Interpretable Deep Learning for Solar Flare Predictions

Linn Abraham[1*], Vishal Upendran[2, 3], Durgesh Tripathi[1], Ninan Sajeeth Philip[4], Nandita Srivastava[5], A. Ramaprakash[1], Sreejith Padinhatteeri[6]

(1) Inter-University Centre for Astronomy and Astrophysics, Pune; (2) Bay Area Environmental Research Institute, Moffett Field, CA, USA; (3) Lockheed Martin Solar and Astrophysics Laboratory, Palo Alto, CA, USA; (4) Artificial Intelligence Research and Intelligent Systems, Kerala; (5) Udaipur Solar Observatory, Physical Research Laboratory, Udaipur; (6) Manipal Academy of Higher Education, Manipal.

Abstract

Within the past decade or more several Machine learning (ML) based models have been developed for the prediction of solar flares predominantly making use of the magnetogram data. Shallow, interpretable, ML models have been historically employed, operating on numerous derived features from magnetograms, with the algorithms showing nearly similar performance upon optimization. This similarity in performance may result from the application of the same features derived from photospheric magnetograms. Progress on using the original magnetogram or coronal imaging measurements has been minimal due to data complexity and suitable computational models. Deep learning models provide us with the avenue to consume multiwavelength data to perform flare forecasting. In this work, we seek to generate an understanding of intensity measurement on the Sun responsible for flares. For this purpose, we develop a Deep Learning (DL) model and train it to classify AIA Active Region data cubes into flaring and non-flaring classes. We then employ interpretable AI tools such as Integrated Gradients to understand the specific features most important for performing such a classification. This helps us open up the black box DL model, giving us insights into the physics of flare trigger mechanisms.

Results

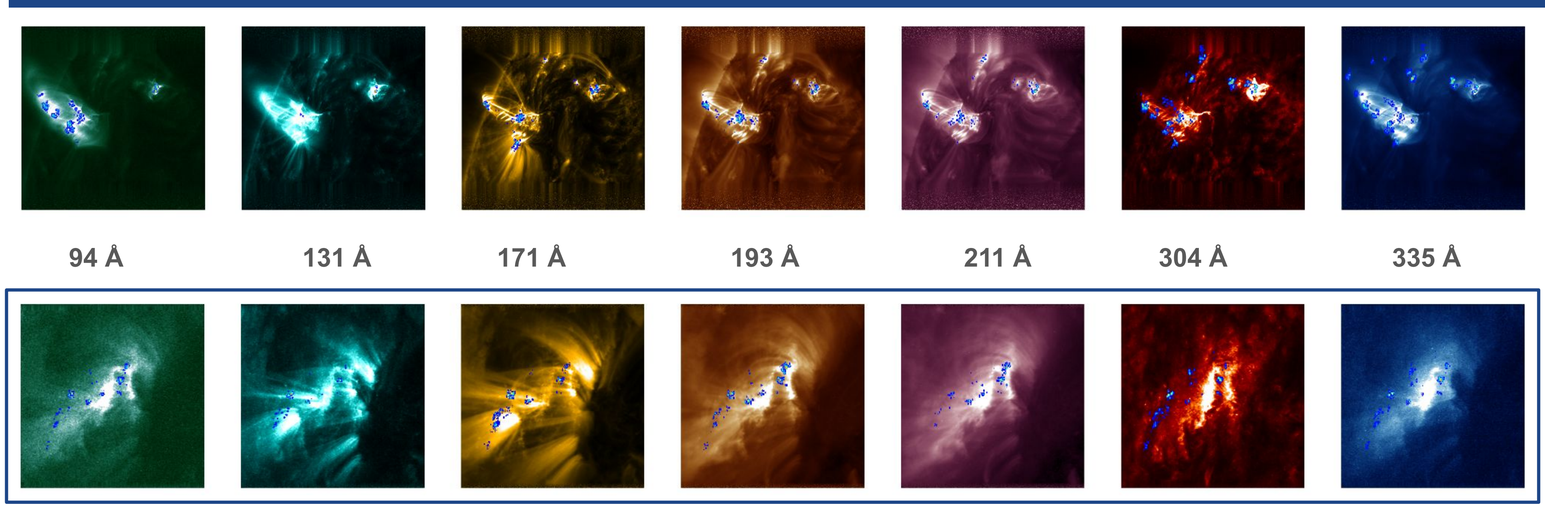

Figure:1 *[**Top row**] Image of a flared active region with contours made from the integrated gradients pixel attribution technique.* ***[Bottom row]*** *The same as for the top row but for a non-flared active region.*

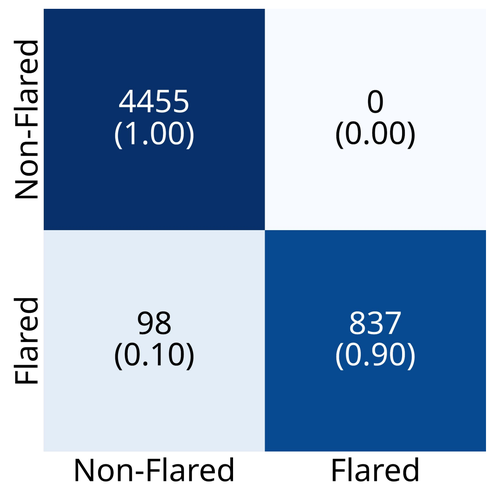

Figure 2: *The confusion matrix showing model performance on test set.*

AIA Active Region Patches (AARPS)

