

Article

Deep-Learning-Based Solar Flare Prediction Model: The Influence of the Magnetic Field Height

Lei Hu, Zhongqin Chen, Long Xu and Xin Huang *

Faculty of Electrical Engineering and Computer Science, Ningbo University, Ningbo 315211, China
* Correspondence: huangxin@nbu.edu.cn

Abstract: Solar flares, caused by magnetic field reconnection in the sun's atmosphere, are intense bursts of electromagnetic radiation that can disrupt the Earth's space environment, affecting communication systems, GPSs, and satellites. Traditional physics-based methods for solar flare forecasting have utilized the statistical relationships between solar activity indicators, such as sunspots and magnetic field properties, employing techniques like Poisson distributions and discriminant analysis to estimate probabilities and identify critical parameters. While these methods provide valuable insights, limitations in predictive accuracy have driven the integration of deep learning approaches. With the accumulation of solar observation data and the development of data-driven algorithms, deep learning methods have been widely used to build solar flare prediction models. Most research has focused on designing or selecting the right deep network for the task. However, the influence of the magnetic field height on deep-learning-based prediction models has not been studied. This paper investigates how different magnetic field heights affect solar flare prediction performance. Active regions were observed using HMI magnetograms from 2010 to 2019. The magnetic field heights were stratified to create a database of active regions, and deep neural networks like AlexNet, ResNet-18, and SqueezeNet were used to evaluate prediction performance. The results show that predictions at around 7200 km above the photosphere outperform other heights, aligning with physical method analysis. At this altitude, the average AUC of the predictions from the three models reaches 0.788.



Academic Editor: Dmitrii Kolotkov

Received: 25 December 2024

Revised: 3 April 2025

Accepted: 8 April 2025

Published: 24 April 2025

Citation: Hu, L.; Chen, Z.; Xu, L.; Huang, X. Deep-Learning-Based Solar Flare Prediction Model: The Influence of the Magnetic Field Height. *Universe* **2025**, *11*, 135. <https://doi.org/10.3390/universe11050135>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: solar flare prediction; magnetic field height; deep learning; convolutional neural networks

1. Introduction

Solar flares are sudden, intense bursts of electromagnetic radiation and high-energy particles in the Sun's atmosphere, primarily caused by magnetic reconnection. These events can severely disrupt the Earth's space environment and technological systems, affecting radio communications, GPS accuracy, and satellite operations. Accurate solar flare prediction is therefore essential for mitigating these impacts and safeguarding critical infrastructure.

Solar flare forecasting has long been a focus of research using traditional physics-based methods. Since the 1930s, statistical relationships between solar flares and sunspot activity have been extensively studied, forming the foundation for early forecasting models. Techniques such as Poisson distributions have been developed to estimate flare probabilities [1], while discriminant analysis has been employed to identify the critical magnetic parameters influencing solar activity [2]. Active regions (ARs), characterized by strong magnetic fields, have been widely analyzed through multi-wavelength observations, with researchers extracting morphological, magnetic, and coronal features—including magnetic gradients,

neutral line lengths, magnetic energy dissipation, and effective magnetic fields—to systematically parameterize flare-productive regions [3–5]. Sunspot classification schemes have further refined the relationship between sunspot morphology and flare occurrence. Advanced techniques, such as Zernike moments derived from magnetograms [6] and fractal analysis of active regions [7], have provided deeper insights into flare conditions. Helioseismology studies have additionally revealed the connections between subsurface flow patterns and flare productivity [8], complemented by recent power spectral analysis demonstrating the predictive capability of magnetic power-law indices in young, active regions [9]. However, despite these advancements, the comparable performance of traditional approaches indicates a persistent bottleneck in predictive accuracy [10]. Collectively, these studies elucidate the complex interplay between magnetic field properties, energy storage mechanisms, and flare-triggering processes, highlighting the continued value of physics-based approaches in solar flare forecasting.

Deep learning has significantly advanced solar flare prediction by integrating diverse data-driven approaches. Self-organized criticality and low-dimensional chaotic dynamics in solar activity have been explored [11], while models leveraging McIntosh classifications, magnetic gradients [12,13], and blackout parameters [14] have improved predictions. Techniques like CNN-GRU [15], density clustering with SMOTE [16], selective upsampling [17], and loss function weighting [18] address class imbalance, with ensemble learning and hybrid methods also proving effective [19–21]. LSTM-based models, including BiLSTM-Attention, excel in multiclass flare forecasting, while the cGAN method enhances magnetic polarity accuracy using $H\alpha$ images [22]. Machine learning applications in solar eruption prediction have been reviewed, summarizing progress and future directions [23]. Studies on magnetogram resolution [24] and EUV imaging [25] have highlighted the insensitivity of deep learning models to resolution before specific thresholds. Research integrating magnetic field height has identified optimal ranges (1000–1800 km) for improving flare onset time predictions [26,27], with synthetic δ -sunspot data and the WG_M method offering insights into 3D magnetic configurations [28,29]. These advances underline the promise of deep learning in refining solar flare forecasting and modeling.

Current research on the influence of magnetic field height variations in deep-learning-based solar flare prediction remains limited. This paper presents the first comprehensive investigation into how different magnetic field altitudes affect forecasting performance. Leveraging SDO/HMI observational data, we systematically incorporated magnetic field measurements at multiple atmospheric heights into deep neural networks, rigorously evaluating the predictive value of this three-dimensional magnetic information for solar flare occurrence.

Building upon this foundation, we implemented three widely adopted CNN architectures—AlexNet [30], ResNet-18 [31], and SqueezeNet [32]—to systematically evaluate how magnetic field height variations influence model performance. Using the AUC metric for quantitative assessment, we conducted comprehensive testing across multiple atmospheric heights. This paper not only reveals previously unexplored relationships between magnetic field altitude and flare prediction accuracy but also provides actionable insights for optimizing deep learning applications in solar activity forecasting.

The remainder of this paper is organized as follows: Section 2 details the data sources and preprocessing methodologies. Section 3 presents the architecture and implementation of the three CNN models employed in this study. Section 4 provides a comprehensive analysis and discussion of the experimental results. Finally, Section 5 concludes this paper with key findings and outlines promising directions for future research.

2. Data

Solar Dynamics Observatory/Helioseismic and Magnetic Imager (SDO/HMI) provides continuous and high-quality photosphere magnetic field observation data. We obtained SHARPs on the photospheric magnetic fields of the solar active region every 96 min from 2010 to 2019 from the Joint Science Operations Center (JSOC) (<http://jsoc.stanford.edu/ajax/lookdata.html>, accessed on 1 July 2022) and selected Stonyhurst longitude maximum and minimum values with an arithmetic average $\leq 30^\circ$ as the original data.

The occurrence of solar flares is caused by the sudden release of magnetic energy stored in the solar active region in the corona, so the magnetic information on the solar active region is very important for the prediction of solar flares. However, due to the lack of conventional coronal magnetic field observation data, we annotated the downloaded the SHARP data products and input the active photosphere magnetic field as the basic boundary condition into the Nonlinear Force-Free Coronal Magnetic Field (NLFFF) Extrapolations to construct a 3D coronal vector magnetic field dataset [33].

