

# Leveraging the mathematics of shape for solar magnetic eruption prediction

Varad Deshmukh<sup>1,\*</sup>, Thomas E. Berger<sup>2</sup>, Elizabeth Bradley<sup>1,3</sup>, and James D. Meiss<sup>4</sup>

<sup>1</sup> Department of Computer Science, University of Colorado 430 UCB, Boulder, 80309-0430 CO, USA

<sup>2</sup> Space Weather Technology, Research, and Education Center (SWx TREC), University of Colorado 429 UCB, Boulder, 80309-0429 CO, USA

<sup>3</sup> Santa Fe Institute, Santa Fe, 87501 NM, USA

<sup>4</sup> Department of Applied Mathematics, University of Colorado 526 UCB, Boulder, 80309-0526 CO, USA

Received 30 September 2019 / Accepted 9 March 2020

**Abstract**—Current operational forecasts of solar eruptions are made by human experts using a combination of qualitative shape-based classification systems and historical data about flaring frequencies. In the past decade, there has been a great deal of interest in crafting machine-learning (ML) flare-prediction methods to extract underlying patterns from a training set – e.g. a set of solar magnetogram images, each characterized by features derived from the magnetic field and labeled as to whether it was an eruption precursor. These patterns, captured by various methods (neural nets, support vector machines, etc.), can then be used to classify new images. A major challenge with any ML method is the *featurization* of the data: pre-processing the raw images to extract higher-level properties, such as characteristics of the magnetic field, that can streamline the training and use of these methods. It is key to choose features that are informative, from the standpoint of the task at hand. To date, the majority of ML-based solar eruption methods have used physics-based magnetic and electric field features such as the total unsigned magnetic flux, the gradients of the fields, the vertical current density, etc. In this paper, we extend the relevant feature set to include characteristics of the magnetic field that are based purely on the geometry and topology of 2D magnetogram images and show that this improves the prediction accuracy of a neural-net based flare-prediction method.

**Keywords:** Solar eruption prediction / machine learning / computational geometry / computational topology

## 1 Introduction

Sunspot active regions manifest as large-scale, high-magnitude, dipolar structures in images of the magnetic field at the surface, or “photosphere,” of the sun. They are the source regions for the largest solar magnetic eruptions, which produce flares, Coronal Mass Ejections (CMEs), and energetic particle events that can drive important space weather events. Photospheric magnetograms are 2D samples of the structure of the full 3D solar magnetic field and thus can provide important clues about the increasing complexity of the magnetic field in the lead-up to a magnetic eruption – information that can potentially be leveraged for the purposes of prediction. While there has been a great deal of recent work on machine-learning based algorithms for predicting solar eruptions from magnetogram data, the *features* used by these algorithms have been predominantly physics-based: taking the curl of the magnetic field to get

currents, computing its gradient, summing up its absolute values, etc.

We propose a new approach to this task: *computing the topology and geometry of the structures in 2D magnetograms*. Instead of deriving physical quantities from these data sets on a per-pixel basis, or attempting to model the full 3D coronal magnetic field or field line connections from the 2D information that is captured in magnetograms, we formally quantify their structure using the fundamental mathematics of shape. We argue that – even though this procedure ignores the 3D structure or connectivity of the full field – it enhances the predictive skill of current methods. Indeed, it is the current operational practice for human forecasters to predict solar eruptions by examining sunspot images and/or magnetograms and using the McIntosh (1990) or Hale et al. (1919) classification systems to categorize active region complexity using alphabetical designations. Each category of active region has a statistical “24-h eruption probability” derived from historical records that is then reported (following forecaster adjustments for factors such as rate of flux

\*Corresponding author: [varad.deshmukh@colorado.edu](mailto:varad.deshmukh@colorado.edu)

emergence) as the eruption forecast for a particular active region (Crown, 2012). Recent statistical methods extend this approach by using historical flaring rates, together with a Poisson process hypothesis, to develop more complicated models. For example, Gallagher et al. (2002) use the McIntosh classification to determine the probabilities of C-, M-, or X-class flares and Wheatland (2004) use a power law distribution of the flare magnitudes to determine an empirical eruption probability. However, these methods have not seen much use in operational forecasting, primarily because they do not show greater predictive capability than the historical forecasts that use look-up tables.

Recognizing that the magnetic reconnection that triggers eruptions takes place in the upper atmosphere (corona) of the sun, attempts have been made to model the coronal magnetic field using the measured photospheric field as a boundary condition. The simplest method to extrapolate the surface field into the corona assumes zero current, so that the field potential is a solution to the Laplace equation (Wang & Sheeley Jr., 1992; Barnes et al., 2005; Barnes & Leka, 2006). However, potential fields cannot store energy – they are lowest energy states – and thus cannot model the build-up of energy leading to a CME. Other strategies include nonlinear force-free field (NLFFF) extrapolations (Priest & Démoulin, 1995; Démoulin et al., 1996; Aulanier et al., 2005; Wiegelmann & Sakurai, 2012), which are known to correlate well with the sites of X-ray flare emission from eruptions. NLFFF models are presently the most promising avenue of coronal magnetic field and eruption modeling (Schrijver et al., 2008), but they have a large number of free parameters that require extensive manual tuning. Note that the photospheric boundary conditions are not sufficient to uniquely determine the solution (Schrijver et al., 2006; Metcalf et al., 2008; DeRosa et al., 2009), and hence the utility of these modeled 3D structures for forecasting appears to be limited at present.

An alternate approach models the connectivity between opposite polarity magnetic patches in the photosphere to identify significant structures such as “nulls” and “separatrix surfaces.” For example, Barnes et al. (2005) develop a “Magnetic Charge Topology” (MCT) metric that is used to characterize eruption potential. Longcope (2005) reviews the application of topology to inferred field line connections and (Tarr & Longcope, 2012; Tarr et al., 2013) apply these methods to analyze the eruption potential of active regions. While these methods are sophisticated and the structures that they extract are meaningful in the context of solar eruptions, their computational complexity limits their operational application.<sup>1</sup>

It is important to note that the use of the term “topology” in the methods described in the previous paragraph refers to the study of the shape and/or connectivity of 3D field lines above the photosphere (e.g., Longcope, 2005). This contrasts with the more general, mathematical definition of the term, which refers to the shape and connectivity of sets of any dimension – of which field lines are one instance. Ideas and techniques from this broader field of topology can be used to address many other important and potentially meaningful properties of solar data. Mathematically, topology distinguishes sets that cannot be transformed into one another by continuous maps with continuous

inverses (homeomorphisms). Though this can obscure much of what is commonly meant by structure, its roughness can also be a virtue in that it will eliminate distinctions that could be due to unimportant distortions, e.g. those due to projection of the sun’s spherical surface onto a 2D magnetogram. Computational topology, also known as topological data analysis (TDA), operationalizes this highly abstract framework for use with real-world data, which may be noisy and poorly sampled. Though computing the abstract topology from this type of data is not feasible, an aspect of shape that can be computed relatively straightforwardly is *homology*. In homology, shapes are distinguished according to their pieces, holes, voids, etc. This calculation can be reduced to linear algebra, essentially the calculation of the dimensions of the ranges and null spaces of certain matrices (Kaczynski et al., 2004). TDA has been used in many applications, including coverage of sensor networks (de Silva and Ghrist, 2007), structures in natural images (Ghrist, 2008), neural spike train data (Singh et al., 2008), and even the large-scale structure of the universe (Xu et al., 2019).

To compute the topology of a collection of points that samples an object requires an interpolation scheme to “fill in” the gaps between the points. The theory of *persistent homology* leverages interpolation to compute a shape as a function of scale. The result is encoded in a plot called a persistence diagram, which can be further processed to yield useful features. For example, such a diagram naturally captures how the structures in a magnetogram change as flux emerges into an active region during the evolution towards a flaring state. By contrast, computational geometry – widely used in computer graphics and computer-aided design – addresses the problem of extracting purely geometric information (line segments, polyhedra, etc.) from a data set. The spatial relationship of the positive and negative polarities in an active region, for example – and particularly their positioning relative to strong “neutral lines” in the photospheric field configuration – are known to be important indicators of flaring (e.g., Schrijver, 2007), and computational geometry can easily capture these properties in formal ways. We will demonstrate in Section 2 that a combination of topological and geometric analyses can extract meaningful information from a series of magnetogram examples.

In the past decade, the large increase in magnetogram data afforded by space missions and advances in data access have shifted the forefront of flare-prediction research from empirical modeling methods to “data analytic” approaches such as machine learning. Camporeale (2019) summarizes the state of the art in machine learning approaches to space weather applications. In ML-based prediction applications, characteristic “features” of the photospheric magnetic field, sometimes combined with features seen in simultaneous extreme ultraviolet (EUV) images of the solar corona, are used in a statistical sense to “train” a computational model to predict the probability of an eruption within a given time period (usually 24 h). One example is a support vector machine (SVM) architecture to perform a binary classification of magnetograms as flaring or non-flaring (Yuan et al., 2010; Yang et al., 2013; Bobra & Couvidat, 2015; Boucheron et al., 2015; Nishizuka et al., 2017). Nishizuka et al. (2017) have also applied decision trees and clustering to the same task. Neural networks, which go beyond simple binary classification by learning complex nonlinear relationships among their inputs, have also been used to great advantage by Nishizuka et al. (2018). A variety of other ML algorithms,

<sup>1</sup> Note that Barnes and Leka have gone on to apply their MCT metric to a potentially operational flare prediction algorithm called DAFFS.

such as Bayesian networks (Yu et al., 2010), radial basis model networks (Colak & Qahwaji, 2009), logistic regression (Yuan et al., 2010), LASSO regression (Campi et al., 2019), and random forests or ERTs (Nishizuka et al., 2017; Campi et al., 2019) have also been implemented for solar flare prediction with some degree of success. Nishizuka et al. (2017) and Jonas et al. (2018) include Solar Dynamics Observatory (SDO) Atmospheric Imaging Assembly and EUV image characteristics in addition to magnetogram data in a fully connected neural network architecture. Benvenuto et al. (2018) use Fuzzy C-Means – an unsupervised machine learning method – in combination with some of the mentioned supervised methods for solar flare prediction. Approaches such as Guerra et al. (2018) and Kontogiannis et al. (2018), while not machine-learning approaches themselves, provide statistical tools for evaluating the engineered features in terms of their potential advantage in machine learning models. Finally, while most of the above methods focus on developing ML models using engineered features, recent methods have employed a convolutional neural network (CNN) approach which automatically extracts features from raw magnetogram data that are important to predicting flares (Huang et al., 2018; Park et al., 2018; Zheng et al., 2019).

The sophisticated methods described in the previous paragraph represent the state-of-the-art in ML-based solar eruption prediction. However, they predominately use physics-based features, and a careful, quantitative comparison study shows that none of them are significantly more skilled, and indeed are typically less skilled, than current human-in-the-loop methods employed at operational forecasting offices (Leka et al., 2019a, 2019b; Barnes et al., 2016). Our goal is to improve upon these results using ideas and algorithms from computational topology and computational geometry to quantify the complexity in magnetograms and/or sunspot images. In this study, we analyze magnetogram images using different flux thresholds – sub- or super-level sets, in mathematical parlance – and extract structural signatures that, we conjecture, can be effectively leveraged as elements in a feature vector for machine-learning methods. As evidence in favor of that conjecture, we use high-fidelity vector magnetic field data from the Helioseismic and Magnetic Imager (HMI) instrument (Scherrer et al., 2011) on the NASA Solar Dynamics Observatory satellite (Pesnell et al., 2011) and show that adding shape-based signatures to existing feature vectors improves the 24-h prediction accuracy of a neural-net based method.

