A Comparative Analysis of Machine Learning Models for Solar Flare Forecasting: Identifying High Performing Active Region Flare Indicators

Suvadip Sinha,[1] Om Gupta,[1,2] Vishal Singh,[1,2] B. Lekshmi,[1,3] Dibyendu Nandy*,[1,2] Dhrubaditya Mitra,[4] Saikat Chatterjee,[5] Sourangshu Bhattacharya,[6] Saptarshi Chatterjee,[6] Nandita Srivastava,[7,1] and Axel Brandenburg[4,8]

[1]Center of Excellence in Space Sciences India, Indian Institute of Science Education and Research Kolkata, Mohanpur 741246, West Bengal, India
[2]Department of Physical Sciences, Indian Institute of Science Education and Research Kolkata, Mohanpur 741246, West Bengal, India
[3]Max Planck Institute for Solar System Research, Göttingen, Germany 37077
[4]Nordita, KTH Royal Institute of Technology and Stockholm University, Hannes Alfvéns väg 12, SE-10691 Stockholm, Sweden
[5]KTH Royal Institute of Technology, Stockholm
[6]Indian Institute of Technology, Kharagpur, India
[7]Udaipur Solar Observatory, Physical Research Laboratory, P.O. Box 198, Badi Road, Udaipur 313001, India
[8]The Oskar Klein Centre, Department of Astronomy, Stockholm University, AlbaNova, SE-10691 Stockholm, Sweden

ABSTRACT

Solar flares create adverse space weather impacting space and Earth-based technologies. However, the difficulty of forecasting flares, and by extension severe space weather, is accentuated by the lack of any unique flare trigger or a single physical pathway. Studies indicate that multiple physical properties contribute to active region flare potential, compounding the challenge. Recent developments in Machine Learning (ML) have enabled analysis of higher dimensional data leading to increasingly better flare forecasting techniques. However, consensus on high-performing flare predictors remains elusive. In the most comprehensive study till date, we conduct a comparative analysis of four popular ML techniques (K-Nearest Neighbor, Logistic Regression, Random Forest Classifier, and Support Vector Machine) by training these on magnetic parameters obtained from the Helioseismic Magnetic Imager (HMI) onboard the Solar Dynamics Observatory (SDO) during the entirety of solar cycle 24. We demonstrate that Logistic Regression and Support Vector Machine algorithms perform extremely well in forecasting active region flaring potential. The logistic regression algorithm returns the highest true skill score of 0.967 ± 0.018, possibly the highest classification performance achieved with any parametric study alone. From a comparative assessment, we establish that the magnetic properties like total current helicity, total vertical current density, total unsigned flux, R_value, and total absolute twist are the top-performing flare indicators. We also introduce and analyze two new performance metrics, namely, Severe and Clear Space Weather indicators. Our analysis constrains the most successful ML algorithms and identifies physical parameters that contribute most to active region flare productivity.

1. INTRODUCTION

Solar flares are sudden bursts of electromagnetic radiation from the solar atmosphere, mainly in the extreme ultraviolet and X-ray regime. They are classified into different categories based on the peak X-ray flux recorded in the 1–8 Å band by the Geostationary Operational Environmental Satellite (GOES). The X-class flares are the most powerful with peak flux

Corresponding author: Dibyendu Nandy
dnandi@iiserkol.ac.in

$\geq 10^{-4}\,\mathrm{W\,m^{-2}}$, followed by M-class flares with peak flux $\geq 10^{-5}\,\mathrm{W\,m^{-2}}$. These classes of flares strongly influence the near-Earth space weather and present a bigger potential hazard to human space endeavors than flares with lower peak intensities, which, in decreasing order of intensities, belong to C, B, and A classes, respectively.

From previous studies we know that solar flares originate in active region (AR) structures, where the magnetic flux system becomes energized due to rapid flux emergence, instability, or topological changes of the magnetic configuration via reconnection processes (Forbes 2000; Priest & Forbes 2002; Schrijver 2007; Leka & Barnes 2003a,b; Nandy et al. 2003; Hahn et al. 2005;

Jing et al. 2006). A solar active region with a potential or near-potential magnetic field builds up magnetic non-potential energy (or free magnetic energy) upon being sheared and twisted. A fraction of this free energy is dissipated during a flare event (e.g., Schrijver et al. 2008), and a typical large solar flare can release large quantities of energy of the order of 10^{32}–10^{33} erg. Simultaneously, solar energetic particles are also released into the solar wind. Solar flares are often accompanied by Coronal Mass Ejections (CMEs), which pose serious threats if directed towards the Earth. Earlier studies have shown that magnetic characteristics of active regions (Yeates et al. 2010; Pal et al. 2018, 2017) and filaments determine their propensity to flare and produce associated CMEs (Sinha et al. 2019).

Solar flares (and CMEs) induce extreme space weather conditions that have the potential to harm satellites and impact communication and navigation sectors. The most energetic solar flares can cause electric power grid failures, radio communication blackouts, impact astronaut health, and expose air passengers to harmful doses of radiation (Eastwood et al. 2017; Schrijver 2015; Schrijver et al. 2015; Hapgood 2011). Proactive measures to mitigate the physical and economic impact of space weather are therefore much sought after, of which early warning systems are of foremost interest. While physical model based studies have demonstrated strong potential of success in recent times towards predicting long-term solar activity variations over decadal timescales (Bhowmik & Nandy 2018; Nandy 2021; Nandy et al. 2021), physical model based assessment of active region flaring probability remains elusive.

The creation of knowledge towards predicting solar flares initiated with statistical approaches applied on observational data well before machine learning techniques found favour. In a set of pioneering studies with vector magnetogram data Leka & Barnes (2003b) and Barnes et al. (2007) conducted a multi-parametric statistical study to classify between flaring and flare-quiet active regions based on discriminant analysis.

One of the early applications of machine learning in solar physics was the automatic real-time detection of solar flares from Hα images (e.g., Fernandez Borda et al. 2002; Qu et al. 2003). Very soon, efforts were directed towards the forecast of solar flares. A number of machine learning methods were trained on sunspot-associated data to forecast solar flares (Li et al. 2007; Qahwaji & Colak 2007; Benvenuto et al. 2018; Cinto et al. 2020). Colak & Qahwaji (2009) used neural networks to make multi-class forecasts based on sunspot area and McIntosh classification data. Line-of-sight full-disk magnetogram data from the Solar and Heliospheric Observa-

tory's (SOHO) Michelson Doppler Imager (MDI) presented the next opportunity in the development of solar flare forecasting methods and several advances were made by using features calculated from them. Decision tree classifiers, Learning Vector Quantization, Ordinal Logistic regression, Support Vector Machine (SVM), and AdaBoost methods were experimented with by Yu et al. (2009), Song et al. (2009), Yuan et al. (2010), Huang et al. (2010) and Lan et al. (2012). Ahmed et al. (2013) applied Cascade-Correlation Neural Network and used feature evaluation algorithms to remove redundant features and show that a smaller set of parameters yielded comparable results to the entire set. Huang et al. (2018) combined the line-of-sight magnetograms from MDI and the Solar Dynamics Observatory's (SDO) Helioseismic and Magnetic Imager (HMI) to create an extensive dataset and evaluated the performance of a Convolutional Neural Network on these data.

Following the launch of SDO in 2010, its HMI instrument (Scherrer et al. 2012) started providing one of the most advanced unhindered full-disk vector magnetograms. To facilitate active region based event forecasting, the Spaceweather HMI Active Region Patch (SHARP) data product (Bobra et al. 2014) provides cutouts of automatically tracked magnetic flux concentrations on the solar disc. Using SHARP data, Bobra & Couvidat (2015) implemented an SVM algorithm to distinguish between Active Regions producing an M or X class flare (in the next 24 hours) and those not producing any flare or low intensity flares. Bobra & Couvidat (2015) presented a significant improvement in the performance of AR-parameter based machine learning algorithms, primarily because of the availability of continuous, high-quality HMI vector-magnetogram data to derive input magnetic features.

These recent advances piqued the interest of both solar physics and computer science communities heralding a close interdisciplinary collaboration in solar flare forecasting. Liu et al. (2017) attempted a multi-class classification using Random Forest, Nishizuka et al. (2017) and Florios et al. (2018) compared various ML algorithms which includes SVM, Multi-Layer Perceptrons, Random Forest and k-Nearest Neighbors (KNN) algorithm, Nishizuka et al. (2018, 2020) trained a deep Neural Network for binary classification, and Campi et al. (2019) used the Hybrid LASSO and the Random Forest algorithms on features derived during the FLARECAST project. In a recent study, Ribeiro & Gradvohl (2021) used LightGBM for flare forecasting and showed a nice comparison with existing ML models. Classification using KNN was attempted by Hamdi et al. (2017) for uni-

variate time series and by Filali Boubrahimi & Angryk (2018) for multivariate time series. Decision trees were used by Ma et al. (2017) for multivariate time series. Liu et al. (2019) implemented time series classification by training Long Short-Term Memory (LSTM) Neural Networks on SHARP features and flare history parameters. A similar approach was followed by Jiao et al. (2020) who built classification models on an LSTM regressor. Chen et al. (2019) compared LSTM models trained on SHARP parameters and autoencoder-derived features. Using wavelet analysis and features derived from SDO HMI magnetograms, Support Vector Regression was applied to forecast the X-ray flux by Muranushi et al. (2015) and Boucheron et al. (2015), while Al-Ghraibah et al. (2015) attempted classification using Relevance Vector Machines. Zernike moments calculated from images were also used for binary classification with SVM (Raboonik et al. 2017; Alipour et al. 2019). Strong-field high-gradient Polarity Inversion Line (PIL) features derived from SHARP images were used by Sadykov & Kosovichev (2017) for classification comparing SVM and a graphical method, while Wang et al. (2019) used SHARP parameters weighed with a PIL mask to improve individual parameter performance on a Random Forest Classifier. Dhuri et al. (2019) and Hazra et al. (2020) studied the time evolution of various magnetic parameters and the correlations between them. They trained and tested Logistic Regression, SVM, Gradient Boost, Random Forest, Multilayer Perceptron, KNN and Naïve Bayes classifier on SHARP feature data with good performance.

