

# An Interpretable Machine Learning Approach to Understanding the Relationships between Solar Flares and Source Active Regions

Huseyin Cavus<sup>1</sup>, Jason T. L. Wang<sup>2</sup>, Teja P. S. Singampalli<sup>2</sup>, Gani Caglar Coban<sup>1</sup>, Hongyang Zhang<sup>2</sup>, Abd-ur Raheem<sup>1</sup>, Haimin Wang<sup>3</sup>

<sup>1</sup>Department of Physics, Canakkale Onsekiz Mart University, 17110 Canakkale, Turkey

<sup>2</sup>College of Computing, New Jersey Institute of Technology, Newark, NJ 07102, USA

<sup>3</sup>Institute for Space Weather Sciences, New Jersey Institute of Technology, Newark, NJ 07102, USA

## Key Points:

- Implementing the random forest algorithm as an interpretable machine learning approach to enable the binary classification of solar flares using the physical features of ARs collected from SolarMonitor.org
- Today's value of Mount Wilson AR type and yesterday's value of Hale class are the most and the least governing features respectively
- The difference in the number of spots in the AR comparing with the previous day, i.e., NoS\_Difference, has a distinctive effect for decision-making in both global and local interpretations

arXiv:2502.15066v1 [astro-ph.SR] 20 Feb 2025

**Abstract**

Solar flares are defined as outbursts on the surface of the Sun. They occur when energy accumulated in magnetic fields enclosing solar active regions (ARs) is abruptly expelled. Solar flares and associated coronal mass ejections are sources of space weather that adversely impact devices at or near Earth, including the obstruction of high-frequency radio waves utilized for communication and the deterioration of power grid operations. Tracking and delivering early and precise predictions of solar flares is essential for readiness and catastrophe risk mitigation. This paper employs the random forest (RF) model to address the binary classification task, analyzing the links between solar flares and their originating ARs with observational data gathered from 2011 to 2021 by SolarMonitor.org and the XRT flare database. We seek to identify the physical features of a source AR that significantly influence its potential to trigger  $\geq C$ -class flares. We found that the features of AR\_Type\_Today, Hale\_Class\_Yesterday are the most and the least prepotent features, respectively. NoS\_Difference has a remarkable effect in decision-making in both global and local interpretations.

**Plain Language Summary**

Solar flares arise from the emission of energy accumulated in the magnetic fields of solar ARs. However, the underlying mechanism for these outbursts remains unidentified. This study employs an interpretable machine learning approach to establish a solar flare prediction model, enabling the binary classification of flares ( $\geq C$  as the positive class and  $< C$  as the negative class) using the physical features of ARs obtained from SolarMonitor.org and flare data supplied from the XRT flare catalog with observational data from 2011 to 2021. Using the Random Forest (RF) algorithm, the performance metric values are acquired as recall of 0.81, precision of 0.82, accuracy of 0.74 and F1 score of 0.82. Our findings indicate that, the AR\_Type\_Today ( $\beta$ ,  $\gamma$  and  $\gamma\delta$  types are effective in more than 90% of positive classifications) is the most influential feature and the Hale\_Class\_Yesterday ( $\beta$ ,  $\beta\gamma$  and  $\beta\gamma\delta$  classes are dominant in over 85% cases of positive classes) is the least influential factor for an AR generating a  $\geq C$ -class flare. The NoS\_Difference feature, which takes values between  $-13$  and  $10$  in more than 90% of positive test samples, significantly influences the decision-making process in both global and local interpretations.

**1 Introduction**

Active regions (ARs) on the solar disk are areas where the Sun's magnetic field is altered. They are often associated with sunspots and are the origin of violent eruptions such as coronal mass ejections (CMEs) and solar flares (Harrison, 1995). The presence of sunspots indicates ARs. Ultraviolet and X-ray images of the Sun reveal these regions to be luminous because of the extraordinarily energetic events associated with ARs (Friedman, 1963). These regions are surrounded by dramatic structures, such as coronal loops and solar prominences.

Different solar activities occur according to the open or closed magnetic field lines formed by the sunspot groups. Although open magnetic field lines cause the solar wind, sudden changes in magnetic field lines and reconnections of closed magnetic field lines cause solar eruptions such as CMEs and flares (Priest, 2014). These eruptive events can initiate other physical activities, such as interplanetary shock waves and geomagnetic storms, which affect the Earth's atmosphere, disrupt short-wave communications and adversely affect satellite and space operations (Miteva & Samwel, 2022; Vemareddy & Mishra, 2015).

Flares emerge in active regions (Abramenko, 2005), especially near sunspots, where powerful magnetic fields penetrate the photosphere and connect the corona to the Sun. Flares are propelled by the immediate (minutes to tens of minutes) discharge of mag-

netic energy from the corona. Andrews (2003) reported that all the X-class flares and 55% of the M-class flares were associated with 311 LASCO-observed CMEs between 1996 and 1999. In a more extensive mathematical investigation, Yashiro et al. (2005) indicated that the association rate of CMEs significantly escalated with the magnitude of X-ray flares, rising from 20% for C-class flares to 100% for very large flares. They showed that all CMEs linked to X-class flares were observed by the LASCO coronagraphs, whereas 25-67% of CMEs linked to C-class flares remained undetected. Youssef (2012) worked with 776 CME-flare-matched activities. They found that 67% of the CMEs occurred after flare events. These findings indicate that although flares and CMEs often occur together, there is no one-to-one correspondence between them. Moreover, flare durations do not guarantee the association of CMEs (Harrison, 1995).

