The Astrophysical Journal, 998:340 (10pp), 2026 February 20 https://doi.org/10.3847/1538-4357/ae3c83

Predicting Associations between Solar Flares and Coronal Mass Ejections Using SDO/HMI Magnetograms and a Hybrid Neural Network

Jialiang Li[1,2], Vasyl Yurchyshyn[3], Jason T. L. Wang[1,2], Haimin Wang[1,3,4], Manolis K. Georgoulis[5,6], Wen He[1,4], Yasser Abduallah[1,2], Hameedullah A. Farooki[7], and Yan Xu[1,3,4]
[1] Institute for Space Weather Sciences, New Jersey Institute of Technology, University Heights, Newark, NJ 07102, USA
[2] Department of Computer Science, New Jersey Institute of Technology, University Heights, Newark, NJ 07102, USA
[3] Big Bear Solar Observatory, New Jersey Institute of Technology, 40386, North Shore Lane, Big Bear City, CA 92314, USA; vasyl.yurchyshyn@njit.edu
[4] Center for Solar-Terrestrial Research, New Jersey Institute of Technology, University Heights, Newark, NJ 07102, USA
[5] Johns Hopkins University Applied Physics Laboratory, 11100 Johns Hopkins Road, Laurel, MD 20723, USA
[6] Research Center for Astronomy and Applied Mathematics, Academy of Athens, 11527, Athens, Greece (on leave), Greece
[7] Department of Astrophysical Sciences, Princeton University, Princeton, NJ 08544, USA
Received 2025 August 29; revised 2026 January 4; accepted 2026 January 20; published 2026 February 19

Abstract

Solar eruptions, including flares and coronal mass ejections (CMEs), have a significant impact on Earth. Some flares are associated with CMEs, and some flares are not. The association between flares and CMEs is not always obvious. In this study, we propose a new deep learning method, specifically a hybrid neural network (HNN) that combines a vision transformer with long short-term memory, to predict associations between flares and CMEs. HNN finds spatio-temporal patterns in the time series of line-of-sight magnetograms of solar active regions collected by the Helioseismic and Magnetic Imager on board the Solar Dynamics Observatory and uses the patterns to predict whether a flare projected to occur within the next 24 hr will be eruptive (i.e., CME-associated) or confined (i.e., not CME-associated). Our experimental results demonstrate the good performance of the HNN method. Furthermore, the results show that magnetic flux cancellation in polarity inversion line regions may well play a role in triggering flare-associated CMEs, a finding consistent with the literature.

Unified Astronomy Thesaurus concepts: Solar coronal mass ejections (310); Solar flares (1496); Solar activity (1475); Solar active regions (1974)

1. Introduction

Solar flares are sudden and large explosions in the Sun. Some flares have a close relationship with coronal mass ejections (CMEs), which are observed in white-light coronagraph images (S. Yashiro et al. 2006; K. A. Alobaid et al. 2023). Considering the occurrence of CMEs, there are two types of flares (R. L. Moore et al. 2001). One type is eruptive flares, which are typically associated with CMEs (Y. Wang & J. Zhang 2007; A. N. Afanasyev et al. 2023; M. D. Kazachenko 2023). The other type is noneruptive or confined flares, which are not associated with any CMEs (J. K. Thalmann et al. 2015; K. Teraoka et al. 2025).

CMEs can release billions of tons of charged particles into space at high speeds (J. Lin & T. G. Forbes 2000; D. F. Webb & T. A. Howard 2012; K. A. Alobaid et al. 2023). They have a profound impact on the near-Earth space environment and on Earth, with potential consequences ranging from technological disturbances to life-threatening scenarios (D. N. Baker et al. 2004). As a result, significant efforts have been made to leverage new technologies for the early detection and forecasting of flares (M. G. Bobra & S. Couvidat 2015; C. Liu et al. 2017; K. Florios et al. 2018; X. Huang et al. 2018; E. Jonas et al. 2018; N. Nishizuka et al. 2018; H. Liu et al. 2019; X. Li et al. 2020; X. Wang et al. 2020; P. Sun et al. 2022; H. Zhang et al. 2022; Y. Abduallah et al. 2023; J. Wen et al. 2025; D. Xu et al. 2025) and CMEs (R. Qahwaji et al. 2008; M. G. Bobra & S. Ilonidis 2016; F. Inceoglu et al. 2018; H. Liu et al. 2020; H. Zhang et al. 2025). A recent review on the prediction of these solar energetic events can be found in M. K. Georgoulis et al. (2024). These new technologies are developed to better prepare and to mitigate the impacts of solar eruptions on Earth's technological systems.

Recently, machine learning (E. Alpaydin 2016, and references therein) has drawn significant interest in the field. This technology is able to learn from historical data to make predictions on new, unseen data. For instance, researchers employed machine learning algorithms and vector magnetic data products from the Helioseismic and Magnetic Imager (HMI; J. Schou et al. 2012) on board the Solar Dynamics Observatory (SDO; W. D. Pesnell et al. 2012) to predict flares within the next 24 hr(M. G. Bobra & S. Couvidat 2015; C. Liu et al. 2017). Others used similar machine learning algorithms and data to predict CMEs (M. G. Bobra & S. Ilonidis 2016; F. Inceoglu et al. 2018). More recently, deep learning (Y. LeCun et al. 2015), a subfield of machine learning, has emerged as a powerful tool to predict solar eruptions. A suite of deep learning methods has been developed, ranging from recurrent neural networks, including long short-term memory and gated recurrent units, to convolutional neural networks (J. J. Hopfield 1982; S. Hochreiter & J. Schmidhuber 1997; Y. LeCun et al. 2015), to predict eruptive events (H. Liu et al. 2019; X. Li et al. 2020; H. Liu et al. 2020; P. Sun et al. 2022; Y. Abduallah et al. 2023; D. Xu et al. 2025).

It is now widely believed that flares and CMEs are manifestations of solar coronal relaxation in terms of excess magnetic energy or magnetic helicity (B. C. Low 1994;

R. A. Harrison 1995; D. M. Rust & A. Kumar 1996; P. F. Chen 2011; S. Berkebile-Stoiser et al. 2012). However, the association between them is not straightforward based on current observations (S. Yashiro & N. Gopalswamy 2008; Y. Kawabata et al. 2018). Many efforts have been made to understand the role of the coronal magnetic field in the initialization, evolution, and eruption mechanisms in confined and eruptive flares (T. Török & B. Kliem 2005; C. R. DeVore & S. K. Antiochos 2008; Y. Liu 2008; C. Baumgartner et al. 2018). In order to improve the understanding of these physical mechanisms, research has also been conducted for CME predictions; e.g., some studies used physical parameters derived from photospheric vector magnetograms to predict the occurrence of CMEs (M. G. Bobra & S. Ilonidis 2016; F. Inceoglu et al. 2018; H. Liu et al. 2020; H. Zhang et al. 2025).

In this work, we attempt to further understand the association between flares and CMEs by harnessing the power of a vision transformer (ViT) and long short-term memory (LSTM) with SDO/HMI line-of-sight (LOS) magnetograms. ViT (A. Dosovitskiy et al. 2021) is an effective deep learning method for image classification, while LSTM (H. Liu et al. 2019) works well for time series forecasting. We develop a hybrid neural network, named HNN, that combines ViT and LSTM to predict associations between flares and CMEs. HNN finds spatiotemporal patterns in the time series of SDO/HMI LOS magnetograms of solar active regions (ARs), and uses the patterns to predict whether a flare projected to occur within the next 24 hr will be eruptive (i.e., CME-associated) or confined (i.e., not CME associated). The use of a time series of HMI magnetogram images of ARs, processed through the well-trained HNN model, enables us to leverage the model's inherent strengths for pattern recognition and spatiotemporal data analysis.

