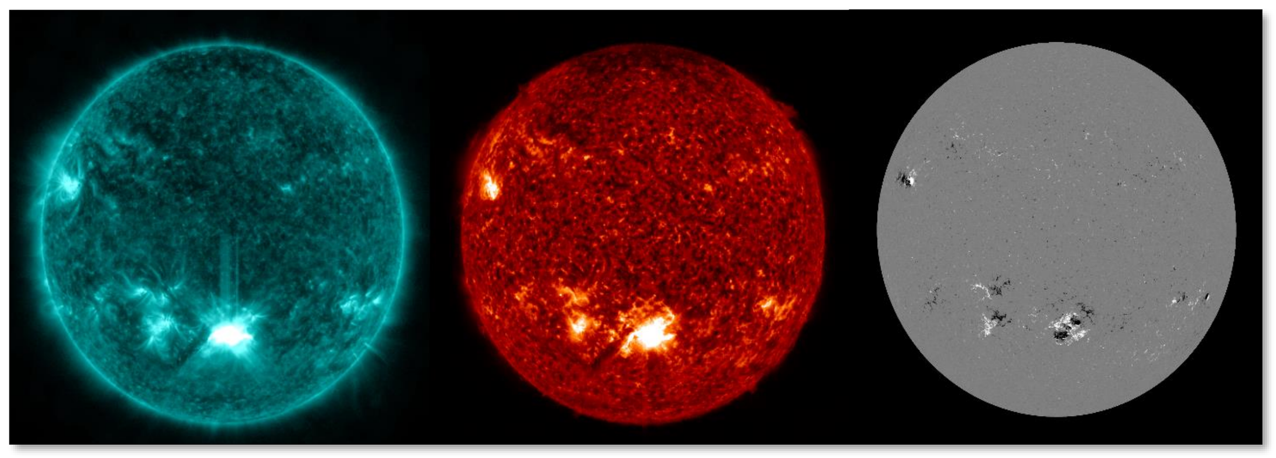
A comparative study of supervised machine learning algorithms to forecast solar flares

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

[1] Center of Excellence in Space Sciences India, IISER Kolkata. [2] Department of Physical Sciences, IISER Kolkata. [3] Max Planck Institute for Solar System Research. [4] Nordita, KTH Royal Institute of Technology and Stockholm University. [5] KTH Royal Institute of Technology, Stockholm. [6] Indian Institute of Technology, Kharagpur. [7] Udaipur Solar Observatory, PRL. [8] Department of Astronomy, AlbaNova University Center, Stockholm University

Introduction

Solar flares are sudden bursts of electromagnetic radiation observed in the solar corona. These flares are typically manifest as rapid enhancements in X-ray and EUV wavelengths.



There are many mechanisms that have been proposed as drivers of solar flares. Almost all of which involve magnetic characteristics of the underlying active region.

These highly energetic events can directly impact human society by damaging our space-borne technologies. Therefore, it is very important to predict and forecast solar flares to take preventive actions to mitigate socioeconomic losses.

In this study, our goal is to make a reliable and accurate Machine Learning (ML) model which can predict an imminent flare by observing the magnetic field properties of the associated active region.

The following four popular Machine Learning (ML) algorithms are used in this study.

1. K-nearest neighbor (KNN)
2. Random Forest Classifier (RFC)
3. Logistic Regression (LR)
4. Support Vector Machine (SVM)

Data Preparation

We use Hinode XRT and GOES flare catalog to get the information of all solar flares and their associated active regions that have appeared between May 2010 and December 2020. We only use those events which have entries in both Hinode and GOES flare catalog.

