The Astrophysical Journal Supplement Series, 287:49 (17pp), 2026 December https://doi.org/10.3847/1538-4365/aeb2ff
© 2026. The Author(s). Published by the American Astronomical Society.
OPEN ACCESS

# Physical Predictors of Halo Coronal Mass Ejection Formation Identified via Explainable Machine Learning: Linking Active Region Evolution to Solar Flares and Type II Radio Bursts

Abd-ur Raheem[1], Huseyin Cavus[2], Jason T. L. Wang[3,4], Gani Caglar Coban[1], and Haimin Wang[4,5,6]
[1] School of Graduate Studies, Canakkale Onsekiz Mart University, 17110 Canakkale, Türkiye; arawan@stu.comu.edu.tr
[2] Department of Physics, Canakkale Onsekiz Mart University, 17110 Canakkale, Türkiye
[3] Department of Computer Science, New Jersey Institute of Technology, University Heights, Newark, NJ 07102, USA
[4] Institute for Space Weather Sciences, New Jersey Institute of Technology, University Heights, Newark, NJ 07102, USA
[5] Center for Solar-Terrestrial Research, New Jersey Institute of Technology, University Heights, Newark, NJ 07102, USA
[6] Big Bear Solar Observatory, New Jersey Institute of Technology, 40386, North Shore Lane, Big Bear City, CA 92314, USA
*Received 2026 June 29; revised 2026 September 22; accepted 2026 October 9; published 2026 November 18*

## Abstract

Solar flares, coronal mass ejections (CMEs), and solar radio bursts are coupled drivers of the most energetic space-weather events, with halo CMEs of particular operational importance for Earth-directed CMEs and intense Type II radio burst activity. This study applies explainable AI, SHapley Additive exPlanations (SHAP) analysis of a random forest classifier, to identify the photospheric magnetic field parameters that most strongly govern whether a CME accompanying a flare-associated Type II burst attains halo morphology. Time-series features are derived from Solar Dynamics Observatory/Helioseismic and Magnetic Imager (SDO/HMI) Space-weather HMI Active Region Patches (SHARP) data spanning up to 24 hr before Type II burst onset. To address nonhalo underrepresentation, three augmentation methods (scaling, jitter, and window slicing) are applied under a leakage-free protocol, and classifiers are trained across 48 dataset variants spanning 24 observation windows. Augmentation improves the F1 score and true skill statistic (TSS) in all 48 experiments, with a representative model (11 hr window, normalized features) achieving F1 = 0.893 and TSS = 0.781. Across 100 repeated train/test splits the discriminative ability is stable (ROC-AUC = 0.987 ± 0.013), with threshold-dependent metrics showing wider spread at this sample size (mean F1 = 0.761). SHAP analysis identifies current helicity (MEANJZH), the twist parameter (MEANALP), and the rate of change of magnetic shear and inclination (MEANSHR, MEANGAM) as the leading predictors, stable across all folds. These quantities trace the build-up of a twisted, current-carrying flux rope toward the Kruskal–Shafranov kink-instability threshold, linking active-region evolution, flare energetics, and the strength of the CME-driven shock that excites the Type II emission.

*Unified Astronomy Thesaurus concepts:* Solar coronal mass ejections (310); Solar flares (1496); Solar active regions (1974); Radio bursts (1339); Random Forests (1935)

## 1. Introduction

Solar flares are intense, transient releases of electromagnetic energy originating from the solar atmosphere and driven by magnetic reconnection processes in magnetically complex active regions (C. Liu et al. 2017). These events are intimately coupled to coronal mass ejections (CMEs), which are large-scale expulsions of magnetized plasma into the heliosphere, and to solar radio bursts, which are generated by the nonthermal particle beams and shock fronts that accompany and follow these eruptions (H. V. Cane et al. 2005). Together, these phenomena constitute the primary drivers of space weather: geomagnetic storms, satellite drag enhancement, radio communication blackouts, disruption of navigation systems, and radiation hazards to astronauts and high-altitude aviation (C. Liu et al. 2017; Y. Abduallah et al. 2022; H. Cavus et al. 2025). Accurate, physically grounded prediction of these events is therefore among the highest-priority goals of operational space-weather science.

The physical connections among solar flares, CMEs, and radio bursts are well established. Flares and CMEs are thought to share a common origin in the free magnetic energy stored in nonpotential field configurations above solar active regions. The disruption and reconnection of large-scale magnetic structures simultaneously releases impulsive flare emission and drives a CME. As the CME-driven shock propagates outward through the corona and interplanetary medium, it excites plasma oscillations that radiate as Type II solar radio bursts detectable from space-based receivers across the decametric–hectometric frequency range (H. V. Cane et al. 2005; H. Cavus 2015). The magnetohydrodynamic processes governing the storage and release of this free energy—from the toroidal field at the base of the convection zone to the shear-flow instabilities that develop in erupting structures—have been studied extensively through MHD modeling (H. Cavus & A. I. Karafistan 2007; G. A. Hoshoudy & H. Cavus 2018). Consequently, the occurrence and properties of solar radio bursts are directly coupled to the characteristics of the originating CME and flare, making halo CME prediction a natural precursor to radio burst forecasting.

Among CME types, halo CMEs are of particular scientific and operational significance. Defined by a projected leading edge that appears to encircle the full solar disk in coronagraph imagery, halo CMEs indicate propagation largely along the Sun–Earth line, making them the primary contributors to intense geomagnetic storms and the strongest

Original content from this work may be used under the terms of the Creative Commons Attribution 4.0 licence. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

drivers of Earth-directed energetic particle events (N. Gopalswamy et al. 2001; L. Lu et al. 2017; K. A. Alobaid et al. 2024; H. Cavus et al. 2025). Statistical studies have shown that halo CMEs are also associated with most energetic radio burst activity. The strongest Type II radio bursts detected in the decametric–hectometric range are produced by fast, wide-angle CMEs that overwhelmingly fall into the halo or partial-halo category (N. Gopalswamy et al. 2005). Predicting halo CME occurrence from pre-eruption photospheric magnetic field observations therefore simultaneously addresses the prediction of geomagnetic disturbances and provides an early indicator of radio burst activity.

Data-driven machine learning approaches to solar eruption prediction have advanced considerably over the past decade (R. Qahwaji et al. 2008; M. G. Bobra & S. Couvidat 2015; C. Liu et al. 2017; H. Liu et al. 2020; H. Cavus et al. 2025). The availability of the Solar Dynamics Observatory/Helioseismic and Magnetic Imager (SDO/HMI) Space-weather HMI Active Region Patches (SHARP) data product (M. G. Bobra et al. 2014), which provides continuous, time-resolved vector magnetogram potential and nonpotential parameters for individual active regions, has enabled the development of time-series models that incorporate pre-event magnetic field evolution. Random forest (RF) classifiers trained on SHARP-derived features have demonstrated robust predictive skill for solar flares (M. G. Bobra & S. Couvidat 2015; G. Barnes et al. 2016; K. Florios et al. 2018) and more recently have been extended to CME property prediction. Deep learning architectures including convolutional neural networks, LSTM-based hybrids, and residual networks have further improved flare detection performance (N. Nishizuka et al. 2018; H. Liu et al. 2019; A.-u. Raheem et al. 2021) and, more recently, have been extended to the prediction of geomagnetic indices such as SYM-H (Y. Abduallah et al. 2024), establishing a rich methodological foundation for solar eruption forecasting.

Despite these advances, two persistent challenges limit the operational usefulness and physical interpretability of these models. The first challenge is class imbalance. Halo CMEs, while being the subset of greatest operational concern, are less common in observational catalogs than standard CMEs, leading to an imbalanced binary classification problem. When trained on imbalanced class distributions, ML algorithms tend to favor the majority class, producing models that achieve high overall accuracy while failing to reliably detect the minority-class events (D. S. Bloomfield et al. 2012). Data-centric strategies including synthetic oversampling and time-series augmentation have been proposed to address this imbalance by generating additional minority-class training samples while preserving the class-specific temporal structure of the data (A. Ahmadzadeh et al. 2021; P. Li et al. 2025). P. Li et al. (2025) conducted a comprehensive evaluation of 12 augmentation methods for deep learning-based solar flare prediction, demonstrating that select strategies, particularly those that preserve local temporal patterns such as scaling (SC) and window slicing (WS), produce measurable gains in TSS and HSS across multiple neural network architectures. The present study adapts this augmentation framework to the halo CME classification problem using an RF classifier.

The second challenge is model interpretability. Ensemble ML classifiers are inherently opaque, which means they produce accurate predictions but offer limited direct insight into the physical mechanisms governing the decisions. In the context of solar physics, where prediction models are expected to not only forecast events but also to advance physical understanding, this limitation is significant (E. Camporeale 2019). Explainable AI (XAI) methods (e.g., H. Zhang et al. 2025), especially SHapley Additive exPlanations (SHAP; S. M. Lundberg et al. 2020) address this by decomposing individual predictions into signed, per-feature contributions grounded in cooperative game theory. SHAP Shapley values have a unique attribution method satisfying local accuracy, consistency, and missingness simultaneously. Applied to an RF trained on SHARP time-series features, SHAP analysis provides a ranked, physically interpretable measure of which magnetic field parameters most strongly drive halo CME classification decisions, offering a quantitative, data-driven perspective on the causal pathways linking photospheric field evolution to energetic solar eruptions and their downstream radio burst signatures (T. Williams et al. 2026).

This study addresses both challenges simultaneously. The problem this study tries to address is given that a flare-associated Type II burst occurred with a CME, was that CME a halo CME or not? The remainder of this paper is structured as follows. Section 2 describes the CME event catalog, SHARP feature extraction, and time-window construction. Section 3 presents the augmentation methodology, training protocol, and evaluation framework. Section 4 reports results. Section 5 discusses the physical implications of the SHAP findings.

## 2. Data

### 2.1. CME Event Catalog

The study uses a catalog of ∼194 CME events (ranging from 185 to 198 depending on the length of the time-series window; see Table 2) identified in STEREO and Wind WAVES Type II bursts and associated CMEs observed by the Solar and Heliospheric Observatory (SOHO), Solar Terrestrial Relations Observatory (STEREO), and SDO missions and matched to solar active regions continuously tracked in the SDO/HMI SHARP dataset (M. G. Bobra et al. 2014). The Type II burst events along with the CME events and their data are obtained from the CDAW Type II burst database (N. Gopalswamy et al. 2019). Each event is characterized by its projected morphology in the STEREO and WIND field of view and its associated SHARP active region record.

### 2.2. SHARP Magnetic Field Parameters

The SHARP data product (M. G. Bobra et al. 2014) provides near-continuous time-series measurements of 17 photospheric vector magnetic field summary parameters for solar active regions observed by SDO/HMI. These parameters characterize distinct aspects of the magnetic field structure that are physically linked to the free energy storage and release processes governing solar eruptions (M. G. Bobra & S. Couvidat 2015). Table 1 lists the full set of parameters used in this study.