The remainder of this paper is organized as follows. Section 2 describes the data used in our study. Section 3 defines our prediction task and presents the HNN model to tackle the task. Section 4 reports the experimental results. Section 5 presents a discussion and concludes the paper.

2. Data

Solar flares are classified according to their peak flux (in watts per square meter, Wm^{-2}) of 1 to 8 Å X-rays as measured by the Geostationary Operational Environmental Satellite (GOES) provided by the National Centers for Environmental Information (H. Liu et al. 2019; Y. Abduallah et al. 2023). In this study, we consider three cumulative flare classes (E. Jonas et al. 2018; N. Nishizuka et al. 2018; H. Liu et al. 2019; Y. Abduallah et al. 2023): ⩾M5.0-class, ⩾M-class, and ⩾C-class. The ⩾M5.0-class contains flares with peak soft X-ray flux above $5 \times 10^{-5}\mathrm{Wm}^{-2}$. Given its cumulative property, the ⩾M-class contains GOES M- and X-class flares, while the ⩾C-class contains GOES C-, M-, and X-class flares. We select flares with identified ARs that occurred in the period between May 2010 and December 2022. We then download line-of-sight (LOS) magnetograms of Space-weather HMI Active Region Patches (SHARPs; M. G. Bobra et al. 2014), produced by the SDO/HMI team, from the Joint Science Operations Center.[8] These HMI magnetograms are downloaded at a cadence of 1 hr. The downloaded magnetogram images have varying dimensions. We crop these images to the same size with 448 × 448 pixels and then resize the images to 224 × 224 pixels for model training purposes. Following the literature (Y. Zheng et al. 2021; P. Sun et al. 2022), we exclude ARs that are outside ±45° of the central meridian to avoid projection effects.

Next, we extract data from the Space Weather Database of Notifications, Knowledge, Information (DONKI),[9] to check whether a given flare is associated with a CME. Combining information from DONKI, we create three datasets separately: a ⩾M5.0 dataset, a ⩾M dataset, and a ⩾C dataset. In the ⩾M5.0 dataset, there are 120 ⩾M5.0-class flares, among which 69 flares are eruptive, and 51 flares are confined. In the ⩾M dataset, there are 619 ⩾M-class flares, comprising 184 eruptive flares and 435 confined flares. In the ⩾C dataset, there are 773 ⩾C-class flares, comprising 301 eruptive flares and 472 confined flares. Note that, in practice, there are many more confined flares than eruptive flares in the ⩾C dataset. DONKI only records a fraction of those confined flares. Our data is mainly collected from DONKI. Thus, we essentially use an undersampling method to tackle the imbalance issue in the ⩾C dataset, which helps avoid bias in model training.

3. Methodology

3.1. Prediction Task

We aim to tackle the following prediction task, which involves a binary classification problem. Given a magnetogram x_t at time point t in an AR where the AR will produce a γ-class flare within the next 24 hr of t, we predict whether x_t is positive or negative. Here, we consider γ to be ⩾M5.0, ⩾M, or ⩾C, separately and individually. Predicting x_t to be positive means that the γ-class flare will be associated with a CME (i.e., the flare is eruptive). Predicting x_t to be negative means that the γ-class flare will not be associated with a CME (i.e., the flare is confined).

Figure 1 illustrates how we label the magnetograms at hand for model training and testing. We label the magnetograms for the ⩾M5.0-, ⩾M-, and ⩾C-class flares separately and individually. For a flare that is associated with a CME, we label the magnetograms collected 24 hr before the flare as positive, as shown in the left panel of Figure 1. For a flare that is not associated with a CME, we label the magnetograms collected 24 hr before the flare as negative, as shown in the right panel of Figure 1. When there are two temporally close flare events that have overlapping magnetograms with opposite labels, we give priority to the positive class, labeling the overlapping magnetograms as positive. Table 1 presents the numbers of positive and negative magnetograms obtained by our labeling scheme for each cumulative flare class. Note that although there are fewer positive/eruptive ⩾C-class flares than negative/confined ⩾C-class flares, there are more positive magnetograms than negative magnetograms for the ⩾C-class flares, as shown in Table 1. This happens because we give priority to the positive class in labeling overlapping magnetograms as indicated above, and there are many such magnetograms for the ⩾C-class flares.

For each cumulative flare class, we employed a stratified 80:20 split for training and testing our model, allocating 80% of the magnetograms in the respective dataset shown in

[8] http://jsoc.stanford.edu/

[9] https://kauai.ccmc.gsfc.nasa.gov/DONKI/

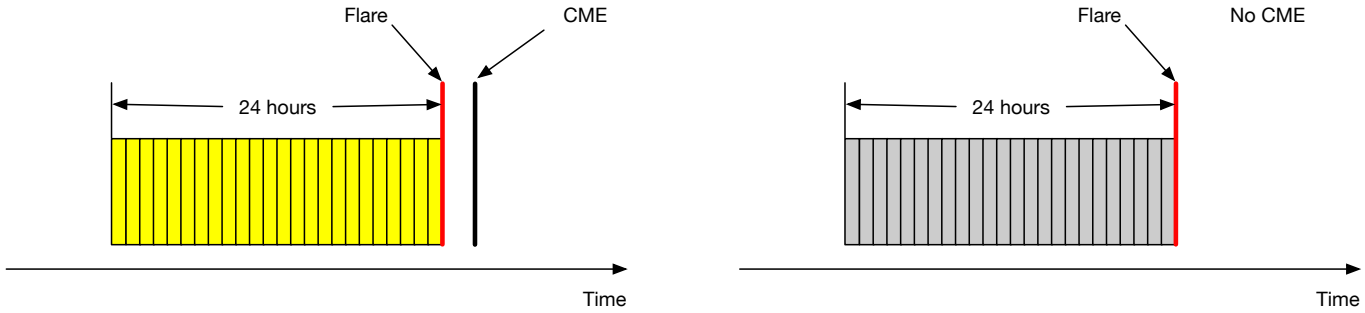


Figure 1. Construction of positive and negative magnetograms used in our prediction task. The magnetograms are collected at a cadence of 1 hr. Each rectangular box corresponds to 1 hr and contains one magnetogram. The red vertical line indicates the peak time of a γ-class flare, where γ implies $\geqslant$M5.0, $\geqslant$M, or $\geqslant$C. The yellow rectangular boxes shown in the left panel contain magnetograms that are within the 24 hr prior to the peak time of a γ-class flare that is associated with a CME; these yellow magnetograms belong to the positive class. The gray rectangular boxes shown in the right panel contain magnetograms that are within the 24 hr prior to the peak time of a γ-class flare that is not associated with a CME; these gray magnetograms belong to the negative class. Notice that in theory, there are 24 magnetograms that are within the 24 hr prior to the peak time of a γ-class flare. However, in practice, since some magnetograms could be missing, there could be fewer than 24 magnetograms collected in our datasets.

Table 1
Sample Sizes for Positive and Negative Classifications as Defined in Our Prediction Task

Cumulative Flare Class	Positive/Eruptive	Negative/Confined
$\geqslant$M5.0	2645	1685
$\geqslant$M	6703	10,631
$\geqslant$C	12,336	12,059

Table 1 to the training set and the remaining 20% to the test set. Magnetograms of the same AR are placed in the training set or in the test set, but not in both. In this way, we ensure that our HNN model is trained with ARs different from the ARs in the test set and makes predictions on the test ARs that the model has never seen during training.

