The development of PIL extraction application in active region of SDO/HMI magnetogram

A Z Utama¹, M Z Nurzaman¹, T Dani¹

¹Research Center for Space, Research Organization for Aeronautics and Space (LAPAN), National Research and Innovation Agency, 133 Djundjunan street, Bandung, Indonesia
E-mail: ahmad.zulfiana@lapan.go.id

Abstract. The Polarity Inversion Line (PIL) on the active region (AR) is a parameter that correlates with the occurrence of the solar flare. This research has developed application to extract the PIL length in the AR. Image processing techniques were used to detect and classify AR from the SDO/HMI Magnetogram image of 1024x1024 pixels. The detection of AR in magnetogram images has its challenges because the ARs are constructed by the white (black) regions represent the magnetic positive (negative) polarities. To improve detection accuracy, the application used The Solar Region Summary (SRS), compiled by SWPC, and the centroid position of sunspot on SDO/HMI Continuum as additional information. The results of AR detection had high accuracy because SRS stores the centroid position of the flaring active region at peak time with the Stonyhurst heliographic on the solar disk which is converted to Cartesian coordinates. The application test used 1450 images and obtained the PIL length on the peak time of the X-class and M-class flares from 2011-2017. The application detected the high distribution of PIL length in the range of 120 to 580 Mm for X class flare, and 120 to 390 Mm for M-class flare. This research provides new data that has the potential to improve the machine learning-based solar flare prediction models.

1. Introduction
Electromagnetic radiation and particle energy caused by increased solar activity can have an impact on technological world today. Particles from energetic events on solar atmosphere likes coronal mass ejection (CME) can also hit crucial electronics on boards a satellite and interfere its system. Electrical and GPS systems are potentially affected by solar activity, and modern society is highly dependent on it [1]. Accordingly, the observation of solar activity is a serious matter. Radiation takes hours or even days to reach the Earth. A reliable early warning system can alert the stack holder to make appropriate and accurate decisions to protect the affected technology infrastructure [2]. Therefore, the worst-case scenarios can be anticipated.

Flare and CME events emit from active region (AR) on the solar surface. An AR is often observed as groups of sunspots that have a higher magnetic complexity compared to other surface areas and at least consist of simple positive and negative polarity. An AR with more complex polarity has larger area, bigger size and a greater number of spots [3]. It can be observed using full disk LOS SDO/HMI Magnetogram image [4]. The image provides information about area of positive/negative polarity and photospheric magnetic field data [5].

Mostly solar flares are related to strong magnetic field gradients along the neutral lines and complex polarity pattern [6]. Haygard et al. (1984) explained that the transverse component of the
magnetic field, known as shear, reach maximum values at the flare onset sites located along the magnetic polarity inversion line (PIL) in the active region [7]. Falconer et al. (2002) also found that flare productivity is highly related with long PIL [8]. Metcalf et al. [1995] and Silva et al. [1996] showed that when magnetic field lines related with solar flares reconnect to a lower energy configuration, the released energy produces coronal mass ejections [9][10]. PIL length of AR is highly an indicative of the potential to open large scale magnetic field and to produce CME and SEP [11]. Active regions for which PIL quickly changes with height into potential field configuration are more favorable for producing non-eruptive events [12]. The PIL is a line that separates the positive and negative polarities of the magnetic flux of the Sun. Research on PIL itself and PIL as training data for machine learning to predict the solar flare has been carried out. Wang et al., (2019) extracted the PIL with a high gradient from the modified HARP data as training input for the prediction model [13]. Sadikov & Kosovichev (2017) has analyzed the relationship between X-ray peak flux and PIL characteristics in the active region. The SVM model was developed to predict the solar flare only based on PIL characteristics. They conclude that PIL characteristics can be used as predictions of solar activity with TSS results of ≥ 0.76 for M1 and ≥ 0.84 for X.10. Sadikov & Kosovichev (2017) claimed the importance of PIL characteristics in solar flare prediction [14]. Wang et.al (2020) used the high gradient of the PIL mask of the ARs data and obtained TSS score 0.79 (24 hr before the flare) and 0.58 (72 hr before the flare) by using a supervised random forest model [15]. Nishizuka et.al (2020) developed Deep Flare Net (DeFN) to predict a solar flare using a deep neural network. DeFN used extracted features from different wavelength images and obtained TSS score of 0.80 for ≥ M-class and ≥ 0.63 for class flares [16]. Zheng (2019) proposed a hybrid Convolutional Neural Network (CNN) model to predict solar occurrence within 24 hr and obtained TSS score with 0.749 ± 0.79 for ≥ M-Class [17].