With rapid developments in the field of Machine Learning (ML) and image processing, it became possible to process images directly using Convolutional Neural Networks (CNNs). Jonas et al. (2018) used vector magnetic field data from HMI as well as multi-wavelength image data of the chromosphere, transition region and corona to train a single-layer CNN, and obtained results comparable to Bobra & Couvidat (2015). Zheng et al. (2019), Li et al. (2020) and Bhattacharjee et al. (2020) used line-of-sight magnetograms to train deep CNN models.

The underlying non-unique and non-deterministic nature of the triggering mechanisms without well defined parametric thresholds make flare forecasting a challenging task making the problem suitable for multi-parametric statistical approaches and computational ML algorithms applied to large databases. Attempts to supplement vector magnetogram data with extreme ultraviolet images have not yielded significant improvement. On comparing CNN models trained with and without multi-wavelength image data from the Atmo-

spheric Imaging Assembly (AIA) on board the SDO, Jonas et al. (2018) found that the best performing model was the one not provided with AIA data as input. Similarly, the implementation of CNN networks have to be developed further for application on flare forecasting. Bhattacharjee et al. (2020) found that the CNN output had spurious dependencies on the magnetogram dimensions.

We limit our comparative analysis to well-studied and successful machine learning algorithms (limited to parametric approach alone for efficiency) to determine their relative performance. This is achieved by applying these algorithms to the largest, single-instrument database suitable for this purpose, i.e., the HMI vector magnetic field observations.

Over the last two decades a wide range of machine learning algorithms have been applied to forecast solar flares. The input data to such algorithms are, most commonly, several AR magnetic parameters derived from magnetograms, magnetogram images and time series data of magnetic parameters. These works have been able to achieve reasonably accurate forecast of whether an active region is going to flare or not and if it does, in which class the flare lies. Furthermore, these works have attempted to extract which magnetic parameters are best correlated with flaring. In general the obtained results are independent of the algorithms used – unsigned current density, unsigned flux and current helicity came up as key parameters in most of the previous studies. However, their relative ranking in terms of which contribute most to the flare potential has not been rigorously explored.

In this paper, we compare several machine-learning algorithms to find out which offers the best flare forecasting capability. Bobra & Couvidat (2015) and subsequent studies have shown that the magnetic twist parameter in the SHARP database does not perform well for machine classification whereas earlier physics-based works suggests twist is a flare-relevant parameter (Linton et al. 1996; Nandy et al. 2003; Hahn et al. 2005; Nandy 2008). We therefore include a global indicator of magnetic twist in our analysis. Furthermore we introduce two new performance metrics, Severe Space Weather and Clear Space Weather indicators, to distinguish between these two equally important conditions. Our analysis, detailed in followings sections, is based on the highest number of unique active regions to date used in training ML algorithms, from the beginning of the SDO era to December 2020, covering the entire solar cycle 24.

2. DATA SELECTION

Depending on the intensities, flares are categorized into five classes: A, B, C, M and X in the ascending order of intensities. In this study, all the active region information are collected from *hmi.sharp_cea_720s* series (Bobra et al. 2014) but for a longer time of observation. We build our dataset considering all the ARs that have appeared on the Sun starting from 2010 May to 2020 December. We divide active regions into two groups – Positive or flaring class and Negative or non-flaring class. The Positive class is defined such that it consists of ARs which have produced at least one M or X class flares in their lifetime. In contrast, the Negative class is formed by those active regions which produce only low intensity flares ($\leq$ B class) or did not flare at all.

We use the XRT Flare Catalog based on the Hinode Flare Catalogue (Watanabe et al. 2012), and the GOES flare catalog to collect information of flare events of the past 10 years, such as flare timing, flare intensities, associated active regions (NOAA number) and their positions on the solar disc. GOES flare events are collected from the sunpy.instr.goes module of the python SunPy library (The SunPy Community et al. 2020). Our dataset covers the entire solar cycle 24 starting from May 2010 to December 2020. We find that not all events in the XRT catalog match with the GOES events list because of slight differences in flare peak-times. For each event in the XRT catalog we search for a similar event in GOES catalog within a time window of 4 minutes centered at the flare peak time of the XRT event. If a similar flare event is found in GOES catalog with a same flare class and NOAA number, we call it a match. Pi-chart in Figure 1 depicts the number of matched and unmatched events in the two catalogs. Following this, a total of 668 matched and 80 unmatched events in the XRT catalog is obtained with flare intensity $\geq$ M class. Manual inspection of these 80 unmatched events with more lenient conditions, for example, allowing flare peak-time difference up to 15 minutes, results in a further reduction in the number of unmatched events by 57. The negative dataset is prepared by excluding flare associated active regions from all recorded SHARP regions during our observational time domain.

All the magnetic parameters representing the flaring AR are either collected from SHARP header keywords or calculated from the vector magnetic field data, 24 hours before the flare peak-time. For the negative/non-flaring class, we choose the magnetogram observation at the central-time snap of their entire lifespan on the visible disc. In addition, if an active region's position vector from the sun center makes an angle greater than 70° with the line of sight, we discard that region from our
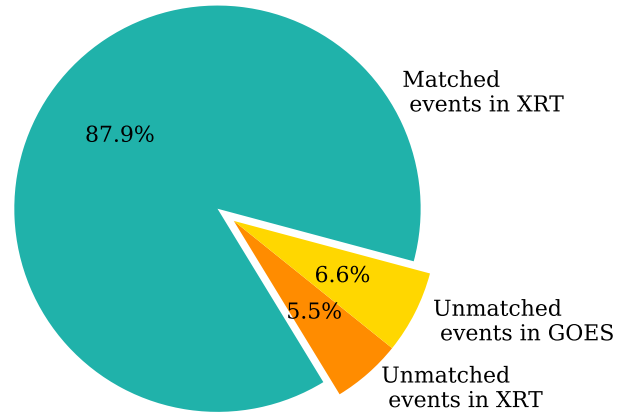

Figure 1. Distribution of matched and unmatched events between XRT and GOES flare catalog

analysis, which is a standard method to avoid high projection effects. We have implemented this 70° positional filter in the last step of our data-preparation process so that it only restricts our domain of analysis, not the domain of our observation. In other words, this ensures that any AR producing an M/X class flare outside this 70° angular region has not been included in our negative class.

Following all these selection criteria, our final positive class contains 503 flaring events and the negative class consists of 3358 non flaring events. Note that in our positive class, recurrent flare events are treated as separate events with different entries. Whereas each non-flaring SHARP region has single entry in the negative class.

3. METHODS AND ANALYSIS

Preparing the input data and proper training of model are the most crucial steps while working with ML algorithms (Ahmadzadeh et al. 2021). SHARP data comes with various magnetic parameters, calculated from the vector-magnetic field maps of ARs. Previous studies showed the importance of these derived parameters in characterizing AR properties and complexities (Hagyard et al. 1984; Leka & Barnes 2003c,d; Georgoulis & Rust 2007; LaBonte et al. 2007; Moore et al. 2012). Bobra & Couvidat (2015) identified 13 such parameters which they found are most useful in describing the flare potential. However, they estimated the magnetic twist using the parameter called MEANALP, whose poor performance led them to exclude any contribution of the magnetic twist from their classification. But, it has been shown that a high twist in the magnetic flux rope can store non-potential magnetic energy and often leads to eruption of the flux rope via the kink instability (Nandy et al. 2008). Motivated

by these physical arguments, we calculate 6 new magnetic parameters related to AR twist, such as TOTAB-STWIST, AVG90PABSTWIST, VTWIST, AVGABST-WIST, AVGTWIST, and MEANALP (see Table 1 for descriptions). We incorporate these 6 twist related features along with those 13 parameters used by Bobra & Couvidat (2015). Here we assume the force–free field approximation for the estimation of magnetic twist – also known as alpha parameter. The vertical twist parameter α_z at each pixel of an AR magnetogram is given as,

$$\alpha_z = \mu_0 \left(\frac{J_z}{B_z} \right), \tag{1}$$

where J_z and B_z are the z-components of current density and magnetic field, respectively, and μ_0 is the permeability of free space. All these 19 features and their description is listed in Table 1. We perform univariate feature selection analysis with the ANOVA Fisher statistics (F-statistics) using the python Scikit-learn library to finalize our set of input magnetic features by eliminating those which are not very useful for this classification. The obtained F-score (see Bobra & Couvidat (2015) for the calculation of F-score) for each feature is represented in Figure 2. It is quite surprising to note that all the magnetic twist related parameters (including MEANALP) are insignificant according to F-statistics except from the TOTABSTWIST, which ranks third in the list. This indicates that flare potential is more coupled to gross/extrinsic quantities rather than magnetic properties at individual pixels. We exclude last five features having the lowest normalized F-scores in Figure 2, i.e., AVG90PABSTWIST, VTWIST, AVGABSTWIST, AVGTWIST and MEANALP (descriptions in Table 1). All further analyses are done with the remaining 14 features which contain 10 SHARP keyword parameters and 4 derived parameters including one newly introduced parameter TOTABSTWIST.

