An Image Processing approach to identify solar plages observed at 393.37 nm by the Kodaikanal Solar Observatory

Sarvesh Gharat,1,* Bhaskar Bose2, Abhimanyu Borthakur3 and Rakesh Mazumder4
1 Centre for Machine Intelligence and Data Science, Indian Institute of Technology Bombay, 400076, Mumbai, India
2 Smart Mobility Group, Tata Consultancy Services, 560067, Bangalore, India
3 Department of Electronics and Communication Engineering, Manipal Institute of Technology, 576104, Karnataka, India
4 Birla Industrial & Technological Museum, Kolkata, National Council of Science Museums, Ministry of Culture, Govt. of India

Accepted XXX. Received YYY; in original form ZZZ

ABSTRACT
Solar plages, which are bright regions on the Sun’s surface, are an important indicator of solar activity. In this study, we propose an automated algorithm for identifying solar plages in Ca K wavelength solar data obtained from the Kodaikanal Solar Observatory. The algorithm successfully annotates all visually identifiable plages in an image and outputs the corresponding calculated plage index. We perform a time series analysis of the plage index (rolling mean) across multiple solar cycles to test the algorithm’s reliability and robustness. The results show a strong correlation between the calculated plage index and those reported in a previous study. The correlation coefficients obtained for all the solar cycles are higher than 0.90, indicating the reliability of the model. We also suggest that adjusting the hyperparameters appropriately for a specific image using our web-based app can increase the model’s efficiency. The algorithm has been deployed on the Streamlit Community Cloud platform, where users can upload images and customize the hyperparameters for desired results. The input data used in this study is freely available from the KSO data archive, and the code and the generated data are publicly available on our GitHub repository. Our proposed algorithm provides an efficient and reliable method for identifying solar plages, which can aid the study of solar activity and its impact on the Earth’s climate, technology, and space weather.

Key words: Sun: chromosphere – Sun: faculae, plages – techniques: image processing

1 INTRODUCTION
The importance of space weather is becoming more apparent in recent years (Schwenn 2006). The Sun’s magnetic activity is essential in regulating space weather, but we only have continuous records of solar magnetic field observations for the last few solar cycles. To better understand the variation of solar magnetic fields, we need to reconstruct magnetic field proxies for past solar cycles. Fortunately, we have data on solar magnetic features such as sunspots, plages, and filaments for over a century, which we can use as proxies to indirectly comprehend the long-term variation of the Sun’s magnetic field. Ca-K images are particularly useful for illustrating the long-term variability of the chromospheric magnetic field (Chatzistergos et al. 2018; Penza et al. 2021). The Mount Wilson (Foukal et al. 2009) and the Kodaikanal Solar Observatory (Raju & Singh 2014) have century-long records of Ca-K images. Solar Plages are regions of high magnetic field concentration that can help trace the Sun’s magnetic activity (Shine & Linsky 1974; Azariadis & Guesnerie 1986; OLSON et al. 1978; Neidig 1989; Mackay et al. 2008; Canfield et al. 2000; Shine & Linsky 1972). Although different authors may use different definitions for the Ca II K plage index, it is commonly defined as the fractional area of the total plage area to that of the full solar disk area (Bertello et al. 2020). Ca II K plage indices show periodic variation with the solar cycle (Bertello et al. 2020; Chatzistergos et al. 2019, 2020a; Chatterjee et al. 2016). Chatterjee et al. (2016) have shown that the temporal evolution of the latitude distribution of the plages depicts a butterfly pattern, similar to the temporal evolution of the latitude distribution of the sunspots.

