Leveraging the Mathematics of Shape for Solar Magnetic Eruption Prediction

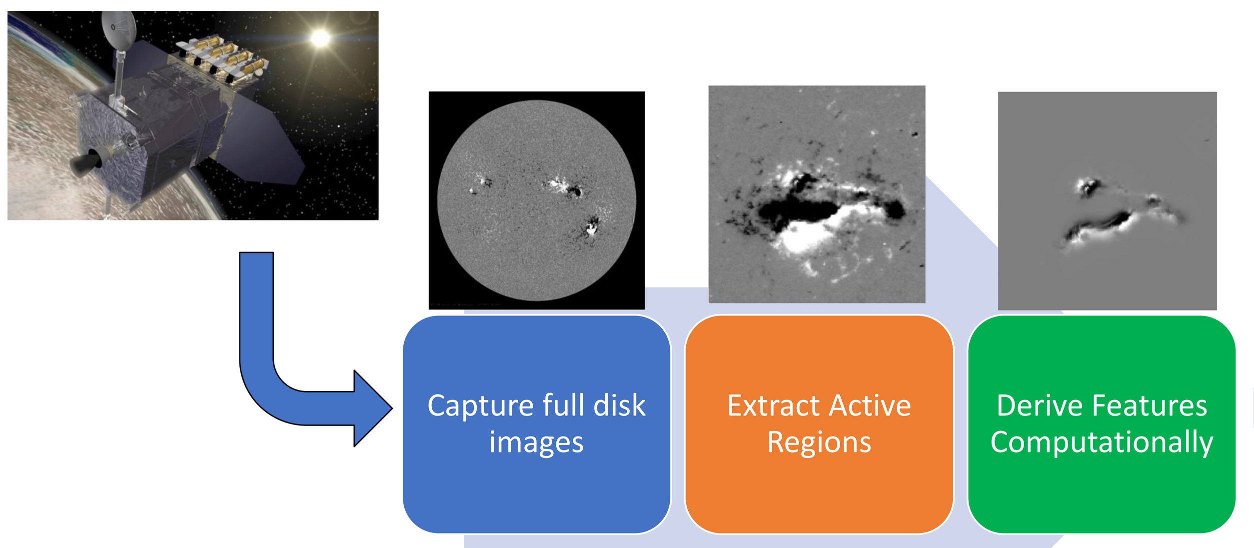
Varad Deshmukh[1], Thomas Berger[2], Elizabeth Bradley[1,3], and James Meiss[4]

1. University of Colorado (CU) Department of Computer Science. 2. University of Colorado (CU) Space Weather Technology, Research, and Education Center (SWx TREC). 3. The Santa Fe Institute. 4. University of Colorado (CU) Department of Applied Mathematics.

Grand Challenge University of Colorado Boulder Space Weather Center

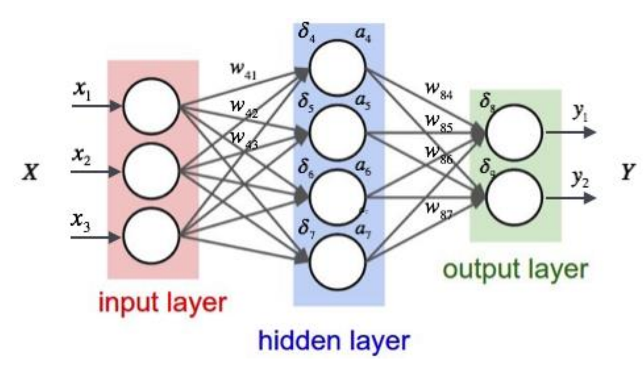
SWx TREC

A Novel Data Driven Approach to Solar Flare Prediction



Usually, physics-based properties of an active region are used as feature inputs to Machine Learning Models.

The magnetic field of the Sun is measured using in-orbit instruments such as the Helioseismic and Magnetic Imager (HMI) and processed further to extract *active regions,* where the magnetic field intensity is high.

