A Time Series Classification-based Approach for Solar Flare Prediction

Conference Paper · December 2017

DOI: 10.1109/BigData.2017.8258213

5 authors, including:

Shah Muhammad Hamdi
Georgia State University
9 PUBLICATIONS **15** CITATIONS
SEE PROFILE

Dustin Kempton
Georgia State University
22 PUBLICATIONS **107** CITATIONS
SEE PROFILE

Ruizhe Ma
Georgia State University
17 PUBLICATIONS **54** CITATIONS
SEE PROFILE

Soukaina Filali Boubrahimi
Georgia State University
23 PUBLICATIONS **53** CITATIONS
SEE PROFILE

A Time Series Classification-based Approach for Solar Flare Prediction

Shah Muhammad Hamdi, Dustin Kempton, Ruizhe Ma, Soukaïna Filali Boubrahimi, Rafal A. Angryk
Department of Computer Science
Georgia State University, Atlanta, GA 30302
Email: {shamdi1, dkempton1, rma1, sfilaliboubrahimi1, rangryk}@cs.gsu.edu

***Abstract*—Solar flare prediction is an important task because of their potential impacts on both space and terrestrial infrastructure. This prediction task can be modeled as a binary classification between flaring and non-flaring Active Regions. Previous works on flare prediction focused on representing flaring and non-flaring Active Region examples in vector space, where the feature space was found from the Active Region magnetic field parameters. We extract time series samples of these Active Region parameters and present a flare prediction method based on the k-NN classification of the univariate time series. We find that, for our classification task, using a statistical summarization on the time series of a single Active Region parameter, called *total unsigned current helicity*, outperforms the use of all Active Region parameters at a single instant of time. Additionally, we present a data model of the flaring/non-flaring Active Regions using multivariate time series.**

***Index Terms*—Flare prediction; Time series; Total unsigned current helicity; k-NN classifier**

I. Introduction

Solar flares are sudden bursts of radiation from Sun's surface. Solar events such as flares and Coronal Mass Ejections (CMEs) can have hazardous impacts on infrastructures both in space and on the ground. X-rays and UV radiation of large flares can cause radio blackout. Energetic particle flux from flares can cause solar radiation storm, which can have negative health effects on astronauts, aircrew and airline passengers, as well as negative technological impacts on electronic devices of the satellites, aircraft, and even the devices located on the ground. Therefore, precise forecasting and prediction of severe space weather conditions such as M-class and X-class flares can save infrastructures in space and on the ground, whose replacement/repairing cost might be trillions of dollars [1]. Fig. 1 is an example of an M-class flare [2].

Since theoretical models of solar flare occurrence, such as the relationship between the photospheric and coronal magnetic field of the Sun during the flare occurrence, are not fully understood, heliophysics community relies on data-driven approaches for flare prediction. As most flares occur in the Active Regions of the Sun, flare prediction can be modeled as a supervised learning problem of machine learning, specifically the binary classification between flaring and non-flaring Active Regions (AR), where flaring Active Regions are considered to be in the positive class and non-flaring Active Regions are considered to be in the negative class. In this work, as positive class examples, we consider the Active Regions that have one or more M-class or X-class flares during their crossing of the observable solar disk. The Active Regions that have never flared during the disk crossing (not even C-class flares) are considered as negative class examples.

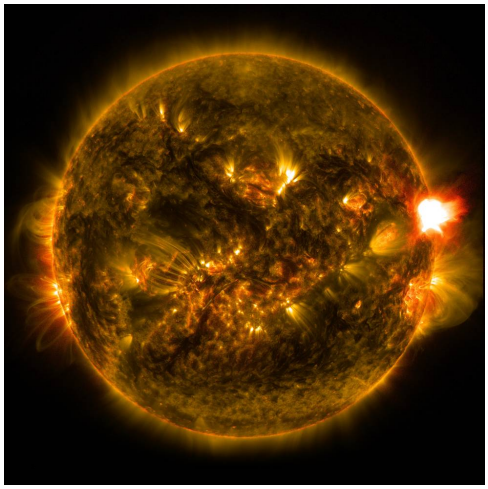


Fig. 1: Image of an M-class flare which erupts from the right side of the Sun at 11:24 p.m. EST on Jan. 12, 2015. Credit: NASA/SDO.

While the previous studies on flare prediction focused on the vector-based representation of flaring and non-flaring Active Regions, where the feature space is formed by the magnetic field-based AR parameters formulated by solar physicists [3]–[5], we focus on the time series properties of the AR parameters. These time series are extracted based on two time windows : *lookback* (the time window before which the flare happens), and *span* (the time window during which the AR parameter values are calculated).

Time series representation of the AR parameters can be used to rank them based on the time series quality in classification. When the time series of each AR parameter is considered, the problem becomes the multivariate time series classification problem. By finding the AR parameter whose time series exhibit the best classification performance, the problem can be simplified into the single-variate time series classification. If time series data of a single AR parameter is utilized for flare prediction, then time series classification algorithms which are successfully applied on other domains such as stock market behavior prediction [6], handwriting recognition [7] and so on, can similarly be applied to flare prediction.

978-1-5386-2715-0/17/$31.00 ©2017 IEEE

The complexity of time series classification greatly depends on the length of the time series, because each time step increase in the time series length is an increase in the dimension of the input vector space, leading to the curse of dimensionality. To overcome this problem, we propose to use a statistical summarization of the time series. The summarized time series of the best AR parameter is considered as the vector-based representation of flaring/non-flaring AR, and k-nearest neighbors (k-NN) classifier is used on this vector space. By considering the time series of only one AR parameter, which is selected based on the classification performance on different datasets, our model reduces the cost of calculating multiple AR parameters, while exhibiting better classification performance than the models where all AR parameters are used with the values at a single instant of time.

The contributions made by this paper are listed below.

1) Data modeling of the flaring/non-flaring Active Regions using multivariate time series based on lookback and span time windows.
2) Finding the best AR parameter in terms of its time series quality in classification.
3) Making the summarized representation of the time series of the best AR parameter to form a new vector space of flaring and non-flaring Active Regions. The performance of k-NN classifier on this vector space is better than the state-of-the-art flare prediction models.
4) Experimentally validating that the consideration of C-class flares in positive class does not improve classification performance.