Cai et at., (2020) has developed a framework for the detection and extraction of PIL features on LOS Magnetogram patch images. Their research has generated a PIL data set to understand the characteristics and structure of PIL [18]. PIL parameters have also been used as input and model for predicting flare activity by Wang et al., (2020). They have developed a prediction model using the KPCA machine learning algorithm to obtain the best two features derived from the PIL Masks. The features are effective for predicting the occurrence of solar flare. Furthermore, a supervised random forest model was developed to classify active areas into non-strong flaring and strong flaring groups. Wang et al., (2019) in another study also stated that the properties of PILs in ARs are strongly correlated to solar flares and CME occurrences [13].

This study focuses on methods to obtain PIL parameters and area of flaring active region from full-disk LOS SDI/HMI magnetogram images. The active region was obtained from the paired magnetic field region according to the position of SRS active region centroid. The SRS provides archives of the flaring active region position. The SRS active region position has been converted into Cartesian coordinates for an image with a size of 1024x1024 pixels. The converted SRS position causes the detection results to reach 100% accuracy. The high accuracy was expected to produce qualified training data for machine learning in future works. This study used M and X-class flare event from 2011-2017 at peak times. The area and PIL length were obtained and extracted from the detected of the flaring active region. The PIL length is converted to Micro Hemisphere (Mm), while the area of flaring active region is converted to Solar Million Hemisphere (SMH).

2. Data and methodology
Pipeline process starts from downloading AR data and magnetogram images, preprocessing phase, grouping and detecting active region. Figure 1 shows the pipeline diagram of developed method purposes. Each process is explained in subsection 2.1 – 2.5.
2.1. Magnetogram and Active Region Data
Collecting full disk magnetogram and intensitygram images were based on data archive of X and M-class flares from 2011-2017 and were downloaded from Joint Science Operations Center (JSOC) server using DRMS Python package. The magnetogram and intensitygram data were produced by Helioseismic and Magnetic Imager (HMI; Schou et al., 2012) on board Solar Dynamics Observatory (SDO; Pesnell, Thompson, and Chamberlin, 2012) and yields continues full-disk observations of components of the photospheric magnetic field. AR detection was performed on full disk SDO HMI/Magnetogram images and used data of AR location from Solar Region Summary (SRS) which provided by National Oceanic and Atmosphere Administration, Space Weather Prediction Center (NOAA/SWPC). Unfortunately, AR location from SRS data is only for 00:30 UT in which almost all X and M class flare taken place at varying time. To handle this issue, we rotated AR location from SRS data using solar differential formula until it reaches peak time of X and M class flare. As a result, we obtained AR location at the X and M class flare’s peak time.

SRS data store centroid location of AR in Stonyhurst Heliographic Coordinate (SHC) format, area, Modified-Zurich classification, magnetic field regional classification, and sunspot number. The Centroid location that used SHC and has been rotated to X and M class flare peak time was converted into Cartesian Coordinates (CC) with 1024x1024 pixel dimensions using the World Coordinate System method. Then, it was used as a reference point to detect the active region position in full-disk magnetogram image.

2.2. Preprocessing Phase
Preprocessing phase was performed before we implemented AR detection and grouping algorithm. This phase aims to obtain contour, centroid, and the bounding box of the contour. In this study, we used OpenCV library in Python as image processing functions. Figure 2 shows stages of the preprocessing phase, starting from reading the image, convert RGB to binary channel, reducing the noise, and obtaining contour, centroid, and bounding box.

Contours can only be obtained from binary images. RGB image was converted to grayscale and binary image with 170 G threshold value for continuum image, 100 G for black area, and 170 G for white area in full disk magnetogram image. The gaussian blur filter was used to reduce the distributed
noise. In addition, small contours were removed based on area values to reduce the computational value during the process of grouping active region.

