Prediction of Large Solar Flares Based on SHARP and HED Magnetic Field Parameters

Xuebao Li [ID],[1] Xuefeng Li [ID],[2] Yanfang Zheng [ID],[1] Ting Li [ID],[3, 4, 5] Pengchao Yan [ID],[1] Hongwei Ye [ID],[1] Shunhuang Zhang [ID],[1] Xiaotian Wang [ID],[2] Yongshang Lv [ID],[1] and Xusheng Huang [ID][1]

[1]School of Computer Science, Jiangsu University of Science and Technology, Zhenjiang, 212100, China;
[2]School of Automation, Jiangsu University of Science and Technology, Zhenjiang, 212100, China;
[3]CAS Key Laboratory of Solar Activity, National Astronomical Observations, Chinese Academy of Sciences, Beijing 100101, China;
[4]School of Astronomy and Space Science, University of Chinese Academy of Sciences, Beijing 100049, China
[5]National Space Science Center, Chinese Academy of Sciences, Beijing 100190, China

ABSTRACT

The existing flare prediction primarily relies on photospheric magnetic field parameters from the entire active region (AR), such as Space-Weather HMI Activity Region Patches (SHARP) parameters. However, these parameters may not capture the details the AR evolution preceding flares. The magnetic structure within the core area of an AR is essential for predicting large solar flares. This paper utilizes the area of high photospheric free energy density (HED region) as a proxy for the AR core region. We construct two datasets: SHARP and HED datasets. The ARs contained in both datasets are identical. Furthermore, the start and end times for the same AR in both datasets are identical. We develop six models for 24-hour solar flare forecasting, utilizing SHARP and HED datasets. We then compare their categorical and probabilistic forecasting performance. Additionally, we conduct an analysis of parameter importance. The main results are as follows: (1) Among the six solar flare prediction models, the models using HED parameters outperform those using SHARP parameters in both categorical and probabilistic prediction, indicating the important role of the HED region in the flare initiation process. (2) The Transformer flare prediction model stands out significantly in True Skill Statistic (TSS) and Brier Skill Score (BSS), surpassing the other models. (3) In parameter importance analysis, the total photospheric free magnetic energy density (E_{free}) within the HED parameters excels in both categorical and probabilistic forecasting. Similarly, among the SHARP parameters, the R_VALUE stands out as the most effective parameter for both categorical and probabilistic forecasting.

1. INTRODUCTION

Solar flares are powerful explosive events characterized by sudden electromagnetic radiation bursts and the rapid release of high-energy particles (Fleishman et al. 2020; Sinha et al. 2022). The solar prediction is important because Ultraviolet (UV) and Soft X-ray (SXR) radiation affect the ionosphere and change the radio communication and Low Earth Orbit (LEO) drag (Qian et al. 2011; Sadykov et al. 2019; Monte-Moreno & Hernández-Pajares 2014; Nwankwo 2018). Moreover, Solar flares often initiate the chain reaction, which includes coronal mass ejections (CMEs) and solar energetic particle events (SEPs), and therefore their prediction aids the prediction of CMEs and SEPs resulting in a variety of other space weather effects (Youssef 2012; García-Rigo et al. 2016; Gou et al. 2020). Hence, establishing a

Corresponding author: Yanfang Zheng
zyf062856@163.com

Corresponding author: Ting Li
liting@nao.cas.cn

reliable prediction model for solar flares is crucial for effectively preventing or mitigating the damage caused by solar flares.

In recent years, with the accumulation of solar flare observational data and the development of machine learning, some researchers have made progress in predicting solar flares using Space-Weather HMI Activity Region Patches (SHARP; Bobra et al. 2014) data. For instance, some researchers employed multi-layer perceptrons, Support Vector Machines (SVM), and random forest algorithms to predict the occurrence or intensity of solar flares (Liu et al. 2017; Florios et al. 2018; Inceoglu et al. 2018; Reep & Barnes 2021). With the advent of the Recurrent Neural Network (RNN), other researchers utilized the Long Short-Term Memory (LSTM) network, taking into account the time dependency of solar flare prediction (Liu et al. 2019; Sun et al. 2019; Wang et al. 2020; Sun et al. 2022). As image recognition and segmentation technology matured, researchers predicted solar flares by inputting SHARP magnetic maps into a deep convolutional neural network (CNN) (Park et al. 2018; Li et al. 2020; Sun et al. 2022; Deshmukh et al. 2022). Recently, Liu et al. (2019) and Zheng et al. (2023a) utilized a combined LSTM-Attention method to enhance the performance of the neural network model for solar flare forecasting. With the continuous iteration and improvement of models, the new Transformer model has emerged as a promising approach for solar flare prediction, leveraging its ability to capture long-range dependencies (Kaneda et al. 2022; Abduallah et al. 2023; Pelkum Donahue & Inceoglu 2024; Grim & Gradvohl 2024).

Previous studies have shown that the photospheric non-potential parameters (such as the vertical current, current helicity, shear angle, and so on) are closely related to the occurrence of large solar flares (Falconer et al. 2002; Su et al. 2014; Toriumi & Wang 2019). Thus numerous photospheric magnetic field parameters are used to develop flare prediction models (Bobra & Couvidat 2015; Liu et al. 2017; Campi et al. 2019). These widely used parameters are mostly calculated within the entire active region (AR), such as the SHARP parameters. However, they may not capture the finer details preceding solar flare eruptions (Barnes et al. 2016; Toriumi & Wang 2019). Indeed, the magnetic structure in the core area of an AR is considered a crucial factor in determining whether a large solar flare would occur (Schrijver 2007; Sun et al. 2015). For instance, Sun et al. (2021) indicated that the features extracted within the polarity inversion line (PIL) region which is a sub-region of the AR exhibit greater effectiveness than the SHARP parameters for solar flare prediction. Recently, Li et al. (2024) used the area of high photospheric free energy density (HED region) as a proxy for the AR core region. They found that the combination of total photospheric free magnetic energy density E_{free} and mean unsigned current helicity H_c presents a good ability to distinguish between C-class and M/X-class flaring ARs. In this paper, we establish a database that comprises two datasets: a SHARP dataset constructed using SHARP magnetic field parameters, and an HED dataset focusing on magnetic field parameters of the HED region. Additionally, we employ five distinct deep learning algorithms (Transformer, BiLSTM-Attention, BiLSTM, LSTM-Attention, and LSTM) as well as a neural network (NN) model to develop solar flare prediction models for forecasting $\geq$ M class flares within 24 hours. We subsequently utilize the parameters from the SHARP and HED datasets to comparatively evaluate their categorical and probabilistic predictive performance. Finally, we conduct an in-depth analysis to assess the parameter importance of both SHARP and HED parameters, respectively.

The rest of this article is organized as follows. The data is described in Section 2, the method is introduced in Section 3. The results are presented in Section 4, and the conclusions and discussions are provided in Section 5.

2. DATA

The database in this paper comprises two datasets: the SHARP dataset and the HED dataset. Specifically, the SHARP dataset incorporates magnetic field parameters derived from SHARP data. Out of the 25 parameters provided in the SHARP data products (Bobra & Couvidat 2015), 10 parameters have been selected based on prior studies (Liu et al. 2017; Tang et al. 2021; Zheng et al. 2023b). These studies make a consensus that these 10 parameters are effective for flare prediction, as depicted in Table 1. The HED dataset comprises magnetic field parameters of the HED region, sourced from Li et al. (2024). There are 11 magnetic field parameters that characterize each AR from Li et al. (2024). However, some of these parameters exhibit strong correlations among themselves, reflected by exceedingly high Spearman rank-order correlation coefficients. Thus, we narrow down the selection to 6 parameters that exhibit relatively weaker correlations, as detailed in Table 2. Solar flares are classified as C, M, or X class according to the peak magnitude of the soft X-ray flux observed by the Geostationary Operational Environment Satellite (GOES). This

classification of solar flare is also applied in our work. The solar flare information is obtained from the GOES X-ray flare listings[1].

In this work, we extract data from the 24-hour window preceding the solar flare of each AR, capturing 40 data samples per AR at a frequency of 36 minutes, to ensure sufficient variability among the samples of ARs (Li et al. 2022; Zheng et al. 2023a). The data of Li et al. (2024) encompasses 323 ARs. However, due to the absence of HED regions in some ARs during this pre-flare period, some ARs may have fewer than the full complement of 40 data samples. We discard those with fewer than 30 samples to maintain the integrity and reliability of our experimental data. Therefore, in this work, we ultimately select 286 ARs, categorizing them into 189 C-class flares and 97 M/X-class flares. The methodology employed for collecting data in the SHARP dataset is aligned with that used for the HED dataset. The SHARP and HED datasets utilized in this paper maintain consistency in the ARs, with start and end times consistent for each AR. Finally, the database for this work comprises both the SHARP and HED datasets, each consisting of 189 C-class flaring ARs, 86 M-class flaring ARs, and 11 X-class flaring ARs.

In this paper, samples of M/X-class ($\geq$ M) flares in an AR are defined as the positive class, whereas C-class samples are defined as the negative class. Since the SHARP and HED parameters have distinct scales and units, they are individually normalized using mean-standard deviation before being input into the flare prediction model. There are three primary normalization strategies: global normalization, local normalization (Ahmadzadeh et al. 2021), and a method that utilizes the mean and standard deviation from the training set to normalize data on the validation and testing sets (Jiao et al. 2020). According to the research by Ahmadzadeh et al. (2021), the global normalization method demonstrates superior performance compared to the local normalization method. Therefore, we adopt the global normalization method to normalize the entire dataset. This normalization standardizes each parameter by subtracting its mean and dividing by its standard deviation, ensuring that all parameters have a mean of 0 and a standard deviation of 1, regardless of their original scales. There is no additional transformation (i.e., logarithmic) performed on the parameters. We employ an AR-based cross-validation (CV) method (Zheng et al. 2019; Li et al. 2020) to partition the SHARP and HED datasets with the distribution of the training, validation, and testing sets at a ratio of 60%, 20%, and 20%, respectively. This process is repeated ten times using 10 different random seeds, as depicted in Figure 1. This partitioning strategy guarantees that all samples from each individual AR can be arranged only into the training, validation, or testing dataset. This method also allows the model to undergo training, validation, and testing on different CV datasets, ensuring that the model is exposed to a sufficiently diverse range of AR samples. The SHARP and HED datasets are then divided into 10 distinct sets, each comprising training, validation, and testing datasets, respectively. Table 3 presents the number of solar magnetogram samples and ARs for a single SHARP/HED dataset, and the other datasets follow the similar distribution pattern. In this paper, the CV sets for the SHARP and HED datasets are precisely aligned to ensure a one-to-one correspondence in their AR distributions. In other words, the distribution of ARs in SHARP Dataset 1 is identical to that in HED Dataset 1, and this consistency extends to all other paired CV sets.