Because the span of magnetic field intensity is very large, we needed to normalize the three-dimensional magnetic field data obtained from extrapolation. Most of the magnetic field intensity in the solar active region is concentrated in a certain range, but some data are far beyond (or far below) this range, the normalization according to the overall data reduce the values in the dataset range to 0, and there is only a small amount of data near 1. Therefore, before normalization, we set the left and right thresholds for the magnetic field data concentration region of the photosphere as ± 500 Gauss, and the data with an absolute value greater than 500 Gauss were uniformly classified as the threshold. We delimited the threshold for each layer of data obtained by extrapolating and kept the proportion of data within the threshold to the total data the same as that in the photosphere layer. In the photosphere, the data within ± 500 Gauss accounted for about 98% of the total data. This is shown in Figure 1.

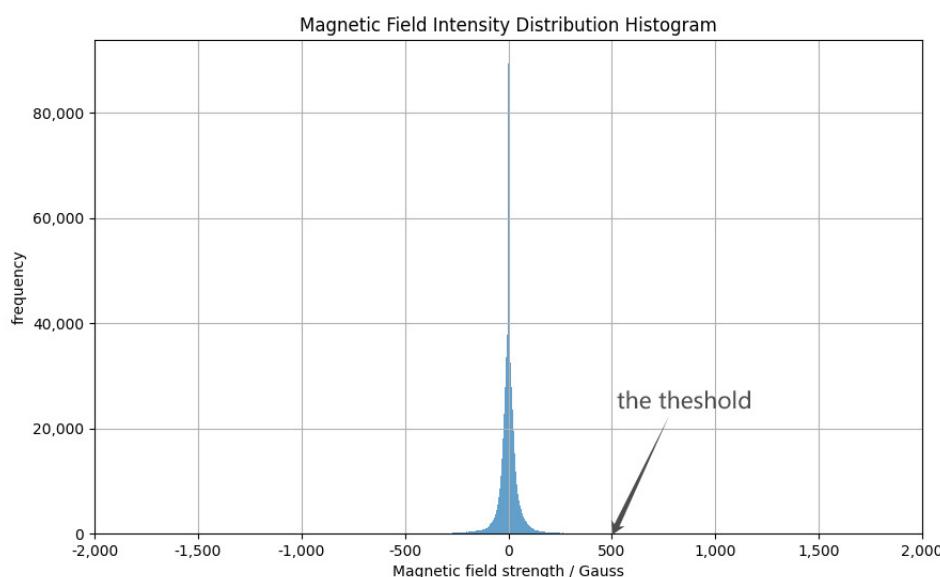


Figure 1. Magnetic field intensity distribution of all sample pixels in the photosphere. The horizontal axis represents the strength of the magnetic field in Gauss, and the vertical axis represents the corresponding frequency. Data with magnetic field strength greater than the threshold value were counted as 500 gauss.

At the end of data processing, we labeled the data samples. We classified the observed data and their extrapolated results into flare and nonflare samples. If at least one M-class flare occurred in the active region within 24 h from the beginning of the observation, the magnetic field data sample of the active region was considered to be a flare sample; conversely, it was considered to be a nonflare sample.

We obtained a dataset of 52,352 samples, to each of which we added an additional third dimension to the traditional two-dimensional magnetic map: magnetic field height. The dataset was composed of 1140 AR, and we used the samples from January to October of each year from 2010 to 2019 as the training set and the remaining samples as the test set. The ratio of the training set to the test set was approximately 4.68:1. The training set consisted of 43,140 samples, of which 882 were flares, and 42,258 were nonflares. The test set consisted of 9212 samples, of which 206 were flares, and 9006 were nonflares. Obviously, we found a positive and negative sample class imbalance, so we needed to resample the sample.

3. Methods

Deep learning methods demonstrate strong interpretability in solar flare prediction applications. Local interpretation techniques, including gradient-based methods (e.g., gradient input) and feature importance approaches (e.g., LIME), enable the analysis of individual prediction decisions by examining input feature contributions [34]. Global interpretation methods characterize overall model behavior, with techniques like Feature Importance Ranking identifying key discriminative features across the entire input space. Visualization of these interpretability results reveals the specific magnetic field characteristics and patterns that models prioritize during flare identification and prediction. These interpretability analyses not only validate model reliability but also provide physical insights into flare-triggering mechanisms and evolutionary processes, ultimately enhancing forecasting accuracy.

Deep learning can automatically extract useful features from raw observational data. In this study, three widely used CNN models were selected, AlexNet, ResNet-18, and SqueezeNet, to examine the flare prediction performance at different altitudes.

Figure 2 illustrates the architecture of our flare prediction network. To evaluate how magnetic field height affects prediction performance, we conducted separate training sessions using datasets from different atmospheric levels. The models generated binary predictions indicating whether each active region will produce flares within 24 h. We then computed standard evaluation metrics to enable the systematic comparison of prediction accuracy across both different network architectures and varying magnetic field heights.

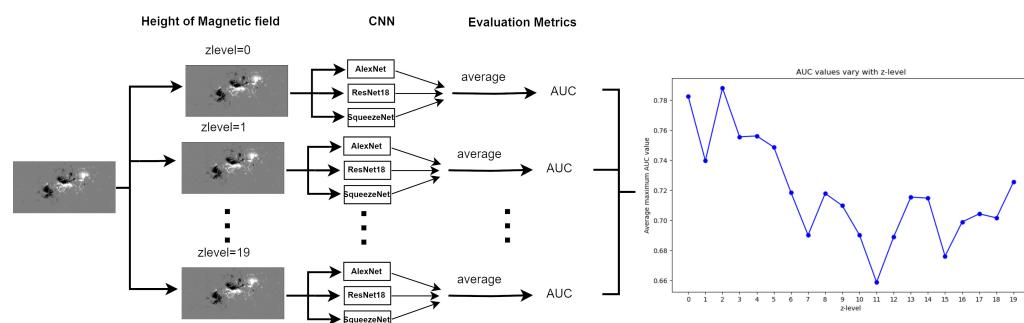


Figure 2. Flow chart of forecasting network. The magnetograms of different heights were input into the three networks for training, and the trend graph of the average AUC of the three networks with height was obtained.

3.1. AlexNet

Figure 3 illustrates the AlexNet architecture, which comprises five convolutional (conv) layers, three fully connected (FC) layers, three max-pooling (MaxPool) layers, and one adaptive average pooling (AdaptiveAvePool) layer. Each convolutional layer is immediately followed by a Rectified Linear Unit (ReLU) activation layer, with additional ReLU layers inserted between the FC layers Appendix A). As the first CNN architecture to implement ReLU activation systematically, AlexNet significantly improved training efficiency by introducing nonlinear transformations. These nonlinearities enable the network to learn complex representations while mitigating gradient vanishing problems during backpropagation. The specific configuration of each network layer is detailed below.