3.2. The HNN Model

Figure 2 shows the overall architecture of the proposed HNN model. The model leverages the strength of a vision transformer (ViT) in capturing global spatial dependencies within an image and the capability of long short-term memory (LSTM) for sequence modeling. Let x_t represent the magnetogram collected at time point t in an AR. The input of HNN is a sequence of m consecutive magnetograms $x_{t-m+1}, x_{t-m+2}, x_{t-m+3}\ldots, x_{t-2}, x_{t-1}, x_t$ where time point t and time point $t-i$ differ by i hr. (In our study, m is set to 10.) The label of these m consecutive magnetograms is defined as the label of the last magnetogram x_t. Thus, if x_t belongs to the positive class, then the input sequence x_{t-m+1}, x_{t-m+2}, $x_{t-m+3}\ldots$, x_{t-2}, x_{t-1}, x_t is defined as positive; otherwise, the sequence is defined as negative. Note that we collect and label 24 magnetograms that are within the 24 hr prior to the peak time of a flare, as illustrated in Figure 1. These collected labeled magnetograms (positive versus negative) constitute the datasets at hand, as shown in Table 1. When we train/test our HNN model, we take each magnetogram x_t in the corresponding training/test set, use x_t and its preceding $(m-1)=9$ magnetograms to form a time series or sequence of 10 magnetograms, and input the sequence to the HNN model as shown in Figure 2. Forming this sequence of 10 magnetograms allows our model to understand the evolution of magnetic fields and capture temporal patterns in the input sequence, hence enhancing model accuracy. When there are more than two consecutive magnetograms missing in the sequence, x_t is excluded from the study. When there is only one or two consecutive magnetograms missing in the sequence, we use linear interpolation to generate synthetic magnetograms, fill in the gaps, and produce a complete, nongapped time series used as the input of HNN.

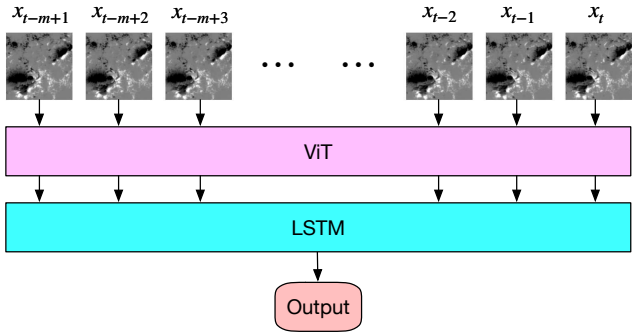


Figure 2. Overall architecture of our HNN model. Given is a magnetogram x_t at time point t in an AR that will produce a γ-class flare within the next 24 hr of t. The HNN model accepts, as input, a time series or sequence of m magnetograms x_{t-m+1}, x_{t-m+2}, $x_{t-m+3}\ldots$, x_{t-2}, x_{t-1}, x_t where m is set to 10. The sequence is first processed image-by-image by a vision transformer (ViT) to extract spatial features in the sequence. The resultant sequence of m embedding vectors is then sent to a long short-term memory (LSTM) network to capture temporal dependencies in the sequence. Finally, the HNN model produces, as output, "1" indicating that x_t is positive (i.e., the γ-class flare is eruptive) or "0" indicating that x_t is negative (i.e., the γ-class flare is confined).

As illustrated in Figure 2, the m input magnetograms are sent to the vision transformer (ViT), which captures the spatial patterns in the input magnetograms and outputs m embedding vectors, one embedding vector corresponding to one input magnetogram. The sequence of m embedding vectors is then sent to the LSTM network, which captures the temporal patterns in the embedding vector sequence. Finally, our HNN model outputs "1" indicating that x_t is positive (that is, a flare-associated CME will be triggered) or "0" indicating that x_t is negative (that is, a flare-associated CME will not be triggered).

Figure 3 presents the architectural design of the vision transformer (ViT). The input of the ViT is a magnetogram, which is divided into image patches of $N \times N$ pixels. (In the study presented here, N is set to 16). Each image patch is flattened into a vector, which is linearly projected onto an embedding space by a linear projection layer. To retain spatial information, which is lost during the patching process, positional embeddings are added to the patch embeddings by

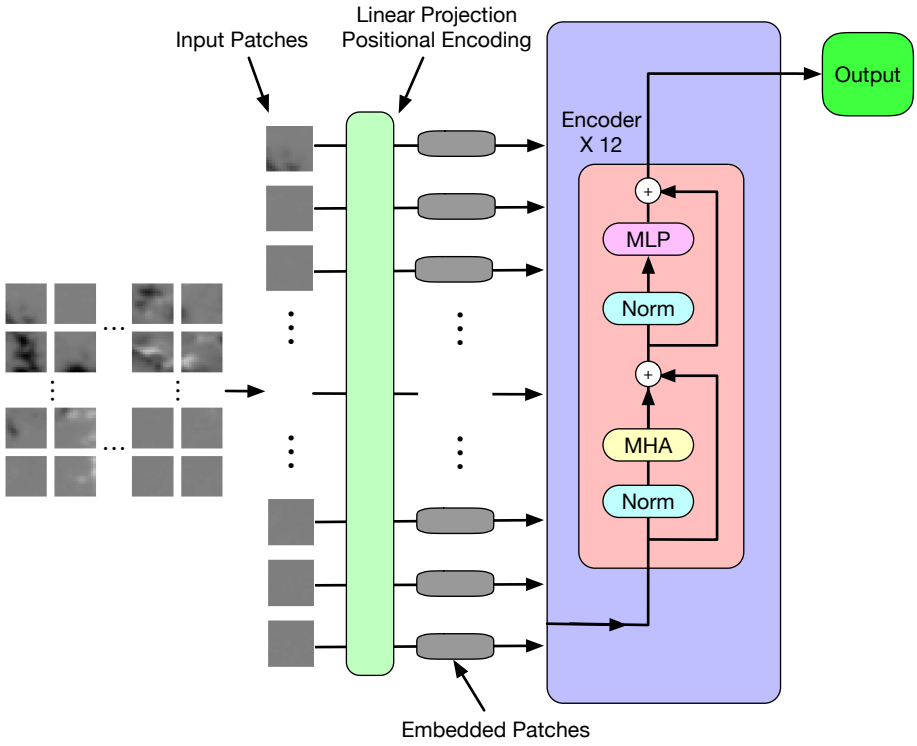


Figure 3. Architectural design of the vision transformer (ViT). Each input magnetogram image is divided into nonoverlapping patches, which are embedded via linear projection and positional encoding. These patch embeddings are fed into a stack of transformer encoder layers, each of which contains a multihead self-attention (MHA) module, layer normalization (Norm), and a multilayer perceptron (MLP) module, to generate the final image representation.

a positional encoding layer. These positional embeddings are typically learnable parameters that allow the ViT to understand the original location of each patch. The resultant sequence of embedding vectors is then sent to the ViT encoders.

The ViT contains E encoders. (In our study presented here, E is set to 12). Each encoder updates the sequence representation based on two modules: a multihead self-attention (MHA) module, where the number of attention heads in an encoder is also set to 12, and a position-wise multilayer perceptron (MLP) module. Layer normalization (Y. LeCun et al. 2015) is applied before each module, and a residual connection (K. He et al. 2016), denoted by $\oplus$, is used after each module. The MHA mechanism allows the ViT to dynamically weigh the relevance of different patches, capturing long-range dependencies throughout the input magnetogram by computing the attention scores between all pairs of patch embeddings. The MLP module consists of two linear layers with a Gelu nonlinearity in between, applied independently to each sequence element. This deep processing allows the ViT to extract complex hierarchical features from the spatial arrangement of patches. To accelerate convergence and improve model performance, we load the pretrained ViT weights of A. Dosovitskiy et al. (2021) to our ViT model. This allows the model to leverage the features learned from large-scale datasets, thereby improving generalization and training efficiency.