This is, to our knowledge, the first time that systematic quantitative measures of the shape of 2D magnetic structures in the photosphere – based on their topology and their geometry – have been developed for the purposes of flare prediction, and tested carefully in the context of a machine-learning method. In a sense, our approach is a mathematical systemization of the current *ad hoc* McIntosh or Hale classification systems. Though it employs topology, it is very different from the work described above that analyzes the magnetic field-line structure. We focus on the shapes of *two-dimensional* sets, restricting our analysis to the photospheric magnetic field structures. Our goal is to extract formal shape characterizations that can be leveraged by ML methods to improve flare prediction. We are not attempting to model the coronal magnetic field or determine field line connections between neighboring opposite polarity structures. Our approach differs from existing work on geometric (Schrijver, 2007; McAteer et al., 2010)

and topological (Knyazeva et al., 2011; Makarenko et al., 2014) analysis of magnetogram features in that the goal is to derive features that improve ML-based methods for flare prediction rather than to discover any one physical property that is more or less predictive of flaring. We believe that this approach has merit since current operational flare prediction methods – the McIntosh and/or Hale classification systems used by human experts – are fundamentally based on active-region shape and geometry. In addition, Tarr & Longcope (2012) state that “topological changes” can be shown to precede flaring activity in a typical sunspot active region, suggesting that active region shape, and its evolution, have a fundamental, meaningful connection to the physics of magnetic eruptions.

The following section goes into more depth on how to formulate and deploy computational topology and geometry in the context of HMI magnetograms. Section 3 presents a study of how features extracted from magnetograms using those techniques can improve the prediction accuracy of a specific machine-learning method for flare prediction. We conclude in Section 4.

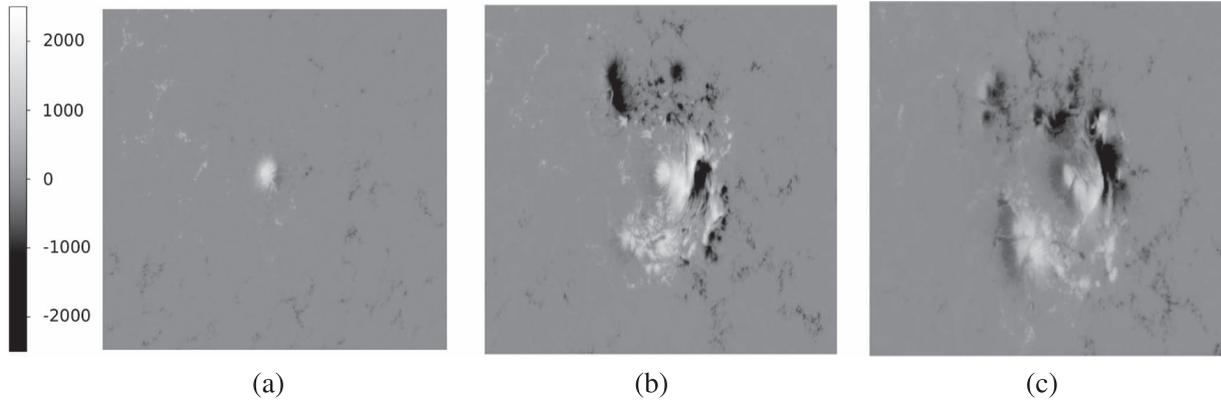
## 2 Capturing the shapes of active regions

Figure 1 shows a series of line-of-sight magnetograms of an active region before and during an eruptive period. In panel (a), the active region is newly emerged and is concentrated into a relatively compact and simple configuration.<sup>2</sup> Such initial emergence configurations store little free energy and are rarely associated with eruptions. However, as more magnetic flux emerges and this active region evolves under the influence of the plasma flows in the photosphere, it is stretched, rotated, and sheared into the complex shape shown in panel (b). While in this complex configuration, the active region produced a strong flare that had major impacts on Earth-based radio reception. The further development shown in panel (c), later in this sunspot’s series of flaring events, is characterized by intense “polarity mixing,” with positive and negative field in close proximity in highly sheared and stretched shapes.

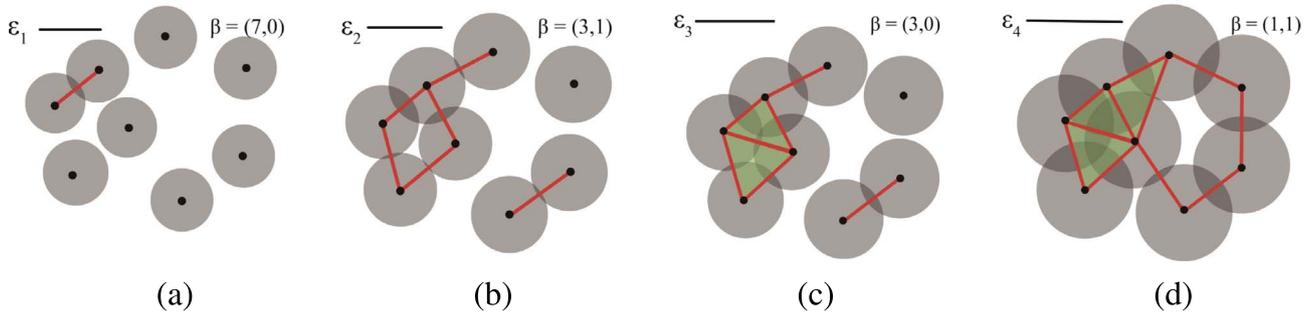
Topological data analysis (TDA) (Kaczynski et al., 2004; Ghrist, 2008; Zomorodian, 2012) is exactly the right foundation for extracting and codifying the spatial richness of these images. Topology is the fundamental mathematics of shape: two sets are topologically equivalent if they are homeomorphic<sup>3</sup>; thus, as is often said, topology does not distinguish between a doughnut and a coffee cup. Unfortunately it is not practical to compute the topology of a set even if one has a complete description of it, let alone when one only has a finite number of samples. Topological data analysis was developed to address these challenges. It formally quantifies the shape of a data set according to the so-called Betti numbers: the number of components ( $\beta_0$ ), holes ( $\beta_1$ ), voids ( $\beta_2$ ), etc., in the data. Note that every topological space has a unique set of Betti numbers, but these do not completely classify the topology; for example, they would not capture the twist of a Möbius strip.

<sup>2</sup> We note that since this longitudinal flux density measurement is taken close to the eastern limb of the Sun, the polarities do not appear to fully balance. This is a well-known effect of strong, often non-radial, fields in emerging active regions observed near the limb.

<sup>3</sup> That is, they can be mapped to one another by a bi-continuous bijection.



**Fig. 1.** A series of HMI magnetograms of sunspot #AR 12673, which produced multiple large eruptions as it crossed the disk of the Sun in September 2017: (a) at 0000 UT on 9/1, (b) at 0900 UT on 9/5, roughly 24 hours before this sunspot produced an X-class solar flare, and (c) at 1000 UT on 9/7, around the time of an M-class flare.



**Fig. 2.** Computational topology: sketch of a data set (the black points) enlarged into balls of diameter  $\varepsilon$ , for four values of  $\varepsilon$ . The resulting connections, shown as lines (red) and triangles (green) form what is called a Rips complex.

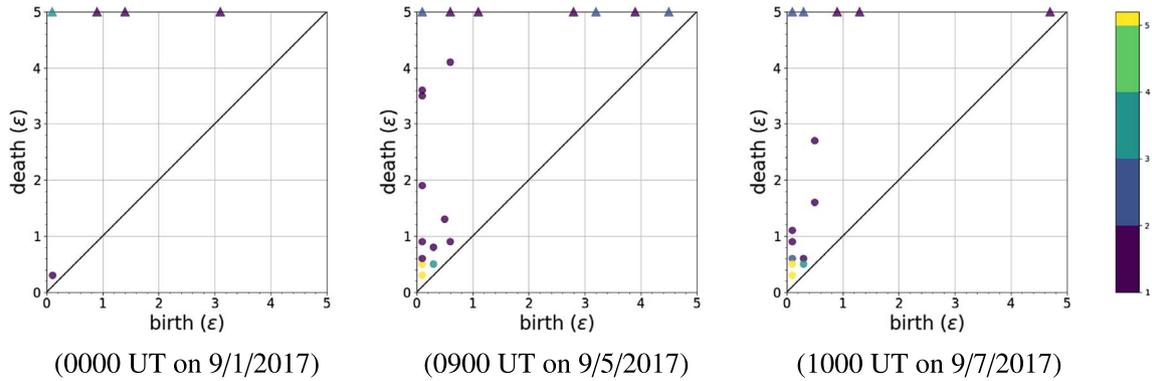
Of course, a finite collection of points does not really have a “shape.” TDA handles this by defining a scale parameter, for example, creating a manifold from a set of disconnected points by enlarging each point into a ball, as sketched in Figure 2. This leads to the notion of  $\varepsilon$ -connectedness: a pair of points is treated as connected if they are within a distance  $\varepsilon$  of each other. A set of points connected by a graph with edges of length no more than  $\varepsilon$  is called an  $\varepsilon$  component. Of course, when  $\varepsilon = 0$ , each point (black dot in the figure) is a component. Conversely, the entire manifold is viewed as connected, from the standpoint of TDA, if the balls have a sufficiently large radius, as for the value  $\varepsilon_4$  in the figure. The procedure of varying scale is familiar from the calculation of fractal dimension: one views dimension as reflecting the scale-dependent growth of the number of data points in a ball. However, TDA does not attempt to compute shape in the limit  $\varepsilon \rightarrow 0$ , as one would for fractal dimension; rather it views the change in shape at finite values  $\varepsilon$  as reflecting macroscopic properties of the data (Robins et al., 1998, 2000).

More generally, the connections give rise to a simplicial complex, called a Rips<sup>4</sup> complex; this is essentially a list of these connections and groups of connections. Connections between pairs of points give edges (red lines in Fig. 2) that correspond to 1-simplices. When there is a cycle with three vertices – i.e. three  $\varepsilon$ -balls pairwise overlap – the associated triangle (green) is filled in; this is a 2-simplex, etc. The shape

of the complex, and the Betti numbers that describe it, vary with the scale parameter: when  $\varepsilon = 0$ ,  $\beta_0$  is equal to the number of points in the data set and the Rips complex also contains no higher-dimensional simplices so  $\beta_1 = 0$ . As  $\varepsilon$  grows, nearby points are successively joined, causing components to grow and holes to form. For example for  $\varepsilon = \varepsilon_2$  in the figure there are three components and one hole, so  $\beta = (\beta_0, \beta_1) = (3, 1)$ , but as  $\varepsilon$  reaches  $\varepsilon_3$ , two triangles have formed to fill in the hole and  $\beta_1$  becomes 0. Every hole eventually vanishes as the complex gets filled in, and so for large enough  $\varepsilon$  there is a single component with no holes, i.e.  $\beta = (1, 0)$ .

All of this information about the spatial scales of the topological features in a data set can be captured in a plot called a *persistence diagram* (Edelsbrunner et al., 2000). Most  $\varepsilon$ -components, for example, have birth and death parameter values – where they appear and disappear, respectively, from the construction. A  $\beta_0$ -persistence diagram has a point at  $(\varepsilon_{\text{birth}}, \varepsilon_{\text{death}})$  for each component. Such a diagram is shown in Figure 3 for the series of magnetograms in Figure 1. Components that still exist at the upper end of the calculation interval – here  $\varepsilon = 5$  pixels – are represented by triangles in the figure. When multiple components have the same  $\varepsilon_{\text{birth}}$  and  $\varepsilon_{\text{death}}$  values, the color of the icon in Figure 3 indicates the number of components with that lifespan. One can also plot persistence diagrams for the other Betti numbers; the large hole to the right of center in Figure 2d, for instance, would give a point near  $(\varepsilon_4, \varepsilon_5)$  on a  $\beta_1$ -persistence diagram, where  $\varepsilon_5$  is the ball radius

<sup>4</sup> An alternative, related complex is the Čech complex.