We use the AIA Active Region Patches dataset (AARPS; Dissauer et al., 2023, HARPS; Bobra et al., 2014). Consisting of active region observations of the Sun in seven EUV passbands 94, 131, 171, 193, 211, 304, 335 between 2010 and 2018. Daily observations for six hours consisting of 11 burst images at 72s cadence taken 1 hour apart. These are labelled as a positive if it has eventually produced an X class flare. If it has produced a flare in neither M or X it becomes a negative.

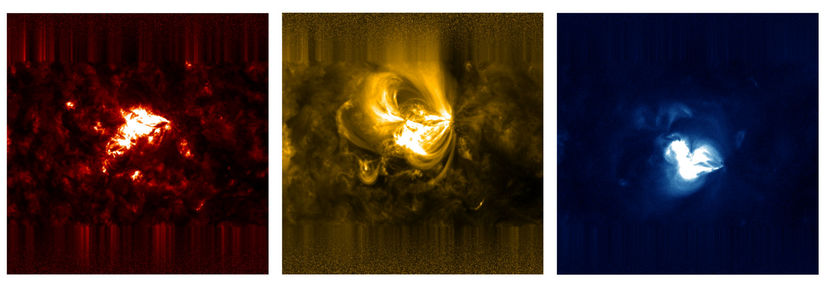

Figure 3 : *Multi-wavelength observation (304 Å, 171 Å, 335 Å) of active region 1449 with padding applied to attain the target size of 512 x 512.*

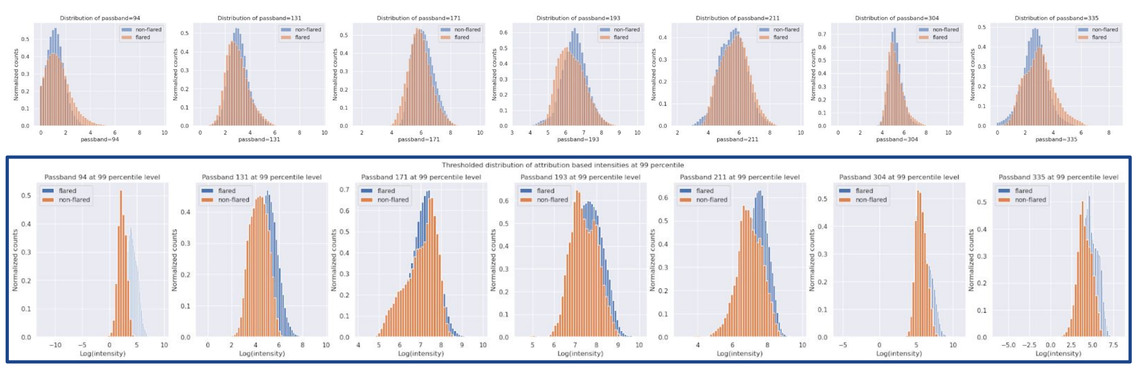

Figure 4 : *(Top) The passband wise distribution of log-transformed intensities for flaring and non-flaring classes across all images in our dataset. (Bottom) The distribution of passband wise log-transformed intensities for flaring and non-flaring corresponding to regions of importance picked by the model attribution technique and thresholded at 99 percentile.*

Data Pre-Processing & Techniques

- Pad all images to the size of the largest AARP in our dataset and then scale everything to the input size of the network.
- Pad using quiet sun values sampled from the image and scaled down in proportion to distance from edge.
- Z-score normalization is done by computing the mean and standard deviation of the training set passband wise.
- A symmetric log transformation is applied on the data.
- Horizontal and vertical flipping is done as part of data augmentation.

- We trained an AlexNet model with 7 input channels.
- We then visualized images from our test set using the integrated gradients technique. This uses the trained model to assign an importance score to each pixel in the image and we then create contours from this map (*see, Figure 1*).

Future Scope

- The HMI data can be used to further investigate physical parameters like current density from the regions picked out by the model.
- The distribution analysis can be done on the image time series.
- We also plan to do a cross validation of the current pixel attribution method to other techniques like Grad-Cam.
- It is also in our interest to extend our dataset by adding M and C

References

1. Properties of Flare-imminent versus Flare-quiet Active Regions from the Chromosphere through the Corona. I. Introduction of the AIA Active Region Patches (AARPs), Dissauer, K., Leka, K. D., & Wagner, E. L. (2023)
2. The Helioseismic and Magnetic Imager (HMI) Vector Magnetic Field Pipeline: SHARPs – Space-Weather HMI Active Region Patches, Bobra, M. G., Sun, X., Hoeksema, J. T., Turmon, M., Liu, Y., Hayashi, K., Barnes, G., & Leka, K. D. (2014)
3. Axiomatic Attribution for Deep Networks, Sundararajan, M., Taly, A., & Yan, Q. (2017)