Henceforth, each AR is represented by a single data point in 14-dimensional feature space except those ARs which have produced multiple M/X class flares. The latter are accounted as separate events for each M/X class flares. Our whole dataset, consisting 3861 events, is randomly divided into two groups, training and testing, with a population ratio of 4:1 respectively. We arrange the data such that the ratio of flaring to non-flaring events is the same for both the training set as well as the test set. We pre-process the data by normalizing it such that the processed data has zero mean and unit standard deviation. For this normalization we solely use the training dataset and then apply the same population mean and standard deviation to normalize the test dataset. To make our classification more robust
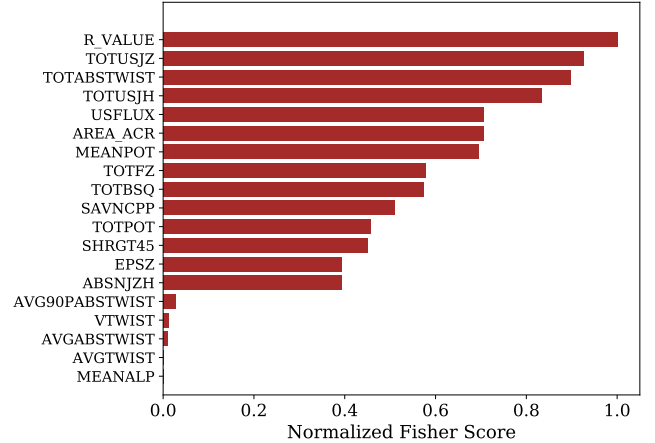

Figure 2. Normalized F-score ranking of the magnetic field parameters.

and independent of any bias, we randomly shuffle our dataset to make 20 similar but differently distributed representative pairs of training and testing data. We denote each such pair with D_i where i is a running index between 1 and 20. Each machine-learning algorithm is evaluated by its average performance over these 20 D_is. The schematic diagram of our analysis method is shown in Figure 3. The following two sections describe how we quantify model performance and compare between different ML algorithms.

3.1. Performance metric

Typically, the performance of a ML model is evaluated from the confusion matrix. It is a 2×2 matrix whose elements are the number of correctly forecasted positive class events (TP), number of correctly forecasted negative class events (TN), number of events falsely forecasted as positive class (FP) and number of events falsely forecasted as negative class (FN). In general, there are various parameters that can be derived from the confusion matrix such as accuracy, recall, f1-score, etc., but their suitability depends on the particular problem. Simple accuracy is defined as the ratio of the number of correct forecast to the total number of forecast. In our data-set, the number of flaring ARs (positive class) is much less than the number of non-flaring (negative class) ARs, which means our data-set is highly imbalanced. Hence we cannot simply use accuracy measure to evaluate the models. To deal with this problem we use Macro Accuracy (MAC)(Pereira et al. 2018) and True Skill Score (TSS) (Woodcock 1976) values to evaluate our ML-models. Bloomfield et al. (2012) showed that TSS is unaffected by class imbalance and

S. No.	Keyword	Description	Normalized F-score
1	R_VALUE	Sum of unsigned flux near polarity inversion line	1.000
2	TOTUSJZ	Total unsigned vertical current	0.927
3	TOTABSTWIST*	Total absolute twist calculated over strong field ($|B| \geq 300\,\mathrm{G}$) regions	0.898
4	TOTUSJH	Total unsigned current helicity	0.833
5	USFLUX	Total unsigned flux	0.707
6	AREA_ACR	Area of strong field pixels in the active region	0.706
7	MEANPOT	Mean photospheric magnetic free energy	0.695
8	TOTFZ*	Sum of z-component of Lorentz force	0.579
9	TOTBSQ*	Total magnitude of Lorentz force	0.573
10	SAVNCPP	Sum of the modulus of the net current per magnetic polarity	0.510
11	TOTPOT	Total photospheric magnetic free energy density	0.456
12	SHRGT45	Fraction of Area with shear $> 45°$	0.451
13	EPSZ*	Sum of z-component of normalized Lorentz force	0.393
14	ABSNJZH	Absolute value of the net current helicity	0.393
15	AVG90PABSTWIST*	Average absolute twist for pixels having twist more than 90 percentile value	0.027
16	VTWIST*	Standard deviation of twist within an AR	0.013
17	AVGABSTWIST*	average absolute value of twist	0.011
18	AVGTWIST*	average value of twist	0.000
19	MEANALP	Mean value of flux weighted twist	0.000

Table 1. Details of AR parameters extracted from SHARP data. Asterisk (*) denotes parameters that are not readily available in the SHARP header keywords and are calculated explicitly from the SHARP vector magnetic field data.
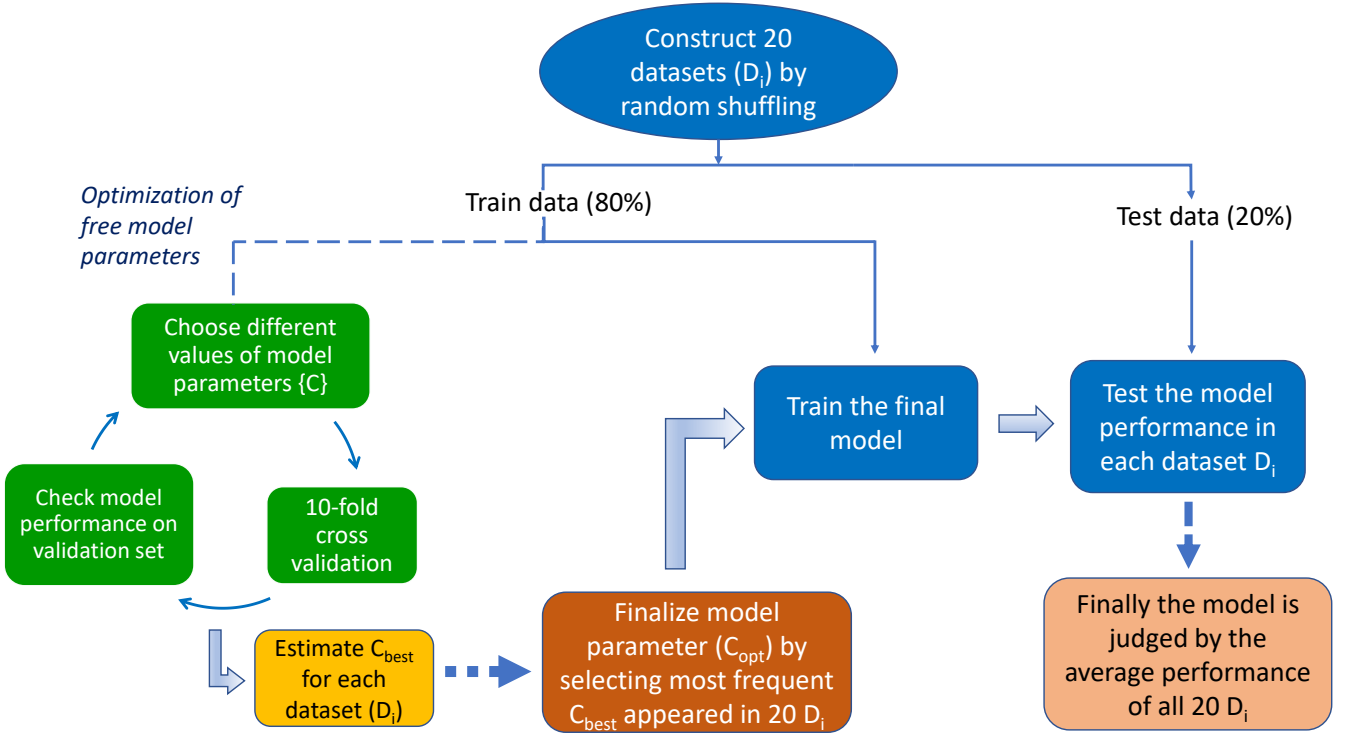

Figure 3. Schematic diagram of our method of analysis.

gives an unbiased result. The MAC and TSS are defined as,

$$\text{MAC} = \frac{1}{2}\left(\frac{\text{TP}}{\text{TP}+\text{FN}} + \frac{\text{TN}}{\text{FP}+\text{TN}}\right), \qquad (2a)$$

$$\text{TSS} = \frac{\text{TP}}{\text{TP}+\text{FN}} - \frac{\text{FP}}{\text{FP}+\text{TN}}. \qquad (2b)$$

The MAC is the average of the accuracy of each individual class, hence its value lies between 0 and 1. On the other hand TSS has two components: the first is the positive class accuracy and the second is the probability of false forecast for the negative class. Using TSS, we penalize the model's performance commensurately by subtracting the false-alarm ratio from the positive class accuracy. This shows its usefulness in the present problem as we are more interested in correctly predicting flaring ARs with minimal amount of false detection. The value of TSS ranges from $+1$ to -1 and we optimize our models to maximize the TSS score.

