Article:
Liu, S., Wang, J., Fludra, A. et al. (2026) Flare prediction modeling based on the time series of SHARP parameters along the polarity inversion line of active regions. Space Weather, 24 (3). e2025SW004687. ISSN: 1542-7390

https://doi.org/10.1029/2025sw004687

Space Weather

RESEARCH ARTICLE
10.1029/2025SW004687

Flare Prediction Modeling Based on the Time Series of SHARP Parameters Along the Polarity Inversion Line of Active Regions

Siwei Liu[1,2], Jingjing Wang[1,2], Andrzej Fludra[3], Robertus Erdélyi[4,5,6], Bingxian Luo[1,2], Tieyan Wang[7], Siqing Liu[1,2], and Qiuzhen Zhong[1,2]

[1]Key Laboratory of Solar Activity and Space Weather, National Space Science Center, Chinese Academy of Sciences, Beijing, China, [2]University of Chinese Academy of Sciences, Beijing, China, [3]STFC RAL Space, Rutherford Appleton Laboratory, Didcot, UK, [4]Solar Physics and Space Plasma Research Centre (SP^{2}RC), School of Mathematics and Statistics, University of Sheffield, Sheffield, UK, [5]Department of Astronomy, Eötvös Loránd University, Budapest, Hungary, [6]Gyula Bay Zoltan Solar Observatory (GSO), Hungarian Solar Physics Foundation (HSPF), Gyula, Hungary, [7]School of Earth Science, Yunnan University, Kunming, China

Key Points:

- The time-series model we utilize outperforms non-sequential models in predicting solar flares within the next 24 hr
- New Space-weather HMI Active Region Patches (SHARP) parameters along the Polarity Inversion Line (PIL) improve prediction performance compared to original SHARP parameters
- The proposed framework provides practical guidance for temporal modeling and feature localization in operational solar flare forecasting

Correspondence to:
B. Luo and A. Fludra,
luobx@nssc.ac.cn;
andrzej.fludra@stfc.ac.uk

Citation:
Liu, S., Wang, J., Fludra, A., Erdélyi, R., Luo, B., Wang, T., et al. (2026). Flare prediction modeling based on the time series of SHARP parameters along the polarity inversion line of active regions. *Space Weather*, *24*, e2025SW004687. https://doi.org/10.1029/2025SW004687

Received 21 AUG 2025
Accepted 23 FEB 2026

Author Contributions:
Conceptualization: Jingjing Wang, Bingxian Luo, Siqing Liu
Data curation: Jingjing Wang, Andrzej Fludra, Bingxian Luo, Tieyan Wang, Siqing Liu, Qiuzhen Zhong
Formal analysis: Siwei Liu
Funding acquisition: Jingjing Wang, Bingxian Luo, Siqing Liu, Qiuzhen Zhong
Investigation: Andrzej Fludra, Tieyan Wang
Methodology: Jingjing Wang
Project administration: Jingjing Wang, Siqing Liu
Resources: Bingxian Luo, Siqing Liu, Qiuzhen Zhong
Software: Siwei Liu
Supervision: Jingjing Wang, Bingxian Luo, Siqing Liu, Qiuzhen Zhong

Abstract The polarity inversion line (PIL) in active regions (ARs) is considered to be closely associated with solar flare eruptions. In this study, we rigorously constructed standardized data sets based on time series of different lengths using Space-weather HMI Active Region Patches (SHARP) parameters calculated along the PIL. We compared the performance of traditional non-sequential models and a time-series model in solar flare prediction tasks, as well as the predictive performance of time-series models with different input lengths within the CNN–BiLSTM–AT framework. The main findings of this study are summarized as follows: (a) SHARP parameters computed along the PIL consistently yield better prediction performance than those calculated over entire active regions. (b) In realistic and highly imbalanced prediction scenarios, the time-series model outperforms non-sequential models, achieving an F1 score of 0.59 for strong-flare prediction. (c) Robustness tests and sliding-window probability forecasts further demonstrate the practical feasibility of the proposed approach. These results provide useful guidance for data representation and model selection in solar flare forecasting.

Plain Language Summary The polarity inversion line (PIL) in active regions (ARs) is closely associated with solar flare eruptions. This study aims to improve solar flare prediction by incorporating time-series modeling and SHARP parameters derived along the PIL. We constructed standardized data sets and compared the performance of traditional non-sequential models, including logistic regression (LR), random forest (RF), and support vector machine, with a CNN–BiLSTM–AT time-series model. The results show that time-series modeling leads to more reliable flare predictions than non-sequential approaches, particularly when PIL-based SHARP parameters are used. These parameters provide a more informative representation of magnetic complexity and improve overall prediction performance. We also investigated the effect of different input sequence lengths and demonstrated the practical applicability of the proposed method through sliding-window probability forecasts. This study highlights the importance of temporal information and localized magnetic features in advancing solar flare forecasting.

1. Introduction

Due to the potential for disastrous space weather events caused by solar eruptions, which can significantly impact aerospace operations and space technology systems, forecasting solar eruption activity has become an essential measure for disaster reduction and protection. Consequently, the development of advanced and reliable forecasting models for solar eruptions remains a critical research objective in space weather studies.

Solar observational data sets obtained from missions such as the Solar Dynamics Observatory (SDO) (Pesnell et al., 2012) and the Geostationary Operational Environmental Satellites (GOES) series (Hill et al., 2005) provide rich repositories of long-term and high-resolution records of solar activity. The open data initiatives adopted by space agencies worldwide have greatly facilitated data accessibility, fostering a favorable environment for scientific research and methodological innovation. Among these data sets, the Space-weather HMI Active Region Patches (SHARP) (Bobra et al., 2014) represent a widely used SDO data product that summarizes vector magnetic

Validation: Siwei Liu, Jingjing Wang, Robertus Erdélyi
Visualization: Siwei Liu
Writing – original draft: Siwei Liu
Writing – review & editing: Siwei Liu, Jingjing Wang, Andrzej Fludra, Robertus Erdélyi

field properties of solar active regions (ARs). The physical quantities captured by SHARP parameters and their temporal evolution are closely associated with solar flare generation, as demonstrated in previous studies (Ran et al., 2022). As a result, SHARP data have been extensively employed in the development of flare prediction models (Georgoulis et al., 2024).

Early flare forecasting studies based on SHARP parameters primarily focused on static or single-time-point machine learning approaches. A seminal work by Bobra and Couvidat (2015) demonstrated that vector magnetic field features derived from SDO/HMI SHARP data can be effectively used for solar flare prediction using machine learning algorithms, laying the foundation for many subsequent studies. Lim et al. (2019) established empirical relationships between SHARP parameters and daily flaring rates to estimate the occurrence probabilities of major flares. Leka et al. (2018) developed the Discriminant Analysis Flare Forecasting System, which predicts solar flare occurrence based on SHARP parameters. A systematic comparison conducted by Barnes et al. (2016) showed that discriminant analysis methods provide competitive and robust performance for operational forecasting. In addition, Kontogiannis et al. (2017) proposed a physically motivated flare prediction approach using magnetic connectivity matrices derived from SHARP data to characterize the non-potentiality of active regions without relying on machine learning techniques.

The growing availability of large-scale solar data sets has enabled the widespread application of machine learning techniques in flare prediction, leading to increasingly mature modeling pipelines (Huang et al., 2024). Campi et al. (2019) utilized 171 flare-predictive active region properties and demonstrated the importance of feature ranking in enhancing predictive capability across different supervised learning methods. Winter and Balasubramaniam (2015) identified a relationship between the maximum ratio of flare flux to non-flare background, enabling the classification of solar flare classes using machine learning. More recently, Li et al. (2024) proposed a hybrid model combining a deep residual network with a support vector machine (SVM) to improve flare classification performance. Liu et al. (2017) further demonstrated the effectiveness of Random Forest algorithms for predicting solar flares using SHARP parameters.

Most of the studies discussed above rely on non-sequential, single-time-point representations of active regions. However, solar flare initiation is inherently a dynamic process, motivating the exploration of time-series-based prediction models. Liu et al. (2019) employed a Long Short-Term Memory (LSTM) network to predict whether an active region would produce an M- or C-class flare within the next 24 hr, using time series of SHARP parameters and flare history data. The interpretability of temporal features learned by the LSTM model was further investigated by Sun et al. (2019). Thibeault et al. (2022) proposed a flare prediction framework based on data assimilation with sandpile models, which capture temporal patterns in GOES X-ray time series and improve the success rate of "All-Clear" forecasts. In addition, Chen et al. (2019) incorporated multivariate SHARP parameter time series together with HMI magnetogram sequences into a deep learning framework to model precursor evolution, while Guastavino et al. (2022) employed video-based magnetogram sequences as direct input to Convolutional Neural Networks (CNN), enabling the extraction of spatiotemporal features from raw observations.