Together, these parameters capture the total magnetic flux content, field complexity, electric current structure, helicity, free energy density, and field-line shear of an active region—all of which are theoretically and observationally linked to the probability and energetics of solar eruptions (M. G. Bobra & S. Couvidat 2015; G. Barnes et al. 2016; K. Florios et al. 2018). The magnetic field parameters such as TOTPOT and MEANSHR have been consistently identified as among the most predictive SHARP quantities in prior flare and CME

**Table 1**
The Helioseismic and Magnetic Imager (HMI) Magnetic Parameters Obtained from JSOC Used as Inputs in This Study

| Parameter | Formula | Description (Unit) |
|---|---|---|
| MEANPOT | $\bar{\rho} \propto \frac{1}{N}\Sigma(B^{\rm Obs} - B^{\rm Pot})^2$ | Mean photospheric excess magnetic energy density (erg cm$^{-3}$) |
| TOTPOT | $\rho_{\rm tot} \propto \Sigma(B^{\rm Obs} - B^{\rm Pot})^2\, {\rm d}A$ | Total photospheric magnetic free energy density (erg cm$^{-1}$) |
| USFLUX | $\Phi = \Sigma\lvert B_z\rvert\, dA$ | Total unsigned flux (Mx) |
| MEANGAM | $\bar{\gamma} \propto \frac{1}{N}\Sigma \arctan\left(\frac{B_h}{B_z}\right)$ | Mean inclination angle (deg) |
| MEANGBT | $\lvert\overline{\nabla B_{\rm tot}}\rvert = \frac{1}{N}\Sigma\sqrt{\left(\frac{\partial B}{\partial x}\right)^2 + \left(\frac{\partial B}{\partial y}\right)^2}$ | Mean gradient of total field (G Mm$^{-1}$) |
| MEANJZD | $\overline{J_z} \propto \frac{1}{N}\Sigma\left(\frac{\partial B_y}{\partial x} - \frac{\partial B_x}{\partial y}\right)$ | Mean vertical current density (mA m$^{-2}$) |
| TOTUSJZ | $J_{Z\ {\rm total}} = \Sigma\lvert J_z\rvert\, dA$ | Total unsigned vertical current (A) |
| MEANGBH | $\lvert\overline{\nabla B_h}\rvert = \frac{1}{N}\Sigma\sqrt{\left(\frac{\partial B_h}{\partial x}\right)^2 + \left(\frac{\partial B_h}{\partial y}\right)^2}$ | Mean gradient of horizontal field (G Mm$^{-1}$) |
| MEANGBZ | $\lvert\overline{\nabla B_z}\rvert = \frac{1}{N}\Sigma\sqrt{\left(\frac{\partial B_z}{\partial x}\right)^2 + \left(\frac{\partial B_z}{\partial y}\right)^2}$ | Mean value of the vertical field gradient (G Mm$^{-1}$) |
| MEANALP | $\alpha_{\rm total} \propto \frac{\Sigma J_z B_z}{\Sigma B_z^2}$ | Mean twist parameter, $\alpha$ (Mm$^{-1}$) |
| MEANJZH | $\overline{H_c} \propto \frac{1}{N}\Sigma B_z J_z$ | Mean current helicity (G$^2$ m$^{-1}$) |
| TOTUSJH | $H_{C\ {\rm total}} \propto \Sigma\lvert B_z\, J_z\rvert$ | Total unsigned current helicity (G$^2$ m$^{-1}$) |
| ABSNJZH | $H_{C\ {\rm abs}} \propto \lvert\Sigma B_z\, J_z\rvert$ | Absolute value of the net current helicity (G$^2$ m$^{-1}$) |
| SAVNCPP | $J_{Z\ {\rm sum}} \propto \lvert\Sigma^{B_z^+} J_z\, dA\rvert + \lvert\Sigma^{B_z^-} J_z\, dA\rvert$ | Sum of the absolute value of the net current per polarity (A) |
| MEANSHR | $\bar{\Gamma} = \frac{1}{N}\Sigma \arccos\left(\frac{B^{\rm Obs} B^{\rm Pot}}{\lvert B^{\rm Obs}\rvert\lvert B^{\rm Pot}\rvert}\right)$ | Mean shear angle (deg) |
| SHRGT45 | Area with shear $> 45°$/Total area | Area with shear angle greater than 45° (per cent of total area) |
| R_VALUE | $\Phi = \Sigma\lvert B_{\rm LOS}\rvert\, dA$ within $R$ mask | Flux along gradient-weighted neutral-line length (Mx) |

studies (M. G. Bobra & S. Couvidat 2015; P. Li et al. 2025). The CDAW database lists 355 total Type II radio bursts events along with their associated CMEs and solar flares. Not all of these events have the CME association or the flare association. Even in the presence of association of all three events (Type II burst, CME, flare) not all of these events have associated source active regions present in the database.

We match the source region's NOAA numbers to the HARP number in the SHARP database. For this we first obtain the SHARP database in its entirety and then create a map of all the NOAA numbers to HARP numbers. In case of ambiguity we discard those events and mark them unusable. Sometimes this association cannot be done due to the absence of data in the SHARP database. Because of this limitation not all of the data could be used. After all the eliminations, only 194 data, depending on the size of the dataset, could be used in this study.

### 2.3. Time-series Window Variants

We create 24 distinct time-series datasets from the SHARP data with respect to the CME and Type II bursts in order to characterize how predictive information evolves. Feature extraction is repeated for 24 distinct event observation windows: Minus windows (minus-1 hr through minus-24 hr) —the observation window ranges from 1 to 24 hr, before Type II burst onset. These windows represent the operational forecasting context in which all observations precede the event.

This systematic sweep over the time windows addresses the operational question of how much magnetic field data is required for reliable halo CME prediction from photospheric data alone, and the scientific question of when discriminative magnetic field signatures of an imminent halo eruption first become statistically detectable.

### 2.4. Feature Extraction

For each CME event and a given time window, SHARP parameter measurements falling within that window are extracted and transformed into a flat feature vector through two complementary representations:

*Statistical features*. For each of the 17 SHARP parameters, four summary statistics are computed over the full observation window. These are, temporal mean, maximum, standard deviation, and linear slope (linear trend). The slope is computed via ordinary least squares across all time steps in the window and captures the rate and direction of magnetic field evolution rather than only its instantaneous state.

*Lagged features*. To preserve the temporal sequence information that would otherwise be lost by aggregation, each SHARP parameter's individual time-step values are explicitly retained as separate features—*t*-0 (most recent time step, closest to Type II burst onset), *t*-1, *t*-2, and so on through *t*-*N* (oldest time step in the window), where *N* is determined by the number of SHARP measurements available within the window. This lagged representation allows the RF classifier to learn directly from the temporal ordering of magnetic field evolution, a design choice motivated by the demonstrated importance of pre-eruption temporal dynamics for solar event prediction (P. Li et al. 2025).

The minimum common window length across all events is used to ensure a consistent feature vector length. Combining both representations yields a flat feature vector of 17 parameters $\times$ (4 statistics + $N$ lags) SHARP-derived features per event.

Two feature representations are constructed: raw SHARP values and $-1$ to $+1$ normalized values, yielding a total of 48 dataset files (24 time windows $\times$ 2 feature representations (one normalized dataset and one raw dataset)).

The normalized datasets use $-1$ to $+1$ normalization for each SHARP feature within each sample, so that it captures the

**Table 2**
Dataset Class Distribution Summary

| Dataset | Total | Halo (+1) | Nonhalo (−1) | Halo (%) | Nonhalo (%) |
|---|---|---|---|---|---|
| minus-1 hr | 185 | 116 | 69 | 62.70 | 37.30 |
| minus-1 hr-norm | 185 | 116 | 69 | 62.70 | 37.30 |
| minus-2 hr | 188 | 117 | 71 | 62.23 | 37.77 |
| minus-2 hr-norm | 188 | 117 | 71 | 62.23 | 37.77 |
| minus-3 hr | 189 | 117 | 72 | 61.90 | 38.10 |
| minus-3 hr-norm | 189 | 117 | 72 | 61.90 | 38.10 |
| minus-4 hr | 189 | 117 | 72 | 61.90 | 38.10 |
| minus-4 hr-norm | 189 | 117 | 72 | 61.90 | 38.10 |
| minus-5 hr | 190 | 118 | 72 | 62.11 | 37.89 |
| minus-5 hr-norm | 190 | 118 | 72 | 62.11 | 37.89 |
| minus-6 hr | 186 | 116 | 70 | 62.37 | 37.63 |
| minus-6 hr-norm | 186 | 116 | 70 | 62.37 | 37.63 |
| minus-7 hr | 191 | 119 | 72 | 62.30 | 37.70 |
| minus-7 hr-norm | 191 | 119 | 72 | 62.30 | 37.70 |
| minus-8 hr | 193 | 119 | 74 | 61.66 | 38.34 |
| minus-8 hr-norm | 193 | 119 | 74 | 61.66 | 38.34 |
| minus-9 hr | 193 | 119 | 74 | 61.66 | 38.34 |
| minus-9 hr-norm | 193 | 119 | 74 | 61.66 | 38.34 |
| minus-10 hr | 193 | 119 | 74 | 61.66 | 38.34 |
| minus-10 hr-norm | 193 | 119 | 74 | 61.66 | 38.34 |
| minus-11 hr | 194 | 120 | 74 | 61.86 | 38.14 |
| minus-11 hr-norm | 194 | 120 | 74 | 61.86 | 38.14 |
| minus-12 hr | 194 | 120 | 74 | 61.86 | 38.14 |
| minus-12 hr-norm | 194 | 120 | 74 | 61.86 | 38.14 |
| minus-13 hr | 195 | 120 | 75 | 61.54 | 38.46 |
| minus-13 hr-norm | 195 | 120 | 75 | 61.54 | 38.46 |
| minus-14 hr | 195 | 120 | 75 | 61.54 | 38.46 |
| minus-14 hr-norm | 195 | 120 | 75 | 61.54 | 38.46 |
| minus-15 hr | 196 | 121 | 75 | 61.73 | 38.27 |
| minus-15 hr-norm | 196 | 121 | 75 | 61.73 | 38.27 |
| minus-16 hr | 196 | 121 | 75 | 61.73 | 38.27 |
| minus-16 hr-norm | 196 | 121 | 75 | 61.73 | 38.27 |
| minus-17 hr | 196 | 121 | 75 | 61.73 | 38.27 |
| minus-17 hr-norm | 196 | 121 | 75 | 61.73 | 38.27 |
| minus-18 hr | 196 | 121 | 75 | 61.73 | 38.27 |
| minus-18 hr-norm | 196 | 121 | 75 | 61.73 | 38.27 |
| minus-19 hr | 196 | 121 | 75 | 61.73 | 38.27 |
| minus-19 hr-norm | 196 | 121 | 75 | 61.73 | 38.27 |
| minus-20 hr | 196 | 121 | 75 | 61.73 | 38.27 |
| minus-20 hr-norm | 196 | 121 | 75 | 61.73 | 38.27 |
| minus-21 hr | 197 | 122 | 75 | 61.93 | 38.07 |
| minus-21 hr-norm | 197 | 122 | 75 | 61.93 | 38.07 |
| minus-22 hr | 198 | 122 | 76 | 61.62 | 38.38 |
| minus-22 hr-norm | 198 | 122 | 76 | 61.62 | 38.38 |
| minus-23 hr | 198 | 122 | 76 | 61.62 | 38.38 |
| minus-23 hr-norm | 198 | 122 | 76 | 61.62 | 38.38 |
| minus-24 hr | 198 | 122 | 76 | 61.62 | 38.38 |
| minus-24 hr-norm | 198 | 122 | 76 | 61.62 | 38.38 |

**Note.** Different dataset configurations; minus-*X* hr represents the number of hours before the Type II burst onset. The -norm suffix denotes normalized versions of each dataset.

change in that feature for each sample. This type of normalization has shown to yield good results in machine learning algorithms (A.-u. Raheem et al. 2021). This normalization is not applied to the derived statistical features.

### 2.5. Labeling of the Datasets

Two binary labels are assigned per event:

1. *Halo CME (label +1)*. The CME exhibits a full or partial-halo morphology in coronagraph imagery, indicating propagation largely along the Sun–Earth line.
2. *Nonhalo CME (label −1)*. The CME does not exhibit halo morphology.

The complete statistics of the dataset and the halo CME versus nonhalo CME are given in Table 2.

The counts vary slightly across time windows because not every CME event had SHARP data available for a specific time window, some events fall out depending on the observation window, and events lacking an associated GOES flare classification were excluded from the catalog. Raw and norm files always have identical counts (same events, just different feature scaling). The shortest window (minus-1 hr,

185 events) and longest (minus-24 hr, 198 events) differ by only 13 events, which means that the datasets are quite stable across all time windows. The class ratio is remarkably consistent at ∼61%–63% halo/∼37%–38% nonhalo throughout, no window introduces significant additional imbalance. This imbalance is sufficient to substantially degrade minority-class recall when standard training procedures are applied, motivating the augmentation approach described in Section 3.

## 3. Method

Figure 1 frames the classification problem addressed in this study. The input is a time series of SHARP magnetic field parameters extracted from an active region in the hours before a Type II burst onset; the output is a binary label distinguishing halo CMEs—Earth-directed events whose ejecta encircle the full solar disk and whose associated shocks propagate deep into the heliosphere—from standard nonhalo CMEs. The two populations differ in their downstream consequences for space weather but share a common photospheric origin, making the distinction a challenging and physically rich classification task. A complete description of the pipeline steps from data ingestion through SHAP explainability analysis is shown in Figure 2.