Depending on the end-user application our priorities for detecting a specific class can change. For example, one may wish to identify only those regions which have the potential to flare with a high degree of confidence without worrying about miss-classifying a non-flaring region as a flaring region. This motivates a new performance indicator which we term as the Sever Space weather (SSW) metric defined as,

$$\text{SSW} = \frac{\text{TP}-\text{FN}}{\text{TP}+\text{FN}}. \qquad (3)$$

Conversely, one may wish to focus on identifying non-flaring regions only with a high degree of confidence. For this we define another parameter called the Clear Space Weather (CSW) metric.

$$\text{CSW} = \frac{\text{TN}-\text{FP}}{\text{TN}+\text{FP}}. \qquad (4)$$

It is important to note that SSW is only linked to flaring class events. It indicates the correct identification ratio combined with a penalty for misidentification within the flaring class. Similarly, the CSW deals with the non-flaring class only. The value of these two metrics lies between -1 and $+1$, where $+1$ indicates perfect identification of all the events within a specific class, whereas -1 indicates the scenario where all events are misclassified. A metric score of 0 denotes the scenario where half of the events in a specific class are correctly identified and the other half are wrongly classified implying no useful classification capability.

Moreover the average value of these two performance metrics returns the TSS and can be demonstrated as follows,

$$\text{SSW} + \text{CSW} = \frac{\text{TP}-\text{FN}}{\text{TP}+\text{FN}} + \frac{\text{TN}-\text{FP}}{\text{TN}+\text{FP}}$$

$$= \frac{\text{TP}}{\text{TP}+\text{FN}} - \frac{\text{FP}}{\text{TN}+\text{FP}} + \frac{\text{TN}}{\text{TN}+\text{FP}} - \frac{\text{FN}}{\text{TP}+\text{FN}}$$

$$= \frac{\text{TP}}{\text{TP}+\text{FN}} - \frac{\text{FP}}{\text{TN}+\text{FP}} + \left(\frac{\text{TN}}{\text{TN}+\text{FP}} - 1\right)$$
$$+ \left(1 - \frac{\text{FN}}{\text{TP}+\text{FN}}\right)$$

$$= 2\text{TSS} \implies \frac{\text{SSW}+\text{CSW}}{2} = \text{TSS}.$$

3.2. *Cross validation*

One of the most important aspect of any machine learning algorithm is the optimization of its hyper-parameters, to achieve best fit on the dataset. If a classifier performs too well on the training dataset, it might fail to capture the overall picture and can badly perform on the test dataset – also known as over-fitting. The optimization of hyper-parameters is done by employing a Grid Search algorithm for finding the optimal hyper-parameters C_{opt} of the training component-set of each D_i. We use 10-fold cross validation on the training dataset of each D_i to avoid the issue of over-fitting. This process divides the training set into 10 groups of equal sample size. Training the model on 9 groups, the validation is done on the 1 remaining group of data points. This happens 10 times such that each data group is made the validation set once. The average validation TSS from this 10-fold cross validation is used to decide the model hyper-parameters for each dataset D_i. We train our models with different values of model hyper-parameters and the optimal values of hyper-parameters C_{opt} are obtained for each D_i by maximizing the average validation TSS. Finally, we choose our operational model hyper-parameter (C_{best}) by selecting the most frequently appearing C_{opt} among these 20 experimental sets D_i. These completes our model optimization process. Now we check performance of the finalized model and judge the model based on the average test performance over 20 D_i.

3.3. *Machine Learning Models*

We use four popular supervised ML algorithms available in the python Scikit-learn library (Pedregosa et al. 2011). These four algorithms are K-Nearest Neighbors (KNN), Logistic regression (LR), Random Forest Classifier (RFC) and Support Vector Machine (SVM); they

are discussed briefly in this section. In all four models, the model parameters are tuned properly to get the best achievable performance, which is TSS in our case. The optimal model parameters of all four models are selected from the histogram plot of Figure 4. A general overview of these ML algorithms can be found in Bishop (2013) or Mehta et al. (2019).

1. KNN: This is an instance-based machine learning technique which uses instances of training data to compute the machine classification based on a simple majority vote of 'k' number of nearest neighbors of each point (Fix & Hodges 1951). When the dataset is not large, as in our case, using the KNN classifier poses no disadvantage as it does not create an internal model which might otherwise use a large memory space. In our model, the weights assigned to each neighbor are equal and the nearest neighbors are calculated using Euclidean distance. The best KNN model is obtained by finding the optimal K, i.e., the optimal value of the number of nearest neighbors to maximize the TSS output. We search for the optimal K value between 1 to 16 in 20 datasets (D_i) to get the maximum TSS. Figure 4(a) shows the histogram plot of optimal K values for 20 different datasets. Since $K = 3$ has the highest number of occurrences, it becomes our final choice.

2. LR: This classifier, also known as the log-linear classifier, is a linear classification model which uses the sigmoid function to classify into discrete categories (Mehta et al. 2019). This makes it extremely suitable for binary classification problems. Our model uses regularized logistic regression, and is implemented using the Logistic Regression classifier available in scikit-learn. The only free parameter of this model, as we have used, is the regularization parameter C, and the most favorable value is estimated from within the range [0.0001, 10000.0], varied with logarithmic increments. In Figure 4(b) we can see that the occurrence is maximum at $C = 1000$, hence we choose 1000 as the optimal C parameter.

3. RFC: This classifier consists of a large number of individual decision tree classifiers which operate as an ensemble (Tin Kam Ho 1995). Each decision tree is trained on a subset of the entire dataset. Generally, decision trees tend to overfit the data and exhibit high variance. Random forests are constructed in such a way so as to decrease the variance. The overall prediction is generated by taking an average of the constituent tree predictions, which tends to cancel out some prediction errors from individual trees. Thus, a large number of uncorrelated trees can produce largely accurate ensemble predictions. Our model uses the RFC available with the scikit-learn package, and the best forest is created by varying the number of trees (also called as 'N-estimator') in the forest from 10 to 1000. We can see from figure 4(c) that occurrence is maximum for $N\text{-}estimator = 120$. Therefore, we select 120 as the optimal value of the number of trees.

4. SVM: It is a powerful classification technique (Cortes & Vapnik 1995), and has previously yielded better results amongst various machine learning models, when applied to solar flare prediction based on SHARP parameters. SVM works by creating a decision boundary, marked by a subset of training points called support vectors, to separate the positive and negative events in the training data. It uses a kernel function to map the data points to higher-dimensional space. Our model uses a Gaussian Radial Basis Function (RBF) as the kernel and assigns the class weight inversely proportional to the class frequency to handle the class imbalance problem. The kernel coefficient gamma (γ) and the regularization parameter C are varied within the ranges [0.0001, 10.0] and [0.001, 100.0] respectively, to get the best SVM model by comparing the TSS scores. The decision bound From Figure 4(d) we can see the optimal values of C and γ are 100 and 0.001 respectively, as this combination produces the highest TSS score in 7 out of 20 D_i.

4. RESULTS

Each ML classifier is trained with the finalized hyperparameters (C_{best}) on the training set of each D_i, and then the trained model is applied on the test set in that D_i. The average and standard deviation of the performance metrics over these 20 test sets are reported in Table 2. We find that all of these models work reasonably well for identifying flaring and non-flaring active regions. The performance of both LR and SVM are very similar, better than that of KNN and RFC. The average TSS of LR and SVM are 0.967 and 0.965 respectively. Therefore, we claim that LR and SVM are equally good in performance. For further analysis we primarily focus on LR because of its marginally higher TSS value. The comparison of the four ML classifiers is depicted in Figure 5. We achieve a remarkable MAC of 0.983 for both LR and SVM. The SSW is also much higher
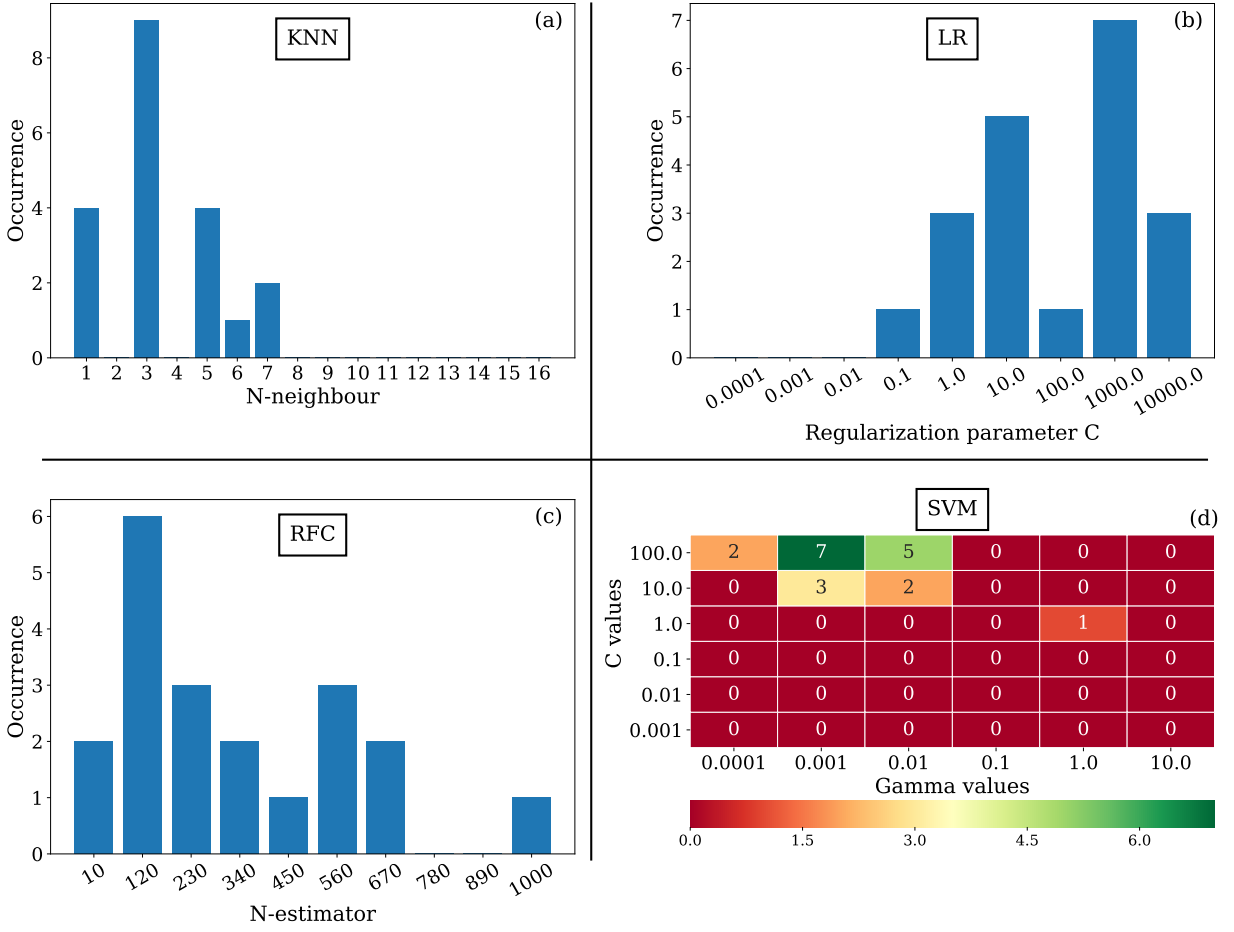
Figure 4. These plots represent the values of optimal parameters which gives best TSS score on the validation dataset for each experimental dataset (D_i)