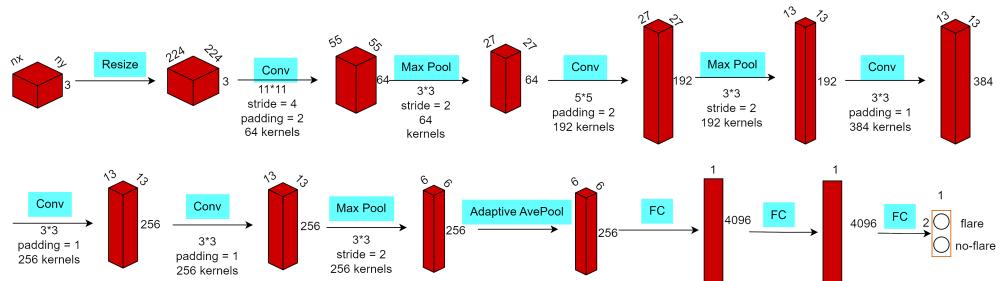


Figure 3. AlexNet structure. AlexNet contains eight layers that can learn weights. The red cube represents the shape of the sample data in different layers, and the blue rectangle represents the operation and layer. Conv represents the convolutional layer, MaxPool represents the maximum pooling layer, and FC represents the fully connected layer. The network output is a two-dimensional one-hot vector representing the probability of a flare outbreak.

3.2. ResNet-18

Model accuracy typically improves with increasing network depth up to a critical threshold. Beyond this point, however, unexpected accuracy degradation occurs due to the vanishing/exploding gradient problem, where backpropagated gradients either diminish or grow exponentially with network depth. The residual module [31] addresses this by introducing skip connections that bypass multiple layers, maintaining gradient flow even through non-transformative layers. This architectural innovation enables successful training of ultra-deep networks (e.g., ResNet variants with 150+ layers) while significantly boosting classification performance. For our study, we implemented ResNet-18, an 18-layer variant offering optimal balance between depth and computational efficiency for our flare prediction task.

When cross-layer composition is carried out, the problem of dimension mismatch is likely to occur. In order to ensure that the input and output can be added correctly, the residuals need to be raised or reduced dimensionally through a convolution layer with a convolution kernel of 1×1 when combining across channels. After the weights are trained, ResNet-18 uses a $512 \times 1 \times 1$ averaging pooling layer to map the residual fast output to a feature vector with a 512 length and finally a linear layer to obtain the output. ResNet-18 has achieved good classification results in flare prediction tasks. Its architecture is shown in Figure 4:

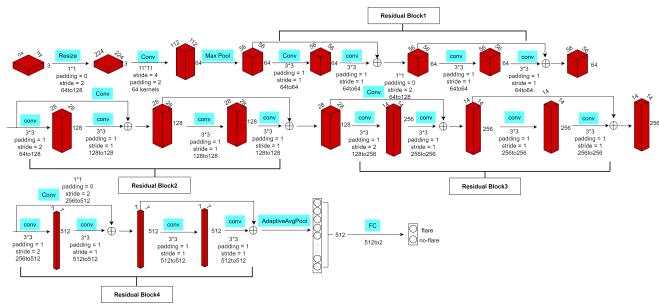


Figure 4. ResNet-18 structure. ResNet-18 consists of 18 weight layers. It introduces a new residual module backpropagation gradient to alleviate the problem of gradient vanishing in deep networks.

3.3. SqueezeNet

The SqueezeNet model architecture is shown in Figure 5. In order to achieve efficiency and accuracy, SqueezeNet adopts the following three strategies (Appendix B).

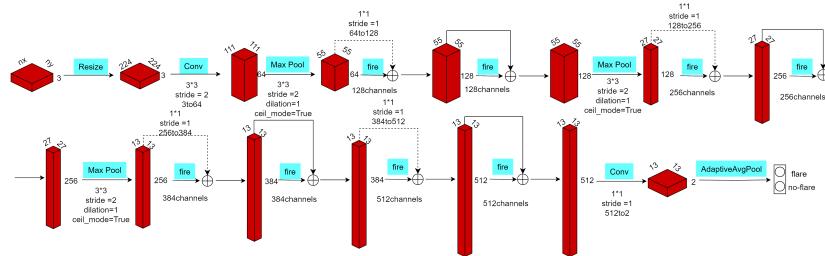


Figure 5. SqueezeZeNet structure. SqueezeNet consists of 10 layers. It introduces the Fire module, which drastically reduces the number of parameters in the model while maintaining accuracy.

3.4. Application of the Model

The data preprocessing pipeline begins by loading all training and test samples into the dataset loader. Each magnetic field map is resized to a standardized $3 \times 512 \times 512$ tensor format, where the three channels (n_x, n_y, n_z) correspond to the vector components of the magnetic field intensity. This spatial normalization ensures dimensional consistency across all inputs while preserving the essential vector field information. The 512×512 resolution was selected to maintain sufficient spatial detail for accurate flare prediction while remaining computationally tractable for the subsequent CNN processing.

Due to the class imbalance problem in the dataset, there were few positive samples and many negative samples. In order to effectively avoid this problem, we defined a self-defined sampler. Assuming that the batch size is x , we randomly sampled 2 negative samples in a cycle; for positive samples, we sampled $x/2$ positive samples sequentially and downsampled most negative samples. Oversampling a small number of positive samples can ensure that positive and negative samples are balanced in each batch, and random sampling also improves the generalization ability of the model.

Regarding the selection of hyperparameters, we mainly considered batch size, learning rate, number of epochs and optimizer. In order to explore the influence of different magnetic field heights on the prediction performance, the same and appropriate hyperparameters were fixed for three different models in this study. We chose a batch size of 64, a learning rate of 1e-3, epochs of 50, and the Adam optimizer [35].

During model training, the input magnetograms were processed to generate flare predictions, which were compared against ground-truth labels. For this binary classification task, we employed the cross-entropy loss function in Equation (1) to quantify the discrepancy between the model's predicted probabilities and the actual flare occurrences. The loss function evaluates how closely the network's output distribution aligns with the true label distribution, providing the gradient signal for backpropagation-based optimization.

$$L(y, p) = - \sum_i y_i \log(p_i), \quad (1)$$

where g is the real label, which is usually a one-hot encoded vector, i.e., the value is 1 on the index position of the real category and 0 on the other position. p is the probability distribution predicted by the model, which is usually a probability vector output by the softmax function. \log is the natural logarithm.

The gradients of the loss function with respect to the network weights are backpropagated through all layers, enabling weight updates in both convolutional filters and fully connected connections via stochastic gradient descent (SGD). This forward–backward propagation cycle iterates continuously until meeting termination criteria. To mitigate overfitting, we employed early stopping—halting training when test set performance plateaued—which simultaneously minimized the loss function and preserved generalization capability. As formalized in Equation (2), SGD updates the network parameters θ using gradients computed on mini-batches of training data:

$$\theta_{new} = \theta_{old} - \alpha \nabla_{\theta} L(\theta_{old} | x_i, y_i), \quad (2)$$

where θ_{old} is the current parameter value, and α is the learning rate, which controls the step size of the parameter update. $\alpha \nabla_{\theta} L(\theta_{old} | x_i, y_i)$ is the loss function with respect to the gradient of the parameter on sample (x, y) . θ_{new} is the updated parameter value. In this study, due to the very large dataset size, we used the Adam optimizer that could automatically adjust the learning rate and converge faster.

