



This is a repository copy of *Short-term solar eruptive activity prediction models based on machine learning approaches: a review*.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/217940/>

Version: Accepted Version

Article:

Huang, X., Zhao, Z., Zhong, Y. et al. (3 more authors) (2024) Short-term solar eruptive activity prediction models based on machine learning approaches: a review. *Science China Earth Sciences*, 67 (12). pp. 3727-3764. ISSN 1674-7313

<https://doi.org/10.1007/s11430-023-1375-2>

© 2024 The Authors. Except as otherwise noted, this author-accepted version of a journal article published in *Science China Earth Sciences* is made available via the University of Sheffield Research Publications and Copyright Policy under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution and reproduction in any medium, provided the original work is properly cited. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>

Reuse

This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here:

<https://creativecommons.org/licenses/>

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.



eprints@whiterose.ac.uk
<https://eprints.whiterose.ac.uk/>

Short-term Solar Eruptive Activity Prediction Models Based on Machine Learning Approaches: A Review

Xin Huang^{1,2*}, Zhongrui Zhao^{2,3,4}, Yufeng Zhong^{2,3}, Long Xu^{1,2}, Marianna B. Korsós^{5,6,7} & R. Erdélyi^{6,7,8}

¹*Faculty of Electrical Engineering and Computer Science, Ningbo University, Ningbo 315211, China;*

²*State Key Laboratory of Space Weather, National Space Science Center, Chinese Academy of Sciences, Beijing 100190, China;*

³*School of Astronomy and Space Science, University of Chinese Academy of Sciences, Beijing 100049, China;*

⁴*Beijing Cellular Exploration Science and Technology Center, Beijing 100049, China;*

⁵*Dipartimento di Fisica e Astronomia "Ettore Majorana", Università di Catania, Via S. Sofia 78, I 95123 Catania, Italy;*

⁶*Department of Astronomy, Eötvös Loránd University, Pázmány Péter sétány 1/A, H-1112 Budapest, Hungary;*

⁷*Hungarian Solar Physics Foundation, Peto"fi tér 3, H-5700 Gyula, Hungary;*

⁸*Solar Physics & Space Plasma Research Center (SP2RC), School of Mathematics and Statistics, University of Sheffield, Sheffield S3 7RH, UK*

Solar eruptive activities, mainly including solar flares, coronal mass ejections (CME), and solar proton events (SPE), have an important impact on space weather and our technosphere. The short-term solar eruptive activity prediction is an active field of research in the space weather prediction. Numerical, statistical, and machine learning methods are proposed to build prediction models of the solar eruptive activities. With the development of space-based and ground-based facilities, a large amount of observational data of the Sun is accumulated, and data-driven prediction models of solar eruptive activities have made a significant progress. In this review, we briefly introduce the machine learning algorithms applied in solar activity prediction, summarize the prediction modeling process, overview the progress made in the field of solar eruptive activity prediction model, and look forward to the possible directions in the future.

Solar flare, Coronal mass ejection, Solar proton event, Machine learning, Prediction model

1 Introduction

With more and more space activities, we are becoming increasingly concerned about the impact and prediction of solar eruptions on our technosphere (Temmer, 2021). Solar eruptive activities mainly include solar flares, coronal mass ejections (CMEs), and solar proton events (SPEs). A flare is a sudden solar activity event that manifests itself as the release of enormous amount of energy and radiation from the surface of the Sun (Benz, 2008). The occurrence of flares has an impact on the Earth's atmosphere, which may trigger near-Earth space weather activities and pose potential risks to human space activities and technological equipment. Therefore, the solar flare prediction has been an important topic in solar physics and space weather research. CMEs are among the significant phenomena within the Sun-Earth system, frequently leading to space environment disturbances and consequential geomagnetic perturbations (Webb and Howard, 2012). Therefore, CMEs have garnered significant attention in space weather research, and predicting whether and when they will impact Earth is a crucial aspect of space weather prediction. SPEs are the strong enhancement in energetic proton flux (typically >10 MeV), commonly from the related solar flare and/or the shock caused by the related CME (Waterfall et al., 2023). The solar eruptive activities and the associated radiations,

magnetized plasma and high energetic particles could have an impact on satellite operation, navigation system, astronaut safety, communication channels on Earth, polar flight and electrical power grid (Temmer, 2021).

Space weather prediction has evolved considerably in the recent and is now entering the era of deep learning (Asensio Ramos et al., 2023), which we have summarised from a machine learning perspective in the hope of providing some inspirations. For solar flare prediction (Toriumi and Wang, 2019), there are modelling methods based on magnetohydrodynamics (MHD) and self-organized criticality (SOC) models from the physical modelling perspective. Both of these approaches have good physical background and interpretability. In operational prediction, more applications are based on data-driven modelling approaches, such as the earliest empirical-based expert systems, and the later statistical methods, machine learning methods for manually extracting features, and deep learning methods for automatically extracting features. In this article, we described in detail the data, the modelling process of flare prediction, the classical algorithms based on machine learning and the deep learning methods, respectively. For CME prediction (Zhao and Dryer, 2014), various physics-based models have been developed over the past decades. In recent years, the continuous advancement of machine learning and deep learning has promoted the emergence of new models. Thus, we present an overview of the current application of machine learning (e.g., support vector machines (SVM) and logistic regression) and deep learning (e.g., neural networks (NN) and convolutional neural networks (CNN)) in predicting the arrival of CMEs and their associated interplanetary shock waves on Earth. Besides, we highlight the shortcomings of existing models and suggest future directions for improvement. For SPEs prediction (Stumpo et al., 2021), both the statistical and the machine learning methods are applied to provide the prediction results for events or their profiles. However, deep learning has not yet been widely used to predict SPEs, because of the limitation of sample size.

In this article, we give an overview of models of flare prediction, CME prediction, and proton event prediction from a machine learning perspective. The paper is organized as follows: Section 2 summarizes the machine learning based modeling process, Section 3 presents the definition of flare prediction, machine learning based modeling process and the progress in flare prediction. Section 4 addresses issues of CME prediction and produces the latest progress in CME prediction. Section 5 describes the definition of SPEs prediction, and the progress in SPEs prediction. Finally, Section 6 discusses some key points of machine learning in establishing prediction models for solar eruptive activities, and Section 7 concludes the paper and provides some future perspectives.

2 Machine learning

2.1 Machine learning development

The development of machine learning can be divided into two stages: traditional machine learning and deep learning. Traditional machine learning, from the perspective of modeling data, mainly deals with low-dimensional data, such as onedimensional data like flux. Therefore, most models require manual feature extraction. The hardware resources for modeling in this stage are typically satisfied by the computing power of a multi-core single machine. In the deep learning stage, from the perspective of data, it can handle higher-dimensional and more complex data, such as image data, video data, etc. In this stage, it is usually possible to automatically extract features and achieve stronger modeling capabilities, but it requires higher training resources, often needing multiple core CPUs, or GPUs, NPU, etc.

As shown in Figure 1, we have listed common models and milestone events. Traditional machine learning, according to the development of neural networks widely used today, can be divided into the early stage, the perceptron era, and the backpropagation era. In the early stage, researchers mainly focused on linear models and statistical methods to explore basic solutions for data classification and regression problems. Alan Turing proposed the Turing test (Turing, 1950) to evaluate whether a machine possesses intelligence, laying the foundation for artificial intelligence. The introduction of simple neural network models (McCulloch and Pitts, 1943) and the proposal of the perceptron (Rosenblatt, 1958) marked the beginning of the perceptron era. The perceptron could solve linearly separable problems, but research progress stalled due to its inability to solve nonlinear problems like XOR. With the proposal of the backpropagation algorithm (Rumelhart et al., 1986), neural network research entered a new stage. The backpropagation algorithm effectively adjusted network weights, solving the training problem of multilayer neural networks and making it possible to solve complex nonlinear problems. During this period, neural

networks achieved significant results in tasks such as speech recognition and image classification, promoting the widespread application and development of machine learning technology.

The deep learning stage can be further divided into the early deep learning period, the convolutional era, the residual era, the attention era, and the large language model era. In the early deep learning period, the main focus was on exploring the structure and training methods of deep neural networks. The convolutional era began with the advent of convolutional neural networks AlexNet ([Krizhevsky et al., 2017](#)) and the use of GPU training. This work and subsequent improvements significantly enhanced the performance of image processing tasks. The residual era, marked by the introduction of residual networks ResNet ([He et al., 2016](#)), solved the gradient vanishing problem in deep network training, making deeper networks possible. The attention era was marked by the proposal of the Transformer model ([Vaswani et al., 2017](#)), which was widely applied to natural language processing tasks and subsequently extended to various other models and tasks. The large language model era saw breakthroughs in natural language processing technology with the introduction of large-scale pre-trained models like BERT and GPT. These models demonstrated strong generative and understanding capabilities, driving the development of artificial intelligence and its applications across various industries.

2.2 Modeling Process

Machine learning and deep learning are used to build models and learn prediction patterns from data, but there are some differences in the modeling process. As shown in Figure 2, the modeling process is compared between machine learning and deep learning. Overall, the modeling process for both machine learning and deep learning involves the steps of problem definition, organizing the dataset, designing the model, training, evaluating, and deploying. The major difference is in feature engineering. Deep learning focuses on automatically learning feature representations from the available data.

2.2.1 Dataset

2.2.1.1 Data collection and pre-processing. Data collection includes the collection of observational data and the calculation of predictors. Data pre-processing includes data cleaning and data labeling. Depending on the modeling method, there may be different cleaning requirements. For example, missing data should be handled in machine learning method. Data labeling is the process of adding one or more meaningful labels to define a prediction problem so that a model can learn from it. For example, we can predict the occurrence of an X-class flare or the soft X-ray peak flux in the next 24 hours.

2.2.1.2 Feature engineering. For machine learning, appropriate feature representations are designed based on the characteristics of the problem and the data. The main problem is what features should be extracted and how to select and combine the features in the prediction model. In deep learning, neural networks automatically learn feature representations from raw data, so there is no need for an explicit feature engineering step, and relatively mature predictors can also be selected based on the actual problem.

Figure 1 The development of machine learning. The horizontal axis represents learning types (supervised learning, self-supervised learning, unsupervised learning), and the vertical axis represents the timeline (including milestone events).

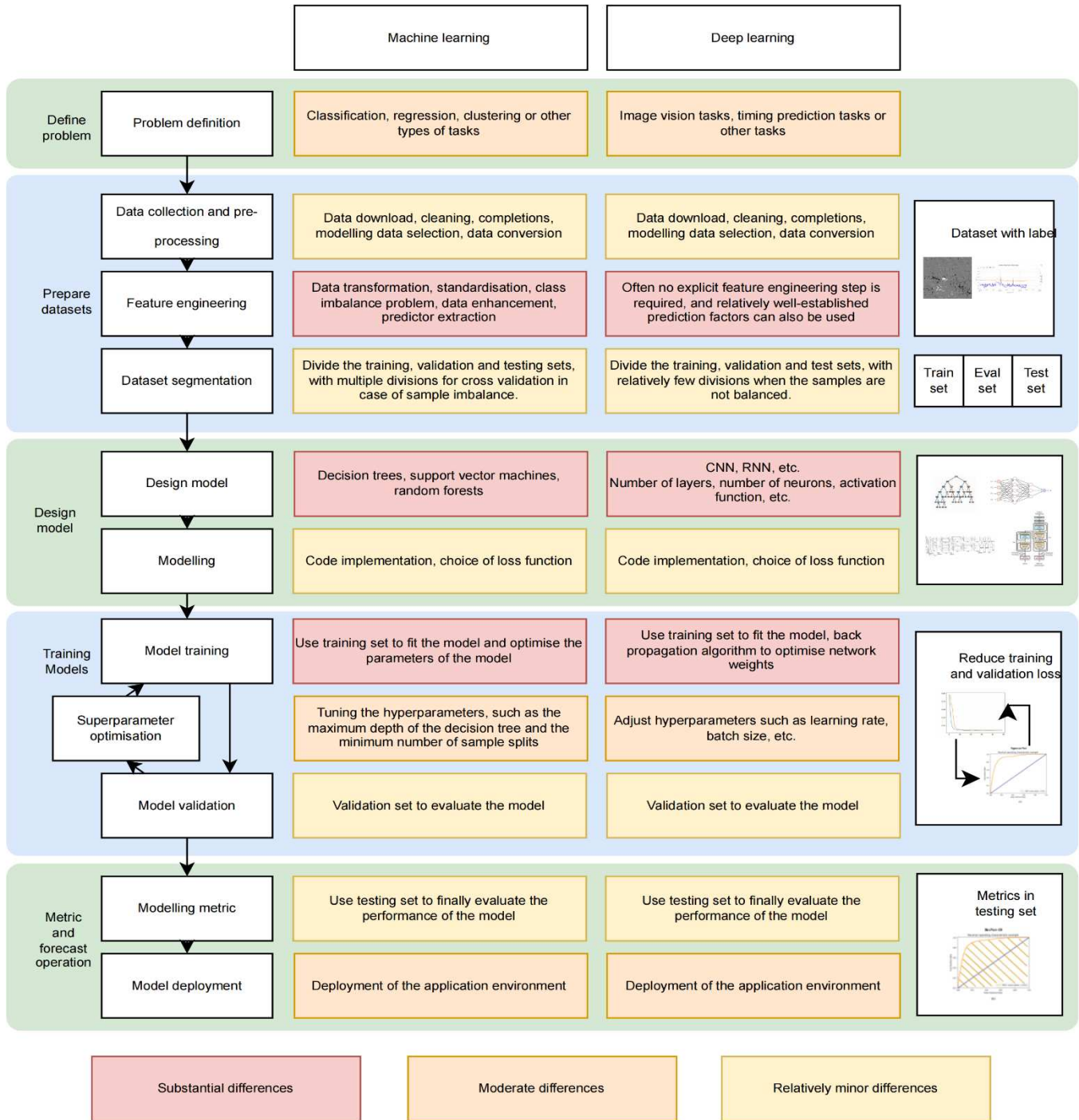


Figure 2 Machine learning and deep learning modeling process and comparison. The machine learning and deep learning modeling process usually consists of five major steps: identifying the problem, organizing the dataset, designing the model, training the model, and evaluating and deploying the model. The major differences are in the phases of feature engineering, designing the model.

2.2.1.3 Data segmentation. Data segmentation is an important step in machine learning, which involves dividing the dataset into training, validation, and testing sets. The training set is applied to learn parameters and train the model. The validation set is used for hyper-parameter tuning and validation of the model to avoid over-fitting. The testing set is used to evaluate the performance of the model. The performance of the model for the testing set is an important metric of its generalization ability. The following are the basic criteria of data segmentation:

Randomness: When the dataset is divided, make sure that the data is randomly disrupted to avoid the impact of the sequential nature of the data.

Data distribution: Ensure that the data distributions of the training, validation and testing sets are similar to those in real applications to ensure the generalization ability of the model.

Data volume: Ensure that there are enough samples in each dataset to represent the various situations and avoid over-fitting.

Time series data: For time series data, make sure that the time period in the training set precedes the validation and testing sets to avoid information leakage problems.

Proportional division: Usually, the dataset could be divided into 70-80% training set, 10-15% validation set, and 10-15% testing set, which are appropriately adjusted according to the data volume and the characteristics of the task. As shown in Figure 3, the dataset is to be divided into a training set and a testing set, with the ratio usually 7:3. The training set is used to train the model, and the testing set is used to evaluate the performance of the model. The training set and the testing set cannot have cross data.

Cross-validation: For smaller amounts of data, cross-validation can be used to more fully utilize the data. For example, k -fold cross-validation divides the data into k portions and uses one of them as the validation set and the other portions as the training set for k validations.

The holdout method and cross validation method are used as model evaluation methods, as shown in Figure 4. The holdout method is a simple method to evaluate the performance of the prediction model. In the holdout method, the dataset is divided into the training set and the testing set, and then the validation set is divided from the training set. The testing score is considered as the generalization measure of the model. The evaluation results of the holdout method is depended on the data partition. The cross validation method can be used to solve this problem to some extent. In the k -fold cross validation method, the training set is split into k subsets. The model is trained by using $k - 1$ subsets, and validated on the remaining one subset. This procedure is repeated k times. The cross validation method seems to give better approximations of generalization, but it is computationally expensive. When the number of samples is very small, the leave one out cross validation method is used. In this method, each sample is used once as a test set, while the remaining samples are used as the training set.

Taking the example of dividing the dataset into the training data and the testing set in solar flare prediction, the data segmentation method is shown in Figure 5. In Figure 5a, one year of data is used for the validation set and the remaining years of data are used for the training set. If multiple divisions are required, data with different year is selected in each time. Data also can be divided by month, for example in Figure 5b, three months data is used for validation and the remaining months for training. If multiple divisions are required, different months could be selected at each time. In Figure 5c, all the data are first divided into different groups according to the flare class. The data selected from different groups is combined to generate the training set and testing set. For example, Li et al. (2020) used a cross-validation method based on Active Regions (AR) data . If multiple divisions are needed, different random seeds can be used for division. Figure 5d shows the simplest division method, which randomly selects a certain percentage of data. However, there may be a problem of data leakage, because of the small amount of data and the periodicity of data distribution and the class imbalance problem. It is important to note that the different samples in the same active region should not appear in both the training set and the testing set.

2.2.2 Model design

The goal of artificial intelligence is to create intelligent systems capable of performing complex tasks. That is, this is also an engineering task, and some classic models and algorithms usually rely on some frameworks. Machine learning code in solar activity prediction rarely is implemented from scratch. Some toolkits, such as sklearn for machine learning, Pytorch, TensorFlow, Mindspore for deep learning, are applied to train the prediction model.

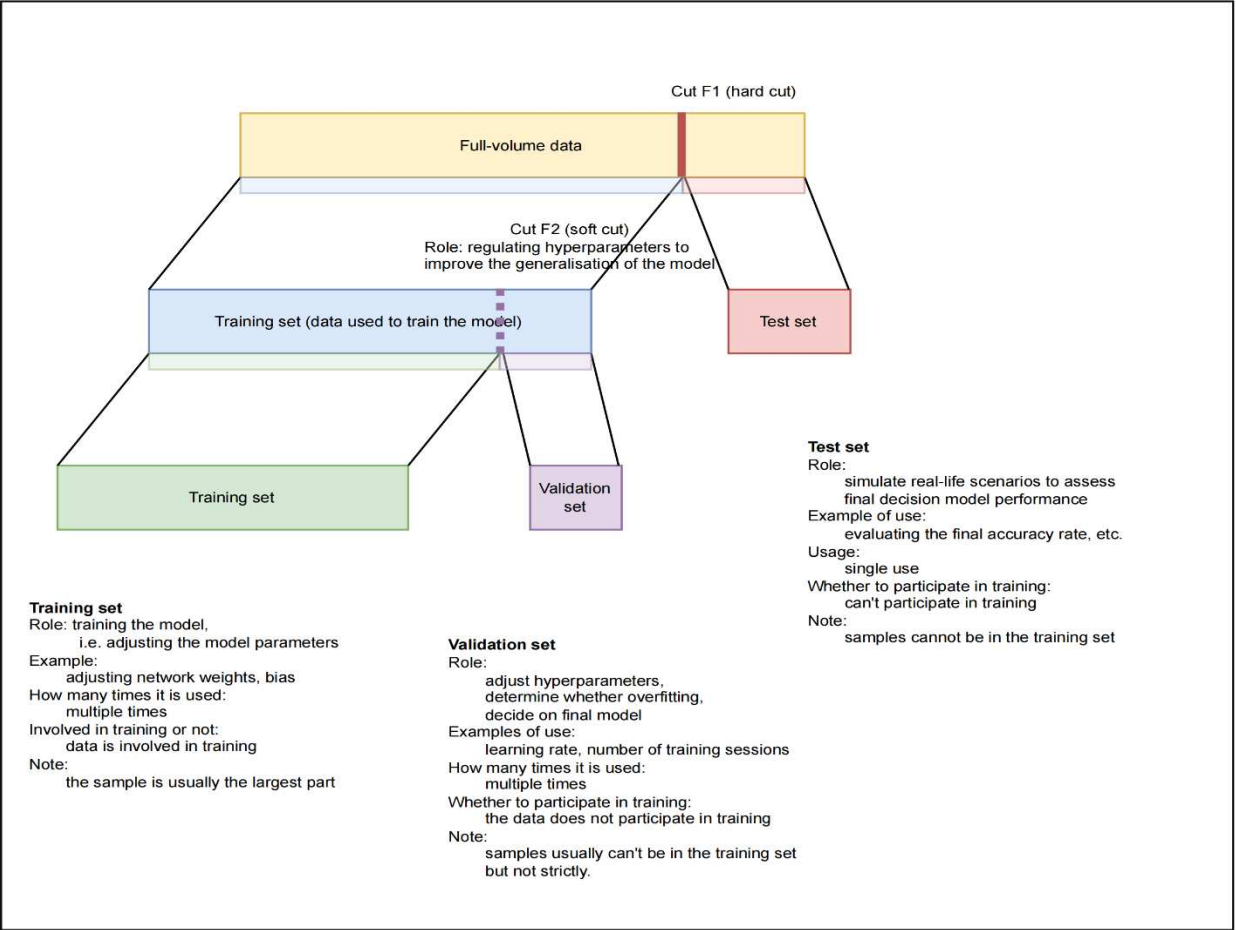
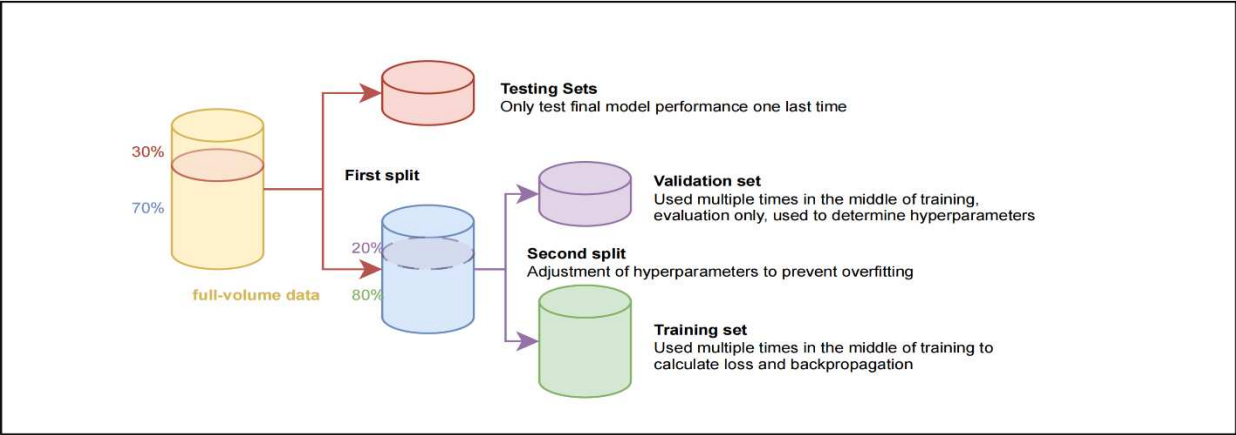


Figure 3 Dataset segmentation. A dataset is usually divided into training, validation, and testing sets.

Figure 4 Holdout method and cross validation. a) is holdout method. b) is k-fold cross validation. c) is leave-one-out cross validation.