Falconer et al. (2012) provide a dual rationale for the increased flare productivity of active regions in their investigation of past flaring as a supplementary indicator of free magnetic energy to predict solar eruptions. Firstly, these ARs typically exhibit a complex multipolar configuration of opposite-polarity magnetic flux instead of a singular bipole arrangement; secondly, they are undergoing rapid evolution through convective flux transfer, flux emergence, and/or flux cancellation. Flares emit energy in a variety of ways, including electromagnetic radiation, particles (protons and electrons), and mass streams (Boudjada et al., 2012). They are distinguished by their X-ray brilliance (X-ray flux). X-class flares are the largest. M-class flares have one-tenth the energy of X-class flares. C-class flares have one-tenth the energy of M-class flares (Bai & Sturrock, 1989). Kusano et al. (2012) and Jaeggli & Norton (2016) worked on magnetic classes of ARs for the period between 1992 and 2015. These authors found that  $\alpha$  and  $\beta$  containing ARs make up 20% and 80%, respectively.

In their study on the development of the determined magnetic complexity of ARs for the 23rd solar cycle, Guo et al. (2010) stated that the most strong flares and rapid coronal mass ejections (CMEs) typically originate in active regions (ARs) having intricate structures. Mohamed et al. (2018) performed a statistical analysis of the frequency of X-class flares during the declining and peak phases of Solar Cycles 23 and 24, respectively. They found a consistency between the number of X-class flares and the number of days of  $\beta\gamma\delta$  group ARs.

In this study, we adopt an interpretable machine learning (ML) approach to examine the relationships between solar flares and their source ARs using observational data collected from 2011 to 2021. By adopting the interpretable ML approach, we attempt to explain how the ML model makes decisions in determining the relationships between flares and their source ARs. This helps us to understand which physical properties of a source AR might play an important role in producing a flare.

The remainder of this paper is organized as follows. Section 2 describes the data used in this study. Section 3 formalizes the problem studied here into a binary classification task and details our approach to solving the task. Section 4 reports the experimental results. Section 5 presents a discussion of the results. Section 6 concludes the paper.

## 2 Data

The database at SolarMonitor.org was utilized to analyze ARs and their magnetic group properties (Gallagher et al., 2002). It aggregates solar activity data from various sources, including the Solar Dynamics Observatory (SDO) with HMI and AIA instruments, the Global H $\alpha$  Network, the Solar and Heliospheric Observatory (SOHO) with EIT and MDI instruments, Synoptic Optical Long-term Investigations of the Sun (SOLIS) with full-disk chromospheric magnetograms, SECCHI, and databases from the National Oceanic and Atmospheric Administration (NOAA). It also features images from

the Hinode XRT team, STEREO’s Extreme UltraViolet Imager (EUVI), and the Solar X-ray Imager (SXI), as well as X-ray images from the X-ray SXT full-disk database.

The SolarMonitor system provides AR data, flare forecasting information, and comprehensive solar disk images of the Sun. Its database offers near-real-time data on solar activity. Specifically, the database includes the following physical properties for an AR with NOAA number, the latest position, and associated flares if available: Hale class, AR type, sunspot area (millionths of the solar disk area), and number of spots.

Hale class has 8 values:  $\alpha$ ,  $\beta$ ,  $\alpha\gamma$ ,  $\alpha\delta$ ,  $\alpha\gamma\delta$ ,  $\beta\gamma$ ,  $\beta\delta$ ,  $\beta\gamma\delta$ . The Hale classes are then regrouped into five groups to obtain the Mount Wilson classification (AR type) with five values:  $\alpha$ ,  $\beta$ ,  $\gamma$ ,  $\delta$ ,  $\gamma\delta$  where

- $\alpha$  refers to Hale class  $\alpha$ ;
- $\beta$  refers to Hale class  $\beta$ ;
- $\gamma$  refers to Hale classes  $\alpha\gamma$  and  $\beta\gamma$ ;
- $\delta$  refers to Hale classes  $\alpha\delta$  and  $\beta\delta$ ;
- $\gamma\delta$  refers to Hale classes  $\alpha\gamma\delta$  and  $\beta\gamma\delta$ .

In summary, our study used 10 physical features, listed in Table 1, obtained from the SolarMonitor database. Each AR record contains values of the 10 features of the corresponding AR.

Feature	Description
Hale_Class_Today	Hale class of the AR in today
Hale_Class_Yesterday	Hale class of the AR in yesterday
AR_Type_Today	Mount Wilson classification of the AR in today
AR_Type_Yesterday	Mount Wilson classification of the AR in yesterday
Spot_Area_Today	Area of spots of the AR in today
Spot_Area_Yesterday	Area of spots of the AR in yesterday
Spot_Area_Difference	Spot_Area_Yesterday – Spot_Area_Today
NoS_Today	Number of spots of the AR in today
NoS_Yesterday	Number of spots of the AR in yesterday
NoS_Difference	NoS_Yesterday – NoS_Today

**Table 1.** Physical Properties or Features of an AR Considered in Our Study

The flare data was sourced from the XRT flare catalog. This catalog contains extensive flare information, including positions, classes, GOES fluxes of the flare source regions, and the source AR’s NOAA numbers. In this catalog, the magnetic group assignments of the flare source ARs were determined using the SolarMonitor database. The data from SolarMonitor.org and XRT were acquired using a web scraping script written in Python. We collected 986 AR-produced flare events in the period between January 2011 and October 2021. After removing ARs with incomplete or missing feature values, we obtained 837 AR records, where each AR record contains values of the 10 physical features of the corresponding AR and the flare event produced by the AR. Table 2 summarizes the types and counts of these flare events.

Flare Type	Counts
A-Class	1
B-Class	231
C-Class	539
M-Class	61
X-Class	5

**Table 2.** Types and Counts of AR-Produced Flares Considered in Our Study

### 3 Methodology

#### 3.1 Classification Task

We consider  $\geq$ C-class flares, which have been widely studied in the literature (Abduallah et al., 2023; Donahue & Inceoglu, 2024; Huang et al., 2018; Li et al., 2020; Liu et al., 2019; Nishizuka et al., 2018; Sun et al., 2022; Tang et al., 2021; Vysakh & Mayank, 2023; Zheng et al., 2023). A  $\geq$ C-class flare refers to a C-class, M-class, or X-class flare and a  $<$ C-class flare refers to an A-class or B-class flare.

