Using Machine Learning Methods to Forecast if Solar Flares Will Be Associated with CMEs and SEPs

Fadil Inceoglu1,2,3, Jacob H. Jeppesen1, Peter Kongstad2, Néstor J. Hernández Marcano1, Rune H. Jacobsen1, and Christoffer Karoff1,2

1 Department of Engineering, Aarhus University, Finlandsvej 22, DK-8200 Aarhus N, Denmark; fadil@eng.au.dk
2 Department of Geoscience, Aarhus University, Høegh-Guldberg Gade 2, DK-8000 Aarhus C, Denmark
3 Stellar Astrophysics Centre, Department of Physics and Astronomy, Aarhus University, Ny Munkegade 120, DK-8000 Aarhus C, Denmark

Received 2018 April 16; revised 2018 May 23; accepted 2018 May 23; published 2018 July 12

Abstract
Among the eruptive activity phenomena observed on the Sun, those that threaten human technology the most are flares with associated coronal mass ejections (CMEs) and solar energetic particles (SEPs). Flares with associated CMEs and SEPs are produced by magnetohydrodynamical processes in magnetically active regions (ARs) on the Sun. However, these ARs do not only produce flares with associated CMEs and SEPs, they also lead to flares and CMEs, which are not associated with any other event. In an attempt to distinguish flares with associated CMEs and SEPs from flares and CMEs, which are unassociated with any other event, we investigate the performances of two machine learning algorithms. To achieve this objective, we employ support vector machines (SVMs) and multilayer perceptrons (MLPs) using data from the Space Weather Database of Notification, Knowledge, Information of NASA Space Weather Center, the Geostationary Operational Environmental Satellite, and the Space-Weather Heliospheric and Magnetic Imager Active Region Patches. We show that True Skill Statistics (TSS) and Heidke Skill Scores (HSS) calculated for SVMs are slightly better than those from the MLPs. We also show that the forecasting time frame of 96 hr provides the best results in predicting if a flare will be associated with CMEs and SEPs (TSS = 0.92 ± 0.09 and HSS = 0.92 ± 0.08). Additionally, we obtain the maximum TSS and HSS values of 0.91 ± 0.06 for predicting that a flare will not be associated with CMEs and SEPs for the 36 hr forecast window, while the 108 hr forecast window gives the maximum TSS and HSS values for predicting that CMEs will not be accompanying any events (TSS = HSS = 0.98 ± 0.02).

Key words: Sun: activity – Sun: coronal mass ejections (CMEs) – Sun: flares

1. Introduction
The Sun potentially endangers modern civilization through large solar eruptions. Predicting and monitoring large solar flares, coronal mass ejections (CMEs), and solar energetic particles (SEPs), which introduce enormous quantities of particles and energy to Earth’s vicinity and into its atmosphere, is therefore of crucial importance.

The strongest observed flare and accompanying CME was the Carrington event that occurred in 1859, which was about twice as big as the strongest events observed during the space era (Carrington 1859; Cliver & Dietrich 2013). Since the 1950s, a series of the most powerful flares and their accompanying CMEs have occurred between mid-October to early 2003 November peaking around October 28 and 29 (the so-called Halloween solar storms), which caused radio blackouts on Earth and problems in different instruments of some satellites, such as star trackers. These flares even killed the power supply of the Japanese Earth-resource satellite, the MIDORI-II (ADEOS-2), and left it inoperative. The effects of the Halloween solar storms extended beyond the Earth to Mars and caused the Mars Odyssey spacecraft to go into deep safe-mode (Lopez et al. 2004).

CMEs originate from large coronal loop structures, which contain plasma and magnetic fields, expanding from the Sun into the interplanetary medium. They occur on a quasi-regular basis and are the largest-scale eruptive phenomena in our solar system (Chen 2011; Webb & Howard 2012; Kilpua et al. 2017). Flares, on the other hand, are eruptive events that occur in the solar atmosphere with energies ranging from $10^{28}$ to $10^{32}$ erg (Shibata & Magara 2011). Based on their peak fluxes in the 1–8 Å range, X-ray flares near Earth, as measured by the XRS instrument on the Geostationary Operational Environmental Satellite (GOES), are classified as A, B, C, M, and X. Each flare class is also divided into subclasses linearly scaled from 1 to 9, such as M4 or X9 (Schrijver & Siscoe 2010). Strong flares and CMEs occasionally cause SEP events, where protons, electrons, and heavier nuclei are accelerated to energies ranging between a few tens of keVs to GeVs (Reames 2013). Strong SEPs cause nuclear cascades in the upper atmosphere of the Earth (Reames et al. 2013). Flares are often associated with CMEs as they are thought to have a common magnetically driven mechanism instead of one causing the other (Webb & Howard 2012). However, there are CMEs and flares that are not associated with one another (Chen 2011; Webb & Howard 2012). Observations show that the speeds and energies of CMEs are higher when they are accompanied by bright flares in comparison to those not accompanied by flares (Chen 2011; Webb & Howard 2012). Magnetically strong regions on the photosphere, the so-called active regions (ARs), are often the source regions of flares and CMEs (Chen 2011; van Driel-Gesztelyi & Green 2015). Strong ARs are generally large and evolve rapidly with lifetimes varying from days to months (van Driel-Gesztelyi & Green 2015), and they exhibit complex magnetic geometry (Benz 2008). Vector magnetograms, which measure the line-of-sight (LoS) magnetic field separately from the image-plane, allow us to calculate the physical indices of the ARs, such as magnetic helicity, magnetic shear angles, proxies for free
energy and magnetic fluxes of the ARs, and polarity-separation lines (Leka & Barnes 2003; Schrijver 2007; Moore et al. 2012).