Table 1. The description and formula of SHARP parameters.

Parameters	Description	Formula
TOTUSJZ	Total unsigned vertical current	$J_{z_{total}} = \sum |J_Z| \, dA$
TOTUSJH	Total unsigned current helicity	$H_{C_{total}} \propto \sum |B_Z \cdot J_Z|$
TOTPOT	Total photospheric magnetic free energy density	$\rho_{tot} \propto \sum (B^{Obs} - B^{Pot})^2 dA$
ABSNJZH	Absolute value of the net current helicity	$H_{C_{abs}} \propto |\sum B_Z \cdot J_Z|$
SAVNCPP	Sum of the modulus of the net current per polarity	$J_{Z_{sum}} \propto \left| \sum^{B_z^+} J_Z dA \right| + \left| \sum^{B_z^-} J_Z dA \right|$
USFLUX	Total unsigned flux	$\varphi = \sum |B_Z| \, dA$
AREA_ACR	Area of strong field pixels in the active region	$Area = \sum Pixels$
MEANPOT	Mean photospheric magnetic free energy	$\bar{\rho} \propto \frac{1}{N}(B^{Obs} - B^{Pot})^2$
R_VALUE	Sum of flux near polarity inversion line	$\varphi = \sum |B_{LoS}| \, dA$ within R mask
SHRGT45	Fraction of area with shear >45°	Area with shear $> 45°$ / total_area

---

[1] https://www.ngdc.noaa.gov/stp/space-weather/solar-data/solar-features/solar-flares/x-rays/goes/xrs/

Table 2. The description and formula of HED parameters.

Parameters	Description	Formula
E_{free}	Total photospheric free energy density	$E_{free} = \sum p_{free} dA$
Ψ	Mean shear angle	$\Psi = \arccos \frac{B_{obs} \cdot B_{pot}}{\left| B_{obs} B_{pot} \right|}$
J_Z	Mean unsigned vertical current density	$J_Z = \frac{1}{\mu_0 N} \sum \left| \frac{\partial B_y}{\partial x} - \frac{\partial B_x}{\partial y} \right|$
H_C	Mean unsigned current helicity	$H_c = \frac{\mu_0}{N} \sum \left| B_Z J_Z \right|$
α	Mean characteristic twist parameter	$\alpha = \mu_0 \frac{\sum J_Z B_Z}{\sum B_Z^2}$
∇B_h	Mean horizontal gradient of horizontal field	$\nabla B_h = \frac{1}{N} \sum \sqrt{(\frac{\partial B_h}{\partial x})^2 + (\frac{\partial B_h}{\partial y})^2}$

Table 3. The number of solar magnetogram samples and ARs in one SHARP/HED dataset.

| Data set	C class sample/ AR numbers	M class sample/ AR numbers	X class sample/ AR numbers
Training set	4520/113	2120/53	240/6
Validation set	1520/38	680/17	80/2
Testing set	1520/38	640/16	120/3
Total	7560/189	3440/86	440/11

3. METHOD

In this paper, we design five solar flare prediction models leveraging five currently popular deep learning algorithms: Transformer, BiLSTM-Attention, BiLSTM, LSTM-Attention, and LSTM. The Transformer model consists of an Input layer, a patch layer, a patch encoder layer, Transformer encoder blocks, and an Output layer. Figure 2 depicts the architecture of this model. On the other hand, the LSTM series models incorporate an Input layer, an LSTM layer, and an Output layer. Notably, the BiLSTM-Attention and LSTM-Attention models incorporate the attention mechanism, and Figure 3 illustrates the architecture of the BiLSTM-Attention model. We use NN model as the baseline model to compare with other deep learning models. The model architecture of the NN model is similar to the output layer of the Transformer and LSTM series models, and uses point-in-time measurements.

3.1. *Input layer*

In solar flare prediction models, the Input layer remains consistent across all deep learning model architectures. The input sequence for these models is typically denoted as $[X_{t-T+1}, X_{t-T+2}, \ldots, X_{t-1}, X_t]$, where X_t represents the vector input captured at time t, X_{t-1} represents the vector input at time $t-1$, and T designates the time step length which is set to 40 in this paper. Specifically, the Input layer processes T samples originating from a single AR, with all samples belonging to either the positive or negative class.

3.2. *Transformer layer and LSTM series layer*

This paper is inspired by the model structure of Vision Transformer (ViT; Yuan et al. 2021), which transforms the data output from the Input layer into dimensional representations that are suitable for processing by the Transformer model. This transformation is achieved through the Patch layer and Patch Encoder layer. In this paper, we design a Transformer encoder block. Each Transformer encoder block consists of a Multi-Head Attention (MHA) layer, two Add Layers, a Batch Normalization (BN) layer, and a Multi-Layer Perceptron (MLP) layer. The MHA layer leverages multiple independent self-attention mechanisms simultaneously focusing on different parts of the input sequence, and capturing diverse dependencies (Vaswani et al. 2017). In this work, the MHA layer is configured with 2 heads, each with 200 dimensions. Add Layer 1 performs residual connections between the input and output of the MHA layer, facilitating better model training and optimization (He et al. 2016). The BN layer enhances the stability and convergence speed of neural network training while providing regularization effects that improve the generalization
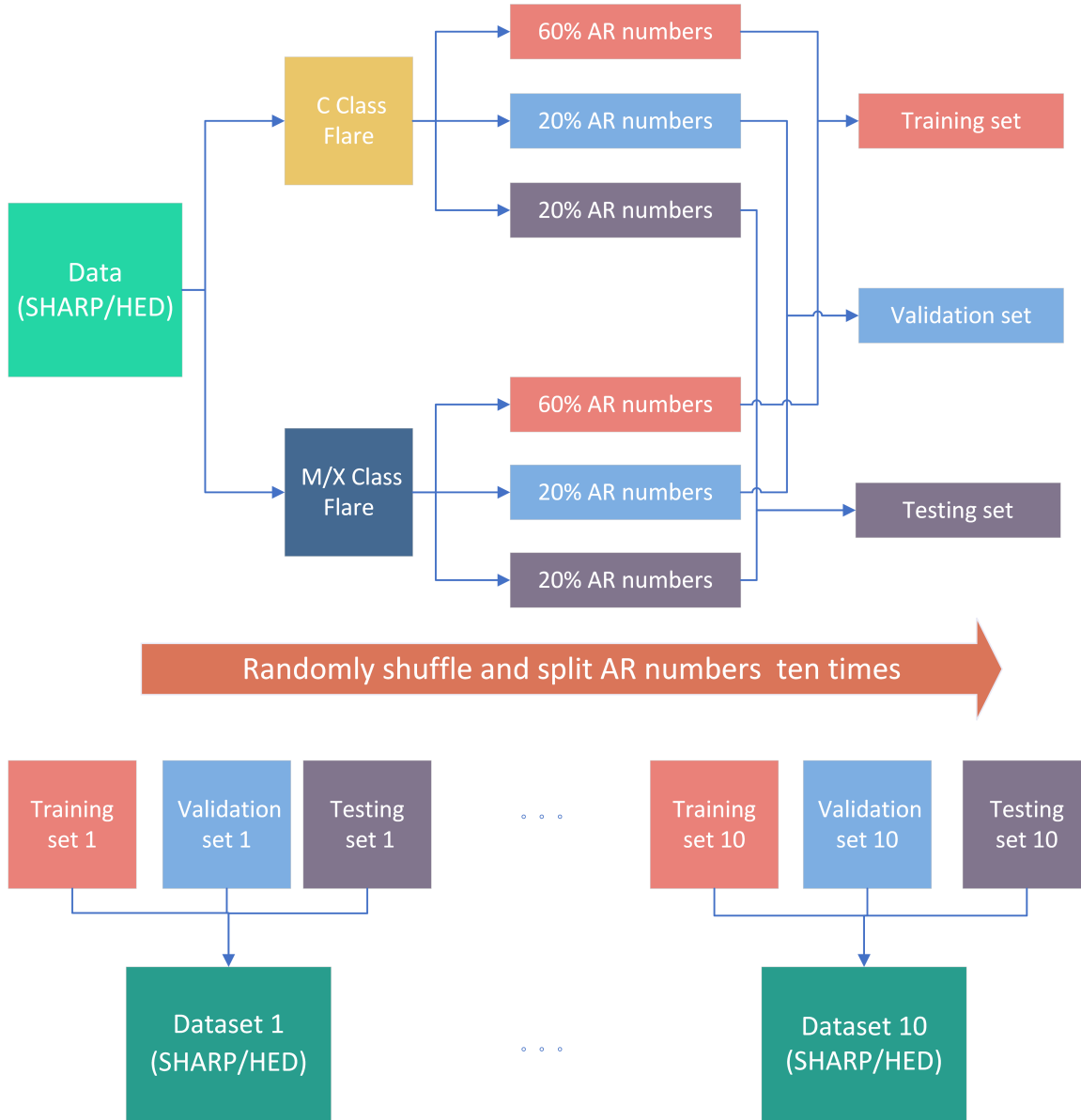
Figure 1. The Procedure for Generating 10-Fold Cross-Validation Set Divisions for the SHARP/HED Datasets.

ability of the model (Zerveas et al. 2021). The MLP layer further transforms and extracts features, resulting in a more comprehensive representation. In this study, the MLP comprises two Dense layers interspersed with two Dropout layers. Add Layer 2 also performs residual connections, connecting the input of the BN layer to the output of the MLP layer, and feeding the combined result into the next transformer encoder block. We employ a total of 4 transformer encoder blocks.

The LSTM layer can be categorized into unidirectional LSTM and bidirectional LSTM. In unidirectional LSTM, the information flows solely in a forward direction, considering only previous data in the sequence to predict the current output. Conversely, in bidirectional LSTM, the flow of information occurs both forward and backward, allowing the model to consider both preceding and following information at each time step. This allows it to produce a more informed output at the current time step. BiLSTM-Attention and LSTM-Attention incorporate an attention mechanism that dynamically assigns varying weights to different inputs of the model. This attention mechanism enables the model to prioritize and focus on the most relevant information, leading to a better understanding and processing of the input data, ultimately resulting in improved model performance.