### 3.1. Class Augmentation

The nonhalo class (label $= -1$) constitutes the minority class for all experiments in this study. To address the resulting class imbalance without discarding majority-class information, which is a strategy shown to be superior to undersampling for deep learning solar event models (P. Li et al. 2025), three time-series augmentation methods are applied to each minority-class event and some majority class events, generating one synthetic variant per method. The methods are selected on the basis that they preserve the local temporal and structural patterns of the original magnetic field time series, a property identified by P. Li et al. (2025) as critical for generating synthetic samples that are beneficial rather than detrimental to model training.

*Scaling (SC).* A single scalar factor $\alpha$ is drawn from a Gaussian distribution with mean 1.0 and standard deviation $\sigma = 0.10$ and multiplied uniformly across all 17 SHARP parameter time series for a given event (T. T. Um et al. 2017; P. Li et al. 2025). Formally, for each time step $t$: $x_t' = \alpha \cdot x_t$, where $\alpha \sim \mathcal{N}(1,\ 0.10)$. Applying the same factor to all parameters preserves interfeature ratios and relative temporal structure while varying the overall amplitude, simulating the genuine amplitude variability of SHARP parameters across active regions of differing flux content (M. G. Bobra & S. Couvidat 2015).

*Jitter (JIT).* Independent Gaussian noise scaled to each parameter's own standard deviation is added to each time step: $x_t' = x_t + \varepsilon_t$, where $\varepsilon_t \sim \mathcal{N}(0,\ \sigma_x \cdot 0.05)$ and $\sigma_x$ is the standard deviation of the feature's time series within the window (T. T. Um et al. 2017). This simulates natural measurement noise inherent in HMI magnetogram data while preserving the temporal trend of each parameter.

*Window slicing (WS).* A contiguous subwindow comprising 85% of the full time-series length is randomly selected, with the same start index applied to all 17 parameters to maintain physical consistency across features (P. Li et al. 2025). The extracted subwindow is then linearly resampled back to the original length using interpolation. This technique simulates temporal uncertainty, but preserves the overall structure of the time series.

After each augmentation step, all four summary statistics—mean, maximum, standard deviation, and linear slope—are recomputed directly from the augmented time series rather than inherited from the original event record. This ensures full internal consistency, that is, augmented feature vectors are equivalent to what would be produced by running the standard feature extraction pipeline described in Section 2.4 on a SHARP time series.

The number of synthetic rows is not a free parameter. It follows directly from the selection rule and from the three augmentation methods, each of which produces different variants per selected event. In the minus-11 hr dataset, 74 nonhalo and 120 halo CMEs are present. 74 nonhalo events are selected, giving $74 \times 3 = 222$ synthetic nonhalo rows. 73 halo CME events are selected to equalize the prediction class balance, giving $73 \times 3 = 219$ synthetic halo rows. The resulting 441 synthetic rows bring the total to 635 rows. Each synthetic row is then marked to later identify synthetic events and their source real events. These labels are later used to identify the synthetic rows to avoid leakage of synthetic events into the test dataset. The counts differ slightly between time windows because the numbers of real events differ (Table 2).

### 3.2. Leakage-free Training Protocol

A strict leakage-free data split is enforced throughout all trainings to prevent synthetic-to-test contamination. This methodological precaution is motivated by the risk identified during preliminary experiments, when the combined real-plus-synthetic dataset is split naively using a standard stratified train-test partition, synthetic rows derived from real training events can appear in the test set. Because these synthetic samples share strong structural similarity with their parent real events, their presence in the test set artificially inflates evaluation metrics, producing optimistic bias that does not reflect true generalization performance. This issue is analogous to the data leakage described in the context of temporal solar datasets (M. G. Bobra & S. Couvidat 2015) and is explicitly guarded against in our evaluation design.

The leakage-free protocol is implemented as follows. First, the real events ($n \approx 194$ per file) are partitioned into a real training set (75%) and a real test set (25%) using stratified random sampling with stratification on the halo label, ensuring that the class ratio is preserved in both subsets. This yields approximately ∼145 real training events and ∼49 real test events per training experiment. Second, all synthetic rows generated from the augmentation procedure described in Section 3.1 are added exclusively to the training set. The test set therefore contains only real, unseen events at all times. The final training set of approximately 586 events (∼145 real + ∼441 synthetic) is passed to model fitting; evaluation is performed solely on the ∼49 real test events. For the baseline training experiments (no augmentation), a standard stratified 75/25 split of the real data is used directly. The number of synthetic events varies slightly depending on the length of the time-series dataset.

### 3.3. Random Forest Classifier and Threshold Optimization

An RF classifier (L. Breiman 2001) is trained independently for each of the 48 dataset files and for each of the two

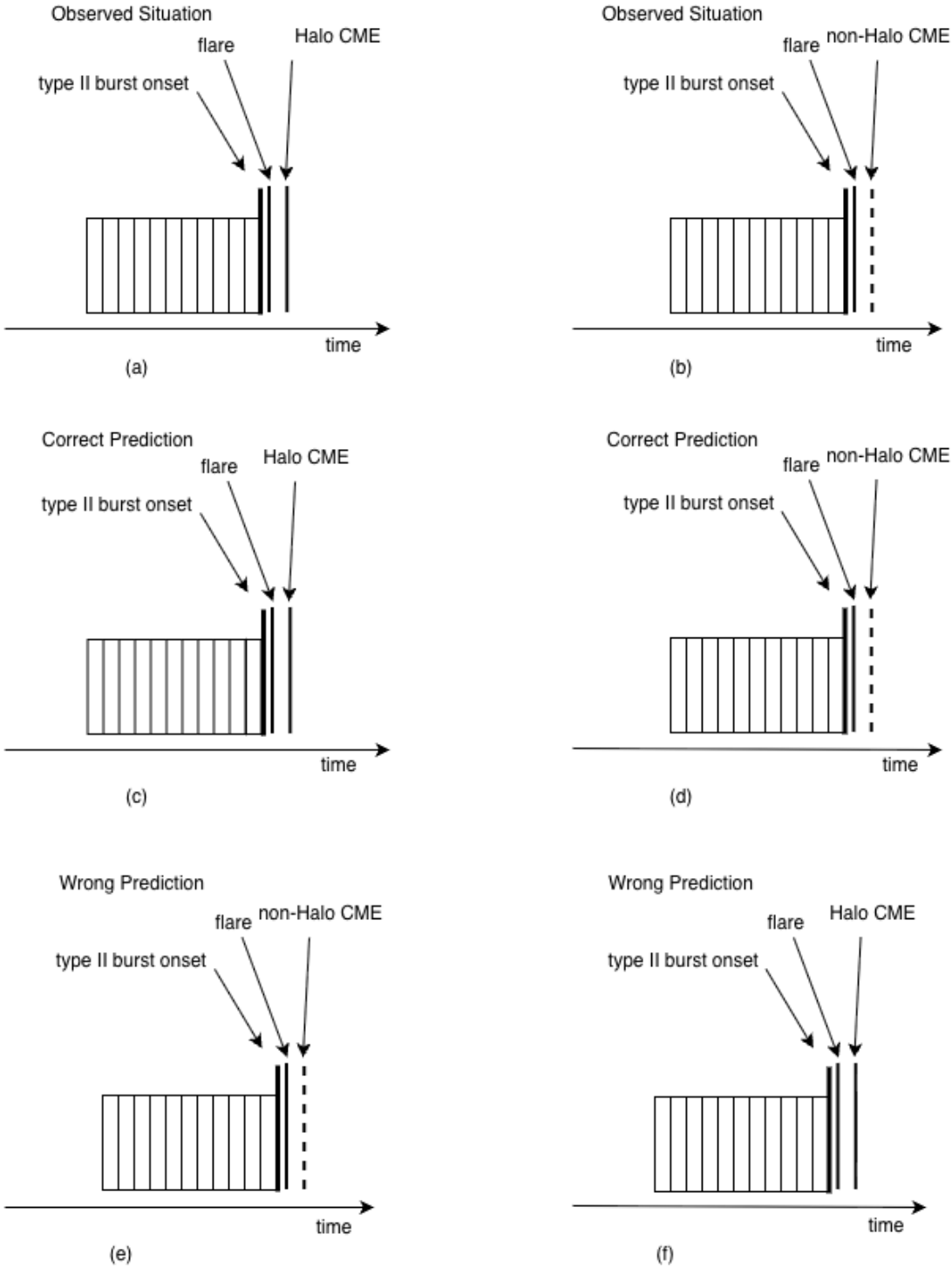


**Figure 1.** Problem definition schematic for the binary classification task addressed in this study. The horizontal axis represents time; the observation window (shaded rectangle, subdivided into individual SHARP measurement time steps at 12 minute cadence) precedes the flare-associated Type II burst onset marker. The 11 hr window shown here is used for illustration only; the study evaluates 48 window lengths ranging from 1 to 24 hr before burst onset. A solid vertical line at the right edge of the observation window denotes a halo CME; a dashed vertical line denotes a nonhalo CME. Panels (a) and (b) show the observed (true) situation: (a) a halo CME accompanying a flare-associated Type II radio burst, and (b) a nonhalo CME accompanying a flare-associated Type II radio burst. The classification task is to determine which of these two cases applies, using only the SHARP magnetic field time series recorded within the observation window before the burst onset. Panels (c) and (d) show correct model predictions: (c) the model correctly predicts halo CME (true positive) and (d) the model correctly predicts nonhalo CME (true negative). Panels (e) and (f) show incorrect model predictions: (e) the model incorrectly predicts nonhalo CME when the true label is halo CME (false negative), and (f) the model incorrectly predicts halo CME when the true label is nonhalo CME (false positive).

classification targets. The RF is selected as the base classifier on the basis of its established performance on SHARP-derived feature sets and its native support for out-of-bag (OOB) probability estimation, which is tuned for threshold optimization without requiring a separate validation set. The hyperparameter configuration is given in Table 3.

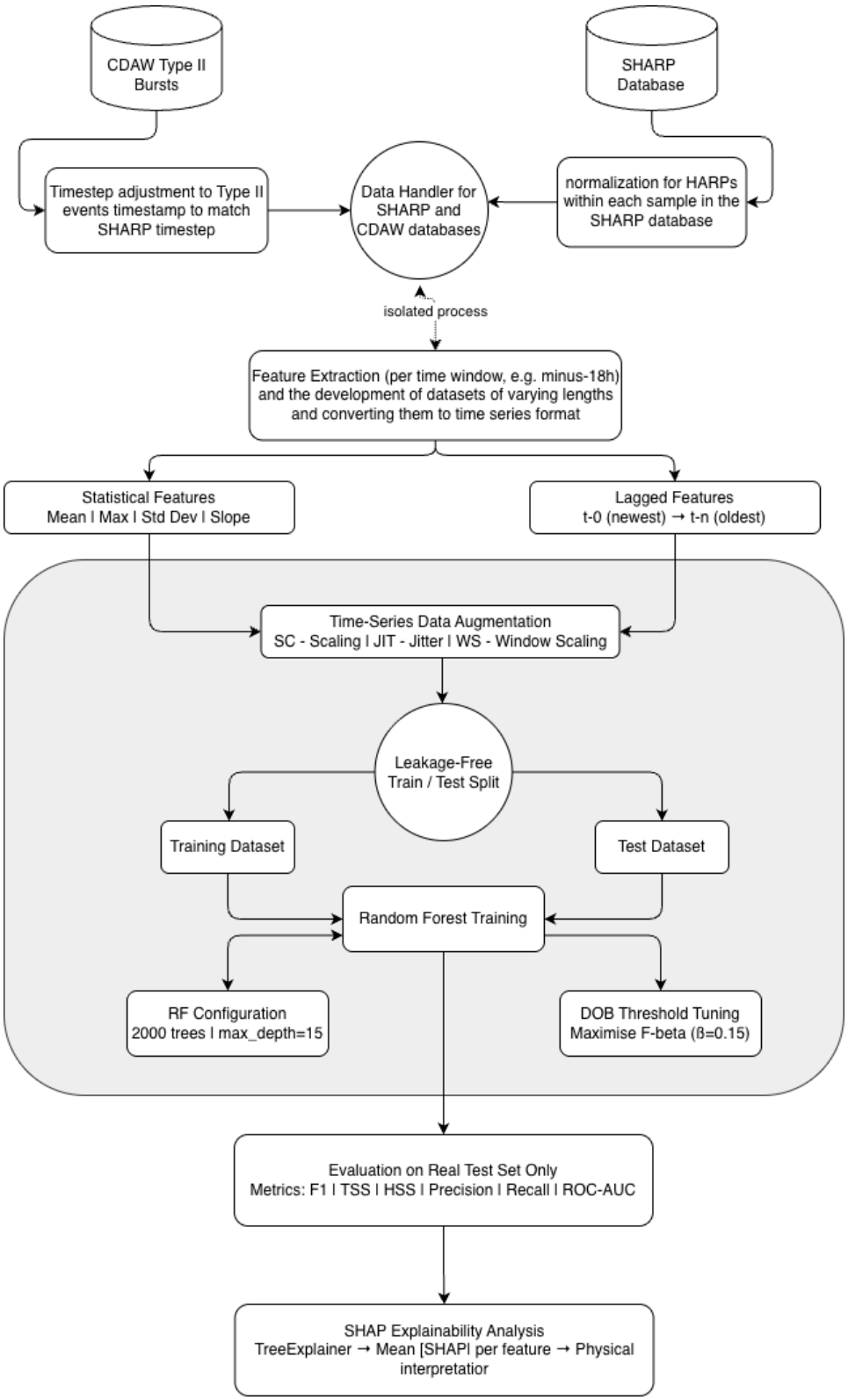


**Figure 2.** Methodology flowchart illustrating the pipeline from data collection through SHAP explainability analysis.

The ensemble size of 2000 trees and the regularization constraints (max_depth, min_samples_split, min_samples_leaf) are set to eliminate overfitting. A class weight of 1.5 is applied to the positive (halo) class to further penalize false negatives in addition to the augmentation-based class balancing.