The findContours() function from the OpenCV library is performed to get contour information for each area. Red dot and Green Lines in figure 3 (c) represent centroid and bounding box for each area. AR grouping used the centroid and the bounding box information.

2.3. Bounding box of SRS AR centroid location data
The converted SRS AR centroid position from SHC to CC doesn’t have bounding box information because the bounding box can only be obtained from the contour. Sunspot allow us to probe magneto convection for strong fields. So, the initialization of the bounding box of the SRS centroid used the bounding box of the detected of sunspot grouping. In this research, we used continuum image to get the bounding box of sunspot group.

The preprocessing pipeline was performed and yield the centroid, and bounding box information in sunspot image. The sunspot grouping uses the minimum distance of the centroid with 20 threshold value. Two area/contours area are combined and produce a new centroid value. This process is repeated and stops when the distance of centroids exceeds the threshold values. The number of iterations were different for each data because it depends by the number of detected contours in preprocessing phase. Figure 4 (a) shows sunspot area that marked by bounding box that drawn in red line and Figure 4 (b) is result of grouping centroid based on the minimum of centroid distance.

The results of grouping based on distance are regrouped according to the intersection of the bounding box. Figure 4 (c) is the final results of sunspot grouping. The initialization of the bounding box of SRS AR centroid adjust to the bounding box of the sunspot group.
2.4. Active Region Grouping

AR grouping uses the bounding box of SRS AR centroid, the centroid and the bounding box of black/white contour. The bounding box of SRS AR centroid is used as a reference to determine the positive (white) and negative (black) polarity regions which are included in an AR. First, the black/white area grouping is performed. The black/white area that intersects with the bounding box of SRS AR centroid is grouped into a region. This is done on both types of black and white areas so this stage produced two new centroid and bounding box for grouping black and white areas. Figure 5 (a) is a black area that intersects with the bounding box of SRS AR centroid and each area is grouped as shown in Figure 5 (b). The yellow line represents the group of the black area, while the green dot is the new centroid value.

The program has already the information of the black/white area group. An AR grouping is grouped by using the minimum distance of the centroid between the SRS AR centroid, the grouped black area, and the grouped white area. Figure 6 depicts each centroid with a different color, where the SRS AR centroid is marked in red, the black area (green), and the white area (blue). An AR is constructed from the grouped black and white area that has a minimum distance with the SRS AR centroid.
Figure 6. (a) Green, red, and blue dot are the centroid of black grouping area, the centroid of AR based on SRS data, and the centroid of white grouping area. (b) the final result of AR grouping.

2.5. Conversion of PIL length and AR area

The contour area of AR is calculated by using the cv.contourArea() function in OpenCV Python, but this function is still in pixel area units. Equation (1) can convert the area into units of the Solar Millionth Hemisphere (SMH) with foreshortening correction (Smart, 1971; Duffet-Smith P, 1988; Meadows, 2020)[19]. Equation (1) is used to calculate the area of sunspots group using a foreshortening correction factor.

\[
A_M = \frac{A_s 10^6}{2\pi R^2(\rho)} \quad (1)
\]

\(A_m\) is area in SMH, and \(\rho\) is the angular distance from the center of the solar disk to the detected area from Equation (2). \(A_s\) is the area measured in pixels\(^2\), and \(R\) is the radius of the solar disk in pixels with \(x\) and \(y\) is the position of centroid group with coordinates \((0,0)\) in the center of the solar disk. The default coordinates \((0,0)\) are always in the upper left corner, but for the calculation of the area, we shift the coordinates \((0,0)\) to the center of the solar disk according to the requirements of Equation (3). \(appdiam\) is the apparent diameter of the Sun (0.53°) and \(diskdiam\) is the diameter of the solar disk in the image (unit is equal to \(x\) and \(y\)).

\[
\rho = \frac{2\rho_1/appdiam}{appdiam/diskdiam} - \rho_1 \quad (2)
\]

\[
\rho_1 = \frac{appdiam}{diskdiam}\sqrt{x^2+y^2} \quad (3)
\]