After processing the data through the E encoders, we produce a final embedding vector, which is the output of the ViT. There are m input magnetograms, which are sent to the ViT. The ViT outputs a sequence of m embedding vectors. The embedding vector sequence is then fed to the LSTM network as shown in Figure 2. The LSTM network processes this sequence to further refine features and capture potential sequential patterns within the embedded spatial data, leveraging its gating mechanisms to model complex dependencies.

The final output of the LSTM network in Figure 2 is sent to a classification layer, which predicts the probability $\hat{y}$ of the occurrence of a flare-associated CME. Our primary goal is binary classification (that is, there will be an eruptive flare versus there will be a confined flare within the next 24 hr). When $\hat{y}$ is greater than or equal to the default probability 0.5, the output is "1" or positive, indicating that there will be an eruptive flare within the next 24 hr. When $\hat{y}$ is smaller than 0.5, the output is "0" or negative, indicating that there will be a confined flare within the next 24 hr. Recognizing the class imbalance inherent in our datasets, as shown in Table 1, we adopt a class-balanced focal loss (T. Lin et al. 2017), which is defined as:

$$\mathrm{FL}(y, \hat{y}) = -\alpha_s (1 - p_s)^{\lambda} \log(p_s), \tag{1}$$

where $y \in \{0, 1\}$ is the ground truth label, $\hat{y}$ is the model's estimated probability for class $y = 1$, and

$$p_s = \begin{cases} \hat{y} & \text{if } y = 1 \\ 1 - \hat{y} & \text{if } y = 0 \end{cases} \tag{2}$$

$$\alpha_s = \begin{cases} \alpha & \text{if } y = 1 \\ 1 - \alpha & \text{if } y = 0 \end{cases} \tag{3}$$

Here, $\alpha \in [0, 1]$ is a balancing factor to address class imbalance (weighting the minority class higher), and $\lambda \geqslant 0$ is the focusing parameter that down-weights well-classified examples, forcing the model to focus on harder, misclassified examples. (In the study presented here, α is set to 0.4 and λ is set to 2). The focal loss function helps mitigate the dominance of the majority class during training. The parameters of the

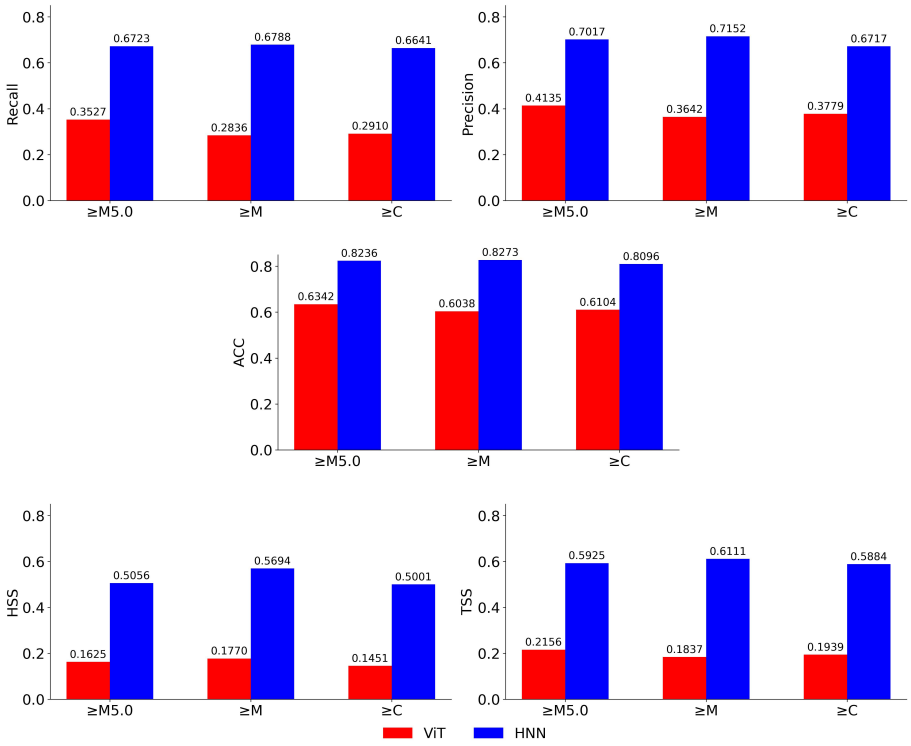


Figure 4. Performance comparison between ViT and HNN.

HNN model are optimized using the Adam optimizer (D. P. Kingma & J. Ba 2015) with a learning rate of 0.0001. The batch size is 32, and the number of epochs is 40.

4. Results

4.1. Experimental Setup and Evaluation Metrics

Given an AR that produces a γ-class flare within the next 24 hr of a time point t and a magnetogram x_t collected at time point t, we define x_t as a true positive (TP) if our HNN model predicts that x_t is positive, and x_t is indeed positive, i.e., the γ-class flare is associated with a CME. We define x_t as a false positive (FP) if our model predicts that x_t is positive while x_t is actually negative, i.e., the γ-class flare is not associated with a CME. We say x_t is a true negative (TN) if our model predicts x_t to be negative and x_t is indeed negative; x_t is a false negative (FN) if our model predicts x_t to be negative while x_t is actually positive. We also use TP (FP, TN, FN, respectively) to represent the total number of true positives (false positives, true negatives, false negatives, respectively) produced by our model.

The evaluation metrics used in the study include the following:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \tag{4}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \tag{5}$$

$$\text{Accuracy (ACC)} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}}, \tag{6}$$

$$\text{Heidke Skill Score (HSS)} = \frac{2(\text{TP} \times \text{TN} - \text{FP} \times \text{FN})}{(\text{TP}+\text{FN})(\text{FN}+\text{TN}) + (\text{TP}+\text{FP})(\text{FP}+\text{TN})}, \tag{7}$$

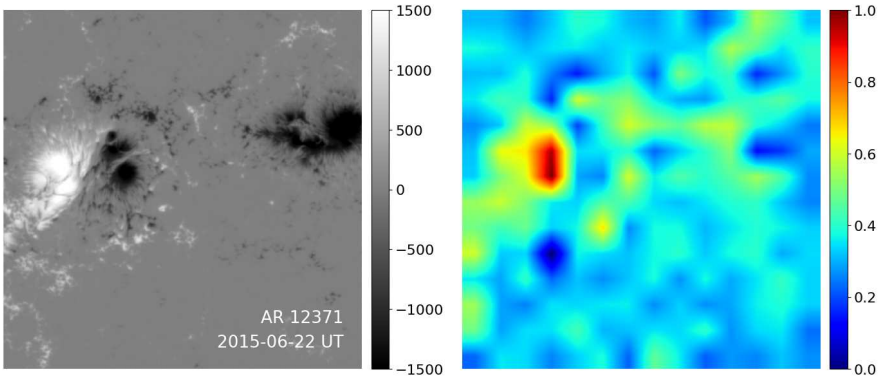
$$\text{True Skill Statistics (TSS)} = \frac{\text{TP}}{\text{TP} + \text{FN}} - \frac{\text{FP}}{\text{TN} + \text{FP}}. \tag{8}$$

In the above, recall, precision, and accuracy range between 0 and 1, with higher values indicating better performance. HSS and TSS range between -1 and 1. HSS shows the performance in comparison to a generic random prediction (negative/positive values meaning worse/better performance), while TSS compares the true positive rate (first term in Equation (8)) with the false positive rate (second term in Equation (8)). HSS $= 0$ implies performance identical to that of a no-skill, generic random prediction, while TSS $= 0$ indicates an exact balance between the true positive rate and the false positive rate. For a comprehensive account, see I. T. Jolliffe & D. B. Stephenson (2011) and D. S. Bloomfield et al. (2012).