Recent advances in time-series flare prediction have increasingly focused on addressing challenges such as extreme class imbalance and data representation. Vural et al. (2024) proposed a contrastive representation learning framework designed for highly imbalanced multivariate time series of photospheric magnetic field parameters, achieving improved discrimination of rare but significant flare events. Extending this work, Vural et al. (2025) conducted a comparative analysis of vector-based, time-series-based, and graph-based representations for solar flare prediction, highlighting the importance of temporal structure and relational information in magnetic field evolution. In parallel, Li et al. (2025) systematically evaluated time-series data augmentation techniques for deep learning-based flare prediction and demonstrated that appropriate augmentation strategies can effectively mitigate data scarcity and imbalance issues.

Despite these advances, current time-series flare prediction models still exhibit notable limitations. While some models achieve high True Skill Statistic (TSS) scores, they often generate an excessive number of false positives, which compromises their operational utility. Consequently, human expert judgment remains indispensable in most operational flare forecasting settings. This situation highlights the need to explore alternative modeling perspectives that can further enhance the reliability and practical value of automated flare prediction systems.

From a physical perspective, Polarity Inversion Line (PIL) regions are widely recognized as critical sites for magnetic energy accumulation and release during solar eruptions (Toriumi & Wang, 2019). For example, Korsós

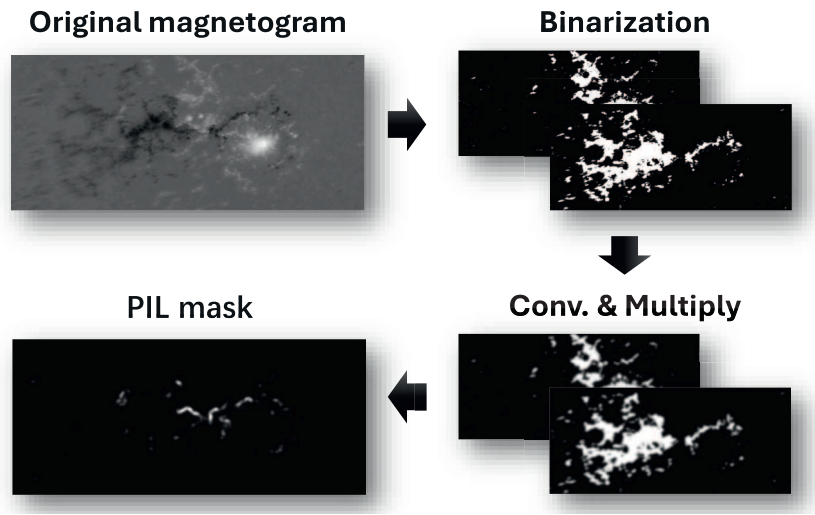

Figure 1. A simple process for extracting a PIL mask.

et al. (2015) found that converging motions of sunspots and variations in magnetic gradients along neutral lines are closely related to solar flare initiation. Cicogna et al. (2021) further demonstrated that flare forecasting algorithms based on high-gradient PILs outperform methods relying on global active region properties, emphasizing the predictive value of magnetic complexity localized around PILs. Building on this physical insight, Wang et al. (2019) utilized the *hmi.sharp_cea_*720*s* data series to construct gradient-based PIL masks and compute a set of new SHARP parameters along the PIL. A simplified workflow for extracting a PIL mask is illustrated in Figure 1. By multiplying the PIL mask element-wise with the original SHARP parameter maps, new PIL-focused parameters are obtained. Their results showed that these localized parameters significantly improve predictive performance compared to SHARP parameters computed over entire active regions.

Based on the work of Wang et al. (2019), this study utilizes SHARP parameters calculated along the PIL to conduct new time-series flare prediction experiments. We adopt time-series models to enable the learning of temporal evolution patterns of active regions. From a temporal perspective, we observe that prior to the occurrence of strong flares (M-class and above), PIL-based SHARP parameters often exhibit enhanced timing characteristics, such as abrupt signal intensification, as illustrated in Figure 2. Consequently, this study focuses on the evolution of PIL-based SHARP parameters. In contrast to previous studies that constructed models using single parameters, this work (a) employs multiple PIL-based SHARP parameters to build more expressive time-series models and (b) systematically analyzes the resulting improvements in predictive efficacy. Our goal is to enhance the practical performance of 24-hr solar flare forecasting and to provide valuable insights from both modeling and data perspectives.

This study makes several contributions to the field of solar flare prediction: (a) we construct a standardized data set comprising over a thousand active regions with rigorous time-series classification; (b) we demonstrate that time-series models substantially outperform non-sequential approaches, achieving an approximately 80% relative increase in the recall of strong flare events; (c) we establish that calculating SHARP parameters along the PIL rather than across entire active regions enhances prediction performance, yielding a 34% relative improvement in F1 score for time-series models; (d) we perform a systematic ablation study to validate the architectural design of the proposed time-series model and to clarify the complementary roles of feature projection, temporal dependency modeling, and attention mechanisms; (e) we identify that a 24-hr input sequence provides an optimal balance between prediction accuracy and model stability; (f) our robustness analysis confirms the model's resilience to feature perturbations while revealing previously underappreciated parameters that gain significance when computed along the PIL; and (g) we develop a sliding-window prediction methodology that enables intuitive visualization of model behavior and reveals how different sequence lengths capture distinct temporal patterns in flare precursors. Collectively, these advances strengthen the feasibility of operational 24-hr solar flare forecasting systems.

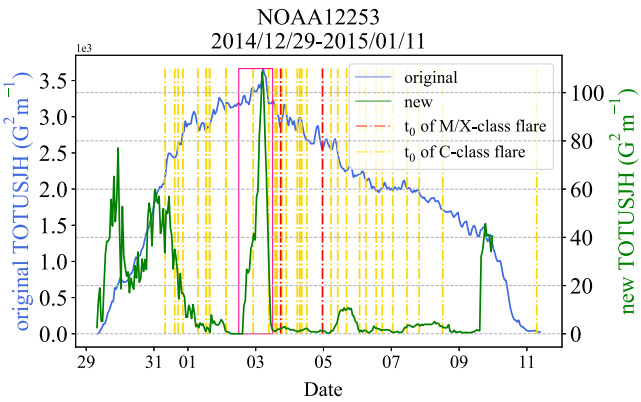

Figure 2. This picture shows the difference between one of the original Space-weather HMI Active Region Patches (SHARP) parameters and the new SHARP parameter along the PIL during an AR period (taking the total unsigned current helicity on AR 12253 as an example). The pink box marks the timing characteristics added by the new SHARP parameters. The blue curve represents the original SHARP parameter, while the green curve displays the new SHARP parameter. Specifically, the SHARP parameter shown in this example is the total unsigned current helicity (TOTUSJH). The red vertical dashed lines indicate the start times of M/X class flare eruptions, and the orange-yellow vertical dashed lines represent the start times of C-class flare eruptions.

2. Data Preparation

The original data are sourced from the SDO data product called Space-weather HMI Active Region Patches (SHARP), which automatically identifies AR patches from the vector magnetic field data.

The total amount of data used in this study exceeds 1 million records. The data covers 1,290 solar active regions (HARPNUM ranging from 1 to 7,325; IDs assigned chronologically) and includes all available HMI SHARP data from this period, spanning approximately one solar cycle from 1 May 2010, to 29 November 2021 (roughly solar cycle 24). In the flare forecasting task, considering the scarcity of data available for training, we set the test-to-training data ratio to 19 to allow the model to learn from more cases. About 130 active regions serve as the test set, while 1,160 are in the training set.

We divided the 1,290 active regions numbered from 1 to 7,325 into 10 groups, each containing 129 active region numbers. Each group is further split into 10 partitions, called "folds" (each fold containing 12 or 13 active region numbers) for cross-validation. For the experiment with the N-th fold, the test set consists of the N-th fold from each group, while the remaining active regions comprise the training set.

This method of splitting the data set has two advantages: (a) It evenly spreads active region numbers of the test set across the entire solar cycle, preventing evaluation distortion caused by uneven flare distribution in high and low years. If the training set is concentrated in high years and the test set in low years, the test set will have very few flare samples, rendering the evaluation unreliable. Conversely, if the training set is in low years and the test set in high years, the model's training on flare samples would be inadequate. (b) As we divide the data set using active region numbers, it also avoids physical overlap between the test and training sets, ensuring data independence.

2.1. Data Partitioning and Class Definition

We define M and above class flares as strong flares, and flares below M-class or quiet periods as non-strong flares. A classification of all data points is conducted using the following criteria:

Class 0 designates data points located in an active region (AR) where no strong flares occur, representing a non-strong flare point. Class 1 indicates a strong flare event, signified by the occurrence of a strong flare within a 24-hr window.

For non-sequential parameter models, training is performed directly using data labeled as class 1 and class 0. In contrast, time-series models utilize the label of the last point in the sequence as the label for the entire sequence.

This classification methodology for time series and non-time series data sets is schematically illustrated in Figure 3, which demonstrates our approach to labeling data for both sequential and non-sequential models in solar flare prediction. The figure clearly distinguishes between point-wise annotation for non-sequential model and segment-wise annotation for time series model, with the critical 24-hr pre-flare window highlighted as the decision boundary for positive classifications.