Classifier Name	Average Performance Measure in 20 trials when the models are optimized for TSS				An Example of Confusion Matrix Elements			
	SSW	CSW	TSS	MAC	TN	TP	FN	FP
KNN	0.887 ± 0.040	0.990 ± 0.006	0.938 ± 0.019	0.969 ± 0.010	663	110	6	3
Random forest	0.898 ± 0.042	0.989 ± 0.008	0.944 ± 0.020	0.972 ± 0.010	664	101	6	2
Logistic regression	0.959 ± 0.033	0.975 ± 0.009	0.967 ± 0.018	0.983 ± 0.009	669	94	1	9
SVM	0.956 ± 0.031	0.974 ± 0.010	0.965 ± 0.017	0.983 ± 0.009	656	105	2	10

Table 2. Performance of classifiers trained with the best hyper-parameters deduced via Grid Search and 10-fold cross-validation over 20 randomly shuffled datasets. The confusion matrix elements correspond to a test dataset in D_i whose TSS is closest to the determined mean.
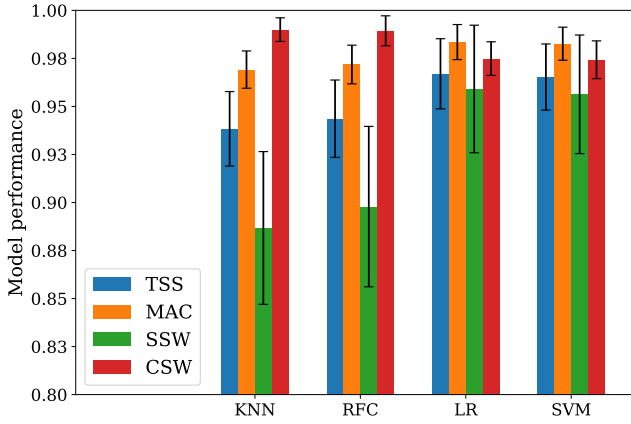
Figure 5. This plot depicts the classification performance of all four models.

for LR (and SVM) compared to RFC or KNN which indicates the suitability of LR (and SVM) in highly active space weather circumstances. Also the close values of SSW and CSW tells us about the unbiased nature of the model predictions. Clearly, LR/SVM is a better choice over KNN/RFC for having a very similar SSW and CSW scores. An example of the confusion matrix elements corresponding to a seed value of D_i with TSS close to the mean value is also shown in Table 2.

To understand which active region parameters are more useful in determining the flaring capability of an AR, we train our models with the 14 AR parameters individually. The outcome of this experiment for LR is presented in Figure 6 where all these AR parameters are plotted along the y-axis in ascending order of their individual TSS scores. This implies that the top most parameter in the y-axis is the most significant one having highest individual classification capability and as we move downward, we find parameters of lesser importance.

The ranking of input features based on the individual TSS scores depends on the ML model used, and can moderately differ for different models. For a particular ML algorithm, feature ranking may also depend on the model hyper-parameters. Hence to get a more general global ranking of features, we follow a marking scheme in which we assign points (ranging from 1 to 14) to each parameter based on their individual TSS-ranking for each of the four models and the univariate F-score ranking. For example, for LR 14 points are assigned to TOTUSJH for its highest TSS score, whereas EPSZ gets 1 point based on Figure 6. Finally we add up all the points for each feature from different models to get a cumulative ranking as shown in Figure 7.

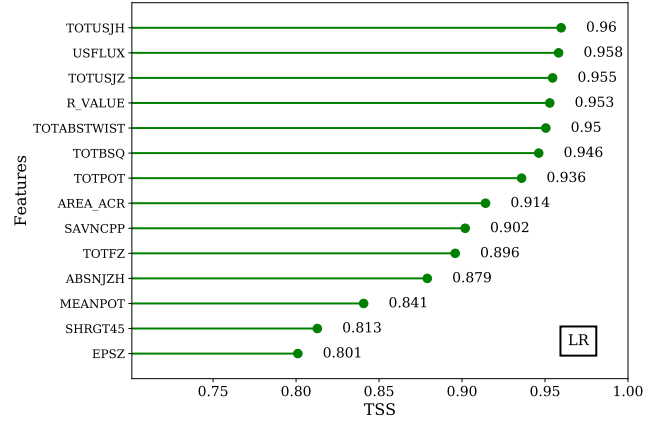We further optimize the LR model by tuning the model hyper-parameter for maximizing SSW and CSW

Figure 6. Average TSS scores of individual features for LR. Each score is obtained by training the LR model with the single parameter as input data, averaged over the output of the 20 experimental datasets.
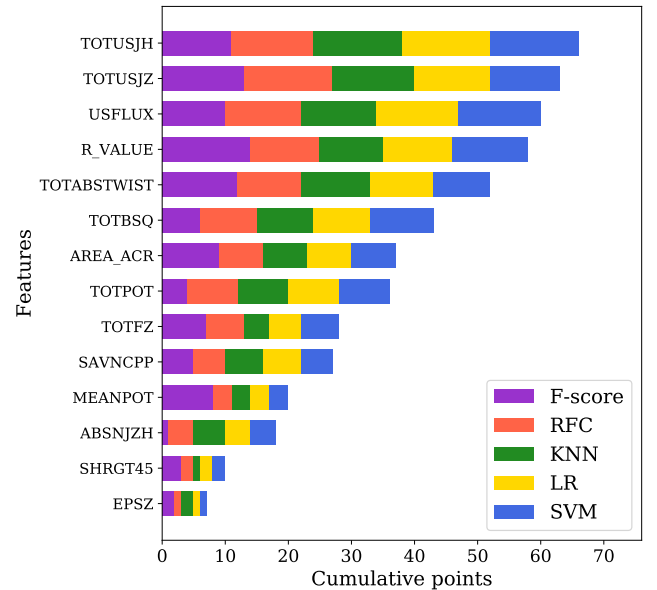

Figure 7. Feature ranking based on cumulative points obtained from F-statistic, RFC, KNN, LR and SVM. Top scoring feature in each model gets 14 points, while lowest scoring feature gets 1 point. The points are added for each feature and then the features are ranked accordingly.

metric to see how the ranking of input magnetic features changes for these two newly introduced space weather metrics. The left panel in Figure 8 shows the ranking of input features according to their individual SSW score and the right panel shows the feature ranking with respect to the CSW score. The major difference is reflected in the ranking of the R_value. It is the top ranked feature when the ranking is based on SSW score but its ranking goes down drastically when it is based on the CSW score. On the other hand, for the CSW metric

extensive parameters (area integrated) appear to be the main contributors.

To study the dependence of model performance on the number of input AR features, we train our LR model by eliminating input features one by one and check the model performance at each step. The elimination is done by following both the ascending and descending order of ranking based on the individual TSS, the result of which is represented in Figure 9. We can see that when we eliminate the features in their ascending order of ranking the model performance does not change much. This is expected because the more important features are eliminated at the last. On the other hand, for descending order we see a drastic fall in model performance when the number of eliminated features increases beyond 9. This indicates a significant loss of correlation with the output labels at each step beyond this point. The plateau in the descending order plot of Figure 9 is only possible if the top ranked features are highly correlated among themselves, causing no significant loss of information when these features are thrown out. A study by Hazra et al. (2015) also confirms the correlation amongst integrated magnetic features to show the connection between AR magnetic properties and coronal X-ray flux. The correlogram presented in Figure 10 confirms this, with all top 8 features, excluding R_VALUE, being highly intercorrelated. One possible reason behind this high correlation could be that they are extrinsic features, or in other words, their values depend on the size of the AR as they represent the sums of physical quantities over the entire AR. As correlated features do not provide new information, we group features with correlation constant > 0.9 and train our model by picking up the top-performing feature from each group. Following this scheme we select 6 features: TOTUSJH, R_VALUE, TOTFZ, SAVNCPP, MEANPOT and SHRGT45. When trained with these features only, the LR classifier gives an average TSS and MAC values of 0.962 and 0.981 respectively (with the SSW and CSW of 0.956 and 0.968), which are close to our primary model performance with 14 features.

