Using Unsupervised Machine Learning to Explore New Classification of Sunspot Active Regions

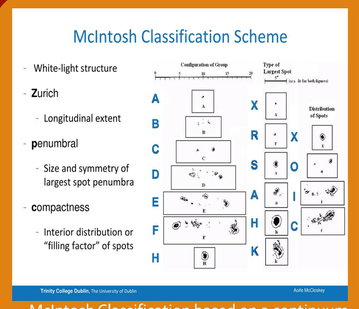
Sara Housseal[1], Tom Berger [2], Varad Deshmukh [2, 3]

[1] Millersville University, [2] CU Boulder SWx-TREC, [3] CU Boulder Computer Science

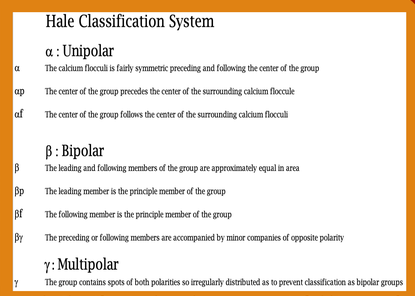
Introduction:

For as long as active regions have been getting classified, it has been done manually by humans, by studying their features and labeling them according to the McIntosh and/or Hale classification systems. There has also been previous work done to automate McIntosh and Hale classification using supervised machine learning but it's time to see if computers can perform unsupervised machine learning to classify active regions in a new quantitative way. We have performed Principle Component Analysis (PCA) on the Solar Dynamics Observatory (SDO) Helioseismic and Magnetic Imager (HMI) Space-weather HMI Active Region Patch (SHARP) images of all sunspot active regions from 2010 to the present. We performed a PCA on 100,000 randomly selected active region images to create 40 PCA "eigenvector images" of active regions. The 40 eigenvector images show that the PCA picked up on the complex magnetic multipole of sunspot active regions and can be narrowed down to 16 eigenvector images to allow compact classification.

Background Info:



McIntosh Classification based on a continuum image of an active region. [2]



Hale Classification based on a magnetic field image of an active region.

TABLE IV

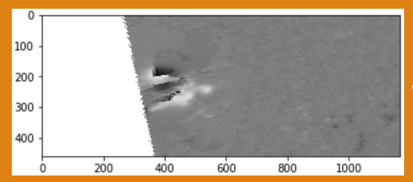
Most flare-productive sunspot types 1969–1976 (from Kildahl, 1980)

McIntosh type	Number of occurrences	Number of class M flares	Number of class X flares	Flares/occurrence per 24 hours	
				M	X
Hsx	1963	99	6	0.05	0.003
Dso	553	51	6	0.09	0.01
Dai	324	58	7	0.18	0.02
Dkc	100	72	10	0.72	0.10
Eki	81	103	11	1.27	0.14
Ekc	63	149	21	2.36	0.33
Fki	47	106	17	2.26	0.36
Fkc	27	39	13	1.44	0.48

Figure above shows that currently, solar flares are being predicted based on the McIntosh classification given to a particular active region. Depending on the classification, a probability of flaring is given to the corresponding active region. [1]

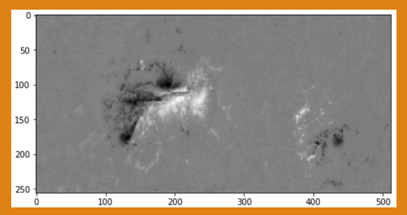
Preparation:

All images were properly formatted in preparation for the PCA, giving us a total of 405,035 Br images to work with.



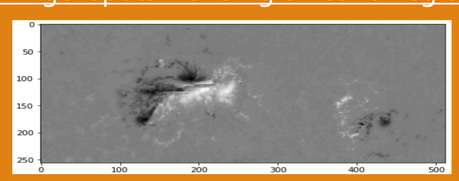
Only images within 60 to -60 degrees longitude were used to avoid issues with the PCA due to distortion to active regions along the limb or the region off the sun where there is no data.

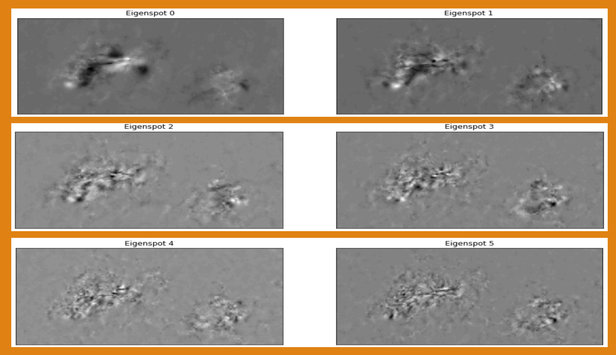
All images were transformed to a size of 512 by 256 pixels to eliminate possible PCA issues with different sized images. The structure of the active regions was preserved using Affine Transformation.



