Research in Astronomy and Astrophysics, 25:035025 (30pp), 2025 March

https://doi.org/10.1088/1674-4527/adbd9f
CSTR: 32081.14.RAA.adbd9f

Research Progress in Solar Flare Prediction Methods

Ke Han[1], Zhen Liu[1], Xian-Yi Zhao[1], Yi-Fei Li[2], De-Quan Zheng[1], and Jie Wan[3]
[1] School of Computer and Information Engineering, Harbin University of Commerce, Harbin 150028, China
[2] Faculty of Computing, Harbin Institute of Technology, Harbin 150001, China
[3] School of Energy Science and Engineering, Harbin Institute of Technology, Harbin 150001, China; wanjie@hit.edu.cn
Received 2025 January 3; revised 2025 February 23; accepted 2025 February 28; published 2025 March 31

Abstract

Solar flares are one of the strongest outbursts of solar activity, posing a serious threat to Earth's critical infrastructure, such as communications, navigation, power, and aviation. Therefore, it is essential to accurately predict solar flares in order to ensure the safety of human activities. Currently, the research focuses on two directions: first, identifying predictors with more physical information and higher prediction accuracy, and second, building flare prediction models that can effectively handle complex observational data. In terms of flare observability and predictability, this paper analyses multiple dimensions of solar flare observability and evaluates the potential of observational parameters in prediction. In flare prediction models, the paper focuses on data-driven models and physical models, with an emphasis on the advantages of deep learning techniques in dealing with complex and high-dimensional data. By reviewing existing traditional machine learning, deep learning, and fusion methods, the key roles of these techniques in improving prediction accuracy and efficiency are revealed. Regarding prevailing challenges, this study discusses the main challenges currently faced in solar flare prediction, such as the complexity of flare samples, the multimodality of observational data, and the interpretability of models. The conclusion summarizes these findings and proposes future research directions and potential technology advancement.

Key words: Sun: activity – Sun: flares – (Sun:) sunspots – Sun: magnetic fields – magnetohydrodynamics (MHD)

1. Introduction

In recent years, numerous studies have explored the relationship between physics and artificial intelligence (AI), offering new directions and methodologies for solar flare prediction (Jiao et al. 2024). Solar flares are violent energy-releasing phenomena of solar activity, usually occurring in active regions of the solar atmosphere and manifesting themselves as rapid reorganization of local magnetic fields and magnetic reconnection processes. This process rapidly converts the magnetic field's free energy into thermal energy, radiant energy and high-energy particle flux, leading to the eruption of a solar flare. Although the specific physical processes within flares have not been fully elucidated, magnetic reconnection and an increase in the magnetic shear angle in the coronal layer are widely believed to be necessary conditions for driving flare eruptions (Reames 2013; Chen et al. 2020; Huang et al. 2024).

A solar flare eruption is accompanied by the release of a wide range of electromagnetic radiation, from radio waves to X-rays and gamma-rays, while heating the plasma in the solar atmosphere to millions of degrees Celsius and ejecting high-speed charged particle flux. Intense flares can significantly affect the Earth's space environment, especially ionospheric disturbances caused by flares, which can interfere with radio communications and GPS navigation, affecting the normal operation of satellites and communication systems. Therefore, accurately predicting solar flares has become a crucial issue for mitigating space weather effects and ensuring the safety of satellite operations, space missions, and ground-based infrastructure (Koskinen et al. 1999; Siscoe 2000; Hayes et al. 2019; Curto 2020).

Currently, widely used solar flare observation methods include optical, magnetic field, X-ray, ultraviolet (UV), and extreme ultraviolet (EUV) observations. The observational data obtained from these methods provide a foundation for validating physical models and are essential for training and testing data-driven models. The most widely used sources of solar observational data currently include satellite platforms such as the Solar and Heliospheric Observatory (SOHO; Domingo et al. 1995), Solar Dynamics Observatory (SDO; Pesnell et al. 2012), and the Geostationary Operational Environmental Satellite (GOES), which provide a variety of predictors such as morphology, magnetic field strength and distribution, as well as the evolutionary processes of solar active regions. These predictors provide critical information for flare prediction models, but their predictability in different prediction tasks still requires further validation. Therefore, how to effectively integrate observational data and combine them with appropriate models to optimize model training and

predictive performance is one of the core challenges of current research.

In recent years, data-driven methods have made significant progress in the field of solar flare prediction, mainly including statistical methods, machine learning methods, deep learning methods, and multi-model fusion methods. Each method has its own advantages, and the choice of model is typically based on the data characteristics and the specific prediction task. For example, statistical methods usually rely on traditional linear regression and classification techniques, which are suitable for simple pattern recognition tasks. In contrast, deep learning and machine learning techniques can handle higher-dimensional and more complicated data, and they are especially well-suited to identifying the underlying nonlinear patterns of flare occurrence. As AI technology advances rapidly, deep learning techniques have shown great promise across various fields, particularly in physics research. Their applications in complex system modeling and data analysis are becoming increasingly prevalent (Huang et al. 2024; Jiao et al. 2024).

In the study of physical models of solar flares, the theory of Self-Organized Criticality (SOC; Bak et al. 1987) has been widely applied in astrophysics (Aschwanden et al. 2016). As a result, most studies have explored the use of SOC theory to develop flare prediction models. The models effectively describe the statistical behavior of solar flares, particularly their power-law distribution. Magnetohydrodynamics (MHD) instability and Double Arc Instability (DAI) are considered key physical mechanisms driving solar flare eruptions. Progress has been made in proposing prediction models based on these mechanisms. However, prediction methods based on MHD and DAI still face significant challenges because the physical mechanisms of solar flares have not been fully elucidated.

Despite the progress made by existing solar flare prediction methods, they still face numerous challenges. First, as the number of features increases, the issue of feature redundancy becomes more pronounced. Such redundancy can degrade the model's predictive accuracy and computational efficiency, leading to suboptimal generalization performance. Second, because solar flare events are rare, there is a severe class imbalance in the flare samples, which can affect the model's prediction performance. Furthermore, multimodal data fusion has steadily emerged as a research focus, and how to effectively address the complementarity of different types of data has become an urgent challenge. Finally, although deep learning methods perform well in handling complex data, they lack sufficient interpretability, and reliance on a single model may reduce the reliability of prediction results. Thus, one of the main concerns in current solar flare prediction research is how to make models more interpretable and reliable while maintaining their effectiveness.

In summary, this paper will discuss the following three aspects: first, analyze the observability, observational data, and their predictability, focusing on the predictive ability of key variables such as morphological parameters, magnetic field parameters, and historical parameters; second, explore the existing solar flare prediction methods, including those based on physical models and data-driven approaches. This paper will focus on analyzing the evolution from traditional machine learning to deep learning models, as well as the application of fusion models in prediction. Third, summarize the main challenges faced in current research, including issues such as sample complexity, multimodal data fusion, and model interpretability. Finally, this paper will give an outlook on future research directions and propose solutions to address the aforementioned challenges.

2. Analysis of the Observability and Predictability of Solar Flare Events

2.1. Observability and Data Sources

Optical observations were one of the earliest methods used to study solar activity, and basic features of solar activity were obtained mainly from images in the visible light spectrum. Sunspots, as important precursors of flares, provide valuable clues for flare prediction through their distribution, size, shape, and variations in activity. In the earliest studies of solar flares, scientists mainly relied on the naked eye and simple optical telescopes to observe the Sun. With advances in telescope technology, later optical observations were able to capture more details, including the evolution of sunspots and more precise solar surface activity. The data that are currently widely used include Hα (Hydrogen-α) images (as shown in panel (a) of Figure 1). These images directly observe the structural changes before and after a flare eruption through the dynamic activity of the solar chromosphere (Chen et al. 2024). The Debrecen Heliophysical Observatory focused on the long-term monitoring of sunspots and compiled a detailed database containing information on the types, numbers, and distribution of sunspots, which provides a solid foundation for the study of solar activity cycles, the evolution of sunspot regions and flare prediction (Baranyi et al. 2016).

One of the main causes of solar flare eruptions is the dynamic changes in the magnetic field, so real-time monitoring of the magnetic field is essential. By observing the magnetic field of the solar active region, it is possible to reveal its non-potential structure and the magnetic reconnection process, which may trigger flares. The European Space Agency and the National Aeronautics and Space Administration (NASA) of the United States of America collaborated to create the solar observation satellite, known as SOHO, which aims to provide high-quality solar observation data. Since its launch in 1996, the Michelson Doppler Imager (MDI; Scherrer et al. 1995) instrument aboard SOHO provides important observational data on the strength and distribution of the photospheric magnetic field, such as line-of-sight (LoS) magnetograms. The evolutionary dynamics of the solar magnetic field revealed by

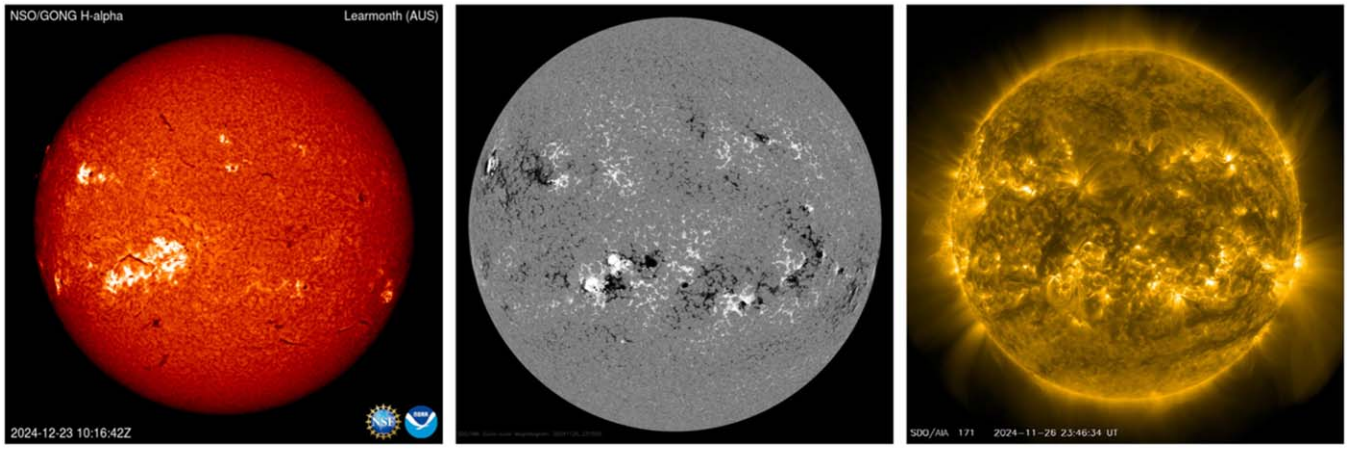

Figure 1. Panel (a) shows the Hα image provided by NSO, panel (b) shows the magnetogram provided by HMI, and panel (c) shows the extreme ultraviolet image in the 171 Å provided by AIA.

these data are crucial for flare prediction and advance our understanding of the cyclic variations of solar activity. The widely used SOHO Magnetic Activity Research Team (Bobra et al. 2021) data set provides more comprehensive magnetic field information, covering data related to solar active regions and the corona. In contrast, SDO, launched in 2010, is equipped with more advanced observational instruments, significantly improving the accuracy of solar activity monitoring. The Helioseismic and Magnetic Imager (HMI; Schou et al. 2012) aboard SDO provides high-precision solar magnetic field data (as shown in panel (b) of Figure 1), which has become a crucial source for flare prediction. Meanwhile, the Space-weather HMI Active Region Patches (SHARP; Bobra et al. 2014) products from SDO provide higher spatial resolution solar active region data, providing more detailed information to reveal the driving mechanism of flares. In addition, some studies have improved the accuracy of flare prediction by combining SOHO and SDO data, maximizing the benefits of both, in terms of spatial and temporal resolution (Sun et al. 2022).

X-rays serve as a key diagnostic of non-thermal processes in solar flares, tracing electron acceleration and magnetic energy release. X-ray observations provide direct radiation intensity data for flare research, especially in the minutes after a flare occurs, when the intensity of X-rays can increase significantly, which is a sign of the flare eruption. The high-precision X-ray sensors carried by the GOES satellite can monitor key parameters such as X-ray radiation, proton flux, and electron flux associated with solar flares. These data are crucial for training predictive models. The GOES satellite's real-time recordings of flare timing, duration, and intensity provide temporally resolved observational data, which can be paired with expert-defined labels (e.g., flare magnitude classes) to train machine learning models. This approach enhances the detection of precursory patterns and the reliability of flare prediction. In addition, the Reuven Ramaty High-Energy Solar Spectroscopic Imager satellite offers precise data on the flare source regions and their radiation energy distribution through high-energy X-ray and gamma-ray imaging, providing rich information to understand the flare mechanisms.

The main radiant energy released by flares is concentrated in the UV and EUV wavelengths. Therefore, UV and EUV observations have important methods for studying solar flares. The corona, chromosphere, and other active regions of the solar atmosphere can be precisely shown by observations at the EUV wavelength. The most widely used instrument is the Atmospheric Imaging Assembly (AIA; Lemen et al. 2012) aboard SDO, which offers high-resolution images across multiple wavelengths (as shown in panel (c) of Figure 1) and can monitor the occurrence and evolution of solar flares in real time. The AIA can capture activity in the active regions of the Sun in various wavelengths (e.g., 171 Å, 304 Å), thus revealing the evolution of the flare and its precursors. Similarly, another instrument on SDO, the Extreme Ultraviolet Variability Experiment, is dedicated to measuring the variations in solar EUV radiation. It provides precise data on solar radiation fluctuations, which are employed to more thoroughly examine how solar flares and other solar activity affect Earth's space weather. In addition, the Extreme Ultraviolet Imaging Telescope on SOHO is another important ultraviolet imaging instrument. It provides detailed observations of the Sun before and after solar flare events, helping researchers analyze the active regions, eruption intensity, and radiation characteristics.

In addition, radio observations offer a unique advantage in the study of solar flares, revealing the process of high-energy particle acceleration and the evolution of radiation sources

during flare eruptions. A more thorough examination of the dynamical processes of the flare is made possible by monitoring radio waves, which provide important information about the energy distribution and propagation patterns of the accelerated particle population. Meanwhile, the physical condition of the high-temperature gas in the flare region can be determined by spectral studies, involving the analysis of variations in spectral lines to provide important parameters such as the temperature, density, and velocity of the flare. These data are crucial for studying the temperature gradient, elemental abundances, and gas motion mechanisms of the flare, providing key clues to understanding its physical mechanism. The Interface Region Imaging Spectrograph satellite utilizes high-resolution spectrum imaging technologies to explore the transition region and the solar chromosphere. It can accurately capture the temperature distribution, variations in elemental abundances, and gas velocity in the flare region, thus revealing the interaction between the flare, the chromosphere, and the transition region. These observational data are valuable for refining flare warning models and improving the prediction of the time and intensity of flare eruptions.

With the advancement of observation technologies, the observability of solar flares has significantly improved. By using various observational methods such as optical, ultraviolet, X-ray, magnetic field, and radio observations, researchers are able to study the flare occurrence mechanism more comprehensively and its interaction with space weather on Earth. With the emergence of new detectors, the capability of observing solar flares will be further enhanced, which will provide more accurate data for early warning of solar activity and space weather prediction. However, how to efficiently integrate data from different observation platforms and wavelengths, reasonably balance the weights of various data types, and successfully overcome the projection effects are still important problems that require further study.

2.2. Analysis of Observational Parameters and Their Predictability

2.2.1. Morphological Parameters

Morphological parameters play an important role in solar flare prediction by describing the geometry and complexity of the solar active region and its internal structure. These parameters, which include the sunspot number, sunspot area, classification of sunspot groups, can provide key information about solar magnetic activity and flare productivity, quantifying the active region's capacity to generate flare events.

The intensity of magnetic activity on the solar surface is directly reflected in the sunspot number, which is often closely associated with the peak of the solar activity cycle. The likelihood of flare eruptions increases as the sunspot number rises, indicating an increase in solar activity. Therefore, the sunspot number has been extensively employed in flare prediction (Hathaway 2015). The sunspot area represents the scale of the active region and the magnetic field strength. Larger sunspot regions are usually accompanied by a stronger magnetic field, which is prone to produce strong flares. Cinto et al. (2020a) extracted features (sunspot number and sunspot area) from NOAA/SWPC data and combined them with machine learning methods for flare prediction. They found that the combination of sunspot number and area provides more comprehensive solar activity information.

The classification of sunspot groups is another important method for flare prediction. By analyzing the morphological distribution, magnetic polarity, and other characteristics of sunspot groups, the potential complexity of solar active regions can be revealed. Mount Wilson classification (Hale et al. 1919) is among the earliest proposed classification methods, which classifies sunspot groups into α-type (unipolar), β-type (bipolar), γ-type (multi-polar), and β-γ-type (complex) based on the symmetry of its magnetic polarity. If the sunspot group contains the δ-types (where opposite polarity umbrae share the same penumbra), it can be further subdivided into β-δ, γ-δ, or β-γ-δ types. This method focuses on the magnetic field symmetry of sunspot groups, and serves as a preliminary classification of sunspot group complexity. Zürich's classification (Cortie 1901) refines the Mount Wilson classification method by introducing nine complexity categories based on the morphological and evolutionary features of sunspot groups. This method can more accurately describe the evolution of the sunspot groups. McIntosh classification (McIntosh 1990) further refines the Zürich method by using three components to represent sunspot group classification and introducing criteria such as the shape of the maximum sunspot penumbra. This makes the classification more refined, allowing for a more accurate correlation between sunspot group complexity and flare potential. Eren et al. (2016) showed that 79% of flares originate from large and complex sunspot groups, and Oloketuyi et al. (2023) also found that more complex sunspot groups such as β-γ-δ types have higher flare production potential, emphasizing the importance of sunspot group complexity in flare prediction. According to another study (Lee et al. 2012), the likelihood of flare eruptions increases dramatically for large and dense sunspot groups as sunspot area grows, based on sunspot classification and its variations.

In summary, morphological parameters like sunspot number, area, and group classification are significant markers of solar activity that can effectively assist prediction models to identify the characteristics of solar magnetic activity and enhance prediction accuracy. Future research should combine these morphological parameters with other observational parameters to further improve the effectiveness of flare prediction.

2.2.2. Magnetic Field Parameters

Although morphological parameters provide a description of the active region to some extent, they are essentially only proxies for the photospheric magnetic field and are unable to adequately reflect the true characteristics of the active region's magnetic field, which is fundamentally associated with solar flare occurrence (Sammis et al. 2000). With advancements in observational technology, an increasing number of high-quality photospheric magnetograms have been obtained, prompting researchers to propose more magnetically based parameters to explore the impact of the magnetic properties on solar flare occurrence.

