



Predicting Solar Flares Using a Convolutional Neural Network with Extreme-ultraviolet Images

Yun Yang¹ , Yi Wei Ni² , P. F. Chen^{2,3} , and Xue Shang Feng^{4,5} ¹ School of Mathematical Sciences, Ministry of Education Key Laboratory for NSLSCS, Nanjing Normal University, Nanjing 210023, People's Republic of China; 05444@njnu.edu.cn² School of Astronomy and Space Science, Nanjing University, Nanjing 210023, People's Republic of China; chenpf@nju.edu.cn³ Key Laboratory of Modern Astronomy and Astrophysics, Nanjing University, People's Republic of China⁴ Institute of Space Science and Applied Technology, Harbin Institute of Technology, Shenzhen 518055, People's Republic of China⁵ SIGMA Weather Group, State Key Laboratory for Space Weather, National Space Science Center, Beijing 100190, People's Republic of China

Received 2024 October 29; revised 2025 March 2; accepted 2025 March 9; published 2025 May 16

Abstract

Solar flares, as one of the most intense solar eruptive activities, can bring a destructive impact on the near-Earth space environment and technological infrastructure. Therefore, accurate and real-time forecast of the occurrence of flares is essential to reduce the potential damage to space facilities and navigation systems. Several types of flare forecasting methods have been proposed, e.g., the statistical methods, numerical methods, traditional machine learning, and the recently developed deep learning. In this paper, we apply one of the deep learning methods, the convolutional neural network ResNet-18, to the tricolor (94, 193, and 304 Å) full-disk extreme-ultraviolet images observed by the Solar Dynamics Observatory Atmospheric Imaging Assembly from 2012 to 2022 with $5 \times 24 \times 365 \times 11$ images and 10026 flares \geq C-class in total to build a solar flare forecasting model to predict whether a flare \geq C-class would occur within the next m hr ($m = 6, 12, 24, 48, \text{ and } 72$). The values of most performance evaluation metrics scores are over 0.8, with some even exceeding 0.9. The main advantage of this model lies in that it can effectively leverage the historical and spatial information of the source active regions as well as the interaction information from the surrounding active regions or other structures. This information, which cannot be provided by the data from a single active region alone, is demonstrated to be crucial in predicting solar flares. The results show that our prediction model performs well in both short-term and long-term forecasting.

Unified Astronomy Thesaurus concepts: [Solar physics \(1476\)](#); [Solar activity \(1475\)](#); [Solar flares \(1496\)](#)

1. Introduction

Solar flares represent dramatic occurrences where significant magnetic energy is abruptly released in the lower regions of the solar corona. These events can span from just a few minutes to several hours, appearing as radiant ribbons in the chromosphere and blazing, heated loops in the corona. When accompanied by formidable coronal mass ejections (CMEs; referenced in P. F. Chen 2011; D. F. Webb & T. A. Howard 2012), these intense flares can have profound and detrimental effects on the near-Earth space environment, endangering satellites and astronauts, disrupting radio communication networks, and impairing navigation systems, alongside other potential repercussions. The extreme events may even result in potential life-threatening consequences (D. Baker et al. 2004). Hence, efficient and accurate forecast is essential for reducing or averting the havoc brought by catastrophic space weather events to human activities.

The prediction methods of solar flares mainly include statistical methods, numerical methods, traditional machine learning methods, and the recent popular deep learning methods. Statistical methods have been widely used in solar flare forecasting since the 1930s. For example, H. Gallagher et al. (2002) and D. S. Bloomfield et al. (2012) developed a Poisson statistical model to estimate the probability of solar flares. Leka and collaborators (K. D. Leka & G. Barnes 2003a, 2003b, 2007; G. Barnes & K. D. Leka 2006; K. D. Leka et al. 2018) have

conducted a series of work on employing a statistical approach—discriminant analysis to ascertain the criticality of magnetic parameters in predicting solar flares through analysis of photospheric magnetic field data from some flare-quiet and flare-imminent active regions. Subsequently, G. Barnes et al. (2007) extended the approach to a probability forecast. K. D. Leka et al. (2023) further extended their previous works for the solar corona and chromosphere, and they utilized the moment-analysis method to produce parameters related to flare for ultraviolet (UV) and extreme-ultraviolet (EUV) image data. J. P. Mason & J. T. Hoeksma (2010) employed the superposed epoch analysis technique to elucidate statistical correlations between magnetic field parameters and the occurrence of solar flares. L. D. Krista (2025) utilized the Detection and EUV Flare Tracking tool (L. D. Krista & M. Chih 2021), which was developed based on a large statistical analysis of flares and automatically identifies flare precursors in EUV observations. Numerical methods, leveraging magnetohydrodynamic (MHD) numerical simulation—an important tool in space physics research (X. S. Feng 2020; Y. Yang et al. 2023)—are applicable for exploring the underlying physical properties. K. Kusano et al. (2004, 2012, 2020) systematically surveyed the possible triggering mechanism of flare by a wide variety of magnetic structures in terms of three-dimensional (3D) MHD simulations. Subsequently, Y. Bamba et al. (2013, 2017) conducted some work to examine the model by K. Kusano et al. (2012) and present some new conclusions regarding the triggering mechanism of solar flares using data from four major flares observed by the Hinode satellite. C. Jiang & X. Feng (2013), C. W. Jiang et al. (2016), and C. Jiang et al. (2018) executed 3D MHD simulations, leveraging the nonlinear force-free field



Original content from this work may be used under the terms of the [Creative Commons Attribution 4.0 licence](#). Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

extrapolation of vector magnetograms or the data-driven active region evolution model (C. W. Jiang et al. 2012) to reproduce the initiation and subsequent development of eruptive flux ropes and successfully mirror some evolutionary characteristics of flares. However, the triggering mechanism of solar flares remains incompletely understood, and the numerical method is substantially hindered to predict the occurrence of solar eruptions (K. Kusano et al. 2012, 2020; D. Sun et al. 2023). In recent years, due to the rapid improvement of computing capability, artificial intelligence algorithms have been introduced to predict solar flares and other space weather problems.

In the early stage, traditional machine learning methods were used more frequently. The typical algorithms include the support vector machine (SVM; R. Li et al. 2007; M. G. Bobra & S. Couvidat 2015; N. Nishizuka et al. 2017), the artificial neural network (R. Qahwaji & T. Colak 2007; H. Wang et al. 2008; T. Colak & R. Qahwaji 2009; O. Ahmed et al. 2013), the ordinal logistic regression (H. Song et al. 2009), the Bayesian network approach (D. Yu et al. 2010), and the ensemble learning method (J. A. Guerra et al. 2015). The emergence of deep learning revolutionized the research of solar flare forecasting. It outcompetes traditional machine learning methods in many aspects, such as the automatic feature extraction, scalability, representational power, generalization capabilities, end-to-end learning, proficiency with unstructured data, and continuous learning abilities. All these characteristics enable deep learning to tackle more complex and diverse problems with greater accuracy and efficiency (G. E. Hinton & R. R. Salakhutdinov 2006; I. Arel et al. 2010; Y. LeCun et al. 2015; J. Schmidhuber 2015). For example, H. Liu et al. (2019) and X. Wang et al. (2020) employed the long-short-term memory (LSTM) method to predict whether an active region would produce a flare within the next 24 hr.