3.3. *Output layer*

The output layer consists of three Batch Normalization (BN) layers, three Dense layers, a Flatten layer, three Dropout functions, and a Softmax activation function. The Dense layer plays a crucial role in neural networks, as it performs linear transformations on the input data, effectively mapping the original feature space into a new one. This transformation facilitates feature extraction and optimizes model training, among other functionalities. The Flatten layer, on the other hand, serves to convert multi-dimensional input data into a flattened one-dimensional representation, ultimately enhancing the utilization of the data. Lastly, the Dropout function randomly deactivates a subset of neurons during training, serving as a regularization technique that prevents overfitting and thereby improves the generalization capability of the model. The Softmax activation function generates two probability values, which sum to 1. Concisely, if the Softmax output is $[1 - \hat{y}_i, \hat{y}_i]$, then $1 - \hat{y}_i$ represents the probability of $<$ M class, while $\hat{y}_i$ represents the probability of $\geq$ M class. Ultimately, the model categorizes the sample as either a positive class or a negative class, based on the relative magnitudes of these output probabilities. The sequential arrangement of these layers and functions is illustrated in Figures 2 and 3.

3.4. *Model Parameter*

To evaluate their predictive performance, we use 10 distinct datasets from the SHARP dataset and HED dataset, respectively, for training, validation, and testing. Specifically, the training and validation sets are used for model training and validation, while the testing set is exclusively used for model evaluation. In this paper, Hyperopt is used for hyperparameter optimization, which can find the optimal combination of parameters within the range set by the experiment (Feurer & Hutter 2019; Yang & Shami 2020). The hyperparameter search space and optimal values for categorical prediction using the Transformer model are presented in Table 4. Each model is trained using 150 epochs. The models in this experiment are trained using the Adam optimizer (Kingma & Ba 2014), and the learning rate is obtained through hyperparameter optimization.

To mitigate the issue of class imbalance in the dataset, we utilize a weighted binary cross-entropy (WBCE) loss function during the model training. By introducing the weight coefficient, WBCE gives different weights to positive and negative samples, so that the loss of minority class accounts for a larger proportion of the total loss, thus guiding the model to pay more attention to minority class samples. This function calculates the loss by considering the weighted binary cross-entropy, which measures the difference between the model predictions and the true labels in each iteration. This guides the training process in the correct direction. The loss function is typically expressed as equations 1 and 2, where ω_i represents the sample class weights, and N is the number of training samples in each batch size. y_i is the actual label for the i_{th} sample (usually 0 or 1 for binary classification). $\hat{y}_i$ represents the probability of $\geq$ M class by the model for the i_{th} sample. n_{count_j} is the count of samples for class j. $n_{classes}$ is the total number of classes. n_{sample} is the number of total training samples.

$$\text{WBCE} = \sum_{i=1}^{N} (\omega_0 y_i \log \hat{y}_i + \omega_1 (1 - y_i) \log (1 - \hat{y}_i)), \tag{1}$$

$$\omega_j = \frac{n_{sample}}{n_{classes} \times n_{count_j}} (j = 0, 1). \tag{2}$$

4. RESULTS

In this paper, we take solar flare forecasting as a binary classification task and evaluate the model performance in categorical and probabilistic forecasting. We leverage a confusion matrix to assess the categorical performance of the model. Specifically, samples accurately classified as the positive class are defined as true positives (TP). Conversely, samples accurately classified as the negative class are defined as true negatives (TN). On the flip side, samples erroneously classified as the positive class are defined as false positives (FP). Similarly, samples erroneously classified as the negative class are defined as false negatives (FN). These four quantities constitute the confusion matrix, which serves as the basis for calculating metrics such as Precision, Recall, Accuracy, False Alarm Ratio (FAR), Heidke Skill Score (HSS), and True Skill Statistics (TSS), as detailed in Table 5. We use Random chance forecast and Negative forecast, as reference points to compare the performance of the categorical prediction. In the Random chance forecast, TP$=\frac{P}{P+N} \times P$, where P represents the number of positive samples and N represents the number of negative samples in the testing set. FN$=P$-TP, TN$=\frac{N}{P+N} \times N$, and FP$=N$-TN. In the Negative forecast, TP and FP

Table 4. The hyperparameter search space and optimal values.

Parameter	Search space	Optimal values	
		SHARP parameters	HED parameters
Number of Transformer Blocks	[1, 2, 3, 4]	4	4
Number of Heads in MHA	[1, 2, 3, 4]	2	2
Learning Rate	[0.0001 ∼ 0.01]	0.0020	0.0001
MLP Dropout Rate	[0.1 ∼ 0.5]	0.3200	0.3000
Dropout Rate 1	[0.1 ∼ 0.5]	0.2000	0.3990
Dropout Rate 2	[0.1 ∼ 0.5]	0.1000	0.1001
Dropout Rate 3	[0.1 ∼ 0.5]	0.1000	0.1000
Batch Size	[5, 10, 15, 20]	10	10
Epochs	[100, 120, 150, 200]	150	150

are both equal to 0, FN is equal to P, and TN is equal to N. By substituting these values for the testing set into the confusion matrix, the corresponding categorical statistic score is obtained. Given that most of these metrics, except for TSS, are susceptible to imbalances in the experimental sample data (Bobra & Couvidat 2015; Bloomfield et al. 2012), we designate TSS as the primary evaluation metric in the categorical performance.

Table 5. Categorical statistic score computed from the confusion matrix.

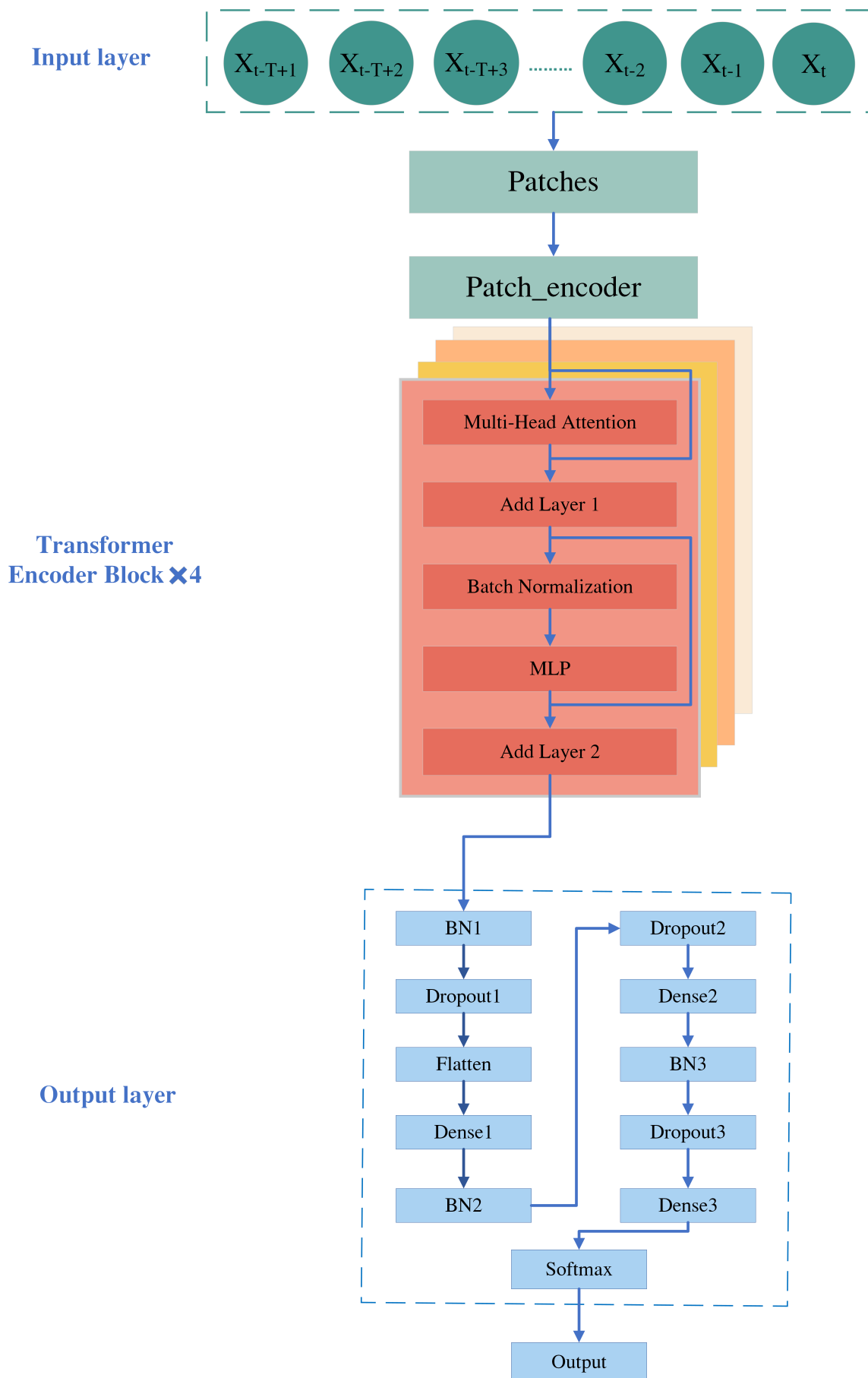| Metrics	Definition	Scale	Perfect	Random chance forecast	Negative forecast
Precision	$\frac{TP}{TP+FP}$	$0 \sim 1$	1	0.333	/
Recall	$\frac{TP}{TP+FN}$	$0 \sim 1$	1	0.333	0
Accuracy	$\frac{TP+TN}{TP+FP+TN+FN}$	$0 \sim 1$	1	0.556	0.667
FAR	$\frac{FP}{TP+FP}$	$0 \sim 1$	0	0.667	/
HSS	$\frac{2[(TP\times TN)-(FN\times FP)]}{(TP+FN)(FN+TN)+(TP+FP)(FP+TN)}$	$-\infty \sim 1$	1	0	0
TSS	$\frac{TP}{TP+FN} - \frac{FP}{FP+TN}$	$-1 \sim 1$	1	0	0

In addition, we utilize the Brier Score (BS; Brier 1950) and the Brier Skill Score (BSS; Wilks 2010) as quantitative metrics to evaluate the probabilistic forecasting performance of our models. The formulations for BS and BSS are presented in equations 3 and 4, respectively. Specifically, N denotes the total number of data samples in the testing dataset. y_i represents the target label, while $\hat{y}_i$ is the predicted probability of the $\geq$ M class for the i_{th} testing sample. $\bar{y} = \frac{1}{N}\sum_{i=1}^{N} y_i$ indicates the average target label value for the $\geq$ M class across all testing samples. BS ranges from 0 to 1, with a score of 0 representing the best probabilistic forecasting performance. Conversely, BSS ranges from $-\infty$ to 1, with a value of 1 indicating the optimal probabilistic forecasting performance.