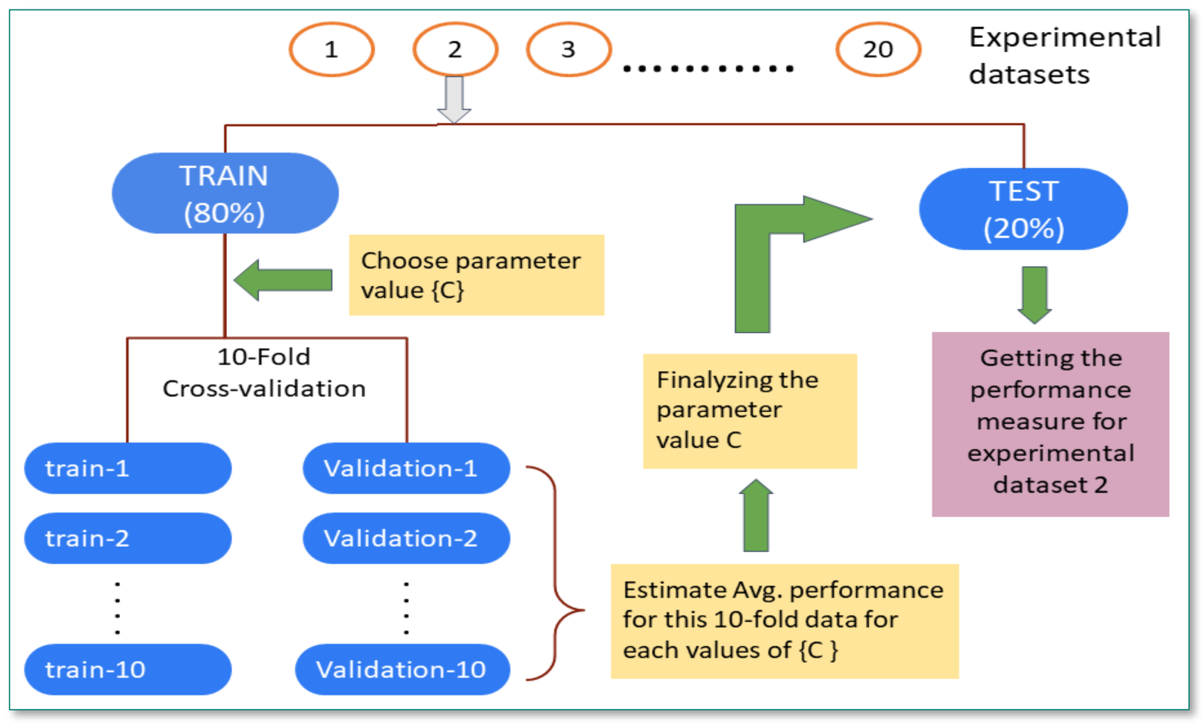
Depending on the flare intensities we divide the active regions into two groups: **flaring** and **non-flaring**. If an AR has ever produced an M or higher intensity flare, it falls in the flaring group whereas, flare intensities below the C class are considered as a non-flaring group.

We have used the Spaceweather HMI Active Region Patch (SHARP) data series to analyze the vector magnetic field properties of the active regions. (Bobra et al. 2015). For flaring class magnetic features are extracted from a SHARP patch captured **24hr** before the flare peak-time. This allows us to predict in advance if a region is going to flare or not.

Our final dataset consists of **503** flaring active regions and **3358** unique non-flaring active regions.

Methodology and input parameters

This schematic diagram demonstrate our method of analysis. We divide our dataset into two part: training and testing in the ratio of 4:1 respectively.



We use these following 14 magnetic features to capture the characteristics of any active region.

No. of features	Description	Keyword
1	Total unsigned current helicity	TOTUSJH
2	Total magnitude of Lorentz force	TOTBSQ
3	Total photospheric magnetic free energy density	TOTPOT
4	Total unsigned vertical current	TOTUSJZ
5	Absolute value of the net current helicity	ABSNJZH
6	Sum of the modulus of the net current per polarity	SAVNCPP
7	Total unsigned flux	USFLUX
8	Area of strong field pixels in the active region	AREA_ACR
9	Sum of z-component of Lorentz force	TOTFZ
10	Mean photospheric magnetic free energy	MEANPOT
11	Sum of flux near polarity inversion line	R_VALUE
12	Fraction of Area with shear >45°	SHRGT45
13	Total absolute twist calculated over strong field regions	TOTABSTWIST
14	Sum of z-component of normalized Lorentz force	EPSZ

Performance Measure

Confusion Matrix	Actual Positive/Flaring events	Actual negative/Non-flaring instance
Predicted Positive/Flaring events	True Positive (TP)	False Positive (FP)
Predicted negative/Non-flaring events	False Negative (FN)	True Negative (TN)

To Assess model performance following performance metrics are calculated from the elements of confusion matrix.

$$\textbf{True Skill Score (TSS)} = \frac{TP}{TP+FN} - \frac{FP}{TN+FP}$$

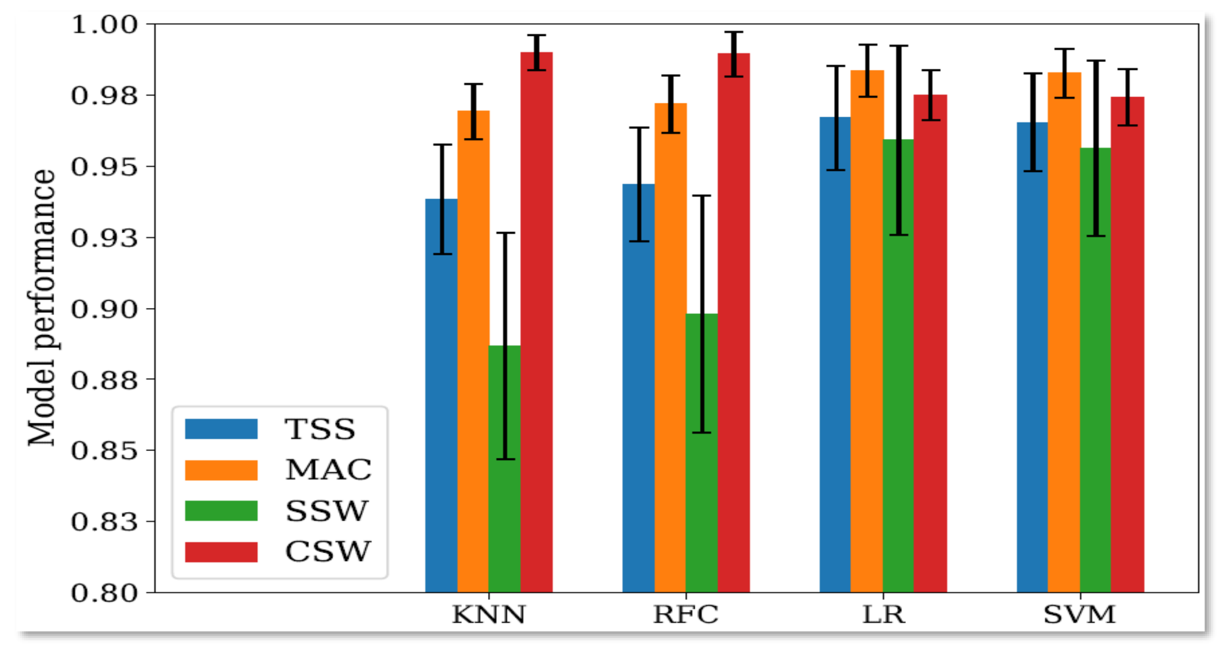
$$\textbf{Macro Accuracy (MAC)} = \frac{1}{2}\left(\frac{TP}{TP+FN} + \frac{TN}{TN+FP}\right)$$

