

Advancing Solar Flare Prediction using Deep Learning with Active Region Patches

Chetraj Pandey^[0000-0002-4699-4050], Temitope Adeyeha^[0000-0001-9512-8733],
Jinsu Hong^[0009-0002-4383-1376], Rafal A. Angryk^[0000-0001-9598-8207], and
Berkay Aydin^[0000-0002-9799-9265]

Georgia State University, Atlanta, GA, 30303, USA
{cpandey1, tadeyeha1, jhong36, rangryk, baydin2}@gsu.edu

Abstract. Solar flares are one of the key space weather phenomena characterized by sudden and intense emissions of radiation from the Sun. The precise and reliable prediction of these phenomena is important due to their potential adverse effects on both space and Earth-based infrastructure. In this paper, we introduce a novel methodology for leveraging shape-based characteristics of magnetograms of active region (AR) patches and provide a novel capability for predicting solar flares covering the entirety of the solar disk (AR patches spanning from -90° to $+90^\circ$ of solar longitude). We create three deep learning models: (i) ResNet34, (ii) MobileNet, and (iii) MobileViT to predict $\geq M$ -class flares and assess the efficacy of these models across various ranges of solar longitude. Given the inherent imbalance in our data, we employ augmentation techniques alongside undersampling during the model training phase, while maintaining imbalanced partitions in the testing data for realistic evaluation. We use a composite skill score (CSS) as our evaluation metric, computed as the geometric mean of the True Skill Score (TSS) and the Heidke Skill Score (HSS) to rank and compare models. The primary contributions of this work are as follows: (i) We introduce a novel capability in solar flare prediction that allows predicting flares for each ARs throughout the solar disk and evaluate and compare the performance, (ii) Our candidate model (MobileNet) achieves a CSS=0.51 (TSS=0.60 and HSS=0.44), CSS=0.51 (TSS=0.59 and HSS=0.44), and CSS=0.48 (TSS=0.56 and HSS=0.40) for AR patches within $\pm 30^\circ$, $\pm 60^\circ$, $\pm 90^\circ$ of solar longitude respectively. Additionally, we demonstrate the ability to issue flare forecasts for ARs in near-limb regions (regions between $\pm 60^\circ$ to $\pm 90^\circ$) with a CSS=0.39 (TSS=0.48 and HSS=0.32), expanding the scope of AR-based models for solar flare prediction. This advancement opens new avenues for more reliable prediction of solar flares, thereby contributing to improved forecasting capabilities.

Keywords: Solar Flares · Deep Learning · Space Weather.

1 Introduction

Solar flares are temporary events characterized by abrupt and massive eruptions of radiation from the Sun's surface. They are critical space weather phenomena

with significant implications for both space-based and Earth-based infrastructures. The National Oceanic and Atmospheric Administration (NOAA) classifies solar flares into five classes based on their peak X-ray flux levels: A, B, C, M, and X, representing the flares from weakest to strongest [8]. Flares weaker than A-class are typically not detected and are thus considered flare-quiet (FQ). M- and X-class flares are the strongest and can cause near-Earth impacts, including disruptions in electrical power grids, the aviation industry, radio and satellite communications, and pose radiation hazards to astronauts in space. Therefore, developing precise and reliable methods to predict solar flares is necessary to mitigate the potential adverse effects of space weather on Earth.

Active regions (ARs) are the areas of high activity on the Sun’s surface, noted for their intense magnetic fields concentrated within sunspots. These magnetic fields often undergo significant distortion and instability, triggering plasma disturbances and releasing energy in the form of flares and other solar phenomena [36]. This makes ARs the regions of interest, emphasizing the importance of utilizing AR-based features for predicting solar flares. However, the magnetic field measurements, which are the dominant feature employed by AR-based methods, are susceptible to severe projection effects caused by the orientation of the observing instrument relative to the solar surface. Therefore, as ARs approach the solar limbs, specifically beyond $\pm 60^\circ$ of solar longitude, the magnetic field readings become distorted [7], which limits the existing models to include data pertaining to central locations only [24], [25], [26]. To address this, we derive images from original line-of-sight (LoS) magnetogram rasters of AR patches with our novel data processing pipeline that captures the overall morphology and spatial distribution of active regions, retaining important shape-based parameters such as size, directionality, sunspot borders, and polarity inversion lines [15], [21]. We recognize the persistence of projection effects also in images of AR magnetogram patches; however, we hypothesize that the complex feature learning capabilities of contemporary deep learning models can potentially learn from the shape-based features retained in images while filtering the distorted readings. Consequently, we include data from ARs beyond $\pm 60^\circ$ as well, thereby providing a novel capability to predict solar flares across the entire disk.

Furthermore, it is important to note that the tracked AR patches vary in size depending on the size of the ARs. Existing approaches have been limited to AR patches in central locations, often resizing rectangular patches to obtain square images. However, this resizing distorts the original aspect ratio, consequently altering the shapes and sizes of ARs. Alternatively, variable-sized AR patches are cropped (using methods like center crop or random crop) to obtain square images, resulting in information loss. In contrast, we propose and utilize a sliding window kernel-based approach. This method selects such a cropped region that maximizes total unsigned flux (USFLUX: the sum of the absolute of the magnetic field strength values), maintaining the original aspect ratios of AR patches and preserving critical spatial features. By maximizing the USFLUX, we ensure that we extract the most representative region with significant magnetic flux build

up. This method adapts to the variability in AR patch shapes and sizes, avoiding distortion and prioritizing the capture of more relevant information.

In this study, we develop deep learning models to predict solar flares of magnitude $\geq M$ -class by leveraging images created from cutouts (patches) of magnetograms corresponding to ARs. We employ these images to train three distinct deep learning architectures: ResNet34 [9], MobileNet [12], and MobileViT [22]. Our contributions can be summarized as follows: (i) we propose a novel methodology for preprocessing magnetograms of AR patches, focusing on preserving shape-based parameters while maximizing the total unsigned flux (USFLUX) [3] at the instance level, (ii) we showcase that our models are capable of predicting flares across the entire solar disk, including often overlooked near-limb regions, improving the comprehensiveness of AR-based solar flare prediction models, and (iii) through rigorous experimentation, our models demonstrate superior performance compared to existing approaches, significantly contributing to ongoing efforts aimed at enhancing space weather forecasting capabilities.