$$BS = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2, \tag{3}$$

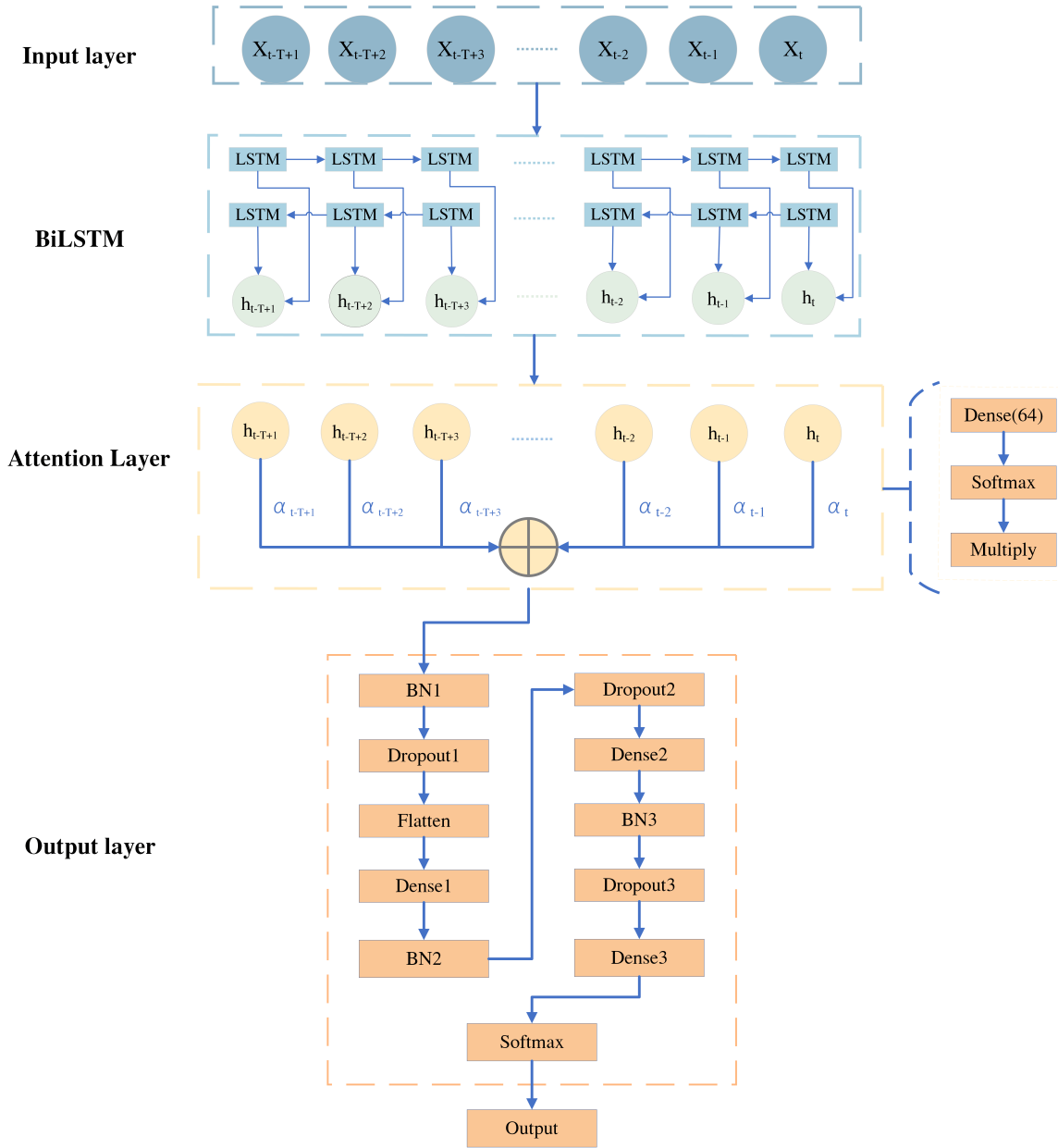$$BSS = 1 - \frac{BS}{\frac{1}{N}\sum_{i=1}^{N}(y_i - \bar{y})^2}. \tag{4}$$

4.1. Model performance evaluation

To ensure that the model does not underfit or overfit the data, we take the Transformer model using HED parameters as an example. In the process of model training for categorical prediction, we monitor the TSS on the validation set for each epoch and save the model that corresponds to the optimal TSS value achieved on the validation set. Similarly, during the process of model training for probability prediction, we track the BSS on the validation set per epoch, and retain the model corresponding to the best BSS value obtained.

Figure 2. The architecture of transformer model.

Input layer

X_{t-T+1} X_{t-T+2} X_{t-T+3} ········· X_{t-2} X_{t-1} X_t

BiLSTM

LSTM → LSTM → LSTM ········· LSTM → LSTM → LSTM

LSTM ← LSTM ← LSTM ········· LSTM ← LSTM ← LSTM

h_{t-T+1} h_{t-T+2} h_{t-T+3} ········· h_{t-2} h_{t-1} h_t

Attention Layer

h_{t-T+1} h_{t-T+2} h_{t-T+3} ········· h_{t-2} h_{t-1} h_t

α_{t-T+1} α_{t-T+2} α_{t-T+3} α_{t-2} α_{t-1} α_t

Dense(64)

Softmax

Multiply

Output layer

BN1 → Dropout2

Dropout1 → Dense2

Flatten → BN3

Dense1 → Dropout3

BN2 → Dense3

Softmax

Output

Figure 3. The architecture of BiLSTM-Attention model.

As the model architectures and hyperparameters utilized for both categorical forecasting and probability forecasting are identical, the only distinction is the specific epochs at which the optimal models are saved. Therefore, the training loss curves for both the categorical forecasting model and the probability forecasting model are consistent, and the validation loss curves between them are also the same, as shown in Figure 4. Although the training loss curves and the validation loss curves exhibit some fluctuations, they are finally convergent, thus indicating no underfitting situation. Notably, despite the observed overfitting in the validation loss curves 9 and 10 in Figure 4, the model we ultimately saved does not exhibit overfitting. In particular, we take the training loss curve 9 as an example. During the training process for categorical forecasting, although the validation loss curve indicates overfitting between epochs 95 and 145, the model is saved at epoch 150 corresponding to the optimal TSS value achieved on the validation set. In this epoch, the validation loss of the model is still very low, indicating that overfitting has disappeared. Similarly, in the training for probability forecasting, the model that holds the optimal BSS values at epoch 84 does not overfit, essentially

equivalent to an earlier termination of the training process. Hence, by monitoring the optimal TSS and BSS values on the validation set, we ensure that the models saved from this experiment exhibit neither underfitting nor overfitting.

To assess the influence of the quantity of unique flares on model training, we conduct an experiment leveraging the Transformer model with HED parameters. Specifically, we randomly reduce the number of ARs within the training set while maintaining the number of ARs in both the validation and testing sets unchanged, and then retrain the model. By observing the changes of TSS in categorical prediction and BSS in probability prediction, we can verify whether the quantity of ARs in the training set is sufficient for training the Transformer model. Figure 5 illustrates the mean and standard deviation of TSS and BSS during the reduction of AR numbers. As depicted in Figure 5(A), a slight fluctuation in the TSS score is observed as the number of ARs in the training set decreases in the range of 0 to 15. However, as the number of ARs in the training set reduces by more than 15, there is a slight decline in TSS score. As shown in Figure 5(B), as the number of ARs in the training set decreases, the variation in BSS remains minimal. These results indicate that the number of ARs in the training dataset is sufficient to support the training of the Transformer model. Subsequently, we evaluate the performance of six solar flare prediction models in predicting solar flares of $\geq$ M class on ten testing datasets to determine whether SHARP parameters or HED parameters are more effective in distinguishing between C-class and M/X-class flares.

Table 6 presents the categorical results of six models for binary-class predictions within a 24-hour window. The mean TSS score of the six solar flare prediction models using SHARP parameters ranges from 0.376 to 0.559, with the Transformer model achieving the best performance, surpassing other models with a score of 0.559 $\pm$ 0.087. On the other hand, when using HED parameters, the mean TSS score of the six solar flare prediction models ranges from 0.643 to 0.721. The Transformer model again outperforms the others, scoring 0.721 $\pm$ 0.084. The BiLSTM model performs the worst with a score of 0.643 $\pm$ 0.074, which is still much better than the best-performing Transformer model using SHARP parameters. We observe that the Transformer model performed the best in both SHARP parameters and HED parameters. In addition, in each of the six models, the categorical forecasting performance using HED parameters is superior to that using SHARP parameters.

Table 7 shows the probabilistic forecasting results of six models. Among the six solar flare prediction models using SHARP parameters, the Transformer model exhibits the best performance in probabilistic forecasting, achieving a BSS score of 0.304 $\pm$ 0.092, while the other models performed comparably with a mean BSS score ranging from 0.080 to 0.161. Meanwhile, the Transformer model stands out in the six solar flare prediction models using HED parameters, achieving a BSS score of 0.522 $\pm$ 0.060, significantly outperforming the other models. The other models also performed comparably, with a mean BSS score ranging from 0.282 to 0.446, which is still almost superior to the probabilistic results using SHARP parameters. It is observed that, when using the same type of parameters, the Transformer model consistently showed superior performance in probabilistic forecasting compared to other models. Furthermore, for the same model, the probabilistic forecasting results using HED parameters are significantly better than those using SHARP parameters.

Overall, the advanced deep learning models, which enable the extraction of intricate relationships among AR samples, can obtain better categorical and probability forecast performance than the NN model, yielding superior forecasting results for solar flares. Moreover, the Transformer model consistently outperforms other models in both categorical and probabilistic predictions, regardless of whether SHARP parameters or HED parameters are used. We infer that this may be because the LSTM model processes time series through recursion, but its ability to handle long-range dependencies is limited by its recursive structure. Although the bidirectional LSTM model is capable of capturing contextual relationships through both forward and backward information, it is still limited to capturing local contextual relationships (Schmidhuber et al. 1997; Sherstinsky 2020). However, the Transformer model, endowed with its unique self-attention mechanism, allows direct access to the relationship between any two positions within the sequence, thereby enabling it to learn comprehensive global contextual information. As a result, the Transformer model excels at capturing not only long-range dependencies but also intricate relationships and patterns that exist between different positions within a sequence (Vaswani et al. 2017). Therefore, the Transformer-type architecture exhibits exceptional performance. Furthermore, the performance of all models utilizing HED parameters surpasses that of models using SHARP parameters in both categorical and probabilistic prediction results.