Rather than using the default 0.5 decision threshold, a per-model optimal threshold is determined from the OOB predicted probabilities on the training set. The precision–recall curve is computed using OOB scores, and the threshold maximizing $F_\beta$ with $\beta = 0.15$ is selected. This small $\beta$ value places substantially greater weight on precision than on recall,

**Table 3**
Random Forest Hyperparameter Configuration

| Hyperparameter | Value |
|---|---|
| n_estimators | 2000 |
| class_weight | {+1:1.5, −1:1.0} |
| max_depth | 15 |
| min_samples_split | 15 |
| min_samples_leaf | 5 |
| oob_score | True |
| random_state | 44 |

biasing the classifier toward issuing only high-confidence positive predictions, which is a preference motivated by the operational space-weather context in which false positive CME alerts carry significant practical costs (M. G. Bobra & S. Couvidat 2015).

### 3.4. SHAP Explainability Analysis

To interpret the trained RF models and identify which SHARP magnetic field parameters most strongly drive the halo CME classification decisions, SHAP TreeExplainer (S. M. Lundberg et al. 2020) is applied to each model on the held-out real test set. SHAP Shapley values decompose each individual prediction into additive feature-level contributions that satisfy local accuracy, consistency, and missingness axioms, providing a theoretically grounded attribution of model behavior (S. M. Lundberg et al. 2020). For each test event, each feature is assigned a signed SHAP value representing its marginal contribution to the model's log-odds prediction. The mean absolute SHAP value across all test events—denoted $|\bar{\varphi}|$—provides a global feature importance ranking that is more robust to the effects of correlated features than the standard RF Gini impurity importance (S. M. Lundberg et al. 2020). This ranking is used to identify the dominant SHARP parameters and temporal lag offsets that are most causally informative for halo CME classification.

### 3.5. Evaluation Metrics

All models are evaluated on the real test set using metrics standard in solar eruption prediction verification (M. G. Bobra & S. Couvidat 2015; P. Li et al. 2025). These metrics, given in Table 4, include

1. *Accuracy*. Overall correct classification rate.
2. *Precision*. (TP/(TP + FP)) where TP = true positives, FP = false positives.
3. *Recall*. (TP/(TP + FN)) where FN = false negatives.
4. *F1*. Harmonic mean of precision and recall.
5. *TSS*. True skill statistic = sensitivity + specificity − 1, ranging from −1 to +1 with 0 indicating no skill beyond climatology.
6. *HSS*. Heidke skill score, measuring skill relative to random chance.
7. *ROC-AUC*. Area under the receiver operating characteristic curve, a threshold-independent measure of discrimination ability.

TSS and HSS are the primary skill scores adopted by operational space-weather forecasting centers and the broader solar event prediction literature (M. G. Bobra & S. Couvidat 2015; G. Barnes et al. 2016; H. Liu et al. 2019; P. Li et al. 2025), and are used as the principal metrics for comparing baseline and augmented model performance in this study.

**Table 4**
Evaluation Metrics and Definitions

| Metric | Definition |
|---|---|
| F1 | Harmonic mean of precision and recall |
| TSS | True skill statistic = sensitivity + specificity − 1 (range: −1 to +1; 0 = no skill) |
| HSS | Heidke skill score—skill relative to random chance |
| Precision | TP/(TP + FP) |
| Recall | TP/(TP + FN) |
| ROC-AUC | Area under the receiver operating characteristic curve |

**Table 5**
Top Five Augmented Halo CME Classification Models

| Metric | Value |
|---|---|
| Dataset | Minus-11 hr |
| Representation | Normalized |
| F1 | 0.893 |
| TSS | 0.781 |
| HSS | 0.752 |
| Precision | 0.962 |
| Recall | 0.833 |
| ROC-AUC | 0.982 |

## 4. Results

### 4.1. Top-performing Augmented Models

Table 5 lists the best-performing augmented dataset trained model by F1 score among training experiments. The minus-11 hr normalized dataset achieves a high F1 (0.893) and TSS (0.781) with realistic, nondegenerate precision and recall, and is adopted as the representative case study for the remainder of this section, with a decision threshold of 0.589 selected via OOB F-beta optimization.

### 4.2. SHAP Feature Importance: Physical Drivers of Halo CME Formation

SHAP TreeExplainer analysis was applied to the augmented RF models on their held-out real test sets. We present four complementary SHAP visualizations for the minus-11 hr normalized model, each illuminating a different dimension of the model's decision-making: a beeswarm summary plot and a bar chart of global feature importance, shown side by side (Figures 3(a) and (b)), a dependence plot for the two highest-ranked features (Figure 4), and a heatmap of SHAP values across all test events (Figure 5). Table 6 lists the top 10 features by mean absolute SHAP value $|\bar{\varphi}|$.

For comparison, Table 7 shows the importance of the RF Gini impurity feature for the same model, revealing a broadly consistent but not identical ranking.

The two rankings overlap strongly: MEANJZH_norm_lag0, MEANJZH_norm_mean, MEANALP_norm_std, MEANSHR_norm_slope, MEANGAM_norm_slope, SHRGT45_norm_slope, SAVNCPP_norm_slope, and

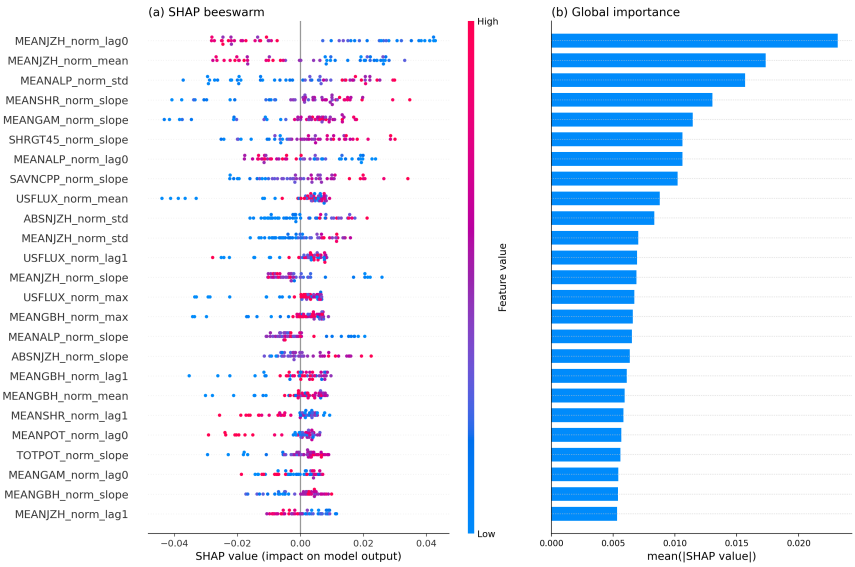


**Figure 3.** SHAP explanation of the augmented random forest model on the minus-11 hr normalized halo CME classification dataset, shown for the top 25 features. The two panels share the same vertical feature axis, ordered by decreasing mean |SHAP|, so each row can be read across both panels. (a) Beeswarm summary plot: each point represents one of the 49 test events; the horizontal axis shows the SHAP value (signed contribution to the halo CME prediction); color indicates the normalized-feature value (red = high, blue = low). (b) Global feature importance, measured by the mean absolute SHAP value $|\bar{\varphi}|$ across all test events. `MEANJZH_norm_lag0` (current helicity at the time step nearest burst onset) is the single most influential feature, followed by `MEANJZH_norm_mean` (mean current helicity over the window) and `MEANALP_norm_std` (variability of the twist parameter $\alpha$). Panel (a) shows the direction of each effect, which panel (b) does not: for example, low values of `MEANJZH_norm_lag0` push the prediction toward halo CME. The prominence of current-helicity features alongside the slopes of the mean shear angle (`MEANSHR`) and mean inclination angle (`MEANGAM`) indicates that both the present helicity state and its recent rate of change in the 11 hr before Type II burst onset are discriminative for halo CME classification.

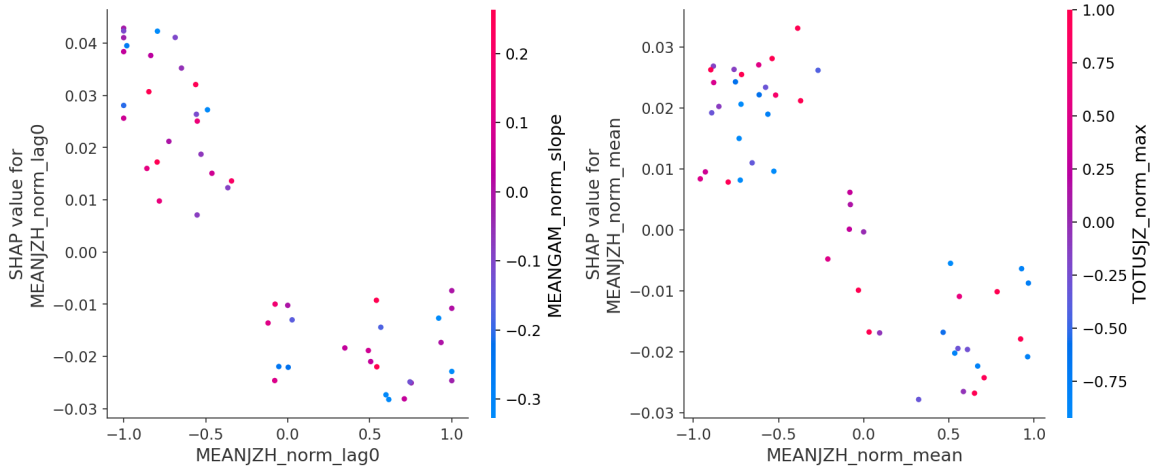


**Figure 4.** SHAP dependence plots for the two highest-ranked features in the minus-11 hr normalized model. Left: `MEANJZH_norm_lag0` (normalized current helicity at the time step nearest burst onset). Events with strongly negative normalized values receive large positive SHAP contributions (pushing the prediction toward halo CME), while events with positive normalized values are pushed toward nonhalo. The color encodes the auto-selected interaction feature `MEANGAM_norm_slope` (rate of change of the mean inclination angle), showing that the relationship is modulated by whether the field inclination is steepening or flattening. Right: `MEANJZH_norm_mean` (normalized mean current helicity over the window), which shows a similar monotonically decreasing relationship, modulated by `TOTUSJZ_norm_max` (peak total unsigned vertical current). Both panels illustrate nonlinear, physically interpretable relationships between the magnetic field parameters and the halo CME classification outcome.

`USFLUX_norm_mean` appear in the top 10 of both the SHAP and Gini rankings, strengthening confidence in the identified feature hierarchy. The beeswarm plot and bar chart (Figures 3(a) and (b)) both confirm the global dominance of current helicity (`MEANJZH`) together with the slopes of the shear, inclination, and net-current-per-polarity parameters. The dependence plots (Figure 4) further reveal that the relationship between `MEANJZH_norm_lag0` and the halo CME prediction is nonlinear, with a sharp transition near zero—events with strongly negative normalized current helicity at the time step nearest burst onset receive disproportionately high positive SHAP values. The heatmap (Figure 5) demonstrates that confident halo CME predictions arise from a coherent pattern of elevated SHAP values across multiple current-helicity and shear/inclination features simultaneously, rather than from any single dominant feature.