The above-mentioned statistical and traditional machine learning methods, as well as the LSTM method, mainly rely on morphological or physical parameters extracted from the source active regions. However, as claimed by C. J. Schrijver & A. M. Title (2011), many flares, especially those associated with CMEs, involve magnetic couplings as long as a full hemisphere. Magnetic parameters of neighboring active regions are nonnegligible in the final eruptions and even in their initiation. It implies that the parameters of the source active region are not sufficient, and part of the information is missing in the traditional approaches. In data mining, according to “Principles of Data Analysis for Beginners,”⁶ it is well acknowledged that data and features determine the upper limit of machine learning, while models and algorithms can only approximate the upper limit. Due to the limitations of the acquired feature information, the relevant methods are considered to have reached a bottleneck for solar flare forecasting. The deep learning methods used to process imaging data can automatically extract the local features and interconnecting features of each image. In particular, the convolutional neural network (CNN) deep learning method has been increasingly applied to imaging data processing. It transforms an input image into a feature vector, which is the image features extracted by the neural network (D. Sun et al. 2023). This extracted precursor information, rich in discriminative characteristics, emerges as a pivotal indicator for enhancing the accuracy of flare forecasting models.

Consequently, the efficacy of flare forecasting is intimately tied to the quality and relevance of the extracted features, as underscored by X. Huang et al. (2018), who applied the CNN method to automatically extract forecasting patterns from magnetograms of active regions to predict flares. Considering that solar flares result from the energy release of the complex coronal magnetic field. However, in magnetogram data, the coronal magnetic field is calculated using extrapolation models based on the photospheric magnetic field, which may involve numerical errors. In contrast, EUV observation data directly contain information about the coronal magnetic field. D. Sun et al. (2023) utilized multiwavelength EUV images of active regions to build a model for flare forecast and obtained some instructive conclusions.

However, in these works, parameters were retrieved from the active regions separately, and the impact of neighboring active regions is missing. Therefore, building forecast models with the full-disk multiwavelength EUV images is demanding. In addition, we use a composite of EUV images as the initial input rather than a simple mixture of individual multiwavelength EUV images to help provide global physical information, which makes it possible to encode active regions. Encoding active regions makes it possible to further integrate and analyze the historical information of these active regions, including the frequency, intensity, and location of the past flares, jets, and other events. This information is helpful to identify the level of activity in active regions and whether they are likely to produce intense solar activities in the future. Moreover, geographical information is also an important factor in predicting solar flare activities. The heliographical locations of active regions and their relative positions to each other can all influence the occurrence and propagation of solar flare events. By integrating this information, it can more accurately predict the occurrence time, intensity, and direction of solar flare activities. In contrast, data from a single active region often provide limited information, and many models did not retrieve the historical and heliographical information of the region. Therefore, utilizing full-disk EUV data and encoding active regions can dig out information more effectively, improving the accuracy and reliability of solar flare activity predictions.

This paper is organized as follows: The data retrieval is described in Section 2. A forecast model based on the deep learning method is proposed in Section 3. The performance of the proposed model is evaluated in Section 4. Section 5 displays the results and discussions. Finally, the conclusions are drawn in Section 6.

2. Data Acquisition and Preprocessing

The Solar Dynamics Observatory (SDO) Atmospheric Imaging Assembly (AIA) telescope provides EUV and UV images of the Sun (J. R. Lemen et al. 2012). It captures 4096×4096 pixels full-disk images in two UV wavelengths (centered at 1600 and 1700 Å), seven EUV wavelengths (centered at 94, 131, 171, 193, 211, 304, and 335 Å), and one visible wavelength (centered at 4500 Å), with 4096×4096 pixels. As shown in Figure 1, we select three EUV wavelengths (centered at 94, 193, and 304 Å) images for the solar flare forecast.

The images of these three channels represent the information of high-, medium-, and low-temperature segments of the low corona, respectively. To alleviate the computational cost, each image, whose original size is 4096×4096 pixels, is resized to lower pixels, i.e., 512×512 pixels (Figure 2). The simultaneous tricolor images are aligned by the world coordinate

⁶ <https://medium.com/@datasciencelove/principles-of-data-analysis-for-beginners-3397e558321d>

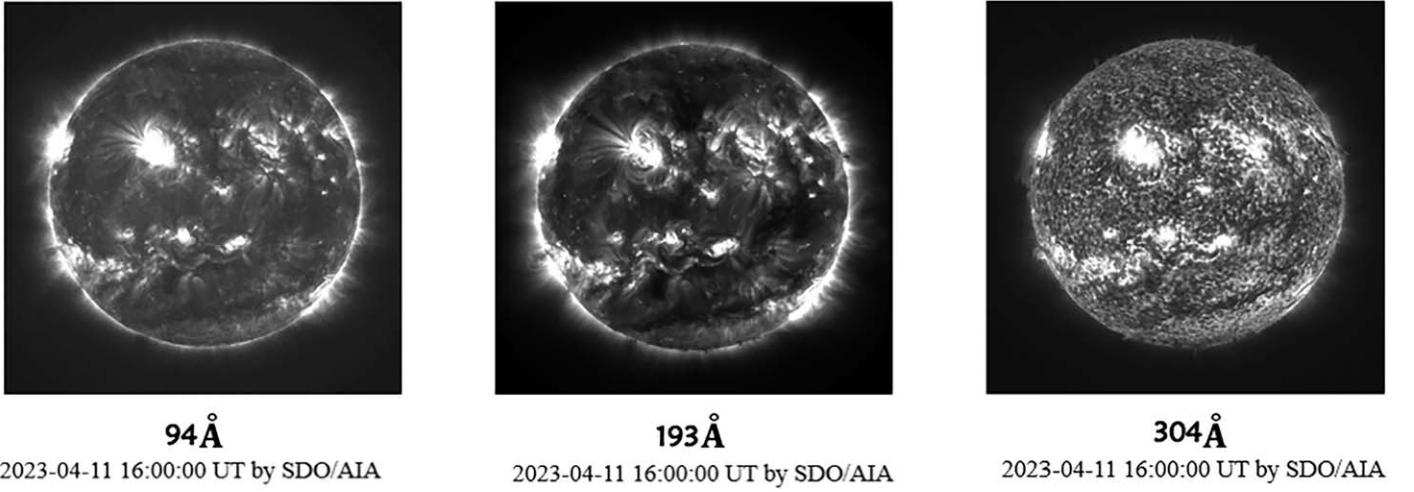


Figure 1. Tricolor (centered at 94, 193, and 304 Å) EUV images observed by the SDO/AIA at 16:00:00 UT of 2023 April 11 were selected to use in our model.