The remainder of this paper is organized as follows: In Sec. 2, we outline the various approaches used in solar flare prediction along with contemporary work using deep learning methods. In Sec. 3, we explain our data processing pipeline, class-wise distribution of the data for binary prediction mode along with description of our flare prediction models. In Sec. 4, we present our experimental design, evaluations, and discuss the implications of our work, lastly, in Sec. 5, we provide our concluding remarks and discuss avenues for future work.

2 Related Work

A range of methodologies, such as human-based prediction techniques (e.g., [4]), statistical approaches (e.g., [18]), and numerical simulations based on physics-based models (e.g., [17]), have been employed to predict solar flares. Recently, the success of data-driven approaches, which leverage machine learning and deep learning techniques, has significantly increased owing to their capacity to exploit extensive datasets [10] and their experimental achievements [23]. As solar flares are phenomena caused by sudden, abrupt changes in the magnetic field in the solar atmosphere, these data-driven approaches most commonly utilize magnetogram-based data which includes solar full-disk magnetograms (e.g., [11], [27], [28], [29], [31]), multivariate time series (MVTs) data extracted from solar photospheric vector magnetograms (e.g., [14], [16]), cutouts or patches of tracked AR (e.g., [13], [20]), and features summarizing each AR patches (e.g., [1], [2]).

A deep learning model based on a multi-layer perceptron to predict solar flares $\geq C$ and $\geq M$ class was presented in [23]. In this study, they used 79 manually selected features extracted from multi-modal solar observations of full solar disk, which included vector magnetograms and extreme ultraviolet (EUV) images to predict $\geq M$ - and $\geq C$ -class flares. In [32], a CNN-based hybrid model to predict the occurrence of a $\geq C$ -class flares. Similarly, in [26], [28] we presented a convolutional neural network (CNN) based model to predict $\geq M$ -class flares utilizing full-disk magnetogram images. While these full-disk models include

near-limb regions, by themselves they are unable to localize the relevant AR which is likely to flare and instead issue one single forecast for entire solar disk.

In [2], a support vector machine (SVM) based model trained with 25 AR summary parameters extracted from vector magnetograms of AR patches within $\pm 68^\circ$ of solar longitude (central meridian distance) was presented. Similarly, in [16], a deep learning based time series classifier and in [14] a sliding window Time Series Forest (TSF) was trained with a MVTs data of 24 space weather related physical parameters primarily calculated from AR magnetograms within $\pm 70^\circ$ of solar longitude. Furthermore, a CNN-based flare forecasting model trained with AR patches (resized to 100×100 pixels) extracted from LoS magnetograms within $\pm 30^\circ$ of the solar longitude to predict $\geq C$ -, $\geq M$ -, and $\geq X$ -class flares was presented in [13]. More recently, [19] proposed a CNN-based model named “CARFFM-4” trained with AR patches (sized to 160×160 pixels) created from R parameter [35] within $\pm 30^\circ$ of solar longitude to predict $\geq M$ -class flares in next 48 hours. It is important to note that, there is variability in literature in terms of type of data modality which includes multiple instruments (HMI/SDO, AIA/SDO, MDI/SOHO) and data types (EUV images, magnetograms and extracted features corresponding to AR and full-disk). Furthermore the variability in prediction targets ($\geq C$ -, $\geq M$ -, $\geq X$ -class flares) and forecasting horizon (24 hours and 48 hours) is also prominent. The predictive capabilities of AR-based models are often limited by observations taken from central locations from $\pm 30^\circ$ to $\pm 70^\circ$. The full-disk models complement the issue of longitudinal coverage in AR-based models; however, they fail to pin-point an active region and issue a single forecast for the entire solar disk. In this work, we introduce a limb-to-limb AR-based flare prediction model encompassing full 180° ($\pm 90^\circ$ of the solar longitude) and evaluate our models efficacy in different longitudinal range and provide a novel capability, to our knowledge, missing in operational systems.

3 Data and Model

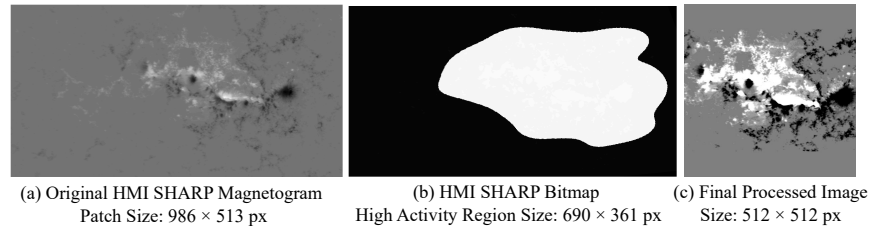


Fig. 1. An illustrative example of (a) Original raw input magnetogram of HMI AR patch corresponding to HARP number: 4781 (NOAA AR number: 12205) at timestamp 2014-11-06T18:00:00 UTC, (b) Bitmap corresponding to HMI AR patch in (a) showing the region of interest indicated by white pixels, (c) Final processed image of AR patch in (a) now sized to 512×512 , that is used to train our models.

The raw input data in our work includes LoS magnetograms of ARs provided by the Helioseismic and Magnetic Imager (HMI) [34] onboard the Solar Dynamics Observatory (SDO) [33], which are publicly available as a data product named Space-Weather HMI Active Region Patches (SHARP) [3] from the Joint Science Operations Center¹. In this work, we utilized magnetograms spanning from May 2010 to 2018, sampling magnetograms at a cadence of one hour. The magnetograms of AR patches contain rasters of magnetic field strength values typically ranging from $\sim \pm 4500$ G. An example of magnetogram of AR patch is shown in Fig. 1(a). Along with magnetograms, we use bitmaps (another data product from the SHARP series) which define the region with pixels located within or outside the ARs, providing the region of interest within the AR patch as shown in Fig. 1 (b). The bitmap are equal in size to the LoS magnetograms of AR patches and contains five unique pixel values: 0, 1, 2, 33, 34, where the values 33 and 34 indicate the pixels that are within the AR region and hence our region of interest [3]. For each AR patch, we assign a binary label using peak X-ray flux converted to NOAA flare classes such that: (i) $\geq M$ indicates Flare (FL) signifying the existence of a relatively strong flaring activity, and (ii) $< M$ indicates No Flare (NF) with a prediction window of 24 hours. To elaborate, from the timestamp of an AR patch to the next 24 hours, if the maximum flare class is $< M$, then we label the AR patch as NF; otherwise, FL.