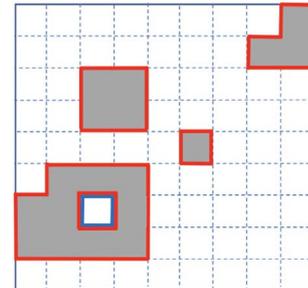


**Fig. 3.**  $\beta_0$  persistence diagrams in a temporal sequence of magnetograms of active region #12673, constructed from the set of points (pixels) with positive magnetic field intensities greater than 200 Gauss. The co-ordinates of each point  $(x, y)$  on a persistence diagram represent the birth and death times of a  $\beta_0$  feature. Note that all the features lie above the  $x = y$  line in the diagrams since a feature cannot die before it is born. In terms of number of connected components at various scales, these diagrams reveal a clear change in the topology of the field structure around the X2.2-class flare that took place at 0910 UT on 6 September 2017.

that causes all of the  $\varepsilon$ -balls around it to overlap, while the short-lived hole in panel (b) would be close to the diagonal near  $\varepsilon_2$ . This rich representation of information about the underlying shape that is sampled by a set of data points can, we conjecture, be effectively leveraged by ML-based flare-prediction methods. In the case of the magnetograms in Figure 1, we have preliminary evidence for this conjecture: the  $\beta_0$ -persistence diagrams in Figure 3 reveal a change in the topology of active region AR #12673 more than 24 h before the X-class flare that took place at 0910 UT on 6 September 2017. To the eye, the change in the overall number of points on these diagrams is quite obvious; more important is the number and location of the points that lie far from the diagonal: i.e. those that persist for wide ranges of the scale parameter  $\varepsilon$ . The rich, multi-scale nature of the information captured in a persistence diagram, and the subjective nature of some of the associated definitions – viz., the notion of “far” from the diagonal – makes it a challenge to develop effective formal metrics for describing that structure. This challenge is a current focus in the TDA research community; see Section 3.2. A full treatment of the associated solutions is beyond the scope of this paper.

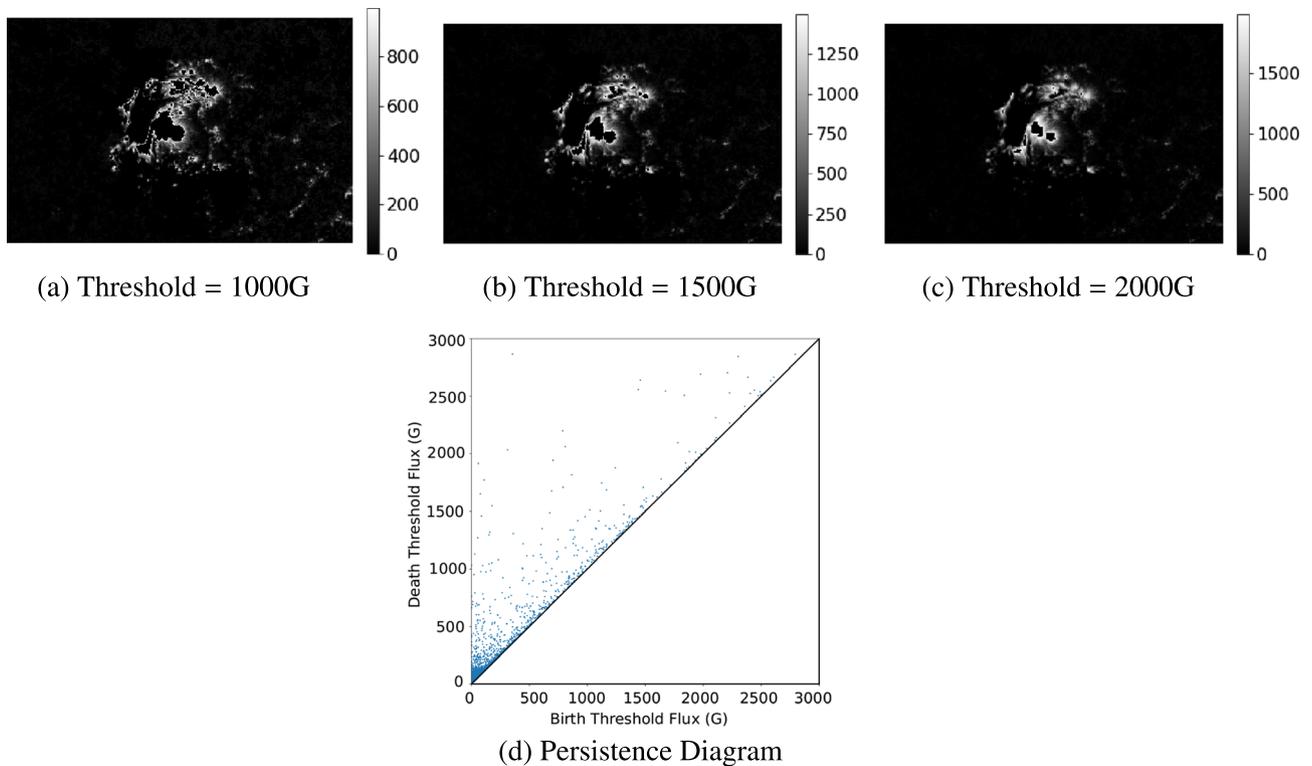
While the pre-flare change in topology is encouraging, there is another issue here: the threshold used in these calculations to choose which pixels in the magnetogram to treat as points in the complex. Thresholding data into categories involves an awkward choice: rarely is there a crisp boundary between “low” and “high.” (Indeed, this is a large part of the drive to use machine learning in data science.) While the radial magnetic field strength in an HMI magnetogram ranges up to about 5000 Gauss, there is no clear notion of what constitutes a good threshold value for defining coronal footpoint boundaries. Moreover, any threshold should be relatively insensitive to the instrumental noise threshold (on the order of 10 G) and the small-scale, background field (on the order of 100 G).

The proof-of-concept persistence diagrams in Figure 3 were constructed from thresholded images containing only pixels with magnetic field intensity greater than 200 G; i.e. a super-level set of the intensity. Instead of some arbitrary choice, it would be preferable to view the threshold itself as a parameter



**Fig. 4.** In a cubical complex, above-threshold pixels are treated as connected iff they are neighbors. This example has four connected components, outlined in red, and one hole, outlined in blue – i.e.  $\beta_0 = 4$  and  $\beta_1 = 1$ .

to vary, thereby obtaining a sequence of super-level sets. An elegant alternative is to simply use the threshold value, rather than the spatial scale, as the persistence parameter. To eliminate the spatial scale parameter  $\varepsilon$  of the Rips complex that is appropriate for arbitrary point clouds like that shown in Figure 2, we note that magnetograms are pixel-based data. The simplest construction of a complex in this case uses a *cubical complex*. Here, one simply treats two pixels as connected if they share an edge or a vertex and each represents a flux level that is below the chosen threshold (i.e., sub-level thresholding). Figure 4 shows a schematic of such a construction; Figure 5a–c shows cubical complexes constructed for three threshold values for AR #12673. Note how the holes in Figure 5a–c form and then fill in as the threshold changes. This gives a different view on persistent homology and a second kind of persistence diagram: one with the threshold value, rather than the connectedness scale, on the  $x$  and  $y$  axes, as shown in Figure 5d. This persistence diagram is far more detailed than the ones in Figure 3 because it captures structures at a *range* of thresholds, and thus is more affected by the complexity of the structure of this active region. The patterns in this plot – the large number of short-lived holes near the diagonal and the long-lived holes



**Fig. 5.** Cubical complexes for AR #12673 at three example threshold values and the full associated  $\beta_1$  persistence diagram for a range of positive magnetic field thresholds.

further above and to the left – are an accurate formal representation of that complexity.<sup>5</sup>

Persistence diagrams are powerful tools, but topology alone is not a complete tool for characterizing the shape of structures in a magnetogram. Simply identifying the number of high-flux regions, for instance, says nothing about their sizes or proximities. A pair of converging high-flux regions looks the same, from the standpoint of topology until they actually touch<sup>6</sup> – and topology cannot measure quantitative features like the total magnetic flux in such a region. In order to capture these important, and potentially predictive, properties so that they can be used by ML algorithms, we extract geometric information as well by computing the sizes of the high-flux regions and the distances between them, summing up the flux inside them, finding their centroids, and the like. Computational geometry algorithms (Forrest, 1971; Preparata & Shamos, 1985) are widely used for these kinds of calculations across many fields of science and engineering, including astronomy – e.g. shape reconstruction for asteroids (Devoegele et al., 2015) and galaxy distribution analysis (Bhavsar & Lauer, 1996).

Our approach to distilling informative, discriminating features for each active region out of solar magnetogram data builds on all of these foundations. Any or all of the topological

and geometric properties described in this section might be meaningful predictors of flaring. Moreover, it is not only the topology and geometry of the positive and negative regions of the field that are indicative, but also their topological and geometric relationships to one another; so we also explore composite features, as described in the following section.

### 3 Results

As demonstration of the utility of shape-based features in machine-learning methods for solar flare prediction, we chose to work with an artificial neural network (ANN) (Haykin, 1998), a machine-learning approach to fitting a repertoire of nonlinear functions to the data. ANNs, also known as multilayer perceptron architectures, are both flexible and powerful; they can generally model more complex nonlinear functions than regression networks like SVMs or decision trees. An ANN is built by stacking together layers of nonlinear elements, known as *neurons*, with weighted interconnections between consecutive layers. The input layer takes the magnetogram features and feeds them into the stack. A layer-by-layer, feed-forward propagation of activations finally results in a prediction at the output layer – in this case, a binary value that classifies the magnetogram as a precursor to a flare (or not), i.e. whether that magnetogram is followed by an eruption above the X-ray class of M-flares within the next 24 h. This classification of magnetograms is equivalent to the M1.0+/0/24 event designation in Leka et al. (2019a).

<sup>5</sup> Some persistence diagram analysis techniques discount or discard points near the diagonal, as they are more sensitive to noise and pixelation of the data.

<sup>6</sup> From the standpoint of *computational* topology, the definition of touch depends on the scale parameter  $\epsilon$ .

The features used in this study were derived from a set of HMI SHARP region magnetograms, described in Section 3.1. We preprocessed each using the techniques described in the previous section to extract a suite of geometry- and topology-based features. These feature sets, described in Section 3.2, were tested both as the sole input vectors to a many-layered fully-connected deep neural network system and in combination with the physics-based feature vectors used in current ML models.

The design of an ANN requires setting various parameters, including the number of layers and the number of neurons in the “hidden layers” between the input and the output layers; these details are discussed in Appendix A. The process of *training* the ANN involves tuning the weights in the layer interconnections so that the ANN generates the correct labels for the data, given values for all of the input features. To implement this, we divided the data into a training set and a testing set; see Section 3.3. The training data were used by standard optimization algorithms to tune the weights. We repeated this using various combinations of the feature sets and evaluated the resulting prediction accuracy using the True Skill Score on the testing set; results are reported in Section 3.4.

