Decreasing False Alarm Rates in ML-based Solar Flare Prediction using SDO/HMI Data

Varad Deshmukh[1], Natasha Flyer[2], Kiera Van Der Sande[3], Thomas Berger[4*]

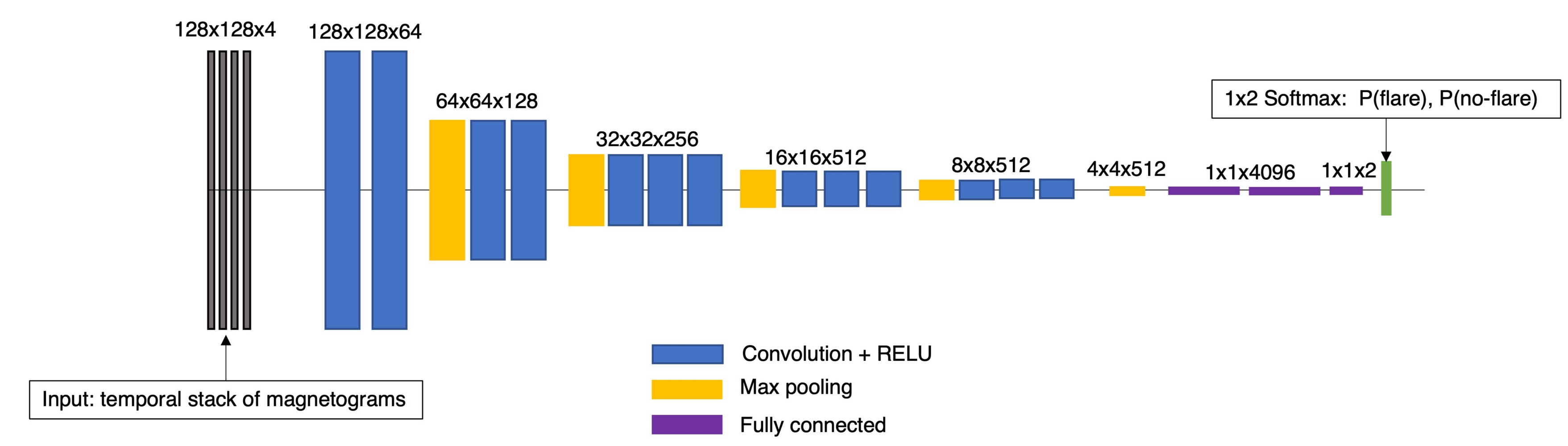
[1] University of Colorado at Boulder Department of Computer Science
[2] Flyer Research LLC
[3] University of Colorado at Boulder Department of Applied Mathematics
[4] University of Colorado at Boulder Space Weather Technology, Research, and Education Center (SWx TREC)
*Presenting author

Funding agencies

The Challenge of Dataset Imbalance in ML-based Solar Flare Prediction

- A primary challenge to ML-based solar flare prediction is *high dataset imbalance*.
- For a SDO/HMI[4] magnetogram dataset labeled flaring/non-flaring in the next 24 hours, approximately 99% of the samples are non-flaring.
- Evaluating these models is traditionally performed using metrics insensitive to dataset imbalance, such as the *True Skill Statistic (TSS).*
- However, deep learning models tuned for optimizing the TSS score on such imbalanced datasets *tend to be overforecasting* (i.e. produce false positives) and affects metrics like precision and HSS_2.
- ***To address overforecasting, we propose a two-stage novel architecture that combines VGG-16 --- a CNN-based deep learning model --- with an extremely randomized trees (ERT) model and tune it using a novel metric: TSS_{scaled}.***