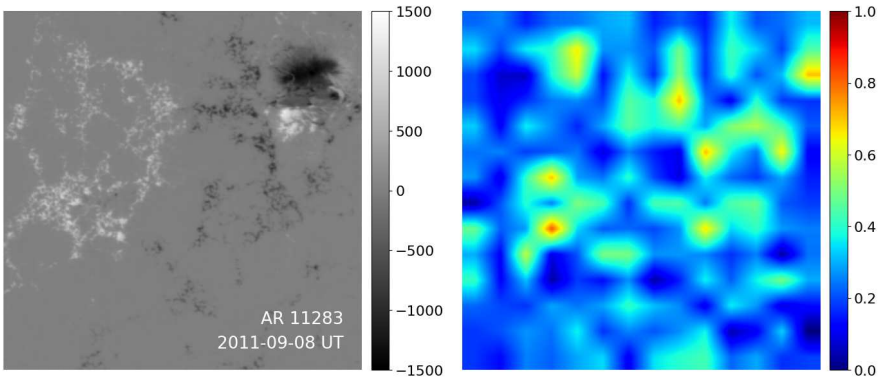
We use the vision transformer (ViT) alone as a baseline method. Unlike our HNN model, which uses a time series or sequence of m consecutive magnetograms x_{t-m+1}, x_{t-m+2}, $x_{t-m+3}\ldots$, x_{t-2}, x_{t-1}, x_t as input to predict whether x_t is positive or negative, the ViT baseline method only uses x_t as input to make the prediction. Thus, the ViT baseline method captures the spatial patterns of the input magnetogram x_t without considering the temporal patterns in the magnetogram sequence used by our HNN model.

4.2. Performance Evaluation of the HNN Model

Figure 4 evaluates the performance of HNN and compares it with ViT. It can be seen in Figure 4 that HNN outperforms ViT in all five metrics. Specifically, for the $\geqslant$M5.0-class flares, HMM achieves a TSS of 0.5925 compared to the TSS of



(a)



(b)

Figure 5. Attention heat maps of the HNN model for two test magnetograms. The gray-scale bar corresponds to the LOS magnetic field, in Gauss. The color bar shows the attention score. A larger attention score at a region indicates that more attention is paid to the region, where large attention scores are represented by dark red and small attention scores are represented by dark blue. (a) A positive prediction (true positive) where the magnetogram on the left is predicted to be positive, related to an eruptive flare. The model pays much attention to the vicinity of a polarity inversion line (PIL) region in the magnetogram. (b) A negative prediction (true negative) where the magnetogram on the left is predicted to be negative, related to a confined flare. The model pays no attention to the clearly weaker PIL region in the magnetogram. Instead, the model pays some attention to the region where no PIL exists.

0.2156 obtained by ViT. For the $\geqslant$M-class flares, HMM achieves a TSS of 0.6111 compared to the TSS of 0.1837 obtained by ViT. For the $\geqslant$C-class flares, HNN achieves a TSS of 0.5884 compared to the TSS of 0.1939 obtained by ViT. These findings underscore the effectiveness of combining ViT with LSTM in our HNN framework. Good performance in recall, precision, accuracy, HSS, and TSS with solar observations as direct input without handpicked features highlights the potential of HNN as an effective tool for operational space weather forecasting.

4.3. Model Interpretation

HNN is a transformer-based model with attention mechanisms. To elucidate the decision-making process of our model, we employ saliency methods (Y. A. D. Djilali et al. 2023; L. Xu et al. 2023) that use the attention weights within the transformer layers of the model to infer importance and directly link the model's predictions to the most influential regions in the input magnetogram. We leverage the attention mechanisms and generate attention heat maps, which reveal the regions where the model focuses the most during inference. Figure 5(a) (Figure 5(b), respectively) presents an attention heat map for a positive (negative, respectively) prediction. In Figure 5(a), according to the attention heat map in the right panel, our model pays much attention to the vicinity of a polarity inversion line (PIL) region in the magnetogram in the left panel. As a result, the model predicts that the magnetogram is positive, i.e., a flare will occur within the next 24 hr and the flare is eruptive. This prediction result is consistent with the established fact that CMEs originate near or along PILs (C. Liu et al. 2019; O. P. M. Aslam et al. 2024). In Figure 5(b), according to the attention heat map in the right panel, our model does not pay any attention to the PIL region in the magnetogram in the left panel. Instead, the model pays some attention to a region where no PIL exists. As a consequence, the model predicts that the magnetogram is negative, i.e., a flare will occur within the next 24 hr and the flare is confined. These attention heat maps help us better understand the decision-making process of the HNN model. The results offer enhanced reliability of HNN for distinguishing between eruptive and confined flare events.

4.4. Physics-based Analysis of Model Predictions

We further pursue a physics-based analysis of model predictions, examining the relationship between magnetic flux cancellation and the occurrence of flare-associated CMEs. Magnetic flux cancellation occurs along PILs where magnetic fields of opposite polarity frequently come into contact. We consider six ARs where each AR is associated with a reference time point t (Figure 6). Our HNN model predicts whether the

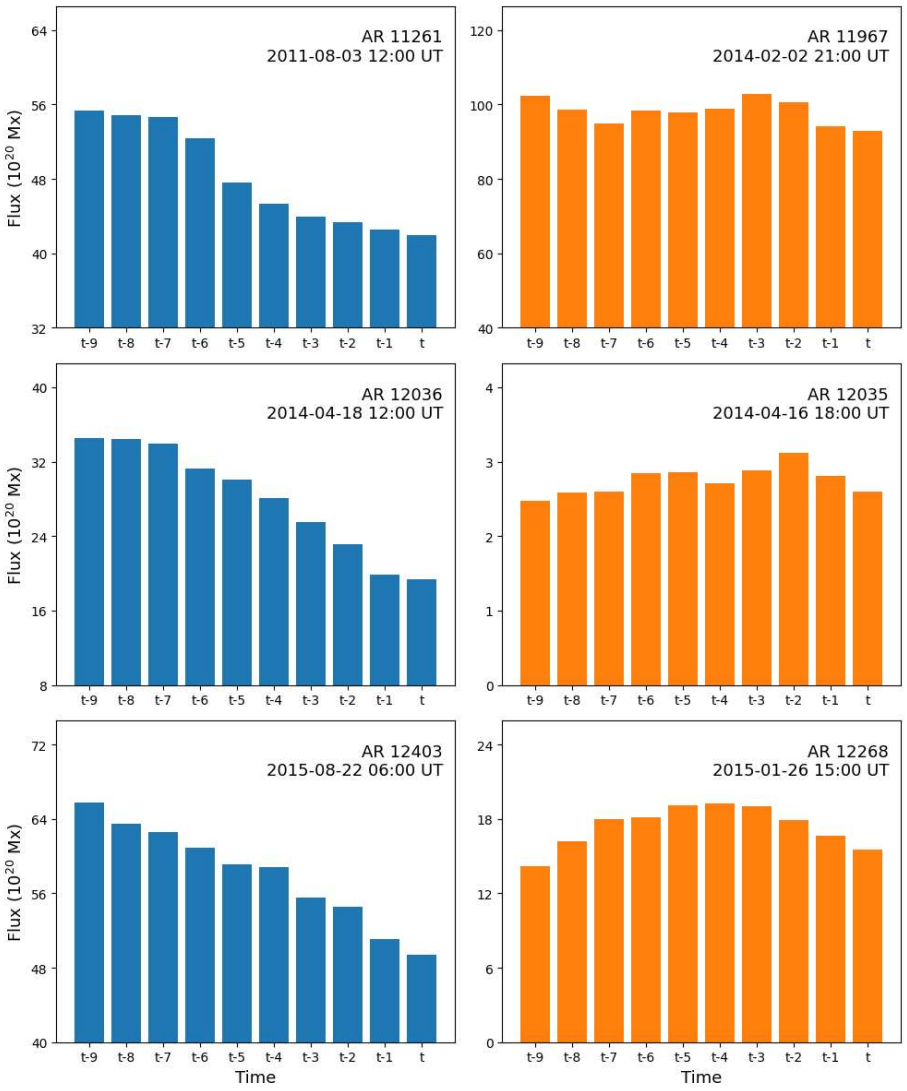


Figure 6. Illustration of the changes of the unsigned magnetic flux in the PIL regions of the 10 magnetograms at time points $t-9$, $t-8$, …, $t-1$, t used as the input of our HNN model in six ARs, where t is as specified in each panel. Left panels: results for three positive predictions (true positives), where a positive prediction indicates that the corresponding AR will produce an eruptive flare within the next 24 hr of t. Right panels: results for three negative predictions (true negatives), where a negative prediction indicates that the corresponding AR will produce a confined flare within the next 24 hr of t. In the three positive predictions, the unsigned magnetic flux in the PIL regions decreases in the course of time, which may point to magnetic flux cancellation.