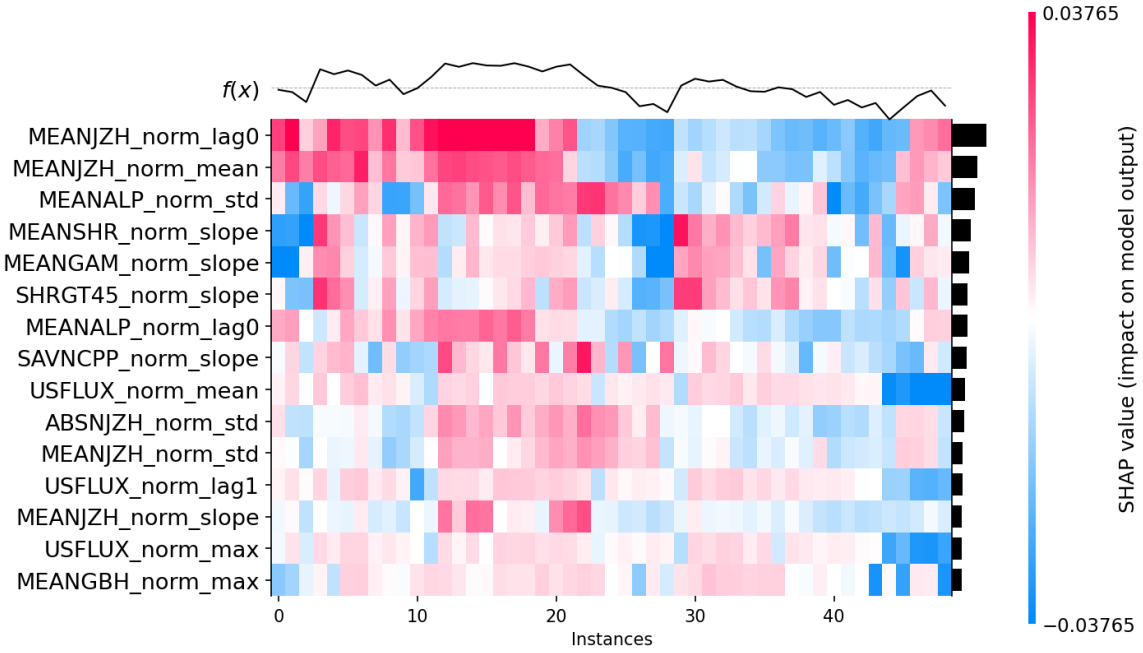


**Figure 5.** SHAP heatmap for the augmented random forest model (minus-11 hr normalized dataset), showing SHAP values for the top 15 features across all 49 test events. Events (columns) are sorted left to right by increasing total SHAP sum, corresponding to increasing predicted probability of halo CME. The black line at the top traces the model output $f(x)$. Red cells indicate a feature value that pushes the prediction toward halo CME (+1); blue cells indicate a push toward nonhalo (−1). The heatmap reveals a coherent group of events (center-left columns) characterized by consistently high positive SHAP contributions from `MEANJZH_norm_lag0`, `MEANJZH_norm_mean`, and the shear/inclination slope features—these are the events the model classifies as halo CMEs with the highest confidence. Conversely, the rightmost events (lowest predicted halo probability) show blue-dominated rows, indicating suppressed current helicity and shear evolution. The structured pattern across all 15 features confirms that halo CME classification is not driven by any single parameter in isolation, but by a coherent ensemble of pre-eruption magnetic field conditions.

**Table 6**
Top 10 SHAP Feature Importances (Mean Absolute SHAP Value $|\varphi|$) for the Minus-11 hr Normalized Augmented Model

| Rank | Feature | Mean \|SHAP\| | Physical Interpretation |
|---|---|---|---|
| 1 | `MEANJZH_norm_lag0` | 0.02316 | Current helicity at the time step nearest burst onset |
| 2 | `MEANJZH_norm_mean` | 0.01733 | Mean current helicity over the window |
| 3 | `MEANALP_norm_std` | 0.01567 | Variability of the twist parameter $\alpha$ over the window |
| 4 | `MEANSHR_norm_slope` | 0.01303 | Rate of increase of the mean shear angle |
| 5 | `MEANGAM_norm_slope` | 0.01143 | Rate of change of the mean inclination angle |
| 6 | `SHRGT45_norm_slope` | 0.01062 | Rate of change of the area with shear angle $>45^{\circ}$ |
| 7 | `MEANALP_norm_lag0` | 0.01060 | Twist parameter $\alpha$ at the time step nearest burst onset |
| 8 | `SAVNCPP_norm_slope` | 0.01021 | Rate of change of the net current per polarity |
| 9 | `USFLUX_norm_mean` | 0.00878 | Mean total unsigned magnetic flux over the window |
| 10 | `ABSNJZH_norm_std` | 0.00832 | Variability of the absolute net current helicity over the window |

**Note.** Feature names follow the convention `PARAMETER_norm_statistic`, where `norm` denotes −1 to +1 normalization and the statistic suffix denotes: `slope` = linear trend over the time window; `mean` = window mean; `std` = standard deviation; `lag0` = value at the time step nearest the event onset.

### *4.3. Overall Impact of Augmentation*

To assess the effect of data augmentation, we trained and evaluated a baseline RF model (no augmentation, standard 75/25 split) and an augmented RF model (leakage-free split with SC, JIT, and WS synthetic samples) for each of the 48 dataset variants. Table 8 reports the average of each metric across all 48 experiments, together with the mean improvement delta.

The F1 score improved in all 48 experiments (100%) and the TSS improved in all 48 (100%). The ROC-AUC increase from 0.599 to 0.977 confirms that the improvement in discrimination extends across all decision thresholds, not only at the selected operating point.

### *4.4. Augmented versus Nonaugmented: Full Comparison by Time Window*

Table 9 provides a comprehensive comparison of baseline and augmented F1, TSS, and HSS for all 24 normalized-feature datasets. This table demonstrates that the improvement is universal, not confined to a subset of time windows and that augmentation consistently rescues models that achieved near-zero or negative skill in the baseline condition.

### *4.5. The Best Results: Minus-11 hr Normalized Dataset*

The minus-11 hr normalized dataset introduced in Section 4.2 is examined here in full to illustrate the complete

**Table 7**
Top 10 RF Gini Impurity Feature Importances for the Minus-11 hr Normalized Augmented Model

| Rank | Feature | Importance |
|---|---|---|
| 1 | MEANSHR_norm_slope | 0.02565 |
| 2 | MEANJZH_norm_lag0 | 0.02494 |
| 3 | MEANGAM_norm_slope | 0.02292 |
| 4 | SHRGT45_norm_slope | 0.02044 |
| 5 | MEANALP_norm_std | 0.02034 |
| 6 | SAVNCPP_norm_slope | 0.01985 |
| 7 | MEANJZH_norm_mean | 0.01924 |
| 8 | USFLUX_norm_lag1 | 0.01701 |
| 9 | USFLUX_norm_mean | 0.01686 |
| 10 | MEANGBH_norm_lag1 | 0.01647 |

**Note.** Feature-naming conventions follow those described in Table 6.

**Table 8**
Mean Performance Metrics across 48 Halo CME Classification Experiments before (Baseline) and after (Augmented) Time-series Data Augmentation

| Metric | Baseline | Augmented | Δ |
|---|---|---|---|
| F1 | 0.537 | 0.815 | +0.278 |
| TSS | 0.131 | 0.689 | +0.557 |
| HSS | 0.127 | 0.631 | +0.504 |
| Precision | 0.665 | 0.998 | +0.334 |
| Recall | 0.486 | 0.691 | +0.205 |
| ROC-AUC | 0.599 | 0.977 | +0.377 |

**Note.** Values reported as average over 48 experiments (24 datasets × 2 representations). Δ denotes the difference (augmented − baseline).

augmentation pipeline and to contrast baseline against augmented behavior on a single representative case.

Baseline performance (minus-11 hr normalized): accuracy = 0.660, precision = 0.710, recall = 0.733, F1 = 0.721, TSS = 0.283, HSS = 0.286, ROC-AUC = 0.648. The baseline model identifies only 73.3% of halo events in the test set, leaving over a quarter of the operationally important halo CMEs undetected.

Augmented performance (minus-11 hr normalized): accuracy = 0.878, precision = 0.962, recall = 0.833, F1 = 0.893, TSS = 0.781, HSS = 0.752, ROC-AUC = 0.982 (decision threshold = 0.589, selected via OOB F-beta optimization). Augmentation increases recall from 73.3% to 83.3% while simultaneously improving precision from 0.710 to 0.962. The confusion matrix for the augmented model is shown in Figure 6.

### *4.6. Type II Burst Frequency and Flare Class Properties of the Event Catalog*

The Type II burst frequency data preserved in the CDAW catalog permit an independent characterization of the physical differences between halo and nonhalo CME populations in our sample. Table 10 summarizes the burst start frequency (FreqA) and burst end frequency (FreqB) statistics for each class.

Table 11 gives the distribution of events across FreqA frequency bands.

Table 12 shows the distribution of burst end frequencies (FreqB). Approximate heliocentric distances are indicated in the table comments using the G. Mann et al. (1999) heliospheric density model.

Table 13 reports the associated flare class distribution for each CME category.

## 5. Discussion

What strikes us most about the SHAP results is not any single feature in isolation, but the coherent physical story they collectively tell. The events in our catalog are not simply CMEs—each one is a manifestation of the full flare–CME–Type II burst chain, in which an impulsive energy release in the low corona (the flare) drives a large-scale plasma ejecta (the CME) that, when fast and wide enough, launches a magnetohydrodynamic shock through the corona that excites the Type II radio emission we use as our event anchor. The question we ask is what photospheric magnetic field conditions, observed up to 11 hr before the Type II burst onset, determine whether the resulting CME achieves the full-halo morphology associated with Earth-directed events. The features the RF found most useful for answering that question are, almost without exception, the same quantities that MHD theory and observational studies have long highlighted as signatures of a region building toward this kind of energetic, coupled eruption. That a data-driven classifier recovers this narrative independently is, we think, a meaningful validation of both the model and the underlying physics. We organize the discussion around several interlocking themes that emerge from the ranking: the role of current helicity as a proxy for flux-rope formation and flare energy release; the significance of the twist parameter and its variability; the dominance of temporal field evolution, expressed through the shear and inclination slopes, over instantaneous state; and the way persistent twist drives the region past the Kruskal–Shafranov kink-instability threshold that determines both eruption speed and shock strength. We then connect these magnetic drivers to the independent observational signatures of the flare–CME–Type II chain in our event catalog and draw the threads together into a unified physical picture with implications for operational forecasting.