In our data processing pipeline, which is illustrated in Fig. 2, we begin by collecting hourly instances of original raw input magnetograms of active region (AR) patches, alongside their corresponding bitmaps. Our initial step involves applying the bitmap as a filter to precisely crop the AR patches, isolating the regions with high activity. Subsequently, we implement a size filter: if the resulting cropped AR patches are smaller than 70 pixels in width, we exclude them from our dataset. It is worth noting that we determine this threshold based on the overall data distribution, ensuring retention of all instances corresponding to ‘FL’ instances while removing those from the ‘NF’ class. Following this filtering stage, we proceed to adjust the magnetic flux. We cap the flux values at ± 256 G, and any flux values within ± 25 G are set to 0 to mitigate noise. Ensuring uniformity in size, we apply zero-padding to patches smaller than 512×512 pixels. Conversely, for larger patches, exceeding 512×512 pixels, we employ a 512×512 kernel to select the patch with the maximum total unsigned flux (namely US-FLUX, which is the sum of the absolute value of magnetic field strength represented as raster values in magnetograms). By doing this, we aim to minimize information loss by picking a spatial window where the total flux is the highest, which is more likely to include the regions of interest. Finally, to standardize the representation, all patches are scaled to fit within the range of 0-255, facilitating the generation of images. An example of final preprocessed image utilizing the magnetogram raster in Fig. 1 (a) and bitmap in Fig. 1 (b) is shown in Fig. 1 (c).

The overall distribution of our binary-labeled AR patches data, with flare classes NF (comprising flare-quiet (FQ), A-, B-, and C-class flares) and FL (including M- and X-class flares), is depicted in Fig. 3(a). In total, we have 501,106

¹ <http://jsoc.stanford.edu>

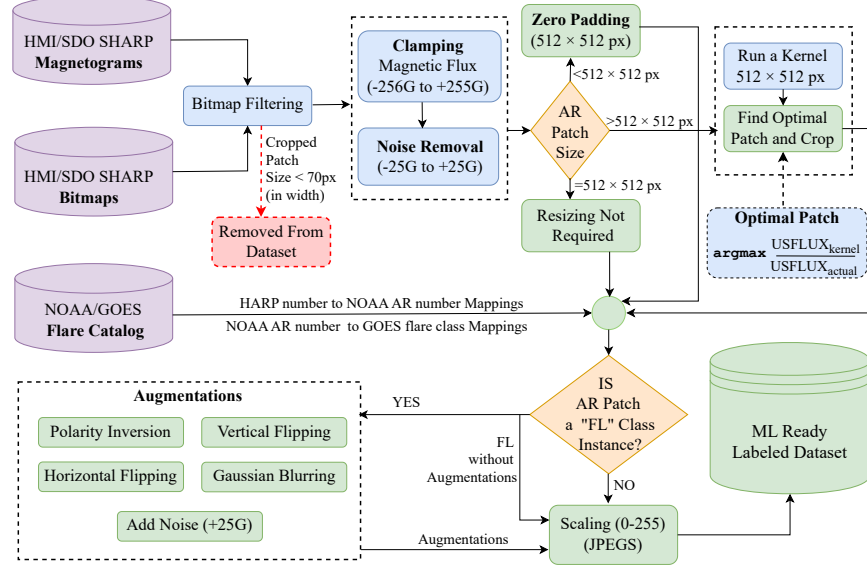


Fig. 2. The process flow diagram of data processing pipeline used in this work. It shows a sequential pipeline for creating JPEG images from magnetogram rasters and corresponding bitmaps along with data augmentation pipeline given the label for the magnetogram patch. Boxes colored in green collectively defines our entire dataset.

instances belonging to the NF class and 10,315 instances belonging to the FL class, resulting in a class imbalance ratio of $\sim 1:49$. We split our entire dataset using temporally non-overlapping tri-monthly partitioning into four partitions based on the onset timestamp of HARP series. To elaborate, we ensure that the data corresponding to an entire AR trajectory is included in only one partition to prevent double-dipping by using the onset timestamp of the HARP series for tri-monthly partitioning, in contrast to using the observation timestamp of magnetograms as mentioned in [24]. The resulting data distribution is shown in Fig. 3 (b), where we use Partitions 1 and 2 as our training set while Partitions 3 and 4 as validation and test set respectively. The data augmentation and undersampling of training set is later described in Sec. 4.1.

The task of solar flare prediction in this work is formalized as a binary image classification problem; therefore, we select three general-purpose deep learning models: (i) ResNet34 [9], (ii) MobileNet [12], and (iii) MobileViT [22]. Recently, attention-based models, notably Vision Transformers (ViTs) [6], have emerged as frontrunners in the task of image classification. They have showcased superior performance on large-scale datasets compared to standard CNNs. However, ViTs typically boast a high number of trainable parameters (ranging from approximately 86 to 632 million), making them demanding in terms of computational resources. This limits their practicality in scenarios with restricted computational capabilities or smaller datasets. In response to these challenges,

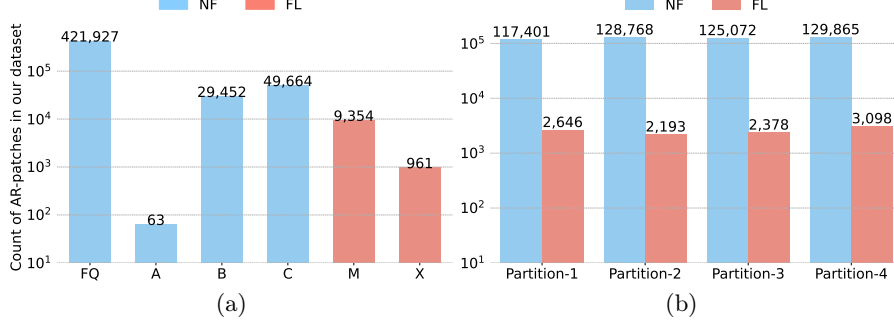


Fig. 3. (a) Overall data distribution (b) Overall data partitioned into four tri-monthly partitions. Note: The height of the bars are in log scale.