4.2. *Parameter importance analysis*

In this paper, given that the SHARP parameters and HED parameters performed optimally in the Transformer model, we utilize the Transformer model and apply Recursive Feature Elimination (RFE; Abhale & Avulapalli 2023)

to conduct separate parameter importance analyses for SHARP and HED parameters, respectively. The study by Bobra & Ilonidis (2016) found that combining certain lower-ranked parameters could enhance model performance. Therefore, we investigate the effects of various parameter combinations on model performance by employing the RFE method. The steps of RFE are as follows. (1) In parameter importance analysis, we solely focus on the TSS score of categorical prediction performance and the BSS score of probabilistic prediction performance for each parameter. Specifically, we train, validate, and test the Transformer model on univariate datasets to evaluate categorical and probabilistic forecast performance, respectively. Therefore, the TSS and BSS scores for the univariate prediction are obtained. Then, we rank the TSS score for each parameter in descending order, and repeat this process for BSS score as well. The TSS and BSS scores for the individual SHARP and HED parameters are ranked in descending order, as shown in Figures 6 and 9, respectively. (2) We sort individual parameters according to their TSS and BSS scores in descending and ascending order, respectively. The descending order is arranged from the highest score (most important) to the lowest score (least important), whereas the ascending order follows the opposite direction. Subsequently, we change the dimension of input parameters sequentially, following descending and ascending order. Taking the categorical forecast performance of the SHARP parameters as an example, in the process of sequentially removing the most important parameters, as shown in Figure 7(A), the R_VALUE on the X-axis represents the initial set of 10 SHARP parameters, SAVNCPP signifies the set with R_VALUE excluded, ABSNJZH represents the set with both R_VALUE and SAVNCPP removed, and so forth. Conversely, in the process of sequentially removing the least important parameters, as shown in Figure 7(B), AREA_ACR on the X-axis represents the original set of 10 SHARP parameters, SHRGT45 denotes the set with AREA_ACR omitted, MEANPOT indicates the set further excluding SHRGT45, and so on. (3) We train, validate, and test the model under different combinations of SHARP and HED parameters. As elucidated in the study conducted by Liu et al. (2020), an exhaustive search for the optimal parameter subset among n parameters would require evaluating all $2^n - 1$ combinations. In light of the extensive parameter space in our study, which comprises 10 SHARP parameters resulting in 1023 combinations and 6 HED parameters yielding 63 combinations, an exhaustive search for the optimal parameter subset becomes computationally intractable. Therefore, we do not consider all possible combinations but rather employ a more targeted RFE approach.

Figure 6 shows a histogram that depicts the mean and standard deviation of TSS and BSS for the single SHARP parameter in $\geq$ M class flare prediction. As depicted in Figure 6, when ranking the parameter importance of individual parameters within the SHARP parameters, the top three most significant parameters for both TSS and BSS scores are consistently R_VALUE, SAVNCPP, and ABSNJZH. Figure 7 shows a line chart displaying the mean and standard deviation of TSS as each SHARP parameter is sequentially reduced in the flare prediction of $\geq$ M class flares. As shown in Figure 7(A), with the removal of the top-ranked parameter, R_VALUE, the TSS score experiences a notable decline. Meanwhile, as illustrated in Figure 7(B), after removing the six least significant parameters, the categorical forecasting performance of the model with the remaining parameter combination remains high. Figure 8 shows a line chart displaying the mean and standard deviation of BSS as each SHARP parameter is sequentially reduced in the flare prediction of $\geq$ M class flares. As observed in Figure 8(A), the BSS score experiences a significant drop when parameters are eliminated down to the less important sets, such as SHRGT45, USFLUX, and AREA_ACR. However, as depicted in Figure 8(B), when removing the least significant parameters from the SHARP parameters sequentially, the BSS curve exhibits minimal fluctuation. This indicates that the categorical forecasting performance and the probabilistic forecasting performance of the parameter R_VALUE are both excellent and play a pivotal role in predicting $\geq$ M class flares, which aligns with previous findings by Schrijver (2007), Welsch et al. (2009), and Liu et al. (2017).

Figure 9 shows a histogram that depicts the mean and standard deviation of TSS and BSS for the single HED parameter in $\geq$ M class flare prediction. As shown in Figure 9, in the ranking of parameter importance among individual parameters in HED parameters, E_{free} and H_C emerge as the top two parameters based on their TSS and BSS scores. Figure 10 shows a line chart displaying the mean and standard deviation of TSS as each HED parameter is sequentially reduced in the prediction of $\geq$ M class flares. In Figure 10(A), a noticeable decline in the TSS score is observed with the removal of E_{free} and Ψ from the parameters, respectively. Conversely, as depicted in Figure 10(B), the TSS curve remains relatively stable when the least significant HED parameters are eliminated sequentially. Figure 11 shows a line chart displaying the mean and standard deviation of BSS as each HED parameter is sequentially reduced in the prediction of $\geq$ M class flares. Similarly, in Figure 11(A), the BSS score undergoes a marked drop with the exclusion of E_{free}. However, in Figure 11(B), the BSS curve remains largely unaffected when the least important HED parameters are removed sequentially. This indicates that E_{free} plays a crucial role in $\geq$ M class flare prediction

and performs well in both categorical forecasting and probabilistic forecasting, aligning with the conclusion drawn by Li et al. (2024).

In general, our analysis shows that R_VALUE is the top performer among SHARP parameters for flare prediction of $\geq$ M class flares, while E_{free} significantly outperforms the other HED parameters. Furthermore, we notice a tendency that the Transformer model, when using parameters with higher TSS scores in categorical prediction, also tends to achieve higher BSS scores in probabilistic prediction.
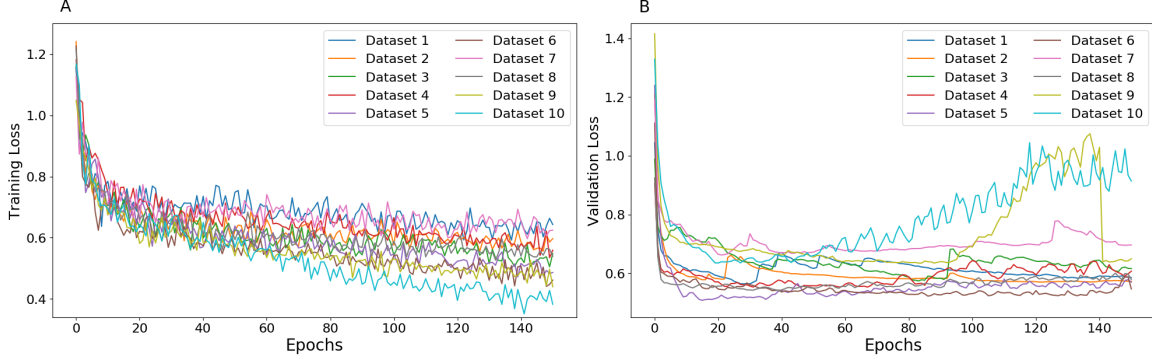
Figure 4. The figure depicts the training loss curves and validation loss curves of a Transformer model using HED parameters in $\geq$ M class flare prediction. Panel (A) signifies the training loss curves, while panel (B) denotes the validation loss curves. The 10 different colored curves represent the changes of training and validation loss with epochs for the model trained and validated by 10 separate datasets.
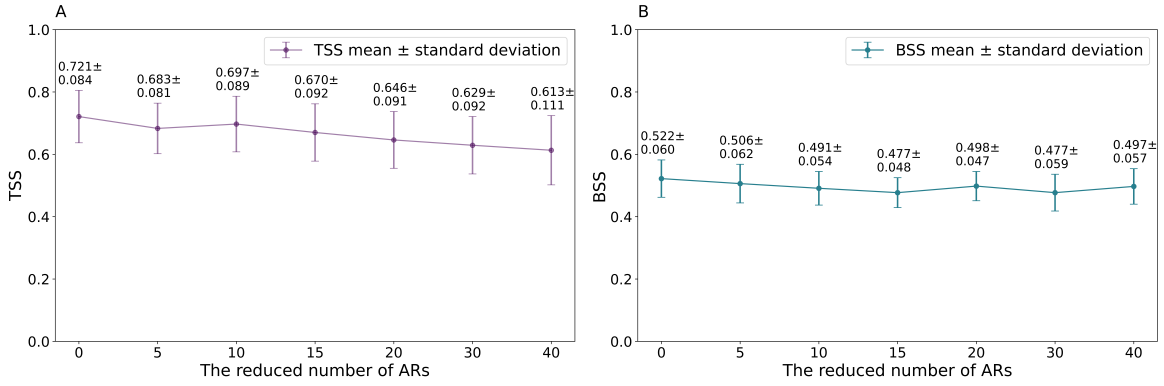
Figure 5. The line chart illustrates the mean and standard deviation of TSS and BSS during the reduction of AR numbers in predicting $\geq$ M-class flares. Panel (A) represents the mean and standard deviation of TSS as the number of AR gradually decreases. Panel (B) shows the mean and standard deviation of BSS as the number of AR gradually decreases. The X-axis coordinates represent the reduced number of ARs.

5. CONCLUSIONS AND DISCUSSIONS

In this paper, we first establish a database comprising two datasets: the SHARP dataset and the HED dataset. The ARs contained in both datasets are identical. Furthermore, the start and end times for the same AR are identical in both datasets. Subsequently, we adopt an AR-based CV method to divide the SHARP dataset and the HED dataset into training, validation, and testing datasets, while maintaining consistency in the AR distribution between the datasets. We develop six models for forecasting solar flares within 24 hours, using both SHARP parameters and HED parameters. We subsequently perform a comparative analysis of these models, evaluating both their categorical and probabilistic forecasting performance. Ultimately, we select the Transformer model, which exhibits superior performance for both SHARP and HED parameters, to conduct a parameter importance analysis. By employing RFE to prune parameters, we aim to uncover commonly significant parameters that influence categorical forecasting performance (TSS) and probabilistic forecasting performance (BSS) for SHARP and HED parameters, respectively.

Table 6. The categorical results of the six models for binary-class predictions within 24 h. The boldface represents the best value in each column. The number before the ± represents the mean, while the number after the ± represents the standard deviation.