$$\textbf{*Severe Space Weather (SSW)} = \frac{TP-FN}{TP+FN}$$

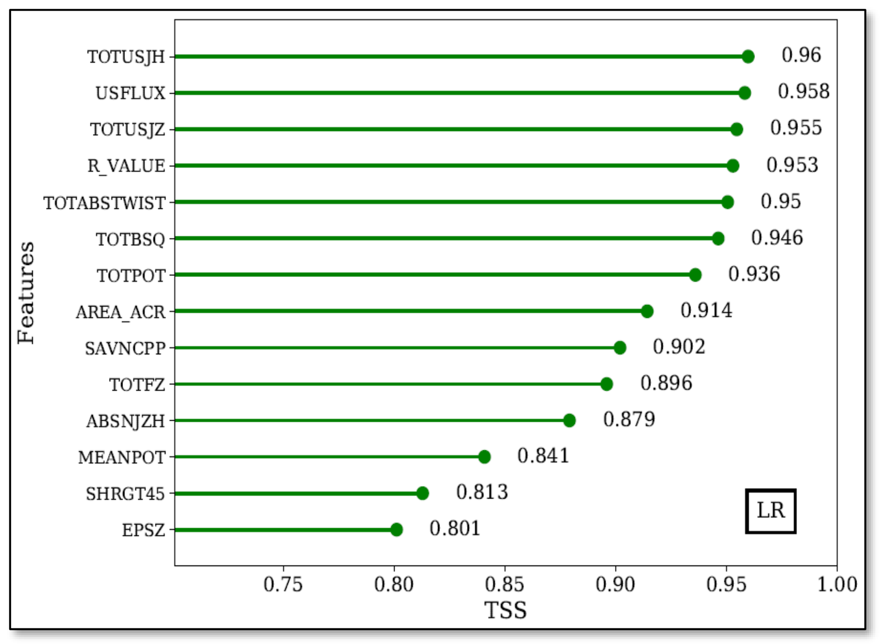
$$\textbf{*Clear Space Weather (CSW)} = \frac{TN-FP}{TN+FP}$$

Results

The following plot graphically demonstrates the performance of all four ML models used in this study.



	TSS	MAC	SSW	CSW
KNN	0.938	0.969	0.887	0.990
RFC	0.944	0.972	0.898	0.989
LR	0.967	0.983	0.959	0.975
SVM	0.965	0.983	0.956	0.974



To know which magnetic feature has a better flare-identification ability, we train our model for each magnetic feature separately and rank them according to their individual TSS score. The plot on the left represents the feature ranking for the LR model which gives the highest TSS score among the four ML models

Conclusions

Our study shows that all four ML models have performed reasonably. LR and SVM deliver comparatively better performance than KNN and RFC.

We have achieved a very high TSS score of 0.967 with our LR model, which is higher than most of the early flare prediction attempts with a parameter-based approach.

We rank the input magnetic features according to their ability of identifying flaring regions and find that current helicity, unsigned flux, vertical current density, R-value, magnetic twist, etc. are useful parameters.

Two new performance indices SSW and CSW are very useful to critically judge the model performance in different space weather scenarios. Both the LR and SVM model has a high SSW and CSW indicating that these models will perform well irrespective of active or quite space weather scenario.

Acknowledgements

The Center of Excellence in Space Sciences India at IISER Kolkata is funded by the Ministry of Education, Government of India. The authors acknowledge funding from the Ministry of Education SPARC Project grant SPARC/2018-2019/P746/SL to IISER Kolkata, NORDITA/Stockholm University, KTH Royal Institute of Technology, and IIT Kharagpur.