(1) *Parameters Based on LoS Magnetograms*. Several studies (Sadykov & Kosovichev 2017; Guennou et al. 2017) have recently validated the effectiveness of the primary polarity inversion line (PIL) in distinguishing flare events and non-flare events using different methods. Using LoS magnetograms, Schrijver (2007) developed a quantitative estimate of the total unsigned flux (R-value) to describe the high gradient and strong magnetic field characteristics of PIL, providing an effective magnetic field topological parameter for flare prediction. With the advancement of observational techniques, and because the resolution of the SDO/HMI magnetograms differs from that of the SOHO/MDI, Cicogna et al. (2021) suggested calibrating the R-value computation. They also introduced a topology-based D index for estimating the magnetic field complexity by calculating the number of PILs.

In PIL-related studies, the gradient-weighted inversion-line length parameter, which combines the length and gradient of the PIL, was proposed by Mason & Hoeksema (2010) and can demonstrate a notable shift prior to a flare. Welsch et al. (2009) found a strong correlation between R-values and the proxy Poynting flux and flare fluxes by analyzing SOHO/MDI LoS data. Song et al. (2009) employed the ordinal logistic regression technique for flare prediction and suggested the total magnetic dissipation ($E_{\rm diss}$) parameter based on LoS magnetograms to describe the active region's global non-potentiality. To more accurately characterize the longitudinal magnetic field pressure in the active regions, Huang & Wang (2013) extracted three parameters from the SOHO/MDI longitudinal magnetic field: the number of pixels with positive P values in the active region ($P_{\rm num}$), the summation of positive P values ($P_{\rm sum}$), and the maximum value of P within the active region ($P_{\rm max}$). These values offer crucial details for a deeper understanding of the magnetic field dynamics and the process underlying flare eruptions. McAteer et al. (2005) discovered a strong correlation between an active region's fractal dimension and its capacity to produce flares, and proposed the lower thresholds of fractal dimension for M-class and X-class flares (1.2 and 1.25, respectively). However, subsequent studies (Georgoulis 2005; Conlon et al. 2009; Giorgi et al. 2015) have pointed out that it is difficult to effectively distinguish between active and inactive regions using the fractal dimension alone, and necessitating its combination with other prediction tools to improve the prediction accuracy.

To explore the impact of different magnetic field representations on the topological parameters of the magnetic field, Guerra et al. (2018) compared the line-of-sight ($B_{\rm los}$) with the spherical-radial (B_r) magnetogram and found that the B_r magnetic field can more accurately reflect the actual situation when estimating the magnetic properties of the active region away from the solar center. Deshmukh et al. (2021) combined the B_r magnetograms from the SDO/HMI with computational geometry and topological techniques to propose new features such as polarity proximity, interaction, and topological characteristics on persistence maps. Based on this, Sun et al. (2021) extended the PIL mask magnetogram analysis method by combining the SHARP parameter and constructed the topological data analysis features and spatial statistical features, thereby significantly improving flare prediction. Abramenko et al. (2002) studied the LoS magnetograms of different active regions and found that the B_z field exhibits multifractal characteristics, with its dissipation spectrum significantly correlated with the flare occurrence rate. Studies have shown that during a solar flares, the dissipation spectrum of B_z tends to flatten.

In summary, magnetic field topological parameters extracted from LoS magnetograms, such as PIL features, R-value, and GWIL, can effectively quantify the magnetic field complexity and identify potential active regions. With the continuous development of observational techniques, the accuracy and reliability of these magnetic field parameters have significantly improved, thereby providing more precise tools and methods for flare prediction. Future research could further optimize the combination of these topological parameters to enhance the performance of flare prediction models.

(2) *Parameters Based on Vector Magnetograms*. The use of magnetic field characteristics in solar flare prediction has advanced significantly in recent years, particularly in the analysis of vector magnetic field data. Leka & Barnes (2003a, 2003b, 2007) made pioneering contributions by extracting several physical parameters from magnetograms (such as total magnetic flux) from vector magnetograms of the photospheric layers as potential flare predictors. They concluded that total photospheric excess magnetic energy was found to be the most effective single variable for predicting large-scale flares. Falconer (2001) derived metrics related to magnetic free energy using magnetograms: the length of PILs with strong shear, vertical current, total unsigned magnetic flux, and current helicity. They demonstrated a strong correlation between these metrics and solar flares. These discoveries laid the foundation for subsequent magnetic field parameters for solar flare prediction. Subsequently, Georgoulis & LaBonte (2007), Georgoulis et al. (2012) calculated the magnetic energy and helicity budgets in the active region solar corona using the

linear force-free field approximation and nonlinear force-free field extrapolation, demonstrating that the enhancement of the magnetic field complexity in the active region and the accumulation of flare energy are typically closely associated with increases in magnetic energy and helicity.

Komm & Hill (2009) further noted that the magnetic flux and vorticity values in regions of X-class flare activity were higher, while regions of M-class and C-class flares exhibited higher vorticity and flux, respectively. This suggests that the introduction of the vorticity parameter can effectively distinguish flare-active regions from non-flare regions. Park et al. (2018b) used the Differential Affine Velocity Estimator for Vector Magnetograms (DAVE4VM) method to analyze the photospheric shear-flow, and proposed the parameters such as mean shear-flow speed (S), maximum shear-flow speed ($S_{\max}$), and the integral of shear-flow speed ($S_{\rm sum}$). Through regression analysis, they revealed a correlation between these shear-flow parameters and the flare occurrence rate. Barnes & Leka (2006), based on the Magnetic Charge Topology model, analyzed vector magnetic field data. They found that parameters describing the coronal field topology and its dynamics were more effective in distinguishing between active regions and quiet regions compared with those based on the photospheric magnetic field alone. This further suggests that coronal field topology information has important potential in flare prediction.

The more complex the magnetic field, the more frequent the evolution of solar flares (Tian 2022). To quantify the complexity of magnetic fields more comprehensively, Georgoulis & Rust (2007) proposed the effective connected magnetic field ($B_{\rm eff}$), which takes into account both the distribution and connectivity of the magnetic field and can effectively reflect the characteristics of the magnetic field in active regions and the potential flare activity. The FORSPEF tool (Anastasiadis et al. 2017) uses $B_{\rm eff}$ as an indicator of the combined strength and connectivity of the magnetic field and determines the flare probability using a sigmoid curve fitting method, as shown in Equation (1)

$$P_{\rm class}(B_{\rm eff_{(norm)}}) = A_2 + \frac{A_1 - A_2}{1 + \exp\left(\frac{\log B_{\rm eff_{(norm)}} - X_0}{W}\right)}. \quad (1)$$

In addition, Cui et al. (2006, 2007) conducted studies on vector magnetograms from the Huairou Solar Observing Station. They found that the gradient length (L_g), the strong shear length (L_s) and the strong gradient-shear length (L_{gs}) show a significant sigmoidal relationship with the flare incidence, provide a new approach for further quantifying magnetic field complexity. Building upon the research on the PIL. Wang et al. (2020) advanced the study of PIL using vector magnetic field data. By combining SHARP parameters with high-gradient PIL masks, they significantly improved flare prediction performance and validated the importance of PIL in distinguishing eruptive active regions. In addition, they also used the Kernel Principal Component Analysis method (Wang et al. 2019) to extract features from the PIL mask, and found that these features performed better than the R value in distinguishing strong flares from non-strong flares.

Current research indicates that various magnetic field parameters extracted from vector magnetograms, such as total magnetic flux, magnetic energy, shear flow velocity, vorticity, and magnetic field topology, are crucial for solar flare prediction. These parameters are primarily included in the SHARP products provided by SDO/HMI, which offer magnetic field data from solar active regions. These data are strong enough to enable magnetic field vector inversion and resolve azimuthal ambiguities (Georgoulis et al. 2024). Machine learning models trained on SHARP data commonly utilize 20 key parameters (Table 1), with 13 of these prioritized by Bobra & Couvidat (2015) through significance testing. Their study demonstrated that using only four parameters, such as unsigned current helicity, photospheric magnetic free energy density, total Lorentz force, and total unsigned vertical current, it was possible to achieve True Skill Statistic (TSS) values comparable to those obtained using all 13 parameters. Campi et al. (2019) applied the Least Absolute Shrinkage and Selection Operator (LASSO) in combination with Random Forest (RF) to validate the importance of unsigned vertical currents, current helicity, and total Lorentz force. Furthermore, current helicity is a key variable influencing flare generation in active regions (Zhang et al. 2022). Aegerter et al. (2020) identified a significant predictive trend between mean vertical current density and B-class, C-class, and M-class flares by analyzing photospheric vector magnetic field data from the SHARP data set. The correlation between these magnetic field parameters plays a crucial role in prediction, and will be further discussed in Section 4.1.1.

2.2.3. Evolutionary and Historical Parameters

In addition to the aforementioned parameters, the dynamic evolution of solar active regions offers critical insights into flare initiation. Time-series and historical data allow analysis of long-term trends and short-term precursors, capturing temporal patterns to improve model accuracy in identifying flare events.

Traditional solar flare prediction models rely on static magnetic field parameters (e.g., shear angle, total magnetic flux). However, recent studies suggest that temporal changes in magnetic field parameters may provide more direct information about flare triggering mechanisms. In their seminal studies, Leka & Barnes (2003a, 2003b, 2007), Barnes & Leka (2006) pioneered the concept of "evolutionary parameters" as a cornerstone of flare prediction through systematic analysis of photospheric vector magnetic field dynamics. By quantifying both instantaneous states and time-dependent variations of

Table 1
20 Important SHARP Parameters and Their Formulas

Keyword	Description	Formula
TOTPOT	Total photospheric magnetic free energy density	$\rho_{\rm tot} \propto \sum(\boldsymbol{B}^{\rm Obs} - \boldsymbol{B}^{\rm Pot})^2 dA$
TOTUSJH	Total unsigned current helicity	$H_{c_{\rm total}} \propto \sum\lvert B_z \cdot J_z\rvert$
USFLUX	Total unsigned flux	$\Phi = \sum\lvert B_z\rvert dA$
ABSNJZH	Absolute value of the net current helicity	$H_{c_{\rm abs}} \propto \lvert\sum B_z \cdot J_z\rvert$
MEANPOT	Mean photospheric magnetic free energy	$\overline{\rho} \propto \frac{1}{N}\sum(\boldsymbol{B}^{\rm Obs} - \boldsymbol{B}^{\rm Pot})^2$
SAVNCPP	Sum of the modulus of the net current per polarity	$J_{z_{\rm sum}} \propto \lvert\sum^{B_z^+} J_z dA\rvert + \lvert\sum^{B_z^-} J_z dA\rvert$
SHRGT45	Fraction of Area with shear >45°	Area with shear >45°/totalarea
TOTUSJZ	Total unsigned vertical current	$J_{z_{\rm total}} = \sum\lvert J_z\rvert dA$
R_VALUE	Sum of flux near polarity inversion line	$\Phi = \sum\lvert B_{LoS}\rvert dA$ within R mask
AREA_ACR	Area of strong field pixels in the active region	Area = ∑Pixels
TOTFZ	Sum of z-component of Lorentz force	$F_z \propto \sum(B_x^2 + B_y^2 - B_z^2)dA$
TOTBSQ	Total magnitude of Lorentz force	$F \propto \sum B^2$
EPSZ	Sum of z-component of normalized Lorentz force	$\delta F_z \propto \frac{\sum(B_x^2 + B_y^2 - B_z^2)}{\sum B^2}$
MEANALP	Mean characteristic twist parameter, α	$\alpha_{\rm total} \propto \frac{\sum J_z \cdot B_z}{\sum B_z^2}$
MEANGAM	Mean angle of field from radial	$\overline{\gamma} = \frac{1}{N}\sum \arctan\left(\frac{B_h}{B_z}\right)$
MEANGBH	Mean gradient of horizontal field	$\overline{\lvert\nabla\ B_h\rvert} = \frac{1}{N}\sum\sqrt{\left(\frac{\partial B_h}{\partial x}\right)^2 + \left(\frac{\partial B_h}{\partial y}\right)^2}$
MEANGBT	Mean gradient of total field	$\overline{\lvert\nabla\ B_{\rm tot}\rvert} = \frac{1}{N}\sum\sqrt{\left(\frac{\partial B}{\partial x}\right)^2 + \left(\frac{\partial B}{\partial y}\right)^2}$
MEANGBZ	Mean gradient of vertical field	$\overline{\lvert\nabla\ B_z\rvert} = \frac{1}{N}\sum\sqrt{\left(\frac{\partial B_z}{\partial x}\right)^2 + \left(\frac{\partial B_z}{\partial y}\right)^2}$
MEANJZD	Mean vertical current density	$\overline{J_z} \propto \frac{1}{N}\sum\left(\frac{\partial B_y}{\partial x} - \frac{\partial B_x}{\partial y}\right)$
MEANJZH	Mean current helicity (B_z contribution)	$\overline{H_c} \propto \frac{1}{N}\sum B_z \cdot J_z$

Note. By analyzing multiple studies, we identified the 20 most widely used SHARP parameters, with the first 13 being the key parameters proposed by Bobra & Couvidat (2015).

magnetic quantities, their work established a foundation for incorporating magnetic field evolution into flare prediction. Korsós et al. (2014, 2015) identified diagnostic parameters (e.g., magnetic flux gradient, weighted horizontal magnetic gradient) whose pre-flare temporal evolution exhibits distinct signatures, particularly near the PIL. Notably, Korsós et al. (2020) demonstrated that periodic magnetic helicity behavior in flare-productive regions serves as a key precursor. In addition, Al-Ghraibah et al. (2015) combined flux evolution features with other snapshots of spacetime features to offer a dynamic perspective on the development of the magnetic field, thereby improving the accuracy of flare predictions. To operationalize evolutionary features, studies employ diverse analytical techniques. Early approaches used sliding windows with traditional machine learning models, including decision trees, Multilayer Perceptron (MLP), and Learning Vector Quantization to extract time-series patterns (Yu et al. 2009; Huang et al. 2010; Li & Zhu 2013; Cinto et al. 2020a). With technological advances, Long Short-Term Memory (LSTM) networks have emerged as superior tools for modeling long-range temporal dependencies in magnetic field data (Liu et al. 2019; Chen et al. 2019), as detailed in Section 3.1.3.

Historical flare activity further refines predictions. Nishizuka et al. (2017) identified past flare records as a key predictor. Falconer et al. (2014) compared prediction methods combining historical flare activity with free energy proxies to those based solely on free energy and found that historical flare activity significantly improved the prediction accuracy of ⩾M-class flares. Similarly, McCloskey et al. (2018) quantified the evolution of the sunspot group using McIntosh classification-derived Poisson probabilities, demonstrating that dynamic behavioral metrics outperform static snapshots.

In conclusion, time-series analysis and historical data synergistically reveal the evolutionary drivers of solar flares. Future research should focus on optimizing temporal feature extraction and addressing challenges such as data sparsity in rare X-class events.

2.2.4. Other Parameters

With the widespread use of high-quality EUV and UV data, an increasing number of researchers (Dissauer et al. 2023; Leka et al. 2023) are seeking new parameters and quantification methods to improve the accuracy of solar flare predictions.

EUV radiation enhancement reflects the activity of high-temperature plasma and can reveal precursors to flare events. Lee et al. (2024) used the solar EUV data for limb flare prediction and discovered a significant positive correlation. This study suggests that EUV data may be superior to traditional magnetograms and white light data in predicting limb flares. Nishizuka et al. (2017) extracted features such as brightening area and intensity from UV and soft X-ray data, providing information on potential flare triggering conditions and dynamics, thus improving the prediction accuracy. Gontikakis et al. (2020) proposed that the evolution of Gaussian parameters over time (EM, $T_{\max}$, σ) could serve as an effective indicator for flare prediction by analyzing the Differential Emission Measure parameters in EUV and UV data. He found that the prediction probability is better than that of unsigned magnetic flux. Sun et al. (2023) used EUV images for flare prediction and found that EUV data at the 94 Å wavelength yielded the best results in single-wavelength prediction. Aschwanden & Aschwanden (2008a, 2008b) analyzed the EUV imaging data and proposed that the fractal dimensions of flares (the relationship between volume fractal dimensions and area fractal dimensions) could improve the prediction of the flare intensity. Massa & Emslie (2022) used sparse spatial Fourier components to represent AIA images, a method that compresses the data while preserving source feature information and is capable of identifying features in the spatial mapping that are otherwise difficult to detect directly. This suggests that the visibility map derived from AIA images could be directly used for flare prediction.

Hao et al. (2024) used solar spectral data in the range of 6400–6700 Å for flare prediction using multiple supervised learning methods. Panos & Kleint (2020) extracted ten features related to flares by parameterizing spectral data from different solar regions (e.g., quiet Sun, sunspots, active regions). The results showed that triplet emission, line core intensity, and line width are the key predictors. Asaly et al. (2020) used Total Electron Content data, along with its spatial variations and the Support Vector Machine (SVM) algorithm, for flare prediction, achieving a high Heidke Skill Score (HSS).

Radio radiation flux also plays an important role in flare prediction. Tlatov et al. (2020) used sunspot observation data and radio flux to predict solar flares. The results showed that the background X-ray flux, 10.7 cm radio flux, and solar flare index were key parameters for predicting intense solar flares. Nagem et al. (2018) proposed a comprehensive analytical prediction system using GOES X-ray flux data combined with the deep convolutional neural network (DCNN), demonstrating the innovative and practical application value of this data in solar flare prediction.

In order to make full use of the positional information of the active area, Huang et al. (2013) proposed an indicator DARAL (distance between active area and predicted active longitude) to describe the positional relationship between the active area and the predicted active longitude. Combining it with magnetic field parameters significantly improves the prediction.

Besides the aforementioned data, other parameters extracted from magnetograms using various tools have also provided new insights for solar flare prediction. For example, Muranushi et al. (2015) used the wavelet transform technique to extract wavelet power from full-disk LoS magnetograms to calculate the probability of flare eruptions. Raboonik et al. (2016) also used Zernike moments extracted from vector magnetograms for flare prediction using an SVM algorithm. Zernike moments have unique and invariant properties, allowing the predictive model to avoid relying on a few global parameters and instead provide more feature-rich information. Based on this research, Alipour et al. (2019) turned to LoS, UV, and EUV images to further extend the prediction horizon.

In summary, radiation data such as EUV, UV, and X-rays, along with feature parameters extracted from magnetograms, have become important tools for improving the accuracy of solar flare prediction. Through improved feature fusion and modeling methodologies, future studies can further investigate how different data sources complement each other and improve the prediction models' accuracy and reliability.