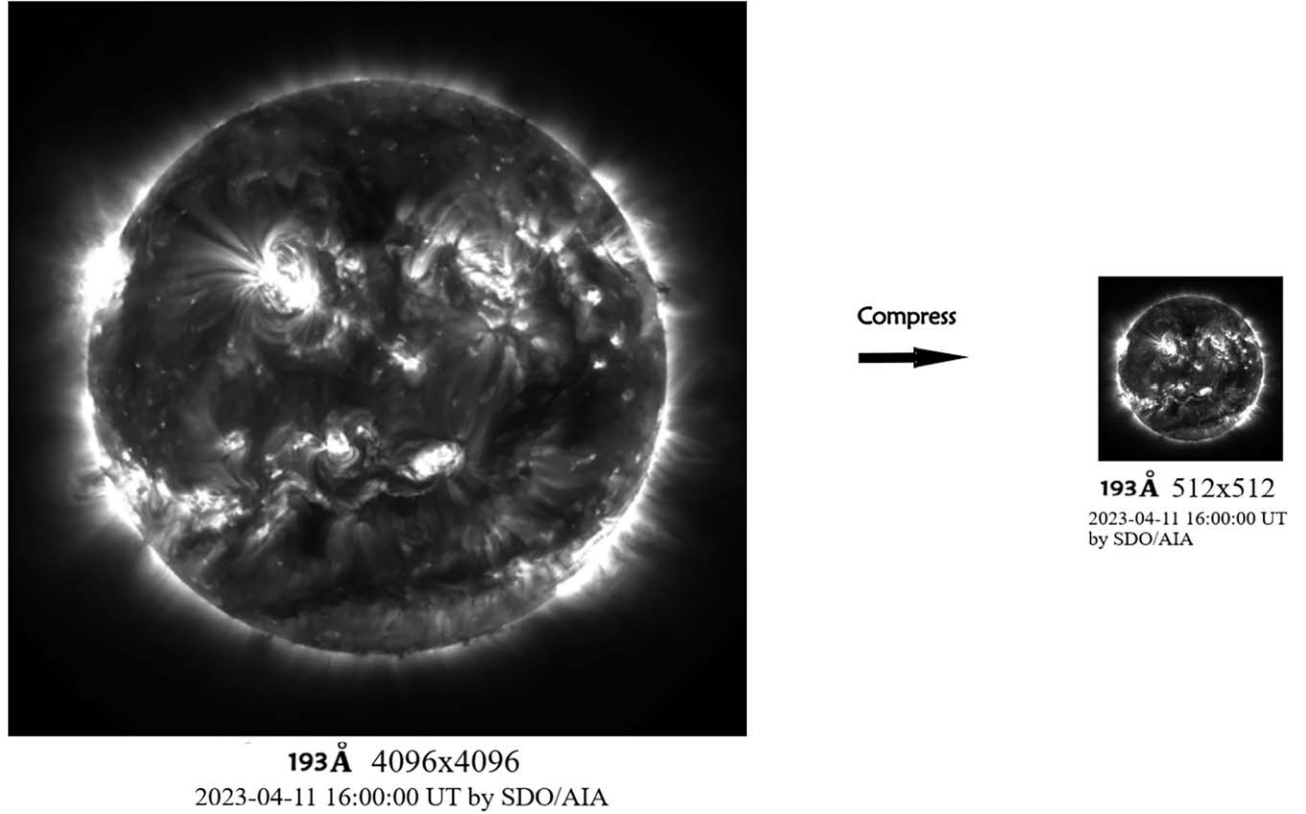


Figure 2. Each image is compressed from 4096×4096 to 512×512 pixels in order to increase the calculation efficiency.

systems standard so as to form a composite map as the input. The intensity of each EUV image is normalized by its maximum value, i.e., for each image, the normalized intensity ranges from 0 to 1.

Solar flares are classified as A-, B-, C-, M- or X-classes according to the peak intensity of the 1–8 Å soft X-ray flux observed by the Geostationary Operational Environmental Satellite system, with higher levels indicating greater energy release. In this study, we focus on the medium and large flares; therefore, the “flare” and “nonflare” events are defined as follows: (1) if there is at least a \geq C-class flare occurring at the observation time, then a “flare” label is assigned to the image;

(2) if no any flare with a magnitude \geq C-class occurs, namely, the A- or B-class or no flares occur at the observation time, then a “nonflare” label is assigned to the image. The flare information, including flare time and class, etc., was obtained from <https://www.ngdc.noaa.gov/stp/space-weather/solar-data/solar-features/solar-flares/x-rays/goes/xrs> and <https://data.ngdc.noaa.gov/platforms/solar-space-observing-satellites/goes>. The flare information during 2012–2016 were taken from the first link. The second link provided the 2017–2022 information. There are 10,026 flares \geq C-class in total during 2012 January 1 and 2022 December 31, including 9134 C-class, 842 M-class, and 50 X-class flares. After

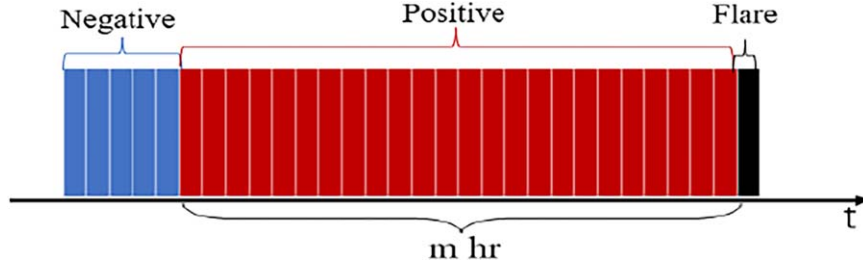


Figure 3. Construction of positive and negative samples used in the prediction task. Each rectangular box represents a data sample (namely, an image at a time). The black box indicates the onset time of a C-/M-/X-class flare. Data samples collected within m hr prior to the flare, represented by red rectangular boxes, belong to the positive class. The other data samples, represented by blue rectangular boxes, belong to the negative class.

obtaining the flare information, positive and negative samples were generated based on the tricolor (at 94, 193, and 304 Å) full-disk EUV images observed by the SDO/AIA from 2012 January 1 to 2022 December 31 with $5 \times 24 \times 365 \times 11$ images in total. We obtained these image data from Joint Science Operations Center (<http://jsoc.stanford.edu/>). These data are continuously collected with a frequency of every 12 minutes. Figure 3 demonstrates the process of constructing positive and negative samples. Each rectangular box represents a data sample (namely, an image at a time). The black box signifies the moment when a C-/M-/X-class flare initiates. Data samples gathered within m hr preceding the flare, depicted by red rectangular boxes, are classified as positive. Conversely, the remaining data samples, denoted by blue rectangular boxes, are categorized as negative. After excluding abnormal data, a total of 476,376 images' data are selected as samples and 276,144 positive samples in total were used for training and testing. To prevent information leakage, we used data from 2012 January 1 to 2019 December 31 as the training set and data from 2020 January 1 to 2022 December 31 as the testing set. It is essential to ensure that there is no overlap between the two sets. The sample sizes with different prediction lead times ($m = 6, 12, 24, 48$, and 72 hr) are listed in Table 1.

3. Deep Learning Model

CNN is a class of deep neural networks that are particularly suited for processing visual imagery. They have extensive application in a wide array of computer vision problems, such as object detection, image segmentation, and image classification. The key components of a CNN model are as follows:

(1) *Convolutional Layer*. This layer performs the convolutional operation, which is a special kind of linear operation. Convolutional filters (or kernels) slide across the input image, performing element-wise multiplication and then summing up the results to produce a feature map. These filters help the network learn different types of visual features, such as edges, corners, or textures.

(2) *Activation Function*. After the convolution operation, an activation function is applied to the feature map. Common activation functions include rectified linear unit (ReLU), sigmoid, and tanh. These functions introduce nonlinearity into the network, allowing it to learn complex patterns.

(3) *Pooling Layer (optional)*. Pooling layers reduce the spatial size of the feature map, thus reducing the number of parameters and computational complexity. The most common type of pooling is max pooling, which selects the maximum value from a specified region of the feature map.

(4) *Fully Connected Layer (optional)*. After several convolutional and pooling layers, the network may flatten the

Table 1
Results of Image Labeling with Different Prediction Lead Times

Total Data Set			
Prediction Lead Time	Positive #	Negative #	Positive Ratio
6 hr	141940	334436	0.298
12 hr	183286	293090	0.3848
24 hr	220701	255675	0.4633
48 hr	255502	220874	0.5363
72 hr	276144	200232	0.5797
Training set			
Prediction lead time	Positive #	Negative #	Positive ratio
6 hr	105256	240966	0.304
12 hr	136763	209459	0.395
24 hr	164416	181806	0.4749
48 hr	189679	156543	0.5479
72 hr	204343	141879	0.5902
Testing set			
Prediction lead time	Positive #	Negative #	Positive ratio
6 hr	36684	93470	0.2819
12 hr	46523	83631	0.3574
24 hr	56285	73869	0.4324
48 hr	65823	64331	0.5057
72 hr	71801	58353	0.5517

feature maps and feed them into fully connected layers. These layers are typically used to classify the input image based on the extracted features.