Metric	Model	Results	
		SHARP parameters	HED parameters
TSS	Transformer	**0.559 ± 0.087**	**0.721 ± 0.084**
	BiLSTM-Attention	0.541 ± 0.127	0.692 ± 0.075
	BiLSTM	0.503 ± 0.135	0.643 ± 0.074
	LSTM-Attention	0.510 ± 0.135	0.705 ± 0.076
	LSTM	0.521 ± 0.118	0.686 ± 0.065
	NN	0.376 ± 0.093	0.666 ± 0.063
HSS	Transformer	0.501 ± 0.091	**0.709 ± 0.084**
	BiLSTM-Attention	**0.504 ± 0.121**	0.655 ± 0.081
	BiLSTM	0.450 ± 0.125	0.591 ± 0.091
	LSTM-Attention	0.457 ± 0.117	0.680 ± 0.087
	LSTM	0.468 ± 0.107	0.663 ± 0.058
	NN	0.380 ± 0.089	0.662 ± 0.061
FAR	Transformer	0.592 ± 0.075	**0.210 ± 0.074**
	BiLSTM-Attention	**0.379 ± 0.086**	0.282 ± 0.076
	BiLSTM	0.437 ± 0.074	0.340 ± 0.094
	LSTM-Attention	0.429 ± 0.064	0.248 ± 0.096
	LSTM	0.425 ± 0.063	0.261 ± 0.051
	NN	0.598 ± 0.058	0.231 ± 0.056
Accuracy	Transformer	0.751 ± 0.054	**0.869 ± 0.038**
	BiLSTM-Attention	**0.763 ± 0.064**	0.838 ± 0.047
	BiLSTM	0.728 ± 0.065	0.801 ± 0.054
	LSTM-Attention	0.735 ± 0.053	0.852 ± 0.045
	LSTM	0.739 ± 0.054	0.846 ± 0.027
	NN	0.730 ± 0.040	0.850 ± 0.028
Precision	Transformer	0.408 ± 0.076	**0.790 ± 0.074**
	BiLSTM-Attention	**0.621 ± 0.086**	0.718 ± 0.076
	BiLSTM	0.563 ± 0.074	0.659 ± 0.094
	LSTM-Attention	0.571 ± 0.064	0.752 ± 0.096
	LSTM	0.575 ± 0.063	0.739 ± 0.051
	NN	0.402 ± 0.058	0.769 ± 0.056
Recall	Transformer	**0.865 ± 0.089**	0.834 ± 0.080
	BiLSTM-Attention	0.791 ± 0.135	0.871 ± 0.077
	BiLSTM	0.821 ± 0.105	**0.882 ± 0.071**
	LSTM-Attention	0.813 ± 0.165	0.855 ± 0.075
	LSTM	0.827 ± 0.121	0.837 ± 0.074
	NN	0.568 ± 0.105	0.784 ± 0.060

The main results of this paper are summarized as follows: (1) In categorical prediction, among the six solar flare forecasting models, the Transformer model performs the best. The Transformer model achieves a TSS of 0.559 ± 0.087 using SHARP parameters, and a TSS of 0.721 ± 0.084 using HED parameters. Additionally, all solar flare forecasting models using HED parameters outperform those using SHARP parameters in terms of categorical

Table 7. The probabilistic results of the six models for binary-class predictions within 24 h. The boldface represents the best value in each column. The number before the $\pm$ represents the mean, while the number after the $\pm$ represents the standard deviation.

Metric	Model	Results	
		SHARP parameters	HED parameters
BSS	Transformer	**0.304 $\pm$ 0.092**	**0.522 $\pm$ 0.060**
	BiLSTM-Attention	0.161 $\pm$ 0.089	0.446 $\pm$ 0.086
	BiLSTM	0.152 $\pm$ 0.124	0.442 $\pm$ 0.090
	LSTM-Attention	0.120 $\pm$ 0.069	0.385 $\pm$ 0.126
	LSTM	0.142 $\pm$ 0.133	0.400 $\pm$ 0.115
	NN	0.080 $\pm$ 0.078	0.282 $\pm$ 0.075
BS	Transformer	**0.154 $\pm$ 0.020**	**0.105 $\pm$ 0.014**
	BiLSTM-Attention	0.185 $\pm$ 0.020	0.122 $\pm$ 0.019
	BiLSTM	0.187 $\pm$ 0.028	0.123 $\pm$ 0.020
	LSTM-Attention	0.194 $\pm$ 0.016	0.136 $\pm$ 0.028
	LSTM	0.189 $\pm$ 0.030	0.132 $\pm$ 0.025
	NN	0.203 $\pm$ 0.018	0.158 $\pm$ 0.016
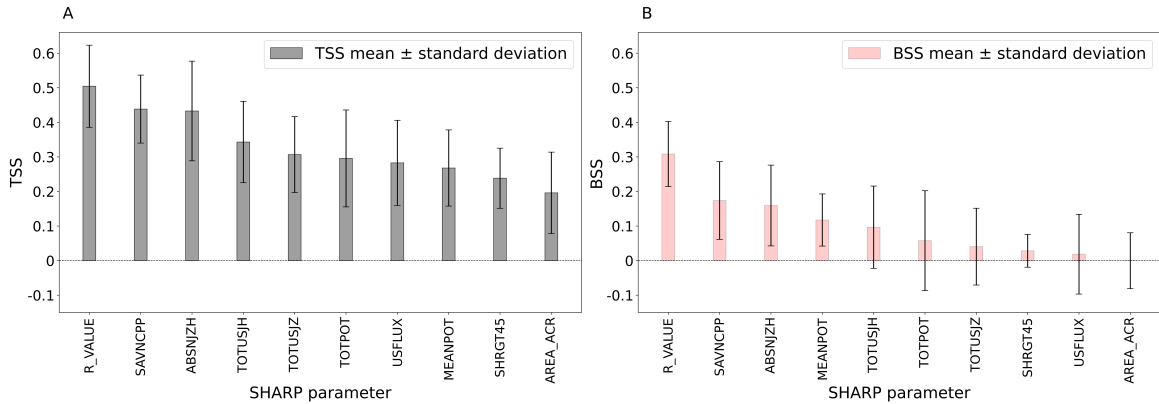
Figure 6. The histogram depicts the mean and standard deviation of TSS and BSS for each individual SHARP parameter in $\geq$ M class flare prediction. Specifically, Panel (A) signifies the mean and standard deviation of TSS, while panel (B) denotes the mean and standard deviation of BSS, both of which are sorted in descending order. The 'bar' denotes the standard deviation.

forecasting performance. (2) In probabilistic prediction, the Transformer model stands out as the best among the six solar flare forecasting models. It achieves a BSS of 0.304 ± 0.092 when using SHARP parameters and a noticeably higher BSS of 0.522 ± 0.060 with HED parameters. Notably, solar flare forecasting models that employ HED parameters outperform those that utilize SHARP parameters in probabilistic forecasting performance. (3) In the analysis of parameter importance utilizing SHARP parameters, R_VALUE, SAVNCPP, and ABSNJZH emerge as the top three parameters in both TSS and BSS scores, while AREA_ACR consistently lags. Notably, R_VALUE performs the best in both categorical and probabilistic forecasting, highlighting its crucial role in predicting $\geq$ M class flares. (4) In the analysis of parameter importance based on HED parameters, E_{free} and H_C stand out as the top two in both TSS and BSS scores, with E_{free} topping the list. This indicates that E_{free} possesses a strong ability to differentiate between C-class and M/X-class flares.

In our study, the R_VALUE stands out as the best performer among SHARP parameters due to its close correlation with solar flares. The large flares frequently occur near strong and highly sheared PILs (Sadykov & Kosovichev 2017; Vasantharaju et al. 2018; Dhakal & Zhang 2023), and the R_VALUE focuses on the regions near the PIL. SAVNCPP and ABSNJZH perform better than TOTUSJH and TOTUSJZ, which implies that the net current and current helicity
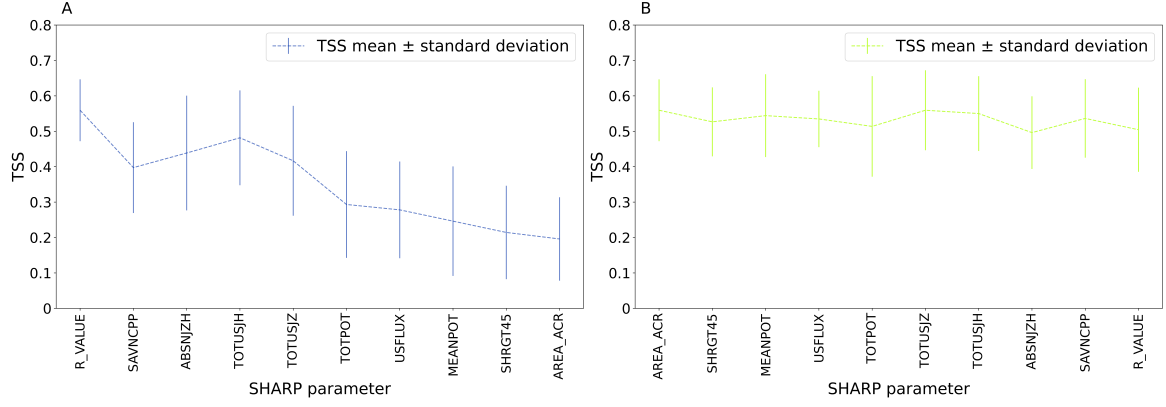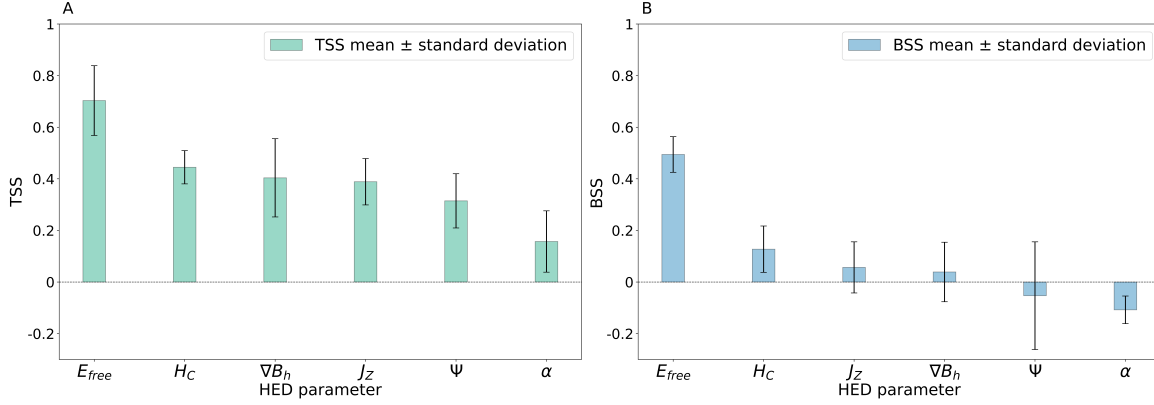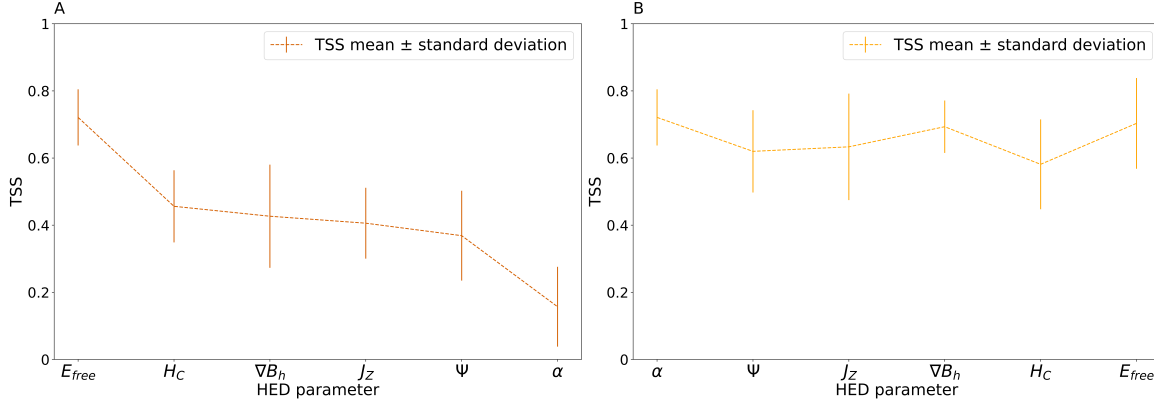
Figure 7. The line chart illustrates the mean and standard deviation of TSS during the sequential reduction of SHARP parameters in predicting $\geq$ M-class flares. Panel (A) represents the mean and standard deviation of TSS as the most important SHARP parameters are sequentially removed, starting from the most important. Panel (B) shows the mean and standard deviation of TSS as the least important SHARP parameters are sequentially removed, starting from the least important. The X-axis coordinates represent the combination of a parameter with all parameters to its right.
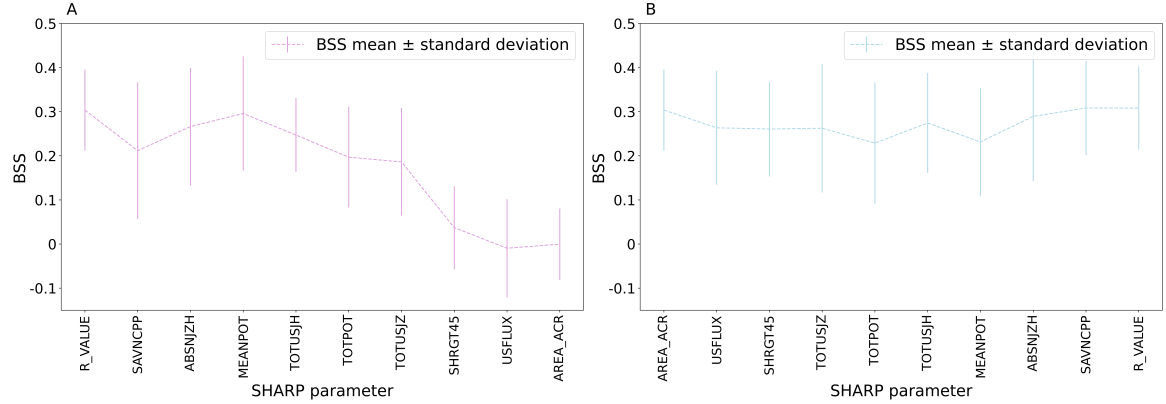
Figure 8. The line chart illustrates the mean and standard deviation of BSS during the sequential reduction of SHARP parameters in predicting $\geq$ M-class flares. Panel (A) represents the mean and standard deviation of BSS as the most important SHARP parameters are sequentially removed, starting from the most important. Panel (B) shows the mean and standard deviation of BSS as the least important SHARP parameters are sequentially removed, starting from the least important. The X-axis coordinates represent the combination of a parameter with all parameters to its right.