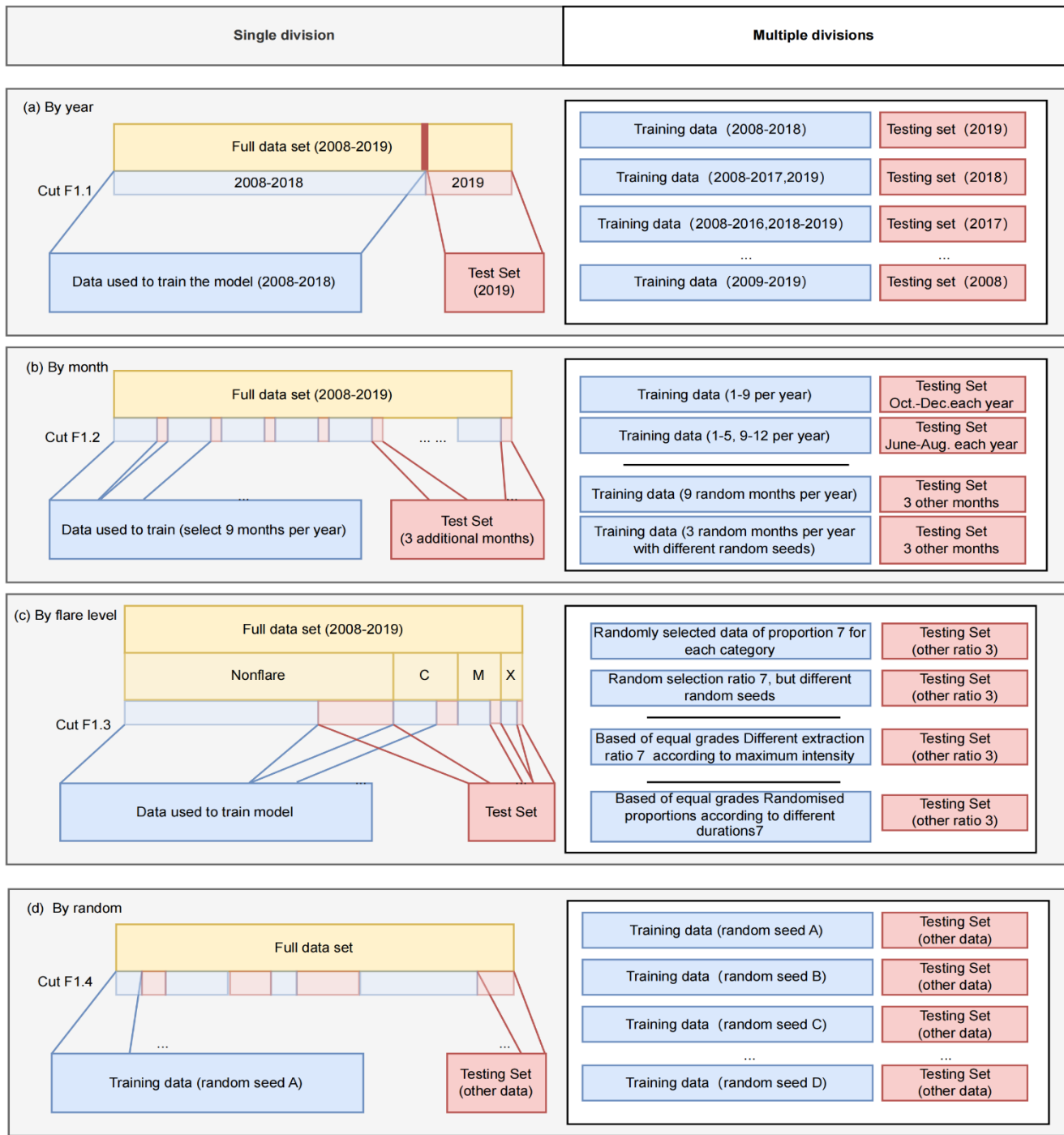


Figure 5 Data segmentation algorithm. (a) aims to split training set and testing set by year. (b) aims to sample training set and testing set by month. (c) aims to split training set and testing set by hierarchical stratification and then stratified sampling. (d) aims to split training set and testing set by random sampling.

2.2.3 Model training

For machine learning, training set is used to fit the model and optimize the parameters of the model so that it can perform well on the training set. For example, the weights of neural network are optimized by a back-propagation algorithm to reduce the loss function on the training set. During the model training process, the hyperparameters of the model are determined according to the validation results.

2.2.4 Model evaluation

Testing sets are used to evaluate the performance of the model. Evaluation metrics are used to estimate the performance of a trained model on a testing set. As shown in Figure 6, Different evaluation metrics are required for the different prediction tasks. The prediction tasks could be considered as classification or regression. The classification is categorized into binary and multi-classification. For regression, the model can output single-point values or multiple-point values.

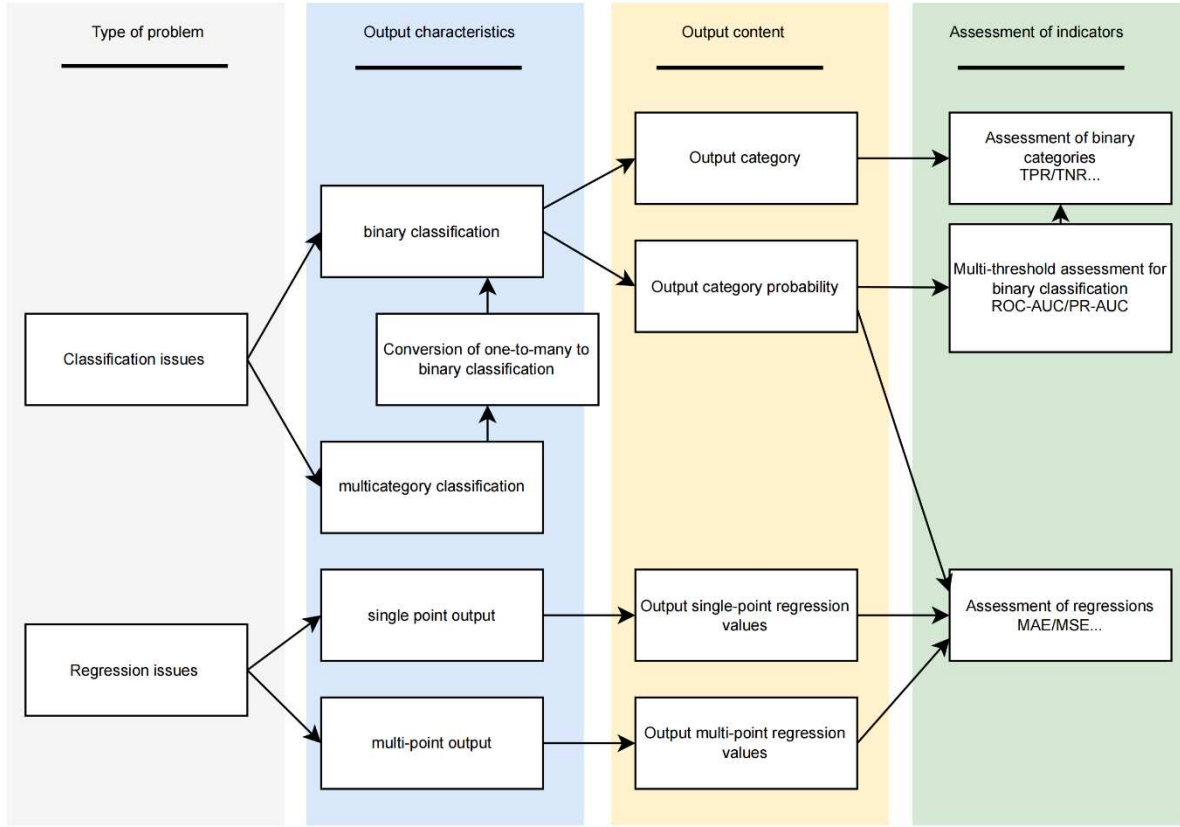


Figure 6 Problem definitions and corresponding metrics.

2.2.4.1 Metrics for binary classification. Binary classification includes positive sample and negative sample. As shown in Figure 7 the calculation of metrics is divided into four steps: sample statistics, statistical summation, calculation of basic metrics and calculation of composite metrics.

The sample statistics include: the number of positive samples that are forecast correctly, called as TP (T means correctly forecast and P means forecast as positive); the number of negative samples that are forecast incorrectly, called as FP (F means incorrectly forecast and P means forecast as positive); the number of positive samples that are forecast incorrectly, called as FN (F means incorrectly forecast and N means forecast as negative); the number of negative samples that are forecast correctly, called as TN (T means correctly forecast and N means forecast as negative).

Statistical summation is to calculate the distribution of data, e.g., the total number of actually positive samples called (P_a) , the total number of actually negative samples called (N_a) . The total number of samples which is forecast to be positive (called

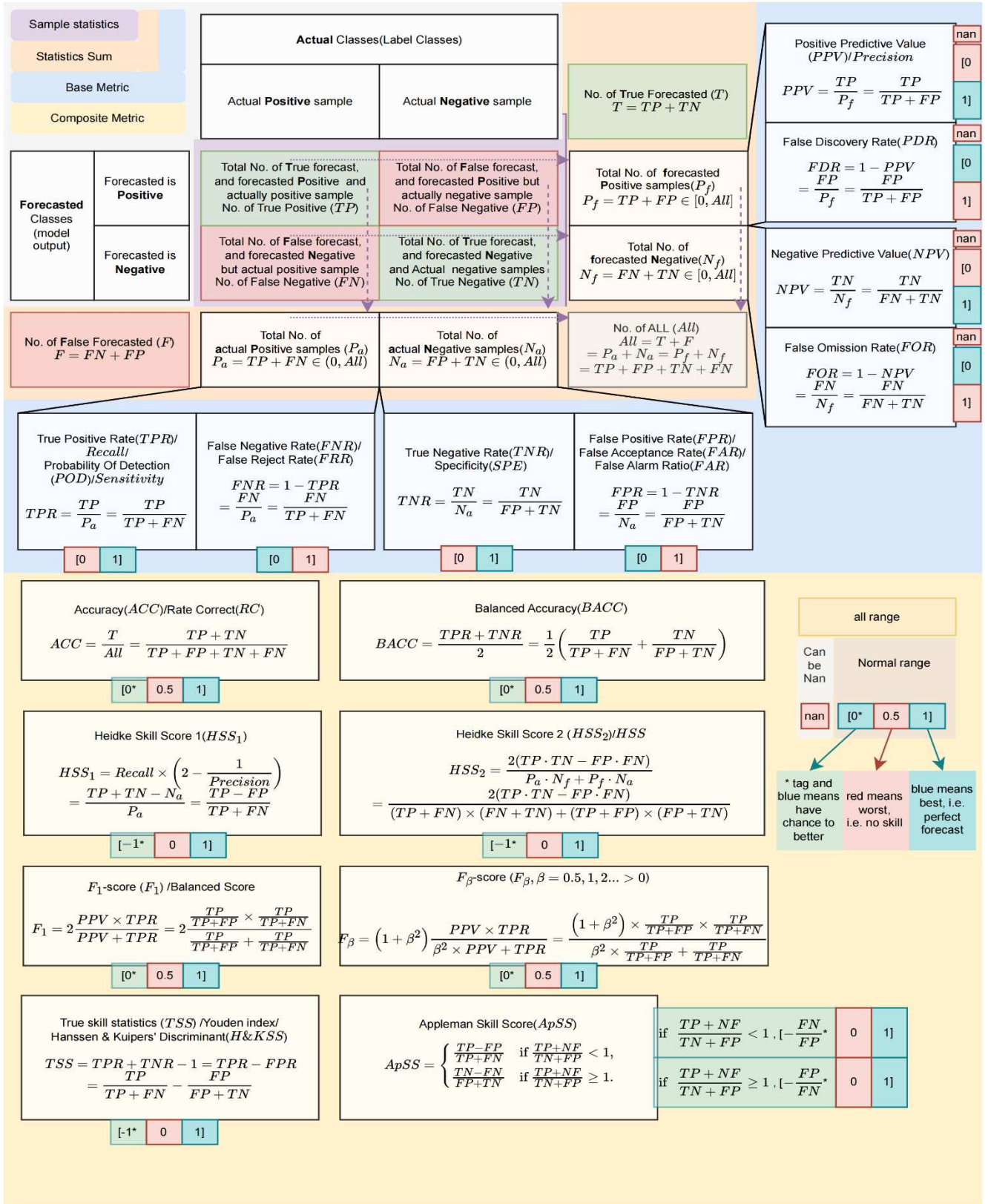


Figure 7 Evaluation metrics for binary classification. The metrics can be divided into four parts, the first part is the confusion matrix in the upper left corner, which shows the basic statistics of the predicted and actual sample cases; then there is the summation of the basic statistics, followed by the calculation of the basic metrics, and finally the calculation of the composite metrics.

P_f) and the total number of samples which is forecast to be negative (called N_f), total number of true forecasts (called T), total number of false forecasts (called F). Total number of samples (denoted as All).

Basic metrics are metrics that can be calculated from the sample statistics and statistical summation. Their calculation methods are shown in Figure 7. The important basic metrics are true positive rate (TPR/Recall/Sensitivity), positive predictive value (PPV/Precision), and false positive/Acceptance/Alarm rate (FPR/FAR). Recall and precision are two fundamental metrics used to evaluate the performance of classification models. Recall measures the proportion of actual positive cases that were correctly identified by the model, while precision measures the proportion of positive forecasts made by the model that were actually correct. For the application scenarios and confidence intervals of these metrics, detailed discussions are provided by (Jolliffe, 2007; Kubo et al., 2017).

Composite metrics are metrics that are calculated from the basic metrics, their calculation methods are shown in Figure 7. When data is balanced, accuracy (ACC) is good metric. However, the number of non-flaring samples is much larger than the number of flaring samples. This is a class imbalance problem. In class imbalance problem, Balanced Accuracy (BACC), Heidke skill scores (Heidke, 1926) (HSSs), True skill statistics (Bloomfield et al., 2012) (TSS), Appleman Skill Score (ApSS), F_1 scores and F_β scores are used to evaluate the model performance (eg. Jiao et al. (2020)). Heidke skill scores (HSSs) are defined in two ways, HSS1 (Barnes and Leka, 2008) and HSS2 (Bobra and Couvidat, 2015; Mason and Hoeksema, 2010). HSS2 is often used to estimate the performance of the solar activity prediction model (Bobra and Couvidat, 2015; Chen et al., 2019).

2.2.4.2 Metrics for multi-threshold binary classification. When the output of the prediction model is probability, the threshold can be manually set. The samples, whose probabilities are larger than the threshold, are considered as positive samples, and other samples are considered as negative samples. The different thresholds produce different outputs. The Receiver Operating Characteristic (ROC) curve and Precision-Recall (PR) curve are used to estimate and compare the model performance when do multi-threshold binary classification, eg. Florios et al. (2018), Liu et al. (2022). The horizontal axis and the vertical axis of the ROC curve are false positive rate (FPR) and true positive rate (TPR), respectively. Selecting different thresholds, a set of (FPR, TPR) value pairs are obtained, and the ROC curves are obtained by concatenating all the value pairs as shown in Figure 8a. The closer the curve to the upper left corner, the better the model performance. To quantitatively characterize how good the model is, the Area Under ROC Curve (AUC) is calculated (Figure 8b).

For the class imbalance problem, the precision-recall (PR) curve is another choose, e.g. Jonas et al. (2018) use the PR curve to evaluate the model performance. The horizontal axis and vertical axis are true positive rate (TPR) and positive predictive value (PPV), respectively. The closer the curve to the upper right corner, the better the curve performance.

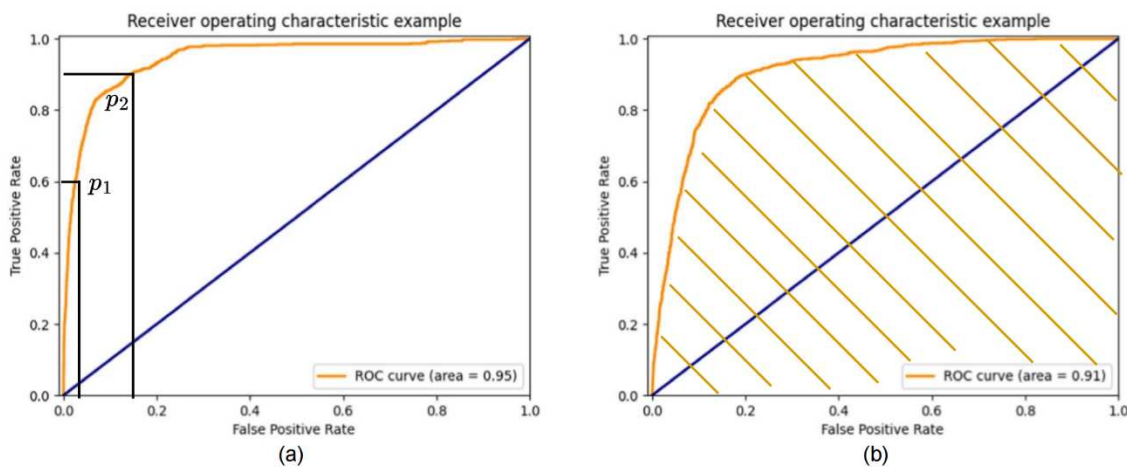


Figure 8 ROC curve.

2.2.4.3 Metrics for regression. Regression model outputs continuous variables, which are commonly used in probabilistic prediction, single-valued prediction, and multiple-valued prediction. The metrics in regression is shown in Figure 9.

Data category	Data value	Total amount of data	Sum of data	Arithmetic mean
Actual Values (Label Values)	$v_{a.all} = v_{a1}, v_{a2}, v_{a3}, \dots, v_{an}$	n	$v_{a.sum} = \sum_{i=1}^n v_{ai}$	$v_{a.avg} = \frac{v_{a.sum}}{n}$
Forecasted Values (model output)	$v_{f.all} = v_{f1}, v_{f2}, v_{f3}, \dots, v_{fn}$	n	$v_{f.sum} = \sum_{i=1}^n v_{fi}$	$v_{f.avg} = \frac{v_{f.sum}}{n}$

Mean absolute error (MAE)/L1 loss	Mean squared error (MSE)/Brier Score (BS)	Root Mean squared error (RMSE)
$MAE(v_{f.all}, v_{a.all}) = \frac{1}{n} \sum_{i=1}^n v_{fi} - v_{ai} $	$MSE(v_{f.all}, v_{a.all}) = \frac{1}{n} \sum_{i=1}^n (v_{fi} - v_{ai})^2$	$RMSE(v_{f.all}, v_{a.all}) = \sqrt{\frac{1}{n} \sum_{i=1}^n (v_{fi} - v_{ai})^2}$

Brier Skill Score (BSS)	the coefficients of determination (R^2)	data	statisticians
$BSS(v_{f.all}, v_{a.all}) = 1 - \frac{MSE(v_{f.all}, v_{a.all})}{\frac{1}{n} \sum_{i=1}^n (v_{ai} - v_{a.avg})^2}$	$R^2(v_{f.all}, v_{a.all}) = 1 - \frac{\sum_{i=1}^n (v_{ai} - v_{fi})^2}{\sum_{i=1}^n (v_{ai} - v_{a.avg})^2}$	evaluation	

Figure 9 Metrics for regression problems.

3 Solar Flares Prediction

3.1 Problem definition

A solar flare is one of the most violent solar activities. A solar flare is a sudden and large-scale release of energy that occurs in a localized region above the surface of the Sun, as shown in Figure 10. Flares are usually accompanied by a series of phenomena: radio enhancement in almost the full range of wavelengths, shower of highly energetic particles with energies ranging from 10^3eV to 10^{11}eV , and coronal mass ejections (Lin, 2000). However, due to the different conditions of formation, not all of these phenomena occur, nor do they occur with exactly the same intensity.

3.1.1 Flare size and classification

The energy of a large flare can reach $4 \times 10^{32}\text{erg}$; of which the electromagnetic radiation part accounts for about 1/4, i.e., 10^{32}erg , and is mainly concentrated in the visible wavelength. The remaining 3/4 of the energy is released in the form of energetic particles and plasma kinetic energy (Ellison, 1963; F. C. Durant, 1965; Lin, 2000; Wentzel and Tidman, 1969). A medium-size flare releases energy at $10^{30} \sim 10^{31}\text{erg}$, and a small flare releases energy at $10^{28} \sim 10^{29}\text{erg}$. Microflares have energy of $10^{26} \sim 10^{27}\text{erg}$, the release of $\leq 10^{25}\text{erg}$ is called Nanoflares.

Flare classification is usually divided into optical flare classification and flare classification in soft X-rays. Flare classification using soft X-rays is now more commonly used. The Earth's ionosphere is sensitive to changes in the intensity of $1 \sim 8 \text{ \AA}$ soft X-rays, so the solar flare X-ray level reflects the ability of that flare to produce ionospheric effects. Based on the $1 \sim 8 \text{ \AA}$ peak flux F_X measured by GOES satellites ((Machol et al., 2020)), flare soft X-ray levels are bundled into A, B, C, M, and X. As shown in Table 1, each major level is subdivided further into nine minor levels. For example, the M3 level indicates the flare soft X-ray peak flux $F_X = 3 \times 10^{-5} \text{ Watts} \cdot \text{m}^{-2}$.

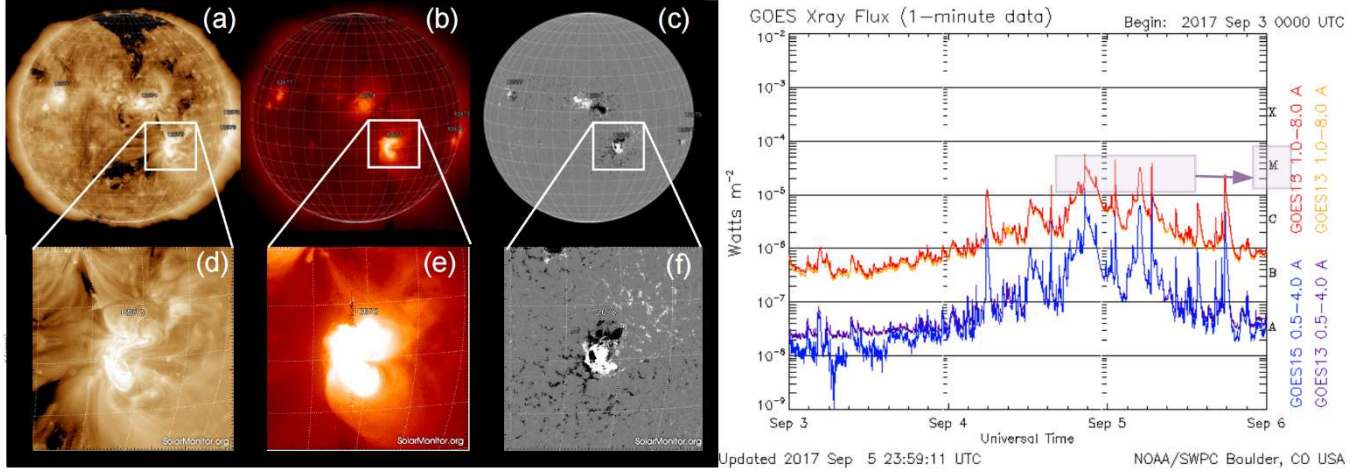


Figure 10 Example of flare. Panels (a)-(f) extracted from the Solarmonitor database (Gallagher et al., 2002), (a)-(c) showing solar full-disk images for September 5, 2017; (d)-(f) are an active region of (a)-(c) that flare with an M-class flare and National Oceanic and Atmospheric Administration (NOAA) Number is 12673; (a) and (d) are Atmospheric Imaging Assembly's (AIA) 193Å observational data at 19:27; (b) and (e) are Hinode X-ray Telescope's (XRT) observational data at 05:55; (c) and (f) show HM magnetograms at 00:46. Right figure is extracted from NOAA, Space Weather Prediction Center (NOAA/SWPC), which is the soft X-ray flux from September 3 to 6, 2017.

Table 1 Soft X-ray classifications of solar flares

Level	$F_X (\text{Watts} \cdot \text{m}^{-2})$
	$[10^{-8}, 10^{-7})$
	$[10^{-7}, 10^{-6})$
	$[10^{-6}, 10^{-5})$
	$[10^{-5}, 10^{-4})$
	$[10^{-4}, \infty)$

3.1.2 Flare Prediction Labels

Short-term solar flare prediction models typically provide results 1-3 days in advance. The prediction results include whether a flare will occur, the class of the flare, the time and location of the flare, and whether it will be accompanied by other solar activity events.

Prediction models usually include two types. One type is defined as a classification problem, predicting discrete values, such as whether a flare will occur in the future. The other type is a regression problem, providing continuous values, such as the peak soft X-ray flux or the flare duration (Reep and Barnes, 2021). The type of data labeling determines the type of modeling problem. Common prediction models include classification problems and regression problems, as shown in Figure 11. Classification problems predict discrete labels, while regression problems predict continuous labels.

3.2 Data and predictor

3.2.1 Data

Machine learning approaches require big data. Continuous multi-wavelength data is an ideal data source. Observational equipment on the satellite is naturally characterized in this way. Solar activity prediction relies on data from the satellites, e.g. see the following popular examples. Solar and Heliospheric Observatory (SOHO; (Domingo et al., 1995)) is a solar satellite established by the European Space Agency (ESA) in cooperation with the National Aeronautics and Space Administration (NASA) of the United States of America, which was launched in 1995. Solar Dynamics Observatory (SDO; (Pesnell et al., 2012a)) is another satellite launched by NASA in 2010, and carries several instruments, which provide high temporal and spatial resolution images of the Sun. Geostationary Operational Environmental Satellite (GOES) comprises a series of satellites operating

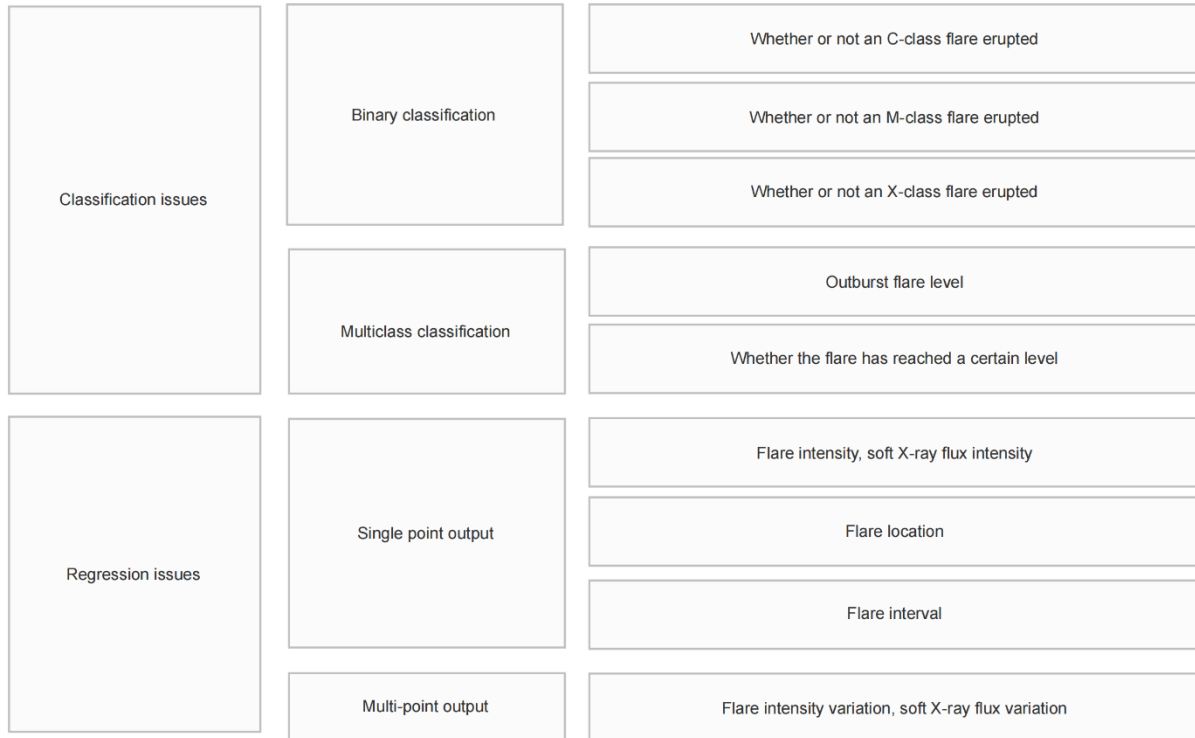


Figure 11 Classification and regression problems in flare forecasting

in geosynchronous orbit for monitoring the effects of solar activity on the Earth's environment. The accumulated data of those satellites can be used to build the machine learning models. The magnetogram and multi-band images can be used as the input of prediction model, and the soft X-ray flux can be used to provide the output of the prediction model.