(5) *Loss Function*. A loss function is used to measure the difference between our predictions and the true labels. Common loss functions for classification tasks include cross-entropy loss and mean squared error loss.

(6) *Optimization Algorithm*. An optimization algorithm, such as stochastic gradient descent or its variants, is used to update the CNN parameters in order to minimize the loss function.

CNNs are highly effective at extracting hierarchical features from images, where lower layers capture simple features like edges and corners, while higher layers capture more complex features like object parts and shapes. This ability to capture characteristic features directly from raw pixels makes CNNs a powerful tool for various computer vision tasks.

ResNet-18 is a classic deep CNN model proposed by Microsoft Research Asia, which was used to participate in the ImageNet image classification competition in 2015. In the name of this model, “18” refers to the number of layers with weights, totaling 18 layers (including convolutional layers and fully connected layers but excluding pooling layers, batch

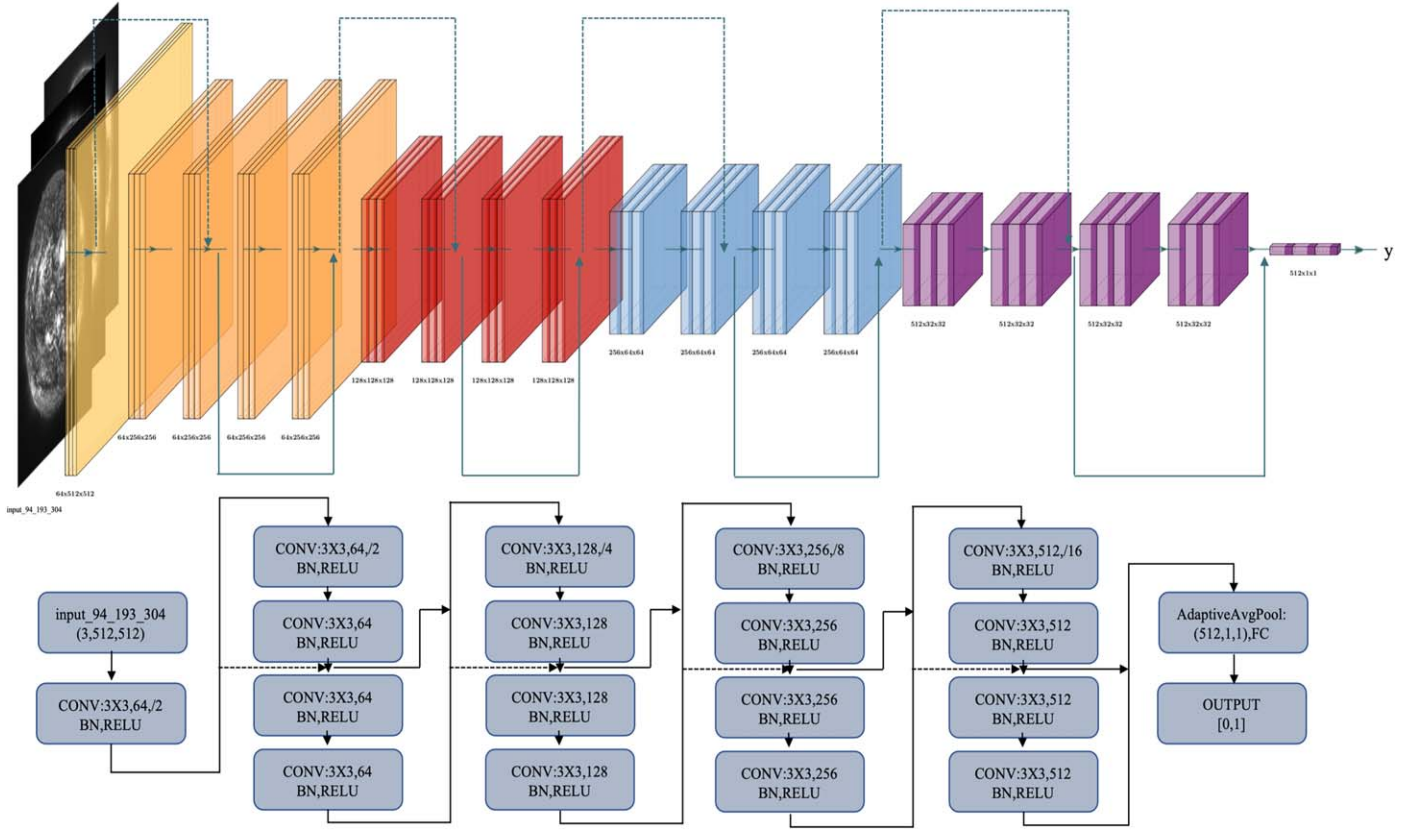


Figure 4. Structural diagram of our forecast model. The CNN network (ResNet-18) was employed to forecast flares with full-disk EUV images of three wavelengths (at 94, 193, and 304 Å) generated by the AIA equipment as the input. The input to each CNN is the composite images of three wavelengths' full-disk EUV images, which have been resized from 4096×4096 to 512×512 pixels. The output of each CNN is the forecast of whether a flare will erupt in the next m hr ($m = 6, 12, 24, 48, \text{ and } 72$). The process in the middle involves layer-by-layer sampling to extract image features and subsequent training. Different colors represent the process of sampling layer by layer, with solid lines indicating sampling at the same image resolution and dotted lines indicating downsampling.

normalization layers, and so on). The basic structure of ResNet-18 includes an input layer, convolutional layers, and residual blocks. The input layer receives tricolor images with a size of 512×512 . The convolutional layers extract local features from the images using 3×3 convolutional kernels and the ReLU activation function. The key component of the model is the residual blocks, where each block consists of two convolutional layers and a skip connection. This design helps address the issues of gradient vanishing and gradient explosion in deep CNNs. The main characteristic of ResNet-18 is the introduction of residual learning, which is crucial for overcoming the challenges of training deep networks. Additionally, ResNet-18 utilizes techniques such as batch normalization and pooling layers to accelerate the training process and improve the model performance. ResNet-18 has achieved excellent performance on the ImageNet data set and has been widely applied to various computer vision tasks, including image classification, object detection, and semantic segmentation, among others. This model provides tools for building, training, and evaluating neural networks, as well as functionalities for data processing. In summary, ResNet-18 is a classic and efficient deep CNN model with powerful feature extraction and classification capabilities.

Figure 4 describes the structural diagram of our forecast model. In this figure, the CNN network (ResNet-18) was employed to forecast flares with full-disk EUV images of three wavelengths (at 94, 193, and 304 Å) generated by the AIA equipment as the input. The input to each CNN is the

composite images of three wavelengths' full-disk EUV images, which have been resized from 4096×4096 to 512×512 pixels. The output of each CNN is the forecast of whether a flare will erupt in the next m hr ($m = 6, 12, 24, 48, \text{ and } 72$). The process in the middle involves layer-by-layer sampling to extract image features and subsequent training. Different colors represent the process of sampling layer by layer, with solid lines indicating the sampling at the same image resolution and dotted lines indicating downsampling. At the start of each iteration, downsampling is implemented to ensure that the training set remains a balanced data set, effectively tackling the problem of insufficient attention being paid to minority class samples. Additionally, it retains the weight boosting of misclassified samples in algorithms, making the model achieve better results in imbalanced data classification. The process of sampling utilizes the shortcut connection technique, which can reduce the loss of image feature information caused by the increase of layers. The concept of the shortcut connection technique is illustrated in Figure 5, where the data from the initial step are added together as the input every few steps.