### 3.1 Data

The Helioseismic and Magnetic Imager (HMI) instrument on the NASA Solar Dynamics Observatory satellite views the entire solar disk at a nominal data cadence of 12 min, and has captured every active region on the Earth-facing side of the Sun since its launch in 2010. The HMI dataset contains *Space Weather HMI Active Region Patches* (SHARPs) that provide cut-out regions around each of these active regions, each with vector magnetic field and other derived quantities calculated for all pixels in the region (Bobra et al., 2014) as it rotates across the Earth-facing hemisphere of the Sun. For this study, we used only  $B_r$ , the radial field component, of SHARPs images along with the associated metadata taken by the HMI instrument between January 2010 and December 2016. The  $B_r$  component image data is available in the JSOC [hmi.sharp\\_cea\\_720s](https://hmi.sharp_cea_720s) dataset, where the magnetic field vector  $B$  is remapped to a Lambert Cylindrical Equal-Area (CEA) projection, and decomposed into the components  $B_r$ ,  $B_\theta$  and  $B_\phi$ . This segment of the SHARPs dataset contains about 2.6 million data records, each approximately 2 MB in size, totaling 5 TB of data. These active regions are known to have produced around 1250 M1.0+ flares within 24 h of the image time – see, for example, Table 1 in Schrijver (2016). We downsampled the  $B_r$  dataset to a one-hour cadence (i.e., taking every 5th magnetogram), then used the NOAA Geostationary Operational Environment Satellite (GOES) X-ray Spectrometer (XRS) flare catalog<sup>7</sup> to label each one as to whether or not the associated active region produced an M1.0+ flare in the 24 h following the time of the sample. Next, we discarded all the magnetogram images that contained invalid pixel data (NaN values). The resulting data set included 3691 active regions, of which 141 produced at least one M1.0+ flare as they crossed the Sun’s disk and 3550 did not. This corresponded to 438,539 total magnetograms, of which 5538 and 432,821 respectively, were labeled as flaring and non-flaring.

<sup>7</sup> [www.ngdc.noaa.gov/stp/space-weather/solar-data/solar-features/solar-flares/x-rays/goes/xrs/](https://www.ngdc.noaa.gov/stp/space-weather/solar-data/solar-features/solar-flares/x-rays/goes/xrs/)

A large positive/negative imbalance like this is a major challenge for any machine-learning algorithm, as described further below.

### 3.2 Features

Values for a number of physics-based features, like the total unsigned flux in the active regions and the vertical current helicity, are included in the image metadata for all SHARPs data products; see Table 1 for a full list of these quantities and [jsoc.stanford.edu/doc/data/hmi/sharp/sharp.htm](https://jsoc.stanford.edu/doc/data/hmi/sharp/sharp.htm) or Bobra et al. (2014) for details about these values and the associated calculations. This feature set – the standard in ML-based flare prediction work – is the base case for our comparison experiments.

Our procedure for computing geometry- and topology-based features from each magnetogram was as follows. To remove “topological” noise, we first filtered out pixels whose magnetic flux magnitude was below 200 G, then aggregated the resulting pixels into clusters. We computed the number and area of these clusters, then discarded all clusters whose area was less than 10% of the area of the maximum cluster. We performed these operations separately for the positive (>200 G) and negative (<−200 G) fields. We then computed an *interaction factor* (IF) between all positive/negative pairs; this quantity is defined a manner similar to the so-called Ising Energy used by Florios et al. (2018) (first introduced in Ahmed et al., 2010):

$$\text{IF} = \frac{B_{\text{pos}} \times B_{\text{neg}}}{r_{\text{min}}^2}$$

where  $B_{\text{pos}}$  and  $B_{\text{neg}}$  are the sums of the flux over the respective components and  $r_{\text{min}}$  is the smallest distance between them.<sup>8</sup> We then chose the pair with the highest IF value and derived a number of secondary features from that “Most Interacting Pair” (MIP). These features are listed in Table 2. For magnetogram images with only one sign of polarity, we assigned all features involving the missing polarity a default value of 0, including those that contain the distance terms between opposite polarities.

To compute topological features from these data, we constructed cubical complexes for each magnetogram across the range of magnetic flux magnitude thresholds between 263.15 G and 5000 G divided into ten equally spaced magnetic flux values (both 263.15 G and 5000 G inclusive): {263.16 G, 789.47 G, 1315.79 G, . . . 5000 G}.<sup>9</sup> This range covers all relevant flux levels from quiet Sun network to sunspot umbral cores. We repeated this operation separately for the positive and negative fields, yielding a total of 10 + 10 = 20 cubical complexes. We then used those complexes to construct  $\beta_1$  persistence diagrams, like that in Figure 5d, that capture the

<sup>8</sup> Florios et al. (2018) compute Ising Energy by aggregating over all pairs of individual monopoles across the two polarity regions:  $-\sum_{ij} \frac{S_i S_j}{d_{ij}}$  ( $S_i = \pm 1$  for positive/negative pairs). On the other hand, IF is computed using the summed fluxes over the two polarity regions and the smallest distance between them.

<sup>9</sup> The specific threshold values come from dividing the interval [−5000 G, 5000 G] into 20 equally spaced flux values, since we are performing the analysis for positive and negative magnetic fluxes as explained further.

**Table 1.** SHARPs magnetic field features: values for these 19 features, as well as estimates of the errors in those calculations, are provided for each magnetogram in the SHARPs database.

Acronym	Description	Units
LAT_FWT	Latitude of the flux-weighted center of active pixels	degrees
LAT_FWT	Latitude of the flux-weighted center of active pixels	degrees
AREA_ACR	Line-of-sight field active pixel area	micro hemispheres
USFLUX	Total unsigned flux	Mx
MEANGAM	Mean inclination angle, gamma	degrees
MEANGBT	Mean value of the total field gradient	G/Mm
MEANGBZ	Mean value of the vertical field gradient	G/Mm
MEANGBH	Mean value of the horizontal field gradient	G/Mm
MEANJZD	Mean vertical current density	mA/m <sup>2</sup>
TOTUSJZ	Total unsigned vertical current	A
MEANALP	Total twist parameter, alpha	1/Mm
MEANJZH	Mean current helicity	G <sup>2</sup> /m
TOTUSJH	Total unsigned current helicity	G <sup>2</sup> /m
ABSJZH	Absolute value of the net current helicity	G <sup>2</sup> /m
SAVNCPP	Sum of the absolute value of the net currents per polarity	A
MEANPOT	Mean photospheric excess magnetic energy density	ergs/cm <sup>3</sup>
TOTPOT	Total photospheric magnetic energy density	ergs/cm <sup>3</sup>
MEANSHR	Mean shear angle (measured using $B_{\text{total}}$ )	degrees
SHRGT45	Percentage of pixels with a mean shear angle greater than 45°	percent

Abbreviations: Mx: Maxwells, G: Gauss, Mm: megameters, A: Amperes.

**Table 2.** Geometry-based features: 16 total for each magnetogram. All distances were measured in terms of pixels for SDO HMI magnetogram images with a pixel resolution of one arcsecond (1 arcsecond = 725 km).

Name	Units	# of features
Total number of positive (negative) clusters	integer	2
Size of largest positive (negative) clusters	arcseconds	2
Interaction factor of MIP	G <sup>2</sup> /arcseconds <sup>2</sup>	1
COM distance between positive and negative elements of MIP	arcseconds	1
Smallest distance between elements of MIP	arcseconds	1
COM distance to smallest distance ratio	arcseconds	1
Total magnetic flux of each element of MIP	G	2
Size of each element of MIP	arcseconds <sup>2</sup>	2
Total magnetic flux per unit area of each element of MIP	G/arcseconds <sup>2</sup>	2
Total magnetic flux of largest elements in the magnetogram	G	2

threshold values at which each hole in that field is born and dies. Finally, we transformed these 2D diagrams to vectors that can be more-effectively leveraged by machine-learning methods, one vector each for the positive and the negative magnetic flux values.

In the past few years, there has been a growing literature on the problem of featurizing persistence diagrams. There are two classes of approaches: finite-dimensional embeddings or kernel-based methods. Persistence landscapes (Bubenik, 2015), persistence images (Adams et al., 2017) and persistence silhouettes (Chazal et al., 2014) are some examples of the first approach that fit template functions to the diagrams. Kernel-based methods (Bubenik, 2015; Reininghaus et al., 2015; Kusano et al., 2016; Carrière et al., 2017; Le & Yamada, 2018) use generalized scalar products that transform the diagrams implicitly into infinite-dimensional Hilbert Spaces. While these methods are useful in defining meaningful relationships between two persistence diagrams, their inherent compression can lead to a loss of information. To address this, Carrière et al. (2019) propose a layer for neural network architectures that

**Table 3.** The three basic feature sets evaluated in this study. The performance of a combination of some of these feature sets is also studied, as described in Section 3.4.

Feature set	Number of features
SHARPs (baseline)	19
Geometry-based	16
Topology-based	20

encodes most of these vector representations using a set of generalized point-wise transformation functions. Here, we chose a very simple version of the first class of approaches, counting the number of “live” holes at each value of the threshold flux. Since we have 10 cubical complexes for each polarity of the field, this produces two vectors with ten entries each. We concatenate these together to create a single vector of length 20 – the topology-based features for our study. Table 3 summarizes the three basic feature sets of interest evaluated in this study.

### 3.3 Training

We followed standard procedures to train and evaluate the ANN model, beginning by splitting the dataset into training and testing sets. During the training phase, the output labels produced by the ANN, working from the features derived from each magnetogram in the training set, were compared with the true labels for that magnetogram. The difference was then propagated backwards through the network to update the weights of the interconnections between the layers. Via multiple passes through the entire dataset (“epochs”), the weights were updated until the ANN learned a set of weights that sufficiently fit the data. Once the training phase was completed, the weights of the model were frozen permanently and the ANN could be used to make predictions on the testing data. Details and citations for all of the steps in this process can be found in [Appendix A](#).

The normal approach to splitting the data into training and testing sets – random shuffling – is not appropriate in this application. Instead, we split the magnetogram images dataset into training and testing sets *based on their SHARPs IDs*. That is, we randomly chose 70% of the SHARPs regions and placed all feature vectors extracted from the associated magnetograms in the training set. The feature vectors extracted from magnetograms from the remaining 30% made up the testing set. This ensured that the features used in evaluating the ANN did not have any similar counterparts that were used by the model during the training phase, thereby avoiding artificial boosting of prediction scores. We repeated this random shuffling procedure with 10 different random seeds to generate statistical results across 10 different training/testing sets.

### 3.4 Evaluation and discussion

To study the utility of geometry- and topology-based features in solar-flare prediction with the ANN described in the previous sections, we performed a number of experiments. After 50 training epochs in each, we computed the True Skill Statistic ([Woodcock, 1976](#)) (also known as the Hanssen-Kuiper skill score) on the testing set, which is widely used in solar-flare prediction studies:

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

where TP and FP (true and false positives) are, respectively, the images that are classified correctly and incorrectly as flaring. Similarly TN and FN are the images classified correctly and incorrectly as non-flaring. Note that  $-1 \leq \text{TSS} \leq 1$  and  $\text{TSS} = -1$  only when  $\text{TP} = \text{TN} = 0$ , so that every prediction is wrong; and  $\text{TSS} = 1$  only when  $\text{FP} = \text{FN} = 0$ , so that every prediction is correct. The skill score reflects an accuracy relative to a reference forecast that is designed such that both random forecasts and unskilled forecasts (always predict majority class) have a score of 0. When  $\text{TSS} = 0$  the prediction no better than chance in the sense that the “hit rate” is the same as the “false alarm rate.”

We performed suites of ten different training/testing experiments with different combinations of these three feature sets. The baseline experiment used the corpus of the 19 SHARPs features of [Table 1](#). The mean TSS score for this was 0.68 with the range of [0.57, 0.76]. This is in line with other

machine-learning based flare-prediction methods ([Leka et al., 2019a](#)), indicating that our neural net is a useful test case for this comparison study.