Given an AR record with 10 physical properties or features listed in Table 1, we want to solve the following binary classification problem: does the corresponding AR produce a  $\geq$ C-class flare? Through binary classification, we attempt to understand which physical properties or features of a source AR play an important role in determining whether the AR produces a  $\geq$ C-class flare.

#### 3.2 Interpretable Classification Model

We adopted the random forest (RF) model to solve the binary classification task described in Section 3.1. RF based models have been shown to be very effective in predicting solar flares (Liu et al., 2017). The hyperparameters were tuned using the randomized search capability available in the Python machine learning library, `scikit-learn` (Pedregosa et al., 2011). There were 100 trees in the random forest. For model interpretation, we adopted the SHAP (SHapley Additive exPlanations) framework (Lundberg & Lee, 2017). SHAP is used to calculate the contribution of each feature to the final output. Let  $F$  be the set of the 10 features in Table 1 and let  $S$  be a subset of  $F$ . The SHAP value of each feature  $i$  in  $F$ ,  $1 \leq i \leq |F|$ , denoted  $\phi_i$ , is defined as:

$$\phi_i = \sum_{S \subseteq F - \{i\}} \frac{|S|!(|F| - |S| - 1)!}{|F|!} (C(S \cup \{j\}) - C(S)). \quad (1)$$

The SHAP value determines the difference in the contribution that the feature  $i$  brings to the prediction if included in a specific subset  $S$ , and averages the differences over every possible combination of possible subsets  $S$  of the features in terms of the contribution function:  $C(S \cup \{j\}) - C(S)$ . SHAP can be used for global interpretation (that is, the interpretation is made over the entire test set) or local interpretation (the interpretation is made for a specific test sample). We adopted `shap.TreeExplainer` in our study.

## 4 Results

### 4.1 Evaluation Metrics

Given an AR record  $R$  with 10 physical feature values, we define  $R$  as a true positive (TP) if our RF model predicts that  $R$  belongs to the positive class (i.e., the corresponding AR produces a  $\geq$ C-class flare) and  $R$  is indeed positive. We define  $R$  as a false positive (FP) if our RF model predicts that  $R$  is positive while  $R$  actually belongs to the negative class (i.e., the corresponding AR produces a  $<$ C-class flare). We say  $R$  is a true negative (TN) if our RF model predicts that  $R$  is negative and  $R$  is indeed negative;  $R$  is a false negative (FN) if our RF model predicts that  $R$  is negative while  $R$  is actually positive. When the context is clear, we also use TP (FP, TN, and FN, respectively) to represent the number of true positives (false positives, true negatives, and false negatives, respectively) produced by our RF model.

The evaluation metrics used in this study include the following:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \quad (2)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \quad (3)$$

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}}, \quad (4)$$

$$\text{F1} = \frac{2 \times \text{TP}}{2 \times \text{TP} + \text{FP} + \text{FN}}. \quad (5)$$

We split the set of 837 AR records at hand into a 80% training set and a 20% test set. There are a total of 605 AR records in the positive class and 232 AR records in the negative class. After splitting, the training set has 484 positive AR records and 185 negative AR records, totaling 669 training AR records or training samples. The test set has 121 positive AR records and 47 negative AR records, totaling 168 test AR records or test samples. Each AR record in the training set contains 10 AR feature values together with a label (positive vs. negative). Each AR record in the test set contains 10 AR feature values without a label. The label in a test sample will be predicted by our classification model. We compute the TP, FP, TN, FN and all evaluation metric values based on the test set.

### 4.2 Performance Assessment

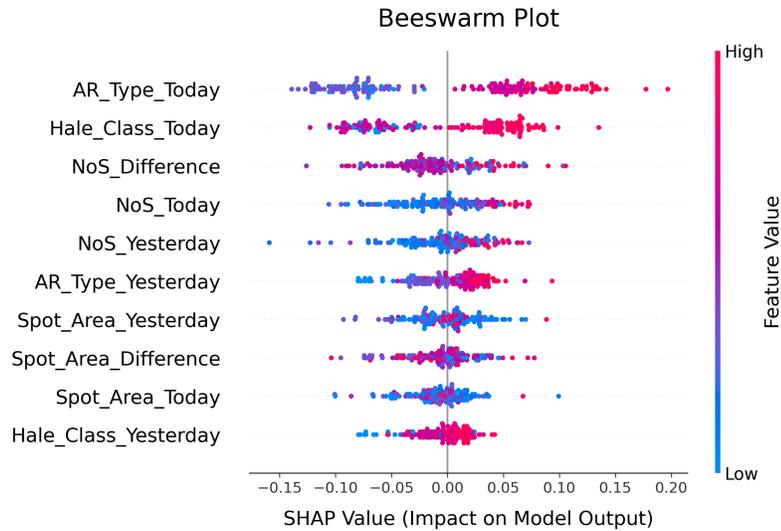
Figure 1 presents the confusion matrix obtained by our RF model, which provides a breakdown analysis of errors that occur when the model makes predictions in the test set. Based on the confusion matrix, we obtain: recall = 0.81, precision = 0.82, accuracy = 0.74, F1 = 0.82. To assess the stability and reliability of our RF model, we also performed a five-fold cross-validation. The average accuracy for five folds is 0.75. These results indicate that our RF model performs reasonably well in solving the binary classification problem described in Section 3.1.