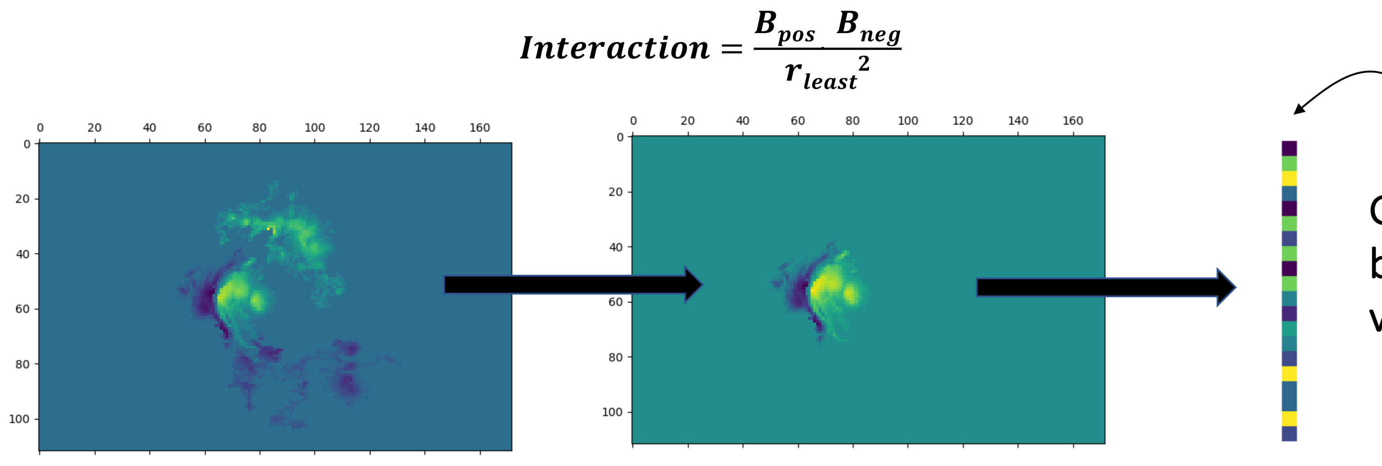


We propose a feature engineering approach based on *formalizing the shape of sunspots* for solar flare prediction using Topological Data Analysis and Computational Geometry.

Geometry-based Features

Geometry-based features study the relationship between the positive and negative polarities in an active region, based on their magnetic flux and the distance between them. We propose an interaction factor, inspired from Ising Energy, and derive various properties with most interacting pair.

$$Interaction = \frac{B_{pos}\ B_{neg}}{r_{least}^{2}}$$

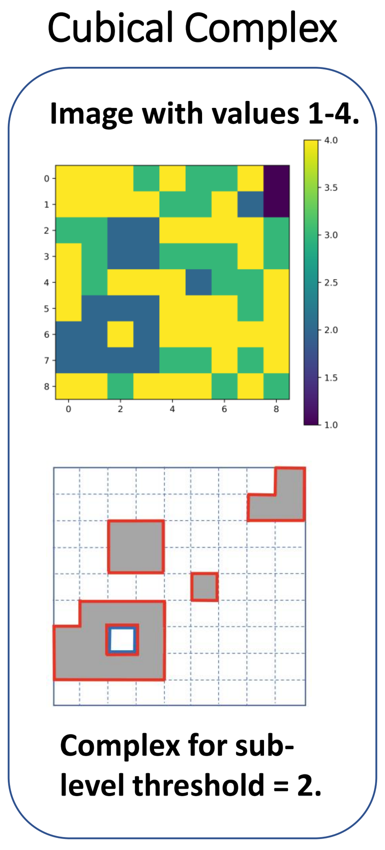


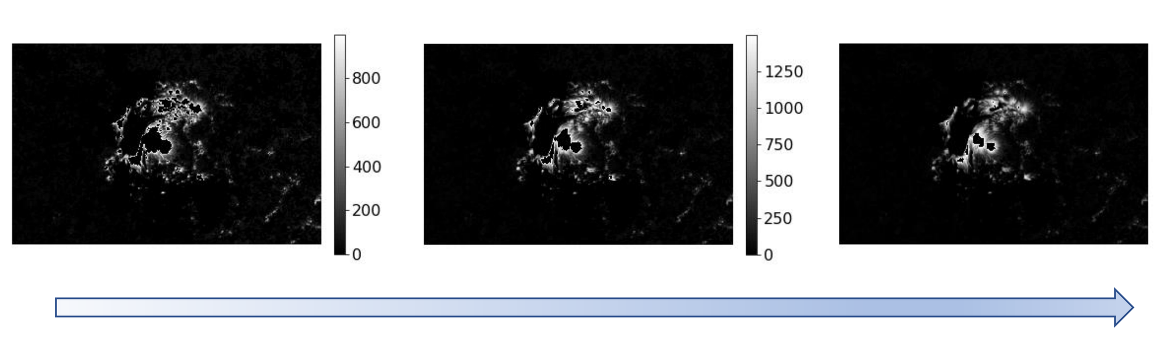
Active Region | Most Interacting Pair (MIP) | Feature vector