Precise measurements of solar irradiance are crucial for studying the climate. The Ca II K observations are highly correlated with the non-sunspot magnetic field, which can provide valuable information on the surface coverage by bright plage and network regions (Babcock & Babcock 1955; Skumanich et al. 1975; Saar & Linsky 1985; Loukitcheva et al. 2009; Pevtsov et al. 2016; Chatzistergos et al. 2016; Kahil et al. 2017). As a result, historical Ca II K spectroheliograms can be used to improve long-term irradiance models significantly, especially with the availability of the digitization of various solar data archives (Chatzistergos et al. 2021; Ribes & Mein 1985; Kariyappa & Pap 1996; Foukal 1996, 1998; Caccin et al. 1998; Worden et al. 1998; Zharkova et al. 2003; Lefebvre et al. 2005; Ermolli et al. 2009a,b; Tlatov et al. 2009; Bertello et al. 2010; Sheeley et al. 2011; Priyal et al. 2014a, 2017; Chatterjee et al. 2016; Chatterjee et al. 2017). Since 1907, the Kodaikanal Solar Observatory has been observing the Sun using photographic plates with a 30 cm objective and f-ratio of f/21 in the Ca K wavelength (Chatterjee et al. 2016) (Hasan et al. 2010). This has resulted in a substantial amount of data over the years, with an effective spatial resolution of around 2 arcsecs for most of the documented time. More recently, the plates have been digitized using a CCD sensor to create 4096 x 4096 raw images with 16-bit resolution (Priyal et al. 2014b).
Manual identification of Solar Plages for century long data is not only time consuming, it also introduces human biasing (Barata et al. 2018) (Benkhalil et al. 2003). In this study, an image processing algorithm is proposed to identify plages from Ca II spectroheliograms obtained from the Kodaikanal Observatory. The data required for this study has been collected from the KSO data archive. Various studies have been conducted to identify different solar features such as sun spots, plages, filaments, etc. (Barata et al. 2018) (Benkhalil et al. 2003) (Aschwanden 2010) (Qahwaji & Colak 2005) (Aboudarham et al. 2008) (Scholl & Habbal 2008), but, there is no exclusive study conducted on the Kodaikanal Solar Observatory. Therefore, this study proposes an image processing algorithm to automate the process of plage identification specifically for the Kodaikanal Solar Observatory.

In their study (Barata et al. 2018), Barata et al. employed morphological transformations to segment Coimbra Observatory spectroheliograms, utilizing dilation, erosion, and the top hat operator before thresholding. Benkhalil et al. (Benkhalil et al. 2003) used thresholding and basic morphological operations similar to (Barata et al. 2018) to identify active regions in Hα and Calcium K images from the Meudon Observatory. However, the variation in images obtained from the KSO data archive across solar cycles 16-22 made it difficult to apply such algorithms to this study. Aschwanden et al. (Aschwanden 2010) provided a comprehensive review of image processing for identifying different solar features and time dependency, including the use of neural networks for feature identification, but their study required a large amount of labeled data and is therefore unsuitable for this approach. Qahwaji and Colak (Qahwaji & Colak 2005) focused on identifying plages, with a proposed algorithm using morphological classification and hole filling on Meudon Observatory data to differentiate between plages and filaments. Similarly, in (Aboudarham et al. 2008) and (Scholl & Habbal 2008), the authors used various morphological transformations to identify multiple solar features. However, due to variations in image properties such as brightness, contrast, and artificial artifacts, a novel algorithm using the OpenCV (Bradski 2000) library is proposed for identifying plages in the Kodaikanal Solar Observatory data archive. The algorithm was tested by analyzing the time series variation of the plage index across solar cycles 16-22 and compared with the plage index reported by (Chatzistergos et al. 2020b) to confirm its efficiency.

The KSO data archive contains Ca-K data up to October 2007 (Chatterjee et al. 2016), which is categorized into three levels. The "Level 0" comprise raw images requiring preprocessing before analysis (Pal et al. 2020), whereas "Level 1" images have undergone basic calibrations, such as bias correction, disc centering, and rotation correction with non-uniform radii (Pal et al. 2020). Neither "Level 0" nor "Level 1" images account for the center-to-limb variation (CLV) (Pötzi et al. 2022), which is necessary to eliminate large-scale intensity variations produced by telescope optics and achieve uniform intensity and contrast. This is only addressed in the "Level 2" images (Pal et al. 2020), which, along with disc centering and uniform radii, are the most suitable for analysis. Consequently, for this study, we limit our investigation to the "Level 2" images.

3 METHODOLOGY
The initial step involves loading the solar images as grayscale images using OpenCV (Bradski 2000), after which the algorithm proceeds with the following steps:

3.1 CLAHE
To enhance the contrast of the images, a technique called Contrast Limited Adaptive Histogram Equalization (CLAHE) was utilized (Reza 2004). This method is an extension of the commonly used histogram equalization technique (Pizer et al. 1987), which has been shown to improve image contrast (Abdullah-Al-Wadud et al. 2007). However, global histogram equalization techniques are not suitable for our study due to inconsistent pixel intensities and varying brightness levels in the images. Therefore, CLAHE was chosen as it is based on the local intensity statistics. To prevent over-amplification of noise, we used a clipping factor (Sundaram et al. 2011) (Liu et al. 2019) as one of the parameters that require experimental tuning to achieve the best results. After experimentation, a clipping factor of 2 was found to produce the most satisfactory outcomes.