Previous studies on predicting solar eruptive phenomena used photospheric magnetic field data to calculate the physical indices of ARs, and they link these physical indices to the occurrences of flares and CMEs (Leka & Barnes 2003; Schrijver 2007; Moore et al. 2012). Recently, machine learning (ML) methods, such as support vector machines (SVMs) and neural networks (NNs), have been used in predicting flares, CMEs, and SEPs (Yu et al. 2009; Yuan et al. 2010; Ahmed et al. 2013; Bobra & Couvidat 2015; Boucheron et al. 2015; Bobra & Ilonidis 2016; Florios et al. 2018).

Yu et al. (2009) used LoS magnetogram data from the Michelson Doppler Imager (MDI) on board the Solar and Heliospheric Observatory (SOHO) to predict flare occurrences within a forecast window of 48 hr. To achieve this objective, the authors used data from ARs, which generated at least one $\geq$C1-class flare. They divided this data into two subclasses as flaring and non-flaring regions based on their total importance threshold value (Equation (1) in Yu et al. 2009) and used it in a learning vector quantization NNs algorithm. Yuan et al. (2010), on the other hand, used the same data to predict flare classes A, B, C, M, and X via SVMs.

To predict $\geq$C1-class flares 24 and 48 hr prior to their occurrences, Ahmed et al. (2013) used magnetic feature data of flaring and non-flaring regions from the Helioseismic and Magnetic Imager (HMI) on board the Solar Dynamics Observatory (SDO) in a cascade correlation NN algorithm. The authors defined an AR as non-flaring if it did not produce a flare within 24 hr, for their 24 hr prediction window, while, for the 48 hr forecast window, they defined an AR as non-flaring if it does not produce a flare within $\pm$48 hr after the sampling time. Boucheron et al. (2015), on the other hand, used the SOHO/MDI data of flaring and non-flaring ARs in an SVM regressor.

To predict occurrences of flares larger than M1 class, Bobra & Couvidat (2015) used SDO/HMI’s definitive flaring and non-flaring AR data, which are defined similarly to those of Ahmed et al. (2013), in SVMs at time delays 48 and 24 hr, respectively. In addition, Bobra & Ilonidis (2016) used the SDO/HMI’s definitive AR data that produce flares and flares with associated CMEs in SVMs to predict whether a flare will be associated with CMEs within a 24 hr forecast window.

Florios et al. (2018), on the other hand, calculated physical features of flaring and non-flaring ARs, which they identified on the SDO/HMI’s near-real-time vector magnetogram data. The flaring and non-flaring ARs are defined based on whether they produce a flare within a 24 hr forecasting window or not. The authors used this data in SVMs, multilayer perceptrons (MLPs), which is based on NNs, and also decision tree algorithms to predict occurrences of $\geq$M1-class and $\geq$C1-class flares.

In this study, we aim to distinguish flares with associated CMEs and SEPs from flares and CMEs without any accompanying events. To predict which event will be produced from an AR, all of which lead to solar eruptions, we investigate the performances of two ML algorithms, SVMs and MLPs, based on the vector magnetic field data observed with the HMI on board the SDO (Schou et al. 2012). The results will also highlight the discriminative potential of the Space-Weather Heliospheric and Magnetic Imager Active Region Patches (SHARPs) data among the three classes. Section 2 describes the data used in this study, while the performed analyses, including brief explanations of the ML methods, are presented in Section 3. Results from the analyses are presented in Section 4, and the discussion and conclusions are given in Section 5.

2. The SDO/HMI Data

To predict which solar eruptive phenomena will be generated from an AR, which is known to have generated only flares, flares with associated CMEs and SEPs, or only CMEs, via SVMs and MLPs, we use data from the Space Weather Database Of Notification, Knowledge, Information (DONKI)4 of NASA Space Weather Research Center, for a period spanning from 2010 January 01 to 2018 January 31. DONKI contains flare data with their classes, source AR numbers, and their start, peak, and end times. This database also shows whether a flare is accompanied by CMEs and/or SEPs. Similar to the flare data, DONKI also provides CME data with their speeds, types, and start times as well as whether they are associated with any flares and/or SEPs. For some events however, DONKI does not provide an AR number although an M- or X-class flare with an accompanying CME is listed, for example, the M3.0-class flare with an accompanying CME and SEP that occurred on 2015 March 06. To fill the unregistered AR numbers in DONKI for these events and also to double check the peak times and the AR numbers of flares, we use flare data from GOES via SunPy Python package v0.8.2 (SunPy Community et al. 2015).

Based on the DONKI and the GOES databases, we have 347 flares that are not associated with CMEs and SEPs, and 179 flares associated with CMEs and SEPs (Figure 1(a)). Further, there are 376 CMEs in total, 174 of which are not associated with any other event (Figure 1(b)).

Figure 1. Top panel shows the distributions of the flare classes of flares (orange) and flares with CMEs and SEPs (red), while the bottom panel shows the distributions of the CME speeds for CME-only events (green) and flares with CMEs and SEPs (red). Note that the y-axes in all panels have logarithmic scales.