particularly in our context with a modest dataset, we explore alternative models that strike a balance between accuracy and efficiency. Our focus lies in selecting lightweight architectures that can deliver competitive performance while being more resource-efficient. ResNet34 [9] and MobileNet [12] are standard CNNs with ~ 21.2 and ~ 4.2 million trainable parameters, respectively, while MobileViT boasts ~ 2 million trainable parameters. ResNet34 excels in capturing fine-grained details within images, potentially aiding in identifying subtle patterns indicative of solar flares. MobileNet, renowned for its efficiency, offers a balance between computational resources and accuracy, making it suitable for deployment in resource-limited environments such as space-based solar observation platforms. Additionally, MobileViT [22], a variant of the Vision Transformer [6] inspired by MobileNet is optimized for computational efficiency, leveraging both efficiency and global context capturing capabilities, potentially enhancing the models' performance in capturing complex patterns relevant to solar flare prediction. This selection of models aims to evaluate their efficacy for solar flare prediction while prioritizing lightweight architectures with varying designs.

4 Experimental Evaluation

4.1 Experimental Settings

As mentioned earlier in Sec. 3, we follow time-segmented tri-monthly partitioning scheme to create four partitions of our entire dataset. Partition-1 and 2 combined are used as the training set. However, due to significant class imbalance in our dataset, we used undersampling together with data augmentation to create a balanced training set. Firstly, we augmented data instances belonging to the FL-class in our training set using five data augmentation techniques: (i) polarity inversion, which swaps the signs of positive polarity to negative and vice versa as shown in Fig. 4 (b), (ii) Gaussian filtering, which applies a Gaussian blur to the image to reduce noise and detail (Fig. 4 (c)), (iii) Horizontal Flipping, which involves flipping the image along a vertical axis (Fig. 4 (d)), (iv) Vertical

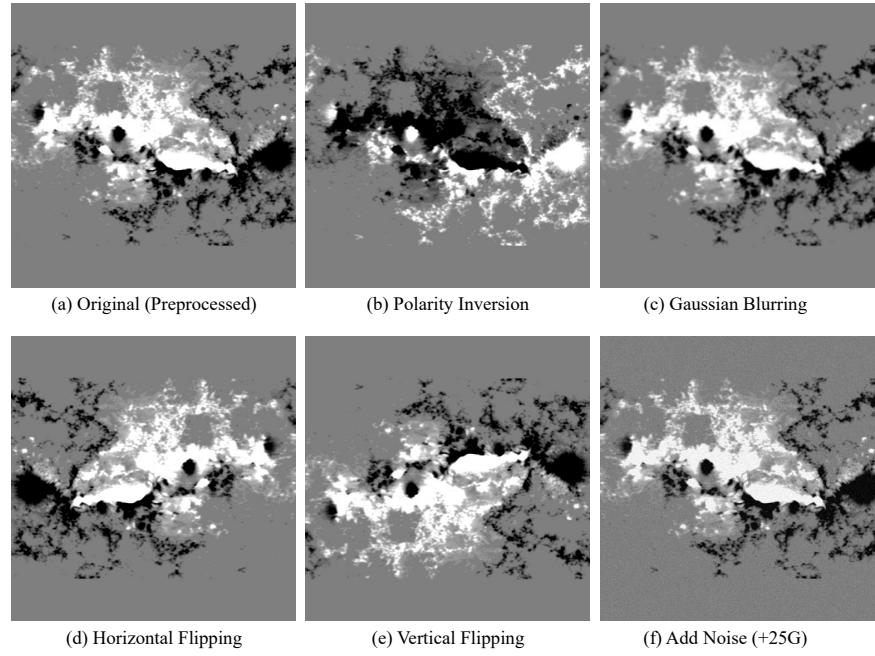


Fig. 4. An illustrative example of (a) input magnetogram of HMI AR patch corresponding to HARP number: 4781 (NOAA AR number: 12205) at timestamp 2014-11-06T18:00:00 UTC, (b-f) five different augmentations applied to AR patch in (a).

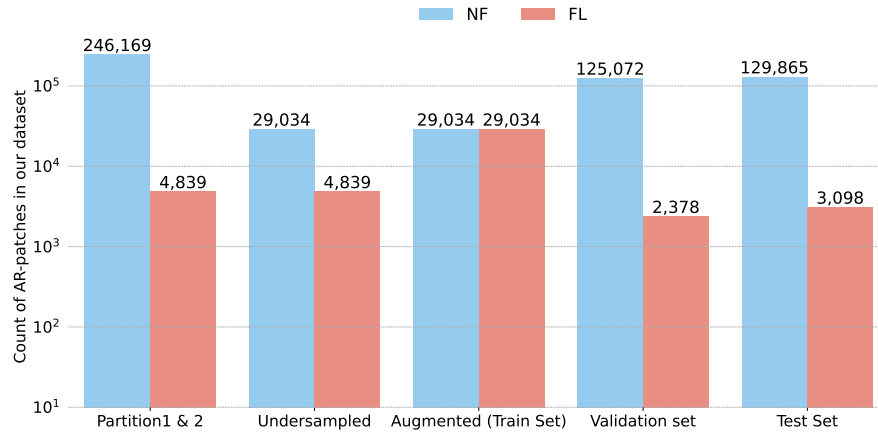


Fig. 5. Overall data partitioned into train set (showing original, undersampled and augmented data counts) , validation set, and test set used in this study.

Flipping , which involves flipping the image along a horizontal axis (Fig. 4 (e)), and (v) Adding random noise within $\pm 25\text{G}$ (Fig. 4 (f)). To balance the FL-Class

instances with NF, we undersampled our training data by randomly selecting 30% of instances belonging to A-, B-, C-class flares each, and $\sim 8\%$ of instances from FQ from both Partition-1 and 2. For realistic evaluation, we maintained the original imbalanced distribution in Partitions 3 and 4, which are our validation and test sets respectively, as shown in Fig. 5.