3.2 Image Thresholding
Image thresholding plays a crucial role in detecting plages by segmenting the image based on a predetermined threshold value (Chowdhury & Little 1995). In our research, a constant threshold value of 180 was employed for the majority of cases, which proved to be effective due to the utilization of CLAHE (Reza 2004). However, it is worth noting that in some instances, a threshold value as low as 100, or as high as 220 may yield superior results. This conclusion was reached after careful analysis of all the archived images. As a result, our web application provides users with a variable threshold slider that enables them to select the most appropriate threshold value for a specific image.

3.3 Opening
Opening (Chen & Haralick 1995) is a morphological image processing technique that involves the use of a structuring element (Song & Delp 1990) to eliminate small objects or noise from an input image. The structuring element is a small binary image that determines the shape and size of the objects to be removed. The opening operation works by sliding the structuring element over the input image and performing a logical AND operation between the structuring element and the corresponding pixels in the input image (Chen & Haralick 1995).
If all the pixels in the structuring element are white (i.e., have a value of 1) and the corresponding pixels in the input image are white, then the output pixel value is set to 1. If not, the output pixel value is black (i.e., has a value of 0). In our study, we used a 3x3 elliptical kernel (Landström & Thurley 2013; Tian & Yang 2016) as the structuring element, which was chosen after experimenting with several other kernels.

### 3.4 Erosion and Dilation

Erosion and dilation are basic image processing transformations commonly used to enhance and remove small contrasting spots in images (Soille 2004). The erosion operation removes regions that are smaller than the size of the structuring element, while dilation joins two regions where the distance between them is less than or equal to the size of the structuring element. The structuring element used in this step is identical to that of section 3.3 - a 3x3 elliptical kernel. In our algorithm, erosion and dilation are important for reducing the noise in the images. In this study, we apply two iterations of erosion and one iteration of dilation, after the opening operation is performed on the thresholded image.

### 3.5 Area Filtering

The image obtained in section 3.4 using pre-built functions available in OpenCV (Bradski 2000) and determining the area of the solar disc. With the area of the solar disc, the radius, center, heliographic latitude, longitude, polar angle, and semi-diameter are calculated according to the methodology outlined in (Hiremath et al. 2019). Using the calculated semi-diameter and solar disc radius, the pixel size and area are computed using the formula given in (Pucha et al. 2016)

Next, the contours are extracted from the segmented image, and the area corresponding to each contour (which either corresponds to a plage or noise incorrectly identified as a plage), is calculated using the steps described in (Hiremath et al. 2019). Contours that are too small or too large are very likely to be noisy components, and are hence, discarded using a lower and an upper threshold, respectively. Only contours with an area in between the two thresholds are retained in the output image. Although this method of hard thresholding can lead to the elimination of some plages, the thresholds have been determined through experimentation on multiple images. Moreover, users can also set their own threshold values based on their prior knowledge of the solar cycle to avoid losing any plages.

In summary, the proposed algorithm is comprised of several steps. Firstly, the input image is read in grayscale and enhanced with Contrast Limited Adaptive Histogram Equalization (CLAHE) to improve contrast. The algorithm allows the user to customize the clipping factor for CLAHE. Then, the image is thresholded using a user-defined threshold value. An elliptical structuring element of size 3x3 is initialized and used for opening, followed by two iterations of erosion and one iteration of dilation. The resulting image is used to extract contours using cv2.findContours, and the area of each contour is calculated while accounting for the projectional effects of the sun’s spherical shape. The algorithm retains only the contours with an area within the user-defined lower and upper area thresholds and discards others. These retained contours are used to calculate the plage index and are overlaid on the input image to highlight the plages using cv2.drawContours. Finally, the output of the algorithm is an image highlighting the plages and the calculated plage index. In the event that the user fails to provide appropriate hyperparameters, the algorithm defaults to the parameters listed in Table 1.

### 4 RESULTS AND DISCUSSION

We used the automated algorithm for identifying solar plages on the Ca K wavelength solar data obtained from the Kodaikanal Solar Observatory, for multiple solar cycles ranging from Cycle 16 to Cycle 22, to confirm the validity of our approach. The number of images utilized for testing the algorithm in each cycle is listed in Table 2.