5. CONCLUSIONS

With the advancement of new technologies, especially in satellite based telecommunications and navigational networks, a significant fraction of our technological assets has become increasingly vulnerable to space weather disturbances. This has resulted growing demand for reliable space weather forecast. Solar flares strongly influence the space weather, which is why we address the problem of predicting solar flares using their source region characteristics. In this work, we have built a high performance operational Logistic Regression (LR) clas-

sifier which can differentiate Solar Active Regions (ARs) based on their flaring capabilities. We have compared four supervised ML models, all of which perform quite well in classifying active regions into Positive/Flaring and Negative/Non-flaring categories. The method we follow is statistically unbiased due to the use of 20 randomly shuffled replicas of primary dataset for measuring model performance. The LR classifier delivers the highest average TSS score of 0.967 ± 0.018 closely followed in performance by the Support Vector Machine (SVM) classifier.

While a direct comparison of the model performances between our algorithms with previous studies may not be appropriate due to subtle differences in the data selection scheme and the size of the database used, we do note that in the context of TSS, we achieve a higher performance score relative to earlier classification attempts with supervised ML algorithms (e.g., Bobra & Couvidat 2015; Nishizuka et al. 2017; Florios et al. 2018).

One possible reason for achieving a high TSS could be the exclusion of C class events in the data preparation stage. The distribution of top five input features in our dataset are shown in figure 11, where we can see a clear separation between two cluster of data points for two different classes. This ensures that our dataset is easily separable with two distinct classes in feature space. Other possible reasons could be the different event selection scheme, larger temporal coverage of our dataset and also it is important to note that each entries in the negative class comes from a different SHARP regions ensuring no repetition of active region patches in the non-flaring class.

In addition to achieving high TSS score, we find that a global indicator of magnetic twist, estimated by the feature TOTABSTWIST, plays an important role in predicting AR flare potential. Although TOTABSTWIST comes in the fifth position of cumulative feature ranking, other twist related parameters including VTWIST and MEANALP, are not found to play a significant role.

We have also introduced two new performance indicators, termed as Severe Space Weather (SSW) and Clear Space Weather (CSW), which are useful in comparing model performances depending on the operational space weather condition one wishes to lay more emphasis on. For example, when the solar activity is high, we may wish to get a reliable all clear forecast for executing specific time-critical tasks that are susceptible to space weather. So, depending on the application and operational space weather scenario, SSW and CSW can provide more meaningful operational intelligence than TSS alone. We can also get an estimate of the model bias towards a specific forecast by examining the difference
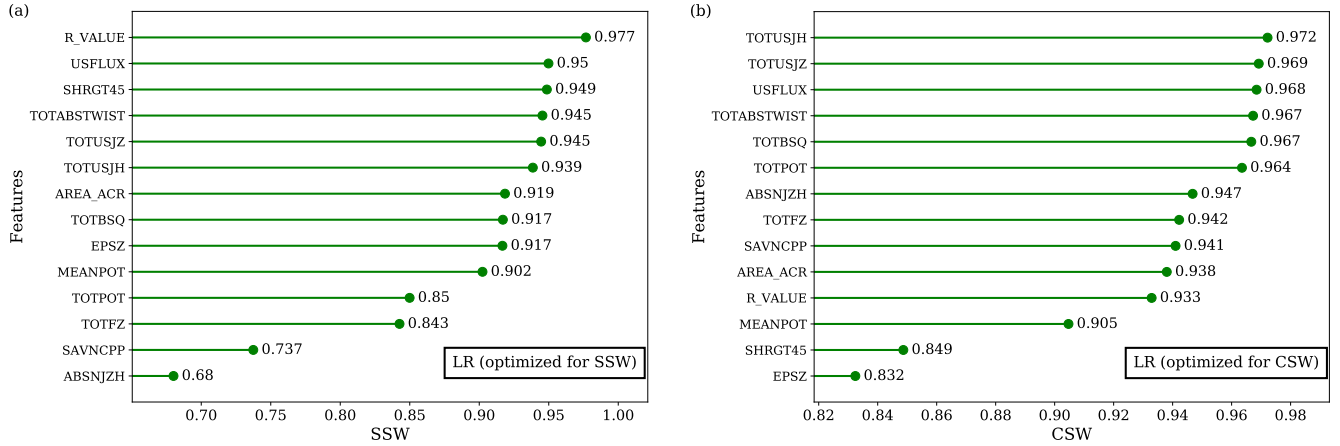
Figure 8. The left panel (a) shows the feature ranking for LR model when the model is optimized for the metric SSW and the panel (b) on the right shows the feature ranking when the LR model is optimized for CSW metric.

between SSW and CSW. With these two indicators, we see that KNN and RFC are more biased towards negative class (as CSW is much higher than SSW) than SVM and LR. Because of the larger size of the negative class, a classifier's forecast may become biased towards it. But our analysis shows that classifiers such as LR and SVM can be suitably optimized to minimize the class imbalance problem significantly.

We have also studied the relative importance of input features in terms of their ability to classify the flaring and non-flaring ARs. Based on the global ranking of Figure 7, we have identified key magnetic features that are responsible for the flare potential of an AR. The total unsigned current helicity, total unsigned vertical current, total unsigned magnetic flux, flux near strong-field high-gradient neutral-line, and total absolute twist are the major deciding factors of AR flare potential. It is important to note that all of highly ranked features in Figure 7 denote extensive or net properties of an AR, except for the R_VALUE. This reaffirms previous findings (Welsch et al. 2009; Hazra et al. 2020) that extensive parameters contribute more to forecasting algorithms than intensive parameters. The only non-extensive feature that performs well is R_VALUE, thus indicating that it contains some unique information regarding flaring potential.

Our analysis shows that for a given classifier, the ranking of input magnetic features differs based on the choice of the performance metric. For example, the ranking of R_value drops from the top to the 11th rank when we use CSW as the performance indicator instead of SSW. The reason behind this drop could be explained from the probability density plot of the R_value in Figure 11 (4th row, right panel), where we can see a very compact

distribution of R_values around the mean for the flaring class. Whereas, for the non-flaring class, R_values are clustered at two different locations resulting in a high intraclass variance.

We also find that the model performances have very low dependency on the number of input features especially when the input features are highly correlated. We have shown that a high model performance could be retained even with a smaller set of input magnetic features, selected carefully to reduce internal correlation. Our work brings to fore key properties of parameter-based ML flare forecasting which can be utilized in future works to develop more robust flare forecasting models. Finally, we anticipate, our comprehensive analysis will lead to operational flare forecasting with higher efficiency and higher precision.
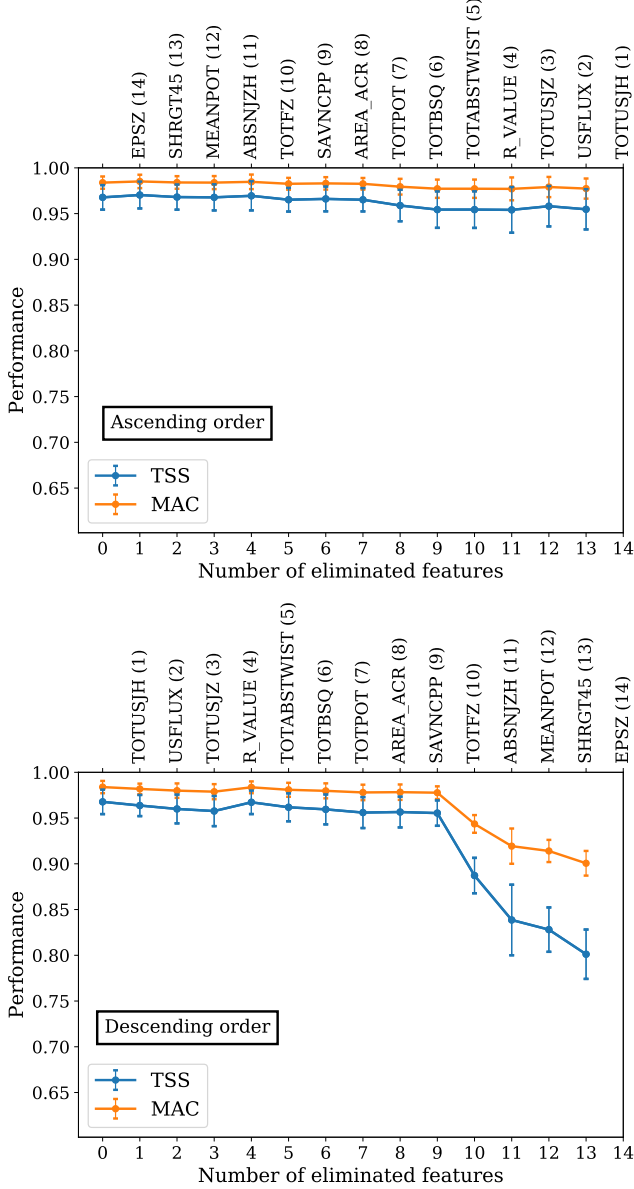
ACKNOWLEDGEMENTS

Figure 9. Variation of LR performance with number of input features. The experiment is performed over all 20 experimental datasets (D_i). In each step, the eliminated feature along with its rank is indicated on the top x-axis whereas the bottom x-axis indicates the total number of eliminated features at that step.

lized the drms open source software package (Glogowski et al. 2019) to access data from HMI. The authors acknowledge usage of data from the XRT flare catalogue and GOES flare database to acquire the information of flaring events and their corresponding active regions. All Machine Learning algorithms have been implemented using the scikit-learn package (Pedregosa et al. 2011) on Python.