4. Model Evaluation Metrics

We treat solar flare forecast as a binary classification problem, so the forecast result contains four cases, as shown in Table 2, which is a confusion matrix. EUV images that have both a true positive label and a corresponding positive prediction are classified as True Positives (TP). Those images that carry a TP label but receive a negative prediction are

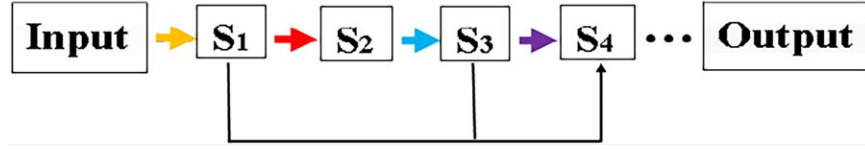


Figure 5. Schematic diagram of the shortcut connection technique. The data from the initial step are added together as the input every few steps. This can reduce the loss of image feature information caused by the increase of layers.

Table 2
Confusion Matrix

	Positive Sample	Negative Sample
Positive sample	TP	FN
Negative sample	FP	TN

termed False Negatives (FN). Images with a true negative label and a matching negative prediction are labeled as True Negatives (TN). Conversely, images that have a TN label but are incorrectly predicted as positive are known as False Positives (FP).

Through the confusion matrix, the model's prediction results on both positive and negative cases can be evaluated separately. Four model performance evaluation metrics, i.e., recall, precision, accuracy (ACC), balanced accuracy (BACC), true skill statistics (TSS), and Heidke skill score (HSS), are calculated using TP, TN, FP, and FN. Recall represents the ratio of TP images among all positive images, describing how many positive images can be successfully predicted from all positive images. Precision represents the ratio of TP images to the total number of images predicted as positive by the model, describing how many positive images predicted by the model are truly positive. Recall and precision evaluate the ability of the model in forecasting positive events, where unity denotes a perfect result and zero is the worst. ACC is the ratio between the number of the correctly forecasted images and the total images. BACC is applicable for imbalanced classification scenarios due to its unbiasedness when dealing with class-imbalance ratios. TSS represents the difference between recall and FP, which examines whether forecasting an image leads to a significant increase of the false alarm or not (N. M. Hansen 1965). HSS measures the accuracy of the correct forecast relative to the random forecast. Both of the TSS and HSS values range from -1 to 1 , where 1 indicates a perfect score and 0 indicates a random or constant forecast. These evaluation metrics are defined as follows:

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

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \quad (1b)$$

$$\text{ACC} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}}, \quad (1c)$$

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

$$\text{TSS} = \frac{\text{TP}}{\text{TP} + \text{FN}} - \frac{\text{FP}}{\text{TN} + \text{FP}}, \quad (1e)$$

Table 3
Model Evaluation for Different Prediction Lead Times

Prediction Lead Time	Recall	Precision	ACC	BACC	TSS	HSS
6 hr	0.786	0.722	0.854	0.834	0.703	0.419
12 hr	0.841	0.755	0.846	0.844	0.704	0.473
24 hr	0.894	0.778	0.844	0.850	0.879	0.579
48 hr	0.908	0.819	0.852	0.851	0.915	0.651
72 hr	0.948	0.786	0.829	0.815	0.898	0.683

$$\text{HSS} = \frac{2(\text{TP} \times \text{TN} - \text{FP} \times \text{FN})}{(\text{TP} + \text{FN}) \times (\text{FN} + \text{TN}) + (\text{TP} + \text{FP}) \times (\text{FP} + \text{TN})}. \quad (1f)$$

Besides the above evaluation metrics, the receiver operating characteristic (ROC) and the area under the ROC curve (AUC) value are also the frequently used evaluation metrics. The ROC curve is a graphical tool used to evaluate the performance of classification models, especially suitable for binary classification problems. The ROC curve demonstrates the classification performance of the model under different decision thresholds by plotting the relationship between the true positive rate (TPR) and the false positive rate (FPR). $\text{TPR} = \text{TP}/(\text{TP} + \text{FN})$ is the same as recall, which represents the proportion of positive cases that are correctly predicted. $\text{FPR} = \text{FP}/(\text{TN} + \text{FP})$ represents the proportion of negative cases that are incorrectly predicted as positive cases. AUC is a standard used to measure the performance of a classification model, whose value ranges from 0 to 1 , with a larger AUC indicating better performance.

5. Results and Discussions

Table 3 shows the prediction performance of the our model, characterized by six evaluation metrics: recall, precision, ACC, BACC, TSS, and HSS. It is seen that for all cases with the lead time being 6, 12, 24, and 48 hr, our model performs good. The values of most evaluation metrics scores in Table 3 are over 0.8 , with some even exceeding 0.9 . Furthermore, Figure 6 demonstrates the numerical statistical results of the ROC. Figure 7 demonstrates the numerical statistical results of the AUC. According to the chart analysis, all the values of the AUC exceed 0.91 , which are rather high scores. These results demonstrate the effectiveness of our prediction model in the applications of both short-term and long-term forecasts.

The performances of our forecast models are further evaluated by comparing with previous works. Table 4 lists the main differences between the models from the following two aspects: (1) The prediction methods include statistics models (K. D. Leka et al. 2023), traditional machine learning models (random forest, RF, and SVM; H. Liu et al. 2019), and