A contour of area is a collection of interconnected pixels. PIL is obtained by calculate the pixel point distance between black and white contours with five points differences. The PIL length from all the connected pixels is converted by correction factors angle between line of sight and field direction into the centre of solar disk image. We converted the pixel units into Micro Hemisphere (Mm) with equation (4), equation (5), and equation (6).

\[
onepix = \frac{\text{solar}\_\text{radius}}{\sqrt{xrad^2 + yrad^2}} \quad (4)
\]

\[
\theta = \arcsin \frac{r}{\sqrt{xrad^2+yrad^2}} \quad (5)
\]

\[
one\_\text{length} = \frac{\text{one}\_\text{pix}}{\cos\theta} \quad (6)
\]
3. Results

![Figure 7](image-url) (a) The result of the detected AR by our application (b) AR location from www.spaceweatherlive.com

The position of the detected AR is an important task to yield the PIL value. We verified our result by comparing it with AR location from www.spaceweatherlive.com. Figure 7 (b) is archive data on 20120706T2300 that obtained from www.spaceweatherlive.com, while Figure 7 (a) is a result of the detected AR in this research. For example, NOAA 11515 locate at 36° on longitude and -17° on latitude and our application was successfully detected the AR with the exact same position as shown on figure 7 (a). The application marked the position of the flaring active region with accurately because the position has been verified with the SRS.

![Figure 8](image-url) Some results for the detected AR that produce X and M-class flare.

Figure 8 shows some results from our application. The first and the second row are for X-class flare case and the third and the fourth row are for M-class flare case. Our application was successfully detected the flaring AR from full solar-disk SDO magnetogram image with size of 1024x1024 pixel.
The green line on figure 8 represents the polarity inversion line between positive and negative region. The yellow and red line are contour for negative (black) and positive (white) polarity region.

**Figure 9.** Some results for the detected AR on full disk SDO magnetogram image for X and M-Class flare.

Figure 9 shows two different chart type with the same time frame. The bar chart describes the total area, otherwise the line chart depicts the PIL Length. The highest of PIL length was 933 Mm on 27 October 2014 from 12192 NOAA and the total area was 60786 SMH. In contrast, the smallest of the PIL length was 76 MM and the total area was 5047 SMH.

**Figure 10.** The PIL length (MM) and the total area (SMH) on the flaring active region with M-Class flare at the peak time during 2011-2017.

The largest PIL length in M-class case was in the active region of 12087 with the PIL length and total area were 822 Mm and 28137 SMH. In contrast, the smallest PIL was 28 Mm and total area was 3233 SMH. We found in some results that the PIL Length in M-class is greater than the length of X-class. For example, the AR of 12087 emitted M-Class with the PIL length of 822 Mm, but the AR of 11283 emitted
X-Class with the PIL length of 136 Mm. This implies that the PIL length and area cannot be used as a single threshold to distinguish X or M-class flare.

Figure 11 (a) describes the high distribution of the PIL on X-Class with range above 120 and below 580 Mm, while for the area is in the range above 10000 SMH and below 32000 SMH as shown in figure 11 (b). The high distribution for M-class is relative same with X-class that is above 120 MM and below 390 Mm for the PIL length and above 10000 SMH and below 32000 SMH.

4. Conclusion
Polarity Inversion Line (PIL) is an important parameter that relates to the solar flare or CME event. We used full disk SDO/HMI magnetogram and intensitygram images that yield X and M-class flares during 2011-2017 to test developed application and to obtain the PIL length for each case. This research has successfully to create application and pipeline process to extract the PIL length and area of detected AR and using SRS data information to verify the position of AR at the peak time of X and M-class flare during 2011-2017. AR that produced X and M-class flare had vary in the length of PIL and area. The application detected the high distribution of PIL length in the range of 120 to 580 Mm for X-class flare, and 120 to 390 Mm for M-class flare. Our data conclude that it is difficult to distinguish between X and M class based on the threshold value of the PIL length. We hope that the data and statistic
that we had obtained can be useful for further research. The possible contribution for future work is that our developed application will obtain more the PIL length and area data by observing active regions that occur from east-west on full solar disk SDO magnetogram image.