magnetogram at t (i.e., x_t) in the AR is positive (i.e., a flare will occur within the next 24 hr of t and the flare is eruptive) or negative (i.e., a flare will occur within the next 24 hr of t and the flare is confined). As described in Section 3.2, the input of the HNN model is a sequence of 10 consecutive magnetograms x_{t-9}, x_{t-8}, …, x_{t-1}, x_t. Magnetic flux cancellation can be quantified by the unsigned magnetic flux contained within the PIL regions in each magnetogram. Using the PIL identification algorithm developed by A. Ji et al. (2023), we locate all PIL regions in a magnetogram.

Figure 6 shows the changes of the unsigned magnetic flux in the PIL regions of the 10 magnetograms used as the input of our HNN model in each of the six ARs. The left panels with blue histograms show three positive predictions (true positives), where a positive prediction indicates that the corresponding AR will produce an eruptive flare. The right panels with orange histograms show three negative predictions (true negatives), where a negative prediction indicates that the corresponding AR will produce a confined flare. It is evident that, in the three positive predictions, the unsigned magnetic flux in the PIL regions decreases in the course of time, which may point to magnetic flux cancellation. We must emphasize, of course, that since we only rely on the LOS field component, similar to the vertical magnetic field, if the target region is relatively close to the central meridian, such a behavior could also be caused by a systematic change in the geometry of the photospheric magnetic field vector, i.e., becoming more horizontal in the course of time. Fully establishing that magnetic flux cancellation plays a role in the behavior seen in Figure 6 (blue histograms) implies a detailed analysis of the full magnetic field vector and EUV/X-ray transient brightenings along the PIL regions, which exceeds the scope of this work. We note, however, that flux cancellation may well play a role, as the behavior reported by Y. Liu et al. (2023) points to a systematic weakening of the photospheric horizontal field in the hours prior to eruptive flares (but not prior to confined flares), only to return to initial preeruption level values immediately after the eruptions. Showing here that the vertical

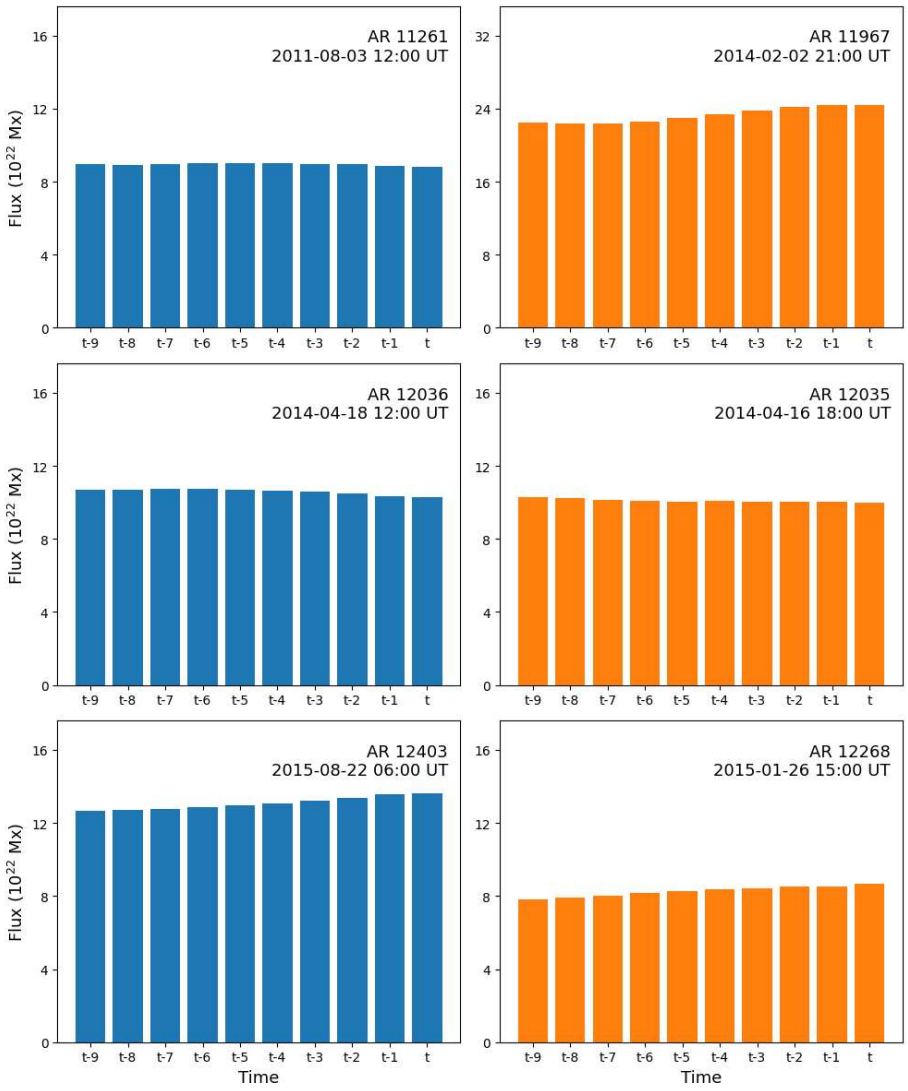


Figure 7. Illustration of the changes of the total unsigned magnetic flux in the 10 magnetograms at time points $t-9$, $t-8$, …, $t-1$, t used as the input of our HNN model in the six ARs shown in Figure 6 where t is as specified in each panel. Left panels: results for the three positive predictions (true positives), where a positive prediction indicates that the corresponding AR will produce an eruptive flare within the next 24 hr of t. Right panels: results for the three negative predictions (true negatives), where a negative prediction indicates that the corresponding AR will produce a confined flare within the next 24 hr of t. There is no clear difference in the trends between the results for the three positive predictions and those for the three negative predictions.

magnetic field becomes weaker instead of strengthening, as would be expected by the results of Y. Liu et al. (2023), indicates true loss of flux, possibly due to flux cancellation. However, for the three negative predictions (orange histograms), no clear trend is observed in the changes of the unsigned magnetic flux in the PIL regions.

We next examine the relationship between the larger-scale magnetic environment in an AR and the occurrence of flare-associated CMEs. Specifically, in each magnetogram, x_{t-9}, x_{t-8}, …, x_{t-1}, x_t, of the AR, we compute the total unsigned magnetic flux of the whole magnetogram. Figure 7 shows the changes of the total unsigned magnetic flux in the 10 magnetograms used as the input of our HNN model in each of the six ARs in Figure 6. There is no clear difference in the trends between the results for the three positive predictions (blue histograms) in the left panels and those for the three negative predictions (orange histograms) in the right panels in Figure 7.