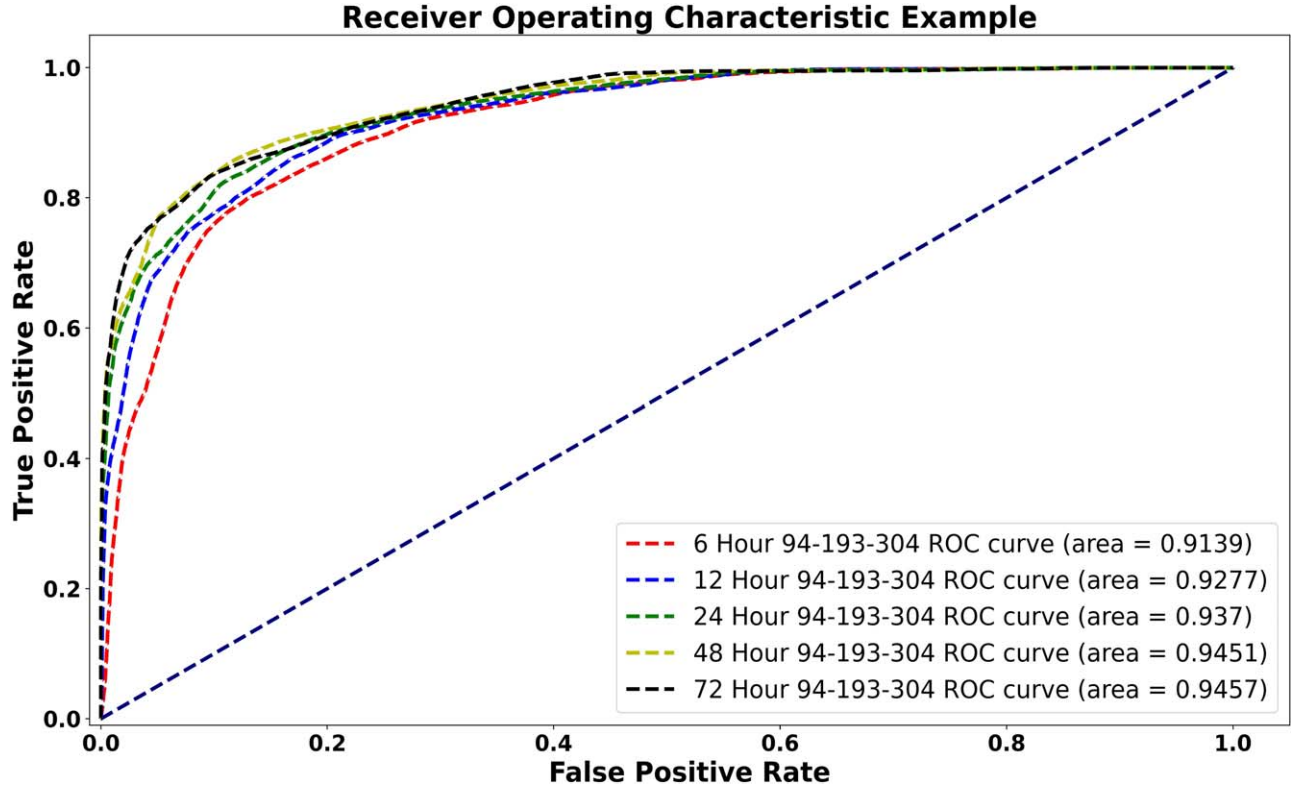


Figure 6. ROC charts of the flare forecast with five lead times ($m = 6, 12, 24, 48$, and 72 hr). The ROC curve demonstrates the classification performance of the model under different decision thresholds by plotting the relationship between the TPR and the FPR.

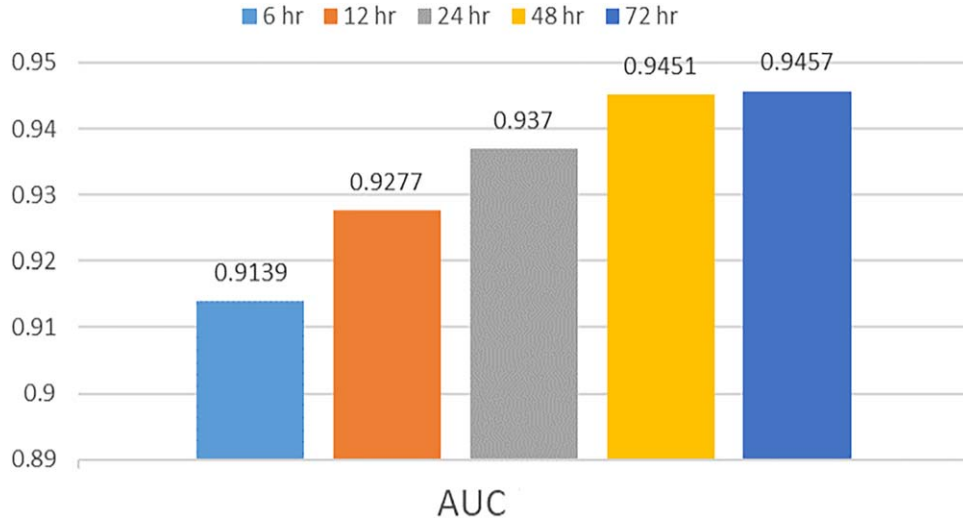


Figure 7. Numerical statistical results of the AUC. The value of the AUC is the area under the ROC curve. The closer its value is to 1, the better the performance of the model.

deep learning models (LSTM; H. Liu et al. 2019; X. Wang et al. 2020) and CNN (X. Huang et al. 2018; D. Sun et al. 2023, this work). (2) Data styles include the Space-weather HMI Active Region Patches (SHARP; M. G. Bobra & S. Couvidat 2015), which is the physical parameters produced from the active region in magnetic image patches (H. Liu et al. 2019; X. Wang et al. 2020). The AIA Active Region Patches (AARPs) database (K. D. Leka et al. 2023), which is the human-constructed physical parameters from the active region in EUV image patches in a manner parallel to the “SHARP parameters,” the magnetic images (X. Huang et al. 2018) and

EUV images (D. Sun et al. 2023) in active region (AR) patches, and the EUV images in full disk (this work). Each model possesses its unique advantages. As shown in Table 4, generally speaking, deep learning methods outperform traditional machine learning and statistical methods in terms of prediction accuracy. Image data can provide more hidden feature information than human-constructed parameters, although these information are currently not well visualized. EUV images complement some information compared to magnetic images, leading to better prediction performance than magnetic images. However, research on predicting flares

Table 4
Performance Evaluation between this Work and Previous Studies

Metric	Reference	Data Type	Data Range	Method	\geq C-class
Recall	X. Huang et al. (2018)	magnetic images	AR Patches	CNN	0.730
	H. Liu et al. (2019)	SHARP	AR Patches	LSTM	0.773
	H. Liu et al. (2019)	SHARP	AR Patches	RF	0.730
	H. Liu et al. (2019)	SHARP	AR Patches	SVM	0.746
	X. Wang et al. (2020)	SHARP	AR Patches	LSTM	0.621
	K. D. Leka et al. (2023)	AARPs	AR Patches	NCI	...
	D. Sun et al. (2023)	EUV images	AR Patches	CNN	...
	This Work	EUV images	Full-disk	CNN	0.894
Precision	X. Huang et al. (2018)	magnetic images	AR Patches	CNN	...
	H. Liu et al. (2019)	SHARP	AR Patches	LSTM	0.541
	H. Liu et al. (2019)	SHARP	AR Patches	RF	0.499
	H. Liu et al. (2019)	SHARP	AR Patches	SVM	0.497
	X. Wang et al. (2020)	SHARP	AR Patches	LSTM	0.635
	K. D. Leka et al. (2023)	AARPs	AR Patches	NCI	...
	D. Sun et al. (2023)	EUV images	AR Patches	CNN	...
	This Work	EUV images	Full-disk	CNN	0.778
ACC	X. Huang et al. (2018)	magnetic images	AR Patches	CNN	...
	H. Liu et al. (2019)	SHARP	AR Patches	LSTM	0.826
	H. Liu et al. (2019)	SHARP	AR Patches	RF	0.804
	H. Liu et al. (2019)	SHARP	AR Patches	SVM	0.803
	X. Wang et al. (2020)	SHARP	AR Patches	LSTM	0.883
	K. D. Leka et al. (2023)	AARPs	AR Patches	NCI	...
	D. Sun et al. (2023)	EUV images	AR Patches	CNN	...
	This Work	EUV images	Full-disk	CNN	0.844
BACC	X. Huang et al. (2018)	magnetic images	AR Patches	CNN	...
	H. Liu et al. (2019)	SHARP	AR Patches	LSTM	0.806
	H. Liu et al. (2019)	SHARP	AR Patches	RF	0.776
	H. Liu et al. (2019)	SHARP	AR Patches	SVM	0.781
	X. Wang et al. (2020)	SHARP	AR Patches	LSTM	0.883
	K. D. Leka et al. (2023)	AARPs	AR Patches	NCI	...
	D. Sun et al. (2023)	EUV images	AR Patches	CNN	...
	This Work	EUV images	Full-disk	CNN	0.850
HSS	X. Huang et al. (2018)	magnetic images	AR Patches	CNN	0.339
	H. Liu et al. (2019)	SHARP	AR Patches	LSTM	0.526
	H. Liu et al. (2019)	SHARP	AR Patches	RF	0.469
	H. Liu et al. (2019)	SHARP	AR Patches	SVM	0.472
	X. Wang et al. (2020)	SHARP	AR Patches	LSTM	0.557
	K. D. Leka et al. (2023)	AARPs	AR Patches	NCI	...
	D. Sun et al. (2023)	EUV images	AR Patches	CNN	...
	This Work	EUV images	Full-disk	CNN	0.579
TSS	X. Huang et al. (2018)	magnetic images	AR Patches	CNN	0.487
	H. Liu et al. (2019)	SHARP	AR Patches	LSTM	0.612
	H. Liu et al. (2019)	SHARP	AR Patches	RF	0.552
	H. Liu et al. (2019)	SHARP	AR Patches	SVM	0.562
	X. Wang et al. (2020)	SHARP	AR Patches	LSTM	0.553
	K. D. Leka et al. (2023)	AARPs	AR Patches	NCI	0.597 ~ 0.68
	D. Sun et al. (2023)	EUV images	AR Patches	CNN	0.84 ~ 0.92
	This Work	EUV images	Full-disk	CNN	0.879
AUC	X. Huang et al. (2018)	magnetic images	AR Patches	CNN	...
	H. Liu et al. (2019)	SHARP	AR Patches	LSTM	...
	H. Liu et al. (2019)	SHARP	AR Patches	RF	...
	H. Liu et al. (2019)	SHARP	AR Patches	SVM	...
	X. Wang et al. (2020)	SHARP	AR Patches	LSTM	...
	K. D. Leka et al. (2023)	AARPs	AR Patches	NCI	0.761 ~ 0.830
	D. Sun et al. (2023)	EUV images	AR Patches	CNN	0.89 ~ 0.94
	This Work	EUV images	Full-disk	CNN	0.937