### 4.3 Model Interpretation

As described in Section 3.2, SHAP incorporates game theory to give each feature a SHAP value. Positive SHAP values have a positive effect, while negative SHAP values have a negative effect. A positive effect increases the probability of predicting that

TN	FP
26	21
FN	TP
23	98

**Figure 1.** Confusion matrix obtained by our RF model on the test set with 168 test samples.

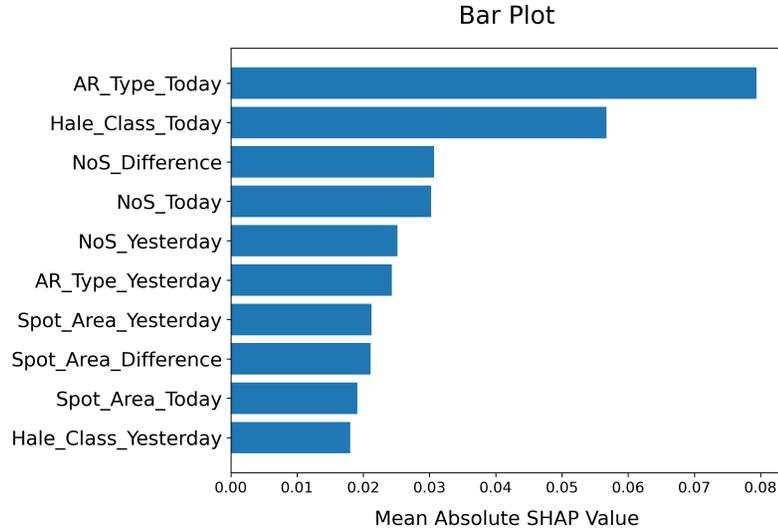


**Figure 2.** Beeswarm plot to visualize the positive or negative effect of a feature for each test sample, represented by a color dot, on the RF model's predictions.

a test sample is in the positive class, whereas a negative effect increases the probability of predicting that a test sample is in the negative class.

Figure 2 presents the beeswarm plot for the RF model. In the beeswarm plot, we can observe, for each feature, the distribution of SHAP values for all test samples. Each test sample corresponds to a dot in each feature row. The placement of each dot on the  $x$ -axis is determined by the SHAP value that the corresponding feature of the corresponding test sample receives. When dots cluster, it shows common test samples based on their SHAP values of the corresponding feature. The color of a dot depends on the value of a feature in the corresponding test sample. Red indicates a high feature value, blue indicates a low feature value, and purple indicates an average or moderate feature value.

Now, focus on the NoS.Today feature in Figure 2. There are more test samples with negative SHAP values than with positive SHAP values. This means that the NoS.Today feature is more likely to push the model's predictions towards a negative class. Furthermore, low feature values (blue dots) tend to lead to negative predictions, as these low



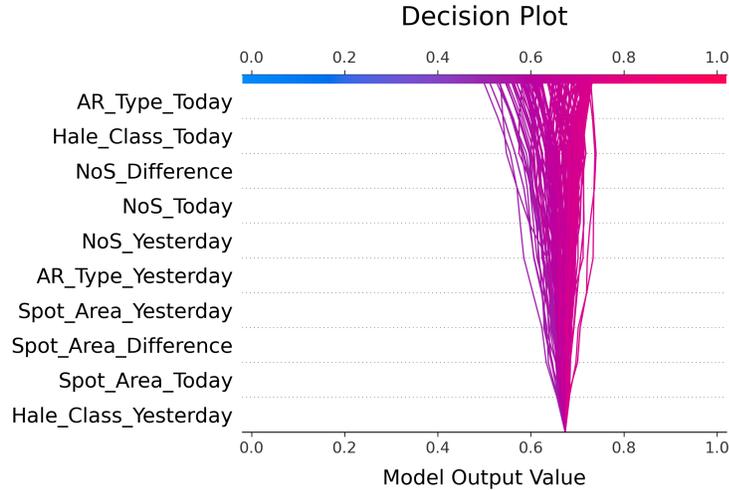
**Figure 3.** Bar plot to display the global importance of each feature on our RF model’s predictions.

feature values have negative SHAP values. High feature values (red dots) tend to lead to positive predictions, as these high feature values have positive SHAP values.

Next, focus on the Hale\_Class\_Yesterday feature in Figure 2. If we compare this feature with the NoS\_Today feature, this feature may appear to have fewer dots. The reason for this appearance is that most of the SHAP values for Hale\_Class\_Yesterday are zero or near zero. This causes the dots to cluster up and makes Hale\_Class\_Yesterday seem to have fewer dots. Furthermore, since most of the SHAP values for the feature are zero or near zero, the Hale\_Class\_Yesterday feature neither pushes the model’s predictions towards a negative class nor pushes the model’s predictions towards a positive class, implying that the Hale\_Class\_Yesterday feature plays an unimportant role in the model’s predictions.

Figure 3 presents the bar plot for the RF model. In the bar plot, each feature is given the mean of the absolute SHAP values across all test samples. This mean of the absolute SHAP values is how the importance of each feature is measured in the bar plot. The longer the bar in the bar plot, the more important the corresponding feature is to the model’s predictions. We can see in Figure 3 that the AR\_Type\_Today feature is of the highest importance to the model’s predictions, while the Hale\_Class\_Yesterday feature is of the lowest importance to the model’s predictions.

In contrast to the bar plot in Figure 3, the beeswarm plot in Figure 2 displays a separate SHAP value for each test sample, and shows the variability of the test samples, represented by color dots, for each feature. The fewer clusters and the more spread out from the zero in the beeswarm plot indicate higher SHAP values (positive or negative), causing a higher mean of absolute SHAP values and ultimately a higher importance. Now, consider again the appearance of Hale\_Class\_Yesterday in the beeswarm plot in Figure 2. The result of having most of its SHAP values equal or near zero reflects how short its bar is in the bar plot in Figure 3. On the other hand, the nature of high variability of the AR\_Type\_Today feature in the beeswarm plot in Figure 2 causes its bar to be significantly longer than those of the other features in the bar plot in Figure 3.