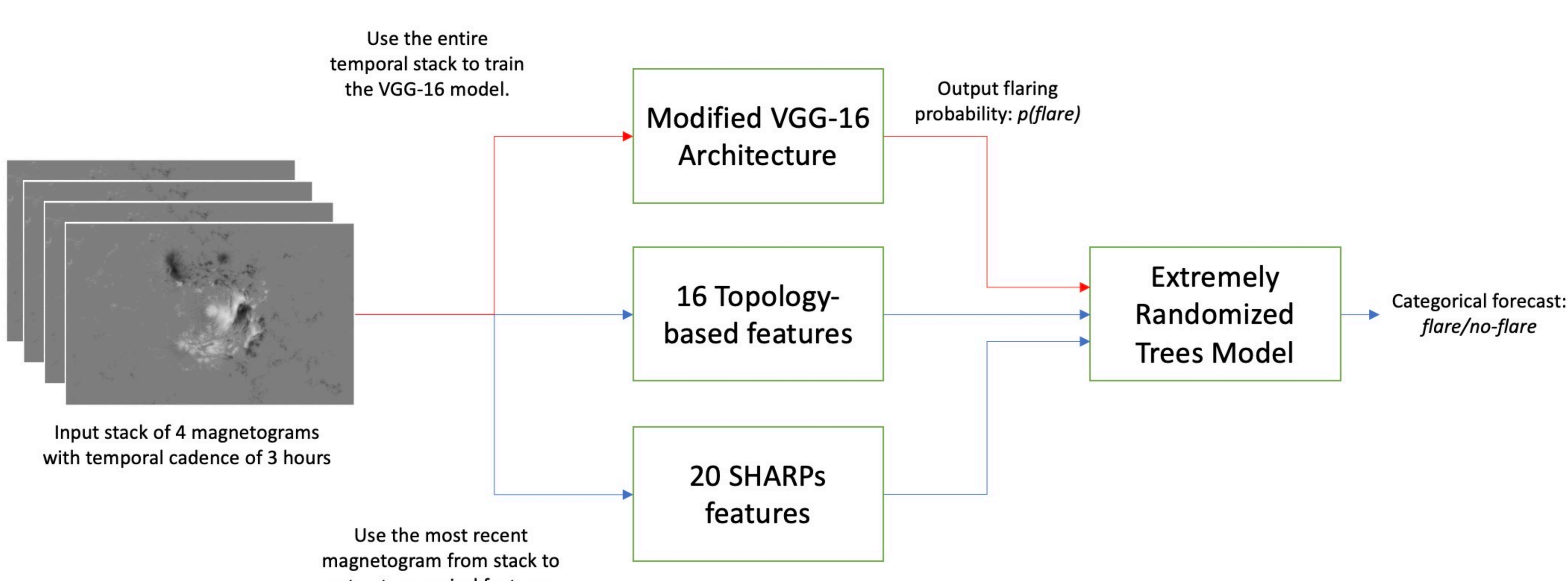
Exploring VGG-16 Variants for Solar Flare Prediction



Configuration	ROC AUC	PR AUC
C_1: $[B_r, B_\phi, B_\theta]$	0.967	0.43
C_2: B_r	0.965	0.43
C_3: B_r stack w/LSTM	0.975	0.43
C_4: B_r stack as channels	0.974	0.46

The first stage of our model is a VGG-16: a standard CNN-based architecture[4]. After experimenting with different input formats and architectural variants, we show that a VGG-16 trained on temporal stacks of B_r works best.

A Hybrid Two-stage Model: Combining CNN-extracted and Engineered Features



We design a two-stage model as follows:

1. **Stage 1** is the modified VGG-16 architecture trained on temporal stacks of magnetogram images, which outputs a flaring probability: *cnn_prob*.

2. **Stage 2** is an ERT model trained on three kinds of features: SHARPs[3], topological[2] and cnn_prob. This outputs a binary prediction for each observation.