The rest of the paper is organized as follows. In Section II, we discuss the related work. We present a data model for flaring and non-flaring Active Regions using multivariate time series in Section III. In Section IV, we present our flare prediction model by time series summarization-based AR parameter selection. We present the experimental findings in Section V. Finally, we discuss our future work and conclude the paper in Section VI.

II. Related Work

While most of the current methods of flare prediction are data-driven approaches, the earliest flare prediction system was THEO [8], which was an expert system that required human input. The system was adopted by Space Environment Center (SEC) of National Oceanic and Atmospheric Administration (NOAA) in 1987. It used a set of sunspot and magnetic field properties to predict different flare classes.

Later efforts of flare prediction are mostly based on data-driven approaches rather than on purely theoretical modeling. Data-driven approaches are divided into two categories - linear statistical and nonlinear statistical (mostly machine learning). These two categories can be subdivided into two subcategories - line-of-sight magnetogram-based models and vector magnetogram-based models.

Active Regions are parameterized either by photospheric magnetic field data that contain only the line-of-sight component of the magnetic field or by the full-disk photospheric vector magnetic field. After the launch of Solar Dynamics Observatory (SDO) by NASA in 2010, its instrument Helioseismic and Magnetic Imager (HMI) has been mapping the full-disk vector magnetic field every 12 minutes [9]. Although the continuous stream of vector magnetogram is a better means for parameterizing the Active Regions, it was not easily available before 2010 and people had to use line-of-sight magnetic data for flare prediction.

Linear statistical studies focus on identifying the AR magnetic properties that are correlated with the flares. Cui et al. [10] and Jing et al. [11] used line-of-sight magnetogram to parameterize Active Regions and studied correlation-based statistical relationships between those AR parameters and flare occurrences. Leka and Barnes [3] calculated vector magnetogram-based AR parameters for the first time and used linear discriminant analysis (LDA) for classification. They collected vector magnetogram data from Mees Solar Observatory Imaging Vector Magnetograph on the summit of Mount Haleakala.

Nonlinear statistical models mainly use machine learning-based classifiers. After parameterizing the Active Regions with line-of-sight magnetograms, Ahmed et al. [12] used the artificial neural network, Yu et al. [13] used C4.5 decision tree, Song et al. [14] used logistic regression, and Al-Ghraibah et al. [15] used relevance vector machine as classification models. Qahwaji et al. [16] considered McIntosh classification of sunspot groups and solar cycle data and used support vector machine (SVM) and Cascade-Correlation Neural Networks (CCNN) for prediction. Bobra et al. [4] used SVM on the AR parameters derived from vector magnetograms. Nishizuka et al. [5] used both line-of-sight and vector magnetograms and compared the performance of three classifiers - k-NN, SVM, and extremely randomized tree (ERT).

Almost all of the abovementioned works focussed on the parameterization of the Active Regions by line-of-sight or vector magnetograms but did not consider the impact of the time series of the individual AR parameters that can be extracted for a particular duration of time before the occurrence of the flare. In this work, we evaluated the vector magnetogram-based AR parameters based on their time series quality to distinguish flaring and non-flaring Active Regions.

III. Data Modeling of the Active Regions

In this section, we define some terminologies and present a formal data model for our flaring and non-flaring Active Regions (Fig. 2). Each Active Region instance is initially represented by six data fields.

$$event = \langle id, timestamp, lookback, span, mvts, label \rangle$$

Here, id is the NOAA Active Region number and $timestamp$ is the occurrence time of the flare (for flaring Active Regions) and the sampling time before which the parameter data are collected (for non-flaring Active Region). The $lookback$ represents the time window before the occurence of the flare. The $span$ is the time window for sampling AR patches from

