMI is used to delete the active area.The data 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.

According to GOES, there were 839 active flares larger than A1.0 that met these requirements between May 2010 and May 2021.05. Since the influence of A and B is not significant and insignificant, we study the flares with flare grades greater than C, M and X, and A total of 5252 flares greater than C1.0 have occurred. The data amount of all kinds of data is shown in Table 1. For the production of

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

Flare Thresholds	Actual number	label	Train\Test
N	5252	0	2299\255
$10^-6 \leq C < 10^-5$	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

labels, previous studies have used flare prediction as a classification task, and there is no relationship between classes. In fact, the flare outburst problem we study has a certain size relationship between classes, so we regard it as a regression problem here. Our label is generated according to the peak flow of soft X-ray, and the specific value is shown in Table 1 . At the same time, due to the imbalance of three types of C, M and X data, setting such a value is also conducive to reducing the result bias caused by data imbalance.

According to the number of flares occurring in each active area, multiple time points are randomly selected from each active area as the starting point of our prediction time, and then the data sets and corresponding labels are synthesized according to the prior and posterior time periods. 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

Based on the data-driven solar flare prediction model, a large number of labeled observation data are used to train the forecast model, and the statistical relationship between magnetic map and solar flare occurrence is learned, so as to forecast the solar 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 training set is used to train the network, and when the training process is complete, the performance of the prediction model can be estimated by the test set.

4.2.1 The distribution 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 on 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 given by the way the tag is made. 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 here 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.

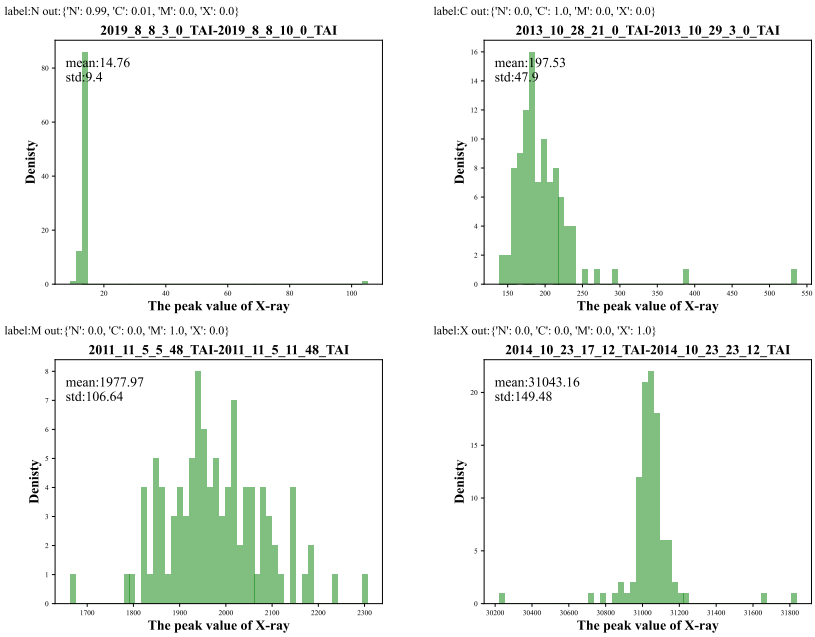

Figure 4: T he results of four types of N, C, M and X prediction

4.2.2 Classification results

We randomly sampled 50 samples from each type of solar flare 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, and further by setting the threshold, the probability of different flares is obtained. In the analysis process, we choose the flare type with the highest flare probability as the output to study the classification results of the network.

For the binary classification task of whether a 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 classifiifiedTN rate is the percentage of negative instances correctly classifiified. Precision 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 different 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.

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 the other elements tell us how many samples in a category are incorrectly assigned to other categories as a proportion of the total number of categories. It tells us intuitively how much data is correctly predicted and which data is not correctly predicted. So here we use confusion matrix to study the multi-classification results of the model. It is worth noting that specific categories are divided according to labels. It can be seen from the table 3 that the model makes good predictions for no flares (N) and large flares (X), while for small flares C is more likely to be predicted as M, and M is more likely to be predicted as C.