During the final evaluation phase, we quantitatively assessed the model’s predictive performance using held-out validation data. The trained convolutional neural network processed input magnetograms from test active regions through forward propagation to generate flare predictions. Model accuracy was systematically evaluated by comparing these predictions with ground-truth flare occurrences, employing standard classification metrics to measure the agreement between predicted and observed events.

4. Experiments and Results

In order to facilitate the subsequent deep learning work, the data observed from January to October each year from 2010 to 2019, along with their extrapolation results, were pre-allocated to the training set, and the remaining samples were used as the test set. At this point, the establishment of the sample dataset was complete. Each sample in the resulting dataset included the SHARP number of the active region, a four-dimensional array of magnetic field data (where the first dimension represented the three components of the magnetic field data, and the following three dimensions represented the space size of the magnetic field), the label of the sample, and the dataset to which the sample belonged (training or test set).

4.1. Evaluation Indices

The output of the model is a two-dimensional vector representing the model’s prediction of whether a flare will occur within a 24 h time window. Table 1 shows the indicators under four prediction conditions: true positive (TPs), false positive (FPs), false negative (FNs), and true negatives (TNs), where TPs represent the number of positive instances classified as positive, FPs represents the number of negative instances classified as positive, FNs represents the number of positive instances classified as negative, and TNs represents the number of negative instances classified as negative. $P = TP + FN$ is the total number of positive samples; $N = FP + TN$ is the total number of negative samples.

Table 1. Confusion matrix for binary solar flare forecasting.

		Forecasted Classes		Total Instances	
Actual Classes		Positive	TP	TN	P
		Negative	FP	FN	N

4.1.1. TPR, FNR, TNR, FPR

True positive rate (TPR), also known as sensitivity or recall, refers to the proportion of all actual positive cases that are correctly identified as positive cases. The false negative rate (FNR) refers to the proportion of all actual positive cases that are incorrectly identified as negative cases. They can be calculated using Equations (3) and (4).

$$TPR = (TP)/(TP + FN), \quad (3)$$

$$FNR = 1 - TPR = (FN)/(TP + FN), \quad (4)$$

TNR (True Negative Rate): true negative rate, also known as Specificity. It refers to the proportion of all actual negative cases that are correctly identified as negative cases. FPR(False Positive Rate): A false positive rate is the proportion of all actual negative cases that are incorrectly identified as positive. Their calculation methods are shown in Equations (5) and (6), respectively.

$$TNR = (TN)/(TN + FP), \quad (5)$$

$$FPR = 1 - TNR = (FP)/(TN + FP), \quad (6)$$

4.1.2. Accuracy

Accuracy is the number of correctly classified samples divided by the total number of samples, again ranging from 0 to 1. Generally speaking, the higher the classification accuracy rate, the closer the result is to one, and the better the classification effect, but the accuracy rate is not applicable to unbalanced classification. If a classifier predicts that all instances of a minority class belong to the majority class, and the majority class also has a high classification accuracy, accuracy remains high. ACC can be obtained through Equation (7).

$$Accuracy = (TP + TN)/(TP + FP + TN + FN), \quad (7)$$

4.1.3. True Skill Score

The true skill score (TSS) ranges from $[-1,1]$, and the closer the result to one, the better the classification. A result of -1 means that all forecasts are wrong, and a result of 1 means that all forecasts are correct, that is, the results of all positive and negative classes of forecasts are the same as the actual situation. The number of positive and negative samples in our solar flare dataset is unbalanced, and the true skill statistics are sensitive to the ratio of class imbalance. Therefore, we used the true skill statistics as the main indicator. The others were used as secondary indicators, which could well reflect the performance of the entire model. The TSS can be obtained through Equation (8).

$$TSS = TPR - FPR, \quad (8)$$

In practice, there are far fewer flaring samples than nonflaring samples. Consider the imbalance between positive and negative samples in the database. If a model performs

flaring judgment for all input active regions, it can also obtain better TPR performance, but the FPR index will be large, and the TSS will be worse.

4.1.4. Receiver Operating Characteristic and Area Under the Curve

The Receiver Operating Characteristic (ROC) curve is a graph that shows how a predictive model behaves at all prediction thresholds, showing the relationship between the model's true rate (TPR) and false positive rate (FPR). The curve has two parameters: TPR and FPR. We can obtain the ROC curve by increasing the threshold from 0 to 1. If the threshold is set to 0, all samples are predicted to be positive. At this point, the TPR and FPR are one. If the threshold is equal to one, all samples are predicted to be negative; At this point, both TPR and FPR are zero. The Area Under the Curve (AUC) score is the area under the ROC curve, which is a general indicator for evaluating the two types of classification models. The higher the AUC value, the better the performance of the model. This score is between 0 and 1.

4.2. Experimental Procedure

During preprocessing, we standardized all input data by resizing the magnetograms to 512×512 resolution using bilinear interpolation. For model training, we exclusively utilized the vertical magnetic field component (nz direction) as the input. To address class imbalance, we implemented balanced batch sampling with equal numbers of positive and negative samples, complemented by three data augmentation techniques: (1) random horizontal flipping, (2) random vertical flipping, and (3) arbitrary rotation (0–360°). These spatial transformations served dual purposes: they artificially expanded our training dataset while improving model robustness to observational variations in solar imagery. Importantly, all augmentation was applied exclusively during training, preserving the integrity of our validation and test sets. The augmentation strategy effectively mitigated the limitations of small datasets in solar physics applications by simulating diverse viewing conditions and increasing the effective training sample size.

We used the network model provided by the PyTorch package 1.13.1 and modified it. The number of output channels of the model was selected as 2. Secondly, we chose the cross-entropy loss function, which is a widely used optimization method in training networks. The stochastic gradient descent algorithm was used to train the model. We trained $\text{batch_size} = 64$, $\text{lr} = 1 \times 10^{-3}$, $\text{num_epochs}=50$ in each network, where batch_size represents batch size, lr represents learning rate, and num_epochs represents number of epochs.

4.3. Experimental Results

After training different models with different magnetic field heights, the corresponding confusion matrix was obtained, and the evaluation indices such as upper loss, AUC, and TSS were calculated. The AUC of the different models was summed and averaged to obtain the comprehensive AUC index. Next, training loss and comprehensive AUC plots were drawn to analyze the influence of the magnetic field height.