The proposed algorithm can only process images from the Kodaikanal Observatory, which should be in .jpg format. To make it more convenient for users, the algorithm has been deployed on the Streamlit Community Cloud platform (https://abhimanyu911-plages-identification-app-yf7yqp.streamlit.app/), where users can upload images and customize the hyperparameters for desired results. The users can experiment with various parameters such as clip limit (ranging from 1 to 10), lower area threshold (ranging from 0 to 500), upper area threshold (ranging from 0 to 10000), and binary threshold (ranging from 0 to 255) and see the corresponding effect on the input image in real-time. This feature allows users to visualize and understand how different hyperparameters impact the output. The algorithm not only overlays the plage contour on the input image but also provides the calculated plage index, which can be compared with the results obtained in (Chatzistegros et al. 2020b)(if available).

The input image in Figure 1 is an example of the proposed algorithm’s performance. The algorithm successfully annotates all visually identifiable plages in the image and also outputs the corresponding calculated plage index of 0.2733. To test the algorithm’s reliability and robustness, we perform a time series analysis of the plage index (rolling mean) across multiple solar cycles. Additionally, we calculate the correlation coefficient, which measures the strength of a linear relationship between two variables, and the $R^2$ score, which indicates the proportion of the variation in the dependent variable (calculated plage index) that can be predicted from the independent variable (plage index from (Chatzistegros et al. 2020b)).

The results obtained for the Solar Cycle 22 are presented in Figure...
5 CONCLUSION

In conclusion, we present an automated algorithm for identifying solar plages in Ca K wavelength solar data obtained from the Kodaikanal Solar Observatory. The proposed algorithm is able to successfully annotate visually identifiable plages and calculate the corresponding plage index, which indicates the level of magnetic activity present in the Sun. To test the reliability and robustness of our algorithm, we performed a time series analysis of the calculated plage index (rolling mean) across multiple solar cycles. The results showed a high correlation between the calculated values and those reported in (Chatzistergos et al. 2020b), as indicated by the high values of the correlation coefficient and $R^2$ scores.

Furthermore, to make the algorithm more accessible to users, we have deployed it on the Streamlit Community Cloud platform, where users can upload images and customize the hyperparameters for desired results. This feature allows users to experiment with different hyperparameters and visualize their impact on the output image in real time. We believe that our proposed algorithm has the potential to improve the accuracy and efficiency of identifying solar plages, which is essential for studying the Sun’s magnetic activity and its effects on space weather. Furthermore, the availability of the input data and the code used in this study at public repositories ensures the reproducibility and transparency of our results.

ACKNOWLEDGEMENTS

We want to thank the Kodaikanal solar observatory, a facility of the Indian Institute of Astrophysics, Bangalore, India for making the archival data available. This data is available for public use at http://kso.iiap.res.in through a service developed at IUCAA under the Data Driven Initiatives project funded by the National Knowledge Network.

The authors would like to express their gratitude to the anonymous referees for their valuable suggestions, which have significantly improved the quality of this study.

DATA AVAILABILITY

The data used in this study can be obtained for free from the Kodaikanal Solar Observatory (KSO) data archive website (https://kso.iiap.res.in/data). In addition, we have made our code and the resulting data publicly accessible on our GitHub repository (https://github.com/SarveshVGharat/Plages-Identification).

REFERENCES

Abdullah-Al-Wadud M., Kabir M. H., Dewan M. A. A., Chae O., 2007, IEEE Transactions on Consumer Electronics, 53, 593
Aboudarham J., Scholl I., Fuller N., Fouesneau M., Galametz M., Gonon F., Maire A., Leroy Y., 2008, in Annales Geophysicae. pp 243–248
Aschwanden M. J., 2010, Solar Physics, 262, 235
Azariadis C., Guesnerie R., 1986, The review of economic studies, 53, 725
Babcock H. W., Babcock H. D., 1955, ApJ, 121, 349
Barata T., Carvalho S., Dorotovič I., Pinheiro F. J., Garcia A., Fernandes J., Lourenço A. M., 2018, Astronomy and Computing, 24, 70
Benkhalil A., Zharkova V., Ipson S., Zharkov S., 2003, in Proceedings of the AISB. pp 66–73
Bertello L., Ulrich R. K., Boyden J. E., 2010, Sol. Phys., 264, 31
Bertello L., Pevtsov A. A., Ulrich R. K., 2020, ApJ, 897, 181
Bradska G., 2000, Dr. Dobb’s Journal of Software Tools