2.2. Data Normalization

The time span of each AR in the original data ranges from 1 to 350 hr, and most series are discontinuous, so normalization is required. First, we interpolate the sequence on a 12-min cadence (the SHARP sampling rate) to fill only the missing points, leaving all originally observed values unchanged. Because magnetograms evolve slowly, we then resample the sequence to a 1-hr cadence to obtain the final series for modeling, which reduces temporal redundancy and computational cost while preserving the relevant evolution.

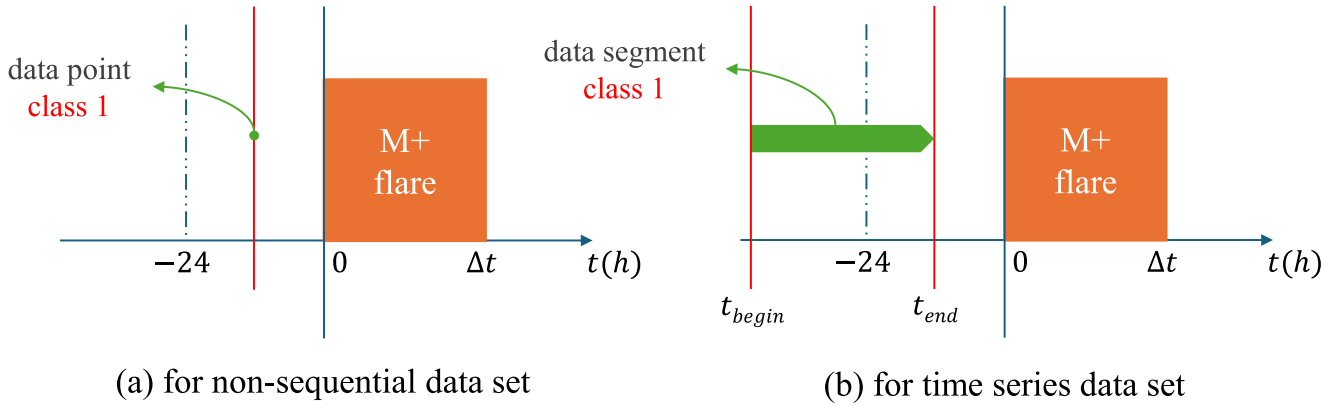

Figure 3. For non-sequential models, we annotate individual points, as shown in (a), where Class 1 (positive) refers to data points within 24 hr before a strong flare occurs. For sequential models, we annotate data segments, as shown in (b), where t_{begin} and t_{end} represent the start and end points of the time series, respectively. Only when t_{end} falls within the 24-hr window before a strong flare event, the entire data segment is considered a positive instance (Class 1).

2.3. Sampling Methods

As one can imagine, when it comes to timing issues, the problem of data imbalance is still very troublesome. But we must make full use of all data and try to avoid unwarranted down-sampling. Due to the significant impact of imbalanced data sets on the efficacy of machine learning Japkowicz and Stephen (2002), strong flare events and non-strong flare events can be processed in the following two ways respectively. For strong flare events with high importance and small amount of data, we use sliding window overlapping sampling. For the less critical and more numerous non-strong flare events, we use non-overlapping sampling. This approach helps to balance the training set while making the most efficient use of all available data. The sampling method is shown in Figure 4. At the same time, we padded the front end of the data with the initial values when the data duration was greater than 5/8 but less than 1 times the model input sequence duration, to avoid losing shorter event data. This balances the data set to a certain extent, making it easier for the machine to learn the precursor information of strong flares.

3. Experimental Design of Flare Prediction Model

3.1. Time Series Model Structure

The overall model employs a one-dimensional CNN combined with a bidirectional LSTM (Long Short-Term Memory) network and an attention mechanism (Zhang et al., 2023). Leveraging the weight sharing and translation invariance properties of CNN convolutional kernels, it effectively extracts local patterns and features from time series data. The CNN output is then provided as input to the bidirectional LSTM network to further capture

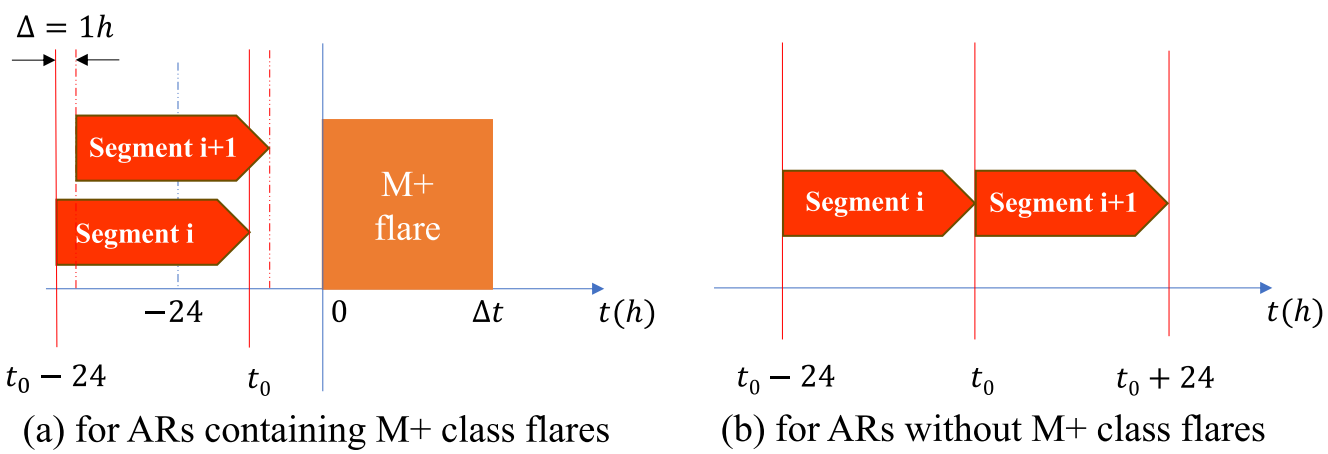

Figure 4. (a) Shows the sampling method of time series data for ARs that contain strong flares; (b) shows the sampling method of time series data that does not contain strong flares.

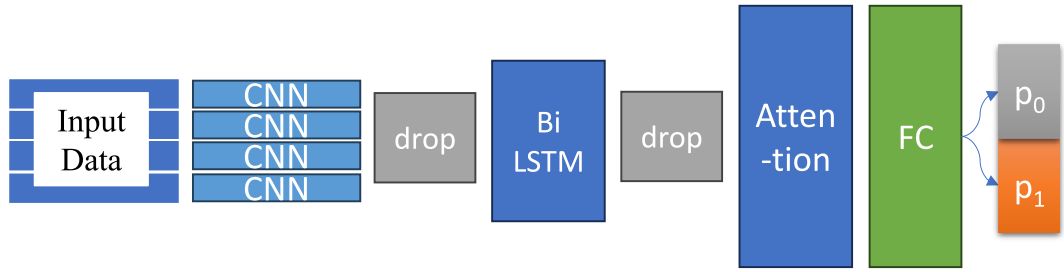

Figure 5. This figure shows the basic architecture of the CNN–BiLSTM–AT model in this study. The output prediction vector is normalized to two categories: 0 and 1 through softmax, where category 0 represents non-strong flare (negative) cases, while category 1 represents strong flare (positive) cases.

the temporal features of the parameters. Within the BiLSTM layer, the temporal information is transformed into a set of tensors, which are subsequently passed to the attention mechanism layer. The attention mechanism layer automatically assigns weights to the input parameters by computing a weighted average of the hidden states from all time steps in the encoder, thereby obtaining the input variables for the next layer. To mitigate overfitting and enhance the feature extraction generalization, dropout layers have been added before and after the convolutional layer. The model diagram is shown in Figure 5.

The model processes sequential data with a dynamic time window of N+1 steps, where N is the number of hours in the input data. This flexible time window allows the model to adapt to different prediction requirements, enabling forecasts for varying future time periods. The input data comprises 16 feature dimensions per time step, capturing comprehensive information about solar activity patterns. Prior to training, all features undergo normalization using MinMaxScaler to ensure consistent scale across different parameters, which significantly improves the convergence and performance of gradient-based optimization algorithms.

The architecture begins with a one-dimensional convolutional layer containing 64 filters with a kernel size of 1 and a ReLU activation function. Rather than capturing temporal patterns across adjacent time steps, this layer acts as a per-time-step feature projection and mixing operation, mapping the original SHARP parameter vectors into a higher-dimensional latent space while preserving the original sequence length. This transformation provides a more expressive and stable feature representation for subsequent temporal modeling by the BiLSTM and attention modules.

The extracted features are then fed into a bidirectional LSTM layer with 64 units that processes the sequence in both forward and backward directions. This bidirectional approach enables the model to capture long-term dependencies and contextual information from both past and future contexts within the sequence window. The BiLSTM layer is configured with return sequences = True to preserve the full sequence output, maintaining the temporal dimension crucial for the subsequent attention processing. This is followed by another dropout layer with a rate of 0.3 to further enhance regularization and prevent overfitting.