Model Parameters: In our model hyperparameter selection process, we define the hyperparameter space, encompassing initial learning rate sets, weight decay sets, batch size sets, and class weight sets as shown in Table 1. We also investigate three sets of class weights, although the training data was balanced, aiming to assess their impact on minimizing false positives. These class weight configurations are indicated as NF:FL, where 1:1 would suggest equal weights in loss minimization. To prioritize the minimization of false positives, we explored increasing the weights for NF-class instances. Following the definition of our hyperparameter space, we conduct a meticulous grid search across this space, evaluating on the validation set for all three models. During this search, we train our models using stochastic gradient descent (SGD) with cross-entropy loss. Additionally, we employ a dynamic learning rate strategy, which reduced the initial learning rate by a factor of 10 every 10 epochs. Upon completing the grid search and evaluating the models, we identified the optimal hyperparameters as shown in Table 1. These parameters exhibited superior performance, we use these to train our final models for 50 epochs and evaluate on the test set.

Table 1. Hyperparameters search space with optimal hyperparameters for each model.

Hyperparameters	Search Space	Optimal Parameters		
		ResNet34	MobileNet	MobileViT
Learning Rate	{0.00001 to 0.01}	0.001	0.001	0.001
Weight Decay	{0.00001 to 0.01}	0.01	0.01	0.001
Batch Size	{48, 64}	48	48	48
Class Weights	{1:1, 3:1, 5:1}	5:1	5:1	5:1

Evaluation Metrics: True Skill Statistic (TSS, in Eq. 1) and Heidke Skill Score (HSS, in Eq. 2), derived from the four elements of confusion matrix: TP, TN, FP, FN are the two forecast skills scores widely used in evaluating flare prediction models.

$$TSS = \frac{TP}{TP + FN} - \frac{FP}{FP + TN} \quad (1)$$

$$HSS = 2 \times \frac{TP \times TN - FN \times FP}{((P \times (FN + TN) + (TP + FP) \times N))} \quad (2)$$

where, $N = TN + FP$ and $P = TP + FN$.

TSS and HSS values range from -1 to 1, where 1 indicates all correct predictions, -1 represents all incorrect predictions (also, it means that all inverse predictions are correct, i.e., there is a skill), and 0 represents no skill. In contrast to TSS, HSS is an imbalance-aware metric that is commonly employed in solar flare prediction models due to the prevalent high class-imbalance ratios. However, choosing a candidate model based on two skill scores becomes difficult, as it demands preference of one metric over another at the end. Therefore, by combining TSS and HSS in a geometric mean as in the Composite Skill Score (CSS, in Eq. 3), we obtain a single metric that balances between discrimination ability and imbalance awareness.

$$CSS = \begin{cases} 0 & \text{if } TSS \times HSS < 0 \\ \sqrt{TSS \times HSS} & \text{otherwise.} \end{cases} \quad (3)$$

CSS considers both the discrimination power of the model (TSS) and its ability to outperform random chance (HSS), offering a more comprehensive evaluation. It provides a single metric where the values range from 0 to 1, with 1 indicating perfect skill. CSS accounts for both aspects of model performance, making it more suitable for assessing forecast models, particularly in scenarios with class imbalance. Therefore, we evaluate and compare our models based on the single metric, CSS, but we also report both TSS and HSS for completeness.

4.2 Evaluation

As explained earlier in Sec. 4.1, we conducted experiments to predict solar flares in a binary setting ($\geq M$ -class flares) using a "train-validation-test split" of our entire dataset, which consists of magnetogram of AR patches covering a solar longitudinal range of $\pm 90^\circ$ (i.e., the entire solar disk). We utilized the validation set to monitor the models' performance every epoch and tuned hyperparameters to optimize the CSS. After training the model with optimal hyperparameters, we employed a threshold tuning approach to calibrate our models by tuning the prediction score thresholds. This involved evaluating the performance of each model on the validation set at different threshold values ranging from 0.01 to 0.99 with an increment of 0.01. We selected the threshold that optimized CSS for each model, resulting in thresholds of 0.36, 0.46, and 0.35 for the ResNet34, MobileNet, and MobileViT models respectively. These thresholds were then applied to the test set for each model, and the models' performance was reported. Additionally, we assessed the performance of all three models on subsets of data representing different longitudinal coverage (within $\pm 30^\circ$, $\pm 45^\circ$, $\pm 60^\circ$, $\pm 75^\circ$, and $\pm 90^\circ$ of solar longitude), with $\pm 90^\circ$ indicating the entire test set. The performance of our models relative to each other in terms of TSS, HSS, and CSS is illustrated in Fig. 6 (a), (b), and (c) respectively.

We observed that, the TSS for MobileViT was higher, while HSS was consistently lower compared to other two models, which empirically highlights our hypotheses of using a composite skill score, as choosing the model based on TSS scores might lead to a false sense of good performance. Therefore based on CSS