3.2.1.1 Magnetic field data. The SOHO/MDI (the Solar and Heliospheric Observatory/the Michelson Doppler Imager) and SDO/HMI (the Solar Dynamics Observatory/the Helioseismic and Magnetic Imager) provide continuous photospheric magnetograms from 1996 and 2010, respectively. Because of the accumulation of big data, they are used to train the machine learning models. For example, Park et al. (2018) applied CNN to the full-disk solar magnetograms without any preprocessing or feature extraction, their model inputs are SOHO/MDI (from 1996 May to 2010 December) and SDO/HMI (from 2011 January to 2017 June) full-disk magnetograms at 00:00 UT. Bobra and Couvidat (2015) used vector magnetograms parameters to forecast solar flares by SVM.

The most solar eruptions occur in the lower solar atmosphere, i.e. in the chromosphere and low corona (Jiang et al., 2021). But, the coronal magnetic field is not yet routinely observed. To overcome this shortage, based on the photospheric magnetogram, the coronal magnetic field is usually calculated by the extrapolation or simulation approach. Aschwanden (2020)

uses HMI's magnetograms and AIA's EUV wave projections with automatically tracked coronal loop coordinates to simulate the force-free coronal magnetic field. Based on the HMI SHARP (Bobra et al., 2021) data, Zhao et al. (2023b) built a large-scale database using nonlinear force-free field extrapolation algorithm. Zhao et al. (2023a) proposed a new mutually embedded perception model (MEPM) based on the 3D magnetohydrodynamic (MHD) equations of the solar wind plasma to reconstruct the structure of the solar corona. Campi et al. (2019) extracted parameters based on 2D magnetograms and 3D potential field extrapolation data, and built a prediction model. Korsós et al. (2020a) used a potential field algorithm to extrapolate the coronal magnetic field and extracts the parameters to probe the optimal height in the upper chromospheric and low coronal magnetic field for the solar flare prediction.

3.2.1.2 Multi-wavelength image. The white-light data contains the information of sunspot, hence it can be used as input of solar flare prediction model (Pesnell et al., 2012b). Abed et al. (2021) automatically extracted features from sunspot images to predict the solar flare level for the next 24 hours. The SDO/AIA provides solar images from photosphere to corona. Liu et al. (2017b) used the Zernike moments method to extract features from extreme ultraviolet images. Krista and Chih (2021) used statistical methods for modelling based on EUV images. Nishizuka et al. (2017) built prediction model by using magnetograms and ultraviolet images. Sun et al. (2023) built the solar flare prediction model by fusing the information from AIA multi-band images.

3.2.1.3 Standard dataset. In order to compare the performance of prediction model built by different methods, it is important to generate a standard dataset. However, solar flare prediction is still an open question, there is not a standard dataset yet defined and established. Some public datasets (Angryk et al., 2020; Zhao et al., 2023b), however, have already been openly constructed.

3.2.2 Predictors

The sunspot parameters or magnetic parameters of active regions can be calculated as the predictors in solar flare prediction model.

3.2.2.1 Sunspot parameters. A sunspot is a localized region of strong magnetic field on the surface of the Sun (Lin, 2000). The characteristics of sunspots are closely related to solar activities (Wang, 2009; Yang et al., 2010; Yin et al., 2007). The complexity of sunspots is also closely related to solar flares (Rong et al., 2023).

In 1919, the Mount Wilson Observatory proposed the Mount Wilson classification by the distribution of magnetic field polarity (Hale et al., 1919). The morphology of the sunspot groups evolved roughly according to the Zurich classification (also known as the Brunner classification), which is, the evolution of the sunspot group usually changes from simple to complex and then to simple again. Based on the Zurich classification of the sunspot group, McIntosh (McIntosh, 1990) proposed the McIntosh classification of the sunspot group with three letters. The first letter of the alphabet still uses the Zurich classification, but combines nine categories (A, B, C, D, E, F, G, H, J) into seven (A, B, C, D, E, F, H). The second letter denotes the shape and complexity of the largest sunspot in the population, which are classified into six categories (x, r, s, a, h, k) according to the penumbra. The third letter denotes the distribution of sunspots in the sunspots group, which is divided into four categories (x, o, f, c).

Related researchers have elaborated on the relationship between the sunspot classifications and flares. Sammis et al. (2000) investigated the relationship between Mount Wilson classification and flares based on 8 years of observations of sunspots, and found that complex magnetic fields are more likely to produce flares. Although large sunspot areas tend to produce large flares, the effect of magnetic classification on solar flare prediction is more pronounced. McIntosh (McIntosh, 1990) stated that the correlation between McIntosh classifications and flares is better than the correlation between Zurich classifications and flares. Atac (1987) synthesized the effects of the McIntosh classifications and the Mount Wilson classifications on flares, and pointed out that sunspots with Dki or Eki, with δ and with maximum magnetic field strengths of 1600-2500 Gauss are more prone to large flares. Falco et al. (2019b) assumed that the flare event frequency follows Poisson statistics and estimated the probability of solar flare prediction model based on the sunspot parameters. Li et al. (2013) has established a flare model by combining the photospheric magnetic field parameter and the sunspot parameter, and the predicting results show that the combined parameters is more effective than using the photospheric magnetic field parameter and the sunspot parameter as the predictors alone. A

series of sunspot databases and on-line tools were compiled at Debrecen Heliophysical Observatory (DHO)([Baranyi et al., 2016](#)).

3.2.2.2 Magnetic field parameters. The sunspot is a manifestation of a strong magnetic field. The most solar flares occur in active regions with complex magnetic field structures. The energy of a flare comes from the magnetic field. Therefore the study of the relationship between flares and magnetic fields is very important for understanding the physical processes of flares or CMEs ([Bobra and Ilonidis, 2016](#); [Georgoulis, 2008](#); [Georgoulis et al., 2019, 2012](#); [Leka and Barnes, 2003](#); [Leka et al., 2012, 2005](#); [Lin, 2000](#); [Rong et al., 2023](#)). The physical parameters and morphological parameters can be extracted from the magnetic field data of active regions.

3.2.2.2.1 Physical predictors. Many magnetic field parameters are proposed to characterize the non-potentiality and complexity of active regions. [Gallagher et al. \(2002\)](#) found that the active regions with significant magnetic field gradient are prone to occur solar flares. [Abramenko et al. \(2002\)](#) found that the dissipation spectrum corresponding to the B_z component of the photospheric magnetic field is flatter when the flare productivity is high. [Komm and Hill \(2009\)](#) found that large vorticity in the active region may produce solar flares. [Leka and Barnes \(2003\)](#) investigated the relationship between a number of photospheric magnetic field parameters and solar flare, such as the horizontal gradient of the magnetic field, the vertical current etc. [Georgoulis and Rust \(2007\)](#) defined an effective connecting magnetic field B_{eff} , which is a robust criterion for identifying flaring and non-flaring active regions. [Cui et al. \(2007, 2006\)](#) calculated the maximum horizontal gradient of the magnetic field, the length of the neutral line, and the number of singular points from photospheric magnetograms of active regions, and the statistical relationships between these magnetic field parameters and flares were analyzed. [Wang et al. \(2008\)](#) developed a shortterm flare prediction system based on three predictors extracted from the magnetograms, the maximum horizontal gradient of the longitudinal magnetic field, the length of the neutral line and the number of singular points. [Cicogna et al. \(2021\)](#) proposed a topological parameter D to represent the complexity of the active region. [Kutsenko et al. \(2021\)](#) concluded that a high flux emergence rate is a necessary condition for an active region to produce strong flares in the future.

Although lots of magnetic parameters are proposed, most of them are correlating with each other. [Leka and Barnes \(2007\)](#) used statistical methods to find strong correlations between some of the parameters and the state of the photospheric magnetic field at any given time has limited bearing on whether that region will be flare productive. [Barnes and Leka \(2008\)](#) quantified total flux Φ_{tot} , the free energy E_e , total unsigned flux high-gradient polarity-separation lines R ([Schrijver, 2007](#)), and the effective connecting magnetic field B_{eff} . The discriminant analysis is used to build solar flare prediction model, and they found that there was no significant difference in the predicting ability of these predictors. [Dhuri et al. \(2019\)](#) and [Lim et al. \(2019b\)](#) also discussed the correlation among the magnetic parameters. The de-correlation operations should be performed when these predictors are applied to build a solar prediction model. On the other hand, [Wang et al. \(2019a\)](#) and [Wang et al. \(2020\)](#) found that their ability to distinguish strong-flaring ARs and non-strong-flaring ARs can also be improved when one calculates them concentrating on high-gradient neutral regions instead of the entire ARs. Some data products, eg. [Bobra et al. \(2014\)](#), added magnetic parameters in the header file of the release data. These parameters can be used in machine learning based modeling process as well. In addition to the solar photospheric magnetic field parameters, [Huang et al. \(2013\)](#) added the distance between active regions and predicted active longitudes as a new predictor, and verified its effectiveness for solar flare prediction.

3.2.2.2.2 Fractal predictors. The solar active regions with much of the energy are the sources of solar flares and coronal mass ejections. Various approaches have been attempted to characterize the complexity of the active regions. From the Mount Wilson magnetic classification to the McIntosh classification, it is indicated that large and complex active regions produce a large numbers of solar flares. However, these sunspot classifications are subjective. Since active regions are characterized by self-similarity, the complexity of active regions can be quantified by fractal methods ([McAteer et al., 2005](#); [Rong et al., 2023](#)). Relative to morphological classifications of sunspots and the magnetic parameters of active regions, the fractal predictors are relatively little studied at present. [McAteer et al. \(2005\)](#) used a counting box method to study the relationship between the calculated fractal dimension and flare productivity for the Wilson classification of each active region. They found that each Mount Wilson classification has a similar fractal dimension frequency distribution, thus indicating self-similarity of active

regions. [Aschwanden and Aschwanden \(2008a,b\)](#) investigated the area fractal dimension and volume fractal dimension of solar flares, and found that the flare region observed at EUV wavelengths exhibits fractal characteristics, and total EUV flux is linearly correlated with the area fractal. [Meunier \(2004\)](#) investigated the fractal dimension variability at different time scales during the solar cycle. [Georgoulis \(2012\)](#) found that both flaring and non-flaring active regions exhibit significant fractality, multifractality, and non-Kolmogorov turbulence but none of the three tested parameters manages to distinguish active regions with major flares from flare-quiet ones. Other researchers ([Conlon et al., 2008](#); [Georgoulis, 2013](#); [Lawrence et al., 1993](#); [Yu et al., 2012](#)) have studied the solar magnetic field and solar flares by using the multi-fractal method.

3.3 Model

The modeling approaches can be divided into physical approaches and data-driven approaches. Data-driven approaches include statistical methods, machine learning methods, and deep learning methods. Statistical methods build the prediction model based on low-dimensional feature parameters. Machine learning methods require manual feature extraction as well, but they can automatically build the prediction model. Deep learning methods can extract features and build the prediction model automatically, supported by the the big data.

3.3.1 Physical approach

Physical approaches include magnetohydrodynamics (MHD) model and self-organizing criticality (SOC) model which was based on principle of firstness ([Cozad et al., 2015](#)).

3.3.1.1 MHD. The magnetohydrodynamic (MHD) equations can be used to dynamically model electromagnetic fields and plasma in time and space ([Piana et al., 2019](#)). On the basis of understanding the physical processes and formation mechanisms of changes in the solar physics, a physical model is built, the initial and boundary conditions are given, and the spatial distribution and evolution of predictors are calculated using numerical simulation ([Rong et al., 2023](#)). Since the physical mechanism of solar flare is still unclear, no applicable and predictable physical model has been established ([Korsós et al., 2018](#); [Lin, 2009](#); [Ning, 2009, 2012](#); [Rong et al., 2023](#); [Wang, 2012](#)).

3.3.1.2 SOC. Self-organized criticality (SOC) is proposed in 1987 ([Bak et al., 1987](#)). A system consisting of a large number of interacting components will naturally evolve towards a self-organized critical state, and when the system reaches the selforganized critical state, even small disturbing events can cause a series of drastic changes in the system. "Self-organization" means that the formation of the state is primarily the result of interactions between organizations within the system, rather than being controlled or dominated by any external factors. "Critical state" means that the system is in a particularly sensitive state, where small local changes can be amplified and extended to the whole system. In other words, when the system is in the critical state, the behaviors of all the components are correlated with each other, and there is a power law relationship between the magnitude of the events and their frequencies in the system ([Bélangier et al., 2007](#)).

A simple example of SOC is the sand pile model ([Aschwanden, 2011](#)). In this model, sands are added to the pile, the average slope of the pile increases until it reaches critical state. Once the critical state is reached, the additional sands will cause an avalanche. It is well known that the frequency of solar flares and their peak intensities follow a power law that is independent of the solar cycle. [Lu and Hamilton \(1991\)](#) suggested that the coronal magnetic field is analogous to a sandpile. The slow generation of photospheric convective motion with random torsion acts as additional sand in the sandpile. Solar flares are triggered by many small events. Magnetic energy stored in the coronal magnetic field is usually considered to be the energy source of flares. Each magnetic reconnection event releases energy from the magnetic field, and through the energy release, the gradient of the magnetic field is less than the critical state value. When another magnetic energy is added, and the magnetic field is driven to a critical state. The process is repeated ([Rong et al., 2023](#)). [Karakatsanis and Pavlos \(2008\)](#) pointed out that the co-existence of two physical processes in the solar activity, the one is the existence of self organized critical state, the other is existence of a low dimensional chaotic dynamics. [Morales and Santos \(2020\)](#) has conducted a similar study. The application of SOC model to solar flare prediction is still in its early stage.

3.3.2 Statistical methods

In the statistical methods, the lower dimensional predictors are applied to build the prediction model estimated the model parameters from the observation data (Giovannelli, 1939). There is a limited dimension of input data and no way to input images directly into the model. Giovannelli (Giovannelli, 1939) proposed the earliest probabilistic prediction model for solar flares. The probability of a flare in an active region is expressed as $P = 2.6akf(i)$, where a is the sunspot area, and k is probability of an eruption per area and $f(i)$ represents the rate of area increase. Drake (1971) proposed a power rate distribution of flare peak flux, which is denoted as $N(S) = \text{const} \times S^\gamma$, where $N(S)$ is the number of events per unit of time per peak flux S , the const denotes a constant. Usually the power exponent γ is slightly smaller than 2 (Aschwanden et al., 1998). Based on 3 parameters of McIntosh classifications, Xin (2010) built a flare prediction model by using a Poisson distribution. The probability of solar flares in a 24-hour period is $P_\mu(N \geq 1) = 1 - \exp(-\mu)$, where μ is the average number of events in per unit of time. Some classical statistical methods, for example Bayesian methods, multiple linear regression, and Fisher's discriminant analysis, are commonly used to build the solar flare prediction model. Wheatland (2004) applied Bayesian event statistics methods to predict the solar flare, based on the GOES soft X-ray data from 1976 to 2003. Bornmann and Shaw (1994) analyzed the contribution of McIntosh classifications to solar flares by using multiple linear regression. Bartkowiak and Jakimiec (1994) built the solar flare prediction model by using multiple regression methods, and they found that the performance of using L1 norm or Gower distance metrics is better than using Euclidean distance. Using the Fisher's discriminant analysis, Barnes and Leka (2006) gave the statistical relationship between magnetic field parameters and flares. Zhang et al. (1994) developed a solar flare prediction model, a multiple linear regression method was applied to get better solar flare predicting results in (Zhu and Wang, 2003). Falco et al. (2019a) has used blackout parameters to build prediction model based on statistical methods. Gyenge et al. (2016) based on the statistical found that main active longitude plays a crucial role in the global position of solar flare occurrence. Park et al. (2020) evaluate consecutive-day predicting patterns. There are also analytical studies based on physical parameters of case active region, such as Korsós et al. (2020b) which studied the role of weighted horizontal magnetic gradients in solar flare events based on typical active regions.

3.3.3 Machine learning methods

The data, model and interpretability are focused in machine learning based solar flare prediction model.

3.3.3.1 Data preprocessing. Feature selection and class imbalance problem are critical to the success of a machine learning task.

3.3.3.1.1 Feature selection. In machine learning algorithms, too many highly correlated features can lead to degradation of model performance, so it is necessary to select important features for the prediction model. This process is known as feature selection. The feature selection can be used to reduce the feature dimension, improve generalization performance of the model, and reduce training time and resource consumption. Classical feature selection methods are as follows:

Filter Methods: Filter methods select features through statistical metrics (variance, mutual information, chi-square test and so on) or feature importance, Filter methods are independent of the model, so their computational efficiency is high. Ahmed et al. (2013) selected the effective predictors by filter methods algorithm (Correlation-based Feature-Selection and Minimum Redundancy Maximum Relevance), and applied the neural network to predict the flare.

Wrapper Methods: Wrapper methods use performance evaluation of the model to determine the importance of feature subset. In wrapper methods, the subset of features is selected by training and evaluating the performance of the model on this subset. Optimal algorithms (recursive feature elimination or genetic algorithm) are applied to find the suitable combination of the features.

Embedded Methods: In embedded methods, feature selection is embedded into the model training process. By considering the importance of features in the model training process, the model can automatically select important features. For example,

decision tree and random forest etc. can automatically select the important subset of features during training process. [Ribeiro and Gradwohl \(2021\)](#) used random forest to select important features.

3.3.3.1.2 Class imbalance problem. Class imbalance means that the number of samples with some classes is much fewer than the number of samples with other classes. For example, the number of flaring samples is much smaller than the number of non-flaring samples. The model, built on the class imbalance data, could tend to provide predicting results where the minor class samples are under-performed. Commonly used methods to deal with class imbalance are as follows:

Resampling: Resampling is to adjust the classes distribution of the dataset by increasing the number of samples with minority class or decreasing the number of samples with majority class. Resampling methods include undersampling and oversampling. In undersampling method, the number of samples with majority class is removed so that it is close to the number of samples with the minority class. It could lead to a risk of information loss or model over-fitting. The oversampling method is commonly used in solar flare prediction. In the over-sampling method, the number of samples with the minority class is increased. For example, [Wan et al. \(2022\)](#) proved that the fusion model based on resampling and the CNN-GRU algorithm is more suitable for solar flare prediction. [Wan et al. \(2023\)](#) adopt the density clustering method combined with the Synthetic Minority Over-sampling Technique (SMOTE) oversampling method, and results proved that the method designed can be well applied to the flare prediction problem. [Liu et al. \(2023\)](#) found that the selective up-sampling method has potential to improve the model performance in strong-flare prediction.

Weighting: When training the model, different weights can be assigned to different samples to deal with the class imbalance problem. The samples with minority class are given large weights to emphasize their impact on the model during the training process. For example, [Deshmukh et al. \(2022\)](#) used loss function weighting to compensate for training set imbalance.

[Wan et al. \(2021\)](#) summarized the characteristics of a series of methods that are typically used to solve the problem of class imbalance. They found that there is not a perfect algorithm that performs best in all aspects. Depending on the particular problem and dataset, it may be necessary to try multiple methods. The selection of the appropriate method should be based on the distribution of data and the assessment of model performance.

3.3.3.2 Models. At an early stage of solar flare prediction, the NOAA's Space Environment Laboratory and the University of Colorado have jointly developed an expert system 'Theo' to predict solar flares based on McIntosh classifications ([McIntosh, 1990](#)). [Fozzard et al. \(1988\)](#) built a solar flare prediction model based on artificial neural network. The same predictors as 'Theo' are used in this model and the same testing accuracy is achieved. [Qahwaji and Colak \(2007\)](#) applied neural network and support vector machine algorithms to establish a hybrid flare short-term prediction system. Based on the McIntosh classifications and sunspot number, they first used support vector machine to predict whether there is a flare, then they used neural network to predict which level of flare is. [Barnes and Leka \(2006\)](#) investigated the parameters extracted from the photospheric magnetic field and applied the Fisher linear discrimination to predict flare occurrence.

[Huang et al. \(2012\)](#); [Yu et al. \(2009a,b, 2010\)](#) applied several machine learning methods to build a solar flare prediction model. The testing results evaluate the effectiveness of the evolutionary information of the magnetic time-series parameters in solar flare prediction. [Cinto et al. \(2020\)](#) proposed the standardization to design the solar flare prediction model. The machine learning-based process is described, and the extreme gradient boosting tree classifiers is applied to build the prediction model. [Raboonik et al. \(2016\)](#) calculated the Zernike moments of magnetograms, and built the solar flare prediction model by using SVM. [Alipour et al. \(2019\)](#) extracted the Zernike moments from multi-band images of active regions, and built a solar flare prediction model by using the support vector machine. [Chen et al. \(2021\)](#) applied K-Nearest Neighbors, Random Forest, and XGBoost to predict the solar flare index based on the magnetic parameters. [Lim et al. \(2019b\)](#) extracted features from photospheric magnetograms, and built the solar flare prediction models based on multilayer perceptron, support vector machine and random forest methods. [Liu et al. \(2017a\)](#) constructed the solar flare prediction model based on random forests and magnetic field parameters. [Song et al. \(2009\)](#) selected total unsigned magnetic flux, the length of the strong-gradient neutral line and the total magnetic dissipation to build a solar flare prediction model by applying logistic regression algorithm. Based on the same

data, [Yuan et al. \(2010\)](#) used a combination of logistic regression and support vector machine to build a flare prediction model. In the model, the logistic regression is used to predict the probability of flare, and then a support vector machine is applied to predict the flare levels using the output probabilities as inputs. Solar flare prediction is usually done using supervised methods. Adopting a different way of thinking, [Benvenuto et al. \(2018\)](#) and [Piana et al. \(2019\)](#) proposed a hybrid supervised and unsupervised machine learning approach to build a solar flare prediction model. The hybrid approach combines supervised and unsupervised learning, and provides effective prediction performance for imbalanced dataset.

The ensemble learning is a commonly used method to improve prediction performance. [Abduallah et al. \(2021\)](#) used voting fusion of random forests, multilayer perceptron, and built an operational solar flare prediction systems. [Li and Du \(2019\)](#) built two solar flare prediction models from the active regions within 30° of solar disk center and outside 30° of solar disk center, then merged these two models into a single one, as a full-disk solar flare prediction model, the performances of these two models are compared and analyzed. [Chen et al. \(2022\)](#) proposed a two-stage model consisting of the unsupervised learning method and supervised learning method. The unsupervised clustering algorithm k-means is used to increase the positive sample rate, and the supervised method CNN is applied to predict whether the solar flares will occur in the next 48 hours. [Guerra et al. \(2020\)](#) constructed ensemble predictions for major solar flares by linearly combining the full-disk probabilistic predictions from a group of operational prediction methods. [Lim et al. \(2019a\)](#) applied ensemble learning method to predict solar flares by considering short-term, mid-term, and long-term properties of active regions. The testing results show that the model performance can be improved by using the ensemble learning method.

3.3.3.3 Interpretability. Interpretability in machine learning aims to make model decisions and predictions easier to understand.

The tree-based model progressively divides the data and generates easy-to-understand rules. The tree-based model like a nested if-else selection structure, each node represents an features test and each leaf node represents a decision. Due to its visual and easy to understand branching process, decision trees are often considered a highly interpretable model. For example, [Engell et al. \(2017\)](#) and [Liu et al. \(2017a\)](#) used tree-based interpretability methods by visualizing structure and node information.

Logistic regression is a linear model used for binary classification problems. It's output can be interpreted as a probability and can be used to predict the probability of a sample belonging to which class. The parameters and feature weights in a logistic regression model can help explain the model's decisions.