4.3.1. Training Loss

Taking ResNet-18 as an example, we obtained the changing trend in the training loss with $zlevel$ under the premise of using the ResNet-18 model. The Figure 6 shows the training loss curve. We found that, on the whole, the training loss showed a decreasing trend with the increase in epochs, which proved that our training was effective, and the training loss showed an increasing trend with the increase in $zlevel$. In relatively high layers, it was difficult to continue to reduce the training loss even after 50 rounds of training epochs, which indicated that with the increase in the number of layers, the training loss would

continue to decrease. The data quality also gradually deteriorates and the model could no longer learn new features well. This conclusion is obviously very reasonable, because the geomagnetic field in the photosphere can be directly observed, while the magnetic field above the photosphere is obtained through extrapolation. As the magnetic field height increases, the data quality gradually deteriorates, and the corresponding indicators also gradually deteriorates.

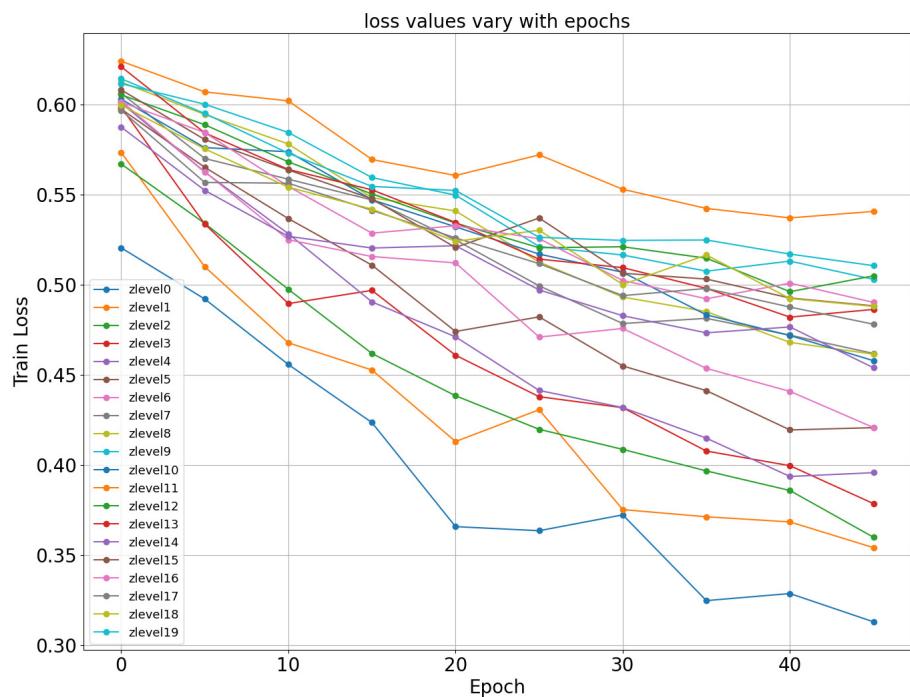


Figure 6. The 20 curves of loss as a function of epoch number. The horizontal axis represents the number of epochs trained, and the vertical axis represents the loss. The 20 curves represent each of the 20 zlevels.

4.3.2. Results

Table 2 shows the contingency table and AUC for different heights in different models. Figure 7 shows the ROC curves of the different models under different magnetic field heights. We conducted pairwise hypothesis tests for different groups of curves with the three models. The results are shown in Table 3. There were significant differences among different groups. It can be seen that with the gradual increase in magnetic field heights, the area under the ROC curve presents a general trend of first increasing and then decreasing and reaches the maximum sum of the areas of the three models when the zlevel is two. The combined AUC curve, AUC bar chart, and standard deviation are shown in Figure 8. As can be seen from the two figures, with the increase in the magnetic field height, the AUC also shows a trend of first increasing and then decreasing; in the first five layers, the two indicators remain at a high level; and then the indicators at the higher levels have an obvious downward trend. Among them, the AUC bar chart on the right is divided into four groups; the first, second, third and fourth groups represent zlevel0–zlevel4, zlevel5–zlevel9, zlevel10–zlevel14, and zlevel15–zlevel19 respectively. It is not difficult to find that the AUC value of the first group, zlevel0–zlevel4, is significantly higher than that of the other groups. This also confirms Korsos's conclusion that the WG_M method has better flare prediction in regions with higher magnetic field heights [27].

From the above, we conclude that the optimal model can be obtained when the zlevel is two; that is, zlevel2 is determined to be the best height. When dividing different layers in the z direction, we divided one layer at an interval of 10 pixels. Therefore, zlevel2 is about 20 pixels in height on the magnetic field of the photosphere. We know that a pixel is about 0.36 Mm, that is, 360 km, so we can find that at a magnetic field height of about 7200 km, there is an optimal magnetic field height, and the flare prediction performance is better than other layers; that is to say, the real eruption location of the flare is likely to be near this optimal height.

From this, we find that our results are not completely consistent with Korsos's conclusion. They believe that the accuracy of solar flare prediction can be improved in a magnetic field height range of 1000–1800 km, while we think that the magnetic field height is about 7200 km, but there is one thing in common that it is above the photosphere. The performance of solar flare prediction in the higher magnetic field is better than that in the photosphere. Magnetic reconnection is an important prerequisite for the eruption of solar flares, and only in the chromosphere and corona are the intensity and complexity of the magnetic field enough to support the occurrence of magnetic reconnection process. This process releases huge energy, and it is possible for solar flares to erupt, which also shows that our conclusion is reasonable.

Table 2. Contingency table and AUC corresponding to different heights in different models.

Net	Group	zlevel	TPs	FNs	TNs	FPs	AUC	
AlexNet	top	0	7813	1193	98	108	0.83	
		1	7310	1696	69	137	0.82	
		3	7410	1596	93	113	0.80	
		6	6951	2055	128	78	0.74	
		16	3951	4188	32	121	0.67	
	middle	19	4325	3346	32	121	0.72	
		12	5559	3430	67	108	0.65	
		bottom	15	6865	1596	90	63	0.62
		18	4544	3276	47	106	0.66	
		3	8430	576	126	80	0.82	
ResNet-18	top	5	7087	1919	92	114	0.79	
		9	7816	1190	104	86	0.79	
		1	8163	843	123	83	0.75	
		8	6630	2376	81	125	0.75	
		10	7748	1246	132	58	0.75	
	middle	7	6326	2680	88	118	0.71	
		16	5750	2389	70	83	0.71	
		17	6146	1928	87	66	0.70	
		0	7866	1140	104	102	0.79	
		top	2	744	8262	0	206	0.81
SqueezeNet	top	4	5860	3146	53	153	0.75	
		7	5860	3146	122	84	0.65	
		8	5780	3226	112	94	0.65	
		12	5126	3863	55	120	0.67	
	middle	9	0	9006	0	190	0.59	
		10	6742	2252	127	63	0.60	
		11	0	8994	0	190	0.56	

Table 3. Different group hypothesis testings with different models.