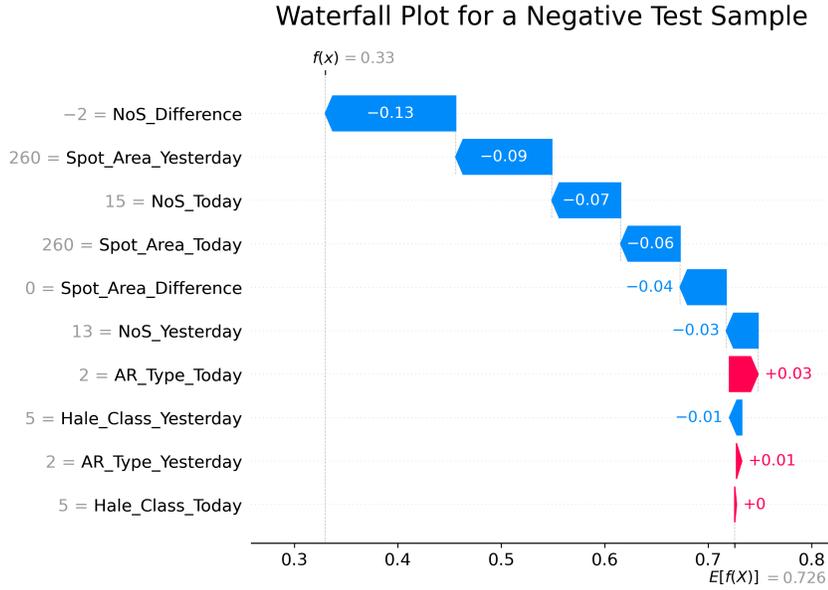
**Figure 4.** Decision plot to understand how our RF model produces its predictions.

Figure 4 presents the decision plot for our RF model. On the  $y$ -axis of the decision plot, the features are displayed, from top to bottom, according to their importance, with the most important feature displayed at the top and the least important feature displayed at the bottom. Each prediction/test sample is represented by a line in the decision plot. The predictions for the test samples begin at the bottom with the same base value, which is approximately 0.67. This base value is the mean of the model’s prediction values over all training samples in the training set. It is used as a starting point before considering any feature contribution. As a prediction/line moves from bottom to top, the SHAP value for each feature is added to the base value. This can help to understand the contribution of each feature in the prediction. The features can have positive or negative contributions, pushing the corresponding line to the right or left, respectively. The  $x$ -axis shows the model output values of the test samples in the test set. The prediction value (i.e., the model output value) of a test sample is the probability that the test sample belongs to the positive class. For example, a prediction value of 0.6 indicates that there is a 60% chance that the test sample belongs to the positive class. At the top, each prediction/line reaches its final prediction value (predicted probability). The prediction value determines the color of the corresponding line. A blue/purple line indicates a lower prediction value (lower predicted probability closer to 0), while a red line indicates a higher prediction value (higher predicted probability closer to 1).

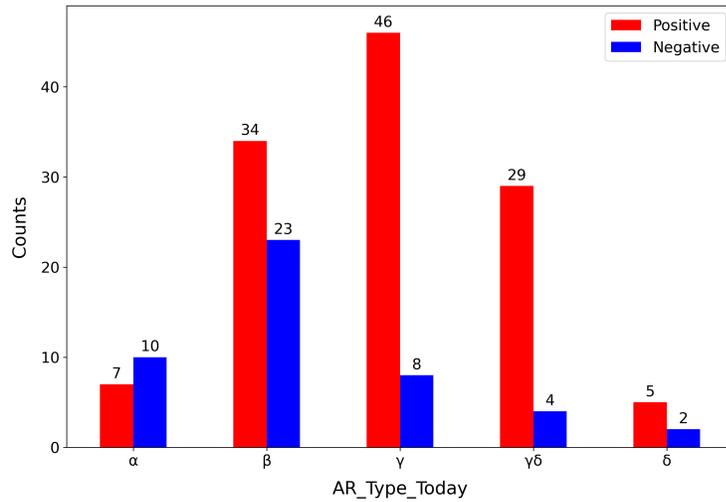
The results shown in the beeswarm, bar, and decision plots are consistent. Based on these results, we conclude that AR\_Type\_Today and Hale\_Class.Today play the most important roles in the model’s decision-making process. The difference between the number of sunspots of an AR in today and the number of sunspots of the AR in yesterday, i.e., NoS\_Difference, also plays a very important role in the model’s predictions. The three plots (beeswarm, bar, and decision) are used mainly for global interpretation based on the entire test set. In what follows, we will present waterfall plots for local interpretation, which look at individual predictions made by our RF model. With waterfall plots, we can better understand each feature and why certain behaviors lead to a specific prediction.

Figure 5 presents the waterfall plot for a test sample predicted to be in the positive class. Figure 6 presents the waterfall plot for a test sample predicted to be in the negative class. These plots display the relative contributions of the different features in order of importance. In Figure 5, we see that the three most important features