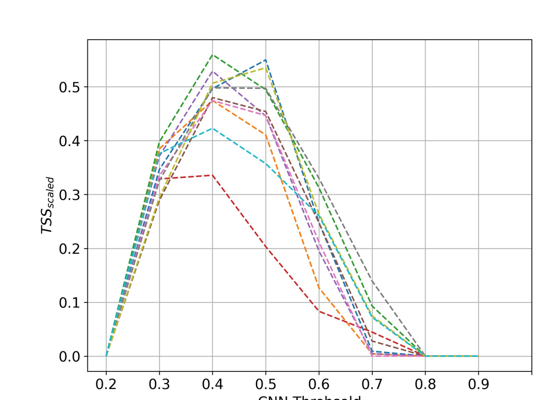
A Novel Metric for Hyperparameter Tuning

For tuning each of the two stages, the TSS metric causes overforecasting:

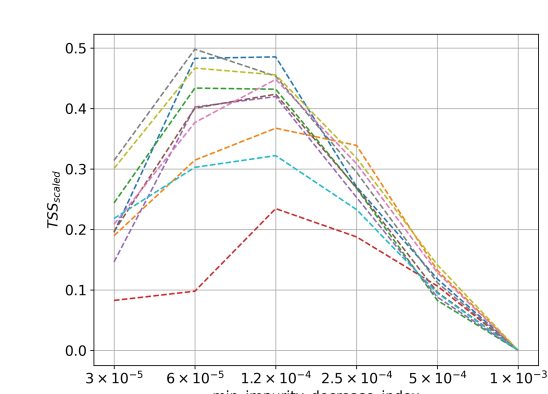
$$\mathrm{TSS} = \mathrm{TPR} - \mathrm{FPR} \quad \in [-1, 1]$$

We instead propose a new metric that additionally penalizes FPs:

$$\mathrm{TSS}_{\mathrm{scaled}} = \mathrm{TPR} - \frac{\mathrm{TPR}_{\max}}{\mathrm{FPR}_{\max}}\mathrm{FPR} \quad \in [-\mathrm{TPR}_{\max}, \mathrm{TPR}_{\max}]$$



VGG-16 optimization on threshold



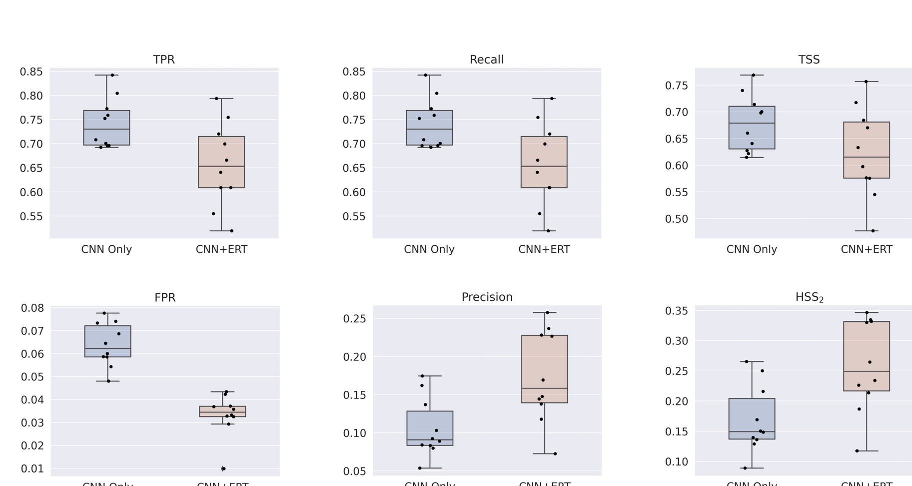
ERT optimization on min_impurity_decrease_index

Model stage	Optimized metric	Optimal hyperparameter	TPR	FPR
CNN-only	TSS	threshold = 0.3	0.90 ± 0.05	0.13 ± 0.02
CNN-only	$\mathrm{TSS}_{\mathrm{scaled}}$	threshold = 0.4	0.75 ± 0.09	0.06 ± 0.01
CNN+ERT	TSS	min_impurity_decrease_index = 0.001	0.90 ± 0.05	0.12 ± 0.02
CNN+ERT	$\mathrm{TSS}_{\mathrm{scaled}}$	min_impurity_decrease_index = 0.00012	0.65 ± 0.11	0.03 ± 0.01