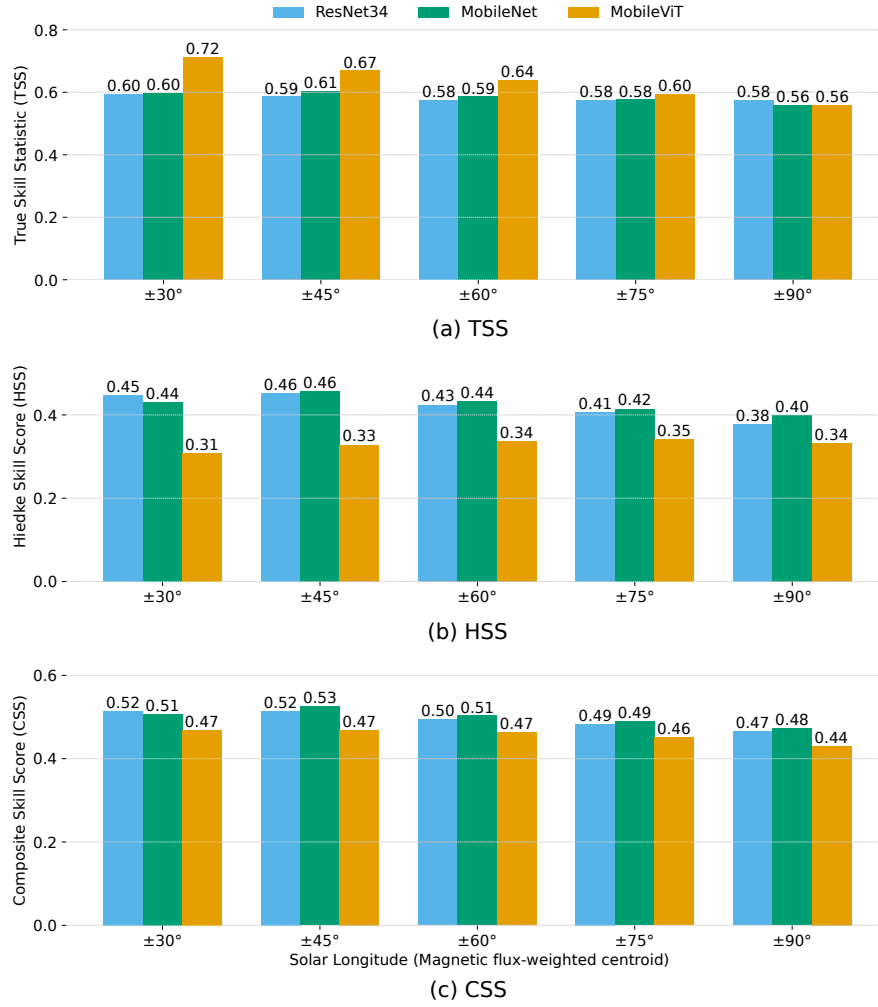


Fig. 6. Performance of our models on test set in terms of (a) TSS, (b) HSS, (c) CSS.

score, we observed that the overall performance corresponding to entire test set ($\pm 90^\circ$), MobileNet, achieved the best performance with CSS=0.48, (TSS=0.56, HSS=0.40) while the lowest CSS was observed with MobileViT with CSS=0.44 (TSS=0.56, HSS=0.34). Furthermore, our analysis revealed a linearly decreasing trend in model performance with increasing longitudinal coverage of ARs.

We conducted a comprehensive comparison of our models' performance against existing literature within the longitudinal coverage of their experiments as presented in Table. 2, focusing on CSS while reporting both the TSS, and HSS met-

Table 2. Comparison with existing literature in terms of TSS, HSS, and CSS with in specific longitudinal coverage.

Longitudinal Coverage	Models	TSS	HSS	CSS
Within $\pm 30^\circ$	Huang et al., 2018 [13]	0.66	0.14	0.31
	Li et al., 2023 [19]	0.45	0.44	0.44
	ResNet34 (This Work)	0.60	0.45	0.52
	MobileNet (This Work)	0.60	0.44	0.51
	MobileViT (This Work)	0.72	0.31	0.47
Within $\pm 60^\circ$	Bloomfield et al., 2012 [1]	0.54	0.19	0.32
	Bobra et al., 2015 [2] ($\pm 68^\circ$)	0.76	0.51	0.62
	Ji et al., 2022 [16] ($\pm 70^\circ$)	0.81	0.22	0.42
	Ji et al., 2023 [14] ($\pm 70^\circ$)	0.81	0.43	0.59
	ResNet34 (This Work)	0.58	0.43	0.50
	MobileNet (This Work)	0.59	0.44	0.51
	MobileViT (This Work)	0.64	0.34	0.47
Within $\pm 90^\circ$	Nishizuka et al., 2018 [23]	0.80	0.26	0.45
	ResNet34 (This Work)	0.58	0.38	0.47
	MobileNet (This Work)	0.56	0.40	0.48
	MobileViT (This Work)	0.56	0.34	0.44

rics as well². Notably, the superior scores are highlighted in bold. Our analysis revealed that all three of our models exhibited superior performance compared to both [13] and [19]. Specifically, ResNet34 surpassed [13] and [19] by 21% and 8% in CSS, respectively, within the $\pm 30^\circ$ range. Similarly, in comparison to models evaluated within $\pm 60^\circ$, our model (MobileNet) outperformed [1] by 19%. While we show the best performance by CSS in bold for [2], it is important to note the challenges in making precise comparisons due to the varying longitudinal ranges covered in studies by [2], [16], [14] (spanning $\pm 68^\circ$, $\pm 70^\circ$, and $\pm 70^\circ$, respectively). Despite these differences, our findings underscore the effectiveness of our approach within the established longitudinal boundaries.

Furthermore, on evaluating with in $\pm 90^\circ$, notably, the MobileNet model demonstrated the best results, outperforming [23] by 3% in CSS. Finally, we emphasize that although there are full-disk models which covers entire solar disk ($\pm 90^\circ$), as mentioned in Sec.2, they are not directly comparable to AR- based models [30], therefore, we exclude full-disk models from comparison. Overall, our study presents a pioneering approach by developing AR-based models that

² It should be noted that the comparisons against the existing literature might not be accurate due to differences in datasets and partitions used across these studies, as already mentioned in Sec. 2. However, they are intended to provide general insights into the standings reported in other studies discussed in this paper.

incorporate AR patches spanning the entire solar disk, a methodology never before explored in solar flare prediction. Our comprehensive evaluation demonstrates the superiority of our models over existing ones within their respective longitudinal coverage. This highlights the effectiveness of our novel approach in advancing the solar flare prediction. To reproduce this work, the source code and experimental results can be accessed from our open source repository [5].

4.3 Discussion

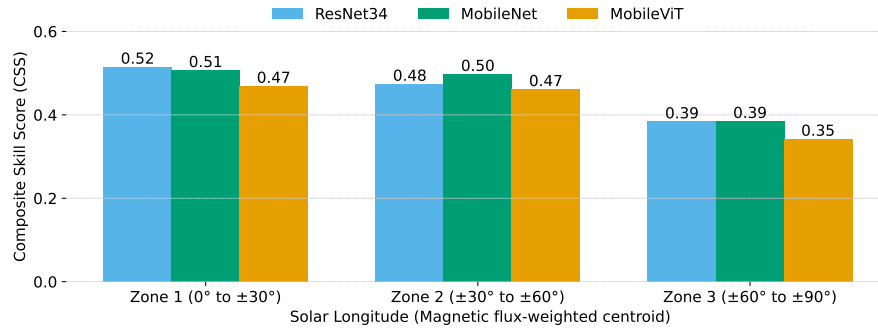


Fig. 7. Models’ efficacy evaluated on three different Zones in terms of CSS. Zone 1 and 2 indicates central locations and Zone 3 represents near-limb location.