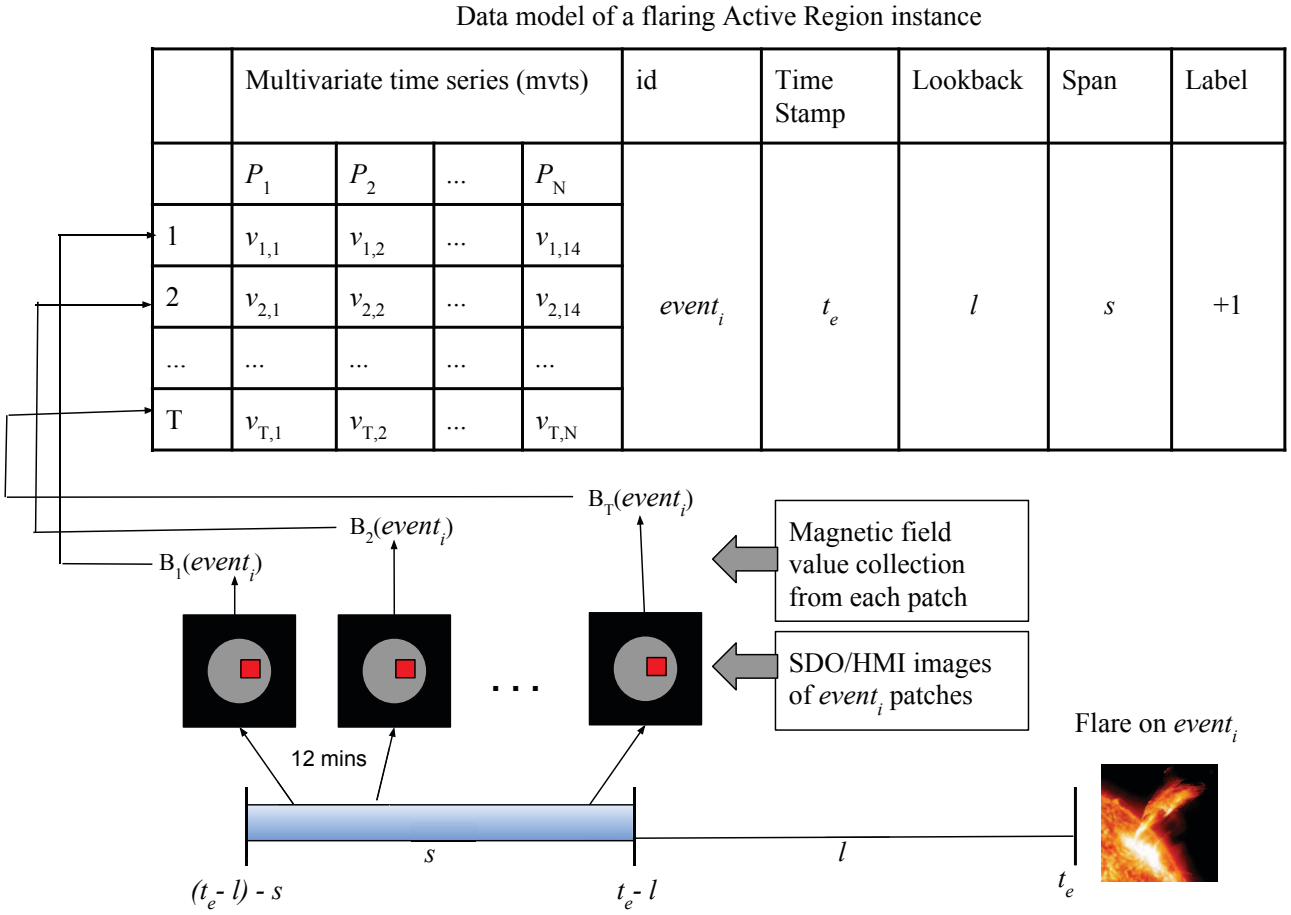


Fig. 2: Data model of a flaring Active Region in terms of id, generation time, lookback time, span time, multivariate time series, and label.

SDO/HMI images in 12 minutes cadence. These AR patches are used to get timewise magnetic field values $B_1, B_2, \ldots, B_T$, where $B_i = [B_{\phi i}, B_{\theta i}, B_{ri}]$ is formed from the components of the vector magnetic field data (see [4] for details). The magnetic field values are used to calculate N parameter values of the Active Region. The $mvts$ is a collection of time series $\{P_1, P_2, \ldots, P_N\}$, where P_j represents the time series of j-th parameter. Each time series has fixed length T. If the time unit of $span$ is hours, and the time unit of $cadence$ is minutes, then $T = \frac{span \times 60}{cadence}$. Therefore, P_j is a vector of length T and represented by $P_j = [v_{1,j}, v_{2,j}, \ldots, v_{T,j}]$, where $v_{k,j}$ is the k-th value of the time series P_j. If an M-class or X-class flare occurs in time t_e of an Active Region, then the $label$ of this example is $+1$. If no M-class or X-class flare occurs during the disk crossing of the Active Region, then the $label$ of that example is -1. The data model of the flaring Active Regions is depicted in Fig. 2.

IV. Flare Classification Model

A. Problem Definition

Given lookback l hours and span s hours, a dataset of M events is represented by $\{(X_i, y_i)\}_{i=1}^{M}$, where $X_i = mvts_i \in \mathbb{R}^{T \times N}$ and $y_i \in \{+1, -1\}$. In binary classification, if Tr is the number of labeled events, i.e., training examples, we train a classifier with Tr labeled examples $\{(X_i, y_i)\}_{i=1}^{Tr}$ and use the classification model to label $M - Tr$ test examples $\{(X_i)\}_{i=Tr+1}^{M}$. In this work, we aim to find a single AR parameter P_j, where $1 \leq j \leq N$, so that its corresponding time series can give the best classifying features. By utilizing only one parameter time series, we reduce the data to $\{(X_i(:, P_j), y_i)\}_{i=1}^{M}$, where $X_i(:, P_j) \in \mathbb{R}^T$, so that the dimensionality decreases and the classification performance increases in comparison with other representations $\{(X_i(:, P_{j'}), y_i)\}_{i=1}^{M}$, where $j' \neq j$.

B. Summarization of Time Series

We represent each time series of length T using 8 summary statistics [17], [18]. First four of them are mean (μ), standard deviation (σ), skewness ($SKEW$), and kurtosis ($KURT$) of the time series. Formulas of these statistics on the time series $P = [v_1, v_2, \ldots, v_T]$ are as follows.

$$\mu(P) = \frac{\sum_{i=1}^{T} v_i}{T} \tag{1}$$

$$\sigma(P) = \sqrt{\frac{\sum_{i=1}^{T}(v_i - \mu(P))^2}{T}} \quad (2)$$

$$SKEW(P) = \frac{\sum_{i=1}^{T}(v_i - \mu(P))^3}{T\sigma(P)^3} \quad (3)$$

$$KURT(P) = \frac{\sum_{i=1}^{T}(v_i - \mu(P))^4}{T\sigma(P)^4} - 3 \quad (4)$$

Then we calculate the first derivative of the time series P, which is given by $P' = [v'_1, v'_2, \ldots, v'_{T-1}]$.

$$v'_i = v_{i+1} - v_i, 1 \leq i \leq T-1$$

Finally, we calculate the same statistics, i.e., mean, standard deviation, skewness, and kurtosis of P'.

$$\mu(P') = \frac{\sum_{i=1}^{T-1} v'_i}{T-1} \quad (5)$$

$$\sigma(P') = \sqrt{\frac{\sum_{i=1}^{T-1}(v'_i - \mu(P'))^2}{T-1}} \quad (6)$$

$$SKEW(P') = \frac{\sum_{i=1}^{T-1}(v'_i - \mu(P'))^3}{(T-1)\sigma(P')^3} \quad (7)$$

$$KURT(P') = \frac{\sum_{i=1}^{T-1}(v'_i - \mu(P'))^4}{(T-1)\sigma(P')^4} - 3 \quad (8)$$

Equations (1) - (8) provide 8 summary statistics of a time series P of length T. These 8 numbers make a vector u which can be thought as a summarized representation of the time series P, and is given by

$$u(P) = [\mu(P), \sigma(P), SKEW(P), KURT(P), \mu(P'), \sigma(P'), SKEW(P'), KURT(P')]$$

This summarization method can be used for feature-based representation of a time series of any length.

C. Parameter Selection and Classification

To assess the classification ability of the time series of each AR parameter, we make datasets $D_j = \{(X_i(:, P_j), y_i)\}_{i=1}^{M}$ for $1 \leq j \leq N$ with given lookback l and span s. Since $D_j \in \mathbb{R}^{T \times M}$, and the performance and runtime of time series-based classifiers depend on the number of dimensions of the vector space, i.e., the length of the time series T, we summarize each time series (column of D_j) by 8 summary statistics. After reducing the data to $D_j \in \mathbb{R}^{8 \times M}$, we use k-NN classifier to distinguish examples of two classes.