2.3. Section Summary

This section analyzes the observability of solar flares and finds that the most widely used observation methods are magnetic field, optical, and X-ray observations. The data obtained from these methods can extract various parameters that reflect flare information, such as morphological, magnetic field, and historical and evolutionary parameters. These parameters enhance the core data for both data-driven models and physical models, but since they provide different types of information, their predictability also varies. Therefore, selecting the appropriate parameters is crucial for flare prediction. By combining the analysis of these parameters, the accuracy of flare prediction can be improved and false alarm rates are reduced. This will provide a more precise basis for space weather prediction.

3. Solar Flare Prediction Methods

3.1. Data-driven Prediction Methods

Current mainstream flare prediction techniques rely on data-driven models because the physical mechanisms underlying solar flares are still poorly understood. Although several reviews have been published on this topic (Han et al. 2023; Huang et al. 2024), there is still a lack of research that systematically analyzes the changes in prediction methods from the perspective of the evolution of data properties. In the early stages, when data were scarce, statistical methods were widely adopted due to their simplicity and efficiency. As the scale and

complexity of observational data have increased, machine learning has gradually shown its advantages. However, traditional machine learning relies on manual feature extraction, which limits its ability to capture complex patterns in the data. In contrast, deep learning can directly process raw data (e.g., magnetograms), significantly improving prediction performance. In addition, the use of a single data type can lead to information loss, which is why fusion models play an important role in flare prediction. In this study, we will systematically analyze the advantages and disadvantages of different prediction models and their applicability across multiple dimensions, such as data types and task objectives.

3.1.1. Statistical-based Prediction Methods

Barnes et al. (2007) introduced a probabilistic prediction approach utilizing photospheric vector magnetic field data and developed a probabilistic prediction model by assuming that the probability distributions of the magnetic field variables follow a Gaussian distribution and evaluating the predictive power of the candidate variables through discriminant analysis. Subsequently, Barnes et al. (2007) further optimized the model on this basis, and by comparing it with Wheatland (2004), they demonstrated that the model has significant advantages in providing prediction confidence and probabilistic information.

To address the heterogeneity between and within active regions, Viet Do et al. (2024) proposed two hybrid models, MM-R and MM-H. The MM-R model uses a hybrid distribution to classify the active regions into three categories: H-class (mainly producing strong flares), I-class (dominated by weak flares), and L-class (with little to no strong flares). Building on this, the MM-H model is further refined to significantly improve the prediction of subsequent flare events by capturing flare activity patterns within the same region.

The Poisson method is widely used for operational prediction of solar flares because of its simplicity, flexibility, and effectiveness in dealing with rare events. Therefore, although different studies (Gallagher et al. 2002; Bloomfield et al. 2012) use different data sets, they all calculate the probability of flare occurrence using the Poisson method.

Stanislavsky et al. (2009) developed the fractional autoregressive integrated moving average (FARIMA) model and incorporated Pareto noise into it to simulate the heavy-tail effect in solar flare data. This approach was found to provide additional information and better predictive power when compared to other prediction methods such as those of NOAA and Wheatland.

Falconer et al. (2011) developed the MAG4 tool to assist NASA/SRAG in predicting flares, coronal mass ejections (CMEs) and solar energetic particles (SEPs). The tool extracts regions of strong magnetic fields from the MDI full-disk magnetograms and identifies them as active regions, calculates the free magnetic energy proxy variable ($^{L}WL_{SG}$), and converts it into flare incidence using an empirical power-law relationship. The power-law formula is given in Equation (2):

$$R = a(^{L}WL_{SG})^{b} \tag{2}$$

where R is the event occurrence rate, $^{L}WL_{SG}$ is the free magnetic energy proxy variable, and a and b are the fitting parameters. Based on this, Falconer introduced flare history data and developed Next MAG4 in conjunction with free energy proxies (Falconer et al. 2014). Compared to methods such as the McIntosh classification, total magnetic flux, and MAG4, the results indicate that Next MAG4 outperforms the other methods.

Lim et al. (2019) empirically fitted the power-law relationship between ten magnetic parameters obtained from SDO/HMI vector data and flare eruption rates, thereby predicting future flare occurrence rates. Murray et al. (2017) compared multiple prediction methods (such as Poisson statistical methods, multivariable regression models, and SVM) and pointed out that the forecasts manually adjusted by Mission Operations Support Wall Operations Center surpass the basic Poisson model results, emphasizing the crucial contribution of forecasters in managing space weather operations. They emphasized the advantages of combining human expertise with complex system predictions during the prediction process.

3.1.2. Machine Learning-based Prediction Methods

Traditional machine learning models typically rely on manually designed features, and their performance is heavily influenced by feature selection. These methods often require high linear separability and specific data distributions, which makes data preprocessing and feature selection crucial. While they tend to have relatively high computational efficiency, they may encounter limitations when applied to large-scale data sets.

Wheatland (2004) analyzed the regularity of solar flare events and their power-law distribution characteristics using a Bayesian approach based on historical statistical data to estimate the likelihood of future flare events. The study demonstrates that the Bayesian method effectively addresses uncertainty and quantifies the probability distribution of flare events by integrating historical data and dynamically updating predictions. Stanislavsky et al. (2020) introduced a Hidden Markov Model to enhance flare prediction performance using GOES solar X-ray data. The model incorporates two dynamic assumptions: Independent and Identically Distributed (IID), and random variables and autoregressive (AR) processes, constructing the IID-IID and AR-AR models, respectively. Performance evaluation based on multiple metrics, reveals that the AR-AR model surpasses the IID-IID model in capturing the dynamic characteristics of flare time-series, demonstrating higher accuracy and stability.

Bobra & Couvidat (2015) used parameters extracted from vector magnetic field data and applied the SVM algorithm to

predict ⩾M1.0-class flares. Domijan et al. (2019) systematically compared the performance of linear classifiers (binary logistic regression) with classifiers that allow nonlinear decision rules (Deep Neural Network (DNN), SVM, RF). Despite the higher complexity of nonlinear models, Logistic Regression (LR) performed best in terms of metrics such as TSS, HSS, and Area Under the Curve (AUC) under feature sparsity conditions.

Florios et al. (2018) extracted 13 key predictors from LoS and vector photospheric magnetogram data and evaluated the effectiveness of machine learning methods (MLP, SVM, RF) in comparison with statistical methods (linear regression, Probit, Logit) for predicting solar flares. The results showed that RF performed the best across all prediction tasks, while statistical methods achieved similar performance to machine learning methods under specific conditions. Sinha et al. (2022) evaluated the effectiveness of K-Nearest Neighbors (KNN), LR, RF, and SVM using SDO/HMI data. Evaluated using the severe space weather and clear space weather, TSS , and macro accuracy indicators, LR showed the best performance, followed by SVM, both of which outperformed KNN and RF, indicating the advantages of LR and SVM in certain prediction scenarios. Nishizuka et al. (2017) integrated vector magnetograms, LoS magnetograms, and UV data to extract multiple predictors and compared the performance of SVM, KNN, and Extreme Randomized Trees (ERT). Based on the TSS indicators, KNN demonstrated the most stable performance in predicting both central and peripheral regions, outperforming both SVM and ERT.

Huang et al. (2010) extracted three parameters (e.g., the maximum horizontal gradient, neutral line length, and the number of singularities) from longitudinal magnetograms. They applied rough set theory to eliminate redundant information and constructed an ensemble model based on the C4.5 decision tree algorithm. The study demonstrated that the ensemble model of decision tree algorithms performed superiorly in short-term flare prediction, effectively capturing the complex relationships between magnetic field features and flare activity.

3.1.3. Deep Learning-based Prediction Methods

A study (Wei et al. 2024) indicates that deep learning models typically achieve higher TSS scores than traditional models in predicting ⩾C-class and ⩾M-class solar flares. The deep neural net (DeFN) developed by Nishizuka et al. (2018) further improves prediction performance by integrating multiple solar observation features, with its TSS metrics outperforming those of methods such as SVM, KNN, and ERT. Additionally, Colak & Qahwaji (2009) compared the effectiveness of three methods-cascade correlation neural network (CCNN), SVM, and radial basis function network—using McIntosh classification as a predictor. The results showed that CCNN performed best in flare type classification, while SVM proved more robust in predicting whether a flare would occur.

Landa & Reuveni (2022) used a 1D Convolutional Neural Network (1D CNN) to build a solar flare prediction model that processes time-series data. Although the model demonstrated certain advantages in predicting major flare categories, there is still room for improvement in accurately distinguishing between different types of flares.

To capture the spatial features of the magnetic field, Huang et al. (2018) developed a CNN model for flare prediction based on LoS magnetogram data and demonstrated the generalization ability of deep learning under various observational conditions. Park et al. (2018) constructed the CNN model combining GoogLeNet and DenseNet for flare prediction using full-disk magnetogram data. Zheng et al. (2019) were the first to apply CNN to multi-class flare prediction by transforming the multi-class task into multiple binary classification problems and combining network structures such as the VGGNet (Qahwaji & Colak 2007; Colak & Qahwaji 2009; Krizhevsky et al. 2012) model. They found that the hybrid CNN model significantly outperforms other methods in predicting >M-class flares, demonstrating the accurate extraction of complex magnetic field features. Li et al. (2020) proposed a shuffle and split cross-validation method based on active region separation to construct data sets and validated the effectiveness and robustness of the CNN in flare prediction.

However, these 1D and 2D CNN models capture only static features and neglect the dynamic evolution of the magnetic field. To address this issue, Sun et al. (2022) developed models to predict flares using a 3D CNN with extended convolutional kernels for time-series data. The experiments show that the 3D CNN outperforms both the 2D CNN and LSTM in capturing dynamic changes along the time dimension, particularly in terms of TSS metrics. This indicates that the 3D CNN can effectively model the spatiotemporal features of the magnetic field and enhance prediction performance. In contrast, LSTM is better suited to handling long-term dependencies in time-series data. Therefore, when building a model, the choice of method should depend on the specific task and data characteristics.

To incorporate the dynamic evolutionary features of magnetic fields, Recurrent Neural Networks (RNN) and their variants have gained significant attention due to their superior performance in handling long-term dependencies. Liu et al. (2019) were the first to introduce LSTM into solar flare prediction, significantly improving model performance by adding an attention layer and a fully connected layer. In their ablation and comparative experiments, they validated the effectiveness of the attention and fully connected layers, demonstrating that the model has more advantages over traditional machine learning methods (such as SVM, MLP, and DeFN Nishizuka et al. 2018). Sun et al. (2022) used CNN and LSTM to construct prediction models based on images and magnetic field parameters, respectively. Their results showed

that the LSTM model outperformed the CNN model in metrics such as TSS and HSS, further demonstrating its ability to capture temporal features effectively.

To enhance the precision of flare intensity prediction, Jiao et al. (2020) developed a hybrid LSTM regression model. In this regression approach, flare intensity was treated as a continuous variable (ranging from −7 to −3) and was used to simultaneously predict both flare intensity and classification. Compared to the model by Chen et al. (2019), this method avoids the loss of intensity details and demonstrates higher prediction accuracy in practical applications. This regression approach offers a more detailed perspective for understanding and predicting flare intensity.

In addition, to explore the effect of time steps on model performance, Wei et al. (2024) compared the performance of eight models, including LSTM, BiLSTM, GRU (Gated Recurrent Unit), and BiGRU, with different time steps. They found that the TSS of all models improved with an increase in time steps and identified the optimal step length as 40. In the evaluation of different RNN models, Platts et al. (2022) demonstrated that LSTM and GRU performed similarly on balanced data sets, whereas Simple RNN exhibited poor performance due to the vanishing gradient problem. After increasing the number of network layers, the performance of GRU improved significantly, demonstrating a better balance between efficiency and accuracy. Although LSTM can capture long-term dependencies, its higher computational complexity and deeper model layers may lead to performance degradation. In contrast, GRU, with fewer parameters and higher computational efficiency, performs more stably in multi-layer models, making it better suited for long-term time series prediction.

Transformer models have unique advantages in capturing complex spatio-temporal relationships. In terms of time-series data, studies (Alshammari et al. 2023; Pelkum Donahue & Inceoglu 2024) have used Transformer models to process time-series data for flare prediction, showing significant improvement over traditional multivariate time-series (MVTS) models, such as LSTM (Liu et al. 2019; Wang et al. 2020) and RF (Alipour et al. 2019). They also indicated that incorporating historical flare data and extending the data window significantly enhanced the prediction of severe flares. For image data, Grim & Gradvohl (2024) introduced a multi-scale Vision Transformer model (MViTv2) based on SDO/HMI magnetogram data, which showed significant improvements through fine-tuning, particularly in handling complex spatiotemporal data. Compared to CNN models (Tang et al. 2021) and DeFN models (Nishizuka et al. 2018), MViTv2 demonstrated significant advantages.

Zhang et al. (2024) pointed out that existing research overfocuses on correlation analysis of data while neglecting the exploration of causality. They argued that effective predictive models should focus more on causal relationships, especially in identifying true precursor features of solar flares. Based on this perspective, they proposed the CausalNet model, which integrates a causal attention mechanism to predict flares and effectively suppress the interference of confounding factors.

Due to the use of different data sets in various studies, comparing different prediction methods solely based on metric scores may introduce bias. To better reflect the characteristics and focus of each method, we conducted a comparative analysis of solar flare prediction studies across multiple dimensions, including prediction window, prediction task, data type, and evaluation metrics. Table 2 presents the specific outcomes. Through the analysis, we found that most of the studies focus on processing magnetograms and time-series data, particularly using HMI and MDI data. In terms of model selection, researchers generally tend to favor LSTM, which is effective at handling time-series data, and CNN, which is effective for processing images. This preference forms the foundation for future research on multimodal fusion models. Regarding the choice of prediction window, flare predictions within 24 and 48 hr are the most common. For prediction tasks, most studies predict whether a flare will occur within a certain time frame (e.g., 24 hr), which is known as a binary classification problem. A few studies predict the continuous variable of flare intensity, which we categorize as a regression problem. This provides new research directions for future flare intensity prediction. Regarding evaluation metrics, TSS and HSS are typically used for flare binary classification problems, followed by Accuracy (ACC), Recall, and Precision. The primary cause of this is the data sets' class imbalance, as TSS and HSS effectively reduce the bias of accuracy toward the majority class, providing a more balanced evaluation score. Section 4.1.2 will go into greater details about the use and importance of these assessment measures.

3.1.4. Prediction Methods Based on Fusion Models

To overcome the limitations of using a single model and data source, many studies have adopted fusion models to enhance prediction performance by integrating information from multiple sources. The fusion process typically consists of three stages: feature fusion, algorithm fusion, and decision fusion. In the feature fusion stage, the model integrates information from multiple data sources to improve solar flare prediction, a process often referred to as multimodal learning. Algorithm fusion overcomes the limitations of a single model by cascading various algorithms. Finally, decision fusion combines predictions from multiple models to generate the final outcome. The simplified structures of the three fusion methods are shown in Figure 2.

(1) *Feature Fusion Model*. Combining data from different sources or features extracted by different models can enhance the information obtained about solar flares, thereby improving performance of solar flare prediction. This multi-source feature fusion strategy helps capture richer physical information, which

Table 2
Comparison of Deep Learning Based Prediction Methods

Model	Prediction Window	Prediction Task	Data Type	Data Details	Evaluation Indicators
DeFN (Nishizuka et al. 2018)	24 hr	Binary classification	Magnetogram	LoS and Vector Magnetogram 600 Å and 130 Å filter image	TSS,HSS,ACC FAR,POD,CSI
CNN (Park et al. 2018)	24 hr	Binary classification	Full-disk Magnetogram	HMI and MDI (1996–2008, 2009–2017,±30°)	TSS,HSS,ACC FAR,POD,CSI
CNN (Huang et al. 2018)	6 hr 12 hr 24 hr 48 hr	Binary classification	LoS Magnetogram	MDI and HMI (1996–2015)	TSS,HSS
3DCNN (Sun et al. 2022)	24 hr	Binary classification	Magnetogram Sequence	HMI (2010–2019, ±45°)	TSS,HSS,ACC Recall,FAR Precision
1DCNN (Landa & Reuveni 2022)	1 hr 3 hr 12 hr 24 hr 48 hr 72 hr 96 hr	Binary classification	Time-Series	X-ray (1998–2019)	TSS,HSS1 HSS2,ACC F1,TPR,PPV
LSTM (Liu et al. 2019)	24 hr	Binary classification	Time-Series	SHARP (2010–2018)	TSS,HSS,ACC Recall,BACC Precision
LSTM (Jiao et al. 2020)	6 hr 12 hr 24 hr 48 hr	Binary classification Regression	Time-Series	SHARP (2010–2018, ±68°)	TSS,HSS1,HSS2 ACC,F1,Recall Precision,MSE
Transformer (Pelkum Donahue & Inceoglu 2024)	24 hr 48 hr	Binary classification	Time-Series	SHARP (2010–2022)	TSS,HSS,ACC Recall,Precision BACC
MViTv2 (Grim & Gradvohl 2024)	24 hr 48 hr	Binary classification	LoS Magnetogram Sequence	HMI (2010–2019)	TSS,HSS,ACC FAR,POD,CSI Precision,F1

Note. The "±30°" in the "data details" column refers to the data selected within 30° of the central meridian. "1996–2008" refers to the data selected from the years 1996 to 2008.

in turn enhances prediction accuracy and robustness. Kaneda et al. (2022) used the self-attention and cross-attention mechanisms of the Transformer to construct magnetogram module (MM) and sunspot feature module (SFM) for processing both images and physical features. A feed-forward neural network and Softmax function produced the final prediction results, as shown in Figure 3. The fusion model outperforms the DeFN model in terms of Geometric Mean Gini Score (GMGS) and TSS.

Tang et al. (2021) constructed three prediction models-DNN, CNN, and BiLSTM-based on magnetogram data and sunspot group magnetic field feature parameters. The DNN and BiLSTM were used to extract magnetic field features, while the CNN focused on extracting magnetogram features. To combine the strengths of each model, the features extracted by all three models were fed into a fully connected layer and a classification layer to obtain the final results. Figure 4 illustrates the specific structure.

Li et al. (2022) compared two methods: using LoS magnetograms alone or combined them with prior knowledge (Mount Wilson classification and active region area), and constructed three models: Pure CNN, Fusion Model 1, and Fusion Model 2. The Pure CNN model was trained and predicted directly using the magnetograms without incorporating any additional prior knowledge. Model 1 uses prior knowledge of active region magnetic type for initial classification, and then applies the CNN model for final classification of the magnetograms, optimizing CNN's prediction performance. Model 2 integrates the active region area as an additional input parameter, further improving prediction performance. Compared to the Pure CNN model, both Model 1 and Model 2 demonstrate better performance in solar flare prediction, with Model 2 exhibiting the best performance. This indicates that integrating prior knowledge and area information can significantly improve prediction accuracy.