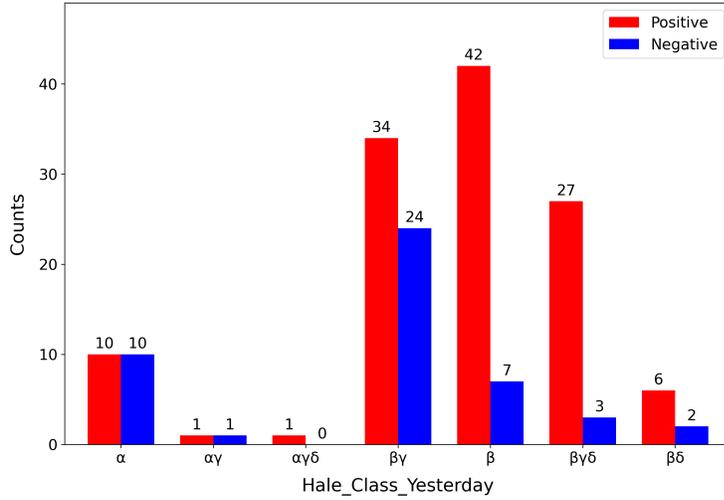




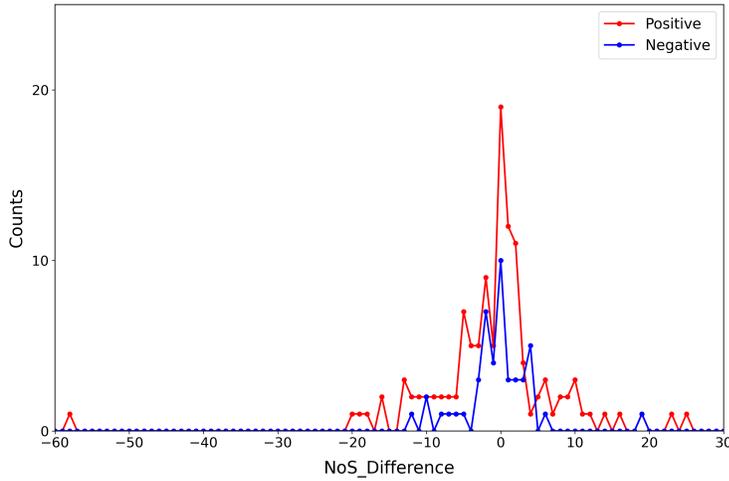
**Figure 6.** Waterfall plot for a test sample predicted to be negative. The plot shows the relative contribution of each feature to the model’s prediction  $f(x) = 0.33$ , starting from the base value  $E[f(x)] = 0.726$ .



**Figure 7.** Number of test samples for the varying feature values (categorical values) of AR\_Type\_Today obtained based on the 121 positive and 47 negative samples in the test set. For each feature value  $v$ , there is a clear difference between the number of positive test samples with the feature value  $v$  and the number of negative test samples with the feature value  $v$ . This AR\_Type\_Today is the most important feature among the 10 physical features considered in this study. Our RF model would prefer to use this feature for flare classification.



**Figure 8.** Number of test samples for the varying feature values (categorical values) of Hale\_Class\_Yesterday obtained based on the 121 positive and 47 negative samples in the test set. No test sample has  $\alpha\delta$ , and hence this feature value is not shown in the figure. Compared to AR\_Type\_Today in Figure 7, there are 3 features ( $\alpha$ ,  $\alpha\gamma$ ,  $\alpha\gamma\delta$ ) based on which it is hard to distinguish between positive test samples and negative test samples. This Hale\_Class\_Yesterday is the least important feature among the 10 physical features considered in this study. Our RF model would prefer not to use this feature for flare classification.



**Figure 9.** Number of test samples for the varying feature values of NoS\_Difference obtained based on the 121 positive and 47 negative samples in the test set. The feature values are numerical values, ranging from  $-58$  to  $25$ . Like AR\_Type\_Today in Figure 7, it is relatively easy to distinguish between positive test samples and negative test samples based on the feature values. This NoS\_Difference is of high importance in both global and local interpretations. Our RF model would prefer to use this feature for flare classification.

8) as the slightest factor, the  $\beta$ ,  $\beta\gamma$  and  $\beta\gamma\delta$  classes are effective over 85% cases in  $\geq C$  predictions, while they are 72% dominant in  $< C$  predictions.

More than 90% of the 121 positive test samples (i.e. in 109 samples) have a NoS\_Difference feature value between  $-13$  and  $10$  as seen in Figure 9. The number of sunspots in the ARs was found to be unchanged for 19 cases. In more than 45% of the samples, the number of sunspots increased, while in 39%, the number of sunspots decreased. In the  $< C$  (negative) predictions, the NoS\_Difference feature takes values between  $-3$  and  $4$  in nearly 81% of the data. In 10 out of 47 samples, the number of spots of ARs remained constant, while it increased in 45% and decreased in 34% of the negative samples.

## 6 Conclusions

In this paper, we employ an interpretable machine learning approach to reveal the relationship between solar flares and their source active regions through a binary classification; in other words, one class is considered positive ( $\geq C$  class) and the other class would be negative ( $< C$  class). We seek to identify the physical features of a source active region that significantly influence the likelihood of its generation of  $\geq C$ -class (i.e. C, M or X classes) flares. We used the Random Forest algorithm together with the SHapley Additive exPlanations (SHAP) method. The performance metric values of recall, precision, accuracy, and the F1 score were obtained as 0.81, 0.82, 0.74, and 0.82, respectively, in our model.

SHAP decision plots for all test data provide a comprehensive view of our model's behavior, allowing us to gain deeper insights into its decision-making process for each test sample. Today's value of AR type is obtained as the most influential, while yesterday's value of Hale class is found to be the least effective in this binary classification (Figures 2 - 4). According to a widely accepted hypothesis in the literature, as the number of sunspots increases, the flare intensity also increases. In our study, as in Sauder (2004), the number of sunspots is not as important as stated. However, we realize that the difference between the number of spots in the AR compared to the previous day is a crucial feature in the decision-making process for positive and negative test samples, as seen in the waterfall plots (Figures 5 and 6). When Figure 9 was carefully analyzed, it was found that this range of difference is between  $-13$  and  $10$  in 90% of the positive class and between  $-4$  and  $3$  in 81% of the negative class. We note that ARs of the types  $\alpha$  and  $\beta$  can be grouped as simple ARs, while other types were classified as ARs of medium and high complexity. Eren et al. (2016) reported that 79% of the flares were generated by medium and high complexity AR groups in their study without binary classification. In the test sample (Figure 7) of our study, it is found around 56%.