The attention mechanism computes attention weights using a dense layer with softmax activation. These attention weights effectively highlight the most important time steps in the sequence for the prediction task, allowing the model to focus on the most relevant temporal patterns when making predictions about future strong solar flares. The attention output is flattened as a Fully Connected Layer (FC) and connected to a final dense layer with 2 units and softmax activation, producing probability distributions for the binary classification of whether a strong solar flare will occur within the forecast horizon. The model is optimized using the Adam optimizer with standard (unweighted) categorical cross-entropy loss function, which is particularly suitable for classification tasks.

3.2. Experiment Content

First, we analyzed the performance of the non-sequential model and the time series model when the original SHARP parameters and the new SHARP parameters along the PIL are used as input respectively.

Wang et al. (2019) compared the prediction performance of these two types of SHARP parameters under a single-parameter model. Our study compares the prediction performance of the original SHARP parameters and the new SHARP parameters under the multi-parameter models.

Table 1
Confusion Matrix for Binary Classification

	Predicted class (forecasts)	
Actual Class (observations)	Positive	Negative
Positive	TP (True Positive)	FN (False Negative)
Negative	FP (False Positive)	TN (True Negative)

Note. This table shows the basic definitions of the confusion matrix elements.

Second, we use the same test set to evaluate the non-sequential model and the time-series model.

Third, in order to find out the optimal length of data for the time-series model, we compare the performance of models with different time series lengths on the same test set.

Fourth, to further evaluate the importance of the parameters, we select different parameter columns to perform global order shuffling and then input them into the model. By comparing its impact on the model performance score, we can see how important different parameters are to this model.

3.3. Evaluation Metrics

For a binary classification task like the flare prediction, the confusion matrix is listed in Table 1. The true positive (TP) is the hit case where the strong-flare events are correctly classified as "strong flare" category. The false positive (FP) is the false alarm case where the non-strong flare events are falsely classified as "strong flare" category. The false negative (FN) is the miss case where strong flare events are falsely classified as "non-strong flare" category. The true negative (TN) is the correct negative samples where the non-strong flare samples are correctly classified as "non-strong flare" category.

Based on the confusion matrix, we adopt recall, precision, F1 score, Heidke Skill Score (HSS), TSS as evaluation metrics. They are calculated according to Equations 1–5. We focus on the prediction performance of strong flare events. If we calculate the precision rate for all samples, the results from the more numerous but less important non-strong flares will dilute the results for the part we are concerned with. Therefore, in the presentation of results, we will primarily showcase the forecasting results for strong flares, while the remaining non-strong flare cases will be presented as "all clear" category in the results.

$$\text{recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{1}$$

$$\text{precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{2}$$

$$F1 = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \tag{3}$$

$$\text{HSS} = \frac{2\,(\text{TP} \times \text{TN} - \text{FP} \times \text{FN})}{(\text{TP} + \text{FN})\,(\text{FN} + \text{TN}) \;+\; (\text{TP} + \text{FP})\,(\text{FP} + \text{TN})} \tag{4}$$

$$\text{TSS} = \frac{\text{TP}}{\text{TP} + \text{FN}} - \frac{\text{FP}}{\text{FP} + \text{TN}} \quad (=\text{sensitivity} - (1 - \text{specificity})) \tag{5}$$

4. Comparison Analysis of Model Performance

4.1. Comparison of Original SHARP Parameters and New SHARP Parameters

The model performance obtained by using the original SHARP parameters and the new SHARP parameters along the PIL as input data is shown in Table 2. Non-sequential data are trained using three commonly adopted baseline models, including logistic regression (LR), random forest (RF), and SVM, while time-series data are trained using the previously mentioned CNN–BiLSTM–AT model. For the non-sequential approaches, LR is implemented as a linear classifier with class-weighted loss, RF is trained using an ensemble of decision trees with bootstrap aggregation, and SVM employs a radial basis function kernel to capture nonlinear decision boundaries.

For point-in-time prediction with non-sequential models (LR, RF, and SVM), we work with a highly imbalanced data set, for which the class weight is the most influential hyperparameter. To ensure that the reported results are close to the performance upper bound of each non-sequential model under imbalance, we perform a sweep over the positive-to-negative class weight ratio from 1:1 up to 1:20 and select the model with the best validation

Table 2
Prediction Results of the Models

Models based on different data sets							
Model	Data	Category	Precision	Recall	F1 score	HSS	TSS
Non-sequential (LR)	Original	all clear	0.99 ± 0.01	0.98 ± 0.02	0.98 ± 0.01	0.30 ± 0.07	0.38 ± 0.06
		strong flare	**0.27 ± 0.06**	**0.40 ± 0.07**	**0.32 ± 0.06**		
Non-sequential (LR)	New	all clear	0.99 ± 0.01	0.99 ± 0.01	0.99 ± 0.01	0.41 ± 0.06	0.42 ± 0.05
		strong flare	**0.43 ± 0.08**	**0.40 ± 0.06**	**0.41 ± 0.08**		
Non-sequential (RF)	Original	all clear	0.99 ± 0.01	0.99 ± 0.01	0.99 ± 0.01	0.36 ± 0.06	0.40 ± 0.05
		strong flare	**0.35 ± 0.06**	**0.42 ± 0.07**	**0.38 ± 0.05**		
Non-sequential (RF)	New	all clear	0.99 ± 0.01	1.00 ± 0.00	0.99 ± 0.01	0.52 ± 0.08	0.45 ± 0.06
		strong flare	**0.71 ± 0.09**	**0.45 ± 0.08**	**0.55 ± 0.06**		
Non-sequential (SVM)	Original	all clear	0.99 ± 0.01	0.98 ± 0.02	0.99 ± 0.01	0.33 ± 0.06	0.38 ± 0.06
		strong flare	**0.30 ± 0.05**	**0.40 ± 0.07**	**0.34 ± 0.09**		
Non-sequential (SVM)	New	all clear	0.99 ± 0.01	1.00 ± 0.00	0.99 ± 0.01	0.36 ± 0.06	0.30 ± 0.05
		strong flare	**0.50 ± 0.08**	**0.30 ± 0.06**	**0.38 ± 0.07**		
Time-series	Original	all clear	0.99 ± 0.01	0.94 ± 0.03	0.96 ± 0.02	0.41 ± 0.09	0.71 ± 0.08
		strong flare	**0.31 ± 0.06**	**0.77 ± 0.09**	**0.44 ± 0.07**		
Time-series	New	all clear	0.99 ± 0.01	0.97 ± 0.03	0.98 ± 0.02	0.58 ± 0.10	0.72 ± 0.08
		strong flare	**0.49 ± 0.08**	**0.75 ± 0.08**	**0.59 ± 0.07**		

Note. This table shows the comparative performance of the original SHARP parameters and the new SHARP parameters in the non-sequential models and the time-series model (sequence length = 24 hr/25 time steps).

performance. Unless otherwise specified, the remaining hyperparameters are set to commonly used default values in the scikit-learn library.

All input features for the non-sequential and time-series models are scaled using Min–Max normalization, in which each feature is linearly mapped to the range (0, 1) based on statistics computed from the training set only. The same scaling parameters are subsequently applied to the validation and test sets to avoid information leakage.

Model selection and evaluation are performed on the validation set using metrics focused on the positive class (strong flares), primarily TSS, HSS, and F1 score. To enable a direct comparison between non-sequential and time-series models, all models are evaluated on the same test set. In the non-sequential setting, a single time point is used as input, whereas in the time-series model this same time point serves as the endpoint of a 24-hr sequence (25 time steps).

The model issues an alert for a strong flare when the predicted probability exceeds 0.5; otherwise, the event is classified as "all clear." The predictive performance of all models is summarized in Table 2, reported as the mean and standard deviation over ten-fold cross-validation, with results for strong flares highlighted in bold. This boldface convention is also applied to the remaining performance tables.

As shown in Table 2, for the non-sequential models, the use of the new SHARP parameters generally leads to a modest increase in recall and a more pronounced improvement in precision for strong-flare prediction, compared to models using AR-wide SHARP parameters. As a result, the overall classification skill of non-sequential models is consistently enhanced. For the time-series model, the improvement is more substantial, with a clear increase in both F1 score and skill scores, indicating a more reliable and balanced prediction performance when temporal information along PILs is incorporated.

The performance gains can be further quantified using relative changes in key evaluation metrics (all percentage changes are computed with respect to models using AR-wide SHARP parameters). For non-sequential models, the F1 score for strong-flare prediction increases by approximately 27%–62% when using PIL-based SHARP parameters, with the largest improvement observed for the Random Forest classifier. Correspondingly, the HSS of non-sequential models improves by about 9%–44% relative across different architectures. For the time-series

model, the adoption of PIL-based parameters results in a consistent relative improvement of approximately 34% in F1 score and 41% in HSS. These results demonstrate that the new SHARP parameters computed along polarity inversion lines provide more discriminative information than AR-wide parameters, benefiting both non-sequential and time-series flare forecasting models.