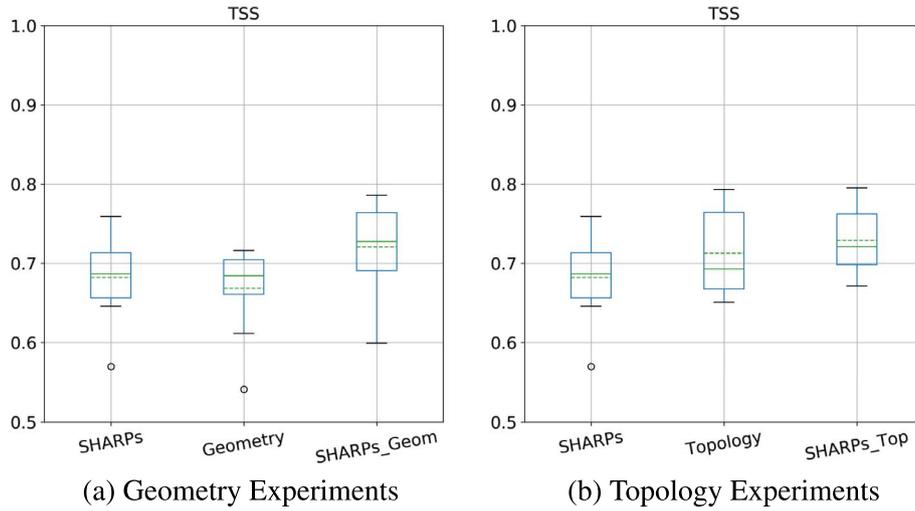
Training and validating the same ANN with the 16 geometry-based features listed in [Table 2](#) yielded a mean TSS value of 0.67 with a range of [0.54, 0.72], suggesting that these geometry-based features, surprisingly, do only slightly worse than the SHARPs features, which were constructed by solar physics experts to be their best characterizations of an active region for the purpose of eruption prediction. A TSS of 0.67 is also in line with the current *ad hoc* prediction methods that employ human experts, qualitative classifications, and historical lookup tables ([Leka et al., 2019a](#)). It is encouraging that abstract geometric features, captured automatically by an algorithm, allow a machine-learning method to match that score.

To determine how the geometry-based features worked in conjunction with the SHARPs features, we repeated these experiments using both the SHARPs metadata features and the geometry-based features (again on the same data sets), which raised the mean TSS score to 0.72, with a range of [0.60, 0.79]. That is, there is a synergy between these two feature sets: the combination works better than either one alone, indicating that when physical properties of the magnetic field are augmented with geometric properties of the active region, the predictive potential of a given magnetogram increases. The box-and-whisker plots in panel (a) of [Figure 6](#) provide a graphical comparison of the TSS scores reported in this paragraph and the previous one.

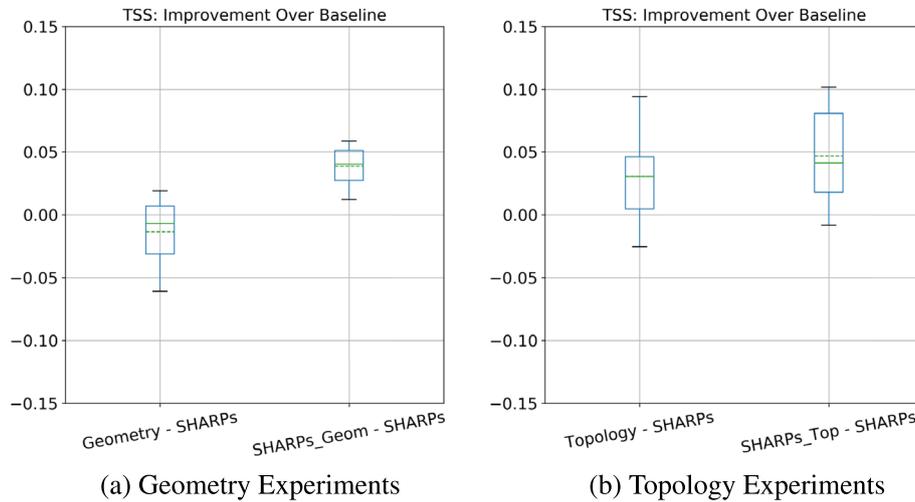
We then trained and tested the ANN using our topology-based features, alone and in combination with the SHARPs features. The results are summarized in [Figure 6b](#). Topological features alone yielded a mean TSS of 0.72 with a range of [0.65, 0.79]. This is a clear improvement over the performance of the ANN with only the baseline SHARPs features. In combination with the SHARPs features, the topological features improved the mean TSS score to 0.73 with a range of [0.67, 0.80]. Comparing the two panels of [Figure 6](#), one can observe that this was also a slight improvement over the SHARPs-geometry combination.

While the above analysis describes the aggregate TSS comparison, it is useful to understand how the developed feature sets (geometry-based, topology-based and their combination feature set counterparts with SHARPs features) perform with respect to the SHARPs-only baseline in each of the ten training/testing splits. The boxplots in [Figure 7](#) show the improvement in TSS scores for the geometry-based, topology-based and the combination feature sets with respect to the baseline TSS. This improvement is determined by subtracting the SHARPs-only TSS score for each generated training/testing data split from the TSS score of the developed feature set using the same training/testing split. For the developed feature sets, the mean improvement of the TSS over the baseline is in line with the above analysis. The geometry-based mean TSS deteriorates by 0.01 while the SHARPs-geometry combination show a mean TSS improvement of 0.04. The topology-based features and the SHARPs-topology combination show a mean TSS improvement of 0.03 and 0.05 respectively.

The results of these experiments confirm our intuition that abstract measures of the shapes of magnetogram structures can be indicative of eruption potential. This is additionally satisfying since the historical method of assessing the eruptive



**Fig. 6.** Box-and-whisker plots of the TSS scores for different feature sets used across the geometry and topology experiments, reported for ten different training/testing splits generated from the same dataset. The green solid and dotted lines indicate the median and the mean of the TSS scores, respectively, for each plot. In plots (a) and (b), the label “SHARPs\_Geom” denotes a combined feature set of SHARPs and geometry-based features, while “SHARPs\_Top” denotes a combined feature set of SHARPs and topology-based features respectively.



**Fig. 7.** Box-and-whisker plots of the TSS score improvements for the developed feature sets over the baseline (SHARPs-only) feature set, reported for ten different training/testing splits generated from the same dataset. The green solid and dotted lines indicate the median and the mean of the TSS score improvements, respectively, for each plot. In plots (a) and (b), the label “SHARPs\_Geom” denotes a combined feature set of SHARPs and geometry-based features, while “SHARPs\_Top” denotes a combined feature set of SHARPs and topology-based features respectively.

potential of an active region relies on human forecasters essentially analyzing “complexity of shape” in a qualitative manner. Our results demonstrate that computational geometry and computational topology have captured this in a quantitative and repeatable algorithm. This not only points the way forward to a more robust set of features for machine-learning based eruption prediction architectures. It may even lead to new and more-effective way to classify sunspot active regions.

In this section, we have relied on TSS as a sole metric for comparing the performance of the model for different feature sets. An extension of this analysis to other metrics is described

in [Appendix B](#). The scores for these alternative metrics, presented in [Tables B.1 and B.2](#), indicate that their trends across the different feature sets are similar to the TSS trends discussed above. Additionally, [Appendix C](#) describes the predictive power of individual features from the different feature sets using the Fisher score – a univariate feature ranking method. As shown in [Figure C.1a and b](#), 9 of the geometry-based features make it to the top 15 highest ranking features when ranking SHARPs and geometry-based features, whereas 6 of the top 15 features belong to the topology feature set in the SHARPs-topology ranking experiment.

## 4 Conclusions

We have shown that abstract spatial properties of magnetograms can be useful in machine-learning methods for solar-flare prediction. We extract these properties from magnetograms using computational geometry and computational topology techniques, producing feature vectors that compete well with the traditional physics-based SHARPs features for the purposes of machine learning methods for flare prediction. A neural net trained on a large corpus of active regions from HMI, preprocessed using these techniques, classifies magnetograms as flare precursors with a slightly higher True Skill Statistic score than the same model using traditional SHARPs features – the educated assessment by human experts as to the best information with which to characterize an active region. Combining shape- and physics-based features further improved the TSS scores, indicating a synergy between the two types of information. Following acceptance of this article, we have recently become aware of further work by the Pulkovo group that has utilized some topological data analysis techniques in a support vector regression model for time-to-flare prediction, trained on a SHARPs data set comprised only of flaring regions – a limitation that is problematic in machine learning, where one wants a training set that includes both positive (flaring) and negative (non-flaring) instances (Knyazeva et al., 2017). Perhaps for this reason, Knyazeva et al. (2017) found that the topological features did not – in contrast to our findings – allow the SVM to outperform the SHARPs features.

The power of these highly abstract classifications of structure may be surprising, particularly in view of the fact that magnetograms are just the boundary conditions of the full fields whose complicated dynamics are what produce eruptions. Even so, active region shape has fundamental, meaningful connections to the physics leading up to an eruption. As an active region emerges, becoming progressively larger and more complex, the shapes of the opposite polarity structures on a magnetogram capture the evolution of the photospheric field. Flares occur when that field forces a rearrangement of the fields in the corona, where magnetic reconnection occurs. Shape is what human forecasters use in their classifications, and topology can codify structural complexity in a formal and yet practical manner that makes it ideally suited for capturing this richness. In current operational practice, an expert examines sunspot images and/or magnetograms, classifies a sunspot active region according to taxonomies developed empirically, and then uses look-up tables of historical probabilities to say whether or not it will erupt within a future time period. In a sense, our approach systematizes this “human-in-the-loop” forecasting approach by applying the mathematical concepts of topology to address the shape and connectivity of *sets* derived from the structure of the photospheric magnetic field.

In a future paper we plan to validate our results using a number of variants of our neural network architecture. In addition, our results can be given a firmer statistical basis by using more training and testing partitions of the HMI data, different initial conditions for the network weights, and different numbers of training epochs. Finally, neural networks are only one type of machine-learning model: we will obviously need to compare with other methods, like SVMs, decision trees, and convolutional neural networks (CNNs), before making broader claims

about the general utility of shape-based features in machine learning.

There are also a number of data issues that require more attention. Firstly, in our current approach, we simply discard the magnetograms containing invalid pixel data. Addressing such pixel-level anomalies in the data – e.g. by smoothing – would be a good next step here. This will help prevent loss of any potentially interesting samples and allow us to work with larger datasets. It will also be important to address the appropriate choice of a temporal resolution: is the selection of every fifth sample (i.e., hourly resolution) optimal? This could be carried out in a purely mathematical fashion, via expert analysis of persistence diagrams, but it would also be useful to evaluate the effects of different temporal cadences on the TSS of the neural-net model.

Another important issue is to address *temporal evolution*: all current ML-based flare prediction methods, including those reported in this paper, use only single snapshots of the field. The dynamics of the structure of an active region – the progression through time of the shapes and their relationships – could be captured using topological approaches that track those structures through time and space, such as the CROCKER plots of Topaz et al. (2015).

Last but not least: the geometric and topological feature sets used here are only a first cut; extending them to capture different aspects of the shape of an active region could improve the results. For example, computational geometry can extract the curvatures of the boundaries of the different polarity regions. In terms of topology, the possibilities include higher-dimensional Betti numbers, alternative complex constructions, and different ways of featurizing persistence diagrams. These could give us new tools to address the structural complexity of active regions that are potentially evolving towards an eruption.

*Acknowledgements.* This material is based upon work sponsored by the National Science Foundation Award (Grant No. CMMI 1537460) opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the NSF. The University of Colorado Space Weather Technology, Research, and Education Center (SWx TREC) is supported by a Grand Challenge grant from the Chancellor of the University of Colorado and hosted in the College of Engineering and Applied Sciences (CEAS) on the Boulder campus. The editor thanks two anonymous reviewers for their assistance in evaluating this paper.

