Physics-driven feature combination for an explainable AI approach to flare forecasting

M. Lampani[1], S. Guastavino[1, 2], M. Piana[1, 2], F. Benvenuto[1], A. M. Massone[1]

1. **MIDA | Università di Genova**, DIMA – Dipartimento di Matematica, Genova (IT), 2. **INAF | Istituto Nazionale di Astrofisica**, Osservatorio Astrofisico di Torino

Introduction

Feature-based machine learning techniques for solar flare forecasting rely on features, representing physical parameters, that are extracted from magnetograms, as from Helioseismic and Magnetic Imager (HMI) images.
The most common approach consists in standardizing such features in order to make them adimensional and applying machine learning techniques to predict the occurrence of solar flares. With this standard approach the machine learning model does not take into account the physical nature of features.
We propose a physics-aware machine learning approach with the aim of constructing physical explainable models: the approach is based on first creating combinations of features accordingly to their dimension driven by the plasma equations and then on applying machine learning techniques on such new combinations in order to explore which ones lead to a more predictive model.

Linear LASSO Regression vs Physics-Driven Regression

In the context of linear regression, we encounter a set of N samples $\{(x_i, y_i)\}_{i=1}^{N}$, where each $x_i = (x_{i1} \dots, x_{ip})$ denotes a p-dimensional vector of features and $y_i$represents the associated response. The objective is to find an estimator function $\hat{g}: \mathcal{X} \to \mathcal{Y}$ that approximates the input-output relation, where $\mathrm{g}(x) = \langle x|\beta\rangle = x^T\beta \; with \; \beta \in \mathbb{R}^p$ signifies the regression weights vector to be estimated.

The Least Absolute Shrinkage and Selection Operator (**LASSO**) method seeks the solution $\hat{\beta}$ to the optimization problem

$$\min_{\beta\in\mathbb{R}^p}\left\{\frac{1}{N}||\boldsymbol{y}-\boldsymbol{X}\beta||_2^2+\lambda|\beta|_1\right\}$$

where $\lambda \geq 0$ represents the regularization parameter and $\beta = (\beta_1, \dots, \beta_p)^T \in \mathbb{R}^p$ signifies the regression weights vector (Hastie et al. (2009)). It is crucial to standardize features in this process involving the summation of various predictors $\boldsymbol{X}$, ensuring they are dimensionless.

Standardization ensures that the choice of units for predictor measurements does not impact the solutions. It is achieved via the following formula, making **the final result is dimensionless**:

$$\tilde{x} = \frac{x_i - \mu_i}{\sigma_i}$$

where x_i denotes the single predictor being standardized, μ_i signifies its mean, σ_i represents its standard deviation, and $\tilde{x}_i$ represents the standardized predictor.

If we consider a set of predictors $x = (m, v, E)$, where m represents mass, v signifies velocity, and E represents energy, the estimator $\mathrm{g}(\tilde{x})$ computation is as follows:

$$\mathrm{g}(\tilde{x}) = \langle\beta|\tilde{x}\rangle_{\mathbb{R}^3} = \beta_1\tilde{m} + \beta_2\tilde{v} + \beta_3\tilde{E}$$

where the features were earlier standardized to render them dimensionless in order to allow the sum.

However, **our approach diverges from this common practice by focusing on the physical units of predictors** instead of disregarding them. The fundamental concept involves leveraging the LASSO to construct an estimator that utilizes non-standardized features, combining them in a manner that shares the same physical unit.
For instance, if we opt to work solely with energies among the predictor set detailed above, **we construct a feature map $\boldsymbol{\psi(x)}$ that accounts for the physical units.**

In this scenario, $\psi: \mathbb{R}^3 \to \mathbb{R}^2$, where $\psi(x) = (m \cdot v^2, E)$ and the computation of $g(x)$ becomes:

$$\mathrm{g}(x) = \langle\beta|\psi(x)\rangle_{\mathbb{R}^2} = \beta_1(m \cdot v^2) + \beta_2 E$$

This approach enables us to consider the inherent physical properties of our features without encountering issues during the summation process.

Feature Selection and Physics-Driven Analysis

As predictors we included features highlighted in two previous works:
- (Bobra et al. (2015)) where **SHARP features** with the highest F-score were selected.
- Campi et al. (2019) where **features from the FLARECAST project** (Georgoulis et al. (2021)) were ranked for importance using LASSO and Random Forest.