Li et al. (2024) developed the STCNet model, which uses the Swin Transformer and CNN models to extract magnetogram and magnetic field features, respectively. After fusion, the features are input into a linear classifier. The specific structure is shown in Figure 5. This model enhances the feature representation of each modality through a cross-modal attention mechanism, and then integrates the features of all modalities by concatenation, significantly improving classification performance.

(2) *Algorithm Fusion Model*. To incorporate the benefits of several models, many studies have adopted a cascading approach, where multiple algorithmic modules are sequentially stacked within a network framework. To predict whether

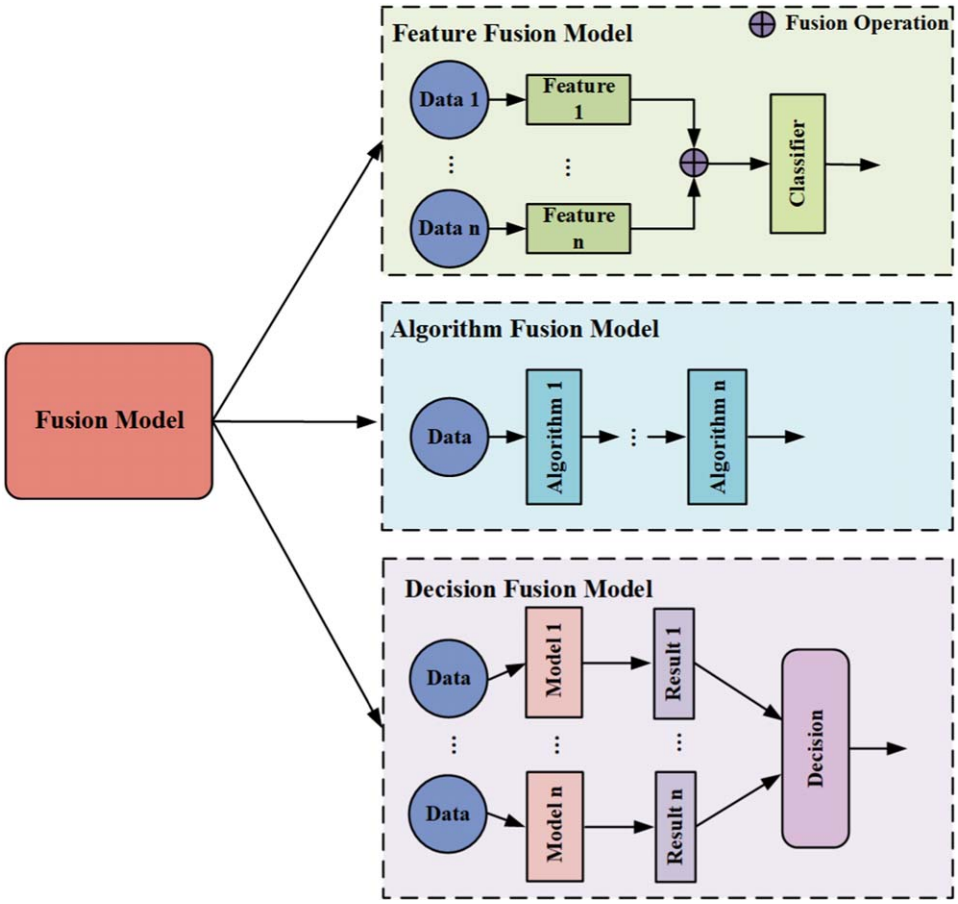

Figure 2. The simplified structures of the three fusion methods. In the Feature Fusion Model, Data 1 and Data n represent data from different sources or of different types, while Feature 1 and Feature n represent the features extracted from the different data. Note that in the Decision Fusion Model, the input data for Model 1 and Model n can either be the same data or data from different sources. Moreover, these models are independently trained.

M-class or stronger flares would occur in the following two days, Li et al. (2007) suggested a flare prediction method that combines the SVM and KNN algorithms, using solar active region data and 10 cm radio flux data. In the model, preliminary classification is performed using SVM, and samples close to the hyperplane are passed to the KNN classifier for further decision-making based on their proximity to the hyperplane, while samples farther from the hyperplane are directly classified by SVM. The SVM-KNN model optimizes SVM's classification performance and outperforms conventional SVM and neural network (NN) techniques in terms of prediction accuracy.

Benvenuto et al. (2018) proposed a hybrid method combining LASSO and Fuzzy C-Means (FCM), where feature weights are computed using LASSO and multiplied by the eigenvalues to generate a real-valued set, which is then classified into two categories by FCM. This method effectively selects important features and optimizes classification decisions through FCM, thereby improving prediction performance. In contrast to conventional machine learning techniques (such as L1-logit, LASSO, SVM, RF, and MLP), the hybrid method not only demonstrates competitive prediction performance but also provides feature weight calculation, a capability that is not available with unsupervised clustering methods.

Deng et al. (2023) proposed a new hierarchical prediction method (HPF). First, the technique predicts whether an active region is part of the "flare region" using Gaussian Naive Bayes (NB) and Balanced Random Forest (BRF) models. Then, a cascaded model combining the Logistic Generalized Additive Model (LogistiGAM), Random UnderSampling Boosting (RUSBoost), and Linear Discriminant Analysis (LDA) is used to predict flare region samples. In the cascaded model, non-flare region samples and negative samples are predicted using conventional baseline models. By using a phased approach, this technique reduces the negative effects of class imbalance on model performance and greatly increases the accuracy of predictions.

As attention mechanisms have evolved, researchers have concentrated on exploiting them to modify the model's focus on various input features so that it can focus more on the most

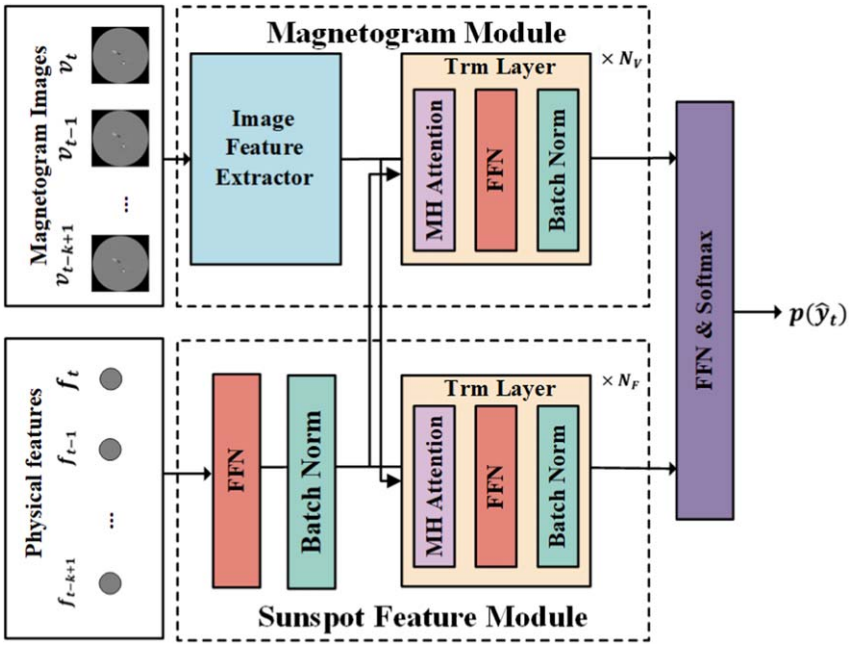

Figure 3. The structure diagram of the Flare Transformer Prediction Model.

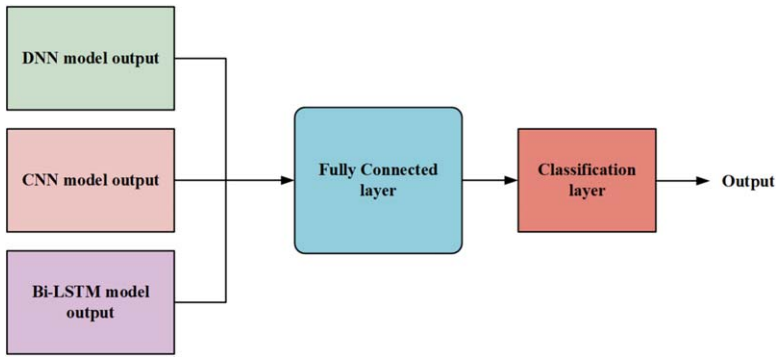

Figure 4. The structure diagram of the DNN-CNN-BiLSTM Fusion Model.

important and pertinent information. Pandey et al. (2023b) introduced an attention module into the CNN model to predict ⩾M1.0-class solar flares using full-disk LoS magnetograms. This method effectively helps the model focus on important regions, reducing false positives and false negatives. Compared to the baseline CNN model, the model with the added attention module demonstrated better prediction performance, especially in predicting flares near the solar limb. Yan et al. (2024) used CNN as the baseline model and compared it with CNN models that included Squeeze and Excitation (SE), Convolutional Block Attention Module (CBAM), and Efficient Channel Attention (ECA) mechanisms. The experimental results show that the model with the added attention mechanism outperforms the baseline CNN model, particularly in terms of the Recall metric.

Zheng et al. (2023a, 2023b) proposed a BiLSTM-A model that incorporates BiLSTM and attention mechanisms, as shown in Figure 6. This model was compared with traditional neural network (NN), LSTM, LSTM-A, and BiLSTM models. The results showed that LSTM and BiLSTM performed better in predicting ⩾C-class flares and outperformed the other models in ⩾M-class flare predictions. However, adding the attention mechanism did not significantly improve prediction performance. Subsequently, they transformed the four-class classification problem into multiple binary classification problems and constructed three binary-class BiLSTM-A models, which together formed the HBiLSTM-A model, as shown in Figure 7. The comparison results with the OAO-CNN (Zheng et al. 2021) and H-CNN (Zheng et al. 2019) models demonstrated that HBiLSTM-A achieved better performance in the BSS metrics.

In Section 3.1.3, we mentioned that there are limitations in using CNN and RNN (and their variants) individually. To

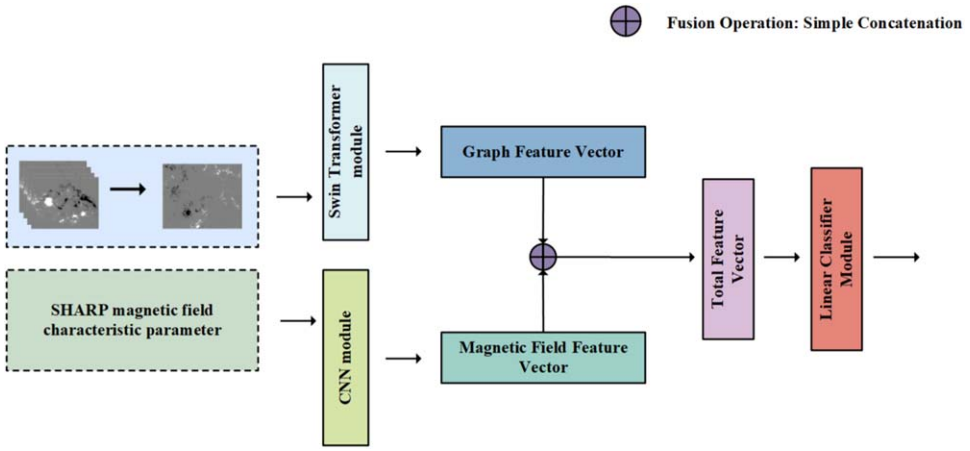

Figure 5. The structure diagram of the STCNet Model.

combine the static spatial features with evolving temporal features of the data, some studies have explored the possibility of integrating CNN and RNN. Guastavino et al. (2022, 2023) introduced video data (each sample consisting of 40 images) and proposed the Long Short Term Memory Convolutional Network (LRCN). This model first processes LoS magnetogram videos through the CNN module, transforming them into a time-series of 64 features, followed by sequence analysis using the LSTM module, which ultimately produces binary classification results. This technique greatly enhances flare prediction performance by fusing CNN's spatial feature extraction with LSTM's time-series modeling capabilities. A CNN-GRU fusion approach was proposed by Wan et al. (2022), which combines GRU's sequence modeling skills with CNN's local feature extraction capabilities. The CNN-GRU model outperforms the SVM, CNN, LSTM, and CNN-LSTM models in flare prediction, with significant improvements in both TSS and HSS indicators.

Abduallah et al. (2023) proposed the SolarFlareNet model, in which the SHARP time-series are first fed into a 1D CNN to enhance local feature extraction. The model then uses an LSTM to handle the complex dependencies within the sequence, and finally, a Transformer encoder captures the global dependencies, as shown in Figure 8. The results of the ablation tests confirm the contribution of each component to the overall model performance and show that the entire SolarFlareNet model performs the best. This demonstrates the effectiveness of combining CNN, LSTM, and Transformer to improve model accuracy.

(3) *Decision Fusion Model*. The decision fusion model integrates the outputs of multiple independently trained models using specific strategies to generate the final prediction. This method aims to combine the advantages of each model, thereby enhancing prediction performance and robustness. In the field of flare prediction, common decision fusion strategies include voting methods and meta-learner-based methods, and others.

In the voting method, common strategies include majority voting, median voting, and weighted voting. The majority voting method determines the final classification result by selecting the category with the most frequent prediction outcomes. Abduallah et al. (2021) and Ribeiro & Gradvohl (2021) both used the majority voting method to integrate RF, MLP, and ELM, as well as RF. Light Gradient Boosting Machine (LightGBM), and SVM, respectively. They both noted that combining different types of machine learning algorithms can effectively enhance the ability to solve complex problems.

A novel deep convolutional neural network (OAR-CNN) based on the H-CNN model (Zheng et al. 2019) was proposed by Zhang et al. (2024), which employs a One-vs-All strategy to decompose the four-class flare prediction problem into four binary classification sub-problems, using independent CNN models for prediction. Majority voting and probability thresholding were the two decision-making techniques used to obtain the final forecast outcomes, with the specific network structure shown in Figure 9. Experimental results revealed that the probability thresholding strategy provided superior classification performance in most cases.

van der Sande et al. (2023) separately trained CNN and LR models to better leverage both the full-disk magnetogram and flare history information, integrating their outputs using median voting. The CNN+LR model outperformed both the LR and CNN models individually, showing that combining magnetogram and history features improves prediction accuracy more than using features from the magnetogram alone.

The weighted voting method is similar to majority voting but optimizes the prediction results by assigning weight to each model. The weights are typically assigned based on model performance metrics such as Recall, Precision. Liu et al. (2017b) proposed a Multi-Model Integration Method (MIM), which uses genetic algorithms to optimize the weights of multiple machine learning models (NB, SVM, Sequential Minimal Optimization,

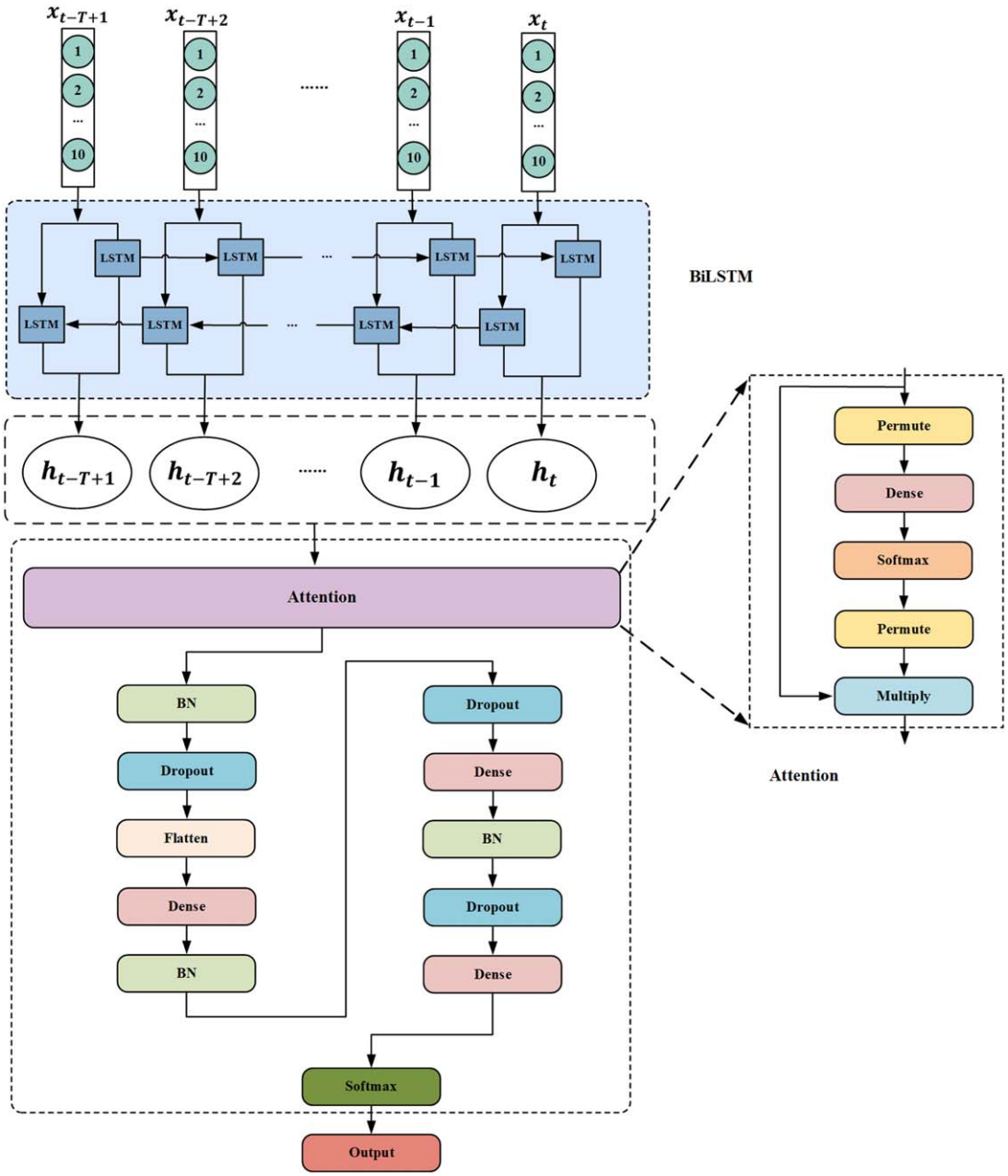

Figure 6. The structure diagram of the BiLSTM-Attention Model.