On the other hand, rules can also be implied in the results, such as the work of [Liu et al. \(2017b\)](#), which adopts a similar idea of clustering, by comparing the samples, which need to be explained, with the samples that have already been explained, based on the similarity.

3.3.4 Deep learning methods

Deep learning models are complex and can be directly fed with the diverse data (eg. images or videos). However more computer power are required. The application of deep learning methods include 3 perspectives: data, models, and interpretability.

3.3.4.1 Data. Regarding the number and distribution of databases, there are problems with small-sample size and class imbalance problem. The small-sample size problem, such as data observed by the newly launched devices, makes it difficult to build the prediction model by using the deep learning methods. [Covas \(2020\)](#) used transfer learning to improve the solar flare prediction performance. Another problem is class imbalance, deep generative models can generate samples to improve the class imbalance problem. [Deng et al. \(2021\)](#) improved the class imbalance problem based generative adversarial network.

When data is missing or unobservable, using deep learning methods to complete or repair is also an idea. [Kim et al. \(2019\)](#) generated farside solar magnetograms from STEREO/Extreme UltraViolet Imager (EUVI) 304-Å images using a deep learning model based on conditional generative adversarial networks (cGANs). But [Liu et al. \(2021\)](#) thought it may be far from providing scientifically reliable magnetograms. [Sun et al. \(2022b\)](#) proposed a novel dynamic deep-learning model by integrating a convolutional gated recurrent units (convGRU) model into a pix2pix baseline to generate more stable magnetograms. [Gao et al. \(2023\)](#) proposed to estimate magnetic field from $H\alpha$ images by using a conditional generative adversarial network (cGAN) as the basic framework, and found that more accurate magnetic polarity can be implicitly generated using magnetic polarity

constraints from time-series observations. It should be pointed out that, it is necessary to decide whether use generation data or not, should be based on the prediction performance.

There are also some researchers who use deep learning methods to repair data, such as Zhao et al. (2019) proposed a deep learning model, namely mask-Pix2Pix, which built on a well-known Pix2Pix network of conditional generative adversarial network (cGAN) for overexposure recovery. Yu et al. (2021) combined partial convolution (PC) and conditional generative adversarial network (GAN) introduce a progressive inpainting model PCGAN to repairing overexposure images. Zhang et al. (2023) found that the attention mechanism in deep learning is well consistent with non-local oversaturation repair theory, thus, an attention augmented convolutional neural network (AANet) is proposed for image desaturation. The inconsistent resolution of magnetograms, eg. observed by different instruments, is also a problem. Jungbluth et al. (2019) reconstructed the resolution of full-disk MDI magnetograms to be consistent with the HMI magnetograms. Dou et al. (2022) enhanced the resolution of the active region of MDI to be consistent. For flare prediction modeling, these methods of repairing data can be used as part of data enhancement, and the physical rationality behind them still needs to be further verified, and whether it is effective needs to be judged according to the actual modeling effect.

3.3.4.2 Model. The deep learning methods can not only directly extract the prediction patterns from raw data, but also get complex non-linear relationships. Huang et al. (2018) first automatically extracted the prediction patterns from magnetograms of active regions by using convolutional neural networks. Li and Huang (2018) automatically extracted the prediction patterns from white-light images of sunspots by using convolutional neural networks. Chen et al. (2019) extracted sequential prediction patterns from time series of magnetic parameters by using the LSTM algorithm. Liu et al. (2019) constructed LSTM prediction model of solar flares based on magnetic parameters. Jiao et al. (2020) built a soft X-ray flux regression model by using LSTM algorithm based on time series of magnetic parameters. Landa and Reuveni (2022) presented a 1D convolutional neural network to predict the probability of solar flares. This model directly extract the predictors from the soft X-ray time-series without handcrafted features from magnetograms. Based on magnetogram, Bhattacharjee et al. (2020) used CNN to build a solar flare prediction model, and they found that the CNN pays attention to regions between the opposite polarities within active regions. Wang et al. (2022) developed an unsupervised network using group train, aiming to understand and memorize quiet periods magnetic features of the non-strong-flare ARs and then identify the strong flares magnetic features of the strong-flare ARs by an anomaly detection algorithm using group test. They found anomaly detection algorithm can be used in precursor identification for strong flares and help in both improving strong-flare prediction accuracy and enlarging the time in advance. Deshmukh et al. (2022) developed a hybrid two-stage machine-learning architecture to address the problem of excessive false alarms in solar flare prediction systems. The first stage uses convolutional neural network model to extract features from magnetograms then produce a flaring probability. The second stage uses those flaring probability and magnetograms feature vector to train an extremely randomized trees model for a solar flare prediction. Sun et al. (2022a) built two flare prediction models using three-dimensional convolutional neural networks (3D CNNs), and used the the gradient-weighted class activation mapping (Grad-CAM) method to visually explain flare prediction models. The Grad-CAM illustrates that the 3D CNNs may extract the spatial distribution and evolution of AR magnetic fields simultaneously for the flare prediction. Zheng et al. (2023b) developed five machine learning models, neural network (NN), long short-term memory (LSTM), LSTM based on attention mechanism (LSTM-A), bidirectional LSTM (BLSTM), and BLSTM based on attention mechanism (BLSTM-A), for predicting whether a C class or M class flare will occur in an active region in the next 24 hr, and found LSTM-A for C class performs better than other models.

Ensemble learning method can be applied to improve the performance of solar flare prediction model. Based on magnetic parameters and magnetograms, Guerra et al. (2020) constructed ensemble predictions for major solar flares by combining the probabilistic prediction results from a group of operational prediction models. Yi et al. (2020) applied two LSTM layers and one fully connected layer for the encoder and decoder to predict solar flare X-ray flux profiles, without any preprocessing to extract features from data, and achieved better performance for the prediction of X-ray flux profiles with low-peak fluxes than those with high-peak fluxes. Li et al. (2022) combined prior knowledge of flare production with CNN structures and build three models, and they results show that the knowledge-informed deep learning technique can improve the performance of flare prediction models. Tang et al. (2021) built a fusion model based on a deep neural network, convolutional neural network, and

bidirectional long short-term memory neural network, found combining the advantages of each independent model can achieve better performance than traditional statistical prediction models or any single machine-learning method. [Guastavino et al. \(2022\)](#) proposed a hybrid deep learning model composed of CNN and LSTM networks based video data of magnetograms. [Zheng et al. \(2023a\)](#) developed a Hybrid Bidirectional Long and Short-Term Memory based on attention mechanism (HBiLSTM-Attention) model and a BiLSTM-Attention model for multiclass flare prediction within 24 h, they found the prediction model based on physical parameters and the prediction model based on autonomous extraction of image features have their own advantages and limitations. [Abduallah et al. \(2023\)](#) presented a transformer-based framework, named SolarFlareNet, which was combined with a one-dimensional convolutional neural network (Conv1D), long short-term memory (LSTM), transformer encoder blocks (TEBs), and additional layers that include batch normalization (BN) layers, dropout layers, and dense layers. Their ablation study indicates that all components have made contributions to the overall performance of the proposed framework.

3.3.4.3 Interpretability. Interpretability of deep learning model aims to understand the decision-making process and internal mechanisms of the deep learning model. Alternative methods of interpretability in deep learning based solar flare prediction model are as follows:

Feature visualization: Feature map visualization or activation heat map can be used to visualize the intermediate layer features in a deep neural network. By visualizing the intermediate layer features in a deep neural network, it is possible to observe and understand the features extracted by the model during the learning process. The visualization method applied in [Zheng et al. \(2021\)](#) work, and the feature visualization results indicate that they model pays attention to the regions with strong gradient, strong intensity, high total intensity, and large range of the intensity in high-level feature maps. [Liu et al. \(2022\)](#) used feature visualization to analyze the impact of magnetogram resolution on solar flare prediction model. [Sun et al. \(2023\)](#) analyzed the region of interest in multi-band images of active regions for solar flare prediction model.

Visualizing network structure: In order to understand the model operation and decision-making, the structure of neural networks could be Visualized, e.g., convolutional kernels and attention mechanisms. [Huang et al. \(2018\)](#) visualized the changes for convolutional kernels during the training process.

Local explanatory methods: The Local Linear Interpretation, eg. LIME ([Ribeiro et al., 2016](#)) can be used to help to understand how the model makes decisions.

Model simplification and rule extraction: These methods attempt to explain the behavior of deep learning models by compressing the model, reducing the number of parameters, or extracting interpretable rules. [Dhuri et al. \(2020\)](#) developed a network-pruning algorithm to interpret the trained CNN. Using synthetic magnetograms, they demonstrated that the CNN output is sensitive to the length scale of the magnetic regions.

The interpretability of deep learning models is a complex and developing research, and there is no universal solution yet. Different methods are applicable to different tasks, and choosing an appropriate interpretive method usually requires evaluation and selection based on specific application requirements and data characteristics.

3.3.5 Methodology comparison and discussion

Some researchers ([Asensio Ramos et al., 2023](#); [Barnes et al., 2016](#); [Camporeale, 2019](#); [Chen et al., 2023](#); [Erdélyi et al., 2022](#); [Georgoulis et al., 2021](#); [Gombosi et al., 2021](#); [Han et al., 2023](#); [Leka et al., 2019a,b](#)) have provided some overviews about solar flare prediction.

Some researchers ([Asensio Ramos et al., 2023](#); [Barnes et al., 2016](#); [Camporeale, 2019](#); [Chen et al., 2023](#); [Erdélyi et al., 2022](#); [Georgoulis et al., 2021](#); [Gombosi et al., 2021](#); [Han et al., 2023](#); [Leka et al., 2019a,b](#)) have provided several reviews on solar flare prediction. To facilitate the application of machine learning in solar flare prediction, we have summarized commonly used

models and their metrics. Table 2 shows the comparison of models in the machine learning stage, and Table 3 shows the comparison of models in the deep learning stage.

Due to differences in data samples, labels, and the division between training and testing sets, model comparisons may have potential unfairness. These results are for reference purposes only (for example, Cinto et al. (2020) discusses the unfairness of model comparisons).

3.4 Potential challenges

The research points and challenges of solar flare prediction mainly include three aspects: Data Selection and Usage; Limited Data Volume; and Interpretability.

Data Selection and Usage: When it comes to the selection and usage of data, unlike in other fields, standard publicly available datasets are not provided in the solar activity prediction. However, common dimensions for selecting data can be chosen from the following two perspectives: (1) Selecting data regions, which commonly include the entire solar disk, data from active

Table 2 Use machine learning models to predict solar flares.

Work ^{a)}		Input	Output		Test Information		Metric ¹⁾				
Author	Model	Region ^{b)}	Class	Time ^{c)}	Test ^{d)}	Table ^{e)}	TPR/Recall	FPR/FAR	ACC	HSS	TSS
Statistical methods											
Bloomfield et al. (2012)	Poisson	AR.F*	C+/M+	h0-24	holdout-tv*	T4	0.753/0.704	0.649/0.854	0.712/0.830	0.315/0.190	0.456/0.539
Machine learning methods											
Colak and Qahwaji (2009)	NN	AR.F ₁	C+/M+/X	h0-6	holdout-tv*	T3	0.363/0.413/0.278	0.680/0.894/0.992		0.378/0.250/0.033	
				h0-12			0.644/0.732/0.667	0.506/0.801/0.982		0.479/0.398/0.096	
				h0-24			0.772/0.865/0.917	0.319/0.688/0.967		0.493/0.470/0.169	
				h0-48			0.829/0.916/0.957	0.171/0.558/0.932		0.474/0.503/0.262	
				h0-72			0.848/0.929/0.968	0.117/0.481/0.908		0.457/0.510/0.310	
Yang et al. (2013)	SVM	AR.C ₁	C1.0+/C5.0+/ M1.0+/M5.0+/ X1.0+	h0-6	cross-tv	T4	0.560/0.595/ 0.523/0.667/ 0.700		0.826/0.898/ 0.939/0.985/ 0.987	0.378/0.306/ 0.286/0.322/ 0.265	0.430/0.508/ 0.474/0.654/ 0.688
				h0-12			0.653/0.485/ 0.554/0.460/ 0.533		0.786/0.896/ 0.934/0.982/ 0.985	0.430/0.379/ 0.382/0.329/ 0.346	0.475/0.419/ 0.505/0.447/ 0.522
				h0-24			0.677/0.626/ 0.550/0.329/ 0.480		0.761/0.847/ 0.907/0.980/ 0.983	0.466/0.437/ 0.424/0.343/ 0.392	0.478/0.507/ 0.531/0.321/ 0.469
				h0-48			0.707/0.627/ 0.642/0.634/ 0.522		0.742/0.825/ 0.860/0.942/ 0.976	0.473/0.474/ 0.438/0.393/ 0.462	0.474/0.500/ 0.531/0.587/ 0.507
Muranushi et al. (2015)	SVM+LR	FD ₁ *	C+/M+/X	h0-24	cross-tv	T5					0.629/0.517/0.846
Liu et al. (2017a)	RF	AR.F* ₁	B+/C+/ M+/X+	h0-24	cross-tv*	T3	0.812/0.526/ 0.671/0.297		0.844/0.712/ 0.778/0.957		0.669/0.328/ 0.500/0.291
Cinto et al. (2020)	XGBoost	AR.C/AR.F ₁	C+	h0-24	cross-tvt	T2	0.86/0.89	0.13/0.13	0.85/0.86		0.69/0.7
				h24-48			0.87/0.87	0.14/0.15	0.84/0.84		0.68/0.67
				h48-27			0.86/0.87	0.15/0.14	0.83/0.85		0.65/0.68

^{a)} Due to space constraints, only some classification models are selected here.

b) For region, AR.C represents modelling use center active regions data and inside some degrees of the central meridian, AR.F represents modelling use all active regions data, FD represents use full-disk data.

1, 2, and 3 represent the dimensions of the data used for modeling,
e.g., 1-dimensional vector features, 2-dimensional image data, 3-dimensional video data.

c) This represents the range of the forecast time window.

d) Test methods is holdout or cross represent validation methods, tv represent splitting as train-val, tvf represent as train-val-test.

e) The latter metric information is extracted from which table in the original paper.

f) The metrics are extracted from the original raw paper, empty spaces represent data not directly given in the original paper.

* Data with '*' represents this information is speculated from the raw paper.

Table 3 Use deep learning models to predict solar flares

Work ^{a)}		Input	Output		Test Information		Metric ^{f)}				
Author	Model	Region ^{b)}	Class	Time ^{c)}	Test ^{d)}	Table ^{e)}	TPR/Recall	FPR/FAR	ACC	HSS	TSS
Deep learning methods											
Huang et al. (2018)	CNN	AR.C 2	C/M/X	h0-6	cross-tv	T4	0.64/0.84/0.88			0.227/0.054/0.006	0.463/0.662/0.723
				h0-12			0.69/0.83/0.91			0.243/0.081/0.014	0.447/0.632/0.758
				h0-24			0.73/0.85/0.87			0.339/0.143/0.025	0.487/0.662/0.714
				h0-48			0.67/0.81/0.97			0.467/0.206/0.042	0.505/0.621/0.791
Nishizuka et al. (2018)	DNN	AR.F 1*	C+/M+	h0-24	holdout-tvt	T3	0.81/0.95	0.47/0.82	0.82/0.86	0.53/0.26	0.63/0.80
Park et al. (2018)	CNN	FD 2	C+*	h0-24*	cross-tv*	T4	0.83	0.11	0.84	0.69	0.69
Li et al. (2020)	CNN	AR.C 2	C+/M+	h0-24	cross-tvt	T2	0.889/0.817	0.094/0.111	0.861/0.891	0.671/0.759	0.679/0.749
Nishizuka et al. (2021)	DNN,t=0.5	AR.F 1*	C+/M+	h0-24	holdout-tvt	T3		0.40/0.24	0.99/0.99	0.64/0.06	0.24/0.24
	DNN,t=0.45							0.47/0.97	0.98/0.98	0.62/0.05	0.78/0.24
	DNN,t=0.40						0.52/0.94	0.98/0.98	0.61/0.10	0.83/0.48	
Zheng et al. (2021)	CNN	AR.C 2	C+/M+	h0-24	cross-tv*	T3	0.895/0.818	0.086/0.126	0.869/0.887	0.692/0.746	0.703/0.739
Deng et al. (2021)	GAN+CNN	AR.C 2	C/M/X	h0-24	holdout-tv*	T8	0.734/0.746/0.830		0.865/0.864/0.909	0.644/0.655/0.770	0.646/0.653/0.762
Sun et al. (2022a)	3D CNN	AR.C 3	C+/M+	h0-24	cross-tv*	T5	0.862/0.925	0.107/0.099	0.879/0.904	0.756/0.667	0.756/0.826
Li et al. (2023)	CNN	AR.F 2	M+	h0-48	cross-tv*	T2	0.672	0.237	0.727		0.454
Abduallah et al. (2023)	Transformer	AR.C 1	C+/M+/M5+	h0-48	cross-tv*	T2	0.722/0.735/0.739		0.896/0.907/0.923		0.719/0.728/0.736
		AR.C 1		h0-72			0.702/0.708/0.717		0.863/0.883/0.906		0.709/0.714/0.729

a) Due to space constraints, only some classification models are selected here.

b) For region, AR.C represents modelling use center active regions data and inside some degrees of the central meridian, AR.F represents modelling use all active regions data, FD represents use full-disk data.

1, 2, and 3 represent the dimensions of the data used for modeling,
e.g., 1-dimensional vector features, 2-dimensional image data, 3-dimensional video data.

c) This represents the range of the forecast time window.

d) Test methods is holdout or cross represent validation methods, tv represent splitting as train-val, tvf represent as train-val-test.

e) The latter metric information is extracted from which table in the original paper.

f) The metrics are extracted from the original raw paper, empty spaces represent data not directly given in the original paper.

** Data with * represents this information is speculated from the raw paper.

regions, or active regions filtered according to specific conditions. (2) Selecting observation data types, where common data types include magnetograms, etc. Many researches focused on flare prediction for each AR individually, while some researches discussed flare prediction of the front side of the Sun for the operational usage, (as in (Nishizuka et al., 2021, 2018, 2017)).

Limited Data Volume: There are two dimensions to be considered here. Firstly, the total amount of data is limited, largely due to constraints in solar observation technologies. Currently, the maximum observation period for solar activity from imaging satellites is typically around two solar activity cycles, which prevents the utilization of machine learning to extract longerterm temporal information. Secondly, the limited sample size includes the scarcity of samples related to solar flare eruptions, leading to what is known as an imbalanced dataset problem. Strong flares are rare events, resulting in an unbalanced dataset that requires special attention in dataset preparation and model verification. This imbalance is one of the reasons why flare prediction remains challenging and in need of improvement.

Interpretability: Machine learning is often considered a black-box model. We are unable to understand the physical meanings of the prediction models. On one hand, this limitation hinders our ability to gain deeper insights from solar physics. On the other hand, it makes challenges to accurately assess reliability of model during engineering deployments.

4 CMEs Prediction

CMEs are a potent solar eruption phenomenon that can induce significant changes in the Earth’s near-Earth space environment. CMEs consist of large-scale eruptions of magnetized plasma originating from the Sun and traveling into interplanetary space (Hundhausen et al., 1984). On average, CMEs propagate outward at speeds ranging from 200 to 3000 km/s and carry masses exceeding 10^{12} kg, resulting in a kinematic energy greater than 10^{25} J (Zhao and Dryer, 2014). During a solar minimum, CMEs occur approximately once every few days, but during a solar maximum, several CMEs can occur per day.

CMEs have gotten a lot of interest from scientists studying the sun and space weather since their discovery in the 1970s (Sheeley Jr et al., 1985). On the one hand, CMEs that have a strong and persistent component of the southern magnetic field can cause intense geomagnetic storms when they collide with the Earth’s magnetosphere (Gosling et al., 1991). On the other hand, fast CMEs propagating in the solar wind generate the interplanetary shock, which serves as the primary source of SPEs (Gopalswamy et al., 2003). Furthermore, Earth-directed CMEs not only affect space missions, aviation, and power networks but also have indirect impacts on other industries, including navigation systems and gas and oil pipelines (Boteler et al., 1998). Given the significant role of CMEs and their interplanetary shocks in space weather, accurately predicting the geo-effectiveness and arrival times of CMEs on Earth is a primary objective.

4.1 Problem definition

CMEs typically spread from the sun to the earth for 1 to 5 days, making it entirely possible to anticipate whether and when they will arrive (Richardson and Cane, 2010). When researchers predict the arrival time of a CME or its associated shock, they use the previously received solar observations as inputs to determine when the event will occur. The observation images are typically from the spacecraft, such as the Solar and Heliospheric Observatory (SOHO; (Domingo et al., 1995)), the Solar TERrestrial Relations Observatory (STEREO; (Socker et al., 2000)), and the Solar Dynamics Observatory (SDO; (Pesnell et al., 2012a)). Furthermore, the physical parameters of CMEs, such as their velocity, acceleration, and angular width, can be obtained from those observation images. Therefore, the current models use these CME parameters derived from observation images as input when predicting CMEs’ arrival.

Researchers have discovered that the CMEs are essentially a plasma cloud, and CMEs continuously interact with the background solar wind while traveling to Earth. Therefore, fast CMEs will compress and heat the slow background solar wind and create a shock wave ahead of it, while slow CMEs entering the interstellar medium won’t cause noticeable solar wind disturbances. As a result, a variety of physics-based models (Corona-Romero et al., 2015, 2017; Dumbovic et al., 2018; Gopalswamy et al., 2001; Kay and Gopalswamy, 2017; Kay et al., 2017; Kilpua et al., 2012; Mays et al., 2015; Millward et al., 2013; Möstl et al., 2017; Riley et al., 2018; Shi et al., 2015; Vršnak et al., 2014; Wold et al., 2018; Wood et al., 2017) simulate the spread and evolution of CMEs in the interplanetary to estimate the time at which they will arrive. Furthermore, the application of machine learning (Choi et al., 2012; Liu et al., 2018; Pricopi et al., 2022; Sudar et al., 2015; Uwamahoro et al.,

2012) and deep learning (Besliu-Ionescu and Mierla, 2021; Besliu-Ionescu et al., 2019; Fu et al., 2021; Shi et al., 2022, 2021; Wang et al., 2019b) techniques to anticipate the time of CMEs' arrival has become a research hotspot with the advancement of computer vision.

Finally, we can define the problem of CMEs' arrival prediction. The CME arrival prediction model simulates the spread and evolution of CMEs in interplanetary space using solar observation data collected by satellites as input to predict whether and when the CMEs will arrive. It is important to note that the CME arrival prediction model can either directly use solar observation images as input or can use the physical parameters of CMEs calculated from those images.

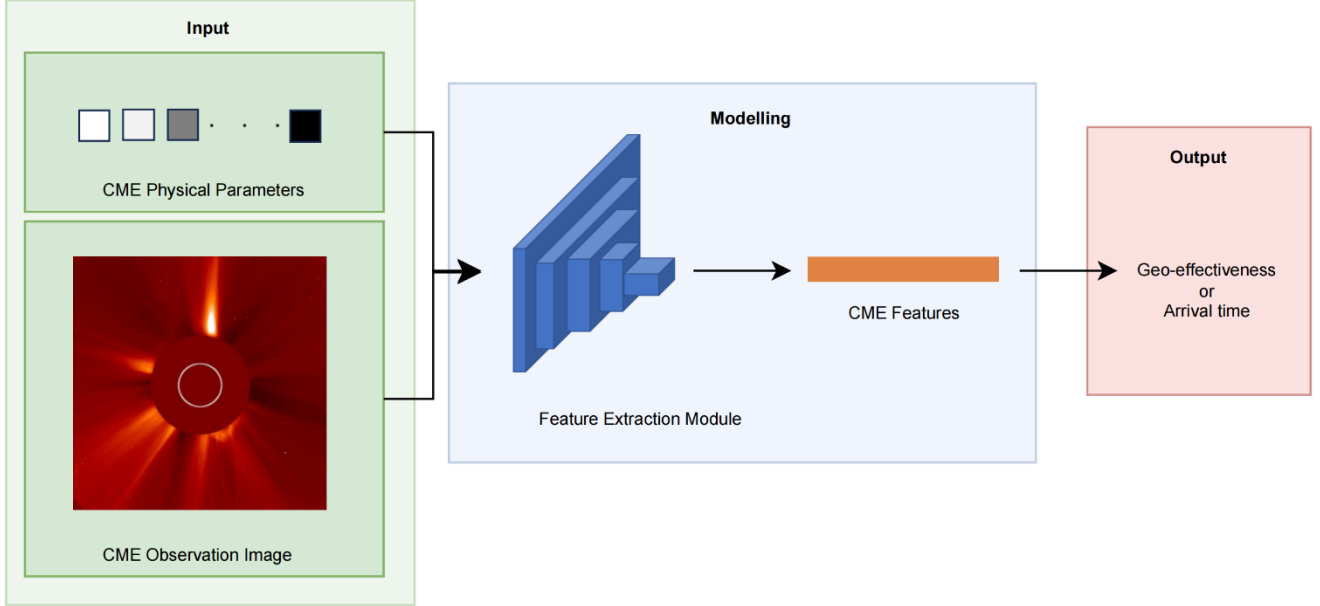


Figure 12 CMEs' arrival prediction modeling pipeline

4.2 Observational instruments and predictors

In this section, we provide a brief overview of the observational instruments and predictors of CMEs (i.e., CMEs' observation images and physical parameters).