Topology-based Features

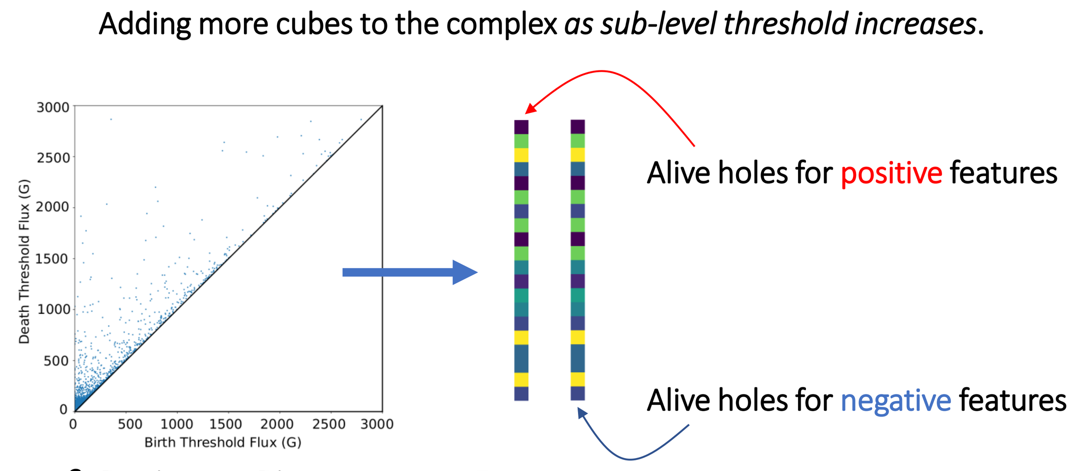
Topological Data Analysis is a means to quantify the shape of an object mathematically. Topology of an object is defined by the number of k-dimensional holes called Betti numbers (β_k).
For pixelated data (like magnetograms), Persistent Homology computes β_k at different scales – in this case, the sub-level magnetic flux threshold, by connecting pixels using a cubical complex.