utilizing EUV images is still relatively scarce. Two typical studies include K. D. Leka et al. (2023), which used 160 parameters derived from EUV images as samples and applied their statistical model, NWRA Classification Infrastructure (NCI; K. D. Leka et al. 2018) for prediction. Their prediction results are slightly lower compared to directly using EUV

images, maybe mainly due to the limited number of manually extracted features or the limited ability of statistical methods to extract information. D. Sun et al. (2023) achieved better prediction results by directly utilizing EUV images, but they use active region patches from full-disk images. Since the lifespan of an active region is limited, isolating them can result

in the loss of historical information, geographical location data, and the influence of surrounding active regions. Therefore, in this work, we use full-disk images as a whole for prediction, resulting in noticeable improvements compared to previous works. D. Sun et al. (2023) tested cases using 1–6 wavelength and found that the more wavelengths used simultaneously, the better the prediction results. We think this is because multi-wavelength images can provide more information. D. Sun et al. (2023) utilized three wavelengths resulting in an AUC of 0.8, whereas in this work, using three wavelengths yields an AUC of 0.937. This implies that full-disk images with a limited number of wavelengths can provide more information than individual active regions. Table 4 only presents results within 24 hr for comparison. As seen from the other results in Table 3, our model demonstrates better prediction performance for 48 and 72 hr.

In this study, we employed full-disk tricolor EUV images, instead of individual active regions, to construct a solar flare forecast model. This represents the most significant distinction in our research. We found that the heliographical location information at the time of flare occurrence is a nonnegligible factor influencing the flare prediction (Y. Yang 2025). First, the likelihood of flare occurrences exhibits latitudinal variability. Similar to the principle of the butterfly diagram that illustrates the distribution of active regions, the number of active regions and the frequency of eruptive activities vary at different latitudes. Therefore, latitude also has a certain potential influence on the occurrence of eruptive activities. Another factor is the size of the active regions. J. Murakoz (2024) pointed out that larger active regions may correspond to a higher number of sunspots, a smaller distance between the positive and negative poles of newly emerging sunspots, and a greater propensity for generating eruptive activities. Second, the eruptive activities of an active region are also contingent upon its ambient environment. Many observations and simulations revealed that there is a prominent correlation between CMEs and ambient magnetic field, especially for global magnetic field (B. J. Lynch & J. K. Edmondson 2013; N. Karna et al. 2021; C. W. Jiang 2024). The flares accompanying CMEs are intimately linked to the reconnection of the current sheet below them. For instance, medium and low latitudes are particularly susceptible to the influence of certain coronal streams, whereas the magnetic fields within medium- and high-latitude active regions are comparatively feeble. Furthermore, potential impacts stemming from generator effects (Z. Y. Li & W. R. Hu 1982; B. Goshu 2024), such as meridional circulations, may also come into play, all of which find expression in the positioning of the active regions. When extracting individual active regions as the input of the forecast model, information about their environment is lost. If an eruptive event occurs at the intersection of two active regions, the information of any single active region is not sufficient for predicting the flare productivity. For active regions with more complex surroundings (M. Falco et al. 2019; H. Hayakawa et al. 2025), more information would be lost if individual active regions are considered separately. This underscores the significance of heliographical location information and indicates that treating each active region in isolation would lead to the loss of essential information, thereby reinforcing the necessity of using global images for flare prediction through image-based methods. In previous flare task processing, each flare active region was isolated from its surrounding

environment. When processed using machine learning methods, each isolated active region was treated uniformly, without distinguishing which active regions were adjacent or correlated. The active regions had no knowledge of the historical events occurring among them. Furthermore, the lifespan of an active region is limited, and for those that have disappeared, their historical information and the impact on surrounding active regions cannot be preserved. However, training data based on the entire solar disk can record the history of flares occurring at various locations. A data-driven machine learning model can capture this historical information, enhancing prediction accuracy.

6. Conclusions





In this study, we utilized the CNN ResNet-18 and tricolor (94, 193, and 304 Å) full-disk EUV images observed by the SDO/AIA from 2012 to 2022 with $5 \times 24 \times 365 \times 11$ images and 10,026 flares \geq C-class in total to build a solar flare forecasting model to predict whether a flare \geq C-class would occur within the next m hr ($m = 6, 12, 24, 48$, and 72). We used several model performance evaluation metrics, such as recall, precision, ACC, BACC, TSS, HSS, ROC, and AUC, to evaluate the performance of our model. The performance scores of our model are good: the values of most performance evaluation metrics scores are over 0.8, with some even exceeding 0.9 and all AUC values exceeding 0.91. These evaluation results indicate that using the full-disk EUV images and CNNs is a fairly effective method to forecast flare eruptions, which is based on the fact that treating each active region in isolation would lead to the loss of some essential information, thereby reinforcing the necessity of using global images for flare prediction through image-based methods. Utilizing full-disk EUV image data and encoding active region history can effectively leverage the ability in predicting solar flares. Currently, methods used for flare forecasting and analysis are mainly divided into two categories: (1) manually constructing feature parameters and (2) automatically extracting feature information from images through machine learning methods. It is generally believed that the parameters obtained through manually constructing are limited, whereas deep features can be automatically extracted from images through deep learning models, which can gradually extract more complex and abstract features through multiple layers of nonlinear transformations. This is also the reason why deep learning methods can achieve higher forecasting accuracy. However, the method of using deep learning to process image information for flare forecasting cannot currently specify the extracted important feature parameters and provide physical explanations for them. To our knowledge, if only the accuracy of forecasting results is considered, machine learning methods have an advantage. Moreover, machine learning methods do not necessarily consume a vast amount of time, as some literature suggests. The time-consuming part lies in the training phase, which is a one-time process. Once the model is trained, it can be used for real-time forecasting. For example, a trained model can predict whether a flare will occur 1 hr later. By inputting an image from a specific moment, it can quickly predict whether a flare will occur 1 hr later, rather than requiring long training data each time before forecasting. Considerable work has been done on feature extraction and analysis for magnetograms, while few studies have been

conducted on flares using EUV images. For instance, K. D. Leka et al. (2023) extracted some features from EUV images using moment analysis and used these features to predict whether a flare would occur. Their evaluation metrics, TSS = 0.68 ~ 0.82 and AUC = 0.8, are a little lower than some machine learning methods, maybe mainly due to the limited manually produced features and the statistical method used to obtain them. Nevertheless, their approach of extracting EUV features by analogy with methods used for magnetograms plays a crucial leading role in this direction and promotes its development. If more advanced image feature analysis methods are employed to quantify the important features extracted during deep learning forecasting, this will advance our understanding of the mechanism of flare generation. As the triggering mechanism of solar flares remains unclear to date, physical methods such as MHD numerical simulations are substantially hindered to predict the occurrence of solar eruptions. In terms of flare analysis and forecasting, image data can provide richer information than manually produced parameter data. Manually produced parameter data are more intuitive in analyzing the mechanism of flare generation and facilitate providing physical explanations. Regarding image data, magnetograms and EUV images are commonly used. Magnetograms have been extensively applied and studied, and the corresponding feature extraction products are relatively mature. However, there has been less research on EUV image data and even fewer directly extracted feature products. This paper marks the first step in our work of this kind, using deep learning for direct prediction. The emphasis of this work is on maximizing the forecasting accuracy rate, rather than delving into the underlying physical principles. Our next step, which is currently underway, is to use deep learning methods for feature extraction from EUV image data.