MLP, Radial Basis Function Networks, and C4.5 Decision Trees) to achieve optimal combination. Their study shows that model diversity significantly impacts the performance of the ensemble model, and combinations with greater diversity perform better after integration. In the future, MIM could further enhance its generalization ability by integrating additional models (e.g., physical, expert, and statistical models). Guerra et al. (2015) developed a solar flare prediction model based on linear weighted integration, which fuses the outputs of several independent prediction methods (MAG4, ASSA, ASAP, and NOAA). However, ASSA, ASAP, and MAG4 are completely automated techniques, while NOAA combines human expertise with automation. Experiments show that linear weighted integration outperforms individual methods in predicting M-class flares, and human expertise remains irreplaceable in some cases, especially when contrasted with fully automated systems.

The ensemble method based on meta-learners is an integration strategy that utilizes meta-learning concepts to

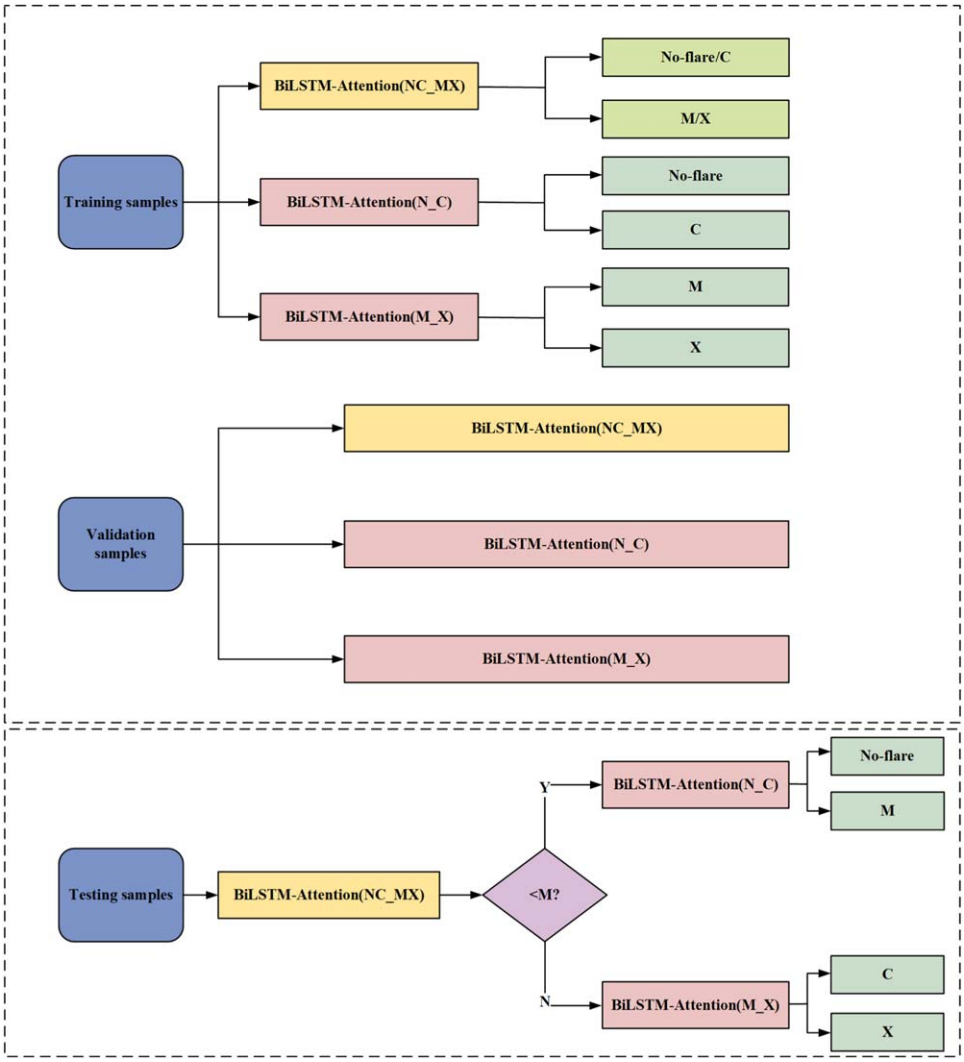

Figure 7. The structure diagram of the HBiLSTM-Attention Model.

improve model performance. These methods function by training a single meta-learner to optimize the prediction results of multiple base models. Common techniques include Boosting, Blending, Bagging, and Stacking methods. To create a more powerful learner, the Boosting approach combines several weak learners, with the core principle being the use of subsequent models to correct the errors of preceding models. Chen et al. (2022) suggested a two-stage early warning system for solar flare prediction, where Boosting was employed in the second stage to integrate the Xception, ResNet18, and ResNet34 models. The system achieved an F1 score of 0.5639, which demonstrates its effectiveness.

Pandey et al. (2022) developed a flare prediction system by combining a full-disk model with an active region model. The full-disk model used AlexNet to predict flares based on LoS magnetograms, while the active region model employed multivariate time-series data to train a Time Series Forest (TSF). These two models are defined as base learners, and their results are fused using a meta-learner based on logistic regression, as detailed in Figure 10. Experimental results showed that the logistic regression meta-learner significantly enhanced prediction performance, highlighting the advantages of the weighted integration strategy in exploiting the complementary information of the base models.

Sun et al. (2022) used the SMARP and SHARP data sets to examine how well CNN and LSTM models performed in flare prediction, and found that the LSTM model based on aggregated parameters outperforms the CNN model that directly uses the magnetogram. Furthermore, they used a meta-learner to fuse the LSTM and CNN models and proposed a weighted stacking method based on the outputs of both models. The meta-learner formula is shown in Equation (3),

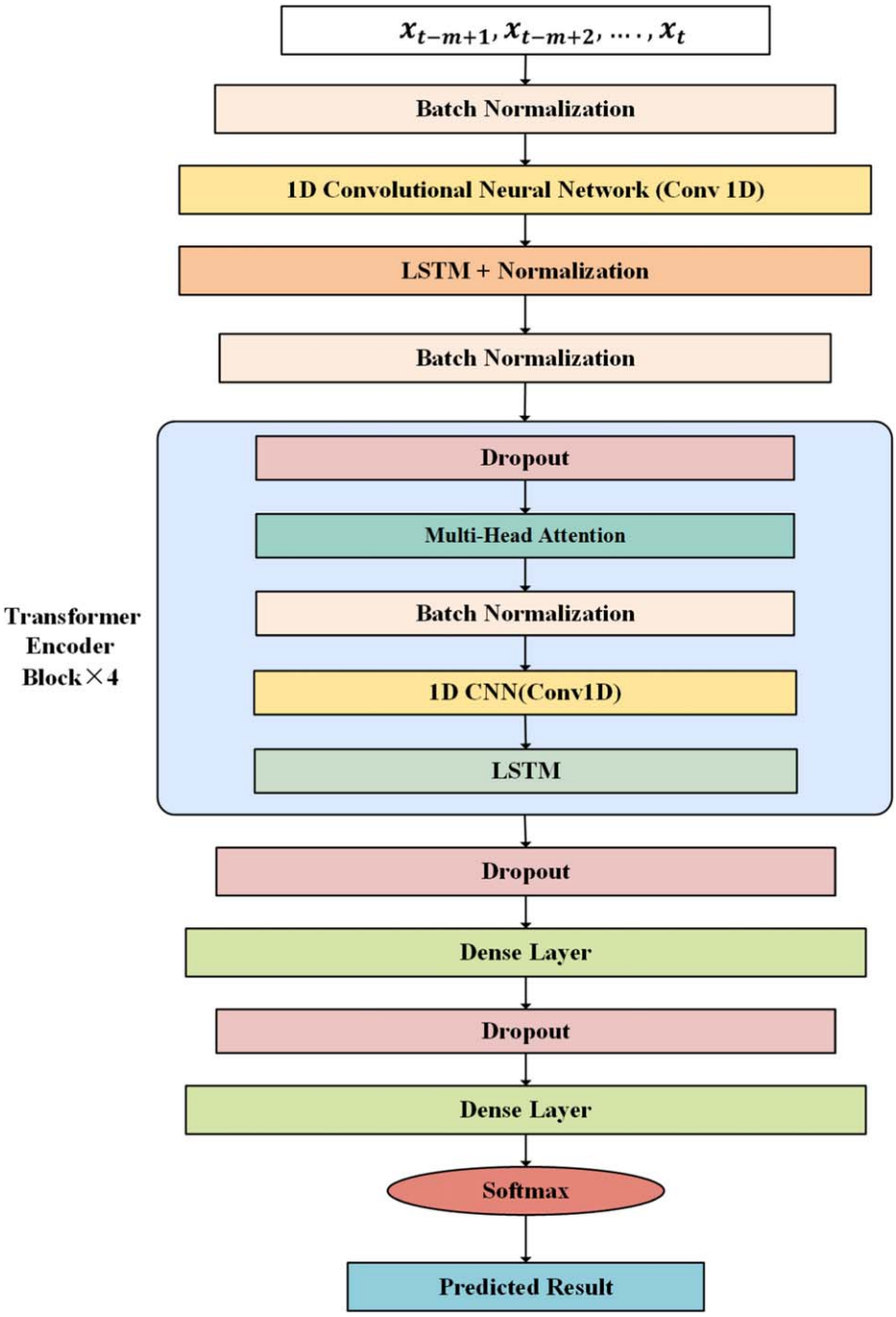

Figure 8. The structure diagram of the SolarFlareNet Model.

where p_i and q_i are the outputs of the LSTM and CNN, respectively, and α is the meta-learner parameter. The experiment showed that the stacked model performs significantly better than the individual model and other meta-learners (such as average and best model selection). Moreover, they pointed out that the fused results rely more on the LSTM model, which proves the importance of time-series information in solar flare prediction.

$$r_i = \alpha p_i + (1 - \alpha) q_i, 0 \leqslant \alpha \leqslant 1 \quad (3)$$

To gain a more comprehensive understanding of fusion methods in solar flare prediction, we conducted a comparative analysis of several studies based on dimensions such as prediction window, prediction tasks, data types, and evaluation

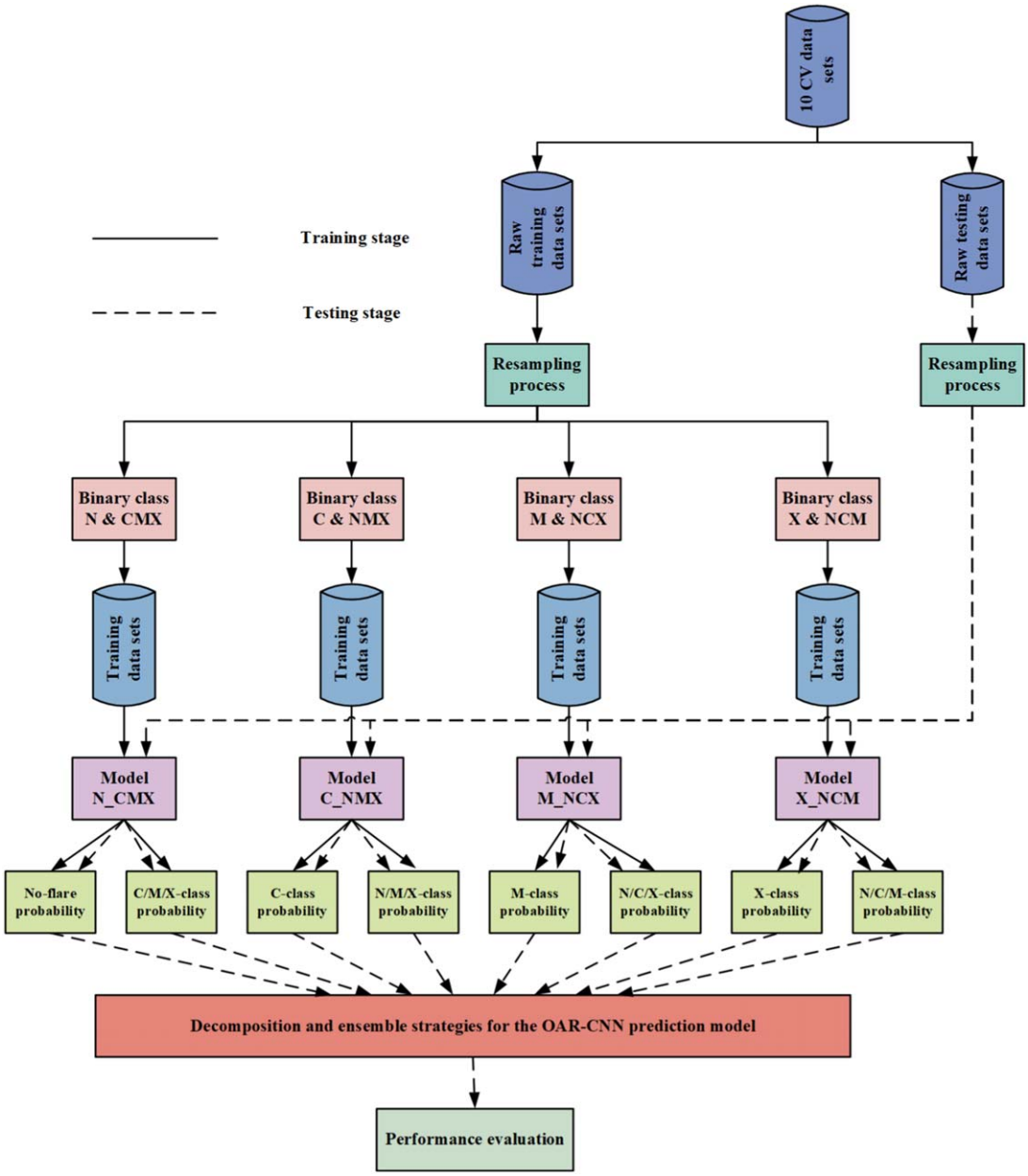

Figure 9. The structure diagram of the OAR-CNN Model.

metrics (see Table 3). The results indicate that current research mainly focuses on two directions: multimodal data fusion and integrating the strengths of different algorithms. In the realm of data fusion, most studies concentrate on combining magnetogram and time-series data. Common fusion strategies include feature fusion and decision fusion, with combinations of CNN and LSTM commonly employed to leverage the complementary information from different modalities. CNN is particularly effective at extracting spatial features from images, while LSTM excels at capturing dynamic changes in time-series data. The integration of these two models allows for the simultaneous processing of both image and temporal data, thereby improving prediction accuracy.

In terms of algorithm fusion, current mainstream approaches include the introduction of attention mechanisms to enhance the model's ability to assign appropriate weights to different types of features, or the construction of more powerful prediction frameworks by stacking deep learning models (e.g., CNN, RNN, Transformer). Specifically, Transformer models are better at handling long-range dependencies, while

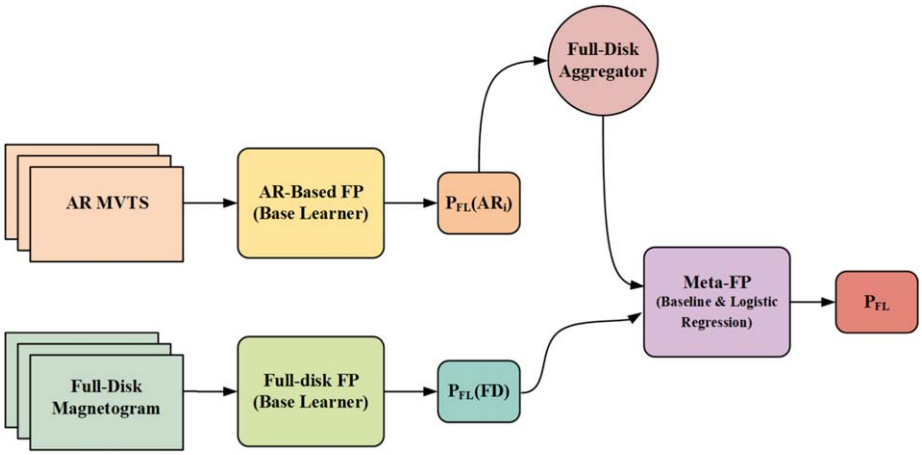

Figure 10. The structure diagram of the Meta-learner-based Ensemble Model.

RNN are well-suited for sequential data. Stacking these models provides stronger feature representations and improved prediction accuracy. Additionally, decision fusion plays a crucial role in optimizing predictive performance. This method not only integrates the advantages of data and algorithms from different modalities but also transforms multi-class problems into multiple binary classification tasks, effectively addressing multi-class prediction challenges. This method effectively reduces the bias of individual classifiers and enhances the robustness and accuracy of the final multi-class decision.

3.2. Physics-based Prediction Methods

Solar flares are thought to be quick releases of magnetic energy that have been stored in the solar atmosphere, according to the MHD hypothesis, and are usually accompanied by violent bursts of solar plasma (Shibata & Magara 2011). Prior research has demonstrated a strong correlation between the solar magnetic loop's stability and the magnetic field's twisting (Hood & Priest 1979; Török et al. 2004; Kliem & Török 2006). Excessive twisting leads to instability of the magnetic loop, which in turn may trigger flares.Thus, MHD instability may act as a driver of solar flares. Therefore, MHD-based numerical simulations can reproduce the dynamical evolution of the solar coronal magnetic field, providing a useful tool for exploring flare eruptions. Guo et al. (2024) developed a data-driven model based on the zero-beta MHD theory and introduced a new method for calculating the decay index, which was used to explain and predict the direction of the flare outburst. However, solar flare prediction methods based on MHD instability still have uncertainties and lack wide practical applications.

Another study (Ishiguro & Kusano 2017) suggested that the DAI mechanism, through the growth of currents and magnetic field twisting, may trigger instability in the early stages of solar flare eruptions, thereby accelerating the eruption process. Therefore, DAI could be the cause of the initial flare eruption event. A flare prediction technique based on DAI was presented by Kusano et al. (2020), called the k-scheme. The method uses the k-value, which represents the degree of magnetic field instability, to predict when solar flares may occur. When the *k*-value exceeds a critical value (approximately 0.1), it suggests that a flare might happen because the solar magnetic field might be getting close to a critical instability. Unlike prediction methods that rely on empirical relationships, regardless of prior flare activity, the *k*-scheme can predict the occurrence, location, and expected strength of flares based on measurements of the magnetic field at the solar surface.