Cubical Complex





Adding more cubes to the complex *as sub-level threshold increases.*

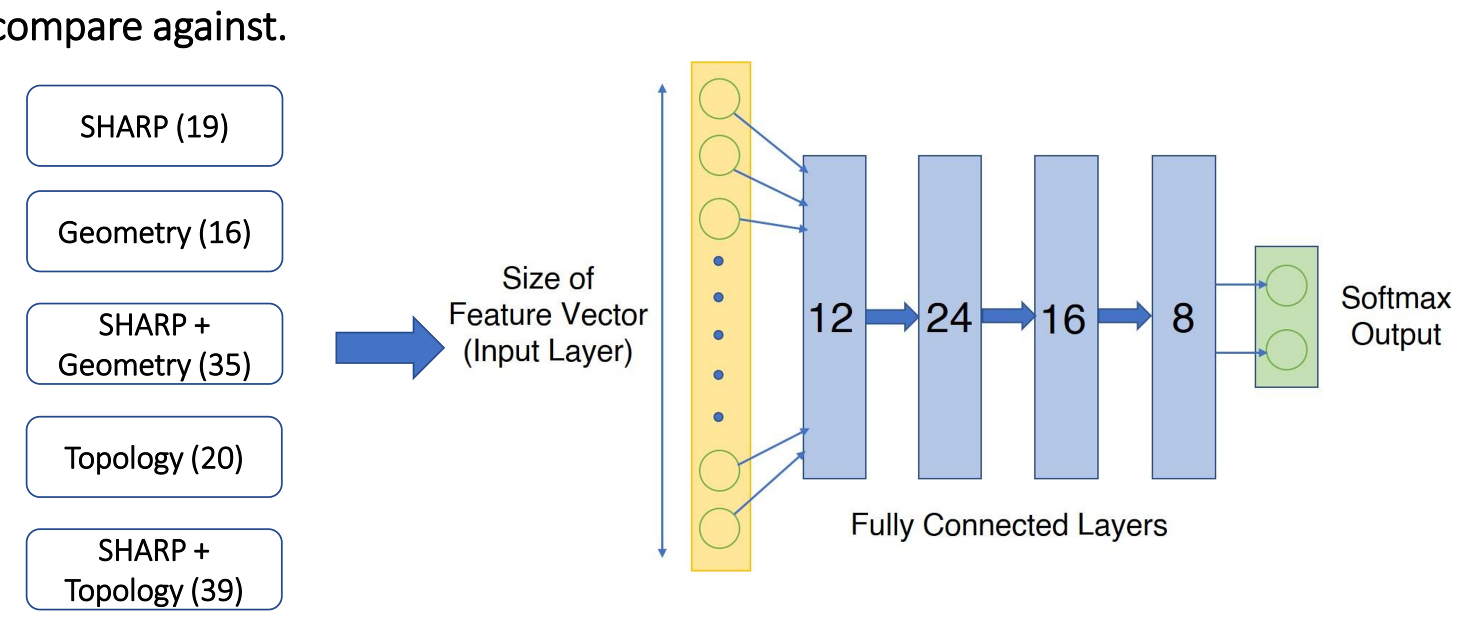


β_1 Persistence Diagram | Feature vector

Persistence Diagram is a way to track the formation and disappearance of β_k features for different sub-level thresholds. We vectorize the PD to generate features for ML models.

Deep Learning Model

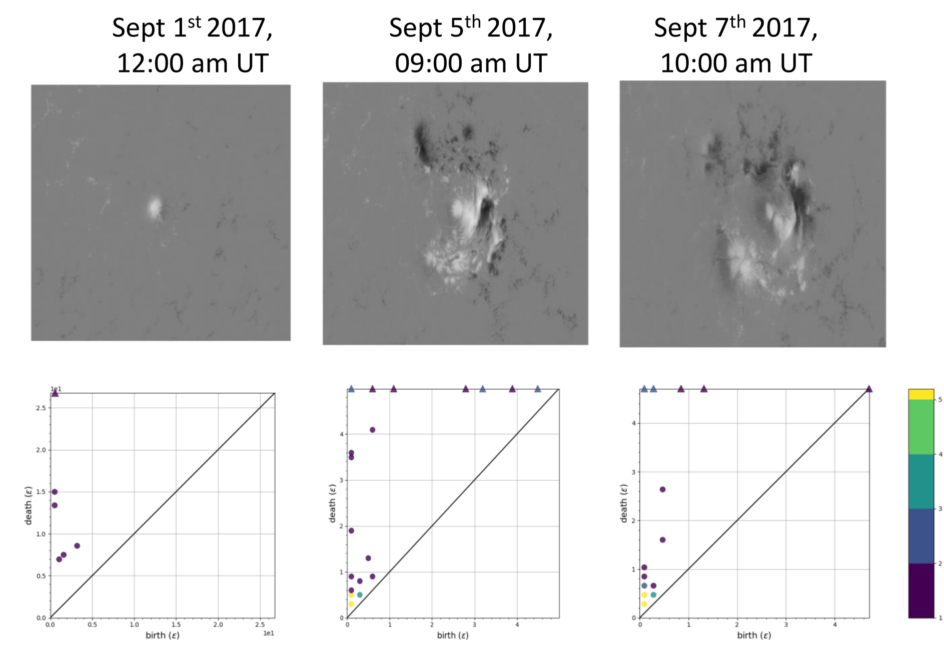
We use a 6-layer Multilayer Perceptron model to evaluate various feature sets. The magnetic flux properties available in the metadata of the SDO HMI SHARP images are used as reference features to compare against.



Feature Sets | Multilayer Perceptron Model

Topological Evolution of Sunspots

Persistence diagrams of positive clusters show increased number of components before the flare occurrence



Temporal evolution of NOAA 12673 active region. An X2.2 flare occurred on Sept 6th at 9:00 am, followed by an X9.3 flare at 12:00 pm.

Experimental Setup

Data
SDO HMI Magnetograms between 2010-2016.