Among the non-sequential approaches, the Random Forest model using the new SHARP parameters achieved the best performance. Nevertheless, the proposed time-series model shows a further improvement of about 10% in overall skill metrics compared to this non-sequential baseline, indicating the added value of explicitly modeling temporal evolution in flare prediction.

The superior performance of PIL-based parameters can be attributed to their ability to capture magnetic field information specifically in regions most relevant to flare genesis. While traditional SHARP parameters measure properties across entire active regions, our approach focuses on polarity inversion lines where magnetic stress and energy accumulation are most pronounced. This targeted approach provides the prediction models with more meaningful precursor signals, allowing them to better distinguish between flare-productive and flare-quiet regions. Furthermore, the evolution of these parameters over time, especially when captured by our sequential models, reveals critical patterns of magnetic field destabilization that often precede major flare events.

4.2. Ablation Study of Temporal Model Components

To examine the role of individual components in the proposed time-series forecasting framework, we performed an ablation study in which key modules of the CNN–BiLSTM–Attention (CNN–BiLSTM–AT) architecture were systematically removed or replaced. All model variants were trained and evaluated using the same data sets, cross-validation folds, input features, and preprocessing steps described in Section X, ensuring that performance differences can be attributed to architectural changes rather than experimental settings.

The full model combines a one-dimensional convolutional layer with kernel size 1, a bidirectional LSTM with 64 hidden units, and a feature-wise attention mechanism. As discussed in Section 3.1, the Conv1D layer acts primarily as a per-time-step feature projection and mixing operation, while the BiLSTM captures temporal evolution in the magnetic parameters and the attention mechanism modulates the relative contribution of latent features. Based on this architecture, we considered several simplified variants, including models without the CNN layer (BiLSTM + AT), without the LSTM layer (CNN + AT), without attention (CNN + BiLSTM), a BiLSTM-only baseline, and a lightweight Transformer encoder with comparable capacity. Except for the modified components, all training hyperparameters and optimization settings were kept identical across experiments.

The results are summarized in Table 3. All temporal models exhibit consistently high recall for strong flares, indicating that PIL-based time-series features provide a robust signal for flare occurrence. Differences between models become apparent when considering precision and skill-based metrics. The full CNN–BiLSTM–AT model achieves the most balanced performance, with a positive-class F1 score of 0.59 and an HSS of 0.58, outperforming all ablation variants. Removing individual components leads to a clear reduction in overall skill, primarily driven by a decrease in positive-class precision.

Among the ablation variants, excluding the attention mechanism results in the smallest performance degradation, suggesting that temporal dependency modeling plays a dominant role, while attention offers an additional but secondary benefit. In contrast, removing either the CNN or the LSTM layer, or replacing the recurrent structure with a Transformer encoder, produces a more pronounced drop in F1, HSS, and TSS. This behavior indicates that neither temporal modeling nor feature projection alone is sufficient to fully exploit the discriminative structure of PIL-related magnetic parameters.

Taken together, these results suggest that the performance gains of the proposed framework arise from the combination of complementary components rather than any single module. The CNN–BiLSTM–AT architecture provides a practical balance between feature transformation, temporal modeling, and adaptive reweighting, leading to improved and more stable forecasting skill for strong solar flares.

Table 3
Ablation Study Results for Different Temporal Model Variants (mean ± std)

Model	Category	Precision	Recall	F1 score	HSS	TSS
CNN + BiLSTM + AT (Full)	all clear	0.99 ± 0.01	0.97 ± 0.03	0.98 ± 0.02	0.58 ± 0.10	0.72 ± 0.08
	strong flare	**0.49 ± 0.08**	**0.75 ± 0.08**	**0.59 ± 0.07**		
No Attention (CNN + BiLSTM)	all clear	0.99 ± 0.01	0.96 ± 0.03	0.97 ± 0.02	0.54 ± 0.09	0.70 ± 0.07
	strong flare	**0.46 ± 0.07**	**0.73 ± 0.08**	**0.56 ± 0.07**		
BiLSTM only	all clear	0.99 ± 0.01	0.95 ± 0.04	0.97 ± 0.03	0.50 ± 0.11	0.68 ± 0.08
	strong flare	**0.43 ± 0.08**	**0.72 ± 0.09**	**0.54 ± 0.09**		
No CNN (BiLSTM + AT)	all clear	0.99 ± 0.01	0.95 ± 0.04	0.97 ± 0.03	0.48 ± 0.12	0.67 ± 0.09
	strong flare	**0.41 ± 0.10**	**0.71 ± 0.09**	**0.52 ± 0.08**		
No LSTM (CNN + AT)	all clear	0.99 ± 0.01	0.94 ± 0.04	0.96 ± 0.03	0.46 ± 0.13	0.66 ± 0.10
	strong flare	**0.39 ± 0.09**	**0.70 ± 0.08**	**0.50 ± 0.10**		
Transformer	all clear	0.99 ± 0.01	0.94 ± 0.04	0.96 ± 0.03	0.47 ± 0.11	0.67 ± 0.09
	strong flare	**0.40 ± 0.07**	**0.71 ± 0.09**	**0.52 ± 0.09**		

Note. All results are reported as mean and standard deviation across cross-validation folds. HSS and TSS are computed from the overall confusion matrix and are therefore identical for the two classes. Variations in standard deviation across ablation variants reflect differences in optimization stability induced by architectural changes.

4.3. Time Series Model Results Evaluation

4.3.1. Optimization and Adjustment of Input Duration Hyperparameter

Table 4 reports the performance of models trained with different sequence lengths and evaluated on a test set constructed analogously to the training set: positive (strong-flare) examples are also generated via overlapping sampling, yielding train–test class proportions that are closely matched and nearly balanced. This setup helps highlight subtle differences between models and better reflects their ability to learn strong-flare events.

Table 4
Prediction Results of the Model (mean ± std)

Tsq	Category	Precision	Recall	F1 score	HSS	TSS
6 h	all clear	0.71 ± 0.12	0.95 ± 0.01	0.81 ± 0.08	0.63 ± 0.08	0.68 ± 0.07
	strong	**0.96 ± 0.02**	**0.72 ± 0.07**	**0.82 ± 0.04**		
12 h	all clear	0.75 ± 0.08	0.93 ± 0.02	0.83 ± 0.04	0.71 ± 0.06	0.74 ± 0.05
	strong	**0.95 ± 0.02**	**0.81 ± 0.05**	**0.87 ± 0.03**		
18 h	all clear	0.75 ± 0.08	0.92 ± 0.02	0.82 ± 0.05	0.69 ± 0.07	0.73 ± 0.06
	strong	**0.94 ± 0.03**	**0.81 ± 0.06**	**0.87 ± 0.042**		
24 h	all clear	0.75 ± 0.08	0.92 ± 0.01	0.82 ± 0.05	0.70 ± 0.05	0.73 ± 0.03
	strong	**0.94 ± 0.02**	**0.82 ± 0.03**	**0.87 ± 0.02**		
30 h	all clear	0.76 ± 0.08	0.90 ± 0.02	0.82 ± 0.04	0.70 ± 0.06	0.73 ± 0.04
	strong	**0.92 ± 0.03**	**0.83 ± 0.05**	**0.88 ± 0.03**		
36 h	all clear	0.79 ± 0.05	0.90 ± 0.02	0.84 ± 0.03	0.73 ± 0.04	0.75 ± 0.04
	strong	**0.93 ± 0.03**	**0.85 ± 0.04**	**0.89 ± 0.03**		
42 h	all clear	0.77 ± 0.07	0.90 ± 0.01	0.83 ± 0.04	0.70 ± 0.07	0.73 ± 0.07
	strong	**0.93 ± 0.02**	**0.83 ± 0.07**	**0.87 ± 0.04**		
48 h	all clear	0.77 ± 0.07	0.91 ± 0.02	0.83 ± 0.04	0.71 ± 0.05	0.74 ± 0.04
	strong	**0.94 ± 0.02**	**0.83 ± 0.04**	**0.88 ± 0.02**		

Note. This table shows the prediction performance of the model for time series data with varying input lengths. Here, "Tsq" represents the duration of the input data segment.