We divide the dataset D_j into a training dataset $D_{j\_train} \in \mathbb{R}^{8 \times Tr}$ and a testing dataset $D_{j\_test} \in \mathbb{R}^{8 \times (M-Tr)}$ (discussed in detail in Section V-B). For each test example in D_{j_test}, k nearest training examples from D_{j_train} are found by calculating the Euclidean distance in 8 dimensional space.

TABLE I: Datasets

Tag (*lLsS*)	Lookback *L* (hours)	Maximum span *S* (hours)	# of events	# of flares	# of non-flares
$l12s24$	12	24	3,436	431	3,005
$l24s24$	24	24	3,292	403	2,889

TABLE II: AR parameters used in the experiments

Tag	Description
USFLUX	Total unsigned flux
MEANGAM	Mean angle of field from radial
MEANGBT	Mean gradient of total field
MEANGBZ	Mean gradient of vertical field
MEANGBH	Mean gradient of horizontal field
MEANJZD	Mean vertical current density
TOTUSJZ	Total unsigned vertical current
MEANALP	Mean characteristic twist parameter α
MEANJZH	Mean current helicity (B_z contribution)
TOTUSJH	Total unsigned current helicity
ABSNJZH	Absolute value of the net current helicity
SAVNCPP	Sum of the modulus of the net current per polarity
MEANPOT	Mean photospheric magnetic free energy
TOTPOT	Total photospheric magnetic free energy density
MEANSHR	Mean shear angle
SHRGT45	Fraction of Area with Shear $> 45°$

The class label of the test example is found from the most common class label among its k neighbors of the training dataset [19], [20]. When $k = 1$, the test example is assigned the class label of its nearest training example. If k is even and there is a tie between the numbers of positive and negative nearest neighbors, then the class label of the nearest neighbor is chosen.

By varying the number of neighbors (k) in k-NN classifier, lookback window size, and span window size, we measure the performance metrics discussed in Section V-C. The AR parameter whose summarized time series get consistently better score with k-NN classifier than the summarized time series of other AR parameters can be considered to be the best AR parameter in distinguishing flaring and non-flaring Active Regions.

V. Experimental Evaluation

In this section, we demonstrate our experimental findings. We used Python's Scikit-learn library for using k-NN classifier. In all the experiments, k-NN classifier uses Euclidean distance. The code of the experiments is available at our Github repository. [1]

A. Dataset description

In the dataset tagged by $lLsS$, the time series of a set of AR parameters are collected before L hours of M/X-class flare occurrence (or before L hours of sampling time of non-flaring Active Regions) and these time series are stacked together to make the $mvts$ of the $event$. The length of each time series of the events of the dataset $lLsS$ is $S \times 5$, since the magnetic field values of the AR patches are calculated

[1] http://github.com/hamdi08/Flare_expts_SABID17/

		Predicted class		
		+1	-1	Total
Actual class	+1	TP	FN	P
	-1	FP	TN	N
	Total	P′	N′	P + N

Fig. 3: Confusion matrix

by SHARP (Spaceweather HMI Active Region Patch) in 12 minutes cadence [21]. When $S = 24$ hours, each time series reach maximum length of $24 \times 5 = 120$. Therefore, $l12s24$ and $l24s24$ are two primitive and full datasets used in the experiments (Table I). Datasets $lLsS$ with other spans, where $S < 24$ can be derived by slicing the last $S \times 5$ values from each time series of the events of $lLs24$, where $L \in \{12, 24\}$. In this work, we have used 16 AR parameters shown in Table II, whose formulas can be found in [4].

B. Train/test splitting methodology

In Table III, we have described the splitting strategy in datasets with fixed lookback time. We trained the model with four years data, sampled from January 2011 to December 2014 and tested the model with events sampled from January 2015 to December 2016. Class imbalance ratio for training is 6.14, and for testing is 11.5. Overall 73% events of a dataset are used for training, while rest 27% are used for testing. Since the train/test splitting is done on the basis of the temporal occurrence of the events, in order to calculate the performance measures, unstratified splitting is performed once. Before running the classifier, both training and test datasets are z-normalized.

C. Performance Measures

To compare our classification results with the existing flare prediction studies ([4], [12], [13], [14], [22], [23]), we evaluate 11 performance measures: accuracy, precision (positive and negative), recall (positive and negative), F1 (positive and negative), HSS_1, HSS_2, GS and TSS. Given a set of test examples, we get a confusion matrix as a result of classification (Fig. 3) [20]. The confusion matrix has four entries - TP, TN, FP and FN, where TP (true positive) is the number of positive examples that are correctly labeled, TN (true negative) is the number of negative examples that are correctly labeled, FP (false positive) is the number of negative examples that are misclassified as positive, and FN (false negative) is the number of positive examples that are misclassified as negative. P and N are the numbers of actual positive and negative examples respectively. Since in flare prediction $P << N$, class imbalance problem exists and therefore accuracy is not a good performance measure. In this subsection, we briefly discuss some performance measures such as HSS_1, HSS_2, GS and TSS which are typically used in the evaluation of flare prediction performance.

Heidke Skill Score and Gilbert Score: To deal with the class imbalance problem, two versions of Heidke Skill Score HSS_1 [22] and HSS_2 [9], and Gilbert score GS [9] have been used in previous solar flare prediction literature [4].

$$HSS_1 = \frac{TP + TN - N}{P}$$
$$HSS_2 = \frac{2 \times [(TP \times TN) - (FN \times FP)]}{P \times N' + N \times P'}$$
$$GS = \frac{TP \times (P + N) - P \times P'}{FN \times (P + N) - N \times P'}$$

While HSS_1 measures the improvement of the prediction over the "*always negative class*" prediction, HSS_2 measures the improvement of the prediction over random prediction. GS considers the number of TP obtained by chance.