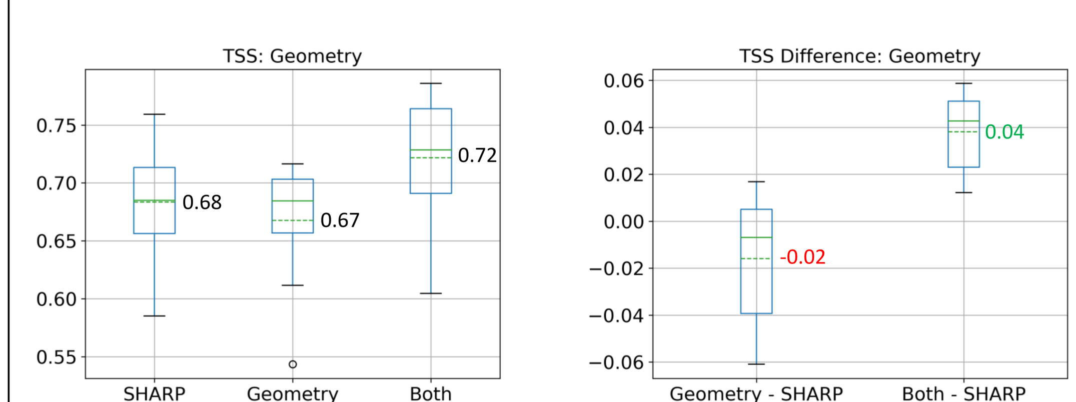
Binary Classification
Label M+ class flares in the next 24 hours as 1, else 0.

Data Shuffling
Random 70/30 splitting based on active region*.

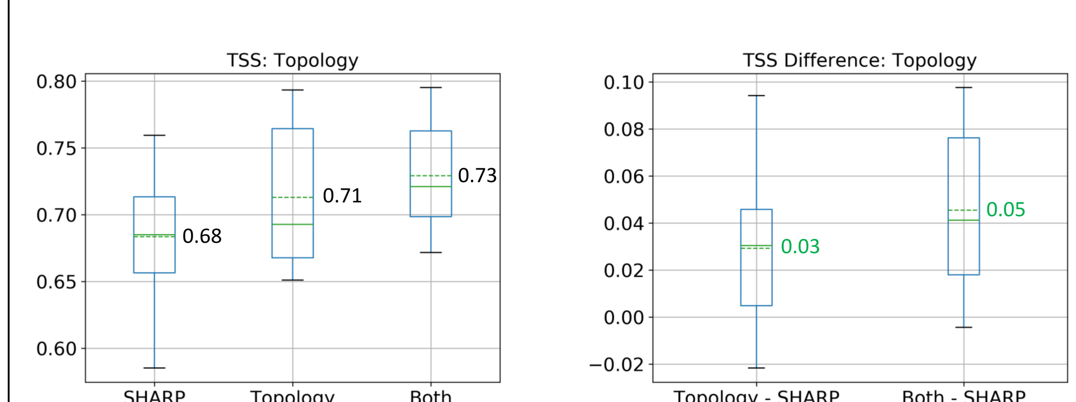
Performance Metric
$$TSS = \frac{TP}{TP + FN} - \frac{FP}{FP + TN}$$

Iterations
10, with different random seeds for shuffling.

Results



While the geometry-based features slightly underperform the SHARP-based features, the combination of both outperforms SHARP-only by 0.04.



The topology-based features outperform the SHARP-based ones by 0.03, while the combination outperforms by 0.05.

Conclusions and Next Steps

1. Complexity of a sunspot, as quantified with TDA and computational geometry, improves flare forecasting accuracy.
2. Machine-learning algorithms can leverage these derived quantities as input features, and determine their relevance to flare occurrence.
3. Improved variations of the geometry and topological features in terms of better interaction metrics and PD vectorization techniques may improve model accuracy.
4. The *temporal evolution* of these features may also be leveraged for flare prediction using CNNs on temporally consecutive feature vector stacks and LSTM models.

References

1. Scherrer, P. H., J. Schou, R. I. Bush, A. G. Kosovichev, R. S. Bogart, et al., 2011. The Helioseismic and Magnetic Imager (HMI) Investigation for the Solar Dynamics Observatory (SDO). Solar Physics, 275(1-619 2), 207–227. 10.1007/s11207-011-9834-2.
2. Florios, K., I. Kontogiannis, S.-H. Park, J. A. Guerra, F. Benvenuto, D. S. Bloomfield, and M. K. Georgoulis, 2018. Forecasting Solar Flares Using Magnetogram-based Predictors and Machine Learning. Sol. Phys., 293(2), 28. 10.1007/s11207-018-1250-4.
3. Zomorodian, A. & Carlsson, G. Discrete Comput Geom (2005) 33: 249. https://doi.org/10.1007/s00454-004-1146-y.
4. Knyazeva, I. S., N. G. Makarenko, and M. K. Livshits, 2011. Detection of New Emerging Magnetic Flux from the Topology of SOHO/MDI Magnetograms. Astronomy Reports, 55(5), 463. 10.1134/S1063772911050040.