Name of list	Units	Features
$\mathcal{L}_{\text{current}}$	A	TOTUSJZ SAVNCPP sharp_kw/snetjzpp/total
$\mathcal{L}_{\text{force}}$	GAm	TOTPOT
$\mathcal{L}_{\text{current-helicity}}$	$\mathrm{G^2m^{-1}}$	TOTUSJH ABSNJZH
$\mathcal{L}_{\text{flux}}$	$\mathrm{Gm^2}$	USFLUX R_VALUE helicity_energy_bvec/abs_tot_dedt_in
$\mathcal{L}_{\text{area}}$	$\mathrm{m^2}$	AREA_ACR SHRGT45 sharp_kw/ggt45fract
$\mathcal{L}_{\text{energy-density}}$	$\mathrm{GAm^{-1}}$	MEANPOT sharp_kw/rho/total
$\mathcal{L}_{\text{gradient-magnetic-field}}$	$\mathrm{Gm^{-1}}$	MEANGBZ sharp_kw/hgradbh/total
$\mathcal{L}_{\text{magnetic-field}}$	G	wlsg_br/value_int wlsg_blos/value_int
$\mathcal{L}_{\text{length}}$	m	mpil_br/max_length

Table 1: Lists of features involved in our study with units of measurement, distinguishing SHARP elements in capital letters from FLARECAST features.

The main idea of this study is to construct a new coherent set of features obtained by suitably combining the original predictors in a such way that they share the same units.

We performed the following steps:

- we made a dimensional analysis of the selected features (Table 1) and we observed that all of the predictors could be described using different combinations of the same fundamental units: gauss (G), amperes (A), and meters (m)
- we divided the features in lists accordingly to the different units (Table 1)
- **we defined admissible operations between the original features that provide predictors in a new shared unit**. Given that the fundamental units used to describe the predictors could be combined in various ways to describe all of our predictors, we adopted a unit based on their product: $\boldsymbol{GAm^2}$**, which is the physical unit of energy**. These operations (Table 2) involve the multiplication and division of features. When a squared unit is present, we either allow the feature to be multiplied by itself or to be multiplied by a different feature that share the same units. To handle such cases, we introduced different versions of the same operation (ex: op6, op6a, op6b, op6c).

By utilizing this approach, **we were able to place a significant emphasis on the physical nature of our data.**

Involved lists	Units	Operations (new features)
$B \in \mathcal{L}_{magnetic-field}$ $I \in \mathcal{L}_{current}$ $A \in \mathcal{L}_{area}$	$\mathrm{G \cdot A \cdot m^2 = GAm^2}$	$\mathrm{op1} = B \cdot I \cdot A$