Acknowledgments
Utama, Nurzaman, and Dani contributed as the main contributor to this paper and had primary responsibility for the final content of the manuscript. All authors read and approved the final paper. We thank Dr. Johan Muhamad from Research Center for Space, LAPAN, for valuable discussion especially on all converted the pixel units into Micro Hemisphere (Mm) equations that appear on the manuscript. We also thank the anonymous reviewer who helps evaluate the manuscript and for their useful suggestions.

References
[1] Schrijver C J, Kauristie K, Aylward A D, Denardini C M, Gibson S E, and et al 2015 Understanding space weather to shield society: A global road map for 2015–2025 commissioned by COSPAR and ILWS Advances in Space Research 55 2745-2807
[2] Sreeja V 2016 Impact and mitigation of space weather effects on GNSS receiver performance Geoscience letters 3(1) pp 1-13
[3] Bobra M G, Sun X, Hoeksema J T, Turmon M, Liu K, Hayashi K, Barnes G, and Leka K D 2014 The Helioseismic and Magnetic Imager (HMI) vector magnetic field pipeline: SHARPs–space-weather HMI active region patches Sol. Phys. 289 3549-3578
[4] Mackay D H, Green L M, and Ballegooijen A D 2011 Modeling the dispersal of an active region: quantifying energy input into the corona The Astrophys. J. 729 97
[5] Guglielmino S L et al. 2012 The frontier between small-scale bipoles and ephemeral regions in the solar photosphere: emergence and decay of an intermediate-scale bipole observed with SUNRISE/IMaX The Astrophys. J. 745 160
[6] Zirin H 1983 The 1981 July 26-27 flares: Magnetic developments leading to and following flares The Astrophys. J. 274 900–909
[7] Hagyard M J, Smith J B, Teuber D and West E A 1984 A quantitative study relating observed shear in photospheric magnetic fields to repeated flaring Sol. Phys. 91 115–126
[8] Falconer D A, Moore R L and Gary G A 2002 Correlation of the coronal mass ejection productivity of solar active regions with measures of their global nonpotentiality from vector magnetograms: Baseline results The Astrophys. J. 569 1016–1025
[9] Metcalf T R, Jiao L,McClymont A N, Canfield R C and Uitenbroek H 1995 Is the solar chromospheric magnetic field force-free? The Astrophys. J. 439 474–481
[10] Silva A V R, White S M, Lin R P, de Pater I, Gary D E, McTiernan J M, Hudson H S, Doyle J G, Hagyard M J and Kundu M R 1996 Comprehensive multiwavelength observations of the 1992 January 7 solar flare The Astrophys. J. Suppl. Ser. 106 621–646
[11] Canfield R C, Hudson H S and McKenzie D E 1999 Sigmoidal morphology and eruptive solar activity Geophysical Research Letters 26(6) 627-630
[12] Baumgartner C, Thalmann J K and Veronig A M 2018 On the factors determining the eruptive character of solar flares The Astrophys. J. 853(2) 105
[13] Wang J, Liu S, Ao X, Zhang Y, Wang T and Liu Y 2019 Parameters derived from the SDO/HMI vector magnetic field data: Potential to improve machine-learning-based solar flare prediction models The Astrophys. J. 884(2) 175
[14] Sadykov V M and Kosovichev A G 2017 Relationships between characteristics of the line-of-sight magnetic field and solar flare forecasts The Astrophys. J. 849(2) 148
[15] Wang J, Zhang Y, Webber S A H, Liu S, Meng X and Wang T 2020 Solar flare predictive features derived from polarity inversion line masks in active regions using an unsupervised machine learning algorithm The Astrophys. J. 892(2) 140
[16] Nishizuka N, Kubo Y, Sugiura K, Den M and Ishii M 2020 Reliable Probability Forecast of Solar Flares: Deep Flare Net-Reliable (DeFN-R) The Astrophys. J. 899(2) 150
[17] Zheng J and Zheng L 2019 A hybrid bidirectional recurrent convolutional neural network attention-based model for text classification *IEEE Access* 7 106673-106685
[18] Cai X, Aydin B, Ji A, Georgoulis M K and Angryk R 2020 A Framework For Detecting Polarity Inversion Lines From Longitudinal Magnetograms *IEEE International Conference On Big Data* 4175-4183
[19] Smart W M 1971 *Text-book on Spherical Astronomy* Cambridge Univ. Press Cambridge