4.2.1 Observational instruments

The Extreme-ultraviolet Imaging Telescope (EIT; (Delaboudiniere et al., 1995)) and the Large Angle and Spectrometric Coronagraph Experiment (LASCO; (Brueckner et al., 1995)) are among the twelve instruments on the SOHO satellite (Domingo et al., 1995) used for observing the Sun. EIT primarily captures high-resolution coronal images in the ultraviolet spectrum. Conversely, LASCO investigates the corona's structure using artificial solar eclipses created by optical systems.

STEREO (Socker et al., 2000) is a dual-satellite system, with the Ahead satellite preceding Earth in orbit and the Behind satellite trailing behind. This configuration allows STEREO (Socker et al., 2000) to capture three-dimensional images of the sun. The Sun-Earth Connection Coronal and Heliospheric Investigation (SECCHI; (Howard et al., 2008)) is the instrument on STEREO (Socker et al., 2000) related to CMEs observations. SECCHI (Howard et al., 2008) is primarily responsible for collecting images of CMEs from the sun's surface to their impact point on Earth, enabling the study of their three-dimensional evolution.

The SDO (Pesnell et al., 2012a) is equipped with three key instruments: the Atmospheric Imaging Assembly (AIA) (Lemen et al., 2012), the Helioseismic and Magnetic Imager (HMI) (Scherrer et al., 2012), and the Extreme Ultraviolet Variability Experiment (EVE) (Woods et al., 2012). Compared to the SOHO (Domingo et al., 1995), the SDO (Pesnell et al., 2012a) offers significantly enhanced shooting efficiency and image clarity.

4.2.2 Predictors

The CME prediction model usually utilizes the physical parameters of CMEs (such as angle width, velocity, acceleration, mass, and energy) as predictors. The physical parameters of CMEs can be determined using CME detection technology to analyze

Table 4 CMEs geoeffectiveness prediction. The experimental results are obtained from the original papers and are based on their own datasets.

Model	Reference	Algorithm	Accuracy
Physics-based	(Mays et al., 2015)	WEC	0.83
	(Wold et al., 2018)	WEC	0.85
	(Möstl et al., 2017)	SSEF	0.90
Machine learning	(Choi et al., 2012)	Support vector machine	0.66
	(Uwamahoro et al., 2012)	Neural network	0.86
	(Besliu-Ionescu et al., 2019)	Logistic regression	1.00
	(Besliu-Ionescu and Mierla, 2021)	Logistic regression	0.98
	(Shi et al., 2022)	Recommendation system	1.00
	(Pricopi et al., 2022)	Machine learning	0.99

observation images. It's worth noting that the commonly used CME kinematic features are projection features of CMEs. For instance, the angular width denotes the projection angle that the CME occupies on the plane of the observing sky. Because of the projection effect, this angle measures the size of the CME in the plane of observation, not its actual three-dimensional size. As the structure of CMEs often appears as a projection on the sky plane, researchers determine the velocity of the entire CME structure by measuring the velocity at its front at various moments. They calculate the average motion speed of CMEs by collecting height-time measurements and fitting the height-time data linearly. Additionally, besides the average projected velocity, the velocity at 20 Rs is also a significant predictor. The velocity at 20 Rs reflects the speed at which CMEs propagate through the solar system. Consequently, researchers commonly integrate the velocity at 20 Rs into space weather models to simulate and predict the behavior of CMEs in the solar system.

It is worth distinguishing between the expansion velocity, the projected velocity, and the propagation velocity of CMEs. The expansion velocity of a CME refers to the speed at which the CME's front is moving outward in a radial direction from the sun. As a CME erupts from the solar corona, the material within it expands outward into space. The expansion velocity quantifies the rate at which this expansion occurs. The projected velocity is the CME velocity in the observer's plane. Since CMEs tend to be three-dimensional structures and observations usually occur in the plane of the solar system, the observed velocity is a projection of the CME motion. The propagation velocity is the speed at which the CME as a whole propagates along its trajectory. This velocity takes into account the overall motion of the CME, including the motion of its leading edge and other structures.

4.3 Model

In this section, we will introduce the application of machine learning and deep learning technology in CME prediction, including CME geoeffectiveness prediction and CME arrival time prediction.

4.3.1 CMEs geoeffectiveness prediction

To predict the precise arrival time of a CME, a forecaster must have a reasonable level of confidence that the event observed in the CME images will hit Earth. Over the past few years, some papers (Mays et al., 2015; Möstl et al., 2017; Wold et al., 2018) have scrutinized hit/miss statistics and devised methods to enhance the hit rate. Möstl et al. (2017) examined the reliability of employing HI observations to determine the arrival time of CMEs as well as assess the precision of hit-and-miss predictions.

Their approach adopted the self-similar expansion model for CME fronts, considering a longitudinal width of 60° , consistent speed, and unvarying propagation direction, mirroring the characteristics of analytical or numerical modeling techniques. [Mays et al. \(2015\)](#) presented the WSA-ENLIL+Cone ensemble model, which offers a probabilistic prediction comprising an assessment of the uncertainty in arrival time resulting from diverse predictions and forecast certainty regarding the probability of CME arrival. [Wold et al. \(2018\)](#) validates the WSA-ENLIL+Cone model and provides a comprehensive analysis of the factors that require careful consideration when interpreting the verification results. These factors pertain to the uncertainties that emerge when determining the input parameters of CMEs based on real-time coronagraph data, the challenges associated with identifying in situ arrivals of ICMEs, as well as cases involving multiple ICME arrivals.

The Support Vector Machine (SVM) has been used for the first time to predict geoeffective halo CMEs by [Choi et al. \(2012\)](#). They utilized data from the SOHO/LASCO CME Catalog, which consisted of halo and partial halo CMEs from January 1996 to April 2010. Moreover, using the X-ray levels of related flares, they identified front-side halo CMEs that are stronger than the B1 class, and using the Dst index, they detected geoeffective halo CMEs stronger than -50 nT. The SVM classification input features are all possible combinations of the angular breadth, speed, and corresponding X-ray flare classes of the CMEs. Utilizing the CME speed and X-ray flare classes as SVM input features, they could determine the statistical parameters for the best model with an accuracy of 0.66.

[Uwamahoro et al. \(2012\)](#) presented a neural network-based model for estimating the probability of geomagnetic storm occurrence following halo CMEs and their associated interplanetary (IP) events. The model incorporates solar and IP variables that define geoeffective halo CMEs. The solar inputs include quantitative variables such as the halo CME angular width (AW), CME speed (V_{cme}), and the flare index, which measures the amount of flaring activity linked to halo CMEs. The IP parameters are the southbound Z-component of the interplanetary magnetic field (IMF) and the peak values of the solar wind speed (V_{sw}). The model accurately predicted all resulting geomagnetic storms with a Dst of less than or equal to -100 nT under the assumption that a halo CME would form. The model's accuracy was 75% for moderate storms with a Dst between -100 and -50 nT. However, it is important to note that the input parameters utilized are directly linked to halo CMEs, and thus, the model is unable to predict the likelihood of geomagnetic storms that are not CME-driven.

[Besliu-Ionescu et al. \(2019\)](#) aimed to predict the geoeffectiveness of a given CME using a modified version of [Srivastava \(2005\)](#)'s logistic regression model with the initial parameters of the CME. Following an eruption, the CME's propagation path is closely related to its origin point on the Sun, its speed, the solar coronal environment, and other factors that influence its ability to reach Earth. The model aims to determine whether a CME will be accompanied by a geomagnetic storm, using a minimum Dst value of -30 nT as a threshold. They used data from April 1, 2010, to June 30, 2011, during the ascending phase of solar cycle 24, and applied a modified logistic regression model to examine CMEs detected by LASCO. This approach differs from that of [Srivastava \(2005\)](#), who restricted their CME database to events associated with strong and super-intense geomagnetic storms over a longer period (1996-2002).

[Besliu-Ionescu and Mierla \(2021\)](#) has revised a logistic regression model based on [Srivastava \(2005\)](#) to predict if a CME will impact the planet and if it will be accompanied by a geomagnetic storm with a minimum Dst value of less than -30 nT. In solar cycle 24, the model is executed for several CMEs listed in the SOHO LASCO CME catalog. Simulating the propagation and evolution of CMEs in interplanetary space requires several independent variables. Specifically, the model relies on five key variables: CME speed at 20 solar radii, angular width, observed position angle, and acceleration. [Besliu-Ionescu and Mierla \(2021\)](#) discovered that the solar wind can occasionally accelerate before the 20 solar radii. Hence, the CME's condition after leaving the solar corona is more precisely revealed by the speed at 20 solar radii. Furthermore, they selected a binary variable for this intent identical to the one used by [Besliu-Ionescu et al. \(2019\)](#). It is defined as follows: 0 if Dst min is > -30 nT (indicating no geomagnetic storm), and 1 if Dst min ≤ -30 nT (suggesting that a storm was detected).

To enhance prediction accuracy, a combination of machine learning techniques and expert forecasters is considered a more effective approach to predict the CME geoeffectiveness. Forecasters must depend on past solar activity events to evaluate current or future events, and the logistic regression and recommendation algorithm can offer forecasters a choice to improve their predictions. Therefore, [Shi et al. \(2022\)](#) first identified and analyzed the correlation coefficients of the distinct CME parameters using logistic regression, such as central position angle, angular width, and linear velocity, which were derived from the CME's

observation images. And then they developed a logistic regression model to predict if a CME would cause an impact on Earth. The model produces an F1 score of 30% and a recall of 53%. Moreover, they utilize the recommendation algorithm to identify the historical event that closely represents each CME. All of this information serves as a benchmark for comparing and predicting the geoeffectiveness of CMEs.

Pricopi et al. (2022) focused on assessing whether machine-learning approaches could determine if newly erupting CMEs have the potential to cause geomagnetic activity. Several approaches were trained on white-light coronagraph datasets of the Sun, including the logistic regression model, k-nearest neighbors, support vector machines, feed-forward artificial neural networks, and ensemble models. Pricopi et al. (2022) limited their forecast to only include solar parameters to ensure extended preparation times, despite evidence that interplanetary factors also influence the geoeffectiveness of CMEs. The main challenges in determining the geoeffectiveness of CMEs were identified as significant imbalances in the number of geoeffective versus ineffective events, as well as a lack of relevant variables. Pricopi et al. (2022) demonstrated that even under such circumstances, these models could achieve sufficient predictive accuracy.

Table 5 CMEs arrival time prediction. The experimental results are obtained from the original papers and are based on their own datasets.

Model	Reference	Algorithm	MAE
Empirical	(Gopalswamy et al., 2001)	ESA	15.1 ± 12.8
	(Kilpua et al., 2012)	GCS+ESA	16.9 ± 11.3
	(Wood et al., 2017)	3D reconstruction	6.3 ± 8.1
Drag-based	(Vršnak et al., 2014)	WEC	13.9 ± 9.7
	(Dumbovic et al., 2018)	DBM	14.3 ± 9.0
	(Shi et al., 2015)	GCS+DBM	12.9
Physics-based	(Corona-Romero et al., 2015)	Type-II; blast-wave model	3.5 ± 2.4
	(Corona-Romero et al., 2017)	Piston shock	6.7 ± 20
MHD	(Millward et al., 2013)	CAT+WEC	7.5 ± 10
	(Riley et al., 2018)	Multiple MHD models	14.1 ± 18.4
	(Wold et al., 2018)	StereoCAT+WEC	10.4 ± 0.9
	(Wang et al., 2018)	WSA-Cone-modified HAF model	18.0
Machine learning	(Sudar et al., 2015)	Neural network	12
	(Liu et al., 2018)	Support vector machine	5.9 ± 4.3
	(Wang et al., 2019b)	Convolutional neural network	12.4
	(Shi et al., 2021)	Recommendation system	11.7
	(Fu et al., 2021)	Convolutional neural network	5.8

4.3.2 CMEs arrival time prediction

In this section, we first provide a brief summary of the classical models based on physical theory, as shown in Table 5. As mentioned in the existing papers (Verbeke et al., 2019; Zhao and Dryer, 2014), these models can be broadly categorized into the following kinds: empirical models, drag-based models, physics-based models, and magnetohydrodynamics (MHD) numerical models. Empirical models (Gopalswamy et al., 2001; Kilpua et al., 2012; Wood et al., 2017) employ straightforward equations for establishing the correlation between the arrival time of CME disturbance at Earth and the corresponding observables detected near the sun, including the initial velocity. Furthermore, CMEs are significantly impacted by the

surrounding solar wind beyond a particular helio-distance. To address this, drag-based models (Dumbovic et al., 2018; Shi et al., 2015; Vršnak et al., 2014) apply the equation of motion to CMEs. These models incorporate acceleration and deceleration, which depend on the relative velocity between the CMEs and the surrounding solar wind. The physics-based models (Corona-Romero et al., 2015, 2017) utilize particular theories or concepts to make accurate predictions, which are generated based on the framework developed from CME-IP disturbance events. The MHD models (Millward et al., 2013; Riley et al., 2018; Wang et al., 2018; Wold et al., 2018) can simulate the motion equation of solar disturbances within the background solar wind. These models incorporate time-varying boundary conditions near the Sun to provide predictions of the expected arrival time of the solar disturbance on Earth. Researchers have proposed numerous physics-based models for simulating the propagation of CMEs. However, only a few have been incorporated into operational space weather forecasting systems, such as the WSA-ENLIL-Cone (WEC) model at NOAA/SWPC (Millward et al., 2013) and the WSA-Cone-modified HAF model at NSSC/SEPC (Wang et al., 2018). The transition from research models to operational prediction systems highlights the significant challenges in CME prediction. It is also worth noting that some models Millward et al. (2013); Vršnak et al. (2014) listed in Table 5 are used for predicting the shock arrival time.

Next, we describe several machine learning techniques that have been applied in predicting CME arrival time. Sudar et al. (2015) presented the first attempt at exploring how the central meridian distance and initial speed of CMEs affect their arrival time using the neural network algorithm. They found a typical drag-like pattern in the solar wind when assessing the arrival time based on CMEs' initial velocity. Also, Sudar et al. (2015) discovered that CMEs related to flares on the western side of the Sun have shorter transit times than those originating on the eastern side. The average time difference between the neural network prediction and observations is consistent with previous research, at 12 hours. Compared with physics-based methods, the biggest advantage of using neural networks is the unnecessary specification of empirical functions or hyper-surfaces that connect input and output parameters. However, the use of more input parameters makes it challenging to obtain the optimal function.

Liu et al. (2018) proposed a novel approach called CAT-PUMA, which utilizes a support vector machine (SVM) algorithm to predict the arrival time of partial- or full-halo CMEs. To develop the prediction engine, they thoroughly investigated CMEs' properties and solar wind factors, employing 182 previously reported geoeffective CME events. By using a feature-selection method, Liu et al. (2018) identified the observable features crucial in predicting CMEs' arrival time. The experimental results show that the CMEs' average speed, final speed, angular width, and mass are the most significant factors in determining transit time in interplanetary space. They also found that the alpha-particle to proton-number density ratio, magnetic fields B_z and B_x , proton temperature, flow speed, flow pressure, and flow longitude are significant solar-wind factors. The model provides a mean absolute prediction error of approximately 6 hours without requiring any a priori assumptions or fundamental physical theory. Even though CAT-PUMA obtains a positive result, its average performance cannot be justified by the best result out of 100,000 experiments.

The models mentioned earlier relied on manually selected parameters as their input. However, gathering the parameters from the original observations is often subjective and could overlook crucial factors, resulting in biases. Furthermore, manual feature selection demands specialized expertise in CMEs. To address these issues, Wang et al. (2019b) utilized input data consisting of 1122 CME images from 223 events detected by the SOHO LASCO C2 Brueckner et al. (1995) to train a CNN regression model, which accurately predicted the arrival times of CMEs. The mean absolute error for all samples is approximately 12.4 hours, which is comparable to the average performance of previous studies on the same topic. Unfortunately, the dataset used in the experiment consisted only of 1122 images, which is considered very small in machine learning. By utilizing the inherent characteristics of CNN, the accuracy of the results can be significantly improved by increasing the number of training samples.

The structure and shape of observed images are not an accurate representation of the CME's actual shape in 3D space due to the projection effect on a 2D plane. This projection effect and observation angle significantly limit research into the physical mechanisms of CMEs and the advancement of prediction. To estimate the geoeffectiveness and arrival time of CMEs, Shi et al. (2021) considered that forecasters' experience can complement CME models. Therefore, they created a machine learning-based recommendation system to effectively analyze the impact of historical CME events and identify comparable events for current prediction work. Additionally, they also constructed a list of associations for 215 distinct observations with 18 criteria, such as CME characteristics, eruption region coordinates, and solar wind parameters. Finally, they used the recommendation algorithm to create a model and predicted the arrival time of CMEs based on these lists. The mean absolute error and root mean square

error of the test set are 11.78 hours and 13.77 hours, respectively. These values are consistent with other CME models published on the CME scoreboard, demonstrating the efficacy of the recommendation algorithm. Furthermore, the innovative recommendation algorithm-based approach may have many potential applications in space weather prediction.

Although the above methods have achieved good results, they are unable to integrate predicting the geoeffectiveness of CMEs and the arrival time of CMEs in a joint framework. Therefore, [Fu et al. \(2021\)](#) proposed a new deep learning framework to predict the geoeffectiveness and arrival time of CME events based on optical observations with satellites. The framework integrates these two requirements for the first time, eliminating the need for manual feature selection. Besides, the deep learning framework takes a series of solar white light observations as its sole input. [Fu et al. \(2021\)](#) gathered 2400 partial or full-halo CME events and related images from 1996 to 2018 for training and testing. The mean absolute error of the arrival time prediction is only 5.8 hours. Moreover, the F1 score and accuracy of the geoeffectiveness prediction reach 0.27% and 75.1%, respectively, on par with or even superior to the best results of conventional methods.

4.4 Potential challenges

While the existing literature has explored many CME prediction methods, significant disparities persist between prediction results and actual observations. These differences appear to be influenced by several factors, most of which are discussed in varying degrees of detail in the existing literature [Vourlidis et al. \(2019\)](#). In the following, we delve into some of the critical issues and potential challenges in CME prediction.

Critical issues in CME prediction encompass CME measurements and their interactions. The uncertainty and errors in CME measurements, particularly in terms of direction, velocity, and shape, significantly impact all CME predictions. Although high spatial resolution and high-sensitivity imaging of thousands of events are already available, measurement uncertainty and error remain high. We believe that there is also a need for a more careful interpretation of the imagery, utilizing existing datasets and analysis methods to reduce uncertainty and error. Regarding the interaction of CMEs with the background solar wind and other CMEs [Manchester et al. \(2017\)](#), the drag on CMEs is influenced by the background solar wind and other CMEs, resulting in diverse kinematic behaviors for different events [Patsourakos et al. \(2016\)](#); [Temmer et al. \(2011\)](#). This variation occurs because CMEs commonly undergo rotation, deflection (in latitude and longitude), fragmentation, or acceleration during propagation [Zhuang et al. \(2017\)](#). Hence, understanding the force balance of CMEs as they traverse the interplanetary medium is also a crucial consideration.

What potential challenges do we encounter in achieving more accurate predictions? Our recommendations include optimizing observing conditions, establishing consistent evaluation criteria, and refining complete physical modeling. To enhance observing conditions, deploying imagers with higher signal-to-noise imaging is suggested. To establish uniform evaluation criteria, the creation of a common CME event dataset and a set of model performance evaluation metrics is essential for facilitating comparisons across different techniques. Regarding complete physical models, enhancing the accuracy of modeling the background solar wind in MHD models is imperative.

5 SPEs Prediction

Solar Energetic Particle events (SEPs) are sharp increases in the particle flux from the Sun as observed in the near-earth environment. Solar Proton Events (SPEs) represent a major subclass of SEP events.

Proton fluxes, measured by GOES spacecraft at Geosynchronous orbit, are integral flux with 5-minute averages for energies greater than 10 MeV. When the proton fluxes equal or exceed the threshold (10 Particle Flux Units (pfu)), they are called Solar Proton Events (SPEs). The protons, whose energy is greater than 500 MeV, can penetrate the atmosphere of the Earth and finally reach the ground. They are called the Ground-Level Enhancement Event. The SPEs are the most dangerous phenomena in solar eruptive activities. They are harmful to the health of astronauts and can lead to single event effects, which are harmful to the safety of satellites. Therefore, the prediction of solar proton events is an important issue.

5.1 Definition of the problem

The SPEs are considered to be accelerated by the reconnection process of solar flares and shocks of CMEs. However, the detailed physical mechanism of SPEs is not yet clear. So far, the operational prediction models of SPEs are mostly based on statistical relationships between the predictors and the events.

The flowchart of SPEs prediction is shown in Figure 13. First of all, the necessary conditions of SPEs are determined. Secondly, all the samples that meet these conditions are selected. When at least one given SPE happens within a given prediction window, the sample is considered to be the positive sample, otherwise, it is considered to be the negative sample. The dataset, which consists of positive samples and negative samples, is divided into the training set and the testing set. And then, the prediction model is built on the training set, and its performance is evaluated on the testing set.

5.2 Predictors

It is well accepted that the SPEs consist of impulsive events, which are related to the acceleration mechanism of the solar flares, and gradual events, which are related to the acceleration mechanism of the CME shocks. Therefore, the predictors of SPEs are usually extracted from the features of solar flares, CMEs and their interplanetary propagation and evolution.

Laurenza et al. (2009) used the location of solar flare, the soft X-ray flux and the radio flux at ~ 1 MHz to predict the occurrence of SPEs. Kahler and Ling (2018) applied the soft and hard X-ray flare spectral variations as the precursors of SPEs. Xia et al. (2015) suggested that the radio type I noise storm, which is related to the magnetic reconnection of solar flares, can be taken as a proper precursors of SPEs. An interpretable machine learning technique was applied to build the SPEs prediction model in Núñez and Paul-Pena (2020). They found that the peak and integrated flux of soft X-ray flares and the radio type III bursts are the most discriminative parameters in the SPEs prediction model. Winter and Ledbetter (2015) pointed that type III radio bursts that occur along with a DH (decameter-hectometer) type II burst can be used as the precursors of SPEs. Kim et al. (2018) proposed the parameters of solar radio flux at 2800 MHz, 1415 MHz and 610 MHz for SPEs prediction. Núñez et al. (2019) used the observations of Extreme Ultraviolet (EUV) to predict SPEs. St. Cyr et al. (2017) investigated the prediction of SPEs by using real time coronagraphic observations. Papaioannou et al. (2018b) applied the characteristics of CMEs, for example, their width, direction and speed, to build the prediction model of SPEs. Considering the higher propagating speed of

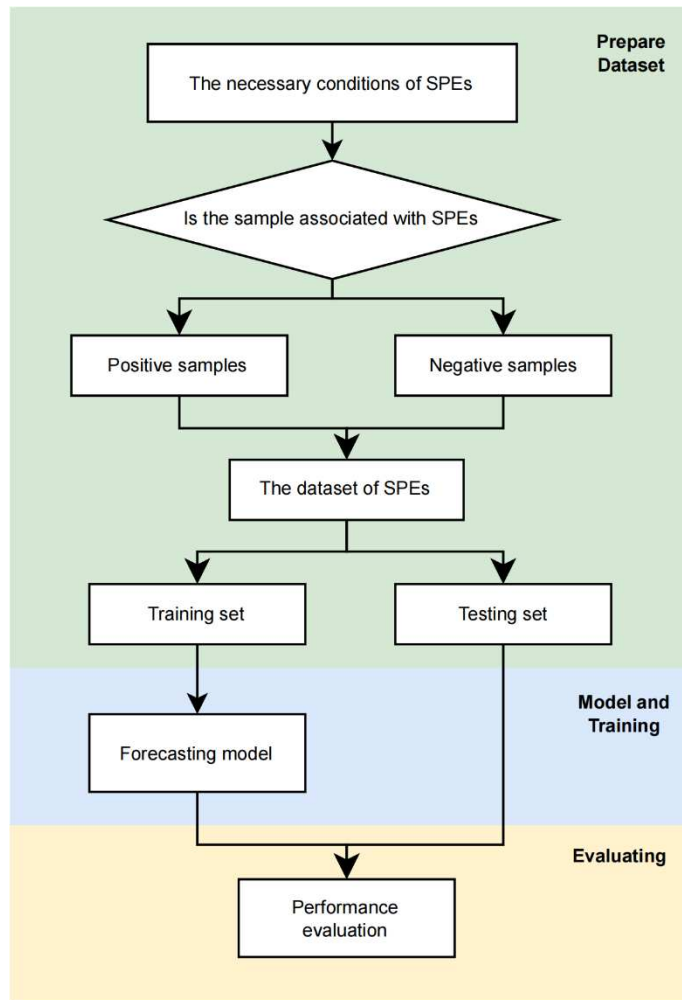


Figure 13 Flowchart Of SPEs prediction model.

Table 6 SPEs prediction models.