$\Phi \in \mathcal{L}_{flux}$ $I \in \mathcal{L}_{current}$	$\mathrm{Mx \cdot A = 10^{-4}\ GAm^2}$	$\mathrm{op2} = \Phi \cdot I$
$F \in \mathcal{L}_{force}$ $\nabla B \in \mathcal{L}_{gradient}$ $B \in \mathcal{L}_{magnetic-field}$ $A \in \mathcal{L}_{area}$	$\mathrm{erg\ cm^{-1} \cdot GMm^{-1}/G \cdot m^2 = 10^{-7}\ GAm^2}$	$\mathrm{op3} = F \cdot \nabla B / B \cdot A$
$F \in \mathcal{L}_{force}$ $B \in \mathcal{L}_{magnetic-field}$ $H \in \mathcal{L}_{current-helicity}$	$\mathrm{erg\ cm^{-1} \cdot G^2\ /\ Gm^{-1} = 10^{-1}\ GAm^2}$	$\mathrm{op4} = F \cdot B^2 / H$
$F \in \mathcal{L}_{force}$ $B_1, B_2 \in \mathcal{L}_{magnetic-field}$ $H \in \mathcal{L}_{current-helicity}$	$\mathrm{erg\ cm^{-1} \cdot G \cdot G\ /\ G^2m^{-1} = 10^{-1}\ GAm^2}$	$\mathrm{op4a} = F \cdot B_1 \cdot B_2 / H$
$H \in \mathcal{L}_{current-helicity}$ $\nabla B \in \mathcal{L}_{gradient-mf}$ $I \in \mathcal{L}_{current}$ $A \in \mathcal{L}_{area}$	$\mathrm{G^2m^{-1}\ /\ GMm^{-1} \cdot A \cdot m^2 = 10^6\ GAm^2}$	$\mathrm{op5} = H / \nabla B \cdot I \cdot A$
$H \in \mathcal{L}_{current-helicity}$ $F \in \mathcal{L}_{force}$ $I_1, I_2 \in \mathcal{L}_{current}$ $A_1, A_2 \in \mathcal{L}_{area}$	$\mathrm{G^2m^{-1}\ /\ erg\ cm^{-1} \cdot A \cdot A \cdot m^2 \cdot m^2 = 10\ GAm^2}$	$\mathrm{op6} = H / F \cdot I_1 \cdot I_2 \cdot A_1 \cdot A_2$
$H \in \mathcal{L}_{current-helicity}$ $F \in \mathcal{L}_{force}$ $I \in \mathcal{L}_{current}$ $A_1, A_2 \in \mathcal{L}_{area}$	$\mathrm{G^2m^{-1}\ /\ erg\ cm^{-1} \cdot A^2 \cdot m^2 \cdot m^2 = 10\ GAm^2}$	$\mathrm{op6a} = H / F \cdot I^2 \cdot A_1 \cdot A_2$
$H \in \mathcal{L}_{current-helicity}$ $F \in \mathcal{L}_{force}$ $I_1, I_2 \in \mathcal{L}_{current}$ $A \in \mathcal{L}_{area}$	$\mathrm{G^2m^{-1}\ /\ erg\ cm^{-1} \cdot A \cdot A \cdot (m^2)^2 = 10\ GAm^2}$	$\mathrm{op6b} = H / F \cdot I_1 \cdot I_2 \cdot A^2$
$H \in \mathcal{L}_{current-helicity}$ $F \in \mathcal{L}_{force}$ $I \in \mathcal{L}_{current}$ $A \in \mathcal{L}_{area}$	$\mathrm{G^2m^{-1}\ /\ erg\ cm^{-1} \cdot A^2 \cdot (m^2)^2 = 10\ GAm^2}$	$\mathrm{op6c} = H / F \cdot I^2 \cdot A^2$
$\nabla B \in \mathcal{L}_{gradient}$ $l \in \mathcal{L}_{length}$ $I \in \mathcal{L}_{current}$ $A \in \mathcal{L}_{area}$	$\mathrm{GMm^{-1} \cdot Mm \cdot A \cdot m^2 = GAm^2}$	$\mathrm{op7} = \nabla B \cdot l \cdot I \cdot A$
$\rho \in \mathcal{L}_{energy-density}$ $l \in \mathcal{L}_{length}$ $A \in \mathcal{L}_{area}$	$\mathrm{erg\ cm^{-3} \cdot Mm^2 = 10^9\ GAm^2}$	$\mathrm{op8} = \rho \cdot l \cdot A$