10.1029/2025SW004687

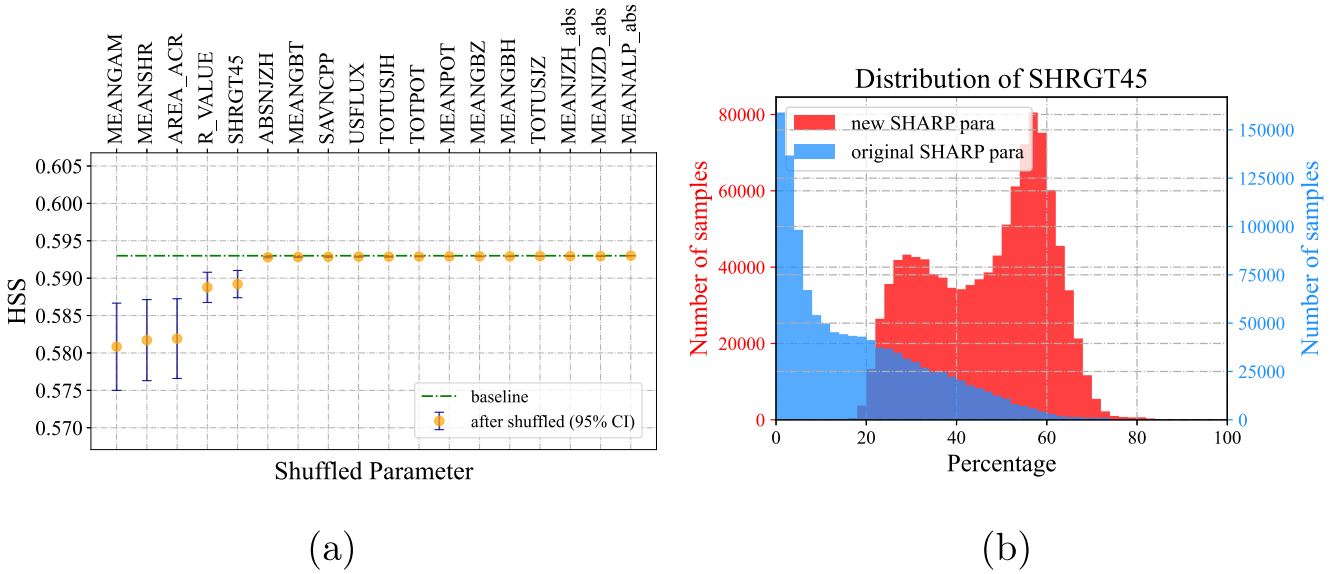

Figure 6. (a) Results of the stability evaluation under correlation-aware feature permutation. The green dashed line denotes the baseline model performance without parameter perturbation. Yellow markers with dark blue confidence intervals indicate model performance after jointly permuting correlated parameter groups. Parameter groups on the *x*-axis are ordered by increasing sensitivity, measured by the relative performance degradation. Panel (b) shows a comparison of the SHRGT45 distribution before and after recalculation along the PIL.

It can be observed from Table 4 that when the input time series length is increased from 6 to 12 hr, the model's forecasting performance is significantly enhanced. However, further increasing the length does not result in substantial performance improvements. As indicated by the lower standard deviation values in Table 4, when the sequence length is set to 24 hr, the prediction performance demonstrates the greatest stability across evaluation metrics.

4.3.2. Input-Parameter Robustness Analysis

We evaluated the model's sensitivity to input perturbations using a correlation-aware feature permutation scheme inspired by Guastavino et al. (2024). Specifically, we first computed the absolute feature–feature correlation matrix on the evaluation data and grouped strongly correlated parameters ($|r| \geq 0.8$). All parameters within each group were then jointly permuted across time steps, and this procedure was repeated 20 times. The impact of the perturbation was quantified by the mean ± standard deviation of the relative decrease in HSS (as well as Accuracy and F1) with respect to the unperturbed baseline. This grouping-based permutation reduces bias arising from collinear predictors and is designed to assess the stability of a trained model under structured feature perturbations.

Figure 6a presents the HSS values (with 95% confidence intervals) on the imbalanced test set for one representative trained model selected from our experiments. While different trained instances may exhibit varying sensitivities to specific parameter groups, the results shown here reflect the behavior of a representative model.

Overall, the model shows stable performance under correlation-aware permutation. Jointly permuting parameter groups that include, for example, the de-projected area of active pixels on the sphere (AREA_ACR), R_value, mean inclination angle (MEANGAM), mean shear angle for B_{total} (MEANSHR), and the fraction of area with shear >45° (SHRGT45) results in only modest (≈1%) performance degradation when these parameters are permuted together with their correlated counterparts.

Some non-typical predictors, such as SHRGT45, still induce measurable sensitivity under permutation in this setting. This may be related to distributional changes introduced by focusing on the polarity inversion line (PIL), which can enhance the temporal informativeness of certain parameters. We also examined the effect of removing individual input parameters during inference and observed a clear performance degradation for all parameters, consistent with the fact that these physical quantities provide complementary information.

4.3.3. Sliding Window Probability Forecast for the Entire Activity Area

To further explore the forecasting ability of the time series flare prediction in real-world tasks, we performed sliding window forecasting for all ARs in the test set throughout their entire AR period. The results were plotted and saved. AR11817, an active region in the test data set with an M-class flare, is shown in Figure 7 as an example.

We focus on the portion preceding the first strong flare eruption. In Figure 7, we define the alarm time as the point where the green probability curve first crosses 0.5 before the first strong flare eruption, and we mark the lead time. As the input duration changes from 6 to 24 hr (a–d), the lead time of successful predictions slightly decreases. This suggests that shorter input models may have a faster response to sudden signals. However, this trend no longer holds as the input duration continues to increase. As temporal features further from the flare are incorporated, the model's response to critical changes becomes more uncertain. Within the current experimental range, setting the input duration hyperparameter to 24 hr seems to achieve the most stable state for the majority of evaluation metrics when forecasting results exhibit little variation across durations of 12–48 hr, making it a favorable choice.

In fact, there were some false alarms when testing the model as shown in Figure 8.

There are two main types of situations that cause false alarms in our flare forecasting model, both involving very active ARs. The first occurs when high-intensity and high-frequency C-class flares are predicted as M/X-class flares. In these cases, the timing of the prediction is accurate, but the predicted intensity is slightly overestimated. This is understandable given the significant overlap in precursor characteristics between high-grade C-class and M-class flares, both of which are important for space weather monitoring.

The second type occurs when hyperactive ARs trigger warnings days before an actual M/X-class flare erupts. These complex regions naturally generate signals that closely resemble the precursors of imminent M/X-class flares, resulting in early warnings rather than false predictions.

To address the first situation, we can implement more detailed multi-classification to better distinguish between C and M-level flares during the training process, or adopt probabilistic forecasts to account for these borderline cases. For the second situation, developing specialized models for complex ARs would allow the system to better adapt to these particular scenarios and adjust its prediction timing accordingly.

5. Conclusion and Discussion

Our study offers a novel perspective on solar flare prediction by addressing certain limitations of previous approaches. Non-sequential flare prediction inherently suffers from limited effectiveness because it cannot utilize timing information crucial for understanding flare dynamics. From a single magnetogram or magnetic parameter snapshot, it is impossible to determine whether a magnetic rope is floating or sinking, or if an active region (AR) is becoming more complex or quieter—dynamic information that often determines flare occurrence. Additionally, we observed that focusing specifically on polarity inversion lines (PILs) rather than entire active regions could potentially provide more relevant precursor signals. While static feature extraction works well for tasks like magnetic type identification, flare prediction requires temporal evolution data and targeted spatial analysis to capture precursor signals effectively.

To address these limitations, we adopted time-series methods for flare forecasting and developed a CNN–BiLSTM–Attention (CNN–BiLSTM–AT) architecture. A rigorously constructed data set was used, incorporating information from more than one thousand active regions with strict time-series labeling, with particular emphasis on magnetic parameters along polarity inversion lines (PILs). When evaluated on real-world, highly imbalanced data, the CNN–BiLSTM–AT time-series model consistently outperformed non-sequential baselines. In particular, the recall of strong-flare events increased from approximately 35%–45% in non-sequential models to about 75% in the time-series framework, accompanied by consistent improvements in F1 score, HSS, and TSS. These results indicate that explicitly modeling temporal evolution enables the model to better capture precursor signatures of strong flares.