Data Availability: The input magnetic feature dataset used for this study is freely available on Zenodo: doi:10.5281/zenodo.5498347.

REFERENCES

Ahmadzadeh, A., Aydin, B., Georgoulis, M. K., et al. 2021, The Astrophysical Journal Supplement Series, 254, 23

Ahmed, O. W., Qahwaji, R., Colak, T., et al. 2013, SoPh, 283, 157

Al-Ghraibah, A., Boucheron, L. E., & McAteer, R. T. J. 2015, A&A, 579, A64

Alipour, N., Mohammadi, F., & Safari, H. 2019, ApJS, 243, 20

Barnes, G., Leka, K. D., Schumer, E. A., & Della-Rose, D. J. 2007, Space Weather, 5, S09002

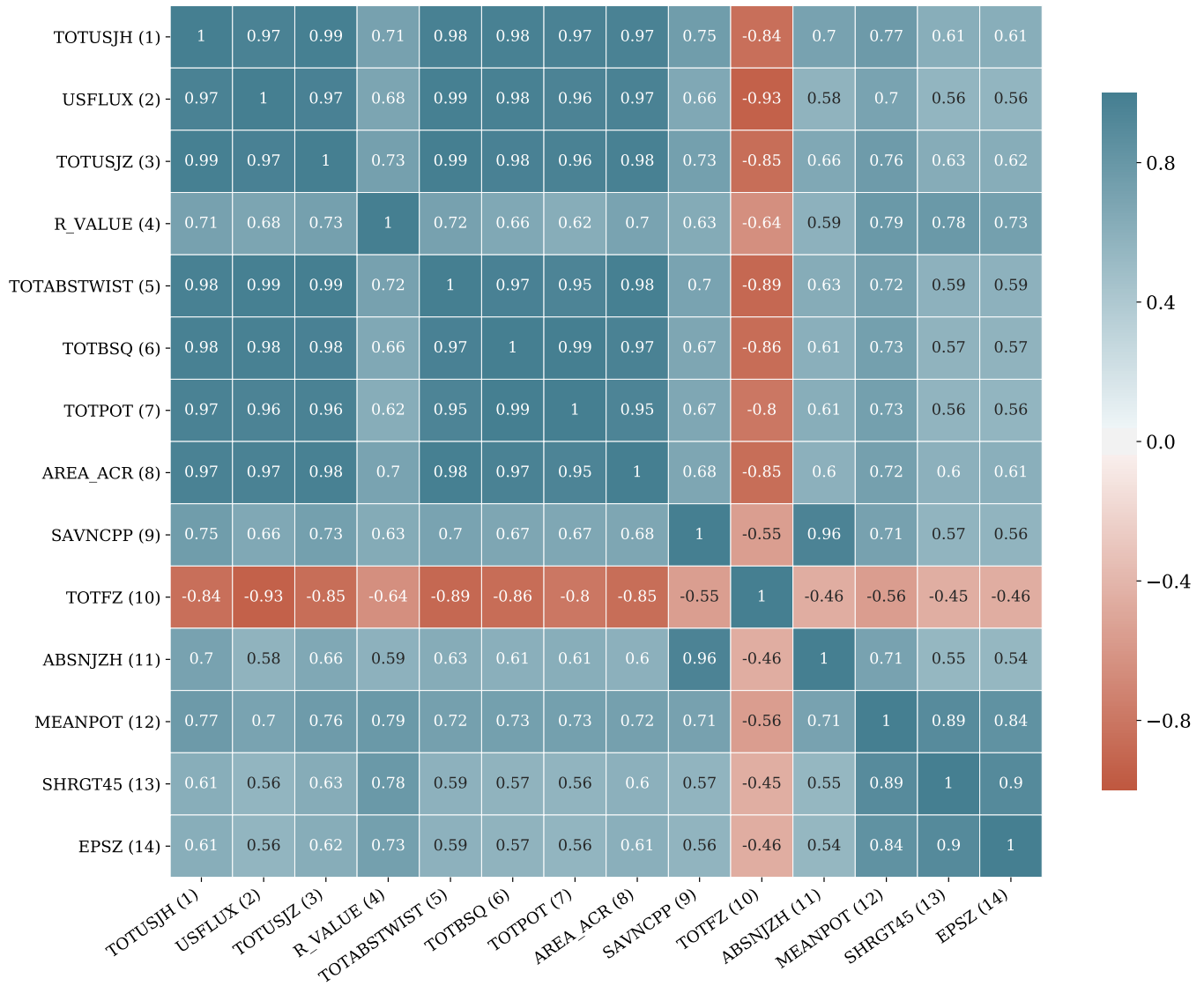Benvenuto, F., Piana, M., Campi, C., & Massone, A. M. 2018, ApJ, 853, 90

Figure 10. Correlation matrix for the input features, which are arranged in the order of their individual TSS-ranking with LR. The ranks are shown in brackets next to the feature keywords.

Bhattacharjee, S., Alshehhi, R., Dhuri, D. B., & Hanasoge, S. M. 2020, ApJ, 898, 98

Bhowmik, P., & Nandy, D. 2018, Nature Communications, 9, 5209

Bishop, C. 2013, Pattern Recognition and Machine Learning: All "just the Facts 101" Material, Information science and statistics (Springer (India) Private Limited). https://books.google.co.in/books?id=HL4HrgEACAAJ

Bloomfield, D. S., Higgins, P. A., McAteer, R. T. J., & Gallagher, P. T. 2012, ApJL, 747, L41

Bobra, M. G., & Couvidat, S. 2015, The Astrophysical Journal, 798, 135. https://doi.org/10.1088%2F0004-637x%2F798%2F2%2F135

Bobra, M. G., Sun, X., Hoeksema, J. T., et al. 2014, SoPh, 289, 3549

Boucheron, L. E., Al-Ghraibah, A., & McAteer, R. T. J. 2015, ApJ, 812, 51

Campi, C., Benvenuto, F., Massone, A. M., et al. 2019, ApJ, 883, 150

Chen, Y., Manchester, W. B., Hero, A. O., et al. 2019, Space Weather, 17, 1404

Cinto, T., Gradvohl, A. L. S., Coelho, G. P., & da Silva, A. E. A. 2020, SoPh, 295, 93

Colak, T., & Qahwaji, R. 2009, Space Weather, 7, S06001

Cortes, & Vapnik. 1995, Machine Learning, 20, 273

Dhuri, D. B., Hanasoge, S. M., & Cheung, M. C. M. 2019, Proceedings of the National Academy of Science, 116, 11141

Eastwood, J. P., Biffis, E., Hapgood, M. A., et al. 2017, Risk Analysis, 37, 206
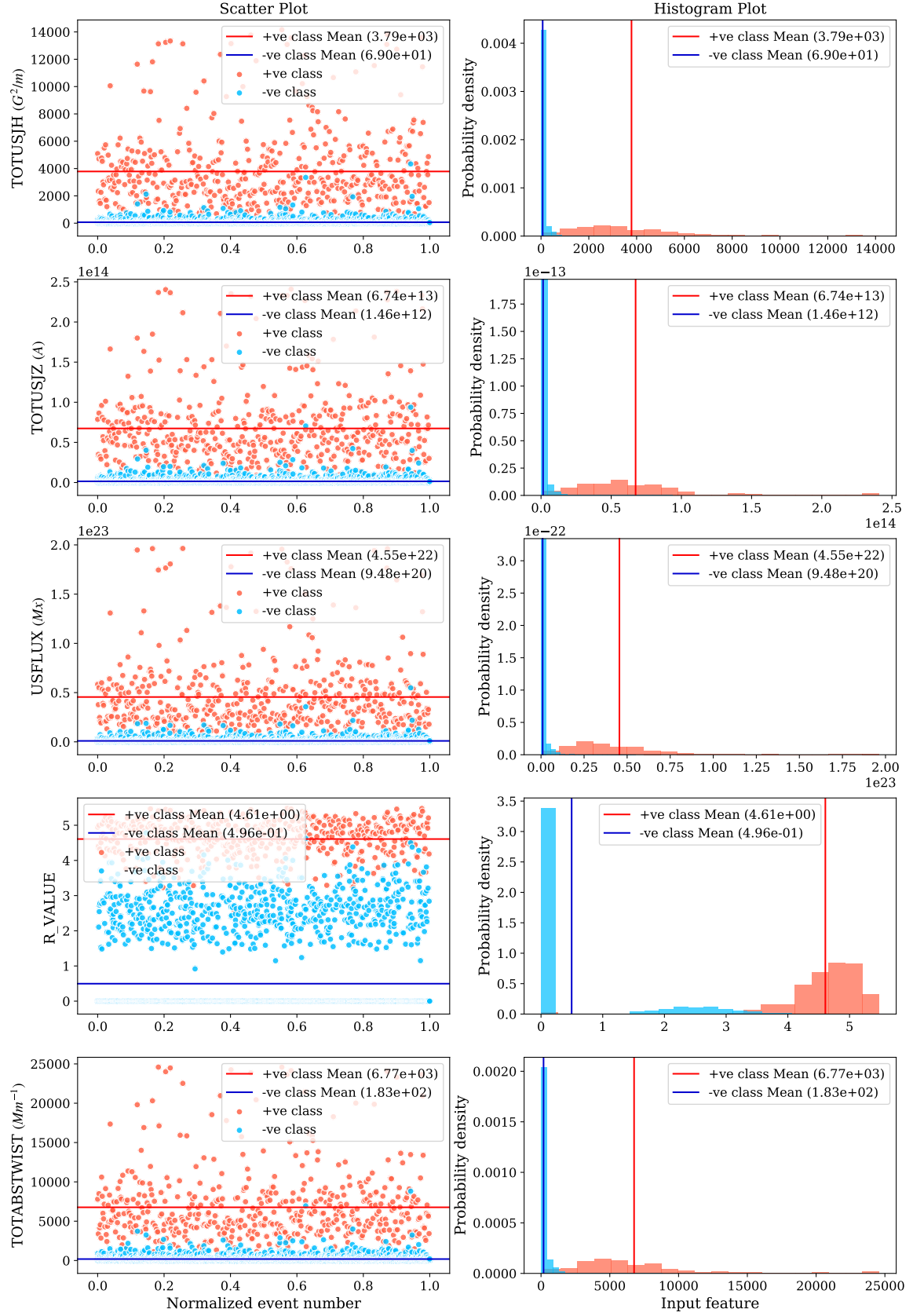
Figure 11. Distribution of magnetic features in positive and negative classes. The left column shows the scatter plot of top five input features according to the cumulative feature-ranking. *X* axis of the scatter plot is the normalized event number which is the number of event divided by the total number of event in that class. The mean values for both the classes are shown. In the right column, histogram plot of probability density is shown for corresponding features.