Net	Group	<i>p</i> -Value
AlexNet	top&middle	0.001
	middle&bottom	0.000
	bottom&top	0.004
	top&middle	0.007
ResNet-18	middle&bottom	0.001
	bottom&top	0.004
	top&middle	0.003
SqueezeNet	middle&bottom	0.001
	bottom&top	0.006

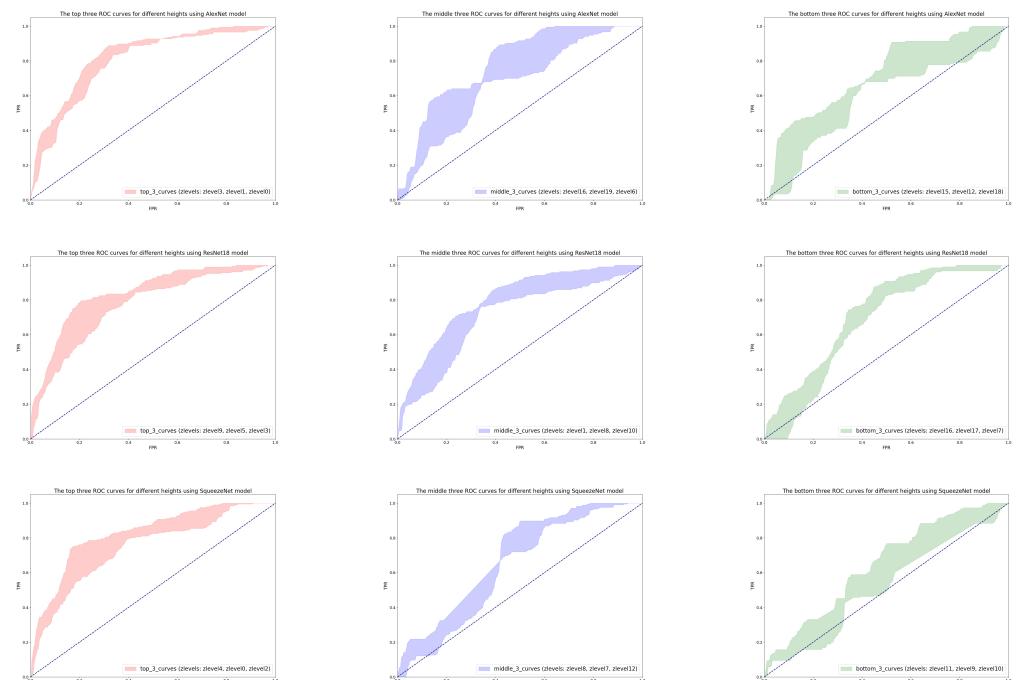


Figure 7. ROC curves for different models at different heights. The left, middle, and right figures represent the ROC curves of the AlexNet, ResNet-18, and SqueezeNet models at different magnetic field heights. The red area represents the area surrounded by the three curves with the highest AUC value, the blue area represents the area surrounded by the three curves in the middle of the AUC value, and the green area represents the area surrounded by the three curves with the lowest AUC value.

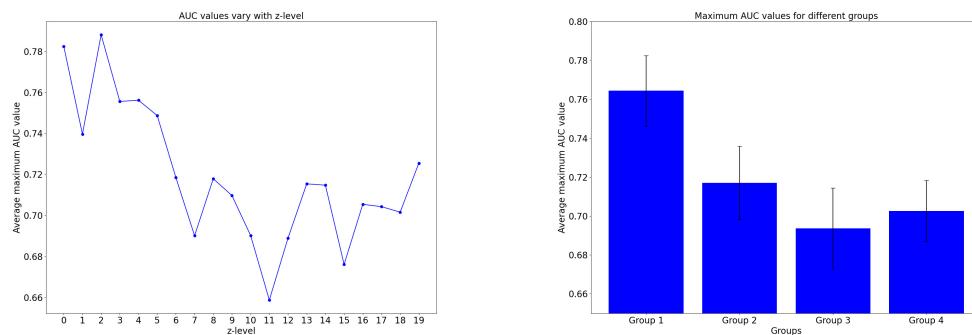


Figure 8. The curve of AUC changes with the height of the magnetic field; the left figure represents the maximum value of the average AUC as a function of the height of the magnetic field; the X-axis in the right figure represents different groups, combining five zlevels into one group; the Y-axis represents the mean value and standard deviation of the AUC of five zlevels.

5. Conclusions and Future Work

To investigate how magnetic field height affects solar flare prediction, we performed magnetic field extrapolations from photospheric magnetograms to obtain data at various atmospheric heights. Using temporally partitioned data (January–October for training, November–December for testing), we evaluated three CNN architectures (AlexNet, ResNet-18, SqueezeNet) across different height levels. Our experiments revealed a consistent inverted-“V” pattern in prediction performance versus height, peaking near typical flare formation heights (consistent with physical expectations) and declining sharply beyond the fourth extrapolation layer. This degradation likely stems from two factors: (1) errors increase in the potential field extrapolation method at greater heights, and (2) data quality filtering during preprocessing disproportionately affects the higher layers due to increased noise and artifacts in the extrapolated data. The resulting sample size reduction in the upper layers may contribute to the observed performance decline.

In this paper, we only discuss the three-dimensional magnetic field under the weak long condition. Future research work can focus on other extrapolation techniques that are closer to real data, using detection equipment for direct observation of the solar chromosphere and corona, etc., to improve the prediction performance by improving the reliability of the data.

Furthermore, the improvements made to the Solar Magnetism and Activity Telescope (SMAT) data processing techniques demonstrate its capability for reliable application to Huairou Solar Observing Station (HSOS) magnetogram datasets [36]. By addressing instrumental challenges such as the zero-level problem and optimizing observational strategies like using both wings of the spectral line, SMAT-derived magnetograms align well with established datasets like those from HMI. These advancements highlight the potential of SMAT data to contribute significantly to studies of global solar magnetic fields and space weather, particularly within the context of the unique observational framework provided by HSOS. Huairou also has high-quality magnetic field data, and our work can be used on those data in the future.

Author Contributions: Conceptualization, L.H. and X.H.; methodology, L.H. and L.X.; software, L.H. and Z.C.; validation, L.H. and Z.C.; formal analysis, L.H.; investigation, X.H. and L.X.; resources, X.H and L.X.; data curation, X.H. and L.X.; writing—original draft preparation, L.H.; writing—review and editing, L.H. and L.X.; visualization, L.H. and Z.C.; supervision, L.H.; project administration, L.H.; funding acquisition, X.H. and L.X. All authors have read and agreed to the published version of this manuscript.

Funding: This work was supported by the Strategic Priority Research Program of the Chinese Academy of Sciences (grant No. XDB0560000), the National Key R&D Program of China (grant No. 2021YFA1600504), and the National Natural Science Foundation of China (grant No. 11873060).

Data Availability Statement: The data used in this study are publicly available [33].

Acknowledgments: We thank the SDO consortium for the data.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A

Appendix A.1. Convolutional Layer

In the convolution layer, the convolution kernel (a set of learnable filters) performs local perception and feature extraction on the input data. Each convolution kernel focuses on different features of the input data, such as edges, corners, and textures. The lower-level convolutional layer, which is closer to the input, tends to learn lower-level features and can learn abstract features such as different shapes. Through the combination of multiple convolutional layers, the deeper convolutional layer in the model can learn a higher level of feature representation. At the same time, the convolution layer has the property of parameter sharing; that is, the same convolution kernel is applied to all positions of the input data. This means that the convolutional layer has fewer parameters, making the model easier to train and reducing the risk of overfitting. The convolutional layer enables the model to recognize features in the image by locally sensing the input data, regardless of the location of the features in the image. This improves the robustness of the model for image transformation.