SOC refers to a dynamical process in which a system spontaneously evolves toward a critical state through local interactions, without external control. In this critical state, small perturbations can trigger cascades of abrupt stochastic events, often exhibiting "avalanche-like" amplification. The sizes and frequencies of these events follow power-law distributions, a feature that has motivated extensive applications of SOC theory to astrophysical phenomena (Osokin et al. 2004; Aschwanden et al. 2016). Lu & Hamilton (1991) were the first to apply SOC theory to analyze the triggering mechanism of solar flares, proposing that solar flares result from a series of local magnetic reconnection events, akin to an "avalanche effect." Their cellular automaton model simulated coronal magnetic field evolution using a lattice of nodes, with each node representing a magnetic field. When the accumulated magnetic energy exceeds the critical threshold of a node, magnetic reconnection occurs, releasing energy and potentially disturbing neighboring nodes, leading to further reconnection events and ultimately triggering a flare eruption. Bélanger et al. (2007) employed the four-dimensional variational data assimilation (4D-VAR) technique to optimize the initial conditions of the SOC model, adjusting its starting state to minimize forecast errors, and applying this approach to flare prediction. Hamon et al. (2002) analyzed the Olami-Feder-Christensen (OFC) model at finite driving rates (Olami et al. 1992), focusing on its non-SOC state and dynamic

Table 3
Comparison of Fusion Models

Model	Prediction Window	Prediction Task	Data Type	Data Details	Evaluation Indicators
RF-MLP-ELM (Abduallah et al. 2021)	24 hr	Multi-class	Time-Series	SHARP (2010–2016, ±70°)	TSS,BASS
RF-LightGBM-SVM (Ribeiro & Gradvohl 2021)	24 hr 48 hr 72 hr	Binary classification	Time-Series	SHARP and DSD (2010–2019)	TSS,HSS,ACC F1,TPR,TNR PPV
CNN-SE,CNN-ECA CNN-CBAM,ViT (Yan et al. 2024)	24 hr	Binary classification	Full-disk Magnetogram	HMI (2010–2023)	TSS,BSS
HBiLSTM-Attention (Zheng et al. 2023a)	24 hr	Multi-class	LoS Magnetogram	SHARP (2010–2018, ±45°)	TSS,HSS,ACC FAR,Recall,Precision
LSTM-Attention and BiLSTM-Attention (Zheng et al. 2023b)	24 hr	Binary classification	Time-Series	SHARP (2010–2018, ±45°)	TSS
OAR-CNN (S. Zhang et al. 2024)	24 hr	Multi-class	LoS Magnetogram	HMI (2010–2018,±45°)	TSS,HSS,ACC FAR,Recall,Precision
CNN-LSTM (Guastavino et al. 2023)	24 hr	Binary classification	Magnetogram Sequence	SHARP (2010–2017)	TSS,HSS,ACC,F1 Recall,Precision
CNN-GRU (Wan et al. 2022)	48 hr	Binary classification	Time-Series	X-ray (1998–2019)	TSS,HSS1,HSS2 ACC,F1,TPR,PPV
1DCNN-LSTM-Transformer (Abduallah et al. 2023)	24 hr 48 hr 72 hr	Binary classification	Time-Series	SHARP (2010–2022, ±70°)	TSS,ACC,BACC Recall, Precision BS,BSS
CNN (Li et al. 2022)	48 hr	Binary classification	LoS Magnetogram and Prior Knowledge	HMI (2010–2018, ±30°)	TSS,HSS,ACC F1,Recall Precision,CSI
1DCNN-Swin Transformer (Li et al. 2024)	48 hr	Binary classification	Time-Series and Longitudinal Magnetogram	HMI (2010–2019, ±45°)	TSS,ACC,FAR F1
CNN-LSTM (Z. Sun et al. 2022)	24 hr	Binary classification	Time-Series and LoS Magnetogram	SMARP and SHARP (2010–2020, ±70°)	TSS,ACC,AUC BSS
DNN-BiLSTM-CNN (Tang et al. 2021)	24 hr 48 hr	Binary classification	Time-Series and LoS Magnetogram	SHARP (2010–2018, ±30°)	TSS,ACC,TPR TNR,F1

Note. The "±30°" in the "data details" column refers to the data selected within 30° of the central meridian. "1996–2008" refers to the data selected from the years 1996 to 2008.

behavior. They found that the dynamic characteristics exhibited by the OFC model under finite driving rates, such as power-law distributions and finite periodicity, are similar to the actual observational features of solar flares. This suggests that the OFC model could provide valuable predictive information and may better align with the observed behavior of solar flares than the traditional SOC model.

Morales & Santos (2020) revisited and simulated the avalanche model of Lu & Hamilton (1991), and first validated its applicability in describing and predicting solar flares. Subsequently, they proposed a new method for flare prediction by using the average value of lattice nodes as an effective indicator for predicting solar flares.

Strugarek & Charbonneau (2014) evaluated the performance of several representative avalanche models in predicting solar flares. Their study found that only avalanche models purely driven by deterministic processes could reliably predict flare events. Therefore, they modified the position of the stochastic process in the avalanche model to develop a deterministic-driven model, and experimentally verified the model's validity. Moreover, this model was combined with data assimilation techniques, which further enhanced prediction accuracy. Morales & Charbonneau (2008) introduced a new solar flare simulation method based on the avalanche model, which generates scale-free size distributions similar to solar flares by simulating the random deformation of magnetic flux lines and reconnection events. These results demonstrated that this method successfully reproduced the energy and its scale distributions of solar flares observed in real data, providing significant theoretical support for understanding the flare triggering mechanism and scale distribution.

Thibeault et al. (2022) proposed an innovative solar flare prediction scheme by combining the avalanche model with data assimilation techniques. Their research showed that the

avalanche model, based on the SOC theory, could effectively predict large flare occurrences, particularly in models with strong non-conservative driving forces. The integration of data assimilation techniques significantly improved prediction skills, especially for large flares.

3.3. Other Prediction Methods

Although machine learning-based prediction methods achieve high accuracy, their results are often difficult for forecasters to interpret effectively. Meanwhile, forecasters' experience contains valuable information. Therefore, a study (Liu et al. 2017c) adopted an image-based case reasoning method, combining it with a genetic algorithm to optimize feature weights and the number of neighbors in case retrieval, thereby improving model performance and enhancing the interpretability of the forecast results. The method effectively facilitates the active participation of forecasters. A comparison with expert systems and traditional machine learning methods reveals that this approach has significant advantages in metrics such as True Positive Rate (TP rate), HSS, and ACC. These findings suggest that integrating human expertise with machine learning methods can significantly enhance forecasting effectiveness and ensure the quality of decision-making.

Instead of using traditional numerical simulations or machine learning algorithms to directly predict solar flares, Gheibi et al. (2017) achieved indirectly flare prediction by constructing and analyzing a complex network of solar flares and investigating the underlying regularities in flare occurrence. The network's topology, degree distribution, and small-world properties, and other characteristics provided predictive clues for flare events. Najafi et al. (2020) proposed a new model for constructing a solar flare network by combining the Abe-Suzuki model with visibility graph conditions to more accurately describe the relationships between flare events. This improved the solar flare network model, along with its network feature analysis, provides new insights and methods for understanding the statistical properties and interrelationships of solar flares.

Steward et al. (2011) used solar magnetograms to threshold the LoS component for automatically identifying active regions. After identifying the active region, the potential for flare eruptions was assessed by locating the Strong Gradient Polarity Inversion Line (SPIL) and calculating the following parameters: the length of the SPIL, the maximum east–west and north–south gradients of the magnetic field in the vicinity, and the line integral of the magnetic field gradient along the SPIL. The potential for flare eruptions was further evaluated based on parameter variations using threshold decision rules.

Wang et al. (2024) explored the impact of the ionospheric D-region on solar flare prediction by analyzing Very Low Frequency (VLF) measurements of solar flares. They found significant differences in the curve-fitting coefficients between the VLF signal and flare X-ray flux under daytime, nighttime, and mixed-path conditions. These findings suggest that the diversity of propagation paths should be considered when making flare predictions.

Silva et al. (2024) aimed to fully exploit the information embedded in the higher-order Balmer series spectra by using radiation-hydrodynamics (RHD) simulation tools to generate synthetic spectra. This approach effectively diagnoses the electron density during flares, providing crucial data for flare prediction. Furthermore, the time-evolutionary features of the spectra, especially the broadening and asymmetry of the spectral lines, reveal the dynamic processes of plasma motion, which could serve as important indicators for predicting flare behavior. Notably, the findings of this work advance the nexus of solar and stellar physics by providing important insights into the study of stellar flares in addition to being applicable to solar flare prediction.

In addition to predicting flare eruptions, some studies focus on prediction the remaining duration of an already erupted flare. Reep & Barnes (2021) used *GOES*/XRS data and employed random forests to predict the remaining duration of solar flares.

3.4. Section Summary

Currently, the models used for solar flare prediction are mainly divided into data-driven models and physical models. In physical models, current research mainly focuses on building flare prediction models based on three theories: SOC theory, MHD theory, and DAI instability theory. Among these, models based on SOC theory are the most widely used. However, due to the unclear physical mechanisms behind flares, these methods still face instability issues.

This work examines the use of statistical techniques, traditional machine learning approaches, deep learning approaches, and ensemble approaches in data-driven models for solar flare prediction. Statistical methods are particularly efficient and computationally simple, especially when data is scarce. However, they are unable to deeply capture the deep features in complex data and are less effective for large-scale data analysis. Traditional machine learning methods (e.g., SVM, RF, and decision trees) rely on manually engineered features and relatively simple algorithms for classification and prediction. While these methods achieved some success in early studies, they exhibit significant limitations when confronted with complex, high-dimensional data.

Currently, flare prediction research focuses on deep learning and fusion methods, particularly the use of CNN for processing image data and methods like LSTM and GRU for capturing time-series information. Most research focuses on prediction windows of 24 or 48 hr, with the primary task being binary classification. Fusion methods currently focus on two main directions: multimodal data fusion and algorithmic merit

integration. Feature fusion and decision fusion are two common components of multimodal data fusion. The former improves the expressiveness of the model by combining features from various data sources, while the latter combines the predictions of several models to boost performance. In terms of algorithmic merit integration, mainstream approaches include introducing attention mechanisms to enhance the model's ability to assign feature weights, as well as constructing a more robust prediction framework by stacking multiple deep learning models (e.g., CNN, RNN, Transformer). However, existing approaches still face challenges in capturing complex feature interactions and inter-model dependencies, which are particularly pronounced in high-dimensional data and complex nonlinear problems, thereby limiting their performance in high-complexity tasks.

Future studies should thus concentrate on investigating more effective feature fusion procedures, including multimodal information fusion based on attention mechanisms, or enhancing the model's capacity for generalization through the use of transfer learning and self-supervised learning methodologies. Additionally, to overcome the drawbacks of deep learning models, researchers should investigate the interpretability and generalization capabilities of these models to better tackle the complex challenges posed by real-world applications, such as multitasking, long time windows, and noisy data.

4. Current Challenges

Physical model-based flare prediction methods are limited by the complexity of magnetic energy release processes, leading to significant uncertainties. In contrast, data-driven flare prediction methods have become mainstream. These methods rely on large amounts of observational data and automatically identify potential patterns through techniques such as machine learning. However, data-driven methods also face several challenges in practical applications, such as data complexity (e.g., feature redundancy and class imbalance), integration of multimodal data, and model interpretability. These challenges are particularly prominent in practical applications of data-driven methods, and addressing them is crucial to improving the accuracy and reliability of flare predictions.

4.1. Complexity of Flare Samples

4.1.1. Feature Redundancy

With continuous advancements in solar observation technology, an increasing amount of observational data is being collected, providing richer information for solar flare prediction, thereby improving prediction accuracy. However, this also introduces new challenges. Solar observational data are typically high-dimensional and encompass multiple features, such as sunspot size, shape, location, magnetic field strength, and the spectral characteristics of solar flares. There are two main issues in processing these high-dimensional features: first, not all features are directly related to flare prediction, and an increase in the number of features may significantly raise model complexity, which in turn affects the model's generalization ability and computational efficiency; second, some features exhibit strong correlations, and studies have shown that feature redundancy can degrade prediction performance (Campi et al. 2019). Therefore, effectively managing these high-dimensional and highly correlated features to enhance model stability and reliability has become a key challenge in current research.

Dimensionality reduction techniques can be applied to reduce data redundancy. The two most widely mentioned methods in flare prediction are selecting more important features through comparison and mapping high-dimensional features to lower-dimensional features. Among the feature selection methods, common approaches include filter methods, wrapper methods, embedding methods, and ensemble methods. Using statistical metrics, filter methods choose features based on how well they relate to the target variable. Ahmed et al. (2013) employed two feature evaluation techniques: Correlation-based Feature Selection (CFS) and Minimum Redundancy Maximum Relevance (mRMR), and selected 11 magnetic features. Cinto et al. (2020b) used a univariate feature selection technique based on the F-score, finding that it helps enhance model generalization and mitigate overfitting. Domijan et al. (2019) determined the ratio of the between-class sum of squares to the within-class sum of squares to rank features using marginal correlation, selecting those with the highest discriminative power. Common embedding methods include techniques such as Lasso regression, decision trees, and RF. Lasso regression compresses the weights of irrelevant features to zero using sparsity constraints, while decision trees and RF rank features based on their importance. Therefore, many studies (Campi et al. 2019; Liu et al. 2017a) evaluate the importance of features using the RF algorithm and select the most important ones. Velanki et al. (2024) constructed an ensemble model for feature selection by assigning weights to three methods-Mutual Information (MI), mRMR, and Euclidean Distance-based on the TSS metric. They discovered that the ideal number of features ranges from 9 to 15, and that the ensemble approach performs noticeably better than individual approaches.

Other methods aim to reduce the data's dimensionality by creating a few new features that replace the original features. Common techniques include PCA (Tenenbaum et al. 2000), Linear Discriminant Analysis (LDA), Kernel PCA and t-Distributed Stochastic Neighbor Embedding (t-SNE). Yuan et al. (2020) employed PCA to decrease the dimensionality of the data and identify the important features, such as sunspot activity region parameters and the 10.7 cm solar radio flux. They highlighted that PCA helps to extract important features from the data by reducing data dimensionality and removing

feature correlation in the process, and helps with feature engineering and data visualization, thus improving model performance and analytical capabilities.

4.1.2. Class Imbalance

The uneven distribution of samples among several classes in a classification job, where some classes contain a noticeably greater number of examples than others, is known as the class imbalance problem. This issue is particularly prominent in solar flare prediction, as large-scale flares (e.g., M-class or X-class) are much less frequent, while smaller-scale flares (e.g., C-class) occur more commonly. This imbalance may cause the model to favor the majority class during training, leading to overfitting its features and neglecting the prediction of the minority class, which ultimately affects the model's overall performance (He & Garcia 2009).

To effectively address the class imbalance problem, researchers have proposed various solutions at the data level, algorithmic level, and evaluation metrics level. At the data level, common approaches include resampling and data augmentation. Resampling methods are primarily categorized into oversampling, undersampling, and hybrid sampling. By oversampling and using data augmentation methods including jittering, smoothing, scaling, and magnitude warping, Ji et al. (2022) created synthetic samples that greatly enhanced the model's capacity to forecast minority class samples. Vysakh & Mayank (2023) employed the SMOTE (Han et al. 2005) method in conjunction with random undersampling. They found that this approach performs best when generating synthetic samples close to the decision boundary, significantly enhancing the classifier's ability to recognize minority class samples. Cinto et al. (2020b) combined SMOTE with ENN (SMOTE-ENN) to generate minority class samples and clean up noisy data, further improving model performance. Wan et al. (2023) introduced density clustering combined with SMOTE, demonstrating its advantages in addressing the class imbalance problem in solar flare prediction. Deng et al. (2021) used Generative Adversarial Networks (GAN) to generate X-class flare samples, effectively alleviating the problem of insufficient minority class samples. Kim et al. (2019) employed Conditional Generative Adversarial Networks (CGAN) to generate SDO/AIA 304 Å images, which enhanced the sample diversity and prediction performance of solar magnetograms.

At the algorithmic level, the model's performance on imbalanced data can be improved by adjusting thresholds, modifying loss functions, and employing ensemble learning. Sun et al. (2022) utilized the focal loss function to control the attention given to minority class samples by adjusting the parameter γ. The experiments showed that the model achieved the highest TSS score when $\gamma = 2$. Deshmukh et al. (2022) applied a weighted loss function to compensate for sample imbalance, avoiding the disruption of the original distribution caused by data resampling, thereby enhancing the model's ability to learn from the minority class. Additionally, Zhang et al. (2024) employed ensemble learning to further mitigate the imbalance problem by dividing negative samples into multiple subsets, combining them with positive samples to train independent classifiers, and generating the final model using a fusion strategy.

Instead of concentrating only on overall classification accuracy, assessment measures must be specifically created to evaluate the model's capacity to forecast minority class samples. Pandey et al. (2023a) introduced two widely used metrics for space weather prediction: TSS and HSS. These measures compare the predictive power of the model with the baseline performance of a random model, in addition to taking into account the model's performance on both positive and negative classes. This makes them more effective in addressing the class imbalance problem than conventional accuracy measures. Also, by analyzing Table 2, we find that the most used indicators for class imbalance are TSS, HSS, BACC, BSS, and BS. Among them, TSS, HSS, and BACC are the most commonly used and directly suitable metrics, as they effectively balance the performance of positive and negative classes and prevent the model from being biased toward the majority class. In contrast, BSS and BS are mainly for probabilistic prediction models, which are suitable for those cases where forecast confidence and probability values need to be taken into account, such as spatial weather prediction, meteorological prediction, and so on.

However, currently, no single method performs optimally in all aspects (Wan et al. 2021). In research, the most suitable methods should be comprehensively evaluated and selected based on the data characteristics, distribution, and research objectives, in order to effectively address the class imbalance problem.

4.1.3. Other Issues

In addition to the feature redundancy issue mentioned in Section 4.1.1, solar flare data also suffer from redundancy in the coding system. To facilitate the systematic observation of sunspots, the scientific community has developed a standardized coding system that identifies sunspots based on features such as size, shape, and position, requiring confirmation from two independent observation stations.

However, since different observers and instruments may record the same sunspot region multiple times, this inevitably leads to the issue of coding redundancy. This redundancy not only increases the complexity of data management, but also impacts the efficiency and accuracy of machine learning models in predicting sunspot regions. Specifically, redundant coding disrupts the model's ability to identify genuine sunspot regions. The accumulation of redundant data raises the data set's complexity, thereby increasing the computational burden.

Although the International Astronomical Union (IAU) revised the coding system in 2011, introducing standardized metrics such as the "relative sunspot number" and the "international sunspot number," the redundancy issue remains unresolved. Therefore, future research could consider evaluating the correlation between sunspot regions through cross-validation and removing highly correlated redundant samples, thereby reducing the impact of encoding redundancy on model predictions.