4 http://kauai.ccmc.gsfc.nasa.gov/DONKI/
Following the verification of the flare data from the DONKI and the \textit{GOES} databases, we use publicly available SHARPs data from the Joint Science Operations Center,\footnote{http://jsoc.stanford.edu} spanning from 2010 January 01 to 2018 January 31. We select the SHARPs data based on four criteria following Bobra & Ilonidis (2016), so that the SHARPs data has to be: (i) disambiguated, (ii) taken while the SDO’s orbital velocity $<\%0{~}\text{m s}^{-1}$, (iii) of a high-quality, meaning data has reliable Stokes vectors (observables during good conditions), and (iv) within $\pm70^\circ$ longitudinal band during \textit{GOES} peak time, as beyond this band, the signal-to-noise ratio in the vector magnetic field data decreases significantly. The SHARPs data contain vector magnetic field measurements of the ARs and 18 keywords, which are listed in Table 1 together with their definitions. These keywords parameterize the measured physical quantities as well as proxies of physical quantities (for details, see Bobra et al. 2014).

The 18 physical features of the ARs (Table 1) for the three subsets are then compiled at a time before the starting times of the events. This time gap is defined as the time delay ($\Delta t$), which ranges from 12 to 120 hr with 12 hr intervals. This means that we use conditions $\Delta t$ hours before the event occurs to predict whether this event will be a flare without associated events, flare with associated CMEs and SEPs or CMEs without associated events. For each time delay iteration, the data from the DONKI and \textit{GOES} databases goes through the same data selection criteria, and this causes the sample size of each of the three class to change (Table 2). This situation is directly related to the temporal evolution and motion of the ARs on the solar disk as well as availability of the SHARPs data for a given time delay.

The flares that occurred during the study period cluster in M-class flares (Figure 1(a)). The flares unassociated with any other events center around M1.0 class, whereas flares associated with CMEs and SEPs cluster around M5.0-class and extend up to X9.9 (Figure 1(a)).

The distribution of the CME speeds ranges between 90 and 2650 km s$^{-1}$. The speeds of the CMEs that are unassociated with an event center around 500 km s$^{-1}$, while the speeds of the CMEs associated with flares and SEPs center around 1000 km s$^{-1}$ and reach up to 2650 km s$^{-1}$ (Figure 1(b)).

The distribution of the flare classes and the CME speeds indicate that we can separate our data into three subsets as (i) flares only, (ii) flares with associated CMEs and SEPs, and (iii) CMEs only. These three subsets represent the classes for our multi-class classification problem, which we aim to sort out using the SVM and MLP algorithms. It must be noted that the size of the underlying data from the SDO/HMI is smaller than those used in previous studies based on the SOHO/MDI, mission duration of which has reached $\sim$22 years. Additionally, the current solar cycle 24 is quiet in nature and does not generate many eruptions compared to previous stronger cycles.

### 3. Analyses

Before using the 18 physical parameters of the ARs in further analyses for each $\Delta t$ hours delay, we standardize all of the data according to their median and standard deviation values. The reason for this approach is to make sure that data with a small sample size is well represented, the distribution of which might sometimes be left or right skewed. For larger sample sizes, the median and the mean values overlap.

#### 3.1. ML Algorithms

To investigate which ML method provides better predictions of our three classes, we use two of the most popular ML algorithms: (i) SVMs (Cortes & Vapnik 1995), and (ii) MLPs (Hornik et al. 1989) provided by the\textit{ scikit-learn} software package v0.19.1 (Pedregosa et al. 2011).

##### 3.1.1. Support Vector Machines

SVMs are initially designed to solve binary ($l = 2$) classification problems, and employing them to multi-class ($l > 2$) classification problems requires different approaches, where they are generally fragmented into series of different binary classification problems (Hsu & Lin 2002). In this study, we use the one-versus-rest approach (Vapnik 1998), which creates $l$ separate binary classifiers for $l$ number of classes. The $m$th binary SVM classifier is then trained using the data from

### Table 1

| Keyword     | Definition                                      |
|-------------|-------------------------------------------------|
| ABSNJZH     | Absolute value of the net current helicity      |
| R_VALUE     | Sum of flux near polarity inversion line        |
| AREA_ACR    | Area of strong field pixels                     |
| MEANSHR     | Mean shear angle                                 |
| TOTPOT      | Total photospheric magnetic free energy density  |
| SAVNCPP     | Sum of the modulus of the net current per polarity |
| TOTUSIZ     | Total unsigned vertical current                 |
| MEANIZID    | Mean vertical current density                   |
| MEANGAM     | Mean angle of field from radial                 |
| MEANALP     | Mean characteristic twist parameter, $\alpha$    |
| MEANGBH     | Mean gradient of horizontal field               |
| MEANGJH     | Mean unsigned current helicity                  |
| SHRGT45     | Fraction of Area with Shear $>$45$^\circ$       |
| MEANPOT     | Mean photospheric magnetic free energy          |
| MEANJZH     | Mean current helicity ($B_c$, contribution)     |
| MEANGBT     | Mean gradient of total field                    |
| USFLUX      | Total unsigned flux                             |

Note: The \textit{Keyword} column indicates the name of the FITS header keyword in the SHARP data series.

#### Table 2

| Time Delay | Flares | Flares w/ CMEs & SEPs | CMEs |
|------------|--------|-----------------------|------|
| 12 hr      | 228    | 103                   | 97   |
| 24 hr      | 237    | 100                   | 94   |
| 36 hr      | 239    | 105                   | 102  |
| 48 hr      | 236    | 100                   | 99   |
| 60 hr      | 226    | 101                   | 93   |
| 72 hr      | 215    | 95                    | 88   |
| 84 hr      | 205    | 91                    | 87   |
| ...        | ...    | ...                   | ...  |
| 120 hr     | 177    | 76                    | 76   |
the $m$th class as positive (+1), whereas the remaining $l - 1$ number of classes are regarded as negative (−1) examples (Hsu & Lin 2002). The SVM then classifies the data by placing a separating hyperplane with the maximum distance between the classes of the data.