### *5.1. Current Helicity and Magnetic Twist*

The dominant feature in both the SHAP (Table 6) and Gini (Table 7) rankings is current helicity, MEANJZH, at both lag0 and its window mean. Defined as $H_c \propto \frac{1}{N}\sum B_z J_z$, it is the pointwise correlation between the vertical field and the vertical current it threads—the photospheric proxy most directly tied to the twist and self-linkage of the coronal field. In a current-free (potential) field $J_z = 0$ and $H_c$ vanishes, the minimum-energy state; any nonzero $H_c$ certifies field-aligned currents, the only way the corona stores the free energy that powers *both* flare and CME—a region of high current helicity is effectively a charged battery. Its sign also encodes *handedness*: a region dominated by one sign of helicity carries a coherent twist, the structural prerequisite for a flux rope that can lose equilibrium as a whole. The dependence plot (Figure 4) shows a strongly nonlinear relation with a transition near zero, so strongly signed normalized helicity yields large positive SHAP contributions. It is thus the magnitude and coherence of the twist—not a weak or mixed-sign field—that predisposes a region to a powerful eruption: a strongly twisted rope sits nearer the kink and torus instability thresholds (B. Kliem &

**Table 9**
Baseline versus Augmented Performance (F1, TSS, HSS) for all 24 Normalized-feature Time-window Datasets

| Dataset | $F1_{Base}$ | $TSS_{Base}$ | $HSS_{Base}$ | $F1_{Aug}$ | $TSS_{Aug}$ | $HSS_{Aug}$ | ΔF1 | ΔTSS |
|---|---|---|---|---|---|---|---|---|
| minus-1 hr | 0.746 | 0.140 | 0.154 | 0.792 | 0.655 | 0.593 | +0.045 | +0.515 |
| minus-2 hr | 0.795 | 0.211 | 0.244 | 0.926 | 0.862 | 0.827 | +0.131 | +0.652 |
| minus-3 hr | 0.229 | 0.033 | 0.027 | 0.824 | 0.700 | 0.636 | +0.595 | +0.667 |
| minus-4 hr | 0.324 | 0.102 | 0.085 | 0.776 | 0.633 | 0.564 | +0.451 | +0.532 |
| minus-5 hr | 0.465 | 0.175 | 0.151 | 0.776 | 0.633 | 0.564 | +0.310 | +0.458 |
| minus-6 hr | 0.667 | 0.040 | 0.042 | 0.766 | 0.621 | 0.556 | +0.099 | +0.581 |
| minus-7 hr | 0.162 | −0.111 | −0.090 | 0.800 | 0.667 | 0.600 | +0.638 | +0.777 |
| minus-8 hr | 0.229 | 0.081 | 0.065 | 0.776 | 0.633 | 0.573 | +0.547 | +0.553 |
| minus-9 hr | 0.318 | −0.135 | −0.117 | 0.889 | 0.800 | 0.756 | +0.571 | +0.935 |
| minus-10 hr | 0.471 | −0.074 | −0.068 | 0.750 | 0.600 | 0.538 | +0.279 | +0.674 |
| **minus-11 hr** | **0.721** | **0.283** | **0.286** | **0.893** | **0.781** | **0.752** | **+ 0.172** | **+ 0.497** |
| minus-12 hr | 0.698 | 0.183 | 0.188 | 0.868 | 0.767 | 0.718 | +0.170 | +0.583 |
| minus-13 hr | 0.000 | −0.050 | −0.040 | 0.909 | 0.833 | 0.795 | +0.909 | +0.883 |
| minus-14 hr | 0.560 | 0.167 | 0.154 | 0.889 | 0.800 | 0.756 | +0.329 | +0.633 |
| minus-15 hr | 0.633 | 0.083 | 0.083 | 0.846 | 0.733 | 0.681 | +0.213 | +0.650 |
| minus-16 hr | 0.596 | 0.067 | 0.065 | 0.868 | 0.767 | 0.718 | +0.271 | +0.700 |
| minus-17 hr | 0.596 | 0.067 | 0.065 | 0.846 | 0.733 | 0.681 | +0.250 | +0.667 |
| minus-18 hr | 0.588 | 0.200 | 0.186 | 0.800 | 0.667 | 0.608 | +0.212 | +0.467 |
| minus-19 hr | 0.604 | 0.183 | 0.173 | 0.800 | 0.667 | 0.608 | +0.196 | +0.483 |
| minus-20 hr | 0.577 | 0.150 | 0.141 | 0.800 | 0.667 | 0.608 | +0.223 | +0.517 |
| minus-21 hr | 0.381 | 0.100 | 0.083 | 0.735 | 0.581 | 0.513 | +0.354 | +0.480 |
| minus-22 hr | 0.263 | 0.061 | 0.051 | 0.830 | 0.710 | 0.650 | +0.567 | +0.648 |
| minus-23 hr | 0.689 | 0.227 | 0.225 | 0.784 | 0.645 | 0.580 | +0.096 | +0.418 |
| minus-24 hr | 0.761 | 0.221 | 0.240 | 0.808 | 0.677 | 0.615 | +0.047 | +0.456 |

**Note.** The bold row (minus-11 hr) is the normalized-feature case-study dataset discussed in Section 4.2. Negative baseline F1/TSS/HSS values indicate near-random or below-chance baseline performance. Δ values are augmented − baseline. This table reports normalized-feature datasets only; the raw-feature versions show similar overall trends.

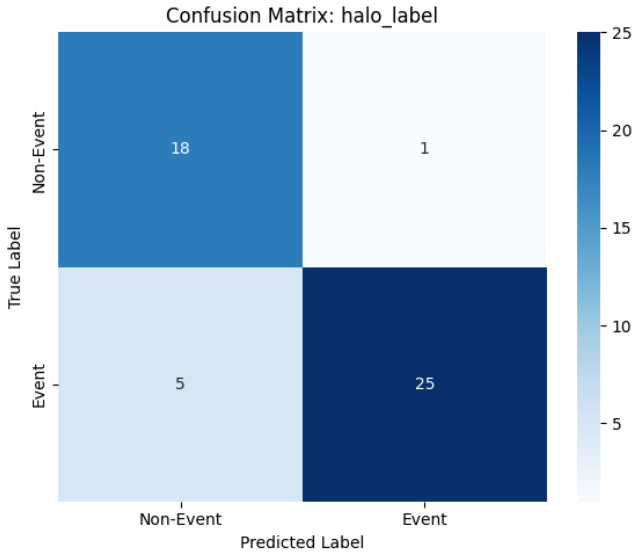


**Figure 6.** Confusion matrix for the augmented random forest model on the minus-11 hr normalized dataset (11 hr, pre-event). Rows: true label, columns: predicted label. Non-Event = Non-Halo CME (−1), Event = Halo CME (+1).

T. Török 2006), releases more energy on reconnection, ejects faster, and drives a stronger shock.

### 5.2. The Twist Parameter α and Its Variability

The third-ranked feature, MEANALP standard deviation, adds a dynamical dimension. The twist parameter $\alpha_{total} \propto \sum J_z B_z / \sum B_z^2$ is the flux-weighted ratio of vertical current to vertical field—the twist per unit field, and the closest

**Table 10**
Type II Burst Start and End Frequency Statistics for Halo and Nonhalo CME Populations

| Quantity | Halo | Nonhalo |
|---|---|---|
| FreqA mean | 12,488 kHz (12.5 MHz) | 11,038 kHz (11.0 MHz) |
| FreqA median | 14,000 kHz | 14,000 kHz |
| FreqB mean | 1809 kHz (1.8 MHz) | 4224 kHz (4.2 MHz) |
| FreqB median | 440 kHz | 2550 kHz |
| Bandwidth mean | 10,679 kHz | 6814 kHz |
| Bandwidth median | 12,900 kHz | 6700 kHz |

**Note.** FreqA = burst start frequency; FreqB = burst end frequency; bandwidth = FreqA − FreqB. FreqA is the frequency at which the Type II burst first appears in the dynamic spectrum; FreqB is the frequency at which it ceases to be detectable.

**Table 11**
Distribution of Type II Burst Start Frequencies (FreqA) across Frequency Bands for Halo and Nonhalo CME Populations

| FreqA Band | Nonhalo | Halo | Total | Halo (%) |
|---|---|---|---|---|
| 500 kHz–1 MHz | 4 | 1 | 5 | 20.0 |
| 1–5 MHz | 15 | 13 | 28 | 46.4 |
| 5–10 MHz | 6 | 13 | 19 | 68.4 |
| 10–15 MHz | 38 | 61 | 99 | 61.6 |
| >15 MHz | 15 | 33 | 48 | 68.8 |

SHARP analog to the force-free constant $\alpha$ in $\nabla \times \boldsymbol{B} = \alpha \boldsymbol{B}$. Where MEANJZH measures *how much* current helicity is present, MEANALP measures *how intensely twisted* the field is, independent of its strength. Crucially, it enters through its variability rather than its mean: a constant $\alpha$ describes a stable

**Table 12**
Distribution of Type II Burst end Frequencies (FreqB) across Frequency Bands for Halo and Nonhalo CME Populations

| FreqB Band | Nonhalo | Halo | Total | Halo (%) |
|---|---|---|---|---|
| <50 kHz | 0 | 5 | 5 | 100.0 |
| 50–100 kHz | 1 | 10 | 11 | 90.9 |
| 100–500 kHz | 12 | 50 | 62 | 80.6 |
| 500 kHz–1 MHz | 6 | 11 | 17 | 64.7 |
| 1–5 MHz | 31 | 30 | 61 | 49.2 |
| 5–10 MHz | 19 | 13 | 32 | 40.6 |
| 10–15 MHz | 9 | 2 | 11 | 18.2 |

**Note.** Using the G. Mann et al. (1999) heliospheric density model: 50 kHz ≈ 215 $R_\odot$ (∼1 au); 115 kHz ≈ 108 $R_\odot$ (∼0.5 au); 585 kHz ≈ 25.4 $R_\odot$; 1 MHz ≈ 15.7 $R_\odot$; 3 MHz ≈ 6 $R_\odot$.

**Table 13**
Distribution of Associated GOES Flare Classes for Halo and Nonhalo CME Events in the Study Catalog

| Flare Class | Nonhalo | Halo | Total | Halo (%) |
|---|---|---|---|---|
| X | 6 | 39 | 45 | 86.7 |
| M | 42 | 69 | 111 | 62.2 |
| C | 27 | 13 | 40 | 32.5 |

**Note.** Flare class is obtained from the CDAW database record associated with each CME event. Events lacking a GOES flare classification were excluded from the catalog.

equilibrium, whereas a fluctuating $\alpha$ signals twist being actively injected, redistributed by reconnection, or driven past the point where equilibrium can hold. The model has learned that the *dynamics* of twist discriminate eruptive regions better than a static twist level.

### 5.3. Shear and Inclination Evolution

A defining pattern across Figures 3 and 5 is that several top features are slopes—temporal rates of change. MEANSHR, the mean shear angle $\bar{\Gamma} \propto \arccos(\boldsymbol{B}^{\rm Obs}\cdot\boldsymbol{B}^{\rm Pot}/|\boldsymbol{B}^{\rm Obs}|\,|\boldsymbol{B}^{\rm Pot}|)$, measures the angular departure of the field from potential; since free energy scales with this departure, a rising slope tracks energy loading along the polarity inversion line, where the most flare-productive filament channels form (C. J. Schrijver 2007). SHRGT45 slope shows the area above 45° of shear expanding, so the current-carrying region is growing. MEANGAM slope tracks the inclination $\bar{\gamma} \propto \arctan(B_h/B_z)$; a field tipping toward horizontal signals a forming rope axis, and its rate how quickly that structure assembles—fittingly, MEANGAM slope is the auto-selected interaction feature modulating the helicity effect in Figure 4. SAVNCPP slope captures the growth of the net unsigned current per polarity that supports and ultimately destabilizes the rope. Together these slopes describe a region *actively driven* toward instability—shear rising, high-shear area spreading, field inclining, net current growing—rather than sitting in a quasi-static state.

### 5.4. The Kink Threshold as the Unifying Mechanism

The prominence of the twist diagnostics—MEANJZH and especially the actively varying MEANALP—invites a mechanistic reading through the Kruskal–Shafranov kink-instability criterion. A line-tied coronal flux rope becomes kink-unstable once its total twist exceeds a critical value, conventionally

**Table 14**
Metric Stability over 100 Repeated 75:25 Random Splits (Minus-11 hr Normalized, halo_label)

| Metric | Mean | Std | Min | Max |
|---|---|---|---|---|
| Accuracy | 0.767 | 0.062 | 0.653 | 0.918 |
| Precision | 0.991 | 0.017 | 0.955 | 1.000 |
| Recall | 0.626 | 0.113 | 0.433 | 0.900 |
| F1 | 0.761 | 0.078 | 0.605 | 0.931 |
| TSS | 0.614 | 0.096 | 0.433 | 0.847 |
| HSS | 0.559 | 0.107 | 0.372 | 0.831 |
| ROC-AUC | 0.987 | 0.013 | 0.937 | 1.000 |

**Note.** Each run uses the same pipeline as the main experiment. Seed 7 corresponds to the canonical result (F1 = 0.893, TSS = 0.781, ROC-AUC = 0.982) reported in Table 5. The wider spread in Recall and F1 relative to ROC-AUC reflects the sensitivity of threshold-dependent metrics to the small test set size (≈49 events per split).