In addition to the inter-class imbalance issue in the class imbalance problem, there is also a certain amount of intra-class imbalance in the data. Intra-class imbalance refers to the uneven distribution of samples within the same class, which may lead to insufficient learning of certain sub-class features by the model. As research progresses, future studies could further investigate the impact of intra-class imbalance on model performance and explore effective strategies for mitigating this issue, such as optimizing sampling strategies or introducing new loss functions, to enhance the model's adaptability to intra-class imbalance.

Ultimately, observational errors due to projection effects may hinder our ability to accurately identify energy accumulation processes and magnetic field instabilities in solar active regions, thereby limiting the accuracy of flare prediction. To mitigate this effect, many studies typically restrict the analysis to a central longitude range, with commonly used ranges such as $\pm 70°$, $\pm 68°$, $\pm 45°$, and $\pm 30°$. This approach reduces observational distortion caused by the solar surface's spherical geometry by limiting the viewing angle of the observed region. However, this method also has certain limitations, particularly in long-term solar activity monitoring. Specifically, restricting the central longitude range may overlook some active regions located on the far side, which may rotate to the center of the solar disk in the coming days and produce significant flare activity. Therefore, in future research, in addition to restricting longitude ranges, it is still necessary to use advanced technologies such as multi-point observation and stereo observation to address the shortcomings of the central longitude screening method and to enhance the overall prediction capability.

In conclusion, the performance of prediction models is also impacted by the problems of encoding redundancy, intra-class imbalance, and projection effects in solar flare data. Future research also needs to explore strategies to address these three issues to optimize data quality and improve predictive accuracy of models.

4.2. Multimodality of Flare Samples

With the continuous advancement of solar observation technology, the variety of solar observation data has expanded, including images, time-series, spectral data, and image sequences. Among these data types, magnetograms and magnetic field features have become the two most commonly used data modalities due to their unique advantages in solar flare prediction. Magnetograms provide high-resolution spatial distribution information, revealing the detailed structure of solar active regions, while magnetic field features offer interpretable physical quantities, such as magnetic field strength and gradients, enhancing the physical interpretability of models. By combining these two data modalities, it is possible to leverage the spatial and temporal distribution advantages of magnetograms while extracting physical quantities closely related to flare eruptions through magnetic field features. This approach compensates for the limitations of using magnetograms alone, improving both prediction accuracy and generalization capability.

In Section 3.1.4, we systematically analyze two main multimodal fusion methods for solar flare prediction. The first method adopts a partitioned feature extraction architecture, where independent modules process heterogeneous data modalities, followed by feature fusion for final prediction. For instance, in the STCNet model proposed by Li et al. (2024), the magnetogram branch employs a Swin Transformer to capture high-dimensional spatiotemporal features, while the magnetic parameter branch utilizes CNN to encode local correlations, with features concatenated for downstream classification. Despite supporting inter-modal feature interaction, this approach faces limitations due to inter-modal heterogeneity-such as feature scale mismatches and distribution shifts-which restrict optimal fusion strategy design. The second method leverages model ensemble frameworks (e.g., Sun et al. 2022), where independently trained single-modality base models generate predictions fused via weighted aggregation or stacked generalization. While this strategy eliminates constraints on model selection, the isolation of base models inherently prevents cross-modal relationship modeling. Both approaches share a common limitation in flare prediction: they fail to effectively capture and integrate the deep interrelationships between features from different modalities, limiting their performance in more complex tasks.

To more effectively fuse multimodal features, future research should focus on introducing attention mechanisms or adopting large models to achieve deeper cross-modal feature interactions. First of all, attention mechanisms enhance the model's ability to model complex feature relationships by assigning different weights to each modality, thereby more accurately capturing correlations between modalities and effectively highlighting the most task-relevant features. To more effectively fuse multimodal features, future research should focus on introducing attention mechanisms or adopting large models to achieve deeper cross-modal feature interactions. First, attention mechanisms enhance the model's ability to model complex feature relationships by assigning different weights to each modality, thereby more accurately capturing correlations between modalities and effectively highlighting the most

task-relevant features. Second, large models, due to their powerful cross-modal representation learning capabilities, can adaptively adjust weights across multiple modalities, facilitating efficient fusion of information through end-to-end training. This adaptive learning approach allows large-scale models to automatically capture and integrate intermodal associations, thus improving the effectiveness of multimodal data processing. In summary, attention mechanisms and large-scale models complement each other: the former strengthens the correlation modeling of features, while the latter improves overall learning across modalities. Combining both is expected to yield better performance in future multimodal learning tasks.

Moreover, although magnetograms and magnetic field features are the most commonly used modalities, the usefulness of spectral data in solar flare prediction has also been shown in recent studies. Spectral data offer complementary observational information, particularly showing improved performance in the early prediction of flare activity. Future research can explore how to integrate spectral data with other modalities to further enhance the predictive performance of models.

4.3. Model Interpretability

Machine learning, particularly neural networks, in the field of solar flare prediction, has shown significant advantages over traditional physical models, particularly in capturing the complex nonlinear relationships between flares and predictors. As a result, neural networks have become the mainstream approach in this field. However, the "black-box" nature of neural networks limits their interpretability due to the lack of transparency in their decision-making process, which presents a major challenge in this area of research. Interpretability refers to the ability to explain the model's internal structure and decision-making processes in a manner that is understandable to humans, which is crucial for enhancing the model's transparency, trustworthiness, and practical applicability.

First, gradient-based interpretability methods are among the most commonly used approaches. Zheng et al. (2021) conducted the first "black-box" analysis of CNN models, employing feature map visualization techniques to reveal the models' prediction mechanisms. The visualizations showed that the model primarily focuses on regions with strong gradients, high intensity, and large total strength in the high-level feature maps when making decisions. Similarly, Yi et al. (2021) used guided backpropagation and Gradient-weighted Class Activation Mapping (Grad-CAM) methods to analyze CNN models, particularly with respect to the influence of SHARP parameters. The research found that these methods validated the model's rationale, and the feature regions that the model focused on were highly consistent with actual observational results, thereby strengthening the connection between the model and the flare activities observed in solar active regions.

In addition, attribution methods based on feature contributions have been widely applied. Pandey et al. (2023a) employed three methods: Guided Grad-CAM, Integral Gradients, and Deep SHapley Additive exPlanations (Deep SHAP), to analyze the prediction results of CNN models. Research has shown that Guided Grad-CAM provides spatially localized interpretability, while Integrated Gradients is suitable for precise analysis of the contribution of individual features, and Deep SHAP effectively handles the complex interactions between features. Alshammari et al. (2023) applied t-SNE to reduce the dimensionality of multi-dimensional time-series features extracted from the final layer of a Transformer model. This reduction facilitates the visualization and analysis of higher-dimensional features, offering deeper insights into the roles of different features and their interactions in the prediction process.

Occlusion-based attribution methods, such as the occlusion map method proposed by Bhattacharjee et al. (2020), are also commonly used interpretability techniques. This method provides intuitive interpretability by gradually occluding local regions of the input image and analyzing the resulting output. However, while the method is easy to understand, it incurs significant computational costs and can only offer local explanations, thus failing to fully reveal the model's global decision-making process.

Sun et al. (2022) conducted a comprehensive analysis of CNN models by combining multiple attribution methods, including the occlusion method, integral gradients, Deep Learning Important FeaTures (DeepLIFT), deconvolution, guided backpropagation, and Grad-CAM. While each of these methods has its advantages, these methods also exhibit certain limitations. For instance, Grad-CAM performs less effectively on high-resolution feature maps, and guided backpropagation and deconvolution show lower sensitivity to network decisions. By employing a combination of these methods, a more thorough understanding of the model's decision-making process and feature selection can be achieved.

Table 4 summarizes the principles, advantages, disadvantages, and applicability of several commonly used interpretability methods. The interpretability of CNN architectures is relatively straightforward due to their spatial feature localization capability (e.g., Grad-CAM highlighting discriminative regions in images). In contrast, RNN and LSTM, designed to model temporal dependencies, present significant challenges. Methods like LIME, SHAP, and Occlusion Sensitivity, while applicable to any model, face challenges when applied to sequential data. Perturbing input data (e.g., masking time steps in LIME) disrupts the recursive structure of RNN/LSTM, leading to unreliable importance estimates. SHAP's assumption of feature independence ignores temporal correlations, resulting in skewed values for time-series data. Traditional visualization methods like Grad-CAM also fail to capture the time-varying importance of hidden states in RNN. Adaptations

Table 4
Comparison of Model Interpretability Methods

Method	Core Principle	Advantage	Disadvantage	Scope of Application
LIME	Approximates local behavior by perturbing inputs and fitting a surrogate model.	Model-agnostic, intuitive for single predictions.	Sensitive to perturbation parameters; explanations may lack global consistency.	Any model
SHAP	Computes Shapley values from cooperative game theory to attribute feature impact.	Rigorous theory, provides global/local explanations.	High computational costs, sensitive to feature correlations.	Any model
DeepLIFT	Compares activation differences between input and reference via backpropagation.	Handles nonlinearities; avoids gradient saturation.	Sensitive to reference values, limited explanation for complex models.	Deep learning model
Integrated Gradients	Integrates gradients along a path from baseline to input.	Satisfies completeness axiom; no dependence on perturbations.	Computationally intensive; sensitive to baseline selection.	Deep learning model
Feature Map Visualization	Direct visualization of intermediate layer activations.	Visually demonstrate the feature extraction process.	Does not explain decision logic; requires manual interpretation.	CNN model
Grad-CAM	Weighted combination of convolutional feature maps by gradient importance.	Highlights class-discriminative regions; works with any CNN.	Limited to convolutional layers; low resolution for fine-grained details.	CNN model
Guided Grad-CAM	Combines Grad-CAM with Guided Backpropagation for pixel-level details.	High-resolution visualizations; class-specific explanations.	Computationally expensive; requires guided backprop implementation.	CNN model
Occlusion Sensitivity	Measures prediction changes by occluding input regions.	Intuitive; no model architecture dependence.	High computational complexity.	Any model

Note. Deep SHAP is a variant of SHAP for deep neural networks, designed to explain feature importance in models with complex architectures. Since it is conceptually based on SHAP, it has been excluded from the table to avoid redundancy. "Any model" refers to machine learning models capable of generating predictions from input data, including both deep learning architectures (e.g., CNN, RNN, LSTM) and traditional models (e.g., SVM, RF). These methods apply to supervised and unsupervised tasks and support diverse data types (e.g., image, text, time-series), though task-specific adaptations may be required.

like Temporal Gradient-weighted Class Activation Mapping (T-Grad-CAM) struggle to address long-range dependencies. Future research should consider integrating attention mechanisms to highlight time-step dependencies or using time-series masking techniques to assess the contribution of different sequence segments. Despite advancements in interpretability, challenges remain, particularly for tasks like solar flare prediction. Combining multiple attribution methods could enhance both precision and efficiency, thus facilitating the practical application of neural networks in such domains.

4.4. Section Summary

In summary, this discussion highlights several common issues in flare prediction: feature redundancy, class imbalance, data multimodality, and model interpretability. In addition, some potential problems: redundancy in the sunspot encoding system, inter-class imbalance, and projection effect problems.

In the issue of feature redundancy, current feature selection research mainly focuses on evaluating the discriminative power of features, particularly their ability to distinguish between classes in classification tasks. However, there are still shortcomings in the methods for assessing feature correlations. Existing standards for evaluating feature correlations lack consistency and struggle to quantify the interactions and dependencies between different features. This limits the comprehensiveness of feature selection, preventing the full exploration of potential interactions between features, which leads to the existence of redundancy issues. In order to better uncover the underlying links between features, future research should concentrate on creating a more efficient quantitative framework for assessing feature significance.

In the class imbalance problem, although many studies have applied resampling techniques to address it, these methods may introduce additional prediction errors by altering the data distribution, thereby affecting model performance. In particular, oversampling and undersampling can lead to overfitting or underfitting, making it important for future research to focus on classifier-based balancing methods, such as cost-sensitive learning or ensemble techniques, to more effectively mitigate the impact of class imbalance on model performance.

In addition, there are issues of intra-class imbalance and encoding system redundancy in solar flare data. Therefore, future research should focus on improving the model's predictive performance in the presence of intra-class imbalance (Han et al. 2023), while also exploring methods to assess the relevance of sunspot regions, remove redundant information, and address the issue of coding redundancy.

In terms of integrating multimodal data, although some studies have attempted to fuse flare observation data features, most methods are limited to simple feature concatenation or ensemble learning, and have not deeply explored the potential relationships between different modalities (e.g., spatial and temporal features). Future research should focus on developing more sophisticated fusion techniques, such as incorporating attention mechanisms or large models, to better capture the dependencies between modalities and thus enhance the model's overall predictive capability.

Although deep learning techniques excel in handling complex nonlinear relationships compared to traditional physical models, their "black-box" nature remains a significant limitation, resulting in a lack of transparency in the model's inference process. To address this issue, attribution methods have emerged as key techniques for enhancing model interpretability. By quantifying the contribution of features to prediction outcomes, attribution methods provide insight into the decision-making process of the model, thereby improving its transparency. However, the inference mechanisms of deep learning models still struggle to align with the physical meanings of traditional models. To address this, incorporating physical information as prior knowledge into neural networks has become an effective approach (Jiao et al. 2024). This not only helps capture complex nonlinear relationships but also provides a certain level of interpretability, enhancing the model's credibility in practical applications.

5. Conclusion and Outlook

Currently, solar flare prediction primarily relies on and magnetic field data from SOHO and SDO, particularly LoS magnetograms and vector magnetograms. However, due to the high dimensionality and redundancy of the data, existing prediction models often struggle to extract optimal results from individual features. Therefore, future research should prioritize optimizing predictors, reducing redundant information, and improving the efficiency of feature selection to enhance model accuracy and robustness. With continuous advancements in observational technologies, new data sources are expected to provide higher resolution and more comprehensive solar activity data, further enhancing flare prediction capabilities (Kontogiannis 2023). Moreover, previous studies have shown that solar flare eruptions are closely linked to CME and solar SEP events. Future research should consider incorporating CME and SEP event data as additional predictors to refine and strengthen the solar activity prediction framework (Han et al. 2023).

Methods for predicting solar flares may be broadly categorized into two groups: physical models and data-driven models. In the case of physical models, MHD and DAI are considered to be the primary drivers of flare eruptions, leading to the development of prediction methods based on these mechanisms. Additionally, SOC models are widely used due to their ability to capture the global statistical behavior of flares. However, physical models still face certain uncertainties and have not been widely adopted due to their complex physical assumptions. Therefore, further investigation into the detailed physical mechanisms underlying flare eruptions is necessary.

Fusion models provide notable benefits over conventional statistical methods, machine learning, and deep learning approaches in data-driven models as the number and complexity of data rise. Fusion models are not only capable of handling multimodal data but also integrate the strengths of different algorithms, such as combining CNN with RNN and incorporating attention mechanisms, thereby enhancing the model's robustness and predictive accuracy. However, current fusion algorithms are still limited, with some suffering from incompatibility issues, highlighting the need for further optimization.

Currently, the main challenges in flare prediction include data complexity, multimodality, and model interpretability. Although there has been some progress in addressing issues such as feature redundancy and class imbalance, limitations still remain. The multimodal data issue has been partially addressed through feature and decision fusion, but achieving deep fusion between different features remains challenging. Additionally, while neural networks show strong performance in prediction tasks, their "black-box" nature limits the interpretability of the inference process, making it difficult to provide transparency regarding the prediction results.

In summary, future research should focus on the following aspects: first, a deeper exploration of the physical mechanisms of solar flare eruptions to provide stronger theoretical support for models; second, optimizing the fusion of different algorithms to improve the predictive performance of fusion models; further evaluating the correlations among features and reducing data redundancy to improve the efficiency of feature selection; simultaneously, developing more effective methods for handling class imbalance, particularly addressing intra-class imbalance issues; exploring more complex deep fusion methods, such as introducing attention mechanisms, to better capture dependencies between multimodal data; finally, exploring more effective methods to improve the interpretability and predictive performance of neural networks, such as incorporating physical information as prior knowledge, can further enhance the accuracy and robustness of solar flare prediction.

Acknowledgments

This work was supported by the National Key Research and Development Program of China (grant No. 2022YFF0503600) and the National Natural Science Foundation of China (NSFC, grant No. 42130202). This work uses GONG data from the NSO Integrated Synoptic Program, managed by the National Solar Observatory under AURA and NSF, with contributions from NOAA. The GONG instruments are hosted by various observatories, including Big Bear, High Altitude, Learmonth, Udaipur, Instituto de Astrofísica de Canarias, and Cerro Tololo. Additionally, we acknowledge the SDO team for providing high-resolution solar data from the AIA and HMI instruments.