Building on this time-series framework, prediction performance was further improved by focusing on magnetic parameters computed along PILs rather than using the original SHARP parameters averaged over entire active regions. The PIL-focused parameters provide a more localized and physically relevant representation of magnetic complexity. Compared with AR-wide SHARP parameters, the use of PIL-based parameters increases the F1 score

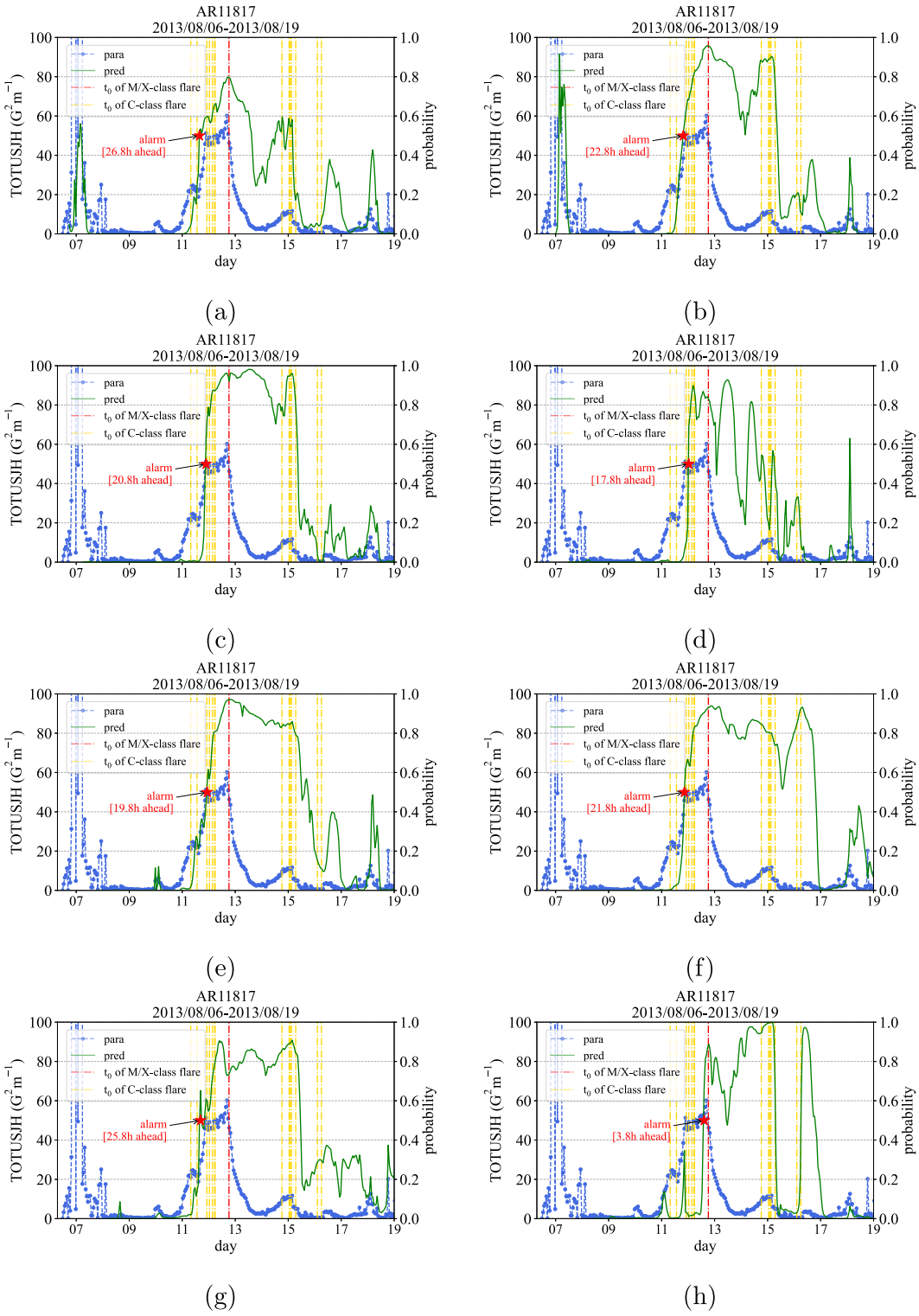

Figure 7. This figure shows the actual results of the sliding window prediction throughout the AR duration of AR11817. Figures (a–h) show the prediction results when the sequence duration is 6, 12, 18, 24, 30, 36, 42, and 48 hr respectively. The royal blue line shows one of the Space-weather HMI Active Region Patches parameters (TOTUSJH here), while the green line shows the strong flare probability output by the model. The red dotted line marks the time point when M/X-class flares erupt, and the yellow dotted line marks the time point when C-class flares erupt. We preset to trigger an alarm when the probability rises to 0.5.

10.1029/2025SW004687

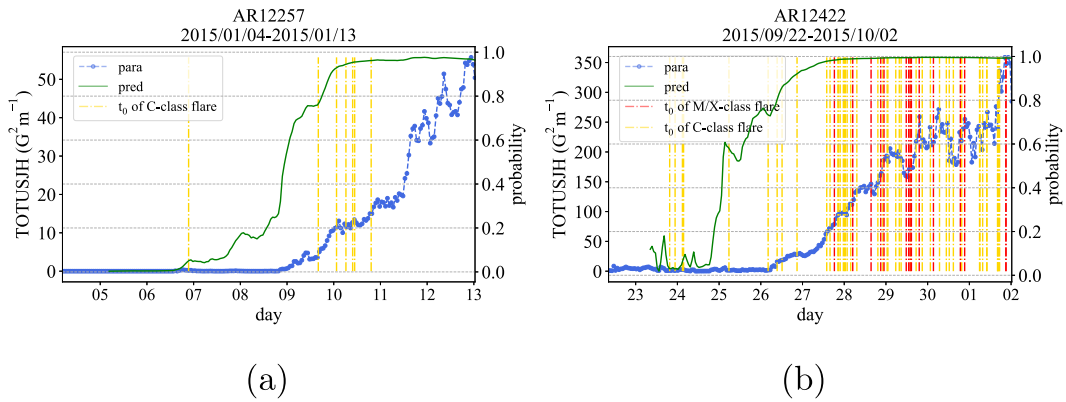

Figure 8. This figure shows two types of false alarms: (a) No M/X-class flares occur in the AR, but the model issue an alert for an impending C-class flare. (b) Hyperactive ARs set off alarms of M-class flares long before they occur.

for strong-flare prediction by approximately 27%–62% in non-sequential models and by about 34% in the time-series model. Correspondingly, HSS improves by roughly 9%–44% for non-sequential approaches and by about 41% for the time-series model. These results demonstrate that magnetic activity along polarity inversion lines contains more discriminative information for strong-flare forecasting than region-averaged magnetic properties.

In addition, an ablation study was conducted to further validate the architectural design choices of the proposed time-series framework. By systematically removing or replacing individual components of the CNN–BiLSTM–AT model, we confirmed that the observed performance gains do not arise from a single module, but rather from the complementary interaction between feature projection, temporal dependency modeling, and adaptive feature reweighting. While all temporal variants maintained high recall for strong flares, the full model consistently achieved superior precision and skill scores, indicating a more balanced and reliable forecasting behavior. These findings provide structural support for the model design adopted in this study and reinforce the importance of jointly modeling temporal evolution and localized magnetic features along polarity inversion lines.

The temporal dimension of our models required careful optimization to balance prediction accuracy with computational efficiency. By comparing models with different sequence durations, we discovered that increasing the input time series length from 6 to 12 hr significantly enhances forecasting performance. Interestingly, extending to 24 hr provides the greatest stability across most evaluation metrics, while further extensions to 48 hr yield diminishing returns without substantial improvements in prediction performance. This finding suggests an optimal observation window for capturing meaningful flare precursors.

Our robustness analysis further validates the model's reliability, showing high resilience to feature perturbations. When we artificially shuffled key parameters such as AREA_ACR, R_value, MEANGAM, and MEANSHR, the impact on performance remained minimal (<2%). This resilience indicates that our sequential model architecture effectively captures complex relationships between features without over-relying on any single parameter. Notably, some parameters that showed lower correlation with flares in previous studies demonstrated unexpected importance in our model when calculated along the PIL region, suggesting these features exhibit unique temporal characteristics when focused on this magnetically significant area.

The practical utility of our approach is demonstrated through comprehensive AR sliding window predictions, which provide intuitive visualization of model performance across different prediction horizons. This method allows us to observe how models with different sequence lengths respond to evolving magnetic conditions and identify distinctive temporal patterns. For instance, in the 6–24 hr prediction window, models using shorter input sequences prove more sensitive to sudden signal changes, potentially providing more timely predictions before strong flares, etc. Through complete forecast curves for all ARs in our test set, we identified two types of false alarms that reveal current model limitations and point toward future improvements.

Overall, our research demonstrates that 24-hr solar flare prediction using time-series models achieves favorable outcomes with practical operational potential. The sliding window probability forecasting system we developed can be readily applied to operational solar flare eruption predictions. By providing detailed documentation of our

data set creation and model training processes, we hope to establish methodological standards that will benefit future research in this domain, ultimately improving space weather forecasting capabilities and our understanding of solar flare dynamics.

Conflict of Interest

The authors declare no conflicts of interest relevant to this study.

Data Availability Statement

All data, code, and software used in this study are publicly available. The SHARP parameter data can be accessed from JSOC (http://jsoc.stanford.edu/ajax/lookdata.html), and the flare list is obtained from the SWPC Solar and Geophysical Event Reports (). The processed data sets, model training data, source code, and partial visualizations of the results generated for this study are publicly archived Zenodo (Zara-Siwei, 2025). The development version of the codebase is also hosted on GitHub for user convenience: https://github.com/Zara-Siwei/PIL_TIME_SERIES. However, the Zenodo repository should be cited for reproducibility and long-term access.