True Skill Statistic: Since HSS_1, HSS_2 and GS still show some dependence on the class imbalance ratio, Bloomfield et al. [23] defined TSS, which is independent on class imbalance ratio and defined as the difference between true positive rate and false positive rate.

$$TSS = \frac{TP}{P} - \frac{FP}{N}$$

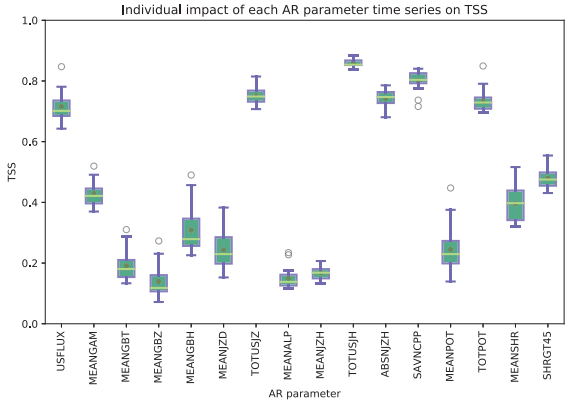
TSS ranges from -1 to $+1$, where random prediction scores 0, perfect prediction scores $+1$, and the prediction that is always wrong scores -1. According to Bobra et al. [4], TSS is the most meaningful measure for performance comparison of different flare prediction studies.

D. Best AR parameter selection

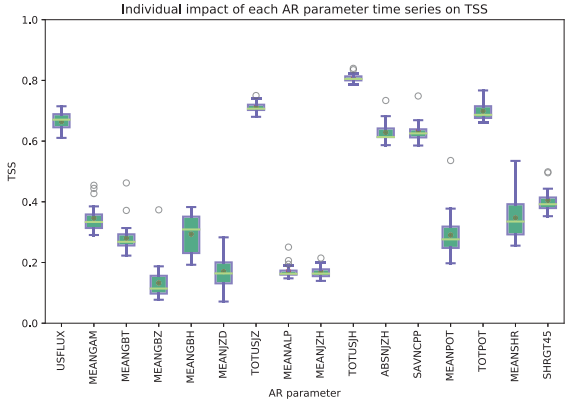
Among the 16 AR parameters, the AR parameter whose corresponding time series give maximum mean TSS with minimum variance after k-NN classification with varying k is considered to be the best AR parameter in terms of the distinguishing ability of the time series. For this experiment, we use both datasets $l12s24$ (Fig. 4a) and $l24s24$ (Fig. 4b) because of the completeness of their time series. For each event of these datasets, we collect the time series of only one parameter P_j at a time, where $1 \leq j \leq 16$ and summarized the time series by 8 summary statistics described in Section IV-B. Then we run k-NN classifier with varying $k = 1, 2, \ldots, 20$ and measure the TSS value for each run. From the boxplots of Fig. 4, it is visible that the summarized time series of *total unsigned current helicity* (TOTUSJH) achieve maximum mean TSS with minimum variance for both lookback settings. This finding is exactly the same as [4], where they found TOTUSJH as the top-1 selected AR parameter based on Fisher criterion. Like [4], our result also indicates that the parameters which calculate sums, e.g., TOTUSJH, TOTUSJZ, TOTPOT, SAVNCPP etc are better than the parameters that calculate the means, e.g., MEANGBZ, MEANGBT, MEANGBH, MEANALP etc. In our later experiments, we only used the time series of the parameter TOTUSJH.

TABLE III: Splitting datasets into train and test sets

Dataset	Training set				Test set			
	Duration	# of events	# of flares	# of non-flares	Duration	# of events	# of flares	# of non-flares
$l12sS$	2011-2014	2,503	349	2,154	2015-2016	933	82	851
$l24sS$		2,408	328	2,080		884	75	809



(a) Time series extracted with lookback 12 hours



(b) Time series extracted with lookback 24 hours

Fig. 4: TSS distributions after 20 k-NN executions on the summarized time series of the individual AR parameters.

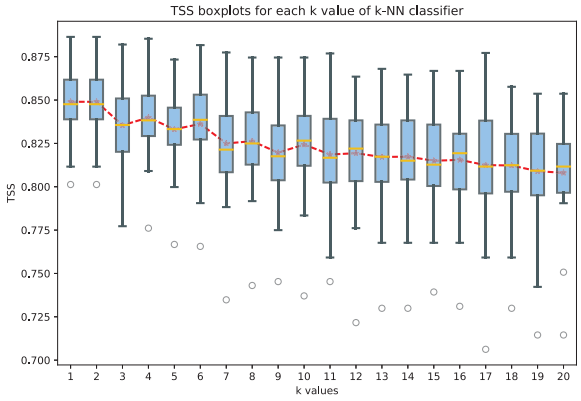


Fig. 5: TSS distributions for different k values of k-NN classification on 24 derived datasets of summarized time series of TOTUSJH.

E. Optimal k in k-NN classifier

Since the number of neighbors (k) in k-NN classifier can affect the classification performance, we look for best performing k value in all datasets. For each lookback L where $L \in \{12, 24\}$ hours, we derive 12 datasets by increasing the span by 2 hours. The derived datasets are $\{l12s2, l12s4, \ldots, l12s24, l24s2, l24s4, \ldots, l24s24\}$. Then for each of these 24 datasets, we extract the summarized time series of the AR parameter TOTUSJH of each event. Finally, for $k = 1, 2, \ldots, 20$, we measure the TSS values in all 24 datasets. From the boxplots of Fig. 5, we see that $k = 1$ and $k = 2$ has the maximum mean TSS. We consider $k = 1$ as optimal k for the later experiments.

F. Comparison with other baselines

In this subsection, we compare our method with two other baselines.

Baseline 1: For each $mvts_i \in \mathbb{R}^{T \times N}$, where $1 \leq i \leq M$, we collect only the last row (latest value of each of the N time series) and consider it as a vector. The resultant vector space has size $N \times M$ (recall that N is the number of AR parameters and M is the number of examples). Unlike other two baselines, this baseline does not consider any span time window. SVM with weighted class weights (cost of the positive class is set greater than that of the negative class since positive examples are comparatively rare) is applied on this vector space. A similar approach was taken in [4]. Since this baseline uses only the ***l***atest ***v***alue of ***e***ach AR ***p***arameter time series, this is tagged as *lvep* in Table IV.