Table 2: Lists of the features involved in our study with units of measurement

Results

1. For the **training and test set separations**, we opted for **two different approaches**:

- the former inspired by the **uniform dataset** created in (Guastavino et al. (2022)), which suggests avoiding chronological splitting of historical data into training and test sets due to the bias introduced by the cyclical nature of solar activity.

- the latter inspired by (van der Sande et al. (2023)) which uses **a cycle-agnostic temporal splitting**, specifically, for each year, our test set comprised samples from November and December, while the remaining data constituted our training set.

To validate our physics-driven study, we conducted a comparison between our results and those obtained through LASSO with Recursive Feature Elimination (RFE) on the same features utilized in creating our operations. However, these features were not combined and were standardized beforehand, allowing us to contrast our outcomes with the conventional approach.

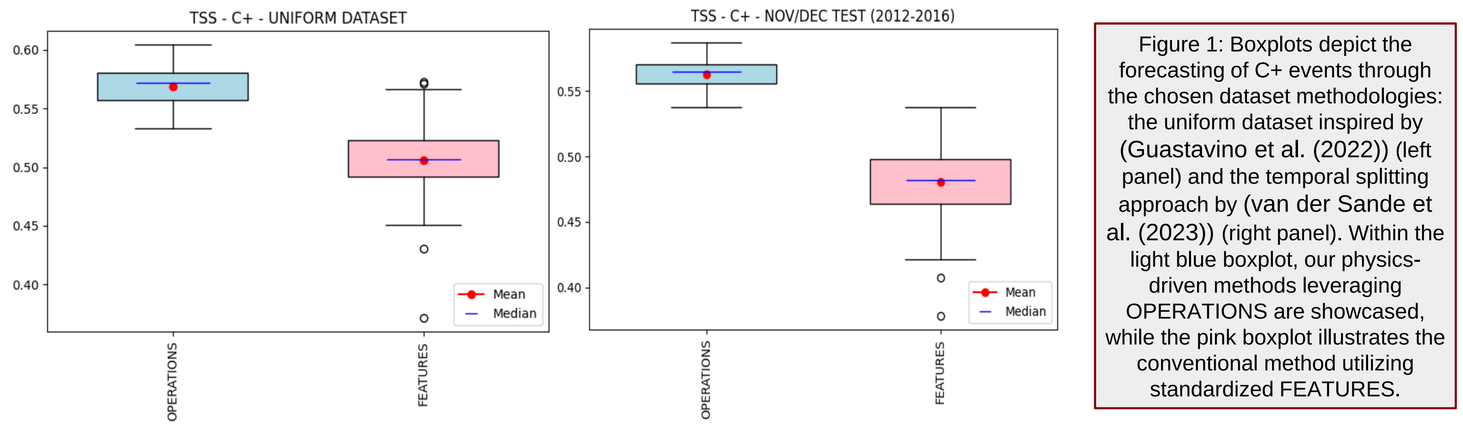

Figure 1: Boxplots depict the forecasting of C+ events through the chosen dataset methodologies: the uniform dataset inspired by (Guastavino et al. (2022)) (left panel) and the temporal splitting approach by (van der Sande et al. (2023)) (right panel). Within the light blue boxplot, our physics-driven methods leveraging OPERATIONS are showcased, while the pink boxplot illustrates the conventional method utilizing standardized FEATURES.

2. Analysis of Figure 1 demonstrates our **physics driven method's** (OPERATIONS) **consistent superiority over the standardized feature approach** (FEATURES). It not only delivers better results consistently but also showcases **greater stability**, as indicated by a narrower spread around the mean value. This emphasizes our method's accuracy in predicting solar flare events.

3. Another important outcome of our analysis was the **extraction of the most relevant and predicting operations** from those defined in Table 2. This was achieved **through the RFE algorithm** and the extraction was conducted using both test sets. The coherence among the extracted relevant operations is evident in Figure 2, where op8, op5, op7 and the various versions of op6 consistently emerge as the most significant operations in both histograms.

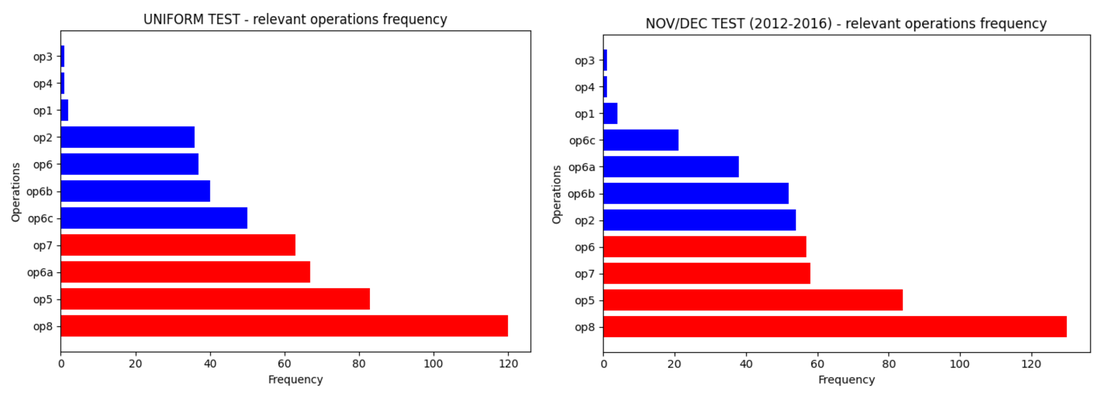

Figure 2: Comparing RFE-identified operations: left depicts a histogram of operations via the uniform dataset method; right showcases operations extracted using temporal splitting. The most predictive operations are represented in red.

REFERENCES

- **Hastie** et al., The Elements of Statistical Learning, Springer, Vol. 2, (2009)
- **Bobra** et al., Solar flare prediction using SDO/HMI vector magnetic field data with a machine-learning algorithm, ApJ, Vol 798, Number 135 (2015)
- **Campi** et al., *Feature Ranking of Active Region Source Properties in Solar Flare Forecasting and the Uncompromised Stochasticity of Flare Occurrence*, ApJ, Vol. 883, Number 2 (2019).
- **Guastavino** et al., *Implementation paradigm for supervised flare forecasting studies: A deep learning application with video data*, Astronomy & Astrophysics, A&A 662, A105 (2022a)
- **van der Sande** et al., Probabilistic solar flare forecasting using historical magnetogram data, ApJ, Vol 955, Number 2 (2023).
- **Georgoulis** et al., The flare likelihood and region eruption forecasting (FLARECAST) project: flare forecasting in the big data & machine learning era. Journal of Space Weather and Space Climate, 2021, 11: 39.

Fondazione Compagnia di San Paolo

margherita.lampani@dima.unige.it
guastavino@dima.unige.it
piana@dima.unige.it

MIDA | UniGe

benvenuto@dima.unige.it
massone@dima.unige.it