Acknowledgments

This research was supported by the National Key Technologies Research and Development Program of the Ministry of Science and Technology of China (2020YFC2201200, 2021YFA1600504), the National Natural Science Foundation of China (42204155, 12127901), and the Natural Science Foundation of Jiangsu, China (BK20210168), as well as the Shenzhen Natural Science Fund (the Stable Support Plan Program GXWD2022 0817152453003) and Shenzhen Key Laboratory Launching Project (No. ZDSYS20210702140800001).

ORCID iDs

Yun Yang  <https://orcid.org/0009-0000-5292-713X>
 Yi Wei Ni  <https://orcid.org/0000-0002-9908-291X>
 P. F. Chen  <https://orcid.org/0000-0002-7289-642X>
 Xue Shang Feng  <https://orcid.org/0000-0001-8605-2159>

References

- Ahmed, O., Qahwaji, R., Colak, T., et al. 2013, *SoPh*, **283**, 157
 Arel, I., Rose, D. C., & Karnowski, T. P. 2010, *MCI*, **5**, 13
 Baker, D., Daly, E., Daglis, I., Kappenman, J., & Panasyuk, M. 2004, *SpWea*, **2**, S02004
 Bamba, Y., Inoue, S., Kusano, K., & Shiota, D. 2017, *ApJ*, **838**, 134
 Bamba, Y., Kusano, K., Yamamoto, T. T., & Okamoto, T. J. 2013, *ApJ*, **778**, 48
 Barnes, G., & Leka, K. D. 2006, *ApJ*, **646**, 1303
 Barnes, G., Leka, K. D., Schumer, E. A., & Della-Rose, D. J. 2007, *SpWea*, **5**, S09002
 Bloomfield, D. S., Higgins, P. A., McAteer, R. T. J., & Gallagher, P. T. 2012, *ApJL*, **747**, L41
 Bobra, M. G., & Couvidat, S. 2015, *ApJ*, **798**, 135
 Chen, P. F. 2011, *LRSP*, **8**, 1
 Colak, T., & Qahwaji, R. 2009, *SpWea*, **7**, S06001
 Falco, M., Costa, P., & Romano, P. 2019, *JSWSC*, **9**, A22
 Feng, X. S. 2020, *Magnetohydrodynamic Modeling of the Solar Corona and Heliosphere* (Singapore: Springer), 1
 Gallagher, H., Jack, A., Roepstorff, A., & Frith, C. 2002, *Nimg*, **16**, 814
 Goshu, B. 2024, *BrJSc*, **3**, 85
 Guerra, J. A., Pulkkinen, A., & Uritsky, V. M. 2015, *SpWea*, **13**, 626
 Hansen, N. M. 1965, *Econ. Inq.*, **4**, 3
 Hayakawa, H., Ebihara, Y., Mishev, A., et al. 2025, *ApJ*, **979**, 49
 Hinton, G. E., & Salakhutdinov, R. R. 2006, *Sci*, **313**, 504
 Huang, X., Wang, H., Xu, L., et al. 2018, *ApJ*, **856**, 7
 Jiang, C., & Feng, X. 2013, *ApJ*, **769**, 144
 Jiang, C., Zou, P., Feng, X., et al. 2018, *ApJ*, **869**, 13
 Jiang, C. W. 2024, *ScChD*, **67**, 3765
 Jiang, C. W., Feng, X. S., Wu, S. T., & Hu, Q. 2012, *ApJ*, **759**, 85
 Jiang, C. W., Wu, S. T., & Q. F. X. H. 2016, *NatCo*, **7**, 11522
 Karna, N., Savcheva, A., Gibson, S., et al. 2021, *ApJ*, **913**, 47
 Krista, L. D. 2025, *ApJ*, **978**, 121
 Krista, L. D., & Chih, M. 2021, *ApJ*, **922**, 218
 Kusano, K., Bamba, Y., Yamamoto, T. T., et al. 2012, *ApJ*, **760**, 31
 Kusano, K., Iju, T., Bamba, Y., & Inoue, S. 2020, *Sci*, **369**, 587
 Kusano, K., Maeshiro, T., Yokoyama, T., & Sakurai, T. 2004, *ApJ*, **610**, 537
 LeCun, Y., Bengio, Y., & Hinton, G. 2015, *Natur*, **521**, 436
 Leka, K. D., & Barnes, G. 2003a, *ApJ*, **595**, 1277
 Leka, K. D., & Barnes, G. 2003b, *ApJ*, **595**, 1296
 Leka, K. D., & Barnes, G. 2007, *ApJ*, **656**, 1173
 Leka, K. D., Barnes, G., & Wagner, E. 2018, *JSWSC*, **8**, A25
 Leka, K. D., Dissauer, K., Barnes, G., & Wagner, E. L. 2023, *ApJ*, **942**, 84
 Lemen, J. R., Title, A. M., Akin, D. J., et al. 2012, *SoPh*, **275**, 17
 Li, R., Wang, H.-N., He, H., Cui, Y.-M., & Zhan-LeDu 2007, *ChJAA*, **7**, 441
 Li, Z. Y., & Hu, W. R. 1982, *ChJSS*, **2**, 137
 Liu, H., Liu, C., Wang, J. T. L., & Wang, H. 2019, *ApJ*, **877**, 121
 Lynch, B. J., & Edmondson, J. K. 2013, *ApJ*, **764**, 87
 Mason, J. P., & Hoeksema, J. T. 2010, *ApJ*, **723**, 634
 Murakozy, J. 2024, *A&A*, **690**, A257
 Nishizuka, N., Sugiura, K., Kubo, Y., et al. 2017, *ApJ*, **835**, 156
 Qahwaji, R., & Colak, T. 2007, *SoPh*, **241**, 195
 Schmidhuber, J. 2015, *NN*, **61**, 85
 Schrijver, C. J., & Title, A. M. 2011, *JGRA*, **116**, A04108
 Song, H., Tan, C., Jing, X., et al. 2009, *SoPh*, **254**, 101
 Sun, D., Huang, X., Zhao, Z., & Xu, L. 2023, *ApJS*, **266**, 8
 Wang, H., Cui, Y., Li, R., Zhang, L., & Han, H. 2008, *AdSpR*, **42**, 1464
 Wang, X., Chen, Y., Toth, G., et al. 2020, *ApJ*, **895**, 3
 Webb, D. F., & Howard, T. A. 2012, *LRSP*, **9**, 3
 Yang, Y. 2025, *FrASS*, **11**, 1509061
 Yang, Y., Liu, J. J., Feng, X. S., Chen, P. F., & Zhang, B. 2023, *ApJS*, **268**, 69
 Yu, D., Huang, X., Wang, H., et al. 2010, *ApJ*, **710**, 869