Upon recognizing a pattern indicating a decline in model performance with increasing longitudinal coverage, we investigated the effectiveness of our models on non-overlapping regions of solar longitudes. To facilitate this analysis, we delineated three zones: Zone 1 representing the region within $\pm 30^\circ$, Zone 2 covering the area between $\pm 30^\circ$ to $\pm 60^\circ$, and Zone 3 spanning $\pm 60^\circ$ to $\pm 90^\circ$. To evaluate the models’ performance across these zones, we computed the CSS and observed a clear decrease in model skill towards the limb, as illustrated in Fig. 7 from all three models. It is worth noting that while existing models are typically designed to predict solar flares up to Zone 2, our model demonstrates capability in the near-limb regions, which includes Zone 3. Despite lower skill scores compared to those in the central region, this study unveils a novel capability that shows skills on the near-limb region, thereby advancing solar flare prediction. This achievement underscores the significance of our research in enhancing our understanding and predictive capabilities in solar phenomena.

5 Conclusion and Future Work

In this study, we introduced three limb-to-limb (i.e., $\pm 90^\circ$) flare prediction models to forecast solar flares of magnitude $\geq M$ -class, which are trained with images

created from LoS magnetic field component of AR patches. The primary focus of our work was to demonstrate the predictive capacity of AR-based models in near-limb regions (beyond $\pm 60^\circ$). On evaluating this novel limb-to-limb model capabilities, the results show that we can satisfactorily predict the flaring activity in the existence of severe projection effects although there is room for improvement (the skill is limited when compared to central locations). Furthermore, the results also support our hypothesis that shape-based features derived from magnetograms are effective when predicting solar flares even when the ARs are close to limbs. Moreover, we also introduced a novel preprocessing pipeline for image transformation of magnetic field data products. This pipeline is remarkable for deep learning-based flare forecasting tools and it provides a systematic workflow for clamping, padding, thresholding and most-relevant window selection; which is critical considering the significant class imbalance in forecasting tasks. Finally, we also introduced a new evaluation metric CSS, which is also important and presents us with a practical index for model comparison and selection. While full-disk models are developed to complement AR-based models in near-limb regions, they lack the ability to localize AR-specific predictions. We define this work as a pioneering step towards fully integrating ARs into solar flare prediction, with significant implications for advancing such predictions. Numerous avenues for future exploration exist, including investigating multimodal solar observations, developing spatiotemporal models, and incorporating explanatory/ interpretative frameworks into the model to enhance reliability.

Ethical Statement: Space weather forecasting research involves several ethical considerations. The data for solar flare prediction, provided publicly by NASA/SDO and the AIA, EVE, and HMI science teams, is free from data privacy and security concerns. Non-commercial use of SDO images for education and public information is encouraged without needing authorization. Ethical and responsible development and use of forecasting models are crucial to prevent biases or negative impacts. These models, despite their advanced capabilities, have limitations due to the rarity of extreme solar events and the methods used to evaluate them. Caution and multiple information sources are essential in decision-making related to space weather events. Transparency about these models' limitations ensures ethical and responsible use, mitigating potential harm. Finally, space weather forecasting should be used for peaceful purposes, such as early detection and reducing vulnerabilities to extreme space weather events.

Acknowledgments. This work is supported in part under two NSF grants (Award #2104004 and #1931555) and a NASA SWR2O2R grant (Award #80NSSC22K0272). The data used in this study is a courtesy of NASA/SDO and the AIA, EVE, and HMI science teams, and the NOAA National Geophysical Data Center (NGDC).

References

1. Bloomfield, D.S., Higgins, P.A., McAteer, R.T.J., Gallagher, P.T.: Toward reliable benchmarking of solar flare forecasting methods. *The Astrophysical Journal*

- 747**(2), L41 (Feb 2012)
2. Bobra, M.G., Couvidat, S.: Solar flare prediction using SDO/HMI vector magnetic field data with a machine-learning algorithm. *The Astrophysical Journal* **798**(2), 135 (Jan 2015)
 3. Bobra, M.G., Sun, X., Hoeksema, J.T., Turmon, M., Liu, Y., Hayashi, K., Barnes, G., Leka, K.D.: The helioseismic and magnetic imager (hmi) vector magnetic field pipeline: Sharps – space-weather hmi active region patches. *Solar Physics* **289**(9), 3549–3578 (Apr 2014)
 4. Crown, M.D.: Validation of the NOAA space weather prediction center's solar flare forecasting look-up table and forecaster-issued probabilities. *Space Weather* **10**(6) (Jun 2012)
 5. DMLab, G.: Arpatch-sfp. <https://bitbucket.org/gsudmlab/arpatch-sfp/src/main/> (2024)
 6. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., Houlsby, N.: An image is worth 16x16 words: Transformers for image recognition at scale. *CoRR abs/2010.11929* (2020), <https://arxiv.org/abs/2010.11929>
 7. Falconer, D.A., Tiwari, S.K., Moore, R.L., Khazanov, I.: A new method to quantify and reduce the net projection error in whole-solar-active-region parameters measured from vector magnetograms. *ApJ* **833**(2), L31 (Dec 2016)
 8. Fletcher, L., Dennis, B., Hudson, H., Krucker, S., Phillips, K., Veronig, A., Battaglia, M., Bone, L., Caspi, A., Chen, Q., Gallagher, P., Grigis, P., Ji, H., Liu, W., Milligan, R., Temmer, M.: An observational overview of solar flares. *Space Science Reviews* **159**(1-4), 19–106 (Aug 2011)
 9. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition (2015)
 10. Hong, J., Ji, A., Pandey, C., Aydin, B.: Enhancing solar flare prediction with innovative data-driven labels. In: 2023 IEEE 5th International Conference on Cognitive Machine Intelligence (CogMI). IEEE (Nov 2023). <https://doi.org/10.1109/cogmi58952.2023.00035>
 11. Hong, J., Pandey, C., Ji, A., Aydin, B.: An innovative solar flare metadata collection for space weather analytics. In: 2023 International Conference on Machine Learning and Applications (ICMLA). pp. 408–413 (2023). <https://doi.org/10.1109/ICMLA58977.2023.00063>
 12. Howard, A., Sandler, M., Chu, G., Chen, L., Chen, B., Tan, M., Wang, W., Zhu, Y., Pang, R., Vasudevan, V., Le, Q.V., Adam, H.: Searching for mobilenetv3. *CoRR abs/1905.02244* (2019), <http://arxiv.org/abs/1905.02244>
 13. Huang, X., Wang, H., Xu, L., Liu, J., Li, R., Dai, X.: Deep learning based solar flare forecasting model. i. results for line-of-sight magnetograms. *The Astrophysical Journal* **856**(1), 7 (Mar 2018)
 14. Ji, A., Aydin, B.: Interpretable solar flare prediction with sliding window multivariate time series forests. In: 2023 IEEE International Conference on Big Data (BigData). IEEE (Dec 2023)
 15. Ji, A., Cai, X., Khasayeva, N., Georgoulis, M.K., Martens, P.C., Angryk, R.A., Aydin, B.: A systematic magnetic polarity inversion line data set from sdo/hmi magnetograms. *The Astrophysical Journal Supplement Series* **265**(1), 28 (2023)
 16. Ji, A., Wen, J., Angryk, R., Aydin, B.: Solar flare forecasting with deep learning-based time series classifiers. In: 2022 26th International Conference on Pattern Recognition (ICPR). IEEE (Aug 2022)
 17. Kusano, K., Iju, T., Bamba, Y., Inoue, S.: A physics-based method that can predict imminent large solar flares. *Science* **369**(6503), 587–591 (Jul 2020)