Model	Input	Output	Algorithm
PPS (Kahler et al., 2007)	Flare parameters	SPE flux	Empirical method
ESPERTA (Laurenza et al., 2018)	Flare parameters and the radio fluence	SPE flux	Logistic regression
UMASOD (Núñez and Paul-Pena, 2020)	Flare parameters and radio burst data	SPE event	Decision tree
FORSPEF (Papaioannou et al., 2015)	Flare parameters and CME parameters	SPE event and its profile	Statistical approach
SPRINTS (Engell et al., 2017)	Flare parameters	SPE profile	Statistical and decision tree models
PROSPER (Papaioannou et al., 2022)	Flare parameters and CME parameters	SPE event	statistical approach

electrons, Posner (2007) proposed the characteristics of electrons propagation can be used to predict the appearance of solar ion events or SPEs. To sum up, we can find that the characteristics of solar flares, CMEs and their related phenomena are usually considered as the predictors of SPEs.

5.3 Model

There are empirical, statistical, and machine learning methods are used to build the prediction model of SPEs, and some models have already been adopted in the practical applications of space weather prediction.

In 1980s, space environment forecast facility in the United States Air Force (USAF) developed the proton prediction system (PPS) (Smart and Shea, 1989). The time-intensity profiles of SPEs, even the heavy ion flux, could be provided in this prediction system. The performance of PPS was validated by Kahler et al. (2007). Laurenza et al. (2018) built the Empirical model for Solar Proton Event Real Time Alert (ESPERTA). The aim of ESPERTA is to provide timely warning for SPEs, therefore, the inputs of the model are parameters extracted from solar flares, but the CMEs related parameters are not be considered in this model. Alberti et al. (2017) tested the performance of ESPERTA from 2006 to 2014. And then, Alberti et al. (2019) applied the ESPERTA model to predict the SPEs from 2015 to 2017, and they found that the performance of the model is consistent with performance of the model tested from 1995 to 2014. Considering the class imbalanced problem in SPEs prediction, Stumpo et al. (2021) reinterpreted the ESPERTA model. The ESPERTA model was optimized by weighted and re-sampled techniques. Núñez (2011) presented a dual-model prediction system of SPEs, called University of Malaga Solar Energetic Particle (UMASEP). Based on the soft X-ray flux, differential proton flux and integral proton flux, UMASEP could output the probability of well and poorly connected SPEs. Further, Zucca et al. (2017) improved UMASEP by replacing the parameters of the soft X-ray time series with the microwave time series. Papaioannou et al. (2015) presented the Forecasting Solar Particle Events and Flares (FORSPEF) model. The proposed model consisted of dual system (flare prediction system and SPEs prediction system), the SPEs prediction system is based on the prediction results of solar flares. Engell et al. (2017) built the Space Radiation Intelligence System (SPRINTS) for SPEs prediction. SPRINTS provided the prediction results of SPEs by using GOES soft X-ray data. Papaioannou et al. (2022) built the Probabilistic Solar Particle Event forecasting model (PROSPER). This model estimated the probability of SPEs by using parameters of flare longitude, flare magnitude, CMEs width and CMEs speed. The SPEs prediction model used at Space Environment Center in National Oceanic and Atmospheric Administration (NOAA) since 1970s was introduced by Balch (1999). In this model, the probability of SPEs was estimated by using the peak flux and rise time of soft X-ray flux. Thereafter, Balch (2008) evaluated the performance of SPEs prediction model used at NOAA's Space Weather Prediction Center (SWPC) from 1986 to 2004. Recently, Bain et al. (2021) estimated the performance of SPEs prediction model currently used at NOAA's Space Weather Prediction Center (SWPC). The Brier Skill Score of the SPEs prediction model with 1 day prediction window is 0.25. And the probability of detection and the false alarm ratio are 91% and 24% for ≥ 10 MeV SPEs warnings. The evaluation index could be used as a benchmark for SPEs prediction model.

In order to improve the performance of the SPEs prediction model, some machine learning techniques, for example, logistic regression, neural networks, support vector machines and decision trees, are applied to build the automatic prediction model. Garcia (1994) applied logistic regression technique to build the prediction model of SPEs. Zhong et al. (2019) developed SPEs prediction model by using time series of soft X-ray flux and proton flux from 1990 to 2017. Richardson et al. (2018) proposed a simple empirical SPEs prediction model by using CMEs speed, CMEs direction and parameters of the related solar flares. The proposed model can be used to predict the intensity of SPEs at any location. Lovelace et al. (2018) built the prediction model of SPEs dose by using the kernel regression algorithm. Papaioannou et al. (2018a) applied the principal component analysis (PCA) algorithm to reduce the dimension of the dataset. Based on the new transformed parameters, a prediction model of SPEs was built. The dimension reduction algorithm is not widely used in SPEs prediction, mainly because the number of variables in the SPEs dataset is not very high. Núñez and Paul-Pena (2020) investigated a decision tree model for predicting SPEs using parameters extracted from solar flare and radio burst data. Decision tree algorithm for the prediction of SPEs is also applied by Laurenza et al. (2018). Jeong et al. (2014) developed a SPEs prediction model by training a multi-layer neural networks on the soft X-ray flux data from 1976 to 2011. Inceoglu et al. (2018) investigated two machine learning algorithms (multi-layer neural networks and support vector machines) to distinguish flares-CMEs with SPEs from flares-CMEs only. Huang et al. (2012) built an ensemble prediction model of SPEs, which combines the information of solar flares and CMEs. Lavasa et al. (2021) compared several machine learning algorithms (logistic regression, multi-layer neural networks, support vector machines, decision trees, random forests and extremely randomized trees) for SPEs prediction by using the parameters extracted from solar flares and CMEs. They found that random forests algorithm gets the best performance.

Driven by the big data, the deep learning model outperforms many shallow machine learning models. With the limitation on the amount of SPEs data, there are not many deep learning methods that can be used for building the prediction model of SPEs. Abdullah et al. (2022) applied deep learning method to build the SPEs prediction model for the first time. A bidirectional long short-term memory networks is used to distinguish active regions with SPEs from other active regions.

For more energetic event, Núñez (2015) built a prediction model of ≥ 100 MeV SPEs, and the integral proton flux can be provided. Ground-level events (GLEs) are SPEs with harder spectrum which can be detectable on Earth's surface by the specially designed detector. Hence they are a subset of SPEs (Shea and Smart, 2012). Núñez et al. (2017) developed the first prediction model of GLEs. Aminalragia-Giamini et al. (2020) studied the relationships of peak and fluence spectra in SPEs, and the fluence spectral is predicted. Last but not least, it shows that the deep learning algorithms have not been widely used to build SPEs prediction model, due to the limitation of data volume.

5.4 Potential challenges

From the point of view of machine learning, we can find that machine learning methods have been widely used to build the prediction model of SPEs, and provide state-of-the-art performance. The deep learning methods have also begun to be used for the SPEs prediction. There are two main challenges to build SPEs prediction model by using deep learning methods. (1) Low occurrence rate of events and limited positive samples. This limits the application of deep learning method to solve the SPEs prediction problem. (2) The source regions of SEPs are uncertain. SPEs are accelerated possibly by flares, CMEs, or the mixture of them. For the challenge (1), generating adversarial networks can be used to generate the event samples, this data augmentation technique is helpful in solving class imbalance problem. For the challenge (2), self supervised learning methods can be used to understand magnetic field of the Sun and extract the prediction patterns from full-disk magnetograms, these patterns can be used to build SPEs prediction model not only from source regions of solar flares, but also from source regions of CMEs.

6 Discussions

We present some important and interesting issues that need to be noted and discussed when machine learning methods are applied to build short-term prediction models of solar activities.

What are the disadvantages and advantages of deep learning prediction model? Compared to physical models, deep learning models cannot provide complete model interpretability and accurate reliability assessment. Generally, if a deep learning model performs well, that's usually considered satisfactory. However, we are not entirely sure why it performs well or whether it will continue to do so. Additionally, the deep learning requires significant computational power and data to train a prediction model. Furthermore, the execution of a deep learning model also requires substantial computational devices, which could limit their application scenarios.

The advantages of deep learning models include good generalization and engineering effectiveness. Deep learning models can be directly trained from a large amount of data, it can learn environmental disturbances as much as possible, thus generally exhibits better generalization. The physical models often need to meet some prerequisites, and due to environmental disturbances, they may not be ideal.

From a data perspective, the reliance of deep learning on the data is a double-edged sword. Deep learning models perform poorly under insufficient data conditions. However, they can model effectively with sufficient data. In the future, With the accumulation of observational data, there will be opportunities to improve the performance of deep learning models. From an engineering perspective, deep learning usually is an efficient modeling tool. After collecting data and preprocessing it with some background knowledge, satisfactory prediction results can be achieved.

What kinds of machine learning methods are suitable? The machine learning is a data-driven approach, hence the types of data is an important element. If the data is small, and some precursors have been extracted, some traditional machine learning methods such as decision tree, random forests, multi-layer perceptions, or even simple linear functions may achieve good results. If the data is one-dimensional time series, such as soft X-ray flux data, the RNN or transformer model which is

suitable for processing sequential data can often yield good results. If the data is images and the quantity is large enough, the convolutional neural networks or vision transformer models can often yield good results.

What needs to be done for building machine learning prediction model? For preparing dataset, you need to gather relevant data sources, clean the data to remove noise and inconsistencies, pre-process the data to make it suitable for the model, and split it into training, validation and testing sets.

For training model, you need to select an appropriate deep learning architecture, initialize the model parameters, define the loss function, choose an optimization algorithm, and train the model using the training dataset. If it's possible to design a suitable loss function by integrating the physical laws, the model can converge faster perform better.

For selecting verification metrics, you need to select evaluation metrics that are appropriate for your specific task, such as accuracy, precision, recall, F1 score, or others. Then, the parameters of the prediction model can be determined by the validation dataset, and the performance of the prediction model is evaluated by using the testing dataset to measure the effectiveness and generalization ability.

Why isn't there a standard dataset? There is still not a standard dataset in short-term solar activities prediction. Standard datasets could impose limitations on scalability, For example, a fixed pre-processing approach may lead to the loss of potential possibilities. However, standard datasets can allow researchers to focus on the study of modeling algorithms. Open accessed data and code are benefited for the long-term development of the entire community.

What's the impact of foundation model on solar activities predictions? From a physical perspective, analysis of the internal structure and learned features of foundation models could provide deeper insights into the underlying physical mechanisms of solar activities.

From an engineering perspective, foundation models may better capture the intricate patterns, resulting in higher prediction performance. Leveraging extensive data and complex model structures, foundation models may excel at identifying anomalous events or sudden phenomena within solar activities, improving the capability to detect unusual occurrences.

7 Conclusions

Solar eruptive activities mainly include solar flares, CMEs and SPEs. The numerical, statistical and machine learning methods have been widely applied to build the prediction models of solar eruptive activities. We gave a brief account to the development of solar eruptive activities prediction from the point of view of machine learning, and find that machine learning algorithms have made significant progress to predict the occurrence of solar flares, the arriving time of CMEs and the occurrence of SPEs. Benefited from the accumulation of solar flare related observational data, the deep learning algorithms have been extensively used to predict the solar flares and automatically extract the predictors of solar flares. Due to limitations in data volume, Deep learning has been used to predict the arriving time of CMEs and the occurrence of SPEs, but it has not been widely applied.

Although some important progresses have been made in solar flare prediction, it remains a challenging task. From a physical standpoint, the trigger of a solar flare is random, although the the energy has been accumulated to reach the critical state. Essentially, the current solar flare prediction model can only provide the probabilistic results. The small sample size problem, the class imbalance problem, the model design and interpretability, the fusion between physical model and deep learning model should be paid more attention.

Deep learning-based methods have been successfully applied to predict the arrival time of CMEs, but they still have substantial problems in several areas, which will also be significant research paths and development trends. Firstly, deep learning-based methods can directly extract features from observed images to construct prediction models. However, these techniques overlook various physical parameters of CMEs, including angular width, center position angle, mass, and kinetic energy. Since the physical parameters of CMEs are estimated by researchers using prior knowledge, they contain the most crucial information about CMEs, which is extremely important for producing more accurate prediction results. Through the

fusion of physical parameters and image data for multimodal learning, there exists the potential to enhance the performance of the model. Second, most available research divides CME detection, CME effectiveness prediction, and CME arrival time prediction into distinct tasks. This step-by-step approach has designed corresponding models for each task, allowing for refined optimization across different aspects. However, the CME effectiveness prediction task depends on the results of CME detection, and the CME arrival time prediction work also depends on the results of CME effectiveness prediction. Therefore, taking into account the interdependencies between these tasks, further integrating these tasks in future research may take us toward a more comprehensive CME prediction system. Furthermore, existing deep learning-based research methods primarily employ CME images collected by the SOHO satellite, ignoring images collected by other satellites like STEREO and SDO. Both the STEREO and SDO satellites have significantly improved their data-gathering efficiency and image clarity compared to the SOHO satellite. As a result, adding more satellite observation images could greatly enhance the model's performance. However, combining multiple observational photos and constructing subsequent models present challenges that require the gradual and sustained efforts of researchers to develop effective solutions.

Machine learning, especially deep learning, has not been widely used for proton event prediction. This is mainly due to the limited number of proton events, which is insufficient to train complex prediction models. In the future, the class imbalance problem and the few shot learning methods may play an important role in SPEs prediction.

Generally speaking, in addition to the improvements of modeling methods, machine learning methods, especially the deep learning methods, may play a greater role, including automatically extracting predictors from new observation data, designing the physics-informed neural networks, applying the foundation model for solar eruptive activities prediction, solving the class imbalance problems and small sample problems in the event prediction. Researches for data-driven prediction include four directions: data, model, interpretability, and engineering. Machine learning usually is a black box, but solar activity prediction is not only an engineering problem, but also a scientific problem. Hence it is essential to study the interpretability of the prediction model. The deployment of the prediction model is also an important issues, There are some good platforms now, eg. [Abduallah et al. \(2021\)](#); [Anastasiadis et al. \(2017\)](#); [Engell et al. \(2017\)](#); [Leka et al. \(2018a,b\)](#); [Muranushi et al. \(2015\)](#); [Murray et al. \(2017\)](#); [Nishizuka et al. \(2021\)](#).

From the data perspective, the large-scale and high-quality data are essential for improving the machine learning model. Data from different observational devices can be processed to form larger datasets. The big data makes it possible to train and test the complex deep neural networks. The standard dataset is necessary to estimate and compare the performance of prediction models. The same testing set is a prerequisite for comparing the performance metrics of the prediction model.

From the arithmetic perspective, the automatic feature extraction of the deep learning method could be used to explore the new predictors from the raw data of various observational devices. The foundation model should be designed to breaking through bottlenecks for the solar eruptive activities prediction. The physical model and machine learning model should be fused to improve the performance of the prediction models.

Acknowledgements. This work was supported by the Strategic Priority Research Program of the Chinese Academy of Sciences (Grant No. XDB0560000), the National Key R&D Program of China (Grant No. 2021YFA1600504), the National Natural Science Foundation of China (Grant No. 11873060). R.E. is grateful to Science and Technology Facilities Council (STFC, grant No. ST/M000826/1) UK. R.E. and M.B.K. acknowledge NKFIH (OTKA, grant No. K142987) Hungary for enabling this research. M.B.K. is grateful to ST/S000518/1, PIA.CE.RI. 2020-2022 Linea 2, CESAR 2020-35-HH.0, and UNKP-224-II-ELTE-186 grants. The authors also acknowledge the support from ISSI-Beijing for their project "Step forward in solar flare and coronal mass ejection (CME) forecasting". Thanks to the reviewers for their valuable comments and suggestions, which played an important role in the improvement and perfection of this article.

Conflict of interest. The authors declare that they have no conflict of interest.

References

Abduallah Y, Jordanova V K, Liu H, Li Q, Wang J T L, and Wang H. 2022. Predicting Solar Energetic Particles Using SDO/HMI Vector Magnetic Data Products and a Bidirectional LSTM Network. *Astrophys J Suppl Ser*, 260(1):16.

- Abduallah Y, Wang J T, Nie Y, Liu C, and Wang H. 2021. DeepSun: machine-learning-as-a-service for solar flare prediction. *Res Astron Astrophys*, 21(7):160.
- Abduallah Y, Wang J T L, Wang H, and Xu Y. 2023. Operational prediction of solar flares using a transformer-based framework. *Sci Rep*, 13(13665):1–11.
- Abed A K, Qahwaji R, and Abed A. 2021. The automated prediction of solar flares from SDO images using deep learning. *Adv Space Res*, 67(8):2544–2557.
- Abramenko V, Yurchyshyn V, Wang H, Spirock T, and Goode P. 2002. Scaling behavior of structure functions of the longitudinal magnetic field in active regions on the sun. *Astrophys J*, 577(1):487.
- Ahmed O W, Qahwaji R, Colak T, Higgins P A, Gallagher P T, and Bloomfield D S. 2013. Solar flare prediction using advanced feature extraction, machine learning, and feature selection. *Sol Phys*, 283:157–175.
- Alberti T, Laurenza M, and Cliver E W. 2019. Forecasting solar proton events by using the ESPERTA model. *Nuovo Cimento C Geophysics Space Physics C*, 42(1):40.
- Alberti T, Laurenza M, Cliver E W, Storini M, Consolini G, and Lepreti F. 2017. Solar Activity from 2006 to 2014 and Short-term Forecasts of Solar Proton Events Using the ESPERTA Model. *Astrophys J Suppl Ser*, 838(1):59.
- Alipour N, Mohammadi F, and Safari H. 2019. Prediction of Flares within 10 Days before They Occur on the Sun. *Astrophys J Suppl Ser*, 243(2):20.
- Aminalragia-Giamini S, Jiggins P, Anastasiadis A, Sandberg I, Aran A, Vainio R, Papadimitriou C, Papaioannou A, Tsiganos A, Paouris E, Vasalos G, Paassilta M, and Dierckxsens M. 2020. Prediction of Solar Proton Event Fluence spectra from their Peak flux spectra. *J Space Weather Space Clim*, 10:1.
- Anastasiadis A, Papaioannou A, Sandberg I, Georgoulis M, Tziotziou K, Kouloumvakos A, and Jiggins P. 2017. Predicting Flares and Solar Energetic Particle Events: The FORSPEF Tool. *Sol Phys*, 292(9):134.
- Angryk R A, Martens P C, Aydin B, Kempton D, Mahajan S S, Basodi S, Ahmadzadeh A, Cai X, Filali Boubrahimi S, Hamdi S M, Schuh M A, and Georgoulis M K. 2020. Multivariate time series dataset for space weather data analytics. *Sci Data*, 7(1):227.
- Aschwanden M. 2011. *Self-Organized Criticality in Astrophysics*. Springer, Berlin, Germany.
- Aschwanden M J. 2020. Global Energetics of Solar Flares. XI. Flare Magnitude Predictions of the GOES Class. *Astrophys J*, 897(1):16.
- Aschwanden M J and Aschwanden P D. 2008a. Solar flare geometries. i. the area fractal dimension. *Astrophys J*, 674(1):530.
- Aschwanden M J and Aschwanden P D. 2008b. Solar flare geometries. II. the volume fractal dimension. *Astrophys J*, 674(1):544.
- Aschwanden M J, Dennis B R, and Benz A O. 1998. Logistic avalanche processes, elementary time structures, and frequency distributions in solar flares. *Astrophys J*, 497(2):972.
- Asensio Ramos A, Cheung M C M, Chifu I, and Gafeira R. 2023. Machine learning in solar physics. *Living Rev Sol Phys*, 20(1):4.
- Asensio Ramos A, Cheung M C M, Chifu I, and Gafeira R. 2023. Machine learning in solar physics. *Living Rev Sol Phys*, 20(1):4–89.
- Atac T. 1987. Statistical relationship between sunspots and major flares. *Astrophys Space Sci*, 129:203–208.

- Bain H M, Steenburgh R A, Onsager T G, and Stitely E M. 2021. A Summary of National Oceanic and Atmospheric Administration Space Weather Prediction Center Proton Event Forecast Performance and Skill. *Space Weather*, 19(7):e2020SW002670.
- Bak P, Tang C, and Wiesenfeld K. 1987. Self-organized criticality: An explanation of the $1/f$ noise. *Phys Rev Lett*, 59(4):381.
- Balch C C. 1999. SEC proton prediction model: verification and analysis. *Radiat Meas*, 30(3):231–250.
- Balch C C. 2008. Updated verification of the Space Weather Prediction Center’s solar energetic particle prediction model. *Space Weather*, 6(1):S01001.
- Baranyi T, Gyori L, and Ludmány A. 2016. On-line Tools for Solar Data Compiled at the Debrecen Observatory and Their Extensions with the Greenwich Sunspot Data. *Sol Phys*, 291(9):3081–3102.
- Barnes G and Leka K. 2006. Photospheric Magnetic Field Properties of Flaring versus Flare-quiet Active Regions. III. Magnetic Charge Topology Models. *Astrophys J*.
- Barnes G and Leka K D. 2008. Evaluating the Performance of Solar Flare Forecasting Methods. *Astrophys J*, 688(2):L107.
- Barnes G, Leka K D, Schrijver C J, Colak T, Qahwaji R, Ashamari O W, Yuan Y, Zhang J, McAteer R T J, Bloomfield D S, Higgins P A, Gallagher P T, Falconer D A, Georgoulis M K, Wheatland M S, Balch C, Dunn T, and Wagner E L. 2016. A comparison of flare forecasting methods. I. results from the all-clear workshop. *Astrophys J*, 829(2):89.
- Bartkowiak A and Jakimiec M. 1994. Distance-based regression in prediction of solar flare activity. *Qüestiió: quaderns d’estadística i investigació operativa*, pages 7–38.
- Bélanger E, Vincent A, and Charbonneau P. 2007. Predicting solar flares by data assimilation in avalanche models: I. model design and validation. *Sol Phys*, 245:141–165.
- Benvenuto F, Piana M, Campi C, and Massone A M. 2018. A Hybrid Supervised/Unsupervised Machine Learning Approach to Solar Flare Prediction. *Astrophys J*, 853(1):90.
- Benz A O. 2008. Flare Observations. *Living Rev Sol Phys*, 5(1):1.
- Besliu-Ionescu D and Mierla M. 2021. Geoeffectiveness prediction of cmes. *Frontiers in Astronomy and Space Sciences*, 8:672203.
- Besliu-Ionescu D, Talpeanu D C, Mierla M, and Muntean G M. 2019. On the prediction of geoeffectiveness of cmes during the ascending phase of sc24 using a logistic regression method. *J Atmos Sol Terr Phys*, 193:105036.
- Bhattacharjee S, Alshehhi R, Dhuri D B, and Hanasoge S M. 2020. Supervised Convolutional Neural Networks for Classification of Flaring and Nonflaring Active Regions Using Line-of-sight Magnetograms. *Astrophys J*, 898(2):98.
- Bloomfield D S, Higgins P A, McAteer R T J, and Gallagher P T. 2012. Toward reliable benchmarking of solar flare forecasting methods. *Astrophys J Lett*, 747(2):L41.
- Bobra M G and Couvidat S. 2015. Solar flare prediction using sdo/hmi vector magnetic field data with a machine-learning algorithm. *Astrophys J*, 798(2):135.
- Bobra M G and Ilonidis S. 2016. Predicting coronal mass ejections using machine learning methods. *Astrophys J*, 821(2):127.
- Bobra M G, Sun X, Hoeksema J T, Turmon M, Liu Y, Hayashi K, Barnes G, and Leka K D. 2014. The Helioseismic and Magnetic Imager (HMI) Vector Magnetic Field Pipeline: SHARPs Space-Weather HMI Active Region Patches. *Sol Phys*, 289(9):3549–3578.

