Enhancing solar flare prediction with physics-informed machine learning and dimensional analysis

S. Guastavino, F. Benvenuto, M. Lampani, M. Catalano, M. Piana, D. Del Moro, S. Chierichini, R. Mugatwala, D. Pedemonte, S. Scardigli, and A. Tacchino

Solar flares remain a central problem for both fundamental solar physics and operational space weather: they are the most energetic explosions in the heliosphere and a primary source of disturbances that can affect satellites, communications and ground infrastructure. Reliable forecasting depends on finding precursors that meaningfully reflect the magnetic energy available in active regions and the processes that can rapidly release it. Building on established magnetogram-based predictors, this work explores a physics-guided route to enrich feature-based machine learning for flare forecasting. Rather than treating every input as a dimensionless, standardized number, as is usually done in traditional machine learning pipelines, the method systematically constructs nonlinear, dimensionally consistent combinations of physically meaningful descriptors (e.g., electric current, magnetic flux, field strength and gradient, area, and energy density) so that each derived quantity has the same target dimension, here chosen as energy. These physics-informed features are then used as inputs to classical machine learning methods (as ridge regression, support vector machine, Lasso) and are selected through the use of ranking methods to identify which combinations carry the most predictive power.

The immediate value of this strategy is twofold. First, by designing features according to dimensional analysis and known physical relationships, the machine learning models operate on inputs that are physically coherent and thus more interpretable. Second, the ranking of these constructed physics-informed features acts as a diagnostic: when certain combinations consistently rank highly, they possibly highlight the underlying physical mechanisms or energetic descriptors that are likely to be relevant precursors.

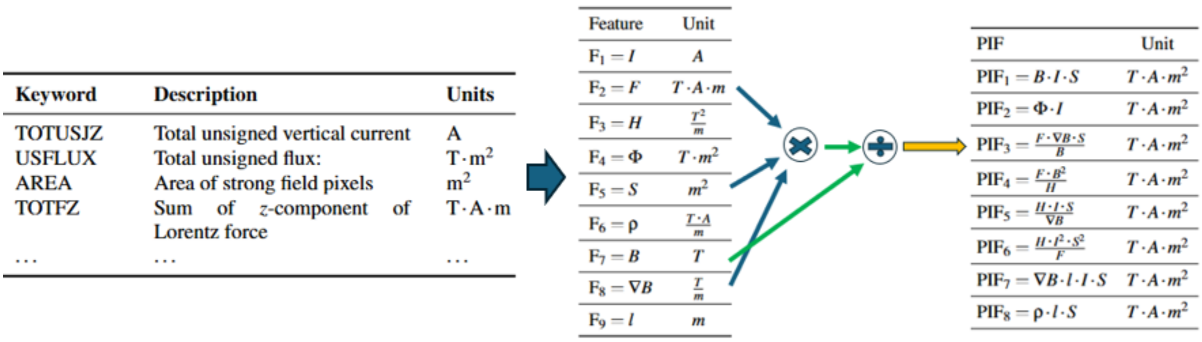

Figure 1: Left: example of Space-weather HMI Active Region Patch (SHARP) features that are considered in the study. Right: example of construction of the physics-informed features (PIFs) which share the same unit starting from a set of features that have different physical meaning.

We consider a portion of the archive from the Helioseismic and Magnetic Imager (HMI) on board the Solar Dynamics Observatory (SDO; magnetogram-derived descriptors spanning 2012–2017) and GOES-flare labels, and we show that introducing physics-informed features improved the model's ability to predict the occurrence of flares compared to using the same machine-learning pipeline with only conventional standardized features, Further, the ranking methods repeatedly identified the product of magnetic flux and electric current as the most influential physics-informed descriptor. This quantity aligns with physical intuition: within current-driven flare models it captures the energy released during a flare and, as such, is a promising candidate for assessing eruptive potential. The prominence of the magnetic flux-electric current product in the ranking therefore both supports the physical consistency of the approach and points to a concrete, interpretable precursor worth investigating further.