Fernandez Borda, R. A., Mininni, P. D., Mandrini, C. H., et al. 2002, SoPh, 206, 347

Filali Boubrahimi, S., & Angryk, R. 2018, in 2018 IEEE First International Conference on Artificial Intelligence and Knowledge Engineering (AIKE), 162–163

Fix, E., & Hodges, J. L. 1951, technical Report 4, USAF School of Aviation Medicine, Randolph Field, Texas, U.S.A.

Florios, K., Kontogiannis, I., Park, S.-H., et al. 2018, SoPh, 293, 28

Forbes, T. 2000, Solar Flare Models, ed. P. Murdin, 2295

Georgoulis, M. K., & Rust, D. M. 2007, The Astrophysical Journal Letters, 661, L109

Glogowski, K., Bobra, M. G., Choudhary, N., Amezcua, A. B., & Mumford, S. J. 2019, Journal of Open Source Software, 4, 1614. https://doi.org/10.21105/joss.01614

Hagyard, M. J., Smith, J. B., J., Teuber, D., & West, E. A. 1984, SoPh, 91, 115

Hahn, M., Gaard, S., Jibben, P., Canfield, R. C., & Nandy, D. 2005, ApJ, 629, 1135

Hamdi, S. M., Kempton, D., Ma, R., Boubrahimi, S. F., & Angryk, R. A. 2017, in 2017 IEEE International Conference on Big Data (Big Data), 2543–2551

Hapgood, M. A. 2011, Advances in Space Research, 47, 2059

Hazra, S., Nandy, D., & Ravindra, B. 2015, SoPh, 290, 771

Hazra, S., Sardar, G., & Chowdhury, P. 2020, A&A, 639, A44

Huang, X., Wang, H., Xu, L., et al. 2018, ApJ, 856, 7

Huang, X., Yu, D., Hu, Q., Wang, H., & Cui, Y. 2010, SoPh, 263, 175

Jiao, Z., Sun, H., Wang, X., et al. 2020, Space Weather, 18, e02440

Jing, J., Song, H., Abramenko, V., Tan, C., & Wang, H. 2006, ApJ, 644, 1273

Jonas, E., Bobra, M., Shankar, V., Todd Hoeksema, J., & Recht, B. 2018, SoPh, 293, 48

LaBonte, B. J., Georgoulis, M. K., & Rust, D. M. 2007, ApJ, 671, 955

Lan, R.-S., Jiang, Y., Ding, L.-G., & Yang, J.-W. 2012, Research in Astronomy and Astrophysics, 12, 1191

Leka, K. D., & Barnes, G. 2003a, ApJ, 595, 1277

—. 2003b, ApJ, 595, 1296

—. 2003c, ApJ, 595, 1277

—. 2003d, ApJ, 595, 1296

Li, R., Wang, H.-N., He, H., Cui, Y.-M., & Zhan-LeDu. 2007, ChJA&A, 7, 441

Li, X., Zheng, Y., Wang, X., & Wang, L. 2020, ApJ, 891, 10

Linton, M. G., Longcope, D. W., & Fisher, G. H. 1996, ApJ, 469, 954

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

Ma, R., Boubrahimi, S. F., Hamdi, S. M., & Angryk, R. A. 2017, in 2017 IEEE International Conference on Big Data (Big Data), 2569–2578

Mehta, P., Bukov, M., Wang, C.-H., et al. 2019, Physics Reports, 810, 1, a high-bias, low-variance introduction to Machine Learning for physicists. https://www.sciencedirect.com/science/article/pii/S0370157319300766

Moore, R. L., Falconer, D. A., & Sterling, A. C. 2012, ApJ, 750, 24

Muranushi, T., Shibayama, T., Muranushi, Y. H., et al. 2015, Space Weather, 13, 778

Nandy, D. 2008, in Astronomical Society of the Pacific Conference Series, Vol. 383, Subsurface and Atmospheric Influences on Solar Activity, ed. R. Howe, R. W. Komm, K. S. Balasubramaniam, & G. J. D. Petrie, 201

Nandy, D. 2021, SoPh, 296, 54

Nandy, D., Hahn, M., Canfield, R. C., & Longcope, D. W. 2003, ApJL, 597, L73

Nandy, D., Mackay, D. H., Canfield, R. C., & Martens, P. C. H. 2008, Journal of Atmospheric and Solar-Terrestrial Physics, 70, 605

Nandy, D., Martens, P. C. H., Obridko, V., Dash, S., & Georgieva, K. 2021, Progress in Earth and Planetary Science, 8, 40

Nishizuka, N., Kubo, Y., Sugiura, K., Den, M., & Ishii, M. 2020, ApJ, 899, 150

Nishizuka, N., Sugiura, K., Kubo, Y., Den, M., & Ishii, M. 2018, ApJ, 858, 113

Nishizuka, N., Sugiura, K., Kubo, Y., et al. 2017, ApJ, 835, 156

Pal, S., Gopalswamy, N., Nandy, D., et al. 2017, ApJ, 851, 123

Pal, S., Nandy, D., Srivastava, N., Gopalswamy, N., & Panda, S. 2018, ApJ, 865, 4

Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2011, Journal of Machine Learning Research, 12, 2825

Pereira, R. B., Plastino, A., Zadrozny, B., & Merschmann, L. H. 2018, Information Processing & Management, 54, 359. https://www.sciencedirect.com/science/article/pii/S0306457318300165

Priest, E. R., & Forbes, T. G. 2002, A&A Rv, 10, 313

Qahwaji, R., & Colak, T. 2007, SoPh, 241, 195

Qu, M., Shih, F. Y., Jing, J., & Wang, H. 2003, SoPh, 217, 157

Raboonik, A., Safari, H., Alipour, N., & Wheatland, M. S. 2017, ApJ, 834, 11

Ribeiro, F., & Gradvohl, A. 2021, Astronomy and Computing, 35, 100468. https://www.sciencedirect.com/science/article/pii/S2213133721000226

Sadykov, V. M., & Kosovichev, A. G. 2017, ApJ, 849, 148

Scherrer, P. H., Schou, J., Bush, R. I., et al. 2012, SoPh, 275, 207

Schrijver, C. J. 2007, ApJL, 655, L117

—. 2015, Space Weather, 13, 524

Schrijver, C. J., DeRosa, M. L., Metcalf, T., et al. 2008, ApJ, 675, 1637

Schrijver, C. J., Kauristie, K., Aylward, A. D., et al. 2015, Advances in Space Research, 55, 2745

Sinha, S., Srivastava, N., & Nandy, D. 2019, ApJ, 880, 84

Song, H., Tan, C., Jing, J., et al. 2009, SoPh, 254, 101

The SunPy Community, Barnes, W. T., Bobra, M. G., et al. 2020, The Astrophysical Journal, 890, 68

Tin Kam Ho. 1995, in Proceedings of 3rd International Conference on Document Analysis and Recognition, Vol. 1, 278–282 vol.1

Wang, J., Liu, S., Ao, X., et al. 2019, ApJ, 884, 175

Watanabe, K., Masuda, S., & Segawa, T. 2012, SoPh, 279, 317

Welsch, B. T., Li, Y., Schuck, P. W., & Fisher, G. H. 2009, ApJ, 705, 821

Woodcock, F. 1976, Monthly Weather Review, 104, 1209

Yeates, A. R., Attrill, G. D. R., Nandy, D., et al. 2010, The Astrophysical Journal, 709, 1238. https://doi.org/10.1088/0004-637x/709/2/1238

Yu, D., Huang, X., Wang, H., & Cui, Y. 2009, SoPh, 255, 91

Yuan, Y., Shih, F. Y., Jing, J., & Wang, H.-M. 2010, Research in Astronomy and Astrophysics, 10, 785

Zheng, Y., Li, X., & Wang, X. 2019, ApJ, 885, 73