## References

- Adams H, Emerson T, Kirby M, Neville R, Peterson C, Shipman P, Chepushtanova S, Hanson E, Motta F, Ziegelmeier L. 2017. Persistence images: A stable vector representation of persistent homology. *J Mach Learn Res* **18**(1): 218–252. <https://dl.acm.org/doi/abs/10.5555/3122009.3122017>.
- Ahmed OW, Qahwaji R, Colak T, Dudok De Wit T, Ipson S. 2010. A new technique for the calculation and 3D visualisation of magnetic complexities on solar satellite images. *Vis Comput* **26**(5): 385–395. <https://doi.org/10.1007/s00371-010-0418-1>.
- Aulanier G,ariat E, Démoulin P. 2005. Current sheet formation in Quasi-separatrix layers and hyperbolic flux tubes. *A&A* **444**: 961–976. <https://doi.org/10.1051/0004-6361:20053600>.

- Barnes G, Leka KD. 2006. Photospheric magnetic field properties of flaring vs. non-flare quiet active regions. III. Magnetic Charge Topology Models. *Astrophys J* **646**: 1303–1318. <https://doi.org/10.1086/510282>.
- Barnes G, Longcope DW, Leka KD. 2005. Implementing a magnetic charge topology model for solar active regions. *Astrophys J* **629**: 561–571. <https://doi.org/10.1086/431175>.
- Barnes G, Leka KD, Schrijver CJ, Colak T, Qahwaji R, et al. 2016. A comparison of flare forecasting methods. I. Results from the “All-Clear” Workshop. *Astrophys J* **829**(2): 1–32. <https://doi.org/10.3847/0004-637x/829/2/89>.
- Benvenuto F, Piana M, Campi C, Massone AM. 2018. A hybrid supervised/unsupervised machine learning approach to solar flare prediction. *Astrophys J* **853**(1): 90. <https://doi.org/10.3847/1538-4357/aaa23c>.
- Bhavsar SP, Lauer DA. 1996. *Analysis of the CFA “Great Wall” using the minimal spanning tree*. Springer, Netherlands, Dordrecht, pp. 517–519. [https://doi.org/10.1007/978-94-009-0145-2\\_66](https://doi.org/10.1007/978-94-009-0145-2_66).
- Bobra MG, Couvidat S. 2015. Solar flare prediction using SDO/HMI vector magnetic field data with a machine-learning algorithm. *Astrophys J* **798**(2): 135. <https://doi.org/10.1088/0004-637x/798/2/135>.
- Bobra MG, Sun X, Hoeksema JT, Turmon M, Liu Y, Hayashi K, Barnes G, Leka KD. 2014. The helioseismic and magnetic imager (HMI) vector magnetic field pipeline: SHARPs – Space-Weather HMI Active Region Patches. *Sol Phys* **289**(9): 3549–3578. <https://doi.org/10.1007/s11207-014-0529-3>.
- Boucheron LE, Al-Ghraibah A, McAteer RTJ. 2015. Prediction of solar flare size and time-to-flare using support vector machine regression. *Astrophys J* **812**: 51. <https://doi.org/10.1088/0004-637x/812/1/51>.
- Bubenik P. 2015. Statistical topological data analysis using persistence landscapes. *J Mach Learn Res* **16**(1): 77–102. <https://dl.acm.org/doi/10.5555/2789272.2789275>.
- Campi C, Benvenuto F, Massone AM, Bloomfield DS, Georgoulis MK, Piana M. 2019. Feature ranking of active region source properties in solar flare forecasting and the uncompromised stochasticity of flare occurrence. *Astrophys J* **883**(2): 150. <https://doi.org/10.3847/1538-4357/ab3c26>.
- Camporeale E. 2019. The challenge of machine learning in space weather: Nowcasting and forecasting. *Space Weather* **17**(8): 1166–1207. <https://doi.org/10.1029/2018SW002061>.
- Carrière M., Cuturi M., Oudot S. 2017. Sliced Wasserstein Kernel for persistence diagrams. In: *Proceedings of the 34th International Conference on Machine Learning – Volume 70, ICML’17*. PMLR, International Convention Centre, Sydney, Australia, pp. 664–673. <https://jmlr.org>, <https://dl.acm.org/doi/10.5555/3305381.3305450>.
- Carrière M, Chazal F, Ike Y, Lacombe T, Royer M, Umeda Y. 2019. PersLay: A simple and versatile neural network layer for persistence diagrams. to appear in *The Proceedings of the International Conference on Artificial Intelligence and Statistics, 2020*.
- Chazal F, Fasy BT, Lecci F, Rinaldo A, Wasserman L. 2014. Stochastic convergence of persistence landscapes and silhouettes. In: *Proceedings of the Thirtieth Annual Symposium on Computational Geometry, SOCG’14*, 474:474–474:483, ACM, New York, NY, USA. <https://doi.org/10.1145/2582112.2582128>
- Colak T, Qahwaji R. 2009. Automated solar activity prediction: A hybrid computer platform using machine learning and solar imaging for automated prediction of solar flares. *Space Weather* **7**: S06001. <https://doi.org/10.1029/2008SW000401>.
- Crown MD. 2012. Validation of the NOAA Space Weather Prediction Center’s solar flare forecasting look-up table and forecaster-issued probabilities. *Space Weather* **10**: S06006. <https://doi.org/10.1029/2011SW000760>.
- de Silva V, Ghrist R. 2007. Coverage in sensor networks via persistent homology. *Algebr Geom Topol* **7**(1): 339–358. <https://doi.org/10.2140/agt.2007.7.339>.
- Demoulin P, Henoux JC, Priest ER, Mandrini CH. 1996. Quasi-separatrix layers in solar fares.I. Method. *A&A* **308**: 643–655.
- DeRosa ML, Schrijver CJ, Barnes G, Leka KD, Lites BW, et al. 2009. A critical assessment of nonlinear force-free field modeling of the solar corona for active region 10953. *Astrophys J* **696**(2): 1780–1791. <https://doi.org/10.1088/0004-637x/696/2/1780>.
- Devogele M, Rivet JP, Tanga P, Bendjoya P, Surdej J, Bartczak P, Hanuš J. 2015. A method to search for large-scale concavities in asteroid shape models. *Mon Notic Roy Astron Soc* **453**: 2232–2240. <https://doi.org/10.1093/mnras/stv1740>.
- Duchi J, Hazan E, Singer Y. 2011. Adaptive subgradient methods for online learning and stochastic optimization. *J Mach Learn Res* **12** (null): 2121–2159. <https://dl.acm.org/doi/10.5555/1953048.2021068>.
- Edelsbrunner H, Letscher D, Zomorodian A. 2000. Topological persistence and simplification. *Discrete Comput Geom* **28**: 511–533. <https://doi.org/10.1007/s00454-002-2885-2>.
- Florios K, Kontogiannis I, Park S-H, Guerra JA, Benvenuto F, Bloomfield DS, Georgoulis MK. 2018. *Forecasting solar flares using magnetogram-based predictors and machine learning*. *Sol Phys* **293**(2): 28. <https://doi.org/10.1007/s11207-018-1250-4>.
- Forrest AR. 1971. II. Current Developments in the Design and Production of Three-Dimensional Curved Objects – Computational Geometry. *Proc Roy Soc Lond A: Math Phys Eng Sci* **321** (1545): 187–195. <https://doi.org/10.1098/rspa.1971.0025>.
- Gallagher P, Moon Y-J, Wang H. 2002. Active-region monitoring and flare forecasting – I. Data processing and first results. *Sol Phys* **209**: 171–183. <https://doi.org/10.1023/A:1020950221179>.
- Ghrist R. 2008. Barcodes: The persistent topology of data. *Bull Am Math Soc* **45**(1): 61–75. <https://doi.org/10.1090/S0273-0979-07-01191-3>.
- Glorot X., Bengio Y. 2010. Understanding the difficulty of training deep feedforward neural networks. In: *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics, Vol. 9 of Proceedings of Machine Learning Research*, Teh Y.W., Titterton M. (Eds.), PMLR, Chia Laguna Resort, Sardinia, Italy, pp. 249–256. <https://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.207.2059>.
- Gu Q., Li, Z., Han, J. 2011. Generalized Fisher score for feature selection. In: *Proceedings of the Twenty-Seventh Conference on Uncertainty in Artificial Intelligence, UAI’11*, AUAI Press, Arlington, Virginia, USA, pp. 266–273. <https://dl.acm.org/doi/abs/10.5555/3020548.3020580>.
- Guerra JA, Park S-H, Gallagher PT, Kontogiannis I, Georgoulis MK, Bloomfield DS. 2018. Active region photospheric magnetic properties derived from line-of-sight and radial fields. *Sol Phys* **293**(1): 9. <https://doi.org/10.1007/s11207-017-1231-z>.
- Hale GE, Ellerman F, Nicholson SB, Joy AH. 1919. The magnetic polarity of sun-spots. *Astrophys J* **49**: 153. <https://doi.org/10.1086/142452>.
- Haykin S. 1998. *Neural networks: A comprehensive foundation*, 2nd edn., Prentice Hall PTR, USA. <https://dl.acm.org/doi/book/10.5555/521706>.
- Huang X, Wang H, Xu L, Liu J, Li R, Dai X. 2018. Deep learning based solar flare forecasting model. I. Results for line-of-sight magnetograms. *Astrophys J* **856**(1): 7. <https://doi.org/10.3847/1538-4357/aaae00>.
- Jonas E, Bobra M, Shankar V, Todd Hoeksema J, Recht B. 2018. Flare prediction using photospheric and coronal image data. *Sol Phys* **293**(3): 48. <https://doi.org/10.1007/s11207-018-1258-9>.