The convolution kernel moves on the input data in a sliding way, places the convolution kernel in the upper left corner of the input data, and performs element multiplication and summation operations with the local region corresponding to the input data to obtain an output value. Then, the convolution kernel is slid to the right and down for a fixed number of steps until the entire width and height of the input data have been traversed.

Assuming the size of the input layer is $W \times W$, the size of the output layer after passing through the convolution layer is expressed as in Equation (A1).

$$(W - F + 2 * P) / S + 1, \quad (\text{A1})$$

where $F \times F$ is the size of the convolution kernel, P is the fill size, and S is the step size of the sliding window.

Appendix A.2. Pooling Layer

The aggregation layer, also known as the pooling layer, can reduce data dimensions by downsampling while retaining important feature information. In the aggregation layer, the input data are divided into several non-overlapping regions, usually using a rectangular window, and the data within each region are aggregated. Pooling layers are generally divided into two categories, such as maximum pooling and average pooling. Maximum pooling takes the maximum value within each region as the output, while average pooling calculates the average of all values within each region as the output. Through this downsampling operation, the aggregation layer not only reduces the data dimension to reduce the computational complexity and storage space but also helps the model to extract important information on the features. For example, maximum pooling preserves the strongest features within a region, helping the model identify key features. At the same time, the convergence layer can also improve the robustness of the model, because important features can be retained even if the input data change slightly.

Appendix A.3. Fully Connected Layer

The fully connected layer is a common layer structure in deep learning and is commonly used in the classifier part of neural networks. In a fully connected layer, all the neurons in the previous layer are connected to every neuron in the next layer. Each connection has a corresponding weight and bias. The fully connected layer is implemented by matrix multiplication and addition. Specifically, the output feature vector of the previous layer is multiplied with the weight matrix, and the bias vector is added. In general, after the fully connected layer, the output needs to be obtained through the activation function to achieve classification or regression of the input data. Through the fully connected layer, models can learn complex relationships between input data and output labels. It can generally be expressed by Equation (A2).

$$FC(x) = wx + b, \quad (A2)$$

where w is the weight matrix, b is the bias vector, and both w and b are learnable parameters

Appendix A.4. Activation Function

Activation functions are a key concept in deep learning, used to introduce nonlinear factors that enable neural networks to learn and model complex functions. The activation function transforms the output of neuron to determine whether to activate the neuron, which can enhance the representation and learning ability of the model and enable the neural network to solve nonlinear problems. Common activation functions include Sigmoid, Tanh, ReLU, and softmax. In AlexNet, we generally use ReLU as the activation function, as shown in Equation (A3).

$$ReLU(x) = \max(0, x), \quad (A3)$$

Appendix B

Appendix B.1. 1×1 Convolutional Kernel Strategy

In SqueezeNet, the 1×1 convolutional kernel is often used because the 1×1 convolutional kernel has fewer parameters than the 3×3 convolution kernel. Through this strategy, the number of parameters and computational effort can be significantly reduced while maintaining the expressive power of the model. For example, a 3×3 convolutional kernel has a parameter quantity of 9, while a 1×1 convolution kernel has a parameter quantity of 1. In SqueezeNet, about 80% of the convolution kernels are 1×1 , which greatly reduces the number of parameters and calculations of the model, making the model more lightweight and efficient.

Appendix B.2. Fire Module Strategy

Each FireModule contains a Squeeze layer with 1×1 convolution cores and an Expand layer with 1×1 and 3×3 convolution cores. This structure can effectively reduce the number of parameters while maintaining a high recognition accuracy. Specifically, the Squeeze layer reduces the number of channels by a 1×1 convolution, thus reducing the number of parameters and the amount of computation. The Expand layer increases the number of channels by 1×1 and 3×3 convolution to maintain the expressiveness of the model. This structure can effectively reduce the number of parameters while maintaining a high recognition accuracy. SqueezeNet further optimizes model performance by adjusting the hyperparameters in FireModule ($s1 \times 1$, $e1 \times 1$, and $e3 \times 3$).

Appendix B.3. Delayed Downsampling Strategy

In SqueezeNet, most convolutional layers use a convolutional kernel with step size 1, while a convolutional or pooled layer with step size greater than 1 is placed at the end of the network. This strategy can make many layers in the network have large activation graphs, so as to improve the classification accuracy.

With these three strategies, SqueezeNet can significantly reduce the number of parameters and computational complexity of the model while maintaining a high accuracy, and SqueezeNet can be compressed to less than 0.5 MB using model compression technology, achieving the goal of high accuracy.

References

1. Gallagher, P.; Moon, Y.; Wang, H. Active-region monitoring and flare forecasting - I. Data processing and first results. *Sol. Phys.* **2002**, *209*, 171–183. [[CrossRef](#)]
2. Leka, K.; Barnes, G. Photospheric magnetic field properties of flaring versus flare-quiet active regions. II. Discriminant analysis. *Astrophys. J.* **2003**, *595*, 1296–1306. [[CrossRef](#)]
3. Lee, K.; Moon, Y.J.; Lee, J.Y.; Lee, K.S.; Na, H. Solar Flare Occurrence Rate and Probability in Terms of the Sunspot Classification Supplemented with Sunspot Area and Its Changes. *Sol. Phys.* **2012**, *281*, 639–650. [[CrossRef](#)]
4. Schrijver, C.J. Driving major solar flares and eruptions: A review. *Adv. Space Res.* **2009**, *43*, 739–755. [[CrossRef](#)]
5. Korsos, M.B.; Ludmany, A.; Erdelyi, R.; Baranyi, T. On Flare Predictability Based on Sunspot Group Evolution. *Astrophys. J. Lett.* **2015**, *802*, L21. [[CrossRef](#)]
6. Raboonik, A.; Safari, H.; Alipour, N.; Wheatland, M.S. Prediction of Solar Flares Using Unique Signatures of Magnetic Field Images. *Astrophys. J.* **2017**, *834*, 11. [[CrossRef](#)]
7. McAteer, R.; Gallagher, P.; Ireland, J. Statistics of active region complexity: A large-scale fractal dimension survey. *Astrophys. J.* **2005**, *631*, 628–635. [[CrossRef](#)]
8. Komm, R.; Howe, R.; Hill, F.; González-Hernández, I.; Toner, C. Kinetic helicity density in solar subsurface layers and flare activity of active regions. *Astrophys. J.* **2005**, *630*, 1184–1193. [[CrossRef](#)]