5. Discussion and Conclusion

We developed a hybrid neural network (HNN) to predict whether a γ-class flare that will occur within the next 24 hr will be associated with a CME, i.e., whether the flare will be eruptive or confined. We considered γ to be defined cumulatively, as $\geqslant$M5.0, $\geqslant$M, or $\geqslant$C. When γ was $\geqslant$M, the HNN model achieved a TSS of 0.6111. A similar result with slightly lower TSS value, albeit using different datasets and prediction methods, was reached by H. Liu et al. (2020), though H. Liu et al. (2020) did not consider $\geqslant$M5.0/$\geqslant$C-class flares. Note that the TSS value is calculated based on the test set at hand. If the methodology presented here were to be used in a near-real time operation scheme, its testing results on eruptive flares versus confined flares may be different than the ones presented here. Note also that, as in the work of H. Liu et al. (2020), we rely on existing tools (e.g., M. G. Bobra & S. Couvidat 2015; C. Liu et al. 2017; K. Florios et al. 2018; X. Huang et al. 2018;

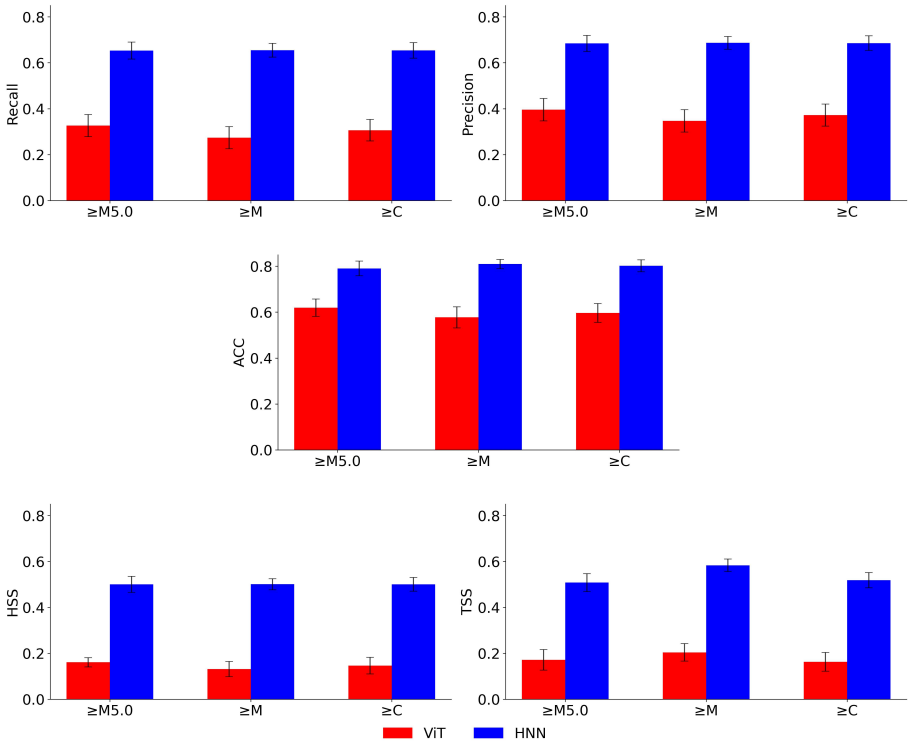


Figure 8. Bar graphs showing the comparison between ViT and HNN based on the five-fold cross-validation scheme.

E. Jonas et al. 2018; N. Nishizuka et al. 2018; H. Liu et al. 2019; X. Li et al. 2020; X. Wang et al. 2020; P. Sun et al. 2022; H. Zhang et al. 2022; Y. Abduallah et al. 2023; J. Wen et al. 2025; D. Xu et al. 2025) to predict whether a γ-class flare will occur within the next 24 hr . If it is predicted that a γ-class flare will occur, then one can use the HNN model developed here to further predict whether that flare will be eruptive (i.e., it is associated with a CME) or confined (i.e., it is not associated with a CME). It should be pointed out that our work is based on DONKI. There are other datasets for flares and CMEs (e.g., A. Ji et al. 2025). It is worthwhile to consider those other datasets in future studies.

Our HNN model employs several parameters and hyperparameters. For example, we use a time series or sequence of 10 magnetograms as input to the HNN model. We tried shorter sequences with six or eight magnetograms and longer sequences with 14 or 18 magnetograms as input. Shorter sequences tended to underutilize the available temporal information in the input, whereas longer sequences increased computational time without yielding better performance. The input magnetogram to the ViT of our model is divided into image patches of 16×16 pixels. We also tried smaller patches of 8×8 pixels and larger patches of 32×32 pixels. Smaller patches yielded more patch embeddings and hence increased processing time, while performance was not improved. Larger patches caused a loss of fine-scale magnetic structure information, yielding worse performance. Likewise, the ViT uses 12 encoders. ViT modules with fewer encoders tended to underfit the training data, while ViT modules with more encoders and deeper configurations increased training time without yielding better performance. These hyperparameter values are optimized using the grid search strategy (F. Pedregosa et al. 2011).

To explain a prediction made by our HNN model, we leveraged the attention mechanisms of the model using saliency methods and generated attention heat maps, which revealed the regions in an input magnetogram where the model focused the most during inference. For a positive prediction (that is, a prediction indicating that the flare related to the input magnetogram was eruptive), the heat map indicated that the model paid much attention to the vicinity of polarity inversion lines (PILs) in the input magnetogram. For a negative prediction (that is, a prediction indicating that the flare related to the input magnetogram was confined), the heat map indicated that the model paid attention to regions where no PIL existed. The heat maps provided insight into how our model produced its predictions, making the model more interpretable. In addition, we aimed to physically understand our model predictions. It appears likely that magnetic flux cancellation in PIL regions plays a role in triggering flare-associated CMEs, a finding consistent with the literature (A. C. Sterling et al. 2018; Y. Liu et al. 2023; S. Patsourakos & V. Archontis 2025). This said, the evolution of the total unsigned magnetic flux in the entire active region does not appear to show a clear relation to eruptive flares.

To further understand the behavior of our HNN model and compare it with the baseline method (ViT), we conducted an additional five-fold cross-validation experiment. In this experiment, there were five runs. In each run, for each cumulative class of flares, 20% of the data in Table 1 were selected from the respective dataset and used for testing; the remaining 80% of the data were used for training. Again, magnetograms of the same AR were placed in the training set or in the test set, but not in both. We recorded recall, precision, accuracy (ACC), Heidke Skill Score (HSS), and True Skill Statistic (TSS) in each run and calculated the mean and standard deviation over the five runs. Figure 8 presents the results, where each colored bar represents the mean of the five runs and its associated error bar represents the standard

deviation divided by the square root of the number of runs. The figure shows that HNN consistently outperforms ViT in all metrics, consolidating the findings in Figure 4. On the basis of these results, we conclude that the HNN model shows promise for predicting associations between solar flares and CMEs.

Acknowledgments

We acknowledge constructive suggestions and comments from an anonymous referee and the scientific editor that have helped to significantly improve the manuscript. We also thank members of the Institute for Space Weather Sciences for fruitful discussions. SDO is the first mission launched for NASA's Living With a Star (LWS) Program. The SDO/HMI data is provided by the Joint Science Operations Center (JSOC) Science Data Processing (SDP). DONKI is developed by the Community Coordinated Modeling Center (CCMC) at NASA. The proposed HNN model is implemented in PyTorch and Astropy packages. V.Y. acknowledges support from NSF AGS grants 2401229, 2408174, 2300341, 2309939, NSF AST-2108235 grant, and NASA 80NSSC24K1914 grant. J.W. and H.W. acknowledge support from NSF AGS grants 2228996, 2149748, NSF OAC grants 2504860, 2320147, and NASA grants 80NSSC24K0548, 80NSSC24K0843, 80NSSC24M0174. M.K.G. acknowledges support from NSF OAC grants 2504861 and 2513887. Y.X. acknowledges support from NSF AGS grants 2228996, 2229064, and NSF RISE-2425602 grant.