$\Phi_{\rm twist} \simeq 2\pi$ (profile- and aspect-ratio-dependent). For a line-tied loop of length $L$ and minor radius $a$, the safety factor $q = 2\pi a B_z/(L B_\phi)$ falls below unity as the azimuthal field $B_\phi$ winds up, the kink setting in at $q < 1$. Since twist per unit length is essentially the force-free parameter, $\Phi_{\rm twist} \approx \alpha L/2$, a region driven toward $\alpha \approx 4\pi/L$ sits on the cusp of marginal stability, and MEANALP is a direct photospheric proxy for exactly this proximity. The classifier has, in effect, learned to read the rate of approach to $\Phi_{\rm twist} = 2\pi$—which is precisely the diagnostic an MHD-informed forecaster would privilege. A region with high and rising MEANALP and strong, coherent helicity (MEANJZH) is one whose rope is being driven toward the threshold from below, which is why the model reads these quantities, and the variability of $\alpha$ in particular, as the strongest signatures of an imminent halo eruption.

Crossing this threshold is a single event that branches into all three observables. The rope kinks, converting internal twist into large-scale writhe, drives itself upward through the torus-unstable regime (B. Kliem & T. Török 2006) and reconnects beneath—a release of free energy that *is* the impulsive flare and *simultaneously* launches the CME: one instability, both phenomena. Table 13 reflects this, with the halo population overwhelmingly X- and M-class while weaker C-class events are predominantly nonhalo—the most supercritical, twist-laden ruptures produce both the hardest flares and the fully Earth-directed halos.

The criterion also bears directly on the Type II burst. A rope crossing the threshold while strongly supercritical carries greater free energy per unit flux and accelerates harder, acting as a faster, more impulsive piston. Shock strength scales with the Alfvén Mach number $M_A = v_{\rm shock}/v_A$, and shock-front electron acceleration—the source of the coherent emission—becomes efficient only once the shock is sufficiently supercritical, so a higher-momentum piston yields a higher-$M_A$ shock radiating as a metric Type II burst. Because the emission tracks the local plasma frequency $f_p \propto \sqrt{n_e}$, a stronger shock forms lower in the denser corona and survives farther out, igniting at higher start frequency and persisting to lower end frequency—exactly the pattern in Tables 10–12, where halos show higher mean FreqA, lower FreqB, and larger bandwidth. The same twist excess that triggers the kink thus sets the shock strength, closing the chain from photospheric $\alpha$ to the metric Type II emission.

### 5.5. Halo CMEs, Accompanying Flares, and Type II Bursts

The frequency statistics in Tables 10–13 offer an independent, observational view of the differences between halo and nonhalo populations that complements the magnetic field analysis above and reinforces the causal narrative from the SHAP results. Where the SHAP rankings identify *which* pre-eruption conditions anticipate a halo, the radio and flare statistics show *what those conditions produce*—a stronger flare, faster CME, and more powerful shock. Both are downstream of the same stored free energy, so their consistency is exactly what the flux-rope picture predicts.

#### 5.5.1. Start Frequencies and Shock Formation Height

The start frequency `FreqA` marks where the Type II shock first becomes a coherent radio source. Since $f_p \propto \sqrt{n_e}$, a higher `FreqA` implies denser ambient plasma and hence a lower, earlier formation height. Halos show a higher mean start frequency than nonhalos (Table 10; 12,488 versus 11,038 kHz), and the banded distribution (Table 11) sharpens the trend: halos are a minority of the lowest-frequency bursts (20% between 500 kHz and 1 MHz) but rise to roughly two-thirds of the highest (68.4% at 5–10 MHz, 68.8% above 15 MHz). A shock igniting low in the dense corona requires a piston fast enough to steepen before it has climbed far—precisely the behavior of the strongly twisted, supercritical ropes flagged by `MEANJZH` and `MEANALP`, whose greater free energy per unit flux drives a faster ejection and an earlier, lower shock.

#### 5.5.2. End Frequencies and Heliospheric Extent

The end frequency `FreqB` records where the shock ceases to emit efficiently, so a lower `FreqB` corresponds to greater heliocentric distance. Using the G. Mann et al. (1999) density model, the band boundaries map to physical heights—50 kHz $\approx 215\,R_\odot$ (∼1 au), 115 kHz $\approx 108\,R_\odot$ (∼0.5 au), 1 MHz $\approx 15.7\,R_\odot$—so a burst persisting to lower `FreqB` is one whose shock survives farther out. The contrast is striking: halos reach a much lower mean end frequency than nonhalos (Table 10; 1809 versus 4224 kHz; median 440 versus 2550 kHz), and Table 12 shows them dominating the longest-lived bursts—all five events below 50 kHz (100%), 90.9% of the 50–100 kHz band, and 80.6% of the 100–500 kHz band—while nonhalos dominate the high-frequency bursts that fade quickly. A shock radiating from $\gtrsim$10 MHz down to tens of kHz has propagated from the low corona to near 1 au as a coherent, electron-accelerating structure, the signature of a high-Mach-number driver. The resulting larger halo bandwidth (10,679 versus 6814 kHz) is a direct observational measure of shock longevity, following from the same supercritical-driver physics inferred from the magnetic field.

#### 5.5.3. Flare Class Association and the Common Energy Reservoir

The GOES flare class distribution (Table 13) closes the loop. Halo association rises monotonically with flare magnitude: halos are 86.7% of X-class but only 32.5% of C-class events, with M-class intermediate at 62.2%. Because flare class measures the radiative energy released by reconnection, drawn from the same reservoir that accelerates the CME, the strong halo–X-class correlation is exactly what the flux-rope picture demands—the most energetic reconnection accompanies the most energetic, fully Earth-directed eruptions, both powered by the same highly nonpotential, twist-laden field. This provides external corroboration for the SHAP finding that current helicity and shear are the dominant predictors of halo formation. Flare class, CME morphology, and Type II shock are thus three faces of a single magnetic event. The decisive point is that these three channels are only indirectly coupled to one another: a merely statistical correlation of the photospheric field would not be expected to align coherently with both the radiative output of reconnection and the interplanetary shock signature. That current helicity and shear evolution anticipate *both* the hardest flares *and* the longest-lived, lowest-end-frequency bursts indicates that the model has isolated a quantity upstream of the common energy reservoir, rather than a downstream symptom of any single channel—the strongest claim the present data support, and one consistent with a genuinely causal pathway from photospheric nonpotentiality to Earth-directed eruption.

### 5.6. Unified Picture and Operational Implications

`USFLUX`, the total unsigned flux $\Phi = \sum |B_z|\, dA$, enters through its mean as a scale term, setting the *capacity* of the energy reservoir while helicity and the slope features decide whether that capacity is released. The model thus reconstructs the standard MHD narrative: a substantial flux reservoir (`USFLUX`) hosting a coherent twisted flux rope (`MEANJZH`, `MEANALP`) whose shear, inclination, and net current all increase before onset. This combination—stored free energy plus active driving toward the kink threshold—simultaneously powers an energetic flare, accelerates a fast wide CME into a halo, and drives the strong shock that produces an intense Type II burst. That a data-driven classifier independently recovers the very quantities MHD links to all three links of the flare–CME–shock chain cross-validates both the model and the physics. Operationally, this suggests that monitoring the temporal evolution of current helicity, twist, and shear in the hours before onset—rather than any single snapshot—carries forecasting value not just for halo CMEs but for the coupled space-weather event as a whole.

### 5.7. Robustness and Limitations

To address the concern that conclusions might be specific to a single train/test partition, we carried out two complementary robustness analyses: 100 repeated random splits and fivefold stratified cross validation.

*Repeated random splits*. The full augmented pipeline (leakage-free 75:25 stratified split, in-memory augmentation of training minority rows, RF with identical hyperparameters, OOB F-beta threshold selection with $\beta = 0.15$) was executed 100 times using split seeds 0, 1, …, 99; the canonical result reported in Table 5 corresponds to seed 7. Table 14 summarizes the metric distributions across all 100 runs, and Figure 7 shows the corresponding histograms.

ROC-AUC is the most stable metric (std = 0.013), as expected for a threshold-free measure. Recall and F1 show wider spread (std = 0.113 and 0.078 respectively), a natural consequence of evaluating on ≈49 real test events per split. The canonical split (seed 7) falls toward the upper end of the F1 distribution, which we acknowledge; the mean across all 100 splits (F1 = 0.761, TSS = 0.614) provides a more

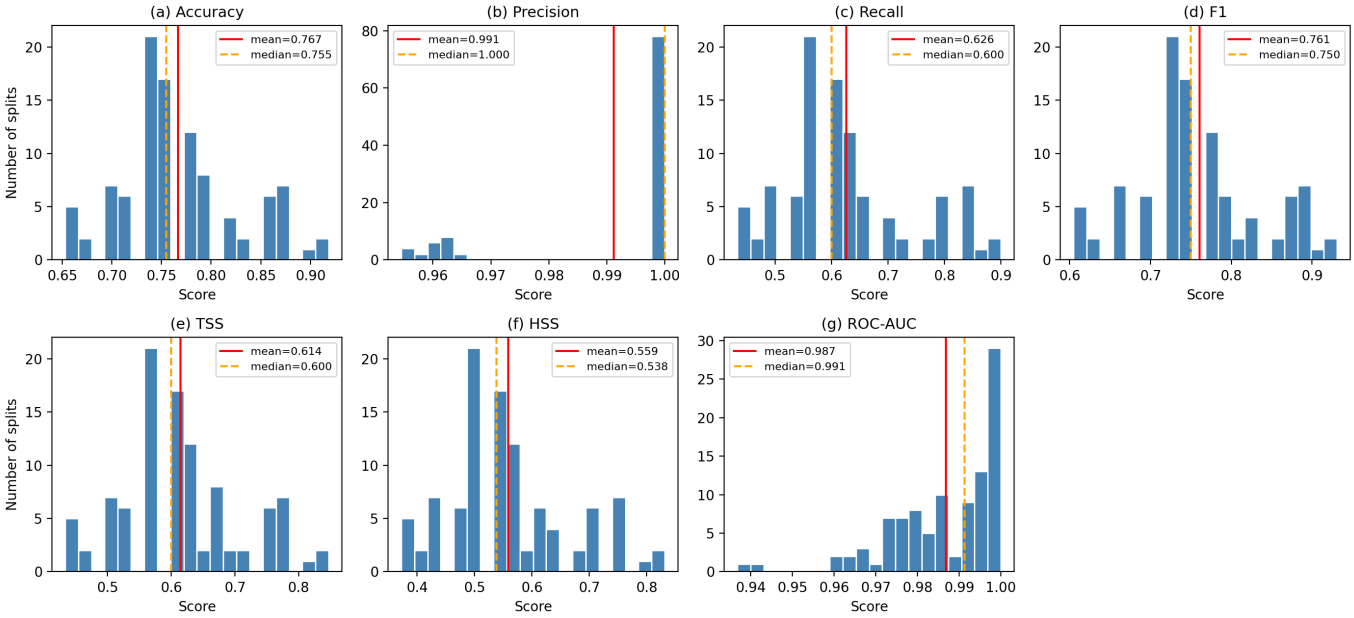


**Figure 7.** Distributions of seven evaluation metrics across 100 repeated 75:25 random splits of the minus-11 hr normalized dataset. The threshold-dependent metrics are shown in the top row, (a) accuracy, (b) precision, (c) recall, and (d) F1. The skill scores and the threshold-free measure are shown in the bottom row, (e) TSS, (f) HSS, and (g) ROC-AUC. Red vertical lines indicate the mean; orange dashed lines indicate the median. ROC-AUC is highly concentrated near 1.0 (mean = 0.987, std = 0.013), confirming stable discriminative ability. Recall and F1 show wider spread due to the small per-split test set size (≈49 events).