Baseline 2: For each $mvts_i \in \mathbb{R}^{T \times N}$, where $1 \leq i \leq M$, we collect the mean value of each time series (AR parameter). Similar to baseline 1, the resultant vector space has size $N \times M$ and SVM with weighted class cost is used as the classifier. One significant difference between baseline 1 and baseline 2 is, baseline 1 does not use the time series of the AR parameters, while baseline 2 uses the mean (one summary statistic) of the time series of each AR parameter. Since this baseline uses the ***m***ean ***v***alue of ***e***ach AR ***p***arameter time series, this is tagged as *mvep* in Table IV.

In the proposed method, each $mvts_i$ is represented by the summarized representation of full span (24 hours) time series of the best AR parameter TOTUSJH. The resultant vector space has size $8 \times M$. Finally, k-NN classifier with $k = 1$ is run on this vector space. Since only the ***b***est AR ***p***arameter is used, we tag the proposed method as *bp* in Table IV.

TABLE IV: Comparison of the proposed method with two other baselines in 11 performance measures

Dataset tag	$l12sS$			$l24sS$		
Baseline tag *	*lvep*	*mvep*	*bp*	*lvep*	*mvep*	*bp*
Span window used	0 hour	24 hours	24 hours	0 hour	24 hours	24 hours
AR parameters used	All	All	TOTUSJH	All	All	TOTUSJH
Classifier	SVM (weighted class cost)	SVM (weighted class cost)	k-NN (k=1)	SVM (weighted class cost)	SVM (weighted class cost)	k-NN (k=1)
Accuracy	0.919	0.919	**0.975**	0.906	0.913	**0.975**
Precision (positive)	0.520	0.521	**0.831**	0.472	0.493	**0.853**
Precision (negative)	**0.994**	0.992	0.991	**0.991**	**0.991**	0.986
Recall (positive)	**0.939**	0.927	0.902	**0.907**	**0.907**	0.853
Recall (negative)	0.917	0.918	**0.982**	0.906	0.913	**0.986**
F1 (positive)	0.670	0.667	**0.865**	0.621	0.638	**0.853**
F1 (negative)	0.954	0.954	**0.986**	0.946	0.950	**0.986**
HSS_1	0.073	0.073	**0.720**	-0.107	-0.027	**0.707**
HSS_2	0.627	0.624	**0.852**	0.573	0.594	**0.840**
GS	0.457	0.454	**0.742**	0.402	0.422	**0.724**
TSS	0.856	0.844	**0.885**	0.813	0.820	**0.840**

* *lvep*: latest value of each parameter, *mvep*: mean value of each parameter, *bp*: best parameter

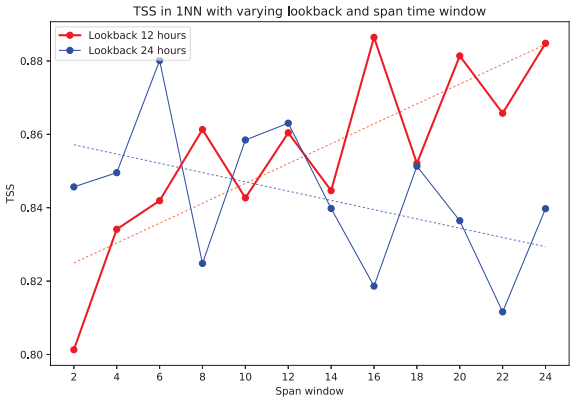


Fig. 6: Variation of TSS after running k-NN (k=1) with different lookback and span settings on the time series of TOTUSJH

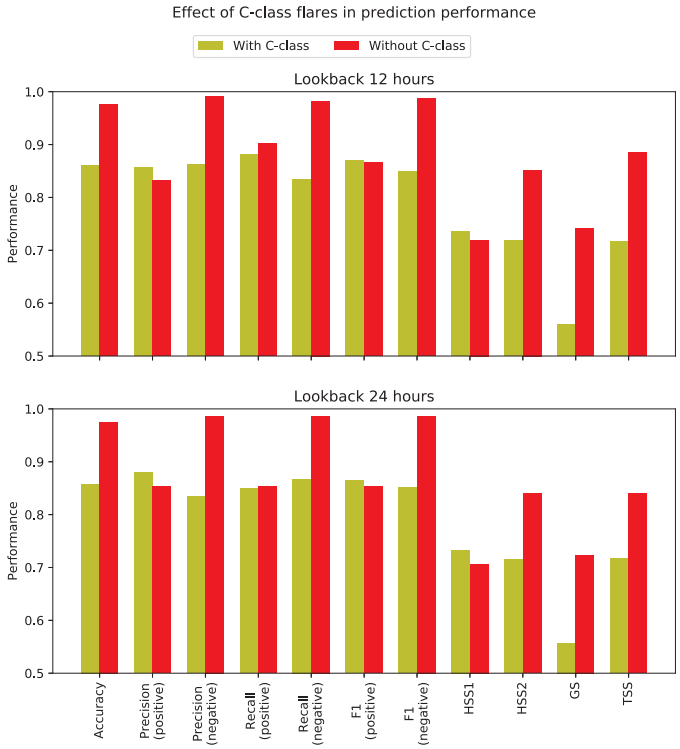


Fig. 7: Performance after k-NN (k=1) execution on the summarized time series of TOTUSJH with/without considering C-class flares

In Table IV, we show the performance comparison of three methods *lvep*, *mvep*, and *bp* using 11 performance measures. For both datasets of lookback 12 and 24 hours, except the precision (negative) and recall (positive), our proposed method *bp* performs best in all other measures. The performance of *lvep* and *mvep* are almost the same in all measures, but *lvep* performs slightly better in small lookback window while *mvep* has slightly better performance in large lookback window. Although *lvep* and *mvep* are almost as good as *bp* in some measures such as TSS, they exhibit poor performance in other measures such as HSS_1, HSS_2, and GS in comparison with *bp*.

G. Span window-based performance

Given a fixed lookback window L, how does the classification performance change if we change the span window size? Fig. 6 shows the change of TSS value after running k-NN ($k = 1$) with varying lookback and span windows of the time series of TOTUSJH. When the lookback is small such as 12 hours, the temporal proximity to the actual event is small, and we observe a linearly increasing trend (dashed red straight line) with the increase of span window. On the contrary, when the lookback is large, e.g., 24 hours, we observe a linearly decreasing trend of TSS (dashed blue straight line). Although the increase or decrease of TSS is not obvious after increasing span since the extension of time series with new values might improve/deteriorate the performance, this overall linear increasing/decreasing trend is an indication of good time series quality of the AR parameter TOTUSJH.