ORCID iDs

Vasyl Yurchyshyn https://orcid.org/0000-0001-9982-2175
Jason T. L. Wang https://orcid.org/0000-0002-2486-1097
Haimin Wang https://orcid.org/0000-0002-5233-565X
Manolis K. Georgoulis https://orcid.org/0000-0001-6913-1330
Yasser Abduallah https://orcid.org/0000-0003-0792-2270
Hameedullah A. Farooki https://orcid.org/0000-0001-7952-8032

References

Abduallah, Y., Wang, J. T. L., Wang, H., & Xu, Y. 2023, NatSR, 13, 13665
Afanasyev, A. N., Fan, Y., Kazachenko, M. D., & Cheung, M. C. M. 2023, ApJ, 952, 136
Alobaid, K. A., Abduallah, Y., Wang, J. T. L., et al. 2023, ApJL, 958, L34
Alpaydin, E. 2016, Machine Learning: The New AI (MIT Press)
Aslam, O. P. M., MacTaggart, D., Williams, T., Fletcher, L., & Romano, P. 2024, MNRAS, 534, 444
Baker, D. N., Daly, E., Daglis, I., Kappenman, J. G., & Panasyuk, M. 2004, SpWea, 2, S02004
Baumgartner, C., Thalmann, J. K., & Veronig, A. M. 2018, ApJ, 853, 105
Berkebile-Stoiser, S., Veronig, A. M., Bein, B. M., & Temmer, M. 2012, ApJ, 753, 88
Bloomfield, D. S., Higgins, P. A., McAteer, R. T. J., & Gallagher, P. T. 2012, ApJL, 747, L41
Bobra, M. G., & Couvidat, S. 2015, ApJ, 798, 135
Bobra, M. G., & Ilonidis, S. 2016, ApJ, 821, 127
Bobra, M. G., Sun, X., Hoeksema, J. T., et al. 2014, SoPh, 289, 3549
Chen, P. F. 2011, LRSP, 8, 1
DeVore, C. R., & Antiochos, S. K. 2008, ApJ, 680, 740
Djilali, Y. A. D., McGuinness, K., & O'Connor, N. E. 2023, in in 34th British Machine Vision Conf., 771, http://proceedings.bmvc2023.org/771/
Dosovitskiy, A., Beyer, L., Kolesnikov, A., et al. 2021, in in 9th Int. Conf. on Learning Representations https://openreview.net/forum?id=YicbFdNTTy
Florios, K., Kontogiannis, I., Park, S.-H., et al. 2018, SoPh, 293, 28
Georgoulis, M. K., Yardley, S. L., Guerra, J. A., et al. 2024, AdSpR, 45, 2903
Harrison, R. A. 1995, A&A, 304, 585
He, K., Zhang, X., Ren, S., & Sun, J. 2016, in in IEEE Conf. on Computer Vision and Pattern Recognition, 770
Hochreiter, S., & Schmidhuber, J. 1997, Neural Computation, 9, 1735
Hopfield, J. J. 1982, PNAS, 79, 2554
Huang, X., Wang, H., Xu, L., et al. 2018, ApJ, 856, 7
Inceoglu, F., Jeppesen, J. H., Kongstad, P., et al. 2018, ApJ, 861, 128
Ji, A., Cai, X., Khasayeva, N., et al. 2023, ApJS, 265, 28
Ji, A., Georgoulis, M. K., & Aydin, B. 2025, DIB, 60, 111539
Jolliffe, I. T., & Stephenson, D. B. 2011, Forecast Verification: A Practitioner's Guide in Atmospheric Science (Wiley)
Jonas, E., Bobra, M., Shankar, V., Todd Hoeksema, J., & Recht, B. 2018, SoPh, 293, 48
Kawabata, Y., Iida, Y., Doi, T., et al. 2018, ApJ, 869, 99
Kazachenko, M. D. 2023, ApJ, 958, 104
Kingma, D. P., & Ba, J. 2015, Adam: A Method for Stochastic Optimization, in 3rd International Conf. on Learning Representations, ICLR 2015
LeCun, Y., Bengio, Y., & Hinton, G. 2015, Natur, 521, 436
Li, X., Zheng, Y., Wang, X., & Wang, L. 2020, ApJ, 891, 10
Lin, J., & Forbes, T. G. 2000, JGR, 105, 2375
Lin, T., Goyal, P., Girshick, R. B., He, K., & Dollár, P. 2017, in in IEEE Int. Conf. on Computer Vision, 2999
Liu, C., Chen, T., & Zhao, X. 2019, A&A, 626, A91
Liu, C., Deng, N., Wang, J. T. L., & Wang, H. 2017, ApJ, 843, 104
Liu, H., Liu, C., Wang, J. T. L., & Wang, H. 2019, ApJ, 877, 121
Liu, H., Liu, C., Wang, J. T. L., & Wang, H. 2020, ApJ, 890, 12
Liu, Y. 2008, ApJL, 679, L151
Liu, Y., Welsch, B. T., Valori, G., et al. 2023, ApJ, 942, 27
Low, B. C. 1994, PhPl, 1, 1684
Moore, R. L., Sterling, A. C., Hudson, H. S., & Lemen, J. R. 2001, ApJ, 552, 833
Nishizuka, N., Sugiura, K., Kubo, Y., Den, M., & Ishii, M. 2018, ApJ, 858, 113
Patsourakos, S., & Archontis, V. 2025, A&A, 699, A87
Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2011, Journal of Machine Learning Research, 12, 2825
Pesnell, W. D., Thompson, B. J., & Chamberlin, P. C. 2012, SoPh, 275, 3
Qahwaji, R., Colak, T., Al-Omari, M., & Ipson, S. 2008, SoPh, 248, 471
Rust, D. M., & Kumar, A. 1996, ApJL, 464, L199
Schou, J., Scherrer, P. H., Bush, R. I., et al. 2012, SoPh, 275, 229
Sterling, A. C., Moore, R. L., & Panesar, N. K. 2018, ApJ, 864, 68
Sun, P., Dai, W., Ding, W., et al. 2022, ApJ, 941, 1
Teraoka, K., Yamasaki, D., Kawabata, Y., Imada, S., & Shimizu, T. 2025, ApJ, 983, 126
Thalmann, J. K., Su, Y., Temmer, M., & Veronig, A. M. 2015, ApJL, 801, L23
Török, T., & Kliem, B. 2005, ApJL, 630, L97
Wang, X., Chen, Y., Toth, G., et al. 2020, ApJ, 895, 3
Wang, Y., & Zhang, J. 2007, ApJ, 665, 1428
Webb, D. F., & Howard, T. A. 2012, LRSP, 9, 3
Wen, J., Ahmadzadeh, A., Georgoulis, M. K., Sadykov, V. M., & Angryk, R. A. 2025, ApJS, 277, 60
Xu, D., Sun, P., Feng, S., Liang, B., & Dai, W. 2025, ApJS, 276, 68
Xu, L., Yan, X., Ding, W., & Liu, Z. 2023, J. Ambient Intell. Humaniz. Comput., 14, 163
Yashiro, S., Akiyama, S., Gopalswamy, N., & Howard, R. A. 2006, ApJL, 650, L143
Yashiro, S., & Gopalswamy, N. 2008, IAUS, 4, 233
Zhang, H., Li, Q., Yang, Y., et al. 2022, ApJS, 263, 28
Zhang, H., Jing, J., Wang, J. T. L., et al. 2025, ApJ, 981, 37
Zheng, Y., Li, X., Si, Y., Qin, W., & Tian, H. 2021, MNRAS, 507, 3519