9. Abramenko, V. Relationship between magnetic power spectrum and flare productivity in solar active regions. *Astrophys. J.* **2005**, *629*, 1141–1149. [[CrossRef](#)]
10. Barnes, G.; Leka, K.D.; Schrijver, C.J.; Colak, T.; Qahwaji, R.; Ashamari, O.W.; Yuan, Y.; Zhang, J.; McAteer, R.T.J.; Bloomfield, D.S.; et al. A Comparison Of Flare Forecasting Methods. I. Results From The “All-Clear” Workshop. *Astrophys. J.* **2016**, *829*, 89. [[CrossRef](#)]
11. Karakatsanis, L.; Pavlos, G. Self organized criticality and chaos into the solar activity. *Nonlinear Phenom. Complex Syst.* **2008**, *11*, 280–284.
12. Bloomfield, D.S.; Higgins, P.A.; McAteer, R.T.J.; Gallagher, P.T. TOWARD RELIABLE BENCHMARKING OF SOLAR FLARE FORECASTING METHODS. *Astrophys. J. Lett.* **2012**, *747*, L41. [[CrossRef](#)]
13. Korsós, M.B.; Romano, P.; Morgan, H.; Ye, Y.; Erdélyi, R.; Zuccarello, F. Differences in Periodic Magnetic Helicity Injection Behavior between Flaring and Non-flaring Active Regions: Case Study. *Astrophys. J. Lett.* **2020**, *897*, L23. [[CrossRef](#)]
14. Falco, M.; Costa, P.; Romano, P. Solar flare forecasting using morphological properties of sunspot groups. *J. Space Weather. Space Clim.* **2019**, *9*. [[CrossRef](#)]
15. Wan, J.; Fu, J.F.; Tan, D.M.; Han, K.; Yu, M.Y. Solar Flare Forecast Model Based on Resampling and Fusion Method. *Res. Astron. Astrophys.* **2022**, *22*, 085020. [[CrossRef](#)]
16. Wan, J.; Fu, J.F.; Wen, R.Q.; Han, K.; Yu, M.Y.; E, P. Flare Forecast Model Based on DS-SMOTE and SVM with Optimized Regular Term. *Res. Astron. Astrophys.* **2023**, *23*, 065004. [[CrossRef](#)]
17. Liu, S.; Wang, J.; Li, M.; Cui, Y.; Guo, J.; Shi, Y.; Luo, B.; Liu, S. A selective up-sampling method applied upon unbalanced data for flare prediction: Potential to improve model performance. *Front. Astron. Space Sci.* **2023**, *10*. [[CrossRef](#)]
18. Deshmukh, V.; Flyer, N.; van der Sande, K.; Berger, T. Decreasing False-alarm Rates in CNN-based Solar Flare Prediction Using SDO/HMI Data. *Astrophys. J. Suppl. Ser.* **2022**, *260*, 9. [[CrossRef](#)]
19. Benvenuto, F.; Piana, M.; Campi, C.; Massone, A.M. A Hybrid Supervised/Unsupervised Machine Learning Approach to Solar Flare Prediction. *Astrophys. J.* **2018**, *853*, 90. [[CrossRef](#)]
20. Abdulla, Y.; Wang, J.T.L.; Nie, Y.; Liu, C.; Wang, H. DeepSun: Machine-learning-as-a-service for solar flare prediction. *Res. Astron. Astrophys.* **2021**, *21*, 160. [[CrossRef](#)]
21. Li, R.; Du, Y. Full-Disk Solar Flare Forecasting Model Based on Data Mining Method. *Adv. Astron.* **2019**, *2019*. [[CrossRef](#)]
22. Huang, X.; Wang, H.; Xu, L.; Liu, J.; Li, R.; Dai, X. Deep Learning Based Solar Flare Forecasting Model. I. Results for Line-of-sight Magnetograms. *Astrophys. J.* **2018**, *856*, 7. [[CrossRef](#)]
23. Huang, X.; Zhao, Z.; Zhong, Y.; Xu, L.; Korsos, M.B.; Erdélyi, R. Short-term solar eruptive activity prediction models based on machine learning approaches: A review. *Sci.-China-Earth Sci.* **2024**, *67*, 3727–3764. [[CrossRef](#)]
24. Liu, S.; Xu, L.; Zhao, Z.; Erdélyi, R.; Korsos, M.B.; Huang, X. Deep Learning Based Solar Flare Forecasting Model. II. Influence of Image Resolution. *Astrophys. J.* **2022**, *941*. [[CrossRef](#)]
25. Sun, D.; Huang, X.; Zhao, Z.; Xu, L. Deep Learning-based Solar Flare Forecasting Model. III. Extracting Precursors from EUV Images. *Astrophys. J. Suppl. Ser.* **2023**, *266*, 8. [[CrossRef](#)]
26. Korsos, M.B.; Chatterjee, P.; Erdélyi, R. Applying the Weighted Horizontal Magnetic Gradient Method to a Simulated Flaring Active Region. *Astrophys. J.* **2018**, *857*, 103. [[CrossRef](#)]
27. Korsos, M.B.; Georgoulis, M.K.; Gyenge, N.; Bisoi, S.K.; Yu, S.; Poedts, S.; Nelson, C.J.; Liu, J.; Yan, Y.; Erdélyi, R. Solar Flare Prediction Using Magnetic Field Diagnostics above the Photosphere. *Astrophys. J.* **2020**, *896*, 119. [[CrossRef](#)]
28. Korsos, M.B.; Ruderman, M.S.; Erdélyi, R. An application of the weighted horizontal magnetic gradient to solar compact and eruptive events. *Adv. Space Res.* **2018**, *61*, 595–602. [[CrossRef](#)]
29. Korsos, M.B.; Yang, S.; Erdélyi, R. Investigation of pre-flare dynamics using the weighted horizontal magnetic gradient method: From small to major flare classes. *J. Space Weather. Space Clim.* **2019**, *9*, A6. [[CrossRef](#)]
30. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet Classification with Deep Convolutional Neural Networks. *Commun. ACM* **2017**, *60*, 84–90. [[CrossRef](#)]
31. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016 ; pp. 770–778. [[CrossRef](#)]
32. Iandola, F.N.; Moskewicz, M.W.; Ashraf, K.; Han, S.; Dally, W.J.; Keutzer, K. SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <1MB model size. *arXiv* **2016**, arXiv:1602.07360.
33. Zhao, Z.; Xu, L.; Zhu, X.; Zhang, X.; Liu, S.; Huang, X.; Ren, Z.; Tian, Y. A Large-Scale Dataset of Three-Dimensional Solar Magnetic Fields Extrapolated by Nonlinear Force-Free Method. *Sci. Data* **2023**, *10*. [[CrossRef](#)] [[PubMed](#)]
34. Molnar, C. *Interpretable Machine Learning: A Guide for Making Black Box Models Explainable*; Lulu.com: Morrisville, NC, USA , 2020.

35. Kingma, D.P.; Ba, J. Adam: A Method for Stochastic Optimization. *arXiv* **2017**, arXiv:1412.6980.
36. Demidov, M.L.; Wang, X.F.; Wang, D.G.; Deng, Y.Y. On the Measurements of Full-Disk Longitudinal Magnetograms at Huairou Solar Observing Station. *Sol. Phys.* **2018**, 293, 146. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.