Our study reveals that (Figure 7) 34% and 66% of positive test samples are formed by simple and complex ARs, respectively. In negative test samples, these rates are 70% and 30% for simple structured and more complex ARs, respectively. In their study on the relationship between magnetic disturbances and flares in active regions, Oloketuyi et al. (2023) reached results similar to ours. According to their results, 45% of the flares in the positive class are formed by ARs with simple structure, while 55% are formed by ARs with complex structure. For flares in the negative class, these rates are found to be 71% and 29% for simple and complex ARs, respectively. In Yang et al. (2017), it was reported that 46% of  $\geq C$  class flares were caused by simple ARs while 54% were caused by more complex ARs in their study of the statistical relationships between flare and AR properties. For  $< C$  class flares, they found that simple structured ARs are 60% effective, while for complex structured ARs they gave a rate of 40%. As can be seen, it would not be erroneous to conclude that the results of our study, in which a binary classification is defined as positive versus negative, are in line with the rates presented in the studies investigating the relationship between AR complexity and flares in the literature.

Our results indicate that the  $\beta$ ,  $\gamma$  and  $\gamma\delta$  types contribute to  $\geq C$  and  $< C$  classifications with more than 90% and 74% effectiveness, respectively. This result is in good agreement with the result obtained in Figure 2 of Sammis et al. (2000). They found that flares having X-ray fluxes from  $10^{-6} Wm^{-2}$  to  $10^{-4} Wm^{-2}$  (i.e. from C to X classes) are mostly produced by ARs having bipolar sunspot groups or more complex structures (namely ARs of  $\beta$  or having higher complexity). Although our study does not show that the surface area of ARs is effective, in a review article by Toriumi & Wang (2019), it is stated that the probability of flare eruption is proportional to the spot area and increases with the spot complexity ( $\beta$ ,  $\gamma$  and  $\gamma\delta$ ) as found in the current study.

However, the causative mechanism for these eruptions remains unidentified. Consequently, traditional solar flare forecast fundamentally relies on the statistical correlation between solar flares and AR features derived from observational data. In this work, an interpretable machine learning approach was employed to elucidate the association between solar flares and their originating ARs using binary classification. The directions of future research include the following.

- Employing a broader spectrum of data, particularly examining data that encompass observational tools that have advanced in recent years, can reveal new findings.
- Integrating data from contemporary technologies, such as sophisticated machine learning analysis tools, through a more diversified data set might yield innovative insights and improve predictive accuracy.
- Examining data over prolonged durations can uncover enduring trends and patterns that may not be evident in shorter datasets. This can aid in comprehending the evolution of the phenomenon and its possible implications.

## Data Availability Statement

The 10 physical features used in this study are taken from the Solar Monitor database (<https://www.solarmonitor.org/>). The GOES solar flare catalog with flare class information is taken from XRT Flare Catalog ([https://xrt.cfa.harvard.edu/flare\\_catalog/](https://xrt.cfa.harvard.edu/flare_catalog/)).

## References

- Abduallah, Y., Wang, J. T. L., Wang, H., & Xu, Y. (2023). Operational prediction of solar flares using a transformer-based framework. *Scientific Reports*, *13*, 13665. doi: 10.1038/s41598-023-40884-1
- Abramenko, V. I. (2005). Relationship between magnetic power spectrum and flare productivity in solar active regions. *The Astrophysical Journal*, *629*(2), 1141-1149. doi: 10.1086/431732
- Andrews, M. D. (2003). A search for CMEs associated with big flares. *Solar Physics*, *218*(1), 261-279. doi: 10.1023/B:SOLA.0000013039.69550.bf
- Bai, T., & Sturrock, P. A. (1989). Classification of solar flares. *Annual Review of Astronomy and Astrophysics*, *27*, 421-467. doi: 10.1146/annurev.aa.27.090189.002225
- Boudjada, M. Y., Schwingenschuh, K., Al-Haddad, E., Parrot, M., Galopeau, P. H. M., Besser, B., ... Voller, W. (2012). Effects of solar and geomagnetic activities on the sub-ionospheric very low frequency transmitter signals received by the DEMETER micro-satellite. *Annals of Geophysics*, *55*(1), 49-55.
- Donahue, K. P., & Inceoglu, F. (2024). Forecasting solar flares with a transformer network. *Frontiers in Astronomy and Space Sciences*, *10*, 1298609. doi: 10.3389/fspas.2023.1298609
- Eren, S., Kilcik, A., Atay, T., Miteva, R., Yurchyshyn, V., Rozelot, J. P., & Ozguc, A. (2016). Flare-production potential associated with different sunspot groups.