react sharper to the solar flares than the total quantities. Previous studies have shown that flare/CME-productive ARs are more likely to exhibit non-neutralized currents than flare/CME-quiet ARs (Avallone & Sun 2020; Liu et al. 2024). Our result suggests that the net current/current helicity in ARs is more closely related to their eruptive potential than the total current/current helicity (Dalmasse et al. 2015; Wang et al. 2023), which probably implies the presence of a magnetic flux rope with net currents before the flare. Among the HED parameters, E_{free} exhibits the optimal performance. In previous studies, the free magnetic energy stored in an AR is a relatively good parameter that is closely related to large flares (Emslie et al. 2012; Liokati et al. 2022; Xu et al. 2022). The HED regions almost coincide with the PILs with a steep horizontal magnetic gradient (see Figure 1 in Li et al. (2024)). Several previous studies have shown that the PIL characteristics are more effective than global characteristics of ARs (Sadykov & Kosovichev 2017; Wang et al. 2019; Sun et al. 2021). Moreover, the E_{free} and H_{C} parameters within HED region perform better than TOTPOT and TOTUSJH calculated within the entire active region. This suggests that the HED parameters that do not scale with the active region size (such as average quantities or quantities relevant to a specific source region) present a better ability to distinguish between C-class and M/X-class flaring ARs than the SHARP parameters that scale with the size of the active region. In total, our results show that the forecasting based on HED parameters demonstrates higher scores than SHARP parameters, which are consistent with previous studies.

Figure 9. The histogram depicts the mean and standard deviation of TSS and BSS for each individual HED parameter in ≥ M class flare prediction. Specifically, panel (A) signifies the mean and standard deviation of TSS, while panel (B) denotes the mean and standard deviation of BSS, both of which are sorted in descending order. The 'bar' denotes the standard deviation.
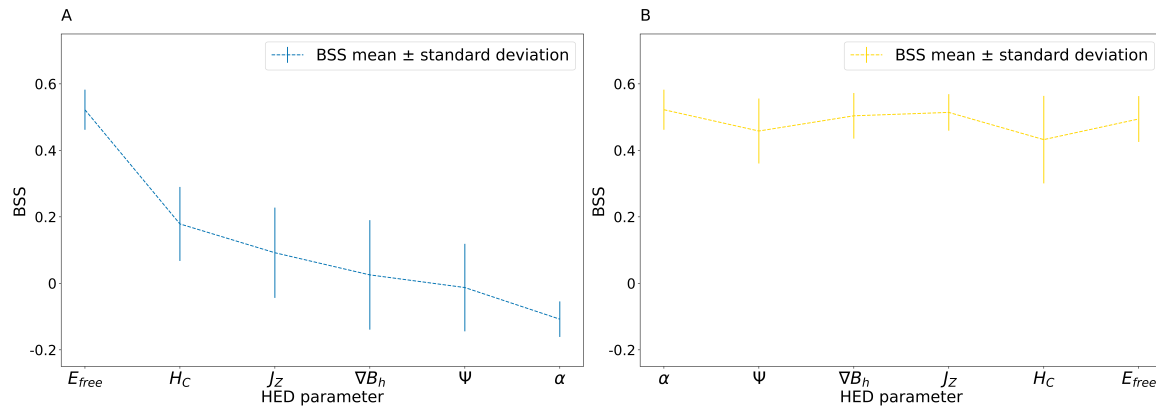

Figure 10. The line chart illustrates the mean and standard deviation of TSS during the sequential reduction of HED parameters in predicting ≥ M-class flares. Panel (A) represents the mean and standard deviation of TSS as the most important HED parameters are sequentially removed, starting from the most important. Panel (B) shows the mean and standard deviation of TSS as the least important HED parameters are sequentially removed, starting from the least important. The X-axis coordinates represent the combination of a parameter with all parameters to its right.

Solar flares significantly affect the production and daily life on Earth. Therefore, accurately predicting their timing and intensity is crucial for the planet's operations. In this paper, we explore the impact of magnetic field characteristics extracted from various regions of the same AR and assess how different models influence the performance of solar flare predictions. Our findings will contribute to the improvement of future research regarding forecasting solar flares. In our future research, we aim to incorporate a broader range of photospheric magnetic field parameters into solar flare predictions. We plan to utilize the data provided by the European Space Agency's (ESA; Akioka et al. 2005; Lu et al. 2019) Vigil L5 mission, as well as Chinese telescopes like the Advanced Space-based Solar Observatory (ASO-S; Gan et al. 2023) and the Chinese H Solar Explorer (CHASE; Li et al. 2019; Cheng & Chuan 2022) satellite. Our ultimate objective is to enhance the accuracy in forecasting the occurrence time and intensity of solar flares. Furthermore, we are dedicated to developing a real-time prediction platform that leverages the latest observational data and advanced deep-learning algorithms to achieve rapid and accurate predictions of solar flares, contributing to global space weather forecasting.

Figure 11. The line chart illustrates the mean and standard deviation of BSS during the sequential reduction of HED parameters in predicting $\geq$ M-class flares. Panel (A) represents the mean and standard deviation of BSS as the most important HED parameters are sequentially removed, starting from the most important. Panel (B) shows the mean and standard deviation of BSS as the least important HED parameters are sequentially removed, starting from the least important. The X-axis coordinates represent the combination of a parameter with all parameters to its right.

DATA AVAILABILITY

The data underlying this paper are available via https://doi.org/10.5281/zenodo.13734431.

REFERENCES

Abduallah, Y., Wang, J. T., Wang, H., & Xu, Y. 2023, Scientific reports, 13, 13665, doi: 10.1038/s41598-023-40884-1

Abhale, A. B., & Avulapalli, J. R. 2023, International Journal of System Assurance Engineering and Management, 14, 2642, doi: 10.1007/s13198-023-02128-3

Ahmadzadeh, A., Aydin, B., Georgoulis, M. K., et al. 2021, The Astrophysical Journal Supplement Series, 254, 23, doi: 10.3847/1538-4365/abec88

Akioka, M., Nagatsuma, T., Miyake, W., Ohtaka, K., & Marubashi, K. 2005, Advances in Space Research, 35, 65, doi: 10.1016/j.asr.2004.09.014

Avallone, E. A., & Sun, X. 2020, The Astrophysical Journal, 893, 123, doi: 10.3847/1538-4357/ab7afa

Barnes, G., Leka, K., Schrijver, C., et al. 2016, The Astrophysical Journal, 829, 89, doi: 10.3847/0004-637x/829/2/89

Bloomfield, D. S., Higgins, P. A., McAteer, R. J., & Gallagher, P. T. 2012, The Astrophysical Journal Letters, 747, L41, doi: 10.1088/2041-8205/747/2/l41

Bobra, M. G., & Couvidat, S. 2015, The Astrophysical Journal, 798, 135, doi: 10.1088/0004-637x/798/2/135

Bobra, M. G., & Ilonidis, S. 2016, The Astrophysical Journal, 821, 127, doi: 10.3847/0004-637X/821/2/127

Bobra, M. G., Sun, X., Hoeksema, J. T., et al. 2014, Solar Physics, 289, 3549, doi: 10.1007/s11207-014-0529-3