References

Abduallah, Y., Wang, J. T., Nie, Y., Liu, C., & Wang, H. 2021, RAA, 21, 160
Abduallah, Y., Wang, J. T., Wang, H., & Xu, Y. 2023, NatSR, 13, 13665
Abramenko, V., Yurchyshyn, V., Wang, H., Spirock, T., & Goode, P. 2002, ApJ, 577, 487
Aegerter, T. R. W., Emmons, D. J., & Loper, R. D. 2020, JASTP, 208, 105375
Ahmed, O. W., Qahwaji, R., Colak, T., et al. 2013, SoPh, 283, 157
Al-Ghraibah, A., Boucheron, L., & McAteer, R. 2015, A&A, 579, A64
Alipour, N., Mohammadi, F., & Safari, H. 2019, ApJS, 243, 20
Alshammari, K., Saini, K., Hamdi, S. M., & Boubrahimi, S. F. 2023, in 2023 Int. Conf. on Machine Learning and Applications (ICMLA), ed. M. A. Wani et al. (Piscataway, NJ: IEEE), 558
Anastasiadis, A., Papaioannou, A., Sandberg, I., et al. 2017, SoPh, 292, 1
Asaly, S., Gottlieb, L.-A., & Reuveni, Y. 2020, IJSTA, 14, 1469
Aschwanden, M. J., & Aschwanden, P. D. 2008a, ApJ, 674, 530
Aschwanden, M. J., & Aschwanden, P. D. 2008b, ApJ, 674, 544
Aschwanden, M. J., Crosby, N. B., Dimitropoulou, M., et al. 2016, SSRv, 198, 47
Bak, P., Tang, C., & Wiesenfeld, K. 1987, PhRvL, 59, 381
Baranyi, T., Győri, L., Ludmány, A., et al. 2016, SoPh, 291, 3081
Barnes, G., & Leka, K. 2006, ApJ, 646, 1303
Barnes, G., Leka, K., Schumer, E., & Della-Rose, D. 2007, SpWea, 5, S09002
Bélanger, E., Vincent, A., & Charbonneau, P. 2007, SoPh, 245, 141
Benvenuto, F., Piana, M., Campi, C., & Massone, A. M. 2018, ApJ, 853, 90
Bhattacharjee, S., Alshehhi, R., Dhuri, D. B., & Hanasoge, S. M. 2020, ApJ, 898, 98
Bloomfield, D. S., Higgins, P. A., McAteer, R. J., & Gallagher, P. T. 2012, ApJL, 747, L41
Bobra, M. G., & Couvidat, S. 2015, ApJ, 798, 135
Bobra, M. G., Sun, X., Hoeksema, J. T., et al. 2014, SoPh, 289, 3549
Bobra, M. G., Wright, P. J., Sun, X., & Turmon, M. J. 2021, ApJS, 256, 26
Campi, C., Benvenuto, F., Massone, A. M., et al. 2019, ApJ, 883, 150
Chen, B., Shen, C., Gary, D. E., et al. 2020, NatAs, 4, 1140
Chen, J., Li, W., Li, S., et al. 2022, SpScT, 2022, 9761567
Chen, Y., Manchester, W., Jin, M., & Pevtsov, A. 2024, Statistics and Data Science in Imaging, 1, 2391688
Chen, Y., Manchester, W. B., Hero, A. O., et al. 2019, SpWea, 17, 1404
Cicogna, D., Berrilli, F., Calchetti, D., et al. 2021, ApJ, 915, 38
Cinto, T., Gradvohl, A., Coelho, G., & Da Silva, A. 2020a, SoPh, 295, 93
Cinto, T., Gradvohl, A. L. S., Coelho, G. P., & da Silva, A. E. A. 2020b, MNRAS, 495, 3332
Colak, T., & Qahwaji, R. 2009, SpWea, 7, S06001
Conlon, P., Gallagher, P., McAteer, R., et al. 2009, in Solar Image Analysis and Visualization, ed. J. Ireland & C. A. Young (New York, NY: Springer), 87
Cortie, A. 1901, ApJ, 13, 260
Cui, Y., Li, R., Wang, H., & He, H. 2007, SoPh, 242, 1
Cui, Y., Li, R., Zhang, L., He, Y., & Wang, H. 2006, SoPh, 237, 45
Curto, J. J. 2020, JSWSC, 10, 27
Deng, H., Zhong, Y., Chen, H., et al. 2023, ApJS, 268, 43
Deng, Z., Wang, F., Deng, H., et al. 2021, ApJ, 922, 232
Deshmukh, V., Berger, T., Meiss, J., & Bradley, E. 2021, in Proceedings of the AAAI Conference on Artificial Intelligence, ed. Qiang Yang, Kevin Leyton-Brown, & Mausam (Menlo Park, CA: AAAI), 15293
Deshmukh, V., Flyer, N., Van der Sande, K., & Berger, T. 2022, ApJS, 260, 9
Dissauer, K., Leka, K., & Wagner, E. L. 2023, ApJ, 942, 83
Domijan, K., Bloomfield, D. S., & Pitié, F. 2019, SoPh, 294, 6
Domingo, V., Fleck, B., & Poland, A. I. 1995, SoPh, 162, 1
Eren, S., Kilcik, A., Atay, T., et al. 2016, MNRAS, 465, stw2742
Falconer, D. 2001, JGRA, 106, 25185
Falconer, D. A., Moore, R. L., Barghouty, A. F., & Khazanov, I. 2014, SpWea, 12, 306
Falconer, D., Barghouty, A. F., Khazanov, I., & Moore, R. 2011, SpWea, 9, S04003
Florios, K., Kontogiannis, I., Park, S.-H., et al. 2018, SoPh, 293, 28
Gallagher, P. T., Moon, Y.-J., & Wang, H. 2002, SoPh, 209, 171
Georgoulis, M. K. 2005, SoPh, 228, 5
Georgoulis, M. K., & LaBonte, B. J. 2007, ApJ, 671, 1034
Georgoulis, M. K., & Rust, D. M. 2007, ApJ, 661, L109
Georgoulis, M. K., Tziotziou, K., & Raouafi, N.-E. 2012, ApJ, 759, 1
Georgoulis, M. K., Yardley, S. L., Guerra, J. A., et al. 2024, AdSpR, In Press
Gheibi, A., Safari, H., & Javaherian, M. 2017, ApJ, 847, 115
Giorgi, F., Ermolli, I., Romano, P., et al. 2015, SoPh, 290, 507
Gontikakis, C., Kontogiannis, I., Georgoulis, M., et al. 2020, arXiv:2011.06433
Grim, L. F. L., & Gradvohl, A. L. S. 2024, SoPh, 299, 33
Guastavino, S., Marchetti, F., Benvenuto, F., Campi, C., & Piana, M. 2022, A&A, 662, A105
Guastavino, S., Marchetti, F., Benvenuto, F., Campi, C., & Piana, M. 2023, FrASS, 9, 1039805
Guennou, C., Pariat, E., Leake, J. E., & Vilmer, N. 2017, JSWSC, 7, A17
Guerra, J. A., Park, S.-H., Gallagher, P. T., et al. 2018, SoPh, 293, 9
Guerra, J. A., Pulkkinen, A., & Uritsky, V. M. 2015, SpWea, 13, 626
Guo, Y., Guo, J., Ni, Y., et al. 2024, RvMPP, 8, 29
Hale, G. E., Ellerman, F., Nicholson, S. B., & Joy, A. H. 1919, ApJ, 49, 153
Hamon, D., Nicodemi, M., & Jensen, H. 2002, A&A, 387, 326
Han, H., Wang, W.-Y., & Mao, B.-H. 2005, Int. Conf. on Intelligent Computing (Berlin: Springer), 878
Han, K., Yu, M.-Y., Fu, J.-F., et al. 2023, RAA, 23, 065002
Hao, N., Flagg, L., & Jayawardhana, R. 2024, ApJ, 973, 109
Hathaway, D. H. 2015, LRSP, 12, 4
Hayes, L. A., Gallagher, P. T., Dennis, B. R., et al. 2019, ApJ, 875, 33
He, H., & Garcia, E. A. 2009, IEEE Transactions on Knowledge and Data Engineering, 21, 1263
Hood, A. W., & Priest, E. 1979, SoPh, 64, 303
Huang, X., & Wang, H.-N. 2013, RAA, 13, 351
Huang, X., Wang, H., Xu, L., et al. 2018, ApJ, 856, 7
Huang, X., Yu, D., Hu, Q., Wang, H., & Cui, Y. 2010, SoPh, 263, 175
Huang, X., Zhang, L., Wang, H., & Li, L. 2013, A&A, 549, A127
Huang, X., Zhao, Z., Zhong, Y., et al. 2024, ScChD, 67, 3727
Ishiguro, N., & Kusano, K. 2017, ApJ, 843, 101
Ji, A., Wen, J., Angryk, R., & Aydin, B. 2022, in 2022 26th Int. Conf. on Pattern Recognition (ICPR) *(Montreal, QC, Canada)* (Piscataway, NJ: IEEE), 2907
Jiao, L., Song, X., You, C., et al. 2024, Artificial Intelligence Review, 57, 256
Jiao, Z., Sun, H., Wang, X., et al. 2020, SpWea, 18, e2020SW002440
Kaneda, K., Wada, Y., Iida, T., et al. 2022, Lecture Notes in Computer Science in Proc. of the Asian Conf. on Computer Vision, ed. L. Wang et al. (Berlin: Springer), 1488
Kim, T., Park, E., Lee, H., et al. 2019, NatAs, 3, 397
Kliem, B., & Török, T. 2006, PhRvL, 96, 255002
Komm, R., & Hill, F. 2009, JGRA, 114, A06105
Kontogiannis, I. 2023, AdSpR, 71, 2017
Korsós, M. B., Baranyi, T., & Ludmány, A. 2014, ApJ, 789, 107
Korsós, M. B., Ludmány, A., Erdélyi, R., & Baranyi, T. 2015, ApJL, 802, L21
Korsós, M., Romano, P., Morgan, H., et al. 2020, ApJL, 897, L23
Koskinen, H., Eliasson, L., Holback, B., et al. 1999, SPEE Final Report (FMI: Helsinki)
Krizhevsky, A., Sutskever, I., & Hinton, G. E. 2012, in Advances in Neural Information Processing Systems *(Vancouver)* ed. F. Pereira et al. (New York, NY: Curran Associates, Inc.), 25

Kusano, K., Iju, T., Bamba, Y., & Inoue, S. 2020, Sci, 369, 587
Landa, V., & Reuveni, Y. 2022, ApJS, 258, 12
Lee, J., Moon, Y.-J., Jeong, H.-J., Yi, K., & Lee, H. 2024, ApJL, 971, L47
Lee, K., Moon, Y.-J., Lee, J.-Y., Lee, K.-S., & Na, H. 2012, SoPh, 281, 639
Leka, K., & Barnes, G. 2003a, ApJ, 595, 1277
Leka, K., & Barnes, G. 2003b, ApJ, 595, 1296
Leka, K., & Barnes, G. 2007, ApJ, 656, 1173
Leka, K., Dissauer, K., Barnes, G., & Wagner, E. L. 2023, ApJ, 942, 84
Lemen, J. R., Title, A. M., Akin, D. J., et al. 2012, SoPh, 275, 17
Li, M., Cui, Y., Luo, B., et al. 2022, SpWea, 20, e2021SW002985
Li, R., Wang, H.-N., He, H., Cui, Y.-M., & Du, Z.-L. 2007, ChJAA, 7, 441
Li, R., Wu, Y., Tian, Q., & Huang, X. 2024, SSPMA, 54, 129611
Li, R., & Zhu, J. 2013, RAA, 13, 1118
Li, X., Zheng, Y., Wang, X., & Wang, L. 2020, ApJ, 891, 10
Lim, D., Moon, Y.-J., Park, J., et al. 2019, arXiv:1907.11373
Liu, C., Deng, N., Wang, J. T., & Wang, H. 2017a, ApJ, 843, 104
Liu, H., Liu, C., Wang, J. T., & Wang, H. 2019, ApJ, 877, 121
Liu, J.-F., Li, F., Wan, J., & Yu, D.-R. 2017b, RAA, 17, 034
Liu, J.-F., Li, F., Zhang, H.-P., & Yu, D.-R. 2017c, RAA, 17, 116
Lu, E. T., & Hamilton, R. J. 1991, ApJL, 380, L89
Mason, J. P., & Hoeksema, J. 2010, ApJ, 723, 634
Massa, P., & Emslie, A. G. 2022, FrASS, 9, 1040099
McAteer, R. J., Gallagher, P. T., & Ireland, J. 2005, ApJ, 631, 628
McCloskey, A. E., Gallagher, P. T., & Bloomfield, D. S. 2018, JSWSC, 8, A34
McIntosh, P. S. 1990, SoPh, 125, 251
Morales, L., & Charbonneau, P. 2008, ApJ, 682, 654
Morales, L. F., & Santos, N. A. 2020, SoPh, 295, 155
Muranushi, T., Shibayama, T., Muranushi, Y. H., et al. 2015, SpWea, 13, 778
Murray, S. A., Bingham, S., Sharpe, M., & Jackson, D. R. 2017, SpWea, 15, 577
Nagem, T. A., Qahwaji, R. S., Ipson, S. S., Wang, Z., & Al-Waisy, A. S. 2018, Int. J. Adv. Computer Sci. App., 9, 492
Najafi, A., Darooneh, A. H., Gheibi, A., & Farhang, N. 2020, ApJ, 894, 66
Nishizuka, N., Sugiura, K., Kubo, Y., Den, M., & Ishii, M. 2018, ApJ, 858, 113
Nishizuka, N., Sugiura, K., Kubo, Y., et al. 2017, ApJ, 835, 156
Olami, Z., Feder, H. J. S., & Christensen, K. 1992, PhRvL, 68, 1244
Oloketuyi, J., Liu, Y., & Elmhamdi, A. 2023, NewA, 100, 101972
Osokin, A., Podlazov, A., Chernetsky, V., & Livshits, M. 2004, Proc. IAUS, 2004, 477
Pandey, C., Angryk, R. A., & Aydin, B. 2023a, Joint European Conf. on Machine Learning and Knowledge Discovery in Databases (Berlin: Springer), 72
Pandey, C., Ji, A., Angryk, R. A., & Aydin, B. 2023b, in 2023 IEEE Sixth Int. Conf. on Artificial Intelligence and Knowledge Engineering (AIKE) *(Laguna Hills, CA, USA)* ed. D. Ostrowski et al. (Piscataway, NJ: IEEE), 83
Pandey, C., Ji, A., Angryk, R. A., Georgoulis, M. K., & Aydin, B. 2022, FrASS, 9, 897301
Panos, B., & Kleint, L. 2020, ApJ, 891, 17
Park, E., Moon, Y.-J., Shin, S., et al. 2018, ApJ, 869, 91
Park, S.-H., Guerra, J. A., Gallagher, P. T., Georgoulis, M. K., & Bloomfield, D. S. 2018b, SoPh, 293, 114
Pelkum Donahue, K., & Inceoglu, F. 2024, FrASS, 10, 1298609
Pesnell, W. D., Thompson, B. J., & Chamberlin, P. 2012, The Solar Dynamics Observatory (SDO) (Berlin: Springer)
Platts, J., Reale, M., Marsh, J., & Urban, C. 2022, JAnSc, 69, 1421
Qahwaji, R., & Colak, T. 2007, SoPh, 241, 195
Raboonik, A., Safari, H., Alipour, N., & Wheatland, M. S. 2016, ApJ, 834, 11
Reames, D. V. 2013, SSRv, 175, 53
Reep, J. W., & Barnes, W. T. 2021, SpWea, 19, e2021SW002754
Ribeiro, F., & Gradvohl, A. L. S. 2021, A&C, 35, 100468
Sadykov, V. M., & Kosovichev, A. G. 2017, ApJ, 849, 148
Sammis, I., Tang, F., & Zirin, H. 2000, ApJ, 540, 583
Scherrer, P. H., Bogart, R. S., Bush, R., et al. 1995, in The Soho Mission, ed. B. Fleck, V. Domingo, & A. Poland (Berlin: Springer), 129
Schou, J., Scherrer, P. H., Bush, R. I., et al. 2012, SoPh, 275, 229
Schrijver, C. J. 2007, ApJ, 655, L117
Shibata, K., & Magara, T. 2011, LRSP, 8, 1
Silva, G. F., Simões, P. J., Osborne, C., et al. 2024, BoSAB, 35, 95
Sinha, S., Gupta, O., Singh, V., et al. 2022, ApJ, 935, 45
Siscoe, G. 2000, JASTP, 62, 1223
Song, H., Tan, C., Jing, J., et al. 2009, SoPh, 254, 101
Stanislavsky, A., Burnecki, K., Magdziarz, M., Weron, A., & Weron, K. 2009, ApJ, 693, 1877
Stanislavsky, A., Nitka, W., Małek, M., Burnecki, K., & Janczura, J. 2020, JASTP, 208, 105407
Steward, G. A., Lobzin, V. V., Wilkinson, P. J., Cairns, I. H., & Robinson, P. A. 2011, SpWea, 9, S11004
Strugarek, A., & Charbonneau, P. 2014, SoPh, 289, 4137
Sun, D., Huang, X., Zhao, Z., & Xu, L. 2023, ApJS, 266, 8
Sun, H., Manchester IV, W., & Chen, Y. 2021, SpWea, 19, e2021SW002837
Sun, P., Dai, W., Ding, W., et al. 2022, ApJ, 941, 1
Sun, Z., Bobra, M. G., Wang, X., et al. 2022, ApJ, 931, 163
Tang, R., Liao, W., Chen, Z., et al. 2021, ApJS, 257, 50
Tenenbaum, J. B., Silva, V. d., & Langford, J. C. 2000, Sci, 290, 2319
Thibeault, C., Strugarek, A., Charbonneau, P., & Tremblay, B. 2022, SoPh, 297, 125
Tian, Y. 2022, J. Phys: Conf. Ser., 2282, 012025
Tlatov, A., Illarionov, E., Berezin, I., & Shramko, A. 2020, CosRe, 58, 444
Török, T., Kliem, B., & Titov, V. 2004, A&A, 413, L27
van der Sande, K., Muñoz-Jaramillo, A., & Chatterjee, S. 2023, ApJ, 955, 148
Velanki, Y., Hosseinzadeh, P., Boubrahimi, S. F., & Hamdi, S. M. 2024, Univ, 10, 373
Viet Do, B., Chen, Y., Nguyen, X., & Manchester, W. 2024, FrASS, 11, 1229092
Vysakh, P., & Mayank, P. 2023, SoPh, 298, 137
Wan, J., Fu, J.-F., Liu, J.-F., et al. 2021, RAA, 21, 237
Wan, J., Fu, J.-F., Tan, D.-M., et al. 2022, RAA, 22, 085020
Wan, J., Fu, J.-F., Wen, R.-Q., et al. 2023, RAA, 23, 065004
Wang, J., Liu, S., Ao, X., et al. 2019, ApJ, 884, 175
Wang, J., Zhang, Y., Webber, S. A. H., et al. 2020, ApJ, 892, 140
Wang, S., Zhou, R., Gu, X., et al. 2024, RemS, 16, 2092
Wang, X., Chen, Y., Toth, G., et al. 2020, ApJ, 895, 3
Wei, J., Zheng, Y., Li, X., et al. 2024, Ap&SS, 369, 48
Welsch, B. T., Li, Y., Schuck, P. W., & Fisher, G. H. 2009, ApJ, 705, 821
Wheatland, M. 2004, ApJ, 609, 1134
Yan, P., Li, X., Zheng, Y., et al. 2024, Ap&SS, 369, 110
Yi, K., Moon, Y.-J., Lim, D., Park, E., & Lee, H. 2021, ApJ, 910, 8
Yu, D., Huang, X., Wang, H., & Cui, Y. 2009, SoPh, 255, 91
Yuan, F., Lin, J., Deng, Y., Guo, J., & Wang, G. 2020, ChSBu, 61, 2316
Zhang, H., Li, Q., Yang, Y., et al. 2022, ApJS, 263, 28
Zhang, S., Zheng, Y., Li, X., et al. 2024, AdSpR, 74, 3467
Zhang, X., Xu, L., Li, Z., & Huang, X. 2024, ApJS, 274, 38
Zheng, Y., Li, X., Si, Y., Qin, W., & Tian, H. 2021, MNRAS, 507, 3519
Zheng, Y., Li, X., & Wang, X. 2019, ApJ, 885, 73
Zheng, Y., Li, X., Yan, S., et al. 2023a, MNRAS, 521, 5384
Zheng, Y., Qin, W., Li, X., et al. 2023b, Ap&SS, 368, 53