Let us consider a multi-class classification reduced to a binary class via one-versus-rest approach for an $m$th class. For a vector of $P$ predictor consisting of training data at observation $i$ is given as a pair $(x_i, y_i)$, where $i = 1, ..., n$, and $x_i \in \mathbb{R}^p$ and its class $y_i \in \{+1, -1\}$, then the SVM solves the optimization problem as follows (Hsu & Lin 2002):

$$\min_{w, b, \xi} \frac{1}{2} w^T w + C \sum_{i=1}^{n} \xi_i,$$

subject to $y_i (w^T \phi(x_i) + b) \geq 1 - \xi_i$,  

$$\xi_i \geq 0, i = 1, ..., n,$$  

(1)

where $w$ is the weight vector, while $\phi(x_i)$ is an unknown function included in a known kernel function that maps the training vectors $x_i$ into a higher dimensional space. $C > 0$ is the regularization parameter, which compromises misclassification of training examples to make the decision surface simpler. Lower $C$ values make the decision surface smooth, while higher values enable the algorithm to minimize the errors on the training examples and maximize the separation margins (Cortes & Vapnik 1995). The kernel function is defined as the inner product of the data with itself for different pairs of observations $i$ and $j$, $K(x_i, x_j) = \phi(x_i)^T \phi(x_j)$ (Florios et al. 2018). In our study, we use a Radial Basis Function (RBF) kernel that is defined as

$$K(x_i, x_j) = \exp(-\gamma ||x_i - x_j||^2),$$  

(2)

where $\gamma > 0$ defines how much influence a single training example has on the classification. If $\gamma$ has a large value, then the other examples need to be closer to be influenced, whereas smaller values can make the model too constrained, which causes the model to not capture the complexity of the underlying data. A more detailed informational study on how to solve Equations (1) and (2) can be found elsewhere (Cortes & Vapnik 1995; Vapnik 1998).

### 3.1.2. Multilayer Perceptrons

The MLP is a feed-forward NN that classifies multi-dimensional data into $l$ different classes. The multi-class classification problem in MLPs can be regarded as a multinomial logistic regression, where the output of the NN is the posterior probability that the input data belongs to a particular class. The estimated posterior probability distribution of a categorical random variable depends not only on a data point from a random feature, but also on the weights of the neurons, which are the basic processing units (MacKay 2005). As a feed-forward network, the MLP performs nonlinear parameterized mapping from an input $I$ to an output that is a continuous function of the input and the weights $O = \{I; \omega, A\}$, where $\omega$ is the weight and bias parameter vector, and $A$ is the architecture of the network defining the number of nodes in every layer (Florios et al. 2018).

In this study, our MLPs have an input layer with $m = 18$ inputs $I_m$ and bias $\theta^{(1)}$, a single hidden layer with $j$ hidden nodes $H_j$ and a bias $\theta^{(2)}$, and an output layer $O_l$ that only post-processes a data point to give an estimate of the posterior probability. The architecture is given below following MacKay (2005) and Florios et al. (2018)

$$a^{(l)}_i = \sum_m \omega^{(l)}_{im} I_m + \theta^{(l)}_i; H_j = f(a^{(l)}_i),$$

$$a^{(2)}_i = \sum_j \omega^{(2)}_{ij} H_j + \theta^{(2)}_i; O_l = g(a^{(2)}_i),$$  

(3)

where $f(a) = \frac{1}{1 + \exp(-\frac{a}{1 + \exp(s(a))})}$ is the logistic activation function, which defines the response of a neuron in the hidden layer to a stimulus obtained from its input (Hinton 1989). The index $m$ used for the inputs, $j$ and $i$, run over the hidden units and outputs, respectively. The weights $\omega$ and biases $\theta$ define the parameter vector $\omega$ (MacKay 2005).

The function $g(a_i)$ coupled with the output layer $O_l$ is called the softmax activation function because our MLP is designed to solve a multi-class classification problem, which is given as (MacKay 2005)

$$g(a_i) = \frac{\exp(a_i)}{\sum_k \exp(a_k)},$$  

(4)

where $a_i$ represents the $i$th element of the input to softmax, corresponding to class $l$, and $k$ is the number of classes.

The MLPs are trained using a data $D = \{I^{(n)}; o^{(n)}\}$ by adjusting $\omega$ to minimize the negative log-likelihood function given as (MacKay 2005)

$$G(\omega) = -\sum_n \sum_i t^{(n)}_i \ln O_l(I^{(n)}; \omega),$$  

(5)

where $I^{(n)}$ is the predictor matrix, $t^{(n)}_i$ is the target vector, and $n = 1, ..., N$ is the observations.

Minimizing the log-loss function is equivalent to obtaining the maximum likelihood estimator of the weights and biases. To find the output and hidden layer weights and biases, we use an L-BFGS unconstrained optimization algorithm.

### 3.2. Training and Tuning of the Algorithms

The size of the underlying data set is limited because of the quiet solar cycle 24 and the mission duration of the SDO/HMI. This situation leads to difficulties in reaching a sufficient number of events to separate the data into training and testing subsets. To overcome this difficulty, we employ the stratified $k$-fold cross-validation (CV) method, following Bobra & Ilonidis (2016), where we change $k$-fold values from 3 to 15. This method divides the data set into $k$ subsets, where one subset is used for test and the remaining are used for training. This process is then repeated $k$ times, such that each subset is used for testing exactly once, before the results are averaged.