- Bobra M G, Wright P J, Sun X, and Turmon M J. 2021. SMARPs and SHARPs: Two Solar Cycles of Active Region Data. *Astrophys J Suppl Ser*, 256(2):26.
- Bornmann P and Shaw D. 1994. Flare rates and the mcintosh active-region classifications. *Sol Phys*, 150:127–146.
- Boteler D, Pirjola R, and Nevanlinna H. 1998. The effects of geomagnetic disturbances on electrical systems at the earth's surface. *Adv Space Res*, 22(1):17–27.
- Brueckner G, Howard R, Koomen M, Korendyke C, Michels D, Moses J, Socker D, Dere K, Lamy P, Llebaria A, et al.. 1995. The large angle spectroscopic coronagraph (lasco) visible light coronal imaging and spectroscopy. *The SOHO mission*, pages 357–402.
- Campi C, Benvenuto F, Massone A M, Bloomfield D S, Georgoulis M K, and Piana M. 2019. Feature Ranking of Active Region Source Properties in Solar Flare Forecasting and the Uncompromised Stochasticity of Flare Occurrence. *Astrophys J*, 883(2):150.
- Camporeale E. 2019. The Challenge of Machine Learning in Space Weather: Nowcasting and Forecasting. *Space Weather*, 17(8):1166–1207.
- Chen A, Ye Q, and Wang J. 2021. Flare Index Prediction with Machine Learning Algorithms. *Sol Phys*, 296(10):150.
- Chen J, Li W, Li S, Chen H, Zhao X, Peng J, Chen Y, and Deng H. 2022. Two-Stage Solar Flare Forecasting Based on Convolutional Neural Networks. *Space: Science & Technology*, 2022:1–10.
- Chen Y, Maloney S, Camporeale E, Huang X, and Zhou Z. 2023. Editorial: Machine learning and statistical methods for solar flare prediction. *Front Astron Space Sci*, 10:1121615.
- Chen Y, Manchester W B, Hero A O, Toth G, DuFumier B, Zhou T, Wang X, Zhu H, Sun Z, and Gombosi T I. 2019. Identifying Solar Flare Precursors Using Time Series of SDO/HMI Images and SHARP Parameters. *Space Weather*, 17(10):1404–1426.
- Choi S, Moon Y J, Vien N A, and Park Y D. 2012. Application of support vector machine to the prediction of geo-effective halo cmes. *J Korean Astron Soc*, 45:31–38.
- Cicogna D, Berrilli F, Calchetti D, Del Moro D, Giovannelli L, Benvenuto F, Campi C, Guastavino S, and Piana M. 2021. Flare-forecasting Algorithms Based on High-gradient Polarity Inversion Lines in Active Regions. *Astrophys J*, 915(1):38.
- Cinto T, Gradwohl A L S, Coelho G P, and da Silva A E A. 2020. Solar Flare Forecasting Using Time Series and Extreme Gradient Boosting Ensembles. *Sol Phys*, 295(7):93.
- Colak T and Qahwaji R. 2009. Automated Solar Activity Prediction: A hybrid computer platform using machine learning and solar imaging for automated prediction of solar flares. *Space Weather*, 7(6).
- Conlon P, Gallagher P, McAteer R, Ireland J, Young C, Kestener P, Hewett R, and Maguire K. 2008. Multifractal properties of evolving active regions. *Sol Phys*, 248:297–309.
- Corona-Romero P, Gonzalez-Esparza J, Aguilar-Rodriguez E, De-la Luz V, and Mejia-Ambriz J. 2015. Kinematics of icmes/shocks: blast wave reconstruction using type-ii emissions. *Sol Phys*, 290:2439–2454.
- Corona-Romero P, Gonzalez-Esparza J, Perez-Alanis C, Aguilar-Rodriguez E, de-la Luz V, and Mejia-Ambriz J. 2017. Calculating travel times and arrival speeds of cmes to earth: An analytic tool for space weather forecasting. *Space Weather*, 15(3):464–483.
- Covas E. 2020. Transfer learning in spatialtemporal forecasting of the solar magnetic field. *Astron Nachr*, 341(4):384–394.
- Cozad A, Sahinidis N V, and Miller D C. 2015. A combined first-principles and data-driven approach to model building. *Comput Chem Eng*, 73:116–127.

- Cui Y, Li R, Wang H, and He H. 2007. Correlation between Solar Flare Productivity and Photospheric Magnetic Field Properties II. Magnetic Gradient and Magnetic Shear. *Sol Phys*, 242(1):1–8.
- Cui Y, Li R, Zhang L, He Y, and Wang H. 2006. Correlation between solar flare productivity and photospheric magnetic field properties: 1. maximum horizontal gradient, length of neutral line, number of singular points. *Sol Phys*, 237:45–59.
- Delaboudiniere J P, Artzner G, Brunaud J, Gabriel A H, Hochedez J F, Millier F, Song X, Au B, Dere K, Howard R A, et al.. 1995. Eit: extreme-ultraviolet imaging telescope for the soho mission. *The SOHO Mission*, pages 291–312.
- Deng Z, Wang F, Deng H, Tan L, Deng L, and Feng S. 2021. Fine-grained Solar Flare Forecasting Based on the Hybrid Convolutional Neural Networks*. *Astrophys J*, 922(2):232.
- Deshmukh V, Flyer N, van der Sande K, and Berger T. 2022. Decreasing False-alarm Rates in CNN-based Solar Flare Prediction Using SDO/HMI Data. *Astrophys J Suppl Ser*, 260(1):9.
- Dhuri D B, Hanasoge S M, Birch A C, and Schunker H. 2020. Application and Interpretation of Deep Learning for Identifying Pre-emergence Magnetic Field Patterns. *Astrophys J*, 903(1):27.
- Dhuri D B, Hanasoge S M, and Cheung M C M. 2019. Machine learning reveals systematic accumulation of electric current in lead-up to solar flares. *Proceedings of the National Academy of Sciences*, 116(23):11141–11146.
- Domingo V, Fleck B, and Poland A I. 1995. The soho mission: an overview. *Sol Phys*, 162:1–37.
- Dou F, Xu L, Ren Z, Zhao D, and Zhang X. 2022. Super-resolution of Solar Magnetograms Using Deep Learning. *Res Astron Astrophys*, 22(8):085018.
- Drake J F. 1971. Characteristics of soft solar x-ray bursts. *Sol Phys*, 16:152–185.
- Dumbovic M, Calogović J, Vršnak B, Temmer M, Mays M L, Veronig A, and Piantchitsch I. 2018. The drag-based ensemble model (dbem) for coronal mass ejection propagation. *Astrophys J*, 854(2):180.
- Ellison M. 1963. Solar flares and associated phenomena. *Planet Space Sci*, 11(6):597–619.
- Engell A J, Falconer D A, Schuh M, Loomis J, and Bissett D. 2017. SPRINTS: A Framework for Solar-Driven Event Forecasting and Research. *Space Weather*, 15(10):1321–1346.
- Erdélyi R, Korsós M B, Huang X, Yang Y, Pizzey D, Wrathmall S A, Hughes I G, Dyer M J, Dhillon V S, Belucz B, Braja R, Chatterjee P, Cheng X, Deng Y, Domínguez S V, Joya R, Gömöry P, Gyenge N G, Hanslmeier A, Kucera A, Kuridze D, Li F, Liu Z, Xu L, Mathioudakis M, Matthews S, McAteer J R, Pevtsov A A, Pötzi W, Romano P, Shen J, Temesváry J, Tlatov A G, Triana C, Utz D, Veronig A M, Wang Y, Yan Y, Zaqarashvili T, and Zuccarello F. 2022. The Solar Activity Monitor Network SAMNet. *J Space Weather Space Clim*, 12:2.
- F. C. Durant I. 1965. *Astronautics and Aeronautics, 1963. Chronology on Science, Technology, and Policy. National Aeronautics and Space Administration Historical Staff | Isis: Vol 56, No 3*. John Hopkins Press.
- Falco M, Costa P, and Romano P. 2019a. Solar flare forecasting using morphological properties of sunspot groups. *J Space Weather Space Clim*, 9:A22.
- Falco M, Costa P, and Romano P. 2019b. Solar flare forecasting using photospheric active region properties. *Il Nuovo Cimento C*, 42(1):1–3.
- Florios K, Kontogiannis I, Park S H, Guerra J A, Benvenuto F, Bloomfield D S, and Georgoulis M K. 2018. Forecasting Solar Flares Using Magnetogram-based Predictors and Machine Learning. *Sol Phys*, 293(2):28.

- Fozzard R, Bradshaw G, and Ceci L. 1988. A connectionist expert system that actually works. *Adv Neural Inf Process Syst*, 1.
- Fu H, Zheng Y, Ye Y, Feng X, Liu C, and Ma H. 2021. Joint geoeffectiveness and arrival time prediction of cmes by a unified deep learning framework. *Remote Sens*, 13(9):1738.
- Gallagher P T, Moon Y J, and Wang H. 2002. Active-Region Monitoring and Flare Forecasting – I. Data Processing and First Results. *Sol Phys*, 209(1):171–183.
- Gao F, Liu T, Sun W, and Xu L. 2023. Generating Space-based SDO/HMI-like Solar Magnetograms from Ground-based Ha Images by Deep Learning. *Astrophys J Suppl Ser*, 266(2):19.
- Garcia H A. 1994. Temperature and Hard X-Ray Signatures for Energetic Proton Events. *Astrophys J Suppl Ser*, 420:422.
- Georgoulis M K. 2008. Magnetic complexity in eruptive solar active regions and associated eruption parameters. *Geophys Res Lett*, 35(6).
- Georgoulis M K. 2012. Are Solar Active Regions with Major Flares More Fractal, Multifractal, or Turbulent Than Others? *Sol Phys*, 276(1):161–181.
- Georgoulis M K. 2013. Toward an Efficient Prediction of Solar Flares: Which Parameters, and How? *Entropy*, 15(11):5022–5052.
- Georgoulis M K, Bloomfield D S, Piana M, Massone A M, Soldati M, Gallagher P T, Pariat E, Vilmer N, Buchlin E, Baudin F, Csillaghy A, Sathiapal H, Jackson D R, Alingery P, Benvenuto F, Campi C, Florios K, Gontikakis C, Guennou C, Guerra J A, Kontogiannis I, Latorre V, Murray S A, Park S H, von Stachelski S, Torbica A, Vischi D, and Worsfold M. 2021. The flare likelihood and region eruption forecasting (FLARECAST) project: flare forecasting in the big data & machine learning era. *J Space Weather Space Clim*, 11:39.
- Georgoulis M K, Nindos A, and Zhang H. 2019. The source and engine of coronal mass ejections. *Phil Trans R Soc A*, 377(2148):20180094.
- Georgoulis M K and Rust D M. 2007. Quantitative forecasting of major solar flares. *Astrophys J*, 661(1):L109.
- Georgoulis M K, Tziotziou K, and Raouafi N E. 2012. MAGNETIC ENERGY AND HELICITY BUDGETS IN THE ACTIVREGION SOLAR CORONA. II. NONLINEAR FORCE-FREE APPROXIMATION. *Astrophys J*, 759(1):1.
- Giovannelli R G. 1939. The relations between eruptions and sunspots. *Astrophys J*, 89:555.
- Gombosi T I, Chen Y, Gloer A, Huang Z, Jia X, Liemohn M W, Manchester W B, Pulkkinen T, Sachdeva N, Al Shidi Q, Sokolov I V, Szente J, Tenishev V, Toth G, van der Holst B, Welling D T, Zhao L, and Zou S. 2021. What sustained multi-disciplinary research can achieve: The space weather modeling framework. *J Space Weather Space Clim*, 11:42.
- Gopalswamy N, Lara A, Yashiro S, Kaiser M L, and Howard R A. 2001. Predicting the 1-au arrival times of coronal mass ejections. *J Geophys Res Space Phys*, 106(A12):29207–29217.
- Gopalswamy N, Yashiro S, Lara A, Kaiser M, Thompson B, Gallagher P, and Howard R. 2003. Large solar energetic particle events of cycle 23: A global view. *Geophys Res Lett*, 30(12).
- Gosling J, McComas D, Phillips J, and Bame S. 1991. Geomagnetic activity associated with earth passage of interplanetary shock disturbances and coronal mass ejections. *J Geophys Res Space Phys*, 96(A5):7831–7839.
- Guastavino S, Marchetti F, Benvenuto F, Campi C, and Piana M. 2022. Implementation paradigm for supervised flare forecasting studies: A deep learning application with video data. *Astronomy & Astrophysics*, 662:A105.

- Guerra J A, Murray S A, Shaun Bloomfield D, and Gallagher P T. 2020. Ensemble forecasting of major solar flares: methods for combining models. *J Space Weather Space Clim*, 10:38.
- Gyenge N, Ludmány A, and Baranyi T. 2016. Active longitude and solar flare occurrences. *Astrophys J*, 818(2):127.
- Hale G E, Ellerman F, Nicholson S B, and Joy A H. 1919. The Magnetic Polarity of Sun-Spots. *Astrophys J*, 49:153.
- Han K, Yu M Y, Fu J F, Ling W B, Zheng D q, Wan J, and Peng E. 2023. Research Progress on Solar Flare Forecast Methods Based on Data-driven Models. *Res Astron Astrophys*, 23(6):065002.
- He K, Zhang X, Ren S, and Sun J. 2016. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778.
- Heidke P. 1926. Berechnung des erfolges und der güte der windstärkevorhersagen im sturmwarnungsdienst. *Geografiska Annaler*, 8:301–349.
- Howard R A, Moses J, Vourlidas A, Newmark J, Socker D G, Plunkett S P, Korendyke C M, Cook J, Hurley A, Davila J, et al.. 2008. Sun earth connection coronal and heliospheric investigation (secchi). *Space Sci Rev*, 136:67–115.
- Huang X, Wang H, and Dai X. 2012. Influences of misprediction costs on solar flare prediction. *Science China Physics, Mechanics and Astronomy*, 55:1956–1962.
- Huang X, Wang H, Xu L, Liu J, Li R, and Dai X. 2018. Deep Learning Based Solar Flare Forecasting Model. I. Results for Line-of-sight Magnetograms. *Astrophys J*, 856(1):7.
- Huang X, Wang H N, and Li L P. 2012. Ensemble prediction model of solar proton events associated with solar flares and coronal mass ejections. *Res Astron Astrophys*, 12(3):313–321.
- Huang X, Zhang L, Wang H, and Li L. 2013. Improving the performance of solar flare prediction using active longitudes information. *Astronomy & Astrophysics*, 549:A127.
- Hundhausen A J, Sawyer C B, House L, Illing R, and Wagner W. 1984. Coronal mass ejections observed during the solar maximum mission: Latitude distribution and rate of occurrence. *J Geophys Res Space Phys*, 89(A5):2639–2646.
- Inceoglu F, Jeppesen J H, Kongstad P, Hernández Marcano N J, Jacobsen R H, and Karoff C. 2018. Using Machine Learning Methods to Forecast if Solar Flares Will Be Associated with CMEs and SEPs. *Astrophys J Suppl Ser*, 861(2):128.
- Jeong E J, Lee J Y, Moon Y J, and Park J. 2014. Forecast of Solar Proton Events with NOAA Scales Based on Solar X-Ray Flare Data Using Neural Network. *Journal of Korean Astronomical Society*, 47(6):209–214.
- Jiang C, Bian X, Sun T, and Feng X. 2021. MHD Modeling of Solar Coronal Magnetic Evolution Driven by Photospheric Flow. *Front Phys*, 9:646750.
- Jiao Z, Sun H, Wang X, Manchester W, Gombosi T, Hero A, and Chen Y. 2020. Solar Flare Intensity Prediction With Machine Learning Models. *Space Weather*, 18(7).
- Jolliffe I T. 2007. Uncertainty and Inference for Verification Measures. *Weather Forecasting*, 22(3):637–650.
- Jonas E, Bobra M, Shankar V, Todd Hoeksema J, and Recht B. 2018. Flare Prediction Using Photospheric and Coronal Image Data. *Sol Phys*, 293(3):48.
- Jungbluth A, Gitiaux X, Maloney S, Shneider C, Wright P, Kalaitzis F, Deudon M, Baydin A G, Gal Y, and Muñoz-Jaramillo A. 2019. Single-Frame Super-Resolution of Solar Magnetograms: Investigating Physics-Based Metrics & Losses. *ArXiv*.
- Kahler S W, Cliver E W, and Ling A G. 2007. Validating the proton prediction system (PPS). *J Atmos Sol Terr Phys*, 69(12):43–49.

- Kahler S W and Ling A G. 2018. Forecasting Solar Energetic Particle (SEP) events with Flare X-ray peak ratios. *J Space Weather Space Clim*, 8:A47.
- Karakatsanis L and Pavlos G. 2008. Soc and chaos into the solar activity. *Nonlinear Phenomena in Complex Systems*, 11(2):280–284.
- Kay C and Gopalswamy N. 2017. Using the coronal evolution to successfully forward model cmes’ in situ magnetic profiles. *J Geophys Res Space Phys*, 122(12):11–810.
- Kay C, Gopalswamy N, Reinard A, and Opher M. 2017. Predicting the magnetic field of earth-impacting cmes. *Astrophys J*, 835(2):117.
- Kilpua E, Mierla M, Rodriguez L, Zhukov A, Srivastava N, and West M. 2012. Estimating travel times of coronal mass ejections to 1 au using multi-spacecraft coronagraph data. *Sol Phys*, 279:477–496.
- Kim K N, Sin S A, Song K A, and Kong J H. 2018. A technique for prediction of SPEs from solar radio flux by statistical analysis, ANN and GA. *Astrophys Space Sci*, 363(8):170.
- Kim T, Park E, Lee H, Moon Y J, Bae S H, Lim D, Jang S, Kim L, Cho I H, Choi M, and Cho K S. 2019. Solar farside magnetograms from deep learning analysis of STEREO/EUVI data. *Nat Astron*, 3(5):397–400.
- Komm R and Hill F. 2009. Solar flares and solar subphotospheric vorticity. *J Geophys Res Space Phys*, 114(A6).
- Korsós M B, Chatterjee P, and Erdélyi R. 2018. Applying the Weighted Horizontal Magnetic Gradient Method to a Simulated Flaring Active Region. *Astrophys J Suppl Ser*, 857(2):103.
- Korsós M B, Georgoulis M K, Gyenge N, Bisoi S K, Yu S, Poedts S, Nelson C J, Liu J, Yan Y, and Erdélyi R. 2020a. Solar Flare Prediction Using Magnetic Field Diagnostics above the Photosphere. *Astrophys J*, 896(2):119.
- Korsós M B, Romano P, Morgan H, Ye Y, Erdélyi R, and Zuccarello F. 2020b. Differences in Periodic Magnetic Helicity Injection Behavior between Flaring and Non-flaring Active Regions: Case Study. *Astrophys J*, 897(2):L23.
- Krista L D and Chih M. 2021. A DEFT Way to Forecast Solar Flares. *Astrophys J*, 922(2):218.
- Krizhevsky A, Sutskever I, and Hinton G E. 2017. ImageNet classification with deep convolutional neural networks. *Commun ACM*, 60(6):84–90.
- Kubo Y, Den M, and Ishii M. 2017. Verification of operational solar flare forecast: Case of Regional Warning Center Japan. *J Space Weather Space Clim*, 7:A20.
- Kutsenko A S, Abramenko V I, and Kutsenko O K. 2021. On the possibility of probing the flare productivity of an active region in the early stage of emergence. *Mon Not R Astron Soc*, 501(4):6076–6082.
- Landa V and Reuveni Y. 2022. Low-dimensional Convolutional Neural Network for Solar Flares GOES Time-series Classification. *Astrophys J Suppl Ser*, 258(1):12.
- Laurenza M, Alberti T, and Cliver E W. 2018. A Short-term ESPERTA-based Forecast Tool for Moderate-to-extreme Solar Proton Events. *Astrophys J Suppl Ser*, 857(2):107.
- Laurenza M, Cliver E W, Hewitt J, Storini M, Ling A G, Balch C C, and Kaiser M L. 2009. A technique for short-term warning of solar energetic particle events based on flare location, flare size, and evidence of particle escape. *Space Weather*, 7(4):S04008.
- Lavasa E, Giannopoulos G, Papaioannou A, Anastasiadis A, Daglis I A, Aran A, Pacheco D, and Sanahuja B. 2021. Assessing the Predictability of Solar Energetic Particles with the Use of Machine Learning Techniques. *Sol Phys*, 296(7):107.

- Lawrence J, Ruzmaikin A, and Cadavid A. 1993. Multifractal measure of the solar magnetic field. *Astrophys J*, 417:805.
- Leka K, Barnes G, and Wagner E. 2018a. The NWRA Classification Infrastructure: description and extension to the Discriminant Analysis Flare Forecasting System (DAFFS). *J Space Weather Space Clim*, 8:A25.
- Leka K D and Barnes G. 2003. Photospheric Magnetic Field Properties of Flaring versus Flare-quiet Active Regions. I. Data, General Approach, and Sample Results. *Astrophys J*, 595(2):1277.
- Leka K D and Barnes G. 2007. Photospheric Magnetic Field Properties of Flaring versus Flare-quiet Active Regions. IV. A Statistically Significant Sample. *Astrophys J*, 656(2):1173.
- Leka K D, Barnes G, Birch A C, Gonzalez-Hernandez I, Dunn T, Javornik B, and Braun D C. 2012. HELIOSEISMOLOGY OF PRE-EMERGING ACTIVE REGIONS. I. OVERVIEW, DATA, AND TARGET SELECTION CRITERIA. *Astrophys J*, 762(2):130.
- Leka K D, Barnes G, and Wagner E. 2018b. The NWRA Classification Infrastructure: description and extension to the Discriminant Analysis Flare Forecasting System (DAFFS). *J Space Weather Space Clim*, 8:A25.
- Leka K D, Fan Y, and Barnes G. 2005. On the Availability of Sufficient Twist in Solar Active Regions to Trigger the Kink Instability. *Astrophys J*, 626(2):1091.
- Leka K D, Park S H, Kusano K, Andries J, Barnes G, Bingham S, Bloomfield D S, McCloskey A E, Delouille V, Falconer D, Gallagher P T, Georgoulis M K, Kubo Y, Lee K, Lee S, Lobzin V, Mun J, Murray S A, Hamad Nageem T A M, Qahwaji R, Sharpe M, Steenburgh R A, Steward G, and Terkildsen M. 2019a. A Comparison of Flare Forecasting Methods. II. Benchmarks, Metrics, and Performance Results for Operational Solar Flare Forecasting Systems. *Astrophys J Suppl Ser*, 243(2):36.
- Leka K D, Park S H, Kusano K, Andries J, Barnes G, Bingham S, Bloomfield D S, McCloskey A E, Delouille V, Falconer D, Gallagher P T, Georgoulis M K, Kubo Y, Lee K, Lee S, Lobzin V, Mun J, Murray S A, Hamad Nageem T A M, Qahwaji R, Sharpe M, Steenburgh R A, Steward G, and Terkildsen M. 2019b. A Comparison of Flare Forecasting Methods. III. Systematic Behaviors of Operational Solar Flare Forecasting Systems. *Astrophys J*, 881(2):101.
- Lemen J R, Title A M, Akin D J, Boerner P F, Chou C, Drake J F, Duncan D W, Edwards C G, Friedlaender F M, Heyman G F, et al.. 2012. The atmospheric imaging assembly (aia) on the solar dynamics observatory (sdo). *Sol Phys*, 275:17–40.
- Li M, Cui Y, Luo B, Ao X, Liu S, Wang J, Li S, Du C, Sun X, and Wang X. 2022. Knowledge-Informed Deep Neural Networks for Solar Flare Forecasting. *Space Weather*, 20(8):e2021SW002985.
- Li M, Cui Y, Luo B, Wang J, and Wang X. 2023. Deep neural networks of solar flare forecasting for complex active regions. *Front Astron Space Sci*, 10:1177550.
- Li R and Du Y. 2019. Full-Disk Solar Flare Forecasting Model Based on Data Mining Method. *Advances in Astronomy*, 2019:1–6.
- Li R and Huang X. 2018. Solar flare forecasting model based on automatic feature extraction of sunspots. *scientia sinica Physica, Mechanica & Astronomica*, 48(11):119601.
- Li R, Zhu J, and Cui Y. 2013. Solar flare prediction model combining photospheric magnetic field coefficients and black spot coefficients in active regions. *Scientific Bulletin*, 58(19):1845–1850.
- Li X, Zheng Y, Wang X, and Wang L. 2020. Predicting Solar Flares Using a Novel Deep Convolutional Neural Network. *Astrophys J*, 891(1):10.
- Lim D, Moon Y J, Park E, Park J, Lee K, Lee J Y, and Jang S. 2019a. Ensemble Forecasting of Major Solar Flares with Short-, Mid-, and Long-term Active Region Properties. *Astrophys J*, 885(1):35.

- Lim D, Moon Y J, Park J, Park E, Lee K, Lee J Y, and Jang S. 2019b. Forecast of daily major flare probability using relationships between vector magnetic properties and flaring rates. *Journal of The Korean Astronomical Society*, 52(4):133–144.
- Lin J. 2009. Studies of solar flares and cmes related to the space solar missions in the future. *Science in China Series G: Physics, Mechanics and Astronomy*, 52(11):1646–1654.
- Lin Y. 2000. Introduction to solar physics.
- Liu C, Deng N, Wang J T L, and Wang H. 2017a. Predicting Solar Flares Using SDO/HMI Vector Magnetic Data Products and the Random Forest Algorithm. *Astrophys J*, 843(2):104.
- Liu H, Liu C, Wang J T L, and Wang H. 2019. Predicting Solar Flares Using a Long Short-term Memory Network. *Astrophys J*, 877(2):121.
- Liu J, Wang Y, Huang X, Korsós M B, Jiang Y, Wang Y, and Erdélyi R. 2021. Reliability of AI-generated magnetograms from only EUV images. *Nat Astron*, 5(2):108–110.
- Liu J, Ye Y, Shen C, Wang Y, and Erdélyi R. 2018. A new tool for cme arrival time prediction using machine learning algorithms: Cat-puma. *Astrophys J*, 855(2):109.
- Liu J F, Li F, Zhang H P, and Yu D R. 2017b. Short-term solar flare prediction using image-case-based reasoning. *Res Astron Astrophys*, 17(11):116.
- Liu S, Wang J, Li M, Cui Y, Guo J, Shi Y, Luo B, and Liu S. 2023. A selective up-sampling method applied upon unbalanced data for flare prediction: potential to improve model performance. *Front Astron Space Sci*, 10:1082694.
- Liu S, Xu L, Zhao Z, Erdélyi R, Korsós M B, and Huang X. 2022. Deep Learning Based Solar Flare Forecasting Model. II. Influence of Image Resolution. *Astrophys J*, 941(1):20.
- Lovelace A M, Rashid A M, de Wet W C, Townsend L W, Wesley Hines J, and Moussa H. 2018. Solar Particle Event Dose Forecasting Using Regression Techniques. *Space Weather*, 16(8):1073–1085.
- Lu E T and Hamilton R J. 1991. Avalanches and the distribution of solar flares. *Astrophys J*, 380:L89–L92.
- Machol J L, Eparvier F G, Viereck R A, Woodraska D L, Snow M, Thiemann E, Woods T N, McClintock W E, Mueller S, Eden T D, Meisner R, Codrescu S, Bouwer S D, and Reinard A A. 2020. Chapter 19 - goes-r series solar x-ray and ultraviolet irradiance. In Goodman S J, Schmit T J, Daniels J, and Redmon R J, editors, *The GOES-R Series*, pages 233–242. Elsevier.
- Manchester W, Kilpua E K, Liu Y D, Lugaz N, Riley P, Török T, and Vršnak B. 2017. The physical processes of cme/icme evolution. *Space Sci Rev*, 212:1159–1219.
- Mason J P and Hoeksema J T. 2010. Testing automated solar flare forecasting with 13 years of michelson doppler imager magnetograms. *Astrophys J*, 723(1):634.
- Mays M, Taktakishvili A, Pulkkinen A, MacNeice P, Rastätter L, Odstrcil D, Jian L, Richardson I, LaSota J, Zheng Y, et al.. 2015. Ensemble modeling of cmes using the wsa–enlil+ cone model. *Sol Phys*, 290:1775–1814.
- McAteer R J, Gallagher P T, and Ireland J. 2005. Statistics of active region complexity: A large-scale fractal dimension survey. *Astrophys J*, 631(1):628.
- McCulloch W S and Pitts W. 1943. A logical calculus of the ideas immanent in nervous activity. *Bull. Math. Biophys.*, 5(4):115–133.
- McIntosh P S. 1990. The classification of sunspot groups. *Sol Phys*, 125:251–267.