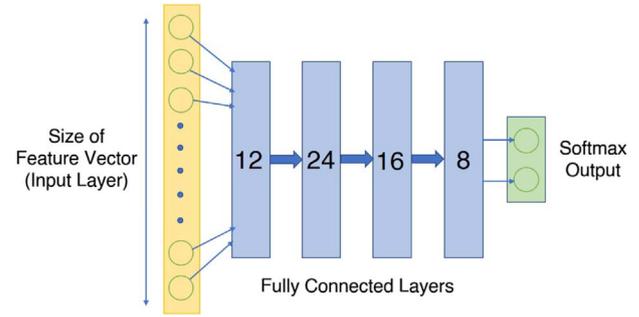
- Kaczynski T., Mischaikow K., Mrozek M. 2004. Computational homology, Vol. 157 of *Applied Mathematical Sciences*, Springer-Verlag, New York. <https://doi.org/10.1007/B97315>.
- Kingma D.P., Ba J. 2014. Adam: A method for stochastic optimization. *arXiv e-prints*. arXiv:1412.6980.
- Knyazeva IS, Makarenko NG, Livshits MA. 2011. Detection of new emerging magnetic flux from the topology of SOHO/MDI magnetograms. *Astron Rep* **55(5)**: 463. <https://doi.org/10.1134/S1063772911050040>.
- Knyazeva IS, Urtiev FA, Makarenko NG. 2017. On the prognostic efficiency of topological descriptors for magnetograms of active regions. *Geomagn Aeron* **57(8)**: 1086–1091. <https://doi.org/10.1134/S0016793217080126>.
- Kontogiannis I, Georgoulis MK, Park S-H, Guerra JA. 2018. Testing and improving a set of morphological predictors of flaring activity. *Sol Phys* **293(6)**: 96. <https://doi.org/10.1007/s11207-018-1317-2>.
- Kusano G., Fukumizu, K., Hiraoka, Y. 2016. Persistence weighted Gaussian Kernel for topological data analysis. In: *Proceedings of the 33rd International Conference on International Conference on Machine Learning – Volume 48, ICML'16, 2004–2013*. PMLR, New York, New York, USA. <https://dl.acm.org/doi/10.5555/3045390.3045602>.
- Le T., Yamada, M. 2018. Persistence Fisher Kernel: A Riemannian manifold Kernel for persistence diagrams. In: *Proceedings of the 32nd International Conference on Neural Information Processing Systems, NIPS'18, 10028–10039*, Curran Associates Inc., Red Hook, NY, USA. <https://dl.acm.org/doi/10.5555/3327546.3327666>.
- Leka KD, Park S-H, Kusano K, Andries J, Barnes G, et al. 2019a. A comparison of flare forecasting methods. II. Benchmarks, metrics, and performance results for operational solar flare forecasting systems. *Astrophys J Suppl Ser* **243(2)**: 36. <https://doi.org/10.3847/1538-4365/ab2e12>.
- Leka KD, Park S-H, Kusano K, Andries J, Barnes G, et al. 2019b. A comparison of flare forecasting methods. III. Systematic behaviors of operational solar flare forecasting systems. *Astrophys J* **881(2)**: 101. <https://doi.org/10.3847/1538-4357/ab2e11>.
- Longcope DW. 2005. Topological methods for the analysis of solar magnetic fields. *Liv Rev Sol Phys* **2(1)**: 7. <https://doi.org/10.12942/lrsp-2005-7>.
- Makarenko N, Malkova D, Machin M, Knyazeva I, Makarenko I. 2014. Methods of computational topology for the analysis of dynamics of active regions of the sun. *J Math Sci* **203(6)**: 806–815. <https://doi.org/10.1007/s10958-014-2170-y>.
- McAteer RTJ, Gallagher PT, Conlon PA. 2010. Turbulence, complexity, and solar flares. *Adv Space Res* **45**: 1067–1074. <https://doi.org/10.1016/j.asr.2009.08.026>.
- McIntosh PS. 1990. The classification of sunspot groups. *Sol Phys* **125**: 251–267. <https://doi.org/10.1007/BF00158405>.
- Metcalf TR, DeRosa ML, Schrijver CJ, Barnes G, van Ballegoijen AA, Wiegmann T, Wheatland MS, Valori G, McTiernan JM. 2008. Nonlinear force-free modeling of coronal magnetic fields. II. Modeling a filament arcade and simulated chromospheric and photospheric vector fields. *Sol Phys* **247(2)**: 269–299. <https://doi.org/10.1007/s11207-007-9110-7>.
- Nishizuka N, Sugiura K, Kubo Y, Den M, Watari S, Ishii M. 2017. Solar flare prediction model with three machine-learning algorithms using ultraviolet brightening and vector magnetograms. *Astrophys J* **835**: 156. <https://doi.org/10.3847/1538-4357/835/2/156>.
- Nishizuka N, Sugiura K, Kubo Y, Den M, Ishii M. 2018. Deep flare net (DeFN) model for solar flare prediction. *Astrophys J* **858**: 113. <https://doi.org/10.3847/1538-4357/aab9a7>.
- Park E, Moon Y-J, Shin S, Yi K, Lim D, Lee H, Shin G. 2018. Application of the deep convolutional neural network to the forecast of solar flare occurrence using full-disk solar magnetograms. *Astrophys J* **869(2)**: 91. <https://doi.org/10.3847/1538-4357/aad40>.
- Paszke A, Gross S, Massa F, Lerer A, Bradbury J, et al. 2019. PyTorch: An Imperative Style, High-Performance Deep Learning Library. *Advances in Neural Information Processing Systems*, **32**, Curran Associates Inc. pp. 8026–8037. <http://papers.nips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library>.
- Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, et al. 2011. Scikit-learn: Machine learning in python. *J Mach Learn Res* **12(null)**: 2825–2830. <https://dl.acm.org/doi/10.5555/1953048.2078195>.
- Pesnell WD, Thompson BJ, Chamberlin PC. 2011. The solar dynamics observatory (SDO). *Sol Phys* **275(1–2)**: 3–15. <https://doi.org/10.1007/s11207-011-9841-3>.
- Preparata FP, Shamos MI. 1985. Computational geometry: An introduction, Springer-Verlag, New York. <https://doi.org/10.1007/978-1-4612-1098-6>.
- Priest ER, Démoulin P. 1995. Three-dimensional magnetic reconnection without null points. I. Basic theory of magnetic flipping. *J Geophys Res* **100**: 23443–23464. <https://doi.org/10.1029/95JA02740>.
- Reininghaus J., Huber S., Bauer U., Kwitt R. 2015. A stable multi-scale kernel for topological machine learning. In: *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4741–4748. <https://doi.org/10.1109/CVPR.2015.7299106>.
- Robins V, Meiss J, Bradley E. 1998. Computing connectedness: An exercise in computational topology. *Nonlinearity* **11(4)**: 913–922.
- Robins V, Meiss J, Bradley E. 2000. Computing connectedness: Disconnectedness and discreteness. *Phys D* **139(3–4)**: 276–300.
- Scherrer PH, Schou J, Bush RI, Kosovichev AG, Bogart RS, et al. 2011. The helioseismic and magnetic imager (HMI) investigation for the solar dynamics observatory (SDO). *Sol Phys* **275(1–2)**: 207–227. <https://doi.org/10.1007/s11207-011-9834-2>.
- Schrijver CJ. 2007. A characteristic magnetic field pattern associated with all major solar flares and its use in flare forecasting. *Astrophys J Lett* **655(2)**: L117. <https://doi.org/10.1086/511857>.
- Schrijver CJ. 2016. The nonpotentiality of coronae of solar active regions, the dynamics of the surface magnetic field, and the potential for large flares. *Astrophys J* **820(2)**: 1–17. <https://doi.org/10.3847/0004-637x/820/2/103>.
- Schrijver CJ, DeRosa ML, Metcalf TR, Liu Y, McTiernan J, Régnier S, Valori G, Wheatland MS, Wiegmann T. 2006. Nonlinear force-free modeling of coronal magnetic fields part I: A quantitative comparison of methods. *Sol Phys* **235(1–2)**: 161–190. <https://doi.org/10.1007/s11207-006-0068-7>.
- Schrijver CJ, DeRosa ML, Metcalf T, Barnes G, Lites B, et al. 2008. Nonlinear force-free field modeling of a solar active region around the time of a major flare and coronal mass ejection. *Astrophys J* **675**: 1637–1644. <https://doi.org/10.1086/527413>.
- Singh G, Memoli F, Ishkhanov T, Sapiro G, Carlsson G, Ringach D. 2008. Topological analysis of population activity in visual cortex. *J Vis* **8(8: 11)**: 1–18. <https://doi.org/10.1167/8.8.11>.
- Tarr L, Longcope D. 2012. Calculating energy storage due to topological changes in emerging active region NOAA AR 11112. *Astrophys J* **749(1)**: 64. <https://doi.org/10.1088/0004-637x/749/1/64>.
- Tarr L, Longcope D, Millhouse M. 2013. Calculating separate magnetic free energy estimates for active regions producing multiple flares: NOAA AR11158. *Astrophys J* **770(1)**: 4. <https://doi.org/10.1088/0004-637X/770/1/4>.

- Topaz CM, Ziegelmeier L, Halverson T. 2015. Topological data analysis of biological aggregation models. *PLoS One* **10**(5): 1–26. <https://doi.org/10.1371/journal.pone.0126383>.
- Wang YM, Sheeley NR Jr. 1992. On potential field models of the solar corona. *Astrophys J* **392**: 310. <https://doi.org/10.1086/171430>.
- Wheatland MS. 2004. A Bayesian approach to solar flare prediction. *Astrophys J* **609**: 1134–1139. <https://doi.org/10.1086/421261>.
- Wiegmann T, Sakurai T. 2012. Solar force-free magnetic fields. *Liv Rev Sol Phys* **9**: 5. <https://doi.org/10.12942/lrsp-2012-5>.
- Woodcock F. 1976. The evaluation of yes/no forecasts for scientific and administrative purposes. *Mon Weather Rev* **104**(10): 1209–1214. [https://doi.org/10.1175/1520-0493\(1976\)104<1209:TEOYFF>2.0.CO;2](https://doi.org/10.1175/1520-0493(1976)104<1209:TEOYFF>2.0.CO;2).
- Xu X, Cisewski-Kehe J, Green SB, Nagai D. 2019. Finding cosmic voids and filament loops using topological data analysis. *Astron Comput* **27**: 34–52. <https://doi.org/10.1016/j.ascom.2019.02.003>.
- Yang X, Lin G, Zhang H, Mao X. 2013. Magnetic nonpotentiality in photospheric active regions as a predictor of solar flares. *Astrophys J* **774**(2): L27. <https://doi.org/10.1088/2041-8205/774/2/L27>.
- Yu D, Huang X, Wang H, Cui Y, Hu Q, Zhou R. 2010. Short-term solar flare level prediction using a Bayesian network approach. *Astrophys J* **710**(1): 869. <https://doi.org/10.1088/0004-637x/710/1/869>.
- Yuan Y, Shih F, Jing J, Wang H. 2010. Solar flare forecasting using sunspot-groups classification and photospheric magnetic parameters. *Proc Int Astron Union* **6**: 446–450. <https://doi.org/10.1017/S1743921311015742>.
- Zheng Y, Li X, Wang X. 2019. Solar flare prediction with the hybrid deep convolutional neural network. *Astrophys J* **885**(1): 73. <https://doi.org/10.3847/1538-4357/ab46bd>.
- Zomorodian A. 2012. Topological data analysis, vol. **70** of *Advances in Applied and Computational Topology*, American Mathematical Society, Providence. <https://doi.org/10.1090/psapm/070>.

## Appendix A

### Artificial neural network: design and implementation

To evaluate the different feature sets, we designed a standard feedforward neural network using PyTorch (Paszke et al., 2019) with six densely connected layers. The input layer size is variable depending on the size of the feature set; the output layer contains two neurons corresponding to the two classes – flaring and non-flaring. The four intermediate layers contain 12, 24, 16 and 8 neurons respectively, when counting from the direction of the input to the output layer. To prevent over-fitting, a Ridge Regression regularization with a penalty



**Fig. A.1.** Representation of the feed-forward Artificial Neural Network (ANN) model used in this work. The number of neurons in the input layer is equal to the size of the feature set being used. The output layer has two neurons – one each for the flaring and non-flaring classes.

factor of 0.01 is used at each layer that limits the  $L_2$  sum of all the weights. For the study reported here, we designed the neural network in a trial-and-error process that balanced complexity of representation against training and testing time. In future work, we play to formalize this approach of fine-tuning the model hyperparameters (learning rates, regularization parameters, loss weights, etc.) using a separate validation and a testing set (as opposed to only using a testing set). The validation set can be used for determining the optimal hyperparameter values, with the testing set used for a final evaluation of the model. The layout of the network is shown in Figure A.1; the network parameters are summarized in Table A.1.