Brier, G. W. 1950, Monthly weather review, 78, 1, doi: 10.1175/1520-0493(1950)078⟨0001:vofeit⟩2.0.co;2

Campi, C., Benvenuto, F., Massone, A. M., et al. 2019, The Astrophysical Journal, 883, 150, doi: 10.3847/1538-4357/ab3c26

Cheng, F., & Chuan, L. 2022, Chinese Journal of Space Science, 42, 546, doi: 10.11728/cjss2022.04.yg05

Dalmasse, K., Aulanier, G., Démoulin, P., et al. 2015, The Astrophysical Journal, 810, 17, doi: 10.1088/0004-637X/810/1/17

Deshmukh, V., Flyer, N., Van der Sande, K., & Berger, T. 2022, The Astrophysical Journal Supplement Series, 260, 9, doi: 10.3847/1538-4365/ac5b0c

Dhakal, S. K., & Zhang, J. 2023, The Astrophysical Journal, 960, 36, doi: 10.3847/1538-4357/ad07d2

Emslie, A., Dennis, B., Shih, A., et al. 2012, The Astrophysical Journal, 759, 71, doi: 10.1088/0004-637X/759/1/71

Falconer, D., Moore, R., & Gary, G. 2002, The Astrophysical Journal, 569, 1016, doi: 10.1086/339161

18

Feurer, M., & Hutter, F. 2019, Automated machine learning: Methods, systems, challenges, 3, doi: 10.1007/978-3-030-05318-5

Fleishman, G. D., Gary, D. E., Chen, B., et al. 2020, Science, 367, 278, doi: 10.1126/science.aax6874

Florios, K., Kontogiannis, I., Park, S.-H., et al. 2018, Solar Physics, 293, 28, doi: 10.1007/s11207-018-1250-4

Gan, W., Zhu, C., Deng, Y., et al. 2023, Solar Physics, 298, 68, doi: 10.1007/s11207-023-02166-x

García-Rigo, A., Núñez, M., Qahwaji, R., et al. 2016, Prediction and warning system of SEP events and solar flares for risk estimation in space launch operations, EDP Sciences, doi: 10.1051/swsc/2016021

Gou, T., Veronig, A. M., Liu, R., et al. 2020, The Astrophysical Journal Letters, 897, L36, doi: 10.3847/2041-8213/ab9ec5

Grim, L. F. L., & Gradvohl, A. L. S. 2024, Solar Physics, 299, 33, doi: 10.1007/s11207-024-02276-0

He, K., Zhang, X., Ren, S., & Sun, J. 2016, in Proceedings of the IEEE conference on computer vision and pattern recognition, 770–778, doi: 10.1109/cvpr.2016.90

Inceoglu, F., Jeppesen, J. H., Kongstad, P., et al. 2018, The Astrophysical Journal, 861, 128, doi: 10.3847/1538-4357/aac81e

Jiao, Z., Sun, H., Wang, X., et al. 2020, Space weather, 18, e2020SW002440, doi: 10.1029/2020SW002440

Kaneda, K., Wada, Y., Iida, T., et al. 2022, in Proceedings of the Asian Conference on Computer Vision, 1488–1503, doi: 10.1007/978-3-031-26284-5_27

Kingma, D. P., & Ba, J. 2014, arXiv preprint arXiv:1412.6980, doi: 10.48550/arXiv.1412.6980

Li, C., Fang, C., Li, Z., et al. 2019, Research in Astronomy and Astrophysics, 19, 165, doi: 10.1088/1674-4527/19/11/165

Li, M., Cui, Y., Luo, B., et al. 2022, Space weather, 20, e2021SW002985, doi: 10.1029/2021SW002985

Li, T., Zheng, Y., Li, X., et al. 2024, The Astrophysical Journal, 964, doi: 10.3847/1538-4357/ad2e90

Li, X., Zheng, Y., Wang, X., & Wang, L. 2020, The Astrophysical Journal, 891, 10, doi: 10.3847/1538-4357/ab6d04

Liokati, E., Nindos, A., & Liu, Y. 2022, Astronomy & Astrophysics, 662, A6, doi: 10.1051/0004-6361/202142868

Liu, C., Deng, N., Wang, J. T., & Wang, H. 2017, The Astrophysical Journal, 843, 104, doi: 10.3847/1538-4357/aa789b

Liu, H., Liu, C., Wang, J. T., & Wang, H. 2019, The Astrophysical Journal, 877, 121, doi: 10.3847/1538-4357/ab1b3c

Liu, H., Liu, C., Wang, J. T., & Wang, H. 2020, The Astrophysical Journal, 890, 12, doi: 10.3847/1538-4357/ab6850

Liu, Y., Török, T., Titov, V., et al. 2024, The Astrophysical Journal, 961, 148, doi: 10.3847/1538-4357/ad11da

Lu, L., Yu, Q.-L., Wang, D., & Lu, Q. 2019, in Conference of the Arabian Journal of Geosciences, Springer, 231–234, doi: 10.1007/978-3-030-72896-0_51

Monte-Moreno, E., & Hernández-Pajares, M. 2014, Journal of Geophysical Research: Space Physics, 119, 9216, doi: 10.1002/2014JA020206

Nwankwo, V. U. 2018, in Exploring the Universe: From Near Space to Extra-Galactic: A Collection of Research Reviews on Contemporary Astrophysics and Space Science, Springer, 637–644, doi: 10.1007/978-3-319-94607-8_49

Park, E., Moon, Y.-J., Shin, S., et al. 2018, The Astrophysical Journal, 869, 91, doi: 10.3847/1538-4357/aaed40

Pelkum Donahue, K., & Inceoglu, F. 2024, Frontiers in Astronomy and Space Sciences, 10, 1298609, doi: 10.3389/fspas.2023.1298609

Qian, L., Burns, A. G., Chamberlin, P. C., & Solomon, S. C. 2011, Journal of Geophysical Research: Space Physics, 116, doi: 10.1029/2011JA016777

Reep, J. W., & Barnes, W. T. 2021, Space Weather, 19, e2021SW002754, doi: 10.1029/2021SW002754

Sadykov, V. M., & Kosovichev, A. G. 2017, The Astrophysical Journal, 849, 148, doi: 10.3847/1538-4357/aa9119

Sadykov, V. M., Kosovichev, A. G., Kitiashvili, I. N., & Frolov, A. 2019, The Astrophysical Journal, 874, 19, doi: 10.3847/1538-4357/ab06c3

Schmidhuber, J., Hochreiter, S., et al. 1997, Neural Comput, 9, 1735, doi: 10.1162/neco.1997.9.8.1735

Schrijver, C. J. 2007, The Astrophysical Journal, 655, L117, doi: 10.1086/511857

Sherstinsky, A. 2020, Physica D: Nonlinear Phenomena, 404, 132306, doi: 10.1016/j.physd.2019.132306

Sinha, S., Gupta, O., Singh, V., et al. 2022, The Astrophysical Journal, 935, 45, doi: 10.3847/1538-4357/ac7955

Su, J., Jing, J., Wang, S., Wiegelmann, T., & Wang, H. 2014, The Astrophysical Journal, 788, 150, doi: 10.1088/0004-637X/788/2/150

Sun, H., Manchester, W., Jiao, Z., Wang, X., & Chen, Y. 2019, arXiv preprint arXiv:1912.12360, doi: 10.48550/arXiv.1912.12360

Sun, H., Manchester IV, W., & Chen, Y. 2021, Space weather, 19, e2021SW002837, doi: 10.1029/2021SW002837

Sun, X., Bobra, M. G., Hoeksema, J. T., et al. 2015, The Astrophysical Journal Letters, 804, L28, doi: 10.1088/2041-8205/804/2/L28

Sun, Z., Bobra, M. G., Wang, X., et al. 2022, The Astrophysical Journal, 931, 163, doi: 10.3847/1538-4357/ac64a6

Tang, R., Liao, W., Chen, Z., et al. 2021, The Astrophysical Journal Supplement Series, 257, 50, doi: 10.3847/1538-4365/ac249e

Toriumi, S., & Wang, H. 2019, Living Reviews in Solar Physics, 16, 3, doi: 10.1007/s41116-019-0019-7

Vasantharaju, N., Vemareddy, P., Ravindra, B., & Doddamani, V. 2018, The Astrophysical Journal, 860, 58, doi: 10.3847/1538-4357/aac272

Vaswani, A., Shazeer, N., Parmar, N., et al. 2017, Advances in neural information processing systems, 30, doi: 10.48550/arXiv.1706.03762

Wang, J., Liu, S., Ao, X., et al. 2019, The Astrophysical Journal, 884, 175, doi: 10.3847/1538-4357/ab441b

Wang, W., Qiu, J., Liu, R., et al. 2023, The Astrophysical Journal, 943, 80, doi: 10.3847/1538-4357/aca6e1

Wang, X., Chen, Y., Toth, G., et al. 2020, The Astrophysical Journal, 895, 3, doi: 10.3847/1538-4357/ab89ac

Welsch, B. T., Li, Y., Schuck, P. W., & Fisher, G. H. 2009, The Astrophysical Journal, 705, 821, doi: 10.1088/0004-637X/705/1/821

Wilks, D. S. 2010, Quarterly Journal of the Royal Meteorological Society, 136, 2109, doi: 10.1002/qj.709

Xu, Z., Yan, X., Yang, L., et al. 2022, The Astrophysical Journal Letters, 937, L11, doi: 10.3847/2041-8213/ac8fef

Yang, L., & Shami, A. 2020, Neurocomputing, 415, 295, doi: 10.1016/j.neucom.2020.07.061

Youssef, M. 2012, NRIAG Journal of Astronomy and Geophysics, 1, 172, doi: 10.1016/j.nrjag.2012.12.014

Yuan, L., Chen, Y., Wang, T., et al. 2021, in Proceedings of the IEEE/CVF international conference on computer vision, 558–567, doi: 10.1109/iccv48922.2021.00060

Zerveas, G., Jayaraman, S., Patel, D., Bhamidipaty, A., & Eickhoff, C. 2021, in Proceedings of the 27th ACM SIGKDD conference on knowledge discovery & data mining, 2114–2124, doi: 10.1145/3447548.3467401

Zheng, Y., Li, X., & Wang, X. 2019, The Astrophysical Journal, 885, 73, doi: 10.3847/1538-4357/ab46bd

Zheng, Y., Li, X., Yan, S., et al. 2023a, Monthly Notices of the Royal Astronomical Society, 521, 5384, doi: 10.1093/mnras/stad839

Zheng, Y., Qin, W., Li, X., et al. 2023b, Astrophysics and Space Science, 368, 53, doi: 10.1007/s10509-023-04209-y