H. Effect of C-class flares in classification performance

In this experiment, we added the Active Regions with having one or more C-class flares (less intense flares in

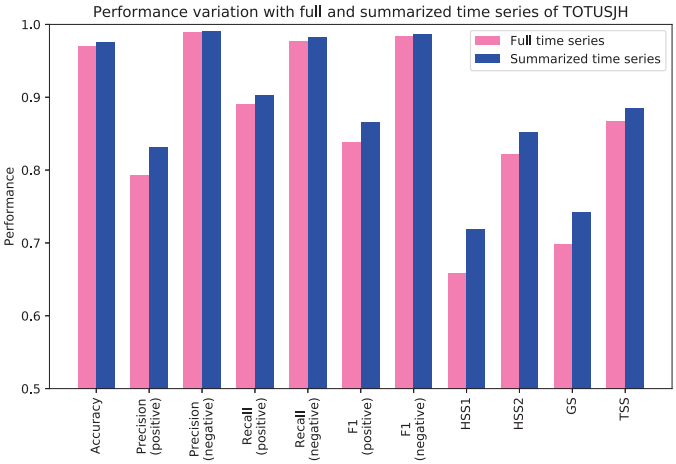


Fig. 8: Comparison of summarized representation and full-length representation of TOTUSJH time series in classification performance

comparison with X/M-class flares) in the positive class. The number of sampled C-class flares for lookback 12 hours is 5,527 and for lookback 24 hours is 5,353. The inclusion of C-class flares changes the class imbalance ratio stated in Section V-B since the number of positive examples (X, M and C-class flares) exceeds the number of non-flares (Active Regions with no flares occurring during the disk crossing). Fig. 7 shows that in both lookback settings with full span window, the inclusion of C-class flares has a slightly negative impact in almost all performance metrics, while we consider k-NN $(k = 1)$ classification on the summarized time series of the AR parameter TOTUSJH. This result agrees with the finding of Bloomfield et al. [23] which says that HSS_1 increases as a result of including of C-class flares in positive class, but it results in the decrease in TSS.

I. Effect of time series summarization

Fig. 8 provides the rationale for our preference of summarized representation of the time series over the full-length time series. This experiment is run on the $l12s24$ dataset. It is visible that if we consider the summarized representation instead of the full-length time series of the AR parameter TOTUSJH, we get an increase in all performance metrics. The number of dimensions in the full-length time series representation is 120, while in case of the summarized representation the number of dimensions is only 8. Since k-NN classifier greatly depends on the number of dimensions in the vector space, the runtime of the classifier is more than 10 times bigger for full-length time series representation than that of summarized representation.

J. Performance using other classifiers

When flaring and non-flaring Active Regions are represented by the summarized representation of the time series of the AR parameter TOTUSJH, other sophisticated classifiers can achieve even better performance than k-NN $(k = 1)$. Table V shows the 11 performance metrics found after running the classifiers SVM, random forest, and naïve Bayes on $l12s24$ and $l24s24$ datasets with the default settings of their Scikit-learn implementations. We also show the performance of k-NN $(k = 1)$ in Table V for comparison. These classifiers and their hyperparameters (according to their Scikit-learn specifications) are listed below.

- **SVM:** C=1.0, kernel=rbf, gamma=1/8, class_weight=None.
- **Random forest:** n_estimators=10, criterion=gini, max_depth=None, class_weight=None.
- **Naïve Bayes:** priors=None.
- **k-NN:** number of neighbors (k) = 1.

Experimental results that are shown in Table V show the robustness of our data representation, i. e., the summarized representation of the time series of AR parameter TOTUSJH, in classifying flaring and non-flaring Active Regions regardless of the classifier. Although some of these classification models have better performance than k-NN, k-NN is found to be more interpretable than these classifiers with respect to lookback and span windows of the time series (Fig. 6).

VI. Conclusion

In this work, we show a novel way of predicting that an Active Region of the Sun might lead to an X-class or M-class flare by leveraging the time series behavior of the magnetic field parameters. We present a formal data model for representing flaring and non-flaring Active Regions using multivariate time series, where each time series, extracted before the *lookback* time of the occurrence of the event and collected throughout the *span* period of time, represents one AR parameter. We use k-NN classifier for classifying the summarized representations of the time series of the AR parameter *total unsigned current helicity*, and exhibit better performance than considering all parameters without time series. We also show the robustness of this representation using other classifiers.

In the future, we aim to work with the augmented dataset which comprises all the vector magnetogram-based AR parameters listed in [4] and analyze the impact of time series on the class imbalance ratio. We are also interested to use the time series of *total unsigned current helicity* in flare regression problem, i.e., predicting the intensity and/or time of flare. The automated learning of the time series features that can be used for time series summarization can be another interesting problem. Ranking the AR parameters by their time series quality for classification, and using the time series of top-K AR parameters can give the solar physicists more insights on the relationship of AR magnetic field and flare occurrence.

Acknowledgment

This project has been supported in part by funding from the Division of Advanced Cyberinfrastructure within the Directorate for Computer and Information Science and Engineering, the Division of Astronomical Sciences within the Directorate for Mathematical and Physical Sciences, and the Division of Atmospheric and Geospace Sciences within the Directorate for

TABLE V: Performance by other classifiers on the summarized time series of TOTUSJH

Dataset tag	$l12s24$				$l24s24$			
Classifier	SVM	Random forest	Naïve Bayes	k-NN (k=1)	SVM	Random forest	Naïve Bayes	k-NN (k=1)
Accuracy	**0.99**	0.97	0.98	0.98	**0.99**	0.97	0.98	0.98
Precision(positive)	**0.96**	0.8	0.8	0.83	**0.98**	0.81	0.84	0.85
Precision(negative)	0.99	0.99	**1.0**	0.99	**0.99**	**0.99**	**0.99**	**0.99**
Recall(positive)	0.93	0.9	**0.95**	0.9	0.87	0.87	**0.91**	0.85
Recall(negative)	**1.0**	0.98	0.98	0.98	**1.0**	0.98	0.98	0.99
F1(positive)	**0.94**	0.85	0.87	0.87	**0.92**	0.84	0.87	0.85
F1(negative)	**0.99**	0.98	**0.99**	**0.99**	**0.99**	0.98	**0.99**	**0.99**
HSS_1	**0.89**	0.67	0.72	0.72	**0.85**	0.67	0.73	0.71
HSS_2	**0.94**	0.83	0.86	0.85	**0.92**	0.82	0.86	0.84
GS	**0.88**	0.71	0.75	0.74	**0.84**	0.7	0.75	0.72
TSS	0.92	0.88	**0.93**	0.89	0.87	0.85	**0.89**	0.84

Geosciences, under NSF award #1443061. It was also supported in part by funding from the Heliophysics Living With a Star Science Program, under NASA award #NNX15AF39G.