4.2.3 Attention area

Based on the data-driven model, the model automatically extracts features related to flares from the data to predict flares. If the model can correctly predict the classification result of flares, it can be

Table 2: The binary classification result of model

	Positive label	TPrate	TNrate	Precision	Allacc
Huang^{2018}	label$\geq C$	0.73	0.76	0.35	
	label$\geq M$	0.85	0.81	0.1	
	label$\geq X$	0.87	0.85	0.015	
Park^{2018}	label$\geq C$			0.84	0.83
Tang^{2021}	label$\geq C$	0.878	0.82	0.131	
	label$\geq C$	0.817	0.84	0.464	
Our	label$\geq C$	0.917	0.959	0.986	0.949
	label$\geq M$	0.910	0.858	0.867	0.884
	label$\geq X$	0.920	0.993	0.979	0.979

Table 3: The multiple classifications result of model

label\ label	N	C	M	X
N	0.92	0.08	0	0
C	0.10	0.58	0.32	0
M	0	0.12	0.86	0
X	0	0	0.02	0.98

considered that the model has extracted effective features. Here, we use Grad-CAM to study the features extracted by the model.

Figure 5 shows the magnetic map of a frame in AR12192, the focus area of the model on the map, and the result of the multiplication of the two. It can be seen that the focus area of the model is mainly in the polarity reversal region and the strong magnetic field region. Figure 6 shows the change of the magnetic map within 24 hours before the occurrence of a large flare (the time interval is 3h) to show the change of the model's focus area. This show the model also focuses on the magnetic field conversion area.

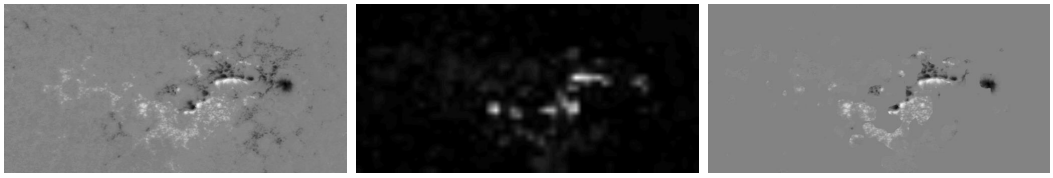

Figure 5: Magnetogram, Heatmap, and Heatmap $\times$ Magnetogram for active regions 12192.

5 Future work

We propose a data-driven solar flare prediction model, which can extract effective features from continuous magnetic maps to achieve distribution prediction of maximum brightness in active regions over the next 24 hours. Forecast model of the test results show that (1) can be the model for a single sample of highest brightness distribution prediction, can realize the characteristics of randomness, forecast model (2) the statistical results, further research model can realize not only predict the flare occurred or not more than 90 $\%$ accuracy, and compared with other algorithms to predict solar flares can be reduced by the rate of false positives. (3) By analyzing the effective features extracted from the model, we can find that the regions that the network pays attention to are mainly polar reversal regions, strong magnetic field regions and the magnetic field conversion area. The data-driven method is used to find out the region concerned by the model, so as to study the change of the magnetic field of the flare from zero to presence, and find some parameters related to the flare generation, to promote the study of the flare mechanism.

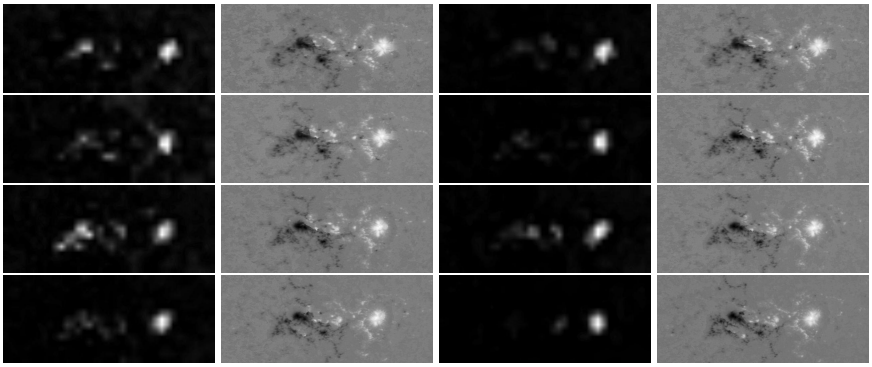

Figure 6: The change of the Magnetogram and Heatmap, within 24 hours before the occurrence of a large flare(the time interval is 3h).

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.