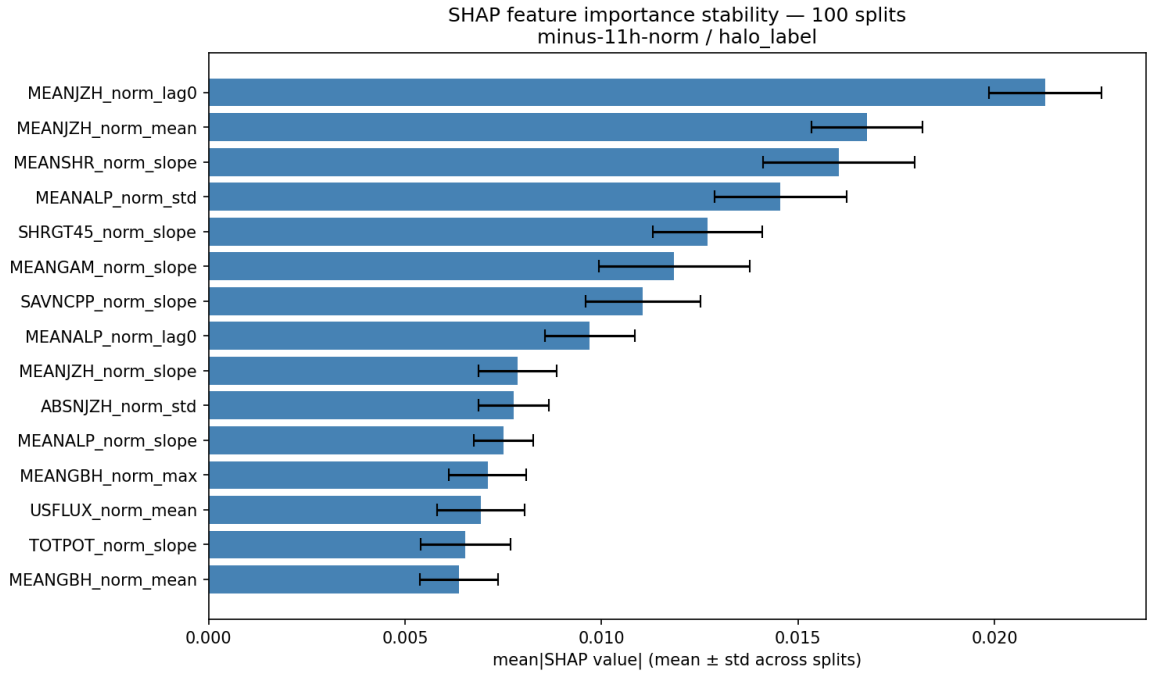


**Figure 8.** Mean absolute SHAP feature importance ± standard deviation across 100 repeated random splits (minus-11 hr normalized dataset). Features are ranked by mean importance. The narrow error bars confirm that the SHAP hierarchy is stable: `MEANJZH_norm_lag0` and `MEANJZH_norm_mean` consistently dominate, followed by shear and inclination slope features, regardless of the particular train/test partition.

conservative but still operationally useful estimate of generalization performance.

*SHAP feature ranking stability*. Figure 8 shows the mean absolute SHAP value ± standard deviation for the top features across all 100 splits. `MEANJZH_norm_lag0` ranks first in every single split (mean $|\bar{\varphi}| = 0.0213 \pm 0.0014$), followed consistently by `MEANJZH_norm_mean` (0.0168 ± 0.0014), `MEANSHR_norm_slope` (0.0160 ± 0.0019), `MEANALP_norm_std` (0.0146 ± 0.0017), and `SHRGT45_norm_slope` (0.0127 ± 0.0014). The small standard deviations relative to the mean values confirm that the physical interpretation of the model is insensitive to which particular events are held out for testing.

*Fivefold stratified cross validation*. As a complementary robustness check, we applied fivefold stratified cross validation with in-fold augmentation: the 194 real events were

partitioned into five folds (`StratifiedKFold`, `random_state` = 42), and for each fold the minority-class rows of the *training* fold were augmented in memory before fitting the RF, ensuring no synthetic copies of held-out events entered the training set. The OOB threshold was selected using $\beta = 1.0$ (standard F1) rather than $\beta = 0.15$, since the stronger precision bias of the latter causes near-zero recall at small training-fold sizes. Across five folds the mean metrics were F1 = 0.718 ± 0.057, Precision = 0.677 ± 0.046, Recall = 0.775 ± 0.124, TSS = 0.168 ± 0.112, and ROC-AUC = 0.598 ± 0.089. The substantially lower ROC-AUC compared with the repeated-split analysis is expected as each test fold contains only ≈39 events. Crucially, the SHAP feature ranking under fivefold CV is identical to that obtained under the repeated splits: `MEANJZH_norm_lag0` and `MEANJZH_norm_mean` rank first and second in all five folds, confirming that the interpretability conclusions are robust regardless of evaluation protocol.

*Limitations*. The primary limitation of this study is sample size. With ≈194 real events, individual metric estimates carry substantial uncertainty (∼0.1 in F1 across splits), and both robustness analyses reflect this. The 100-split repeated-split analysis is the more informative robustness measure for this sample size because it preserves the same 75:25 train/test ratio as the reported results and uses larger test sets (≈49 events) than the CV folds (≈39 events). Future work incorporating additional Type II burst catalogs or longer LASCO observation baselines would substantially reduce this uncertainty.

A methodological caveat attaches to the slope features that figure prominently in the SHAP ranking. The trend term for each SHARP parameter is an ordinary least-squares fit over the whole observation window, so it summarizes the evolution of that parameter by a single average rate of change. This first-order description is a reasonable approximation over the shorter windows, but it is not a claim that these quantities evolve linearly: shear, twist, inclination, and net current can change abruptly or in bursts associated with flux emergence and reconnection, particularly in the hours immediately preceding onset. Over the longer windows a single slope therefore averages together intervals of quiescence and intervals of rapid driving and may either dilute a short-lived but physically decisive acceleration or, conversely, register a spurious trend produced by a steplike change. The lagged features mitigate this in part, since they retain the individual time step values and allow the classifier to recover structure that the linear fit discards, and the fact that the SHAP hierarchy is unchanged across 100 repeated splits and all five cross-validation folds indicates that the ranking itself is not an artifact of the fitting choice. Nevertheless, the slope features should be interpreted as measuring the net tendency of a parameter over the window rather than the functional form of its evolution. We regard this, together with the sample-size limitation noted above, as a principal direction for future work.

## 6. Conclusions

We have applied an explainable machine learning framework—an RF classifier interpreted through SHAP—to the question of whether a flare-associated, Type II-producing CME attains halo morphology, using only photospheric SHARP magnetic field time series recorded up to 24 hr before burst onset. Our principal findings are as follows.

1. Leakage-free time-series augmentation (SC, JIT, WS) improves both F1 and TSS in all 48 experiments, raising mean TSS from 0.131 to 0.689 and mean F1 from 0.537 to 0.815, and rescuing windows that exhibited near-zero or below-chance baseline skill.
2. The representative minus-11 hr normalized model attains F1 = 0.893, TSS = 0.781, and ROC-AUC = 0.982, with realistic, nondegenerate precision and recall. Across 100 repeated random splits this configuration is stable in its discriminative ability (ROC-AUC = 0.987 ± 0.013), with the threshold-dependent metrics showing the wider spread expected at this sample size (mean F1 = 0.761, TSS = 0.614); the representative split lies toward the upper end of that distribution.
3. SHAP analysis identifies current helicity (`MEANJZH`), the variability of the twist parameter (`MEANALP`), and the temporal slopes of shear and inclination (`MEANSHR`, `MEANGAM`) as the dominant predictors—the same nonpotentiality diagnostics that MHD theory associates with a flux rope approaching the Kruskal–Shafranov kink threshold.
4. These magnetic drivers are corroborated by independent observational signatures in the same catalog: halo events are preferentially associated with X-class flares and with Type II bursts of higher start frequency, lower end frequency, and larger bandwidth—the expected signature of a faster, more supercritical shock driven by a more energetic eruption.
5. The SHAP feature hierarchy is robust to the choice of train/test partition: `MEANJZH_norm_lag0` ranks first in every one of 100 repeated splits and in all five folds of stratified cross validation, confirming that the physical interpretation does not depend on which events are held out. The principal limitation is sample size (≈194 real events), which leaves individual skill estimates uncertain at the ∼0.1 level in F1; enlarging the Type II/CME catalog is the clearest path to tighter estimates.

Taken together, the results indicate that a data-driven classifier independently recovers the physical narrative of the coupled flare–CME–Type II chain, and that the temporal *evolution* of helicity, twist, and shear—rather than any single instantaneous snapshot—carries the discriminative information. Operationally, this motivates forecasting schemes that track the rate of approach to instability in the hours before onset, with value extending beyond halo CME prediction to the coupled space-weather event as a whole.

## Acknowledgments

This study made use of the CDAW database that is maintained and provided by the CDAW Data Center by NASA and The Catholic University of America in cooperation with the Naval Research Laboratory. Data used in this study also came from SOHO, which is a project of international cooperation between ESA and NASA. This study also utilized the SHARP database maintained and provided by JSOC at Stanford. This research used version 2.1.2 of the SunPy open-source software package. J.T.L.W. and H.W. acknowledge support from NASA under grant No. 80NSSC24M0174. This study is part of the doctorate thesis of A.-u.R.

## ORCID iDs

Huseyin Cavus https://orcid.org/0000-0003-4224-7039
Jason T. L. Wang https://orcid.org/0000-0002-2486-1097
Haimin Wang https://orcid.org/0000-0002-5233-565X

## References

Abduallah, Y., Alobaid, K. A., Wang, J. T. L., et al. 2024, SpWea, 22, e2023SW003824
Abduallah, Y., Jordanova, V. K., Liu, H., et al. 2022, ApJS, 260, 16
Ahmadzadeh, A., Aydin, B., Georgoulis, M. K., et al. 2021, ApJS, 254, 23
Alobaid, K. A., Wang, J. T. L., Wang, H., et al. 2024, SoPh, 299, 159
Barnes, G., Leka, K. D., Schrijver, C. J., et al. 2016, ApJ, 829, 89
Bloomfield, D. S., Higgins, P. A., McAteer, R. T. J., & Gallagher, P. T. 2012, ApJL, 747, L41
Bobra, M. G., & Couvidat, S. 2015, ApJ, 798, 135
Bobra, M. G., Sun, X., Hoeksema, J. T., et al. 2014, SoPh, 289, 3549
Breiman, L. 2001, MachL, 45, 5
Camporeale, E. 2019, SpWea, 17, 1166
Cane, H. V., Erickson, W. C., & Prestage, N. P. 2005, ApJ, 623, 1180
Cavus, H. 2015, AstBu, 70, 117
Cavus, H., Coban, G. C., Wang, H., et al. 2025, AdSpR, 76, 3726
Cavus, H., & Karafistan, A. I. 2007, MNRAS, 380, 142
Florios, K., Kontogiannis, I., Park, S.-H., et al. 2018, SoPh, 293, 28
Gopalswamy, N., Aguilar-Rodriguez, E., Yashiro, S., et al. 2005, JGRA, 110, A12S07
Gopalswamy, N., Lara, A., Yashiro, S., Kaiser, M. L., & Howard, R. A. 2001, JGRA, 106, 29207
Gopalswamy, N., Mäkelä, P., & Yashiro, S. 2019, SunGe, 14, 111
Hoshoudy, G. A., & Cavus, H. 2018, JApA, 39, 39
Kliem, B., & Török, T. 2006, PhRvL, 96, 255002
Li, P., Bahri, O., Filali Boubrahimi, S., & Hamdi, S. M. 2025, ApJS, 280, 52
Liu, C., Deng, N., Wang, J. T. L., & Wang, H. 2017, ApJ, 843, 104
Liu, H., Liu, C., Wang, J. T. L., & Wang, H. 2019, ApJ, 877, 121
Liu, H., Liu, C., Wang, J. T. L., & Wang, H. 2020, ApJ, 890, 12
Lu, L., Inhester, B., Feng, L., Liu, S., & Zhao, X. 2017, ApJ, 835, 188
Lundberg, S. M., Erion, G., Chen, H., et al. 2020, NatMI, 2, 56
Mann, G., Jansen, F., MacDowall, R., Kaiser, M., & Stone, R. 1999, A&A, 348, 614
Nishizuka, N., Sugiura, K., Kubo, Y., Den, M., & Ishii, M. 2018, ApJ, 858, 113
Qahwaji, R., Colak, T., Al-Omari, M., & Ipson, S. 2008, SoPh, 248, 471
Raheem, A.-u., Cavus, H., Coban, G. C., et al. 2021, MNRAS, 506, 1916
Schrijver, C. J. 2007, ApJL, 655, L117
Um, T. T., Pfister, F. M. J., Pichler, D., et al. 2017, in Proc. of the 19th ACM Int. Conf. on Multimodal Interaction, ed. E. Lank et al. (Association for Computing Machinery), 216
Williams, T., Prior, C. B., MacTaggart, D., Aslam, O. P. M., & Bloomfield, D. S. 2026, ApJ, 999, 87
Zhang, H., Jing, J., Wang, J. T. L., et al. 2025, ApJ, 981, 37