The SDO data is available courtesy of NASA/SDO (https://sdo.gsfc.nasa.gov/) and the Atmospheric Imaging Assembly (AIA), Extreme Ultraviolet Variability Experiment (EVE), and Helioseismic and Magnetic Imager (HMI) science teams.

REFERENCES

[1] J. Eastwood, E. Biffis, M. Hapgood, L. Green, M. Bisi, R. Bentley, R. Wicks, L.-A. McKinnell, M. Gibbs, and C. Burnett, "The economic impact of space weather: Where do we stand?" *Risk Analysis*, vol. 37, no. 2, pp. 206–218, 2017.

[2] "Solar flare captured by SDO," https://www.nasa.gov/content/goddard/nasa-releases-images-of-1st-notable-solar-flare-of-2015, [Online; accessed 30-may-2017].

[3] K. Leka and G. Barnes, "Photospheric magnetic field properties of flaring versus flare-quiet active regions. ii. discriminant analysis," *The Astrophysical Journal*, vol. 595, no. 2, p. 1296, 2003.

[4] M. G. Bobra and S. Couvidat, "Solar flare prediction using SDO/HMI vector magnetic field data with a machine-learning algorithm," *The Astrophysical Journal*, vol. 798, no. 2, p. 135, 2015.

[5] N. Nishizuka, K. Sugiura, Y. Kubo, M. Den, S. Watari, and M. Ishii, "Solar flare prediction model with three machine-learning algorithms using ultraviolet brightening and vector magnetograms," *The Astrophysical Journal*, vol. 835, no. 2, p. 156, 2017.

[6] K.-j. Kim, "Financial time series forecasting using support vector machines," *Neurocomputing*, vol. 55, no. 1, pp. 307–319, 2003.

[7] T. M. Rath and R. Manmatha, "Word image matching using dynamic time warping," in *Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on*, vol. 2. IEEE, 2003, pp. II–II.

[8] P. S. McIntosh, "The classification of sunspot groups," *Solar Physics*, vol. 125, no. 2, pp. 251–267, 1990.

[9] J. P. Mason and J. Hoeksema, "Testing automated solar flare forecasting with 13 years of michelson doppler imager magnetograms," *The Astrophysical Journal*, vol. 723, no. 1, p. 634, 2010.

[10] Y. Cui, R. Li, L. Zhang, Y. He, and H. Wang, "Correlation between solar flare productivity and photospheric magnetic field properties," *Solar Physics*, vol. 237, no. 1, pp. 45–59, 2006.

[11] J. Jing, H. Song, V. Abramenko, C. Tan, and H. Wang, "The statistical relationship between the photospheric magnetic parameters and the flare productivity of active regions," *The Astrophysical Journal*, vol. 644, no. 2, p. 1273, 2006.

[12] O. W. Ahmed, R. Qahwaji, T. Colak, P. A. Higgins, P. T. Gallagher, and D. S. Bloomfield, "Solar flare prediction using advanced feature extraction, machine learning, and feature selection," *Solar Physics*, pp. 1–19, 2013.

[13] D. Yu, X. Huang, H. Wang, and Y. Cui, "Short-term solar flare prediction using a sequential supervised learning method," *Solar Physics*, vol. 255, no. 1, pp. 91–105, 2009.

[14] H. Song, C. Tan, J. Jing, H. Wang, V. Yurchyshyn, and V. Abramenko, "Statistical assessment of photospheric magnetic features in imminent solar flare predictions," *Solar Physics*, vol. 254, no. 1, pp. 101–125, 2009.

[15] A. Al-Ghraibah, L. Boucheron, and R. McAteer, "An automated classification approach to ranking photospheric proxies of magnetic energy build-up," *Astronomy & Astrophysics*, vol. 579, p. A64, 2015.

[16] R. Qahwaji and T. Colak, "Automatic short-term solar flare prediction using machine learning and sunspot associations," *Solar Physics*, vol. 241, no. 1, pp. 195–211, 2007.

[17] A. Nanopoulos, R. Alcock, and Y. Manolopoulos, "Feature-based classification of time-series data," *International Journal of Computer Research*, vol. 10, no. 3, pp. 49–61, 2001.

[18] E. J. Keogh, "A decade of progress in indexing and mining large time series databases," in *Proceedings of the 32nd International Conference on Very Large Data Bases, Seoul, Korea, September 12-15, 2006*, 2006, p. 1268. [Online]. Available: http://dl.acm.org/citation.cfm?id=1164262

[19] T. Cover and P. Hart, "Nearest neighbor pattern classification," *IEEE transactions on information theory*, vol. 13, no. 1, pp. 21–27, 1967.

[20] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques, 3rd edition*. Morgan Kaufmann, 2011. [Online]. Available: http://hanj.cs.illinois.edu/bk3/

[21] "Joint Science Operations Center (JSOC)," http://jsoc.stanford.edu/, [Online; accessed 01-October-2017].

[22] G. Barnes and K. Leka, "Evaluating the performance of solar flare forecasting methods," *The Astrophysical Journal Letters*, vol. 688, no. 2, p. L107, 2008.

[23] D. S. Bloomfield, P. A. Higgins, R. J. McAteer, and P. T. Gallagher, "Toward reliable benchmarking of solar flare forecasting methods," *The Astrophysical Journal Letters*, vol. 747, no. 2, p. L41, 2012.