- Monthly Notices of the Royal Astronomical Society*, 465(1), 68-75. Retrieved from <https://doi.org/10.1093/mnras/stw2742> doi: 10.1093/mnras/stw2742
- Falconer, D. A., Moore, R. L., Barghouty, A. F., & Khazanov, I. (2012). Prior flaring as a complement to free magnetic energy for forecasting solar eruptions. *The Astrophysical Journal*, 757(1), 32. Retrieved from <https://dx.doi.org/10.1088/0004-637X/757/1/32> doi: 10.1088/0004-637X/757/1/32
- Friedman, H. (1963). Ultraviolet and X rays from the Sun. *Annual Review of Astronomy and Astrophysics*, 1, 59. doi: 10.1146/annurev.aa.01.090163.000423
- Gallagher, P. T., Moon, Y. J., & Wang, H. (2002). Active-region monitoring and flare forecasting I. Data processing and first results. *Solar Physics*, 209(1), 171-183. doi: 10.1023/A:1020950221179
- Guo, J., Zhang, H. Q., Chumak, O. V., & Lin, J. B. (2010). The evolution of the quantified magnetic complexity of active regions in the 23rd solar cycle. *Monthly Notices of the Royal Astronomical Society*, 405(1), 111-114. Retrieved from <https://doi.org/10.1111/j.1365-2966.2010.16465.x> doi: 10.1111/j.1365-2966.2010.16465.x
- Harrison, R. A. (1995). The nature of solar flares associated with coronal mass ejection. *Astronomy & Astrophysics*, 304, 585.
- Huang, X., Wang, H., Xu, L., Liu, J., Li, R., & Dai, X. (2018). Deep learning based solar flare forecasting model. I. Results for line-of-sight magnetograms. *The Astrophysical Journal*, 856(1), 7. doi: 10.3847/1538-4357/aaae00
- Jaeggli, S. A., & Norton, A. A. (2016). The magnetic classification of solar active regions 1992-2015. *The Astrophysical Journal Letters*, 820(1), L11. doi: 10.3847/2041-8205/820/1/L11
- Kusano, K., Bamba, Y., Yamamoto, T. T., Iida, Y., Toriumi, S., & Asai, A. (2012). Magnetic field structures triggering solar flares and coronal mass ejections. *The Astrophysical Journal*, 760(1), 31. doi: 10.1088/0004-637X/760/1/31
- Li, X., Zheng, Y., Wang, X., & Wang, L. (2020). Predicting solar flares using a novel deep convolutional neural network. *The Astrophysical Journal*, 891(1), 10. doi: 10.3847/1538-4357/ab6d04
- Liu, C., Deng, N., Wang, J. T. L., & Wang, H. (2017). Predicting solar flares using SDO/HMI vector magnetic data products and the random forest algorithm. *The Astrophysical Journal*, 843(2), 104. doi: 10.3847/1538-4357/aa789b
- Liu, H., Liu, C., Wang, J. T. L., & Wang, H. (2019). Predicting solar flares using a long short-term memory network. *The Astrophysical Journal*, 877(2), 121. doi: 10.3847/1538-4357/ab1b3c
- Lundberg, S. M., & Lee, S. (2017). A unified approach to interpreting model predictions. In I. Guyon et al. (Eds.), *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017* (pp. 4765-4774). Retrieved from <https://proceedings.neurips.cc/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html>
- Miteva, R., & Samwel, S. W. (2022). M-class solar flares in solar cycles 23 and 24: Properties and space weather relevance. *Universe*, 8(1), 39. doi: 10.3390/universe8010039
- Mohamed, W., Yousef, S., & Shaltout, M. (2018). On the classifications of the solar active regions (ars); case study for solar cycles 23 and 24. *The Egyptian Journal of Remote Sensing and Space Science*, 21(3), 249-253. Retrieved from <https://www.sciencedirect.com/science/article/pii/S1110982316301296> doi: <https://doi.org/10.1016/j.ejrs.2017.04.004>
- Nishizuka, N., Sugiura, K., Kubo, Y., Den, M., & Ishii, M. (2018). Deep flare net (DeFN) model for solar flare prediction. *The Astrophysical Journal*, 858(2), 113. doi: 10.3847/1538-4357/aab9a7
- Oloketuyi, J., Liu, Y., & Elmhamdi, A. (2023). Investigating the associations between solar flares and magnetic complexity of active regions. *New As-*

- tronomy*, 100, 101972. Retrieved from <https://www.sciencedirect.com/science/article/pii/S1384107622001543> doi: <https://doi.org/10.1016/j.newast.2022.101972>
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... Duchesnay, E. (2011). Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.*, 12, 2825–2830. doi: <https://dl.acm.org/doi/10.5555/1953048.2078195>
- Priest, E. (2014). *Magnetohydrodynamics of the Sun*. Cambridge University Press.
- Sammis, I., Tang, F., & Zirin, H. (2000). The dependence of large flare occurrence on the magnetic structure of sunspots. *The Astrophysical Journal*, 540(1), 583. Retrieved from <https://dx.doi.org/10.1086/309303> doi: 10.1086/309303
- Sauder, E. (2004). The relationship between sunspots and solar flares. In *American Astronomical Society Meeting Abstracts* (Vol. 205, p. 10.02).
- Sun, P., Dai, W., Ding, W., Feng, S., Cui, Y., Liang, B., ... Yang, Y. (2022). Solar flare forecast using 3D convolutional neural networks. *The Astrophysical Journal*, 941(1), 1. doi: 10.3847/1538-4357/ac9e53
- Tang, R., Liao, W., Chen, Z., Zeng, X., Wang, J.-s., Luo, B., ... Wu, Z. (2021). Solar flare prediction based on the fusion of multiple deep-learning models. *The Astrophysical Journal Supplement Series*, 257(2), 50. doi: 10.3847/1538-4365/ac249e
- Toriumi, S., & Wang, H. (2019). Flare-productive active regions. *Living Reviews in Solar Physics*, 16(1), 3. doi: 10.1007/s41116-019-0019-7
- Vemareddy, P., & Mishra, W. (2015). A full study on the Sun-Earth connection of an Earth-directed CME magnetic flux rope. *The Astrophysical Journal*, 814(1), 59. doi: 10.1088/0004-637X/814/1/59
- Vysakh, P. A., & Mayank, P. (2023). Solar flare prediction and feature selection using a light-gradient-boosting machine algorithm. *Solar Physics*, 298(11), 137. doi: 10.1007/s11207-023-02223-5
- Yang, Y.-H., Hsieh, M.-S., Yu, H.-S., & Chen, P. F. (2017). A statistical study of flare productivity associated with sunspot properties in different magnetic types of active regions. *The Astrophysical Journal*, 834(2), 150. Retrieved from <https://dx.doi.org/10.3847/1538-4357/834/2/150> doi: 10.3847/1538-4357/834/2/150
- Yashiro, S., Gopalswamy, N., Akiyama, S., Michalek, G., & Howard, R. A. (2005). Visibility of coronal mass ejections as a function of flare location and intensity. *Journal of Geophysical Research (Space Physics)*, 110(A12), A12S05. doi: 10.1029/2005JA011151
- Youssef, M. (2012). On the relation between the CMEs and the solar flares. *NRIAG Journal of Astronomy and Geophysics*, 1(2), 172-178. doi: 10.1016/j.nrjag.2012.12.014
- Zheng, Y., Li, X., Yan, S., Huang, X., Lou, H., & Li, Z. (2023). Multiclass solar flare forecasting models with different deep learning algorithms. *Monthly Notices of the Royal Astronomical Society*, 521(4), 5384-5399. doi: 10.1093/mnras/stad839