Acknowledgments
We thank the SDO/HMI team members for their contributions to the SDO mission. J.W. was supported by the Strategic Priority Research Program of the Chinese Academy of Sciences (Grant XDB0560000), China's Space Origins Exploration Program (Grant GJ11020405), the National Natural Science Foundation of China (Grant 42074224), the Pandeng Program of the National Space Science Center, CAS, and the Youth Innovation Promotion Association, CAS. R.E. is grateful to the Science and Technology Facilities Council (STFC, Grant ST/M000826/1), UK, acknowledges NKFIH (OTKA, Grant K142987 and Excellence Grant, Grant TKP2021-NKTA-64), Hungary, and PIFI (China, Grant 2024PVA0043) for enabling this research. T.W. was supported by the National Natural Science Foundation of China (Grant 42104157) and the Xingdian Youth Project of Yunnan. This work was also supported by the International Space Science Institute—Beijing (ISSI-BJ, ID 24-604) project on "Small-scale eruptions in the Sun."

References

Barnes, G., Leka, K., Schrijver, C., Colak, T., Qahwaji, R., Ashamari, O., et al. (2016). A comparison of flare forecasting methods. I. Results from the "all-clear" workshop. *The Astrophysical Journal*, *829*(2), 89. https://doi.org/10.3847/0004-637x/829/2/89

Bobra, M. G., & Couvidat, S. (2015). Solar flare prediction using sdo/hmi vector magnetic field data with a machine-learning algorithm. *The Astrophysical Journal*, *798*(2), 135. https://doi.org/10.1088/0004-637x/798/2/135

Bobra, M. G., Sun, X., Hoeksema, J. T., Turmon, M., Liu, Y., Hayashi, K., et al. (2014). The helioseismic and magnetic imager (Hmi) vector magnetic field pipeline: Sharps–space-weather hmi active region patches. *Solar Physics*, *289*(9), 3549–3578. https://doi.org/10.1007/s11207-014-0529-3

Campi, C., Benvenuto, F., Massone, A. M., Bloomfield, D. S., Georgoulis, M. K., & Piana, M. (2019). Feature ranking of active region source properties in solar flare forecasting and the uncompromised stochasticity of flare occurrence. *The Astrophysical Journal*, *883*(2), 150. https://doi.org/10.3847/1538-4357/ab3c26

Chen, Y., Manchester, W. B., Hero, A. O., Toth, G., DuFumier, B., Zhou, T., et al. (2019). Identifying solar flare precursors using time series of sdo/hmi images and sharp parameters. *Space Weather*, *17*(10), 1404–1426. https://doi.org/10.1029/2019sw002214

Cicogna, D., Berrilli, F., Calchetti, D., Del Moro, D., Giovannelli, L., Benvenuto, F., et al. (2021). Flare-forecasting algorithms based on high-gradient polarity inversion lines in active regions. *The Astrophysical Journal*, *915*(1), 38. https://doi.org/10.3847/1538-4357/abfafb

Georgoulis, M. K., Yardley, S. L., Guerra, J. A., Murray, S. A., Ahmadzadeh, A., Anastasiadis, A., et al. (2024). Prediction of solar energetic events impacting space weather conditions. *Advances in Space Research*. https://doi.org/10.1016/j.asr.2024.02.030

Guastavino, S., Bahamazava, K., Perracchione, E., Camattari, F., Audone, G., Telloni, D., et al. (2024). Forecasting geoffective events from solar wind data and evaluating the most predictive features through machine learning approaches. *The Astrophysical Journal*, *971*(1), 94. https://doi.org/10.3847/1538-4357/ad5b57

Guastavino, S., Marchetti, F., Benvenuto, F., Campi, C., & Piana, M. (2022). Implementation paradigm for supervised flare forecasting studies: A deep learning application with video data. *Astronomy & Astrophysics*, *662*, A105. https://doi.org/10.1051/0004-6361/202243617

Hill, S., Pizzo, V., Balch, C., Biesecker, D., Bornmann, P., Hildner, E., et al. (2005). The noaa goes-12 solar x-ray imager (Sxi) 1. Instrument, operations, and data. *Solar Physics*, *226*(2), 255–281. https://doi.org/10.1007/s11207-005-7416-x

Huang, X., Zhao, Z., Zhong, Y., Xu, L., Korsós, M. B., & Erdélyi, R. (2024). Short-term solar eruptive activity prediction models based on machine learning approaches: A review. *Science China Earth Sciences*, *67*(12), 3727–3764. https://doi.org/10.1007/s11430-023-1375-2

Japkowicz, N., & Stephen, S. (2002). The class imbalance problem: A systematic study. *Intelligent Data Analysis*, *6*(5), 429–449. https://doi.org/10.3233/ida-2002-6504

Kontogiannis, I., Georgoulis, M. K., Park, S.-H., & Guerra, J. A. (2017). Non-neutralized electric currents in solar active regions and flare productivity. *Solar Physics*, *292*(11), 159. https://doi.org/10.1007/s11207-017-1185-1

Korsós, M. B., Ludmány, A., Erdélyi, R., & Baranyi, T. (2015). On flare predictability based on sunspot group evolution. *The Astrophysical Journal Letters*, *802*(2), L21. https://doi.org/10.1088/2041-8205/802/2/l21

Leka, K., Barnes, G., & Wagner, E. (2018). *The nwra classification infrastructure: Description and extension to the discriminant analysis flare forecasting system (daffs)*. EDP Sciences.

Li, P., Bahri, O., Boubrahimi, S. F., & Hamdi, S. M. (2025). Evaluating time-series augmentation techniques for deep learning–based solar flare prediction. *The Astrophysical Journal Supplement Series*, *280*(2), 52. https://doi.org/10.3847/1538-4365/adfa2a

Li, Y., Huang, S., Xu, S., Yuan, Z., Jiang, K., Xiong, Q., & Lin, R. (2024). Solar flare forecasting based on a fusion model. *Earth and Planetary Physics*, *9*(1), 171–181. https://doi.org/10.26464/epp2024058

Lim, D., Moon, Y.-J., Park, J., Park, E., Lee, K., Lee, J.-Y., & Jang, S. (2019). Forecast of daily major flare probability using relationships between vector magnetic properties and flaring rates. *arXiv preprint arXiv:1907.11373*.

Liu, C., Deng, N., Wang, J. T., & Wang, H. (2017). Predicting solar flares using sdo/hmi vector magnetic data products and the random forest algorithm. *The Astrophysical Journal*, *843*(2), 104. https://doi.org/10.3847/1538-4357/aa789b

Liu, H., Liu, C., Wang, J. T., & Wang, H. (2019). Predicting solar flares using a long short-term memory network. *The Astrophysical Journal*, *877*(2), 121. https://doi.org/10.3847/1538-4357/ab1b3c

Pesnell, W. D., Thompson, B. J., & Chamberlin, P. (2012). *The solar dynamics observatory (sdo)*. Springer.

Ran, H., Liu, Y. D., Guo, Y., & Wang, R. (2022). Relationship between successive flares in the same active region and space-weather hmi active region patch (sharp) parameters. *arXiv preprint arXiv:2207.07254*.

Sun, H., Manchester, W., Jiao, Z., Wang, X., & Chen, Y. (2019). Interpreting lstm prediction on solar flare eruption with time-series clustering. *arXiv preprint arXiv:1912.12360*.

Thibeault, C., Strugarek, A., Charbonneau, P., & Tremblay, B. (2022). Forecasting solar flares by data assimilation in sandpile models. *Solar Physics*, *297*(9), 125. https://doi.org/10.1007/s11207-022-02055-9

Toriumi, S., & Wang, H. (2019). Flare-productive active regions. *Living Reviews in Solar Physics*, *16*(1), 3. https://doi.org/10.1007/s41116-019-0019-7

Vural, O., Hamdi, S. M., & Boubrahimi, S. F. (2024). Contrastive representation learning for predicting solar flares from extremely imbalanced multivariate time series data. In *2024 international conference on machine learning and applications (icmla)* (pp. 1077–1082).

Vural, O., Hamdi, S. M., & Boubrahimi, S. F. (2025). Solar flare prediction using multivariate time series of photospheric magnetic field parameters: A comparative analysis of vector, time series, and graph data representations. *Remote Sensing*, *17*(6), 1075. https://doi.org/10.3390/rs17061075

Wang, J., Liu, S., Ao, X., Zhang, Y., Wang, T., & Liu, Y. (2019). Parameters derived from the sdo/hmi vector magnetic field data: Potential to improve machine-learning-based solar flare prediction models. *The Astrophysical Journal*, *884*(2), 175. https://doi.org/10.3847/1538-4357/ab441b

Winter, L. M., & Balasubramaniam, K. (2015). Using the maximum x-ray flux ratio and x-ray background to predict solar flare class. *Space Weather*, *13*(5), 286–297. https://doi.org/10.1002/2015sw001170

Zara-Siwei. (2025). Zara-Siwei/PIL_TIME_SERIES: Pil_time_series (solar_flare_prediction) [Software]. *Zenodo*. https://doi.org/10.5281/zenodo.17717096

Zhang, J., Ye, L., & Lai, Y. (2023). Stock price prediction using cnn-bilstm-attention model. *Mathematics*, *11*(9), 1985. https://doi.org/10.3390/math11091985