Single AR Results:

PCA :Eigenspots" for Single Active Region

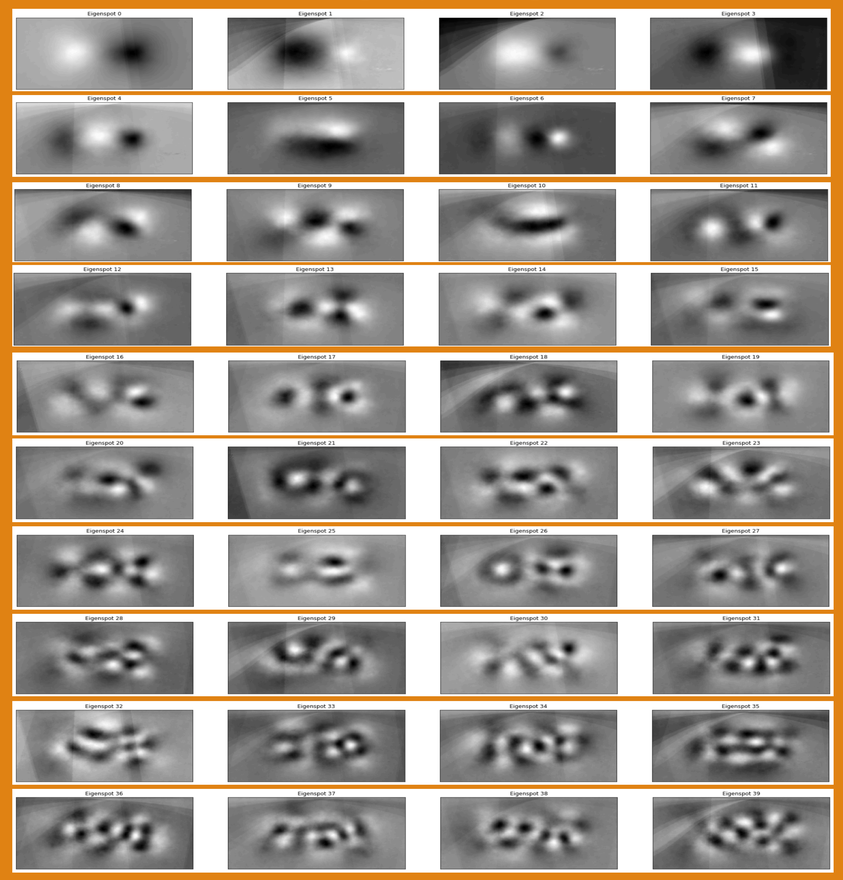




The PCA was tested using 158 images of a single active region and results show that the PCA was easily able to pick up on features as well as small scale details of a single active region as not much change occurred.

Large Random Batch Results:

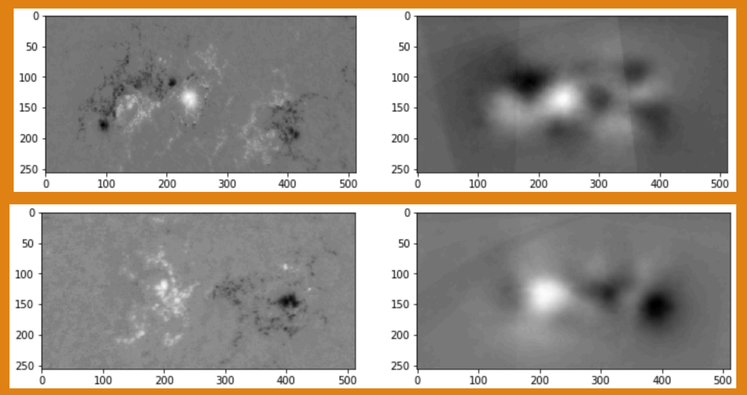
40 PCA "Eigenspot" Images



It is shown that results become less detailed for 25,000 randomly selected images, but the PCA is still able to pick up on key structures of the active regions such as the varying polarities. "Eigenspots" progressively become more complex as more are generated.

Reconstruction:

The overall structure of active regions are successfully reconstructed using the 40 "eigenspots" in different weights.



Currently, small scale details are not achieved due to the "eigenspots" lacking those details, but as shown above, the different polarities and surrounding regions are detected and reconstructed. Work will be continued in an effort to run the entire data set and produce more "eigenspots" to see if that features more small scale details within the "eigenspots".

New Classification:

New classification system will be based on different components of the active regions consisting of:

- Eigenvector Image Weights
- Evolution of the Active Region
- Maximum Strength of the Magnetic Field within the Polarity
- Rate of Evolution of the Active Region Properties

Conclusions:

- Using Unsupervised Machine Learning, the computer is able to pickup on the basic components and patterns of different sunspot active regions.
- The PCA is then able to be used to create a reconstruction of different sunspot active regions using different weights of the generated "eigenspots".
- Batched/Incremental PCA will be attempted in an effort to run the PCA on the entire data set of 405,000 images.
- These results can lead to the possibility for a new classification system that is based on "eigenspot" weight, evolution, and magnetic field strength.
- Work will be continued towards creating the new classification system.

Acknowledgements:

- This research was supported by the National Science Foundation REU program, Award #1659878
- 1, McIntosh, P. S. (1989, August 21). *The Classification of Sunspot Groups.* Retrieved from http://adsabs.harvard.edu/full/1990SoPh..125..251M
- 2, McCloskey, A. *McIntosh Classification Scheme* [Digital image]. Retrieved from https://slideplayer.com/slide/12722702/
- All FITS/PNG images are from JSOC: http://hmi.stanford.edu