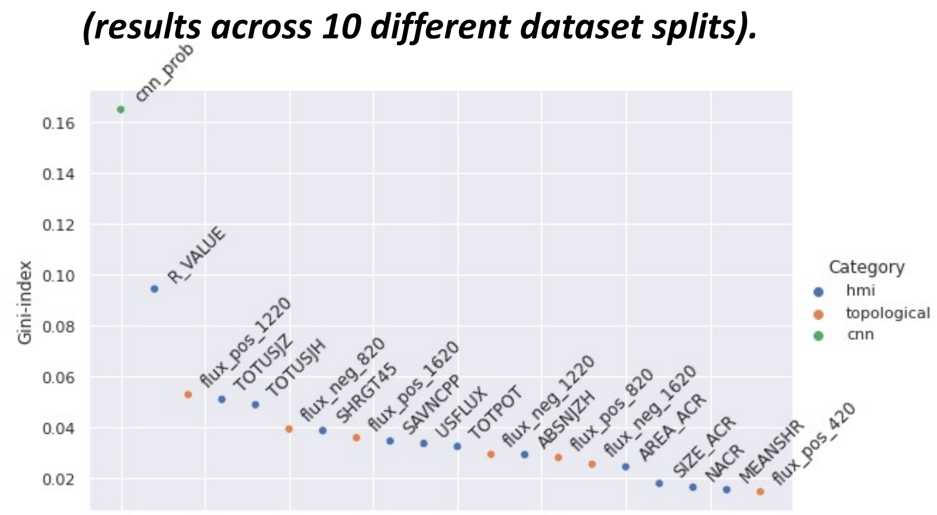
The TSS_{scaled} metric significantly reduces FPR in the prediction, while slightly impacting the TPR.

Results: False Positive Reduction and Feature Ranking

Metric	Formula
Recall (TPR)	$\frac{TP}{TP+FN}$
False Positive/Alarm Rate (FPR)	$\frac{FP}{FP+TN}$
Precision	$\frac{TP}{TP+FP}$
True Skill Statistic (TSS)	$\frac{TP}{TP+FN} - \frac{FP}{FP+TN}$
Heidke Skill Score (HSS_2)	$\frac{2(TP \times TN - FP \times FN)}{(TP+FP)(FP+TN)+(TP+FN)(FN+TN)}$



Metrics sensitive to FPs improve significantly in the CNN+ERT model (results across 10 different dataset splits).



The ERT model can also be used for feature ranking. The cnn_prob output from the VGG-16 model ranks highest, followed by the R_VALUE feature due to Shrijver.

Conclusions

We propose a hybrid two-staged CNN+ERT model for solar flare prediction using SDO/HMI magnetogram. Important findings from this paper are:

1. The CNN model performs best when trained on temporal sequences of the Br component of magnetograms.
2. The two-staged model is shown to be effective in lowering the false positives, thus reducing overforecasting.
3. The proposed metric --- TSS_{scaled} --- for optimizing the hybrid model selects hyperparameters that further reduce false positives.
4. The ERT component of the model is useful for feature ranking, showing that the VGG-16 prediction is the best feature for discriminating flaring and non-flaring magnetograms.

References

[1] Deshmukh, V., Flyer, N., Van Der Sande, K., and Berger, T., "Decreasing False Alarm Rates in ML-based Solar Flare Prediction using SDO/HMI Data", arXiv e-prints, 2021 (To appear in ApJ).
[2] Deshmukh. V., Berger T.E., Bradley E., Meiss J.D. "Leveraging the mathematics of shape for solar magnetic eruption prediction." J. Space Weather Space Clim. 10 13 (2020). DOI: 10.1051/swsc/2020014
[3] Scherrer, P.H., Schou, J., Bush, R.I. *et al.* The *Helioseismic and Magnetic Imager* (HMI) Investigation for the *Solar Dynamics Observatory* (SDO). *Sol Phys* **275**, 207–227 (2012). DOI: 10.1007/s11207-011-9834-2
[4] Simonyan, K. and Zisserman, A., "Very Deep Convolutional Networks for Large-Scale Image Recognition", arXiv e-prints, 2014.