18. Lee, K., Moon, Y.J., Lee, J.Y., Lee, K.S., Na, H.: Solar flare occurrence rate and probability in terms of the sunspot classification supplemented with sunspot area and its changes. *Solar Physics* **281**(2), 639–650 (Sep 2012)
19. Li, M., Cui, Y., Luo, B., Wang, J., Wang, X.: Deep neural networks of solar flare forecasting for complex active regions. *Frontiers in Astronomy and Space Sciences* **10** (Jun 2023)
20. Li, X., Zheng, Y., Wang, X., Wang, L.: Predicting solar flares using a novel deep convolutional neural network. *The Astrophysical Journal* **891**(1), 10 (Feb 2020)
21. McIntosh, P.S.: The classification of sunspot groups. *Solar Physics* **125**, 251–267 (1990)
22. Mehta, S., Rastegari, M.: Mobilevit: Light-weight, general-purpose, and mobile-friendly vision transformer. *ArXiv abs/2110.02178* (2021)
23. Nishizuka, N., Sugiura, K., Kubo, Y., Den, M., Ishii, M.: Deep flare net (DeFN) model for solar flare prediction. *The Astrophysical Journal* **858**(2), 113 (May 2018)
24. Pandey, C., Angryk, R.A., Aydin, B.: Solar flare forecasting with deep neural networks using compressed full-disk hmi magnetograms. In: 2021 IEEE International Conference on Big Data (Big Data). pp. 1725–1730 (2021). <https://doi.org/10.1109/BigData52589.2021.9671322>
25. Pandey, C., Angryk, R.A., Aydin, B.: Deep Neural Networks Based Solar Flare Prediction Using Compressed Full-disk Line-of-sight Magnetograms, p. 380–396. Springer International Publishing (2022). https://doi.org/10.1007/978-3-031-04447-2_26
26. Pandey, C., Angryk, R.A., Aydin, B.: Explaining Full-Disk Deep Learning Model for Solar Flare Prediction Using Attribution Methods, p. 72–89. Springer Nature Switzerland (2023). https://doi.org/10.1007/978-3-031-43430-3_5
27. Pandey, C., Angryk, R.A., Aydin, B.: Unveiling the potential of deep learning models for solar flare prediction in near-limb regions. In: 2023 International Conference on Machine Learning and Applications (ICMLA). pp. 703–708 (2023)
28. Pandey, C., Angryk, R.A., Georgoulis, M.K., Aydin, B.: Explainable Deep Learning-Based Solar Flare Prediction with Post Hoc Attention for Operational Forecasting, p. 567–581. Springer Nature Switzerland (2023). https://doi.org/10.1007/978-3-031-45275-8_38
29. Pandey, C., Ji, A., Angryk, R.A., Aydin, B.: Towards interpretable solar flare prediction with attention-based deep neural networks. In: 2023 IEEE Sixth International Conference on Artificial Intelligence and Knowledge Engineering (AIKE). pp. 83–90 (2023). <https://doi.org/10.1109/AIKE59827.2023.00021>
30. Pandey, C., Ji, A., Angryk, R.A., Georgoulis, M.K., Aydin, B.: Towards coupling full-disk and active region-based flare prediction for operational space weather forecasting. *Frontiers in Astronomy and Space Sciences* **9** (Aug 2022). <https://doi.org/10.3389/fspas.2022.897301>
31. Pandey, C., Ji, A., Nandakumar, T., Angryk, R.A., Aydin, B.: Exploring deep learning for full-disk solar flare prediction with empirical insights from guided grad-cam explanations. In: 2023 IEEE 10th International Conference on Data Science and Advanced Analytics (DSAA). pp. 1–10 (2023). <https://doi.org/10.1109/DSAA60987.2023.10302639>
32. Park, E., Moon, Y.J., Shin, S., Yi, K., Lim, D., Lee, H., Shin, G.: Application of the deep convolutional neural network to the forecast of solar flare occurrence using full-disk solar magnetograms. *The Astrophysical Journal* **869**(2), 91 (Dec 2018)

33. Pesnell, W., Thompson, B.J., Chamberlin, P.C.: The solar dynamics observatory (SDO). *Solar Physics* **275**(1-2), 3–15 (Oct 2011)
34. Schou, J., Scherrer, P.H., Bush, R.I., Wachter, R., Couvidat, S., Rabello-Soares, M.C., Bogart, R.S., Hoeksema, J.T., Liu, Y., Duvall, T.L., Akin, D.J., Allard, B.A., Miles, J.W., Rairden, R., Shine, R.A., Tarbell, T.D., Title, A.M., Wolfson, C.J., Elmore, D.F., Norton, A.A., Tomczyk, S.: Design and ground calibration of the helioseismic and magnetic imager (HMI) instrument on the solar dynamics observatory (SDO). *Solar Physics* **275**(1-2), 229–259 (Oct 2011)
35. Schrijver, C.J.: A characteristic magnetic field pattern associated with all major solar flares and its use in flare forecasting. *The Astrophysical Journal* **655**(2), L117–L120 (Jan 2007). <https://doi.org/10.1086/511857>, <http://dx.doi.org/10.1086/511857>
36. Toriumi, S., Wang, H.: Flare-productive active regions. *Living Reviews in Solar Physics* **16**(1) (May 2019)