Limitations of this method do exist, however, as it assumes that the data set follows the same distribution over time. In our case, the number of events varies in-phase with the 11 yr sunspot cycle, which is expected considering the formation of the ARs on the solar photosphere by emergence of deep-seated toroidal magnetic flux ropes, which are generated by the shearing of the poloidal magnetic fields by solar differential rotation, through the photosphere to the corona (Charbonneau 2010; van Driel-Gesztelyi & Green 2015). However, significant flares can occur at all phases of the sunspot cycle, as three of the last four solar cycles showed X-class flares around their minima (Hathaway 2015). The date of the event is not
The Astrophysical Journal, 861:128 (10pp), 2018 July 10

Incceoglu et al.

Table 3
A Confusion Matrix

|      | $O_1$ | ... | $O_l$ |
|------|-------|-----|-------|
| $P_1$ | $n_{11}$ | ... | $n_{1l}$ |
| $P_2$ | $n_{21}$ | ... | $n_{2l}$ |

Note. The term $n_{ij}$ denotes the number of instances predicted in class $i$ by the classifier ($P_i$), where it is observed in class $j$ ($O_j$), and $1 \leq i, j \leq l$.

included in the classifier; thus, this bias should be accounted for. This is done by shuffling the data before the $k$-fold CV is carried out.

In addition to employing the $k$-fold CV, we tune the hyper-parameters of the SVMs and MLPs in each $k$-fold. To achieve this, we use the grid search algorithm provided by the scikit-learn software package (Pedregosa et al. 2011). This algorithm performs an exhaustive search over a predefined set of hyper-parameter values, and the best combination of them is then retained. For example, two hyper-parameters with lists of predefined values would result in a two-dimensional grid, where each element is tested, and the pair of values that gives the best result is saved. While training the SVM, the regularization parameter, $C$, varies between $10^{-3}$ and $10^{3}$, while the RBF coefficient, $\gamma$, varies from $10^{-6}$ to $10^{4}$ in 11 equidistant values in log-space. For the MLP, we have an input, a hidden, and an output layer, where the size of the hidden layer is tested with [18, 36, 54, 72, 90, 108] neurons. Additionally, we vary both the regularization and tolerance parameters between $10^{-3}$ and $10^{-3}$ in five equidistant values in log-space.

3.3. Comparison of the Algorithms

Most of the metrics that are used to characterize and quantify the predictive power of classifiers are calculated directly from the confusion matrices, which are built based on the results from the raw classifier outputs. A confusion matrix for $l$ number of classes shows how $n$ number of instances are distributed over predicted $P_i$ and observed $O_i$ classes, where $1 \leq i \leq l$.

The diagonal terms in a confusion matrix ($i = j$) show correctly predicted instances, while the off-diagonal terms ($i \neq j$) represent incorrectly predicted classes (Table 3).

Let us now consider only one class $i$. The confusion matrix obtained from a classifier gives four types of instances, which are true positives (TP), false positives (FP), false negatives (FN), and true negatives (TN). TP and FP refer to the events that are predicted and observed, and those that are predicted but not observed, respectively. FN and TN, on the other hand, represent the events that are not predicted but observed, and those that are not predicted and not observed, respectively. For the class $i$, these metrics are calculated as $TP = n_{i,i}$, $FP = n_{i,+} - n_{i,i}$, $FN = n_{+,i} - n_{i,i}$, and $TN = n - TP - FP - FN$, where $n_{i,+}$ and $n_{+,i}$ denote the sums of the confusion matrix elements over row $i$ and column $i$, respectively (Labatut & Cherifi 2011).

To evaluate the predictive power of the algorithms, (i) SVMs for different $k$-folds, (ii) MLPs for different $k$-folds, and (iii) SVMs versus MLPs for the same $k$-fold, we use the True Skill Statistics (TSS; Hanssen & Kuipers 1965) and the Heidke Skill Score (HSS; Heidke 1926).

The TSS compares the probability of detection (POD), to the probability of false detection (POFD) and is calculated based on the confusion matrices obtained from ML algorithms. The TSS is defined as,

$$TSS = POD - POFD, \quad \text{(6)}$$

The TSS varies between $-1$ and $+1$, and the value of 0 means that the algorithm is incompetent. High positive values indicate that the algorithm performs well, while negative values show a contradictory behavior, suggesting that the positive and negative classes are mixed around and therefore give a reversed score (Florios et al. 2018).

The HSS, on the other hand, is a method of measuring the fractional improvement of the forecast over the random
The TSS and HSS values are calculated based on the confusion matrices obtained from the SVMs for different \( k \)-fold CV. A sample of the confusion matrix for the 36 hr delay and \( k = 3 \)-fold CV shows that 74 instances of the flares, 31 of the flares with CMEs and SEPs, and again 33 of the CMEs, are correctly classified (Table 4). The results from TSS values suggest that we can predict if a flare will not be associated with any other event 96 hr prior to its occurrence with maximum TSS and HSS values of 0.91 ± 0.07 (Table 7). The TSS and HSS levels are above 0.85 on average for the time delays between 48 hr and 72 hr and increases to around 0.88 ± 0.08 ± 0.08 level for the 108 hr and 120 hr delays. The maximum TSS values for the 24 hr and 12 hr delays are 0.85 ± 0.10 and 0.73 ± 0.10, respectively (Figure 2 and Table 5). A similar variation pattern can also be observed in the HSS values (Figure 3 and Table 5).

The maximum TSS and HSS values of 0.92 ± 0.09 and 0.92 ± 0.08, respectively, show that the SVMs can successfully predict if a flare will be accompanied with CMEs and SEPs 96 hr prior to their occurrences. For the 36 hr forecasting window, the TSS and HSS values are 0.91 ± 0.05 and 0.90 ± 0.08, respectively (Table 5 and Figures 2 and 3). The maximum TSS and HSS values for this class are around ~0.72 on average at time delays between 48 and 72 hr, where they increase up to 0.83 for the 84 hr time delay. For time delays longer than 96 hr, the maximum TSS and HSS values are around 0.85. For the 24 hr forecast window, the TSS and HSS values are 0.85 ± 0.14 and 0.84 ± 0.12, respectively. The performance of the SVMs on predicting if a flare will be accompanied with CMEs and SEPs is at a minimum for the time delay of 12 hr (Table 5 and Figures 2 and 3).

The results also show that the SVMs can predict if a CME will not accompany flares and SEPs for the time delays between 120 hr and 84 hr as well as 36 hr and 24 hr prior to their occurrences as suggested by the maximum TSS and HSS levels of above 0.90 ± 0.10 (Table 5). For the 12 hr time delay, the maximum TSS and HSS values are 0.82 ± 0.11 and 0.83 ± 0.10, respectively. The TSS and HSS values decrease down to around ~0.80 for the time delays between 48 hr and 72 hr (Table 5, and Figures 2 and 3).

### 4.1. ML Algorithms

#### 4.1.1. Support Vector Machines

The TSS and HSS values are calculated based on the confusion matrices obtained from the SVMs for different \( k \)-fold CV. A sample for the confusion matrix for the 36 hr delay and \( k = 3 \)-fold CV show that 74 instances of the flares, 31 of the flares with CMEs and SEPs, and again 33 of the CMEs, are correctly classified (Table 6). The TSS and HSS values for different \( k \)-fold CV suggest that the MLPs can predict that a flare will not be associated with any other event 96 hr prior to its occurrence with maximum TSS and HSS values of 0.91 ± 0.07 (Table 7). The TSS and HSS levels are above 0.70 for the 108 hr and 120 hr time delays, while they vary between 0.60 and 0.85 for the time delays between 24 hr and 84 hr. The TSS and HSS decrease below 0.70 for the 12 hr time delay (Figures 4(a) and 5(a)).
The results show that the MLPs can predict whether a flare will be accompanied by CMEs and SEPs with maximum TSS and HSS values of 0.93 ± 0.04 for the 96 hr time delay. Further, these values are above 0.80 levels for the 108 hr and 120 hr time delays, while they are generally below 0.85 levels for the rest (Table 7 and Figures 4 and 5).

The maximum TSS and HSS values above 0.90 ± 0.10 show that the MLPs can successfully predict that a CME will not be associated with flares and SEPs 120, 108, 36, and 24 hr prior to their occurrences (Figures 4 and 5). For the 108 hr delay, we obtained maximum TSS and HSS values of 0.99 ± 0.02 and 0.99 ± 0.01, respectively. These values gradually decrease down below 0.80 until the 48 hr delay. For the 12 hr time delay, the maximum TSS and HSS are 0.78 ± 0.07 and 0.79 ± 0.11, respectively (Table 7).

4.2. Comparison of the SVMs and MLPs

The resulting TSS and HSS values show that the SVMs generally perform better than MLPs, having the TSS and HSS values mostly above 0.50, while the MLPs reach down to around 0.40.

The TSS and HSS values, which are calculated to predict that flares will not be associated with CMEs and SEPs, are generally higher in SVM for all k-fold values, while the MLP values reach down to ~0.55. The two methods show overlapping high TSS and HSS values for the 96 hr delay (Figures 2–5).

The TSS and HSS values show that SVMs can predict that a flare will be accompanied with CMEs and SEPs in the forecast window of 96 hr (TSS = 0.92 ± 0.08, HSS = 0.92 ± 0.09), which also coincides with the highest TSS and HSS values from the MLPs (TSS = 0.93 ± 0.04, HSS = 0.93 ± 0.04). In addition to the 96 hr forecast window, the SVMs also show high TSS and HSS values for the 36 hr time delay (TSS = 0.91 ± 0.05, HSS = 0.90 ± 0.08), while the MLPs cannot reach these levels (Tables 5, and 7).

The SVMs can predict that a CME will not be associated with any other event with TSS and HSS values above 0.90 ± 0.10 for the time delays continuously between 120 hr and 84 hr, as well as 36 hr and 24 hr (Table 5). On the other hand, the MLPs give TSS and HSS values above 0.90 ± 0.10 for the forecast windows of 120, 108, 36, and 24 hr (Table 7).

5. Discussion and Conclusions

Florios et al. (2018) used data from flaring and non-flaring ARs in an SVM algorithm to predict occurrences of >M1-class and >C1-class flares within a 24 hr forecasting window. For prediction of >M1-class flares, the authors reported TSS and HSS values of ~0.72 and ~0.55, respectively, while for >C1-class flares, their TSS and HSS values decreased down to ~0.57 and ~0.50, respectively. To predict occurrences of flares >M1 class, Bobra & Couvidat (2015) used SDO/HMI’s definitive flaring and non-flaring AR data in SVMs at time delays 48 and 24 hr, and reported TSS values of 0.82 and 0.76, respectively. In addition, based on the SDO/HMI’s definitive AR data that produce flares and flares with associated CMEs in SVMs, Bobra & Ilonidis (2016) calculated TSS = ~0.80 ± 0.20 to predict whether a flare will be associated with CMEs in a 24 hr forecast window. On the other hand, Boucheron et al. (2015) used the SOHO/MDI data of flaring and non-flaring ARs in an SVM regressor and calculated TSS = 0.55 and HSS = 0.46 in their size regressions for >C1-class flares, which is used to predict the size of a flare. To predict flare classes A, B, C, M, and X, Yuan et al. (2010) used SVMs and achieved maximum TSS and HSS values of 0.63 and 0.64 for predictions of A and B class flares, 0.09 and 0.11 for only C class flares, 0.05 and 0.06 for only M-class flares, and 0.14 and 0.18 for only X-class flares, respectively. We recovered the TSS and HSS values based on the confusion matrices given in their work (Figures 3(b), 4(b), 5(b), and 6(b) in Yuan et al. 2010).

Using SVMs for the 24 hr forecast window, we reached maximum TSS and HSS values of 0.85 ± 0.10 to predict whether a flare will be unassociated with any other event. The maximum TSS and HSS values to predict whether a flare will be accompanied by CMEs and SEPs are 0.85 ± 0.14 and 0.84 ± 0.12, respectively. As for predicting whether an AR will produce only CMEs, our maximum TSS and HSS values are 0.91 ± 0.10 (Table 5). Additionally, our maximum TSS and HSS values from the SVMs for the 48 hr prediction window that a flare will not be associated with any other event.
is $0.77 \pm 0.08$. For predicting whether a flare will be associated with CMEs and SEPs, our TSS = 0.80 ± 0.10 and HSS = 0.80 ± 0.09, while for forecasting that a CME will not be associated with any other event, the maximum TSS and HSS values are 0.79 ± 0.19 and 0.80 ± 0.09, respectively. Among the time delays we used in our study, which range from 120 hr to 12 hr, the forecast windows of 96 hr and 36 hr consistently provided maximum TSS and HSS values above 0.90 with standard deviations smaller than 10% (Table 5). These results are superior to those found by previous studies. However, it must be noted that the scope of this study is to distinguish the three classes of solar eruptions observed on the Sun; therefore, non-flaring ARs are not included. Also, the choice of the parameters used in this study differs from most of the previous studies.

Using MLPs to predict the occurrences of >M1 and >C1-class flares within a 24 hr forecast window, separately, Florios et al. (2018) obtained maximum TSS and HSS values of ~0.73 and ~0.55 for >M1-class flares, and ~0.57 and ~0.55 for >C1-class flares, respectively (Figures 3 and 5 in Florios et al. 2018). Using a cascade correlation NN algorithm to predict the occurrences of >C1-class flares within 24 hr and 48 hr forecast windows, Ahmed et al. (2013) calculated the maximum TSS and HSS as 0.45 and 0.54, respectively. Yu et al. (2009), on the other hand, used a learning vector quantization NNs algorithm to predict flare occurrences within a forecast window of 48 hr and reached maximum TSS = 0.67, which we recovered using the TSS = TNrate/($N_\text{total}$ - TNrate) relationship.

Our maximum TSS and HSS values obtained from the MLPs for the 24 hr forecast window are $0.79 \pm 0.11$ for forecasting...
that a flare will not be associated with any other event. To predict whether a flare will be associated with CMEs and SEPs, we obtained TSS = 0.74 ± 0.13 and HSS = 0.77 ± 0.05, while these values are 0.93 ± 0.07 and 0.91 ± 0.08 for forecasting that a CME will not be associated with any other event (Table 7). As for the forecast window of 48 hr, we obtained TSS = 0.80 ± 0.09 and HSS = 0.81 ± 0.09 for predictions of flare only events. The MLPs gave maximum TSS and HSS values of 0.71 ± 0.07 and 0.72 ± 0.05 for predicting whether a flare will be accompanied by CMEs and SEPs, while these numbers are 0.79 ± 0.12 and 0.80 ± 0.07 for predictions of CME-only events. Although our maximum TSS and HSS values are higher than those found in previous studies, we must again note that we do not include data from non-flaring ARs, as our main focus here is to distinguish the three classes of solar eruptions observed on the Sun.

In our study, we used SVMs and MLPs to predict whether an AR, which is known to generate solar eruptions, will produce only flares, flares with associated CMEs and SEPs, or only CMEs for time delays extending from 12 hr to 120 hr. To achieve this objective, we used data provided by the SHARPs, GOES, and DONKI databases. The size of the data used in this study is limited due to the mission duration of SDO/HMI and also because of the quietness of the current solar cycle 24, which leads to fewer flares, CMEs, and SEPs. To overcome this difficulty, we employed the stratified k-fold CV method, which ensures that each subset of data is used to train and test the ML algorithms. However, this method assumes that the underlying data follows the same distribution throughout the study period. It is shown that the number of occurrences of flares, CMEs, and SEPs, although sporadic, follow a trend in-phase with the 11 yr solar cycle (Aschwanden & McTiernan 2010; Chen 2011; Hathaway 2015). To decrease this bias in our calculations, we shuffled the data before performing the k-fold CV method. However, to remove this bias substantially, to avoid overfitting, and to train and validate the ML algorithms further, longer data sets are needed.

To optimize the hyper-parameters used in our SVM and MLP algorithms, we employed an embedded grid search for each iteration of k-fold CV values ranging from 3 to 15. Our results show that we can achieve TSS and HSS values greater than 0.90 with standard deviations smaller than 10%, which shows that both our SVM and MLP are good classifiers, though the former is slightly better than the latter. Our results also show that we can predict that flares will not be accompanied with any associated event at the earliest 96 hr before they happen, with maximum TSS and HSS values of 0.90 ± 0.08 for the SVMs and 0.91 ± 0.07 for the MLPs. The earliest forecast that a flare will be associated with CMEs and SEPs can be made at 96 hr time delay with TSS = 0.92 ± 0.09 and HSS = 0.92 ± 0.08 for the SVMs and TSS = HSS = 0.93 ± 0.04 for the MLPs. We also showed that we can predict if an AR will produce only CMEs unassociated with any other events 108 hr before they occur with maximum TSS and HSS values of 0.98 ± 0.02 for the SVMs and 0.99 ± 0.02 for the MLPs. Our results indicate that the discriminative potential of the physical features of ARs in SHARPs data is very high.

We also calculated the Fisher scores of the 18 physical parameters for each time delay, which are not shown in this study, as we do not include them as feature selection criteria. However, we need to note that the results show that the variation in the Fisher scores of the physical parameters of the ARs provide insights into why the predictive powers of the SVMs and MLPs change with different time delays. Additionally, the calculated Fisher scores indicate that the physical features are highly complementary across the different time delays, meaning that all features are relevant but not necessarily at the same time delay.

In conclusion, our results show that the SVMs are slightly better than the MLPs. However, more extensive future work is necessary for SVM classifiers, where ARs that do not produce any eruptive phenomena are planned to be included in the analyses. This will, however, introduce imbalance problems because the number of ARs that do not produce an eruptive event is overwhelmingly high compared to those that do produce eruptions (Bobra & Couvidat 2015). It is therefore important to find the optimum values for the class weight ratios to sort out this problem. Additionally, we plan to investigate the optimum number of features by combining the predictions from the individual time delays, to search a larger grid for the C and γ values, and to employ different kernel functions and grid search their related parameters (such as polynomial functions with different degrees). Furthermore, another direction we plan to exploit is using deep NNs with temporal evolutionary algorithms on the time-series aspect of the available SHARPs data.

Funding for the Stellar Astrophysics Centre is provided by the Danish National Research Foundation (grant agreement No. DNRF106). The project has been supported by the Villum Foundation.

ORCID iDs
Fadil Inceoglu @ https://orcid.org/0000-0003-4726-3994

References
Ahmed, O. W., Qahwaji, R., Colak, T., et al. 2013, SoPh, 283, 157
Aschwanden, M. J., & McTiernan, J. M. 2010, ApJ, 717, 683
Bobra, M. G., & Couvidat, S. 2015, ApJ, 798, 135
Bobra, M. G., & Ionidis, S. 2016, ApJ, 821, 127
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
Carrington, R. C. 1859, MNRAS, 20, 13
Charbonneau, P. 2010, LRSP, 7, 3
Chen, P. F. 2011, LRSP, 8, 1
Cliver, E. W., & Dietrich, W. F. 2013, JSWSC, 3, A31
Cortes, C., & Vapnik, V. 1995, Mach. Learn., 20, 273
Florios, K., Kontogiannis, I., Park, S.-H., et al., 2018, SoPh, 293, 28
Hansen, A., & Kuipers, W. 1965, On the Relationship Between the Frequency of Rain and Various Meteorological Parameters (The Hague: Staatsdrukkerij-en Uitgeverijbedrijf)
Hathaway, D. H. 2015, LRSP, 12, 4
Heidke, P. 1926, Geogr. Ann., 8, 301, http://www.jstor.org/stable/519729
Hinton, G. E. 1989, Artif. Intell., 40, 185
Hornik, K., Stinchcombe, M., & White, H. 1989, NN, 2, 359
Hsu, C., & Lin, C. 2002, J. Mach. Learn. Res., 3, 415425
Kilpu, E., Koskinen, H. E. J., & Pulkkinen, T. I. 2017, LRSP, 14, 5
Labatut, V., & Cherifi, H. 2011, in The 5th Int. Conf. on Information Technology, ed. A-D. Ali (Amman: Al-Zaytoonah Univ. Jordan), 1
Leka, K. D., & Barnes, G. 2003, ApJ, 595, 1277
Lopez, R. E., Baker, D. N., & Allen, J. 2004, EOS, 85, 105
MacKay, D. J. C. 2005, Information Theory, Inference and Learning Algorithms (Cambridge: Cambridge Univ. Press)
Moore, R. L., Falcomer, D. A., & Sterling, A. C. 2012, ApJ, 750, 24
SunPy Community, Mumford, S. J., Christie, S., et al. 2015, CS&D, 8, 014009
Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2011, J. Mach. Learn. Res., 12, 2825, http://www.jmlr.org/papers/volume12/pedregosa11a/pedregosa11a.pdf

Inceoglu et al.
Reames, D. V. 2013, SSRv, 175, 53
Reames, D. V., Ng, C. K., & Tylka, A. J. 2013, SoPh, 285, 233
Schou, J., Scherrer, P. H., Bush, R. I., et al. 2012, SoPh, 275, 229
Schrijver, C. J. 2007, ApJL, 655, L117
Schrijver, C. J., & Siscoe, G. L. 2010, Heliophysics: Space Storms and Radiation: Causes and Effects (Cambridge: Cambridge Univ. Press)

Shibata, K., & Magara, T. 2011, LRSP, 8, 6
van Driel-Gesztelyi, L., & Green, L. M. 2015, LRSP, 12, 1
Vapnik, V. 1998, Statistical Learning Theory (New York: Wiley)
Webb, D. F., & Howard, T. A. 2012, LRSP, 9, 3
Yu, D., Huang, X., Wang, H., & Cui, Y. 2009, SoPh, 255, 91
Yuan, Y., Shih, F. Y., Jing, J., & Wang, H.-M. 2010, RAA, 10, 785