- Meunier N. 2004. Complexity of magnetic structures: Flares and cycle phase dependence. *Astronomy & Astrophysics*, 420(1):333–342.
- Millward G, Biesecker D, Pizzo V, and De Koning C. 2013. An operational software tool for the analysis of coronagraph images: Determining cme parameters for input into the wsa-enlil heliospheric model. *Space Weather*, 11(2):57–68.
- Morales L F and Santos N A. 2020. Predicting Extreme Solar Flare Events Using Lu and Hamilton Avalanche Model. *Sol Phys*, 295(11):155.
- Möstl C, Isavnin A, Boakes P, Kilpua E, Davies J, Harrison R, Barnes D, Krupar V, Eastwood J, Good S, et al.. 2017. Modeling observations of solar coronal mass ejections with heliospheric imagers verified with the heliophysics system observatory. *Space Weather*, 15(7):955–970.
- Muranushi T, Shibayama T, Muranushi Y H, Isobe H, Nemoto S, Komazaki K, and Shibata K. 2015. UFCORIN: A fully automated predictor of solar flares in GOES Xray flux. *Space Weather*, 13(11):778–796.
- Murray S A, Bingham S, Sharpe M, and Jackson D R. 2017. Flare forecasting at the Met Office Space Weather Operations Centre: MOSWOC FLARE FORECASTING. *Space Weather*, 15(4):577–588.
- Ning Z. 2009. The investigation of the neupert effect in two solar flares. *Science in China Series G: Physics, Mechanics and Astronomy*, 52(11):1686–1690.
- Ning Z. 2012. Power conversion factor in solar flares. *Chinese Chin Sci Bull*, 57:1397–1404.
- Nishizuka N, Kubo Y, Sugiura K, Den M, and Ishii M. 2021. Operational solar flare prediction model using Deep Flare Net. *Earth, Planets and Space*, 73(1):64.
- Nishizuka N, Sugiura K, Kubo Y, Den M, and Ishii M. 2018. Deep Flare Net (DeFN) Model for Solar Flare Prediction. *Astrophys J*, 858(2):113.
- Nishizuka N, Sugiura K, Kubo Y, Den M, Watari S, and Ishii M. 2017. Solar Flare Prediction Model with Three Machinelearning Algorithms using Ultraviolet Brightening and Vector Magnetograms. *Astrophys J*, 835(2):156.
- Núñez M. 2011. Predicting solar energetic proton events ($E > 10$ MeV). *Space Weather*, 9(7):S07003.
- Núñez M. 2015. Real-time prediction of the occurrence and intensity of the first hours of >100 MeV solar energetic proton events. *Space Weather*, 13(11):807–819.
- Núñez M, Nieves-Chinchilla T, and Pulkkinen A. 2019. Predicting well-connected SEP events from observations of solar EUVs and energetic protons. *J Space Weather Space Clim*, 9:A27.
- Núñez M and Paul-Pena D. 2020. Predicting >10 MeV SEP Events from Solar Flare and Radio Burst Data. *Universe*, 6(10):161.
- Núñez M, Reyes-Santiago P J, and Malandraki O E. 2017. Real-time prediction of the occurrence of GLE events. *Space Weather*, 15(7):861–873.
- Papaioannou A, Anastasiadis A, Kouloumvakos A, Paassilta M, Vainio R, Valtonen E, Belov A, Eroshenko E, Abunina M, and Abunin A. 2018a. Nowcasting Solar Energetic Particle Events Using Principal Component Analysis. *Sol Phys*, 293(7):100.
- Papaioannou A, Anastasiadis A, Sandberg I, Georgoulis M K, Tsiropoula G, Tziotziou K, Jiggins P, and Hilgers A. 2015. A Novel Forecasting System for Solar Particle Events and Flares (FORSPEF). In *Journal of Physics Conference Series*, volume 632 of *Journal of Physics Conference Series*, page 012075.
- Papaioannou A, Anastasiadis A, Sandberg I, and Jiggins P. 2018b. Nowcasting of Solar Energetic Particle Events using near real-time Coronal Mass Ejection characteristics in the framework of the FORSPEF tool. *J Space Weather Space Clim*, 8:A37.

- Papaioannou A, Vainio R, Raukunen O, Jiggins P, Aran A, Dierckxsens M, Mallios S A, Paassilta M, and Anastasiadis A. 2022. The probabilistic solar particle event forecasting (PROSPER) model. *J Space Weather Space Clim*, 12:24.
- Park E, Moon Y J, Shin S, Yi K, Lim D, Lee H, and Shin G. 2018. Application of the Deep Convolutional Neural Network to the Forecast of Solar Flare Occurrence Using Full-disk Solar Magnetograms. *Astrophys J*, 869(2):91.
- Park S H, Leka K D, Kusano K, Andries J, Barnes G, Bingham S, Bloomfield D S, McCloskey A E, Delouille V, Falconer D, Gallagher P T, Georgoulis M K, Kubo Y, Lee K, Lee S, Lobzin V, Mun J, Murray S A, Hamad Nageem T A M, Qahwaji R, Sharpe M, Steenburgh R A, Steward G, and Terkildsen M. 2020. A Comparison of Flare Forecasting Methods. IV. Evaluating Consecutive-day Forecasting Patterns. *Astrophys J*, 890(2):124.
- Patsourakos S, Georgoulis M, Vourlidas A, Nindos A, Sarris T, Anagnostopoulos G, Anastasiadis A, Chintzoglou G, Daglis I, Gontikakis C, et al.. 2016. The major geoeffective solar eruptions of 2012 march 7: comprehensive sun-to-earth analysis. *Astrophys J*, 817(1):14.
- Pesnell W D, Thompson B J, and Chamberlin P. 2012a. *The solar dynamics observatory (SDO)*. Springer.
- Pesnell W D, Thompson B J, and Chamberlin P C. 2012b. The Solar Dynamics Observatory (SDO). *Sol Phys*, 275(1):3–15.
- Piana M, Campi C, Benvenuto F, Guastavino S, and M. Massone A. 2019. Flare forecasting and feature ranking using SDO/HMI data. *Il Nuovo Cimento C*, 42(1):1–4.
- Posner A. 2007. Up to 1-hour forecasting of radiation hazards from solar energetic ion events with relativistic electrons. *Space Weather*, 5(5):05001.
- Pricopi A C, Paraschiv A R, Besliu-Ionescu D, and Marginean A N. 2022. Predicting the geoeffectiveness of cmes using machine learning. *Astrophys J*, 934(2):176.
- Qahwaji R and Colak T. 2007. Automatic short-term solar flare prediction using machine learning and sunspot associations. *Sol Phys*, 241:195–211.
- Raboonik A, Safari H, Alipour N, and Wheatland M S. 2016. Prediction of solar flares using unique signatures of magnetic field images. *Astrophys J*, 834(1):11.
- Reep J W and Barnes W T. 2021. Forecasting the Remaining Duration of an Ongoing Solar Flare. *Space Weather*, 19(10).
- Ribeiro F and Gradwohl A. 2021. Machine learning techniques applied to solar flares forecasting. *Astronomy and Computing*, 35:100468.
- Ribeiro M T, Singh S, and Guestrin C. 2016. "Why Should I Trust You?": Explaining the Predictions of Any Classifier. *arXiv*.
- Richardson I G and Cane H V. 2010. Near-earth interplanetary coronal mass ejections during solar cycle 23 (1996–2009): Catalog and summary of properties. *Sol Phys*, 264:189–237.
- Richardson I G, Mays M L, and Thompson B J. 2018. Prediction of Solar Energetic Particle Event Peak Proton Intensity Using a Simple Algorithm Based on CME Speed and Direction and Observations of Associated Solar Phenomena. *Space Weather*, 16(11):1862–1881.
- Riley P, Mays M L, Andries J, Amerstorfer T, Biesecker D, Delouille V, Dumbovic M, Feng X, Henley E, Linker J A, et al.. 2018. Forecasting the arrival time of coronal mass ejections: Analysis of the ccmc cme scoreboard. *Space Weather*, 16(9):1245– 1260.
- Rong L, Jie Z, and Xin H. 2023. Progress in solar flare forecasting. [Online; accessed 6. Jul. 2023].

- Rosenblatt F. 1958. The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, 65(6):386–408.
- Rumelhart D E, Hinton G E, and Williams R J. 1986. Learning representations by back-propagating errors. *Nature*, 323(6088):533–536.
- Sammis I, Tang F, and Zirin H. 2000. The dependence of large flare occurrence on the magnetic structure of sunspots. *Astrophys J*, 540(1):583.
- Scherrer P H, Schou J, Bush R, Kosovichev A, Bogart R, Hoeksema J, Liu Y, Duvall T, Zhao J, Title A, et al.. 2012. The helioseismic and magnetic imager (hmi) investigation for the solar dynamics observatory (sdo). *Sol Phys*, 275:207–227.
- Schrijver C J. 2007. A Characteristic Magnetic Field Pattern Associated with All Major Solar Flares and Its Use in Flare Forecasting. *Astrophys J*, 655(2):L117.
- Shea M A and Smart D F. 2012. Space Weather and the Ground-Level Solar Proton Events of the 23rd Solar Cycle. *Space Sci Rev*, 171(1-4):161–188.
- Sheeley Jr N, Howard R, Koomen M, Michels D, Schwenn R, Muehlhaeuser K, and Rosenbauer H. 1985. Coronal mass ejections and interplanetary shocks. *J Geophys Res Space Phys*, 90(A1):163–175.
- Shi T, Wang Y, Wan L, Cheng X, Ding M, and Zhang J. 2015. Predicting the arrival time of coronal mass ejections with the graduated cylindrical shell and drag force model. *Astrophys J*, 806(2):271.
- Shi Y, Wang J, Chen Y, Liu S, Cui Y, and Ao X. 2022. Impacts of cmes on earth based on logistic regression and recommendation algorithm. *Space: Science & Technology*.
- Shi Y R, Chen Y H, Liu S Q, Liu Z, Wang J J, Cui Y M, Luo B, Yuan T J, Zheng F, Wang Z, et al.. 2021. Predicting the cme arrival time based on the recommendation algorithm. *Res Astron Astrophys*, 21(8):190.
- Smart D F and Shea M A. 1989. PPS-87: A new event oriented solar proton prediction model. *Adv Space Res*, 9(10):281–284.
- Socker D G, Howard R A, Korendyke C M, Simnett G M, and Webb D F. 2000. Nasa solar terrestrial relations observatory (stereo) mission heliospheric imager. In *Instrumentation for UV/EUV Astronomy and Solar Missions*, volume 4139, pages 284–293. SPIE.
- Song H, Tan C, Jing J, Wang H, Yurchyshyn V, and Abramenko V. 2009. Statistical assessment of photospheric magnetic features in imminent solar flare predictions. *Sol Phys*, 254:101–125.
- Srivastava N. 2005. A logistic regression model for predicting the occurrence of intense geomagnetic storms. *Ann Geophys*, 23(9):2969–2974.
- St. Cyr O C, Posner A, and Burkepile J T. 2017. Solar energetic particle warnings from a coronagraph. *Space Weather*, 15(1):240–257.
- Stumpo M, Benella S, Laurenza M, Alberti T, Consolini G, and Marcucci M F. 2021. Open Issues in Statistical Forecasting of Solar Proton Events: A Machine Learning Perspective. *Space Weather*, 19(10):e2021SW002794.
- Sudar D, Vršnak B, and Dumbovic M. 2015. Predicting coronal mass ejections transit times to earth with neural network. *Mon Not R Astron Soc*, 456(2):1542–1548.
- Sun D, Huang X, Zhao Z, and Xu L. 2023. Deep Learning–based Solar Flare Forecasting Model. III. Extracting Precursors from EUV Images. *Astrophys J Suppl Ser*, 266(1):8.

- Sun P, Dai W, Ding W, Feng S, Cui Y, Liang B, Dong Z, and Yang Y. 2022a. Solar Flare Forecast Using 3D Convolutional Neural Networks. *Astrophys J*, 941(1):1.
- Sun W, Xu L, Ma S, Yan Y, Liu T, and Zhang W. 2022b. A Dynamic Deep-learning Model for Generating a Magnetogram Sequence from an SDO/AIA EUV Image Sequence. *Astrophys J Suppl Ser*, 262(2):45.
- Tang R, Liao W, Chen Z, Zeng X, Wang J s, Luo B, Chen Y, Cui Y, Zhou M, Deng X, Li H, Yuan K, Hong S, and Wu Z. 2021. Solar Flare Prediction Based on the Fusion of Multiple Deep-learning Models. *Astrophys J Suppl Ser*, 257(2):50.
- Temmer M. 2021. Space weather: the solar perspective. *Living Rev Sol Phys*, 18(1):4.
- Temmer M, Rollett T, Möstl C, Veronig A M, Vršnak B, and Odstreil D. 2011. Influence of the ambient solar wind flow on the propagation behavior of interplanetary coronal mass ejections. *Astrophys J*, 743(2):101.
- Toriumi S and Wang H. 2019. Flare-productive active regions. *Living Rev Sol Phys*, 16(1):3.
- Turing A M. 1950. I.—COMPUTING MACHINERY AND INTELLIGENCE. *Mind*, LIX(236):433–460.
- Uwamahoro J, McKinnell L A, and Habarulema J B. 2012. Estimating the geoeffectiveness of halo CMEs from associated solar and IP parameters using neural networks. *Ann Geophys*, 30(6):963–972.
- Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez A N, Kaiser L, and Polosukhin I. 2017. Attention Is All You Need. *arXiv*.
- Verbeke C, Mays M, Temmer M, Bingham S, Steenburgh R, Dumbovic M, Núñez M, Jian L, Hess P, Wiegand C, et al.. 2019. Benchmarking cme arrival time and impact: Progress on metadata, metrics, and events. *Space Weather*, 17(1):6–26.
- Vourlidis A, Patsourakos S, and Savani N. 2019. Predicting the geoeffective properties of coronal mass ejections: current status, open issues and path forward. *Philos Trans Royal Soc A: Math Phys Eng Sci*, 377(2148):20180096.
- Vršnak B, Temmer M, Žic T, Taktakishvili A, Dumbovic M, Möstl C, Veronig A, Mays M, and Odstreil D. 2014. Heliospheric propagation of coronal mass ejections: comparison of numerical wsa-enlil+ cone model and analytical drag-based model. *Astrophys J Suppl Ser*, 213(2):21.
- Wan J, Fu J F, Liu J F, Shi J K, Jin C G, and Zhang H P. 2021. Class imbalance problem in short-term solar flare prediction. *Res Astron Astrophys*, 21(9):237.
- Wan J, Fu J F, Tan D M, Han K, Yu M Y, and E P. 2022. Solar Flare Forecast Model Based on Resampling and Fusion Method. *Res Astron Astrophys*, 22(8):085020.
- Wan J, Fu J F, Wen R Q, Han K, Yu M Y, and E P. 2023. Flare Forecast Model Based on DS-SMOTE and SVM with Optimized Regular Term. *Res Astron Astrophys*, 23(6):065004.
- Wang H, Cui Y, Li R, Zhang L, and Han H. 2008. Solar flare forecasting model supported with artificial neural network techniques. *Adv Space Res*, 42(9):1464–1468.
- Wang J. 2012. Solar activity studies: From a magnetohydrodynamics description to a plasma perspective. *Chinese Chin Sci Bull*, 57:1362–1368.
- Wang J, Ao X, Wang Y, Wang C, Cai Y, Luo B, Liu S, Shen C, Zhuang B, Xue X, et al.. 2018. An operational solar wind prediction system transitioning fundamental science to operations. *J Space Weather Space Clim*, 8:A39.
- Wang J, Liu S, Ao X, Zhang Y, Wang T, and Liu Y. 2019a. Parameters derived from the sdo/hmi vector magnetic field data: Potential to improve machine-learning-based solar flare prediction models. *Astrophys J*, 884(2):175.

- Wang J, Luo B, and Liu S. 2022. Precursor identification for strong flares based on anomaly detection algorithm. *Front Astron Space Sci*, 9:1037863.
- Wang J, Zhang Y, Hess Webber S A, Liu S, Meng X, and Wang T. 2020. Solar Flare Predictive Features Derived from Polarity Inversion Line Masks in Active Regions Using an Unsupervised Machine Learning Algorithm. *Astrophys J*, 892(2):140.
- Wang J L. 2009. Will the solar cycle 24 be a low one? *Chin Sci Bull*, 54(23):3664–3668.
- Wang Y, Liu J, Jiang Y, and Erdélyi R. 2019b. Cme arrival time prediction using convolutional neural network. *Astrophys J*, 881(1):15.
- Waterfall C O G, Dalla S, Raukunen O, Heynderickx D, Jiggins P, and Vainio R. 2023. High Energy Solar Particle Events and Their Relationship to Associated Flare, CME and GLE Parameters. *Space Weather*, 21(3):e2022SW003334.
- Webb D F and Howard T A. 2012. Coronal Mass Ejections: Observations. *Living Rev Sol Phys*, 9(1):3.
- Wentzel D and Tidman D. 1969. Plasma instabilities in astrophysics. [Online; accessed 30. Aug. 2023].
- Wheatland M. 2004. A bayesian approach to solar flare prediction. *Astrophys J*, 609(2):1134.
- Winter L M and Ledbetter K. 2015. Type II and Type III Radio Bursts and their Correlation with Solar Energetic Proton Events. *Astrophys J Suppl Ser*, 809(1):105.
- Wold A M, Mays M L, Taktakishvili A, Jian L K, Odstreil D, and MacNeice P. 2018. Verification of real-time wsa- enlil+ cone simulations of cme arrival-time at the ccmc from 2010 to 2016. *J Space Weather Space Clim*, 8:A17.
- Wood B E, Wu C C, Lepping R P, Nieves-Chinchilla T, Howard R A, Linton M G, and Socker D G. 2017. A stereo survey of magnetic cloud coronal mass ejections observed at earth in 2008–2012. *Astrophys J Suppl Ser*, 229(2):29.
- Woods T N, Eparvier F G, Hock R, Jones A R, Woodraska D, Judge D, Didkovsky L, Lean J, Mariska J, Warren H, McMullin D, Chamberlin P, Berthiaume G, Bailey S, Fuller-Rowell T, Sojka J, Tobiska W K, and Viereck R. 2012. Extreme Ultraviolet Variability Experiment (EVE) on the Solar Dynamics Observatory (SDO): Overview of Science Objectives, Instrument Design, Data Products, and Model Developments. *Sol Phys*, 275(1):115–143.
- Xia Z, Gao G, Wang M, and Lin J. 2015. Prediction of Solar Proton Events and the Radio Type I Noise Storms. *Chinese Astronomy and Astrophysics*, 39(1):54–65.
- Xin H. 2010. *Research on solar flare prediction method based on sequence data*. PhD thesis, Harbin: Harbin Institute of Technology.
- Yang R, Cao J, Huang W, and Nian A. 2010. Cross wavelet analysis of the relationship between total solar irradiance and sunspot number. *Chin Sci Bull*.
- Yang X, Lin G, Zhang H, and Mao X. 2013. Magnetic nonpotentiality in photospheric active regions as a predictor of solar flares. *Astrophys J Letters*, 774(2):L27.
- Yi K, Moon Y J, Shin G, and Lim D. 2020. Forecast of Major Solar X-Ray Flare Flux Profiles Using Novel Deep Learning Models. *Astrophys J*, 890(1):L5.
- Yin Z, Ma L, Han Y, and Han Y. 2007. Long-term variations of solar activity. *Chinese Chin Sci Bull*, 52(20):2737–2741.
- Yu D, Huang X, Hu Q, Zhou R, Wang H, and Cui Y. 2009a. Short-term solar flare prediction using multiresolution predictors. *Astrophys J*, 709(1):321.
- Yu D, Huang X, Wang H, and Cui Y. 2009b. Short-Term Solar Flare Prediction Using a Sequential Supervised Learning Method.

- Sol Phys*, 255(1):91–105.
- Yu D, Huang X, Wang H, Cui Y, Hu Q, and Zhou R. 2010. Short-term solar flare level prediction using a bayesian network approach. *Astrophys J*, 710(1):869.
- Yu X, Xu L, and Yan Y. 2021. Image Desaturation for SDO/AIA Using Deep Learning. *Sol Phys*, 296(3):56–14.
- Yu Z, Anh V, Eastes R, and Wang D L. 2012. Multifractal analysis of solar flare indices and their horizontal visibility graphs. *Nonlinear Process Geophys*, 19(6):657–665.
- Yuan Y, Shih F Y, Jing J, and Wang H M. 2010. Automated flare forecasting using a statistical learning technique. *Res Astron Astrophys*, 10(8):785.
- Zhang G, Wang J, and Li D. 1994. A new scheme used for the short-term prediction of x-ray flares. *Publications of the Beijing Astronomical Observatory*, 24:24–31.
- Zhang X, Xu L, Ren Z, Yu X, and Li J. 2023. Attention-Based Deep Learning Model for Image Desaturation of SDO/AIA. *Res Astron Astrophys*, 23(8):085004.
- Zhao D, Xu L, Chen L, Yan Y, and Duan L Y. 2019. Mask-Pix2Pix Network for Overexposure Region Recovery of Solar Image. *Adv Astron*, 2019.
- Zhao J, Feng X, Xiang C, and Jiang C. 2023a. A mutually embedded perception model for solar corona. *Mon Not R Astron Soc*, 523(1):1577–1590.
- Zhao X and Dryer M. 2014. Current status of CME/shock arrival time prediction. *Space Weather*, 12(7):448–469.
- Zhao X and Dryer M. 2014. Current status of cme/shock arrival time prediction. *Space Weather*, 12(7):448–469.
- Zhao Z, Xu L, Zhu X, Zhang X, Liu S, Huang X, Ren Z, and Tian Y. 2023b. A Large-Scale Dataset of Three-Dimensional Solar Magnetic Fields Extrapolated by Nonlinear Force-Free Method. *Sci Data*, 10(178):1–15.
- Zheng Y, Li X, Si Y, Qin W, and Tian H. 2021. Hybrid deep convolutional neural network with one-versus-one approach for solar flare prediction. *Mon Not R Astron Soc*, 507(3):3519–3539.
- Zheng Y, Li X, Yan S, Huang X, Lou H, and Li Z. 2023a. Multiclass solar flare forecasting models with different deep learning algorithms. *Mon Not R Astron Soc*, 521(4):5384–5399.
- Zheng Y, Qin W, Li X, Ling Y, Huang X, Li X, Yan P, Yan S, and Lou H. 2023b. Comparative analysis of machine learning models for solar flare prediction. *Astrophys Space Sci*, 368(7):53–12.
- Zhong Q, Wang J, Meng X, Liu S, and Gong J. 2019. Prediction Model for Solar Energetic Proton Events: Analysis and Verification. *Space Weather*, 17(5):709–726.
- Zhu C L and Wang J L. 2003. Verification of short-term predictions of solar soft x-ray bursts for the maximum phase (2000–2001) of solar cycle 23. *Chinese Journal of Astronomy and Astrophysics*, 3(6):563.
- Zhuang B, Wang Y, Shen C, Liu S, Wang J, Pan Z, Li H, and Liu R. 2017. The significance of the influence of the cme deflection in interplanetary space on the cme arrival at earth. *Astrophys J*, 845(2):117.
- Zucca P, Núñez M, and Klein K. 2017. Exploring the potential of microwave diagnostics in SEP forecasting: The occurrence of SEP events. *J Space Weather Space Clim*, 7:A13.