We used the standard back-propagation algorithm built into PyTorch to train the weights of the ANN, which are initialized using the Glorot uniform initializer (Glorot & Bengio, 2010) at the beginning of the training process. Since our problem is a two category classification one, we chose a weighted binary cross-entropy function to measure the training loss between the model output and the true target that is propagated backwards to update the network weights. Over a training instance for  $N$  data samples over  $C$  classes (here  $C = 2$ ), this loss function is given by,

$$L = \sum_{i=1}^N \sum_{c=1}^C w_c y'_{ic} \log p(y_{ic}), \quad (\text{A.1})$$

where  $p(y_{ic})$  represents the target prediction probability for class  $c$  of the  $i$ th data point, whose real target value is  $y'_{ic}$ . The loss function class weights  $w_c$  in this formula determine the penalty of incorrectly predicting the probability for the

**Table A.1.** Parameter settings for the Artificial Neural Network (ANN) model implementation.

Model parameters	Settings
Total hidden layers	4
Hidden layer activation function	Rectified linear unit
Output layer activation function	Softmax
Weight regularization	L2 with a penalty of $10^{-2}$
Training batch size	64
Loss function	Weighted Binary Cross-Entropy (weights: $w_{\text{no-flare}} = 1$ , $w_{\text{flare}} = 100$ )
Optimization method	Adagrad (parameters: learning rate = $5 \times 10^{-4}$ )
Number of Epochs	50

**Table B.1.** Performance evaluations of the various feature sets for the geometry experiments using various metrics. The top three rows report the mean metric score (with the standard deviation) across ten different training/testing sets, while the bottom two rows report the mean improvement with the standard deviation of the metric scores with respect to the SHARP feature set.

Feature set	Accuracy	Precision	Recall	F-10 score	HSS
SHARP	$0.89 \pm 0.01$	$0.09 \pm 0.01$	$0.79 \pm 0.06$	$0.15 \pm 0.01$	$0.13 \pm 0.01$
Geometry	$0.88 \pm 0.01$	$0.08 \pm 0.01$	$0.79 \pm 0.06$	$0.14 \pm 0.01$	$0.12 \pm 0.01$
SHARP + Geometry	$0.90 \pm 0.01$	$0.09 \pm 0.01$	$0.82 \pm 0.06$	$0.17 \pm 0.01$	$0.15 \pm 0.01$
Geometry improvement	$-0.01 \pm 0.00$	$-0.01 \pm 0.00$	$0.00 \pm 0.03$	$-0.01 \pm 0.01$	$-0.02 \pm 0.01$
SHARP + Geometry improvement	$0.01 \pm 0.00$	$0.01 \pm 0.00$	$0.03 \pm 0.02$	$0.01 \pm 0.00$	$0.01 \pm 0.00$

**Table B.2.** Performance evaluations of the various feature sets for the topology experiments using various metrics. The top three rows report the mean metric score (with the standard deviation) across ten different training/testing sets, while the bottom two rows report the mean improvement with the standard deviation of the metric scores with respect to the SHARP feature set.

Feature set	Accuracy	Precision	Recall	F-1 score	HSS
SHARP	$0.89 \pm 0.01$	$0.09 \pm 0.01$	$0.79 \pm 0.06$	$0.15 \pm 0.01$	$0.13 \pm 0.01$
Topology	$0.90 \pm 0.01$	$0.10 \pm 0.01$	$0.81 \pm 0.06$	$0.18 \pm 0.02$	$0.16 \pm 0.02$
SHARP + Topology	$0.90 \pm 0.01$	$0.09 \pm 0.01$	$0.83 \pm 0.05$	$0.17 \pm 0.02$	$0.15 \pm 0.02$
Topology improvement	$0.01 \pm 0.01$	$0.01 \pm 0.01$	$0.02 \pm 0.05$	$0.02 \pm 0.01$	$0.02 \pm 0.01$
SHARP + Topology improvement	$0.01 \pm 0.01$	$0.01 \pm 0.00$	$0.04 \pm 0.04$	$0.01 \pm 0.01$	$0.02 \pm 0.01$

associated classes  $c$ . As demonstrated in Section 3.1, our dataset is highly unbalanced, with the non-flaring class over-represented by two orders of magnitude. By choosing  $w_{\text{no-flare}}:w_{\text{flare}} = 1:100$ , we over-penalize the minority class (flaring magnetograms) to offset the effect of its size. This is one of the many ways to mitigate imbalance in training sets, and has been used in the flare-prediction literature (Bobra & Couvidat, 2015; Nishizuka et al., 2018). Finally, we use the Adagrad optimizer (Duchi et al., 2011) to perform the weight update during the backpropagation phase. An extension of the Stochastic Gradient Descent algorithm, the Adagrad optimizer adapts the learning rate to the individual parameters, so that more sensitive parameters which significantly affect the output are assigned a lower learning rate, whereas less sensitive parameters are updated with a greater learning rate. (Our initial choice was the Adam optimizer, Kingma & Ba, 2014, but we observed jitters in the prediction scores after convergence when using it.) We used the default parameters for the PyTorch implementation with the learning rate multiplier set to  $5 \times 10^{-4}$ .

All experiments were run on a NVIDIA TITAN V graphics processing unit (NVIDIA Driver Version 440.31, CUDA Version 10.2). Across all feature sets, the training time per epoch was approximately 22 s. For a total of 50 epochs, the total training time then amounted to approximately 19 min. The average validation time across all the feature sets was approximately 13 s.

## Appendix B

### Alternative evaluation metrics

Here, we report the performance of the various feature sets using some other standard metrics. Comparing the forecast

against the actual event, we populate the contingency table values: true positives (TP), false positives (FP), true negatives (TN) and false negatives (FN). Using these four classes, we compute the five additional metrics for each experiment: Accuracy, Precision, Recall, the F-1 Score (F1) and the Heidke Skill Score (HSS):

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (\text{B.1})$$

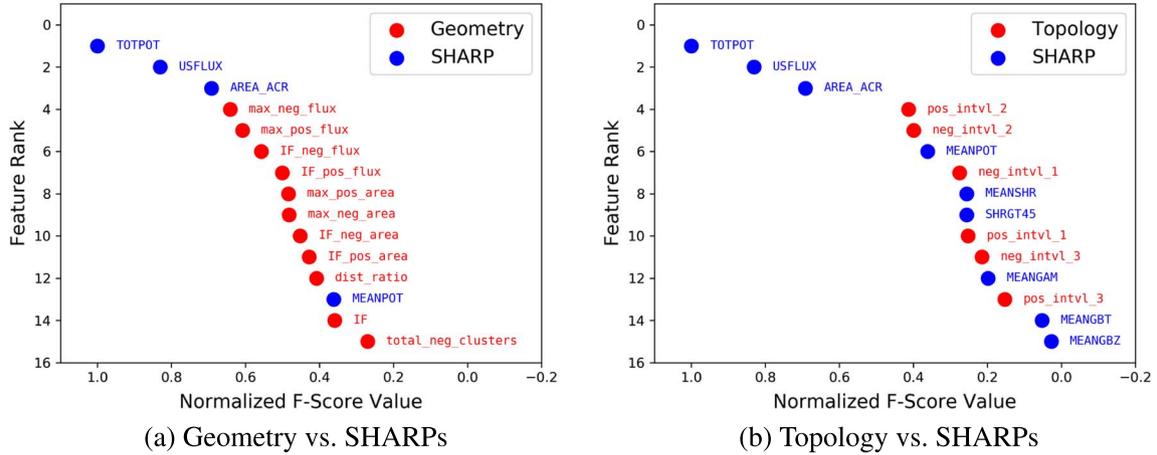
$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (\text{B.2})$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (\text{B.3})$$

$$\text{F1} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (\text{B.4})$$

$$\text{HSS} = \frac{2(\text{TP} \times \text{TN} - \text{FP} \times \text{FN})}{(\text{TP} + \text{FN})(\text{FN} + \text{TN}) + (\text{TP} + \text{FP})(\text{FP} + \text{TN})}. \quad (\text{B.5})$$

For a detailed explanation of these metrics, refer to Bobra & Couvidat (2015) (note that the  $\text{HSS}_2$  variant of the Heidke Skill Score from Bobra & Couvidat, 2015 is used here). The mean metric scores and the mean score improvements along with the standard deviations over the baseline the geometry and topology experiments are summarized in Tables B.1 and B.2 respectively. The trends for these metrics match the trends in the TSS scores described in Section 3.



**Fig. C.1.** Feature ranking, as determined via the Fisher score method, for the experiments reported in this paper. The  $x$ -axis is the normalized Fisher score and the  $y$ -axis is the feature rank. Red identifies features based on topology and geometry; blue indicates SHARPs features. The features are plotted in descending order of the F-Score value on the  $x$ -axis and ascending order of the rank on the  $y$ -axis (both axes are inverted). The topological features described by the formats `pos_intvl_n` and `neg_intvl_n` represent the number of holes for the  $n$ th magnetic flux threshold for the positive and negative polarities respectively.

## Appendix C

### Feature ranking

Evaluating the model performance scores using different feature sets is one way of determining how effective the engineered features are in providing an accurate model prediction. An alternative evaluation strategy is to compute the Fisher score (Gu et al., 2011) individually for each feature, as done in Bobra & Couvidat (2015). A method of univariate feature ranking, the Fisher score (or the F-score) determines the ability of the feature to separate the distributions of the classes in the dataset. For a dataset with two labels, the F-score of a feature  $i$ , as defined in Bobra & Couvidat (2015), is given by,

$$F(i) = \frac{(\bar{x}_i^+ - \bar{x}_i)^2 + (\bar{x}_i^- - \bar{x}_i)^2}{\frac{1}{n^+ - 1} \sum_{k=1}^{n^+} (x_{k,i}^+ - \bar{x}_i)^2 + \frac{1}{n^- - 1} \sum_{k=1}^{n^-} (x_{k,i}^- - \bar{x}_i)^2}, \quad (\text{C.1})$$

where  $\bar{x}_i$ ,  $\bar{x}_i^+$  and  $\bar{x}_i^-$  represent the average feature value over samples of the full dataset:  $n^+$  samples belonging to the positive class and  $n^-$  samples belonging to the negative class respectively. The numerator in the above equation represents the inter-class distance or separability, while the denominator computes the intra-class variance. Thus, a feature with a smaller spread within each of the two classes and a higher separation between the two class means would generate a higher F-score.

For all the features in the SHARPs-plus-geometry and SHARPs-plus-topology combination feature sets from the geometry and the topology experiments, we compute the normalized

F-score scaled with respect to the highest scoring feature. We use the `f_classif` functionality in the Python Scikit-Learn package (Pedregosa et al., 2011) to compute the F-score, and report the top 15 ranking features. The results are shown in Figure C.1. In all experiments, the top three ranking features – total photospheric energy density (TOTPOT), total unsigned flux (USFLUX) and the total area of the active pixels (AREA\_ACR) – belong to the SHARPs feature set. In the geometry-based experiments (Fig. C.1a), all but one of the top 15 properties belong to the geometry feature set. The higher-ranking of these pertain to the magnetic flux: the largest area (`max_pos_flux`, `max_neg_flux`) and the polarities belonging to the MIP (`IF_pos_flux`, `IF_neg_flux`), followed by the areas of the respective polarities (geometry features with the “\_area” suffix). In the topology-based experiments (Fig. C.1b), the feature ranking is more mixed. Six of the top 15 features belong to the topology feature set: the number of live holes at the magnetic flux values  $\pm 263.16$  G (`pos_intvl_1`, `neg_intvl_1`),  $\pm 789.47$  G (`pos_intvl_2`, `neg_intvl_2`) and  $\pm 1315.79$  G (`pos_intvl_3`, `neg_intvl_3`). In terms of univariate feature ranking, the geometry (nine of the top 15 features) and topology features (six of the top 15) perform well individually when compared with SHARPs features.

These scores demonstrate the predictive power of *individual* features in terms of their ability to discriminate the flaring and non-flaring datasets. However, this cannot be directly correlated with the TSS scores of the different feature sets described above. The Fisher score is a univariate feature ranking method and does not take into account the correlation between the different features. We leave this investigation of the correlation within and between the various feature sets as future work.

**Cite this article as:** Deshmukh V, Berger TE, Bradley E & Meiss JD. 2020. Leveraging the mathematics of shape for solar magnetic eruption prediction